THE FORMAL THEORY OF PRICING AND INVESTMENT FOR ELECTRICITY

by

ANDRZEJ JERZY WROBEL

A Thesis Submitted to the

DEPARTMENT OF ECONOMICS

In Fulfillment of the Requirements For the Degree of

DOCTOR OF PHILOSOPHY

IN THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

THE UNIVERSITY OF LONDON

2006

THESES
F
8576

10/9/95

ABSTRACT

The Thesis develops the framework of competitive equilibrium in infinite-dimensional commodity and price spaces, and applies it to the problems of electricity pricing and investment in the generating system. Alternative choices of the spaces are discussed for two different approaches to the price singularities that occur with pointed output peaks.

Thermal generation costs are studied first, by using the mathematical methods of convex calculus and majorisation theory, a.k.a. rearrangement theory. Next, the thermal technology, pumped storage and hydroelectric generation are studied by duality methods of linear and convex programming. These are applied to the problems of operation and valuation of plants, and of river flows. For storage and hydro plants, both problems are approached by shadow-pricing the energy stock, and when the given electricity price is a continuous function of time, the plants' capacities, and in the case of hydro also the river flows, are shown to have definite and separate marginal values. These are used to determine the optimum investment.

A short-run approach to long-run equilibrium is then developed for pricing a differentiated good such as electricity. As one tool, the Wong-Viner Envelope Theorem is extended to the case of convex but nondifferentiable costs by using the short-run profit function and the profit-imputed values of the fixed inputs, and by using the subdifferential as a multi-valued, generalised derivative. The theorem applies readily to purely thermal electricity generation. But in general the short-run approach builds on solutions to the primal-dual pair of plant operation and valuation problems, and it is this framework that is applied to the case of electricity generated by thermal, hydro and pumped-storage plants. This gives, as part of the long-run equilibrium solution, a sound method of valuing the fixed assets—in this case, the river flows and the sites suitable for reservoirs.

Table of Contents

TABLE OF CONTENTS—*Continued*

TABLE OF CONTENTS—*Continued*

List of Figures

LIST OF FIGURES—*Continued*

LIST OF FIGURES—*Continued*

List of Tables

There is also a List of Notation for Chapter 5 (on pages 140–148).

PREFACE AND ACKNOWLEDGMENTS

Early work on this Thesis was supported by The LSE 1980s Fund Graduate Studentship, and before this by the Studienstiftung des deutschen Volkes (in the framework of the European Doctoral Programme in Quantitative Economics). I am grateful to Martin Hellwig for persuading me to go to London when it felt safer to do a *Doktorarbeit* at Bonn. It turned out to be the right decision because it gave me a chance to work with Tony Horsley, who not only supervised me but also backed me to the hilt and was on my side when times got rough and friends just couldn't be found.

This work develops and re-jigs some ideas of Boiteux and Koopmans, as well as a few of my own. But it is permeated even more by Horsley's ideas and, what is even more important, by his way of thinking about scientific problems. His fundamental conviction, grounded in his training and research in elementary particle physics, is that new mathematical frameworks can offer an opportunity for theories of greater verisimilitude with new insights and results. I could not agree more. Rigour is, of course, *de rigeur* these days, but it becomes *rigor mortis* if all it serves is a formal extension of existing knowledge. I hope that this Thesis has, if only in a small way, helped to vindicate Tony's stance, which, with its elements of Schumpeterian destruction, is rarely one that is popular with the academic establishment.

In this Thesis, it is perhaps the chapters on valuation of storage plants that best illustrate the potential of the new framework: a continuous-time analysis yields a result that cannot even be formulated, or guessed at, in discrete time. If I have managed to find a small gem in my searches, this is it. It has been set twice, once in the context of pumped storage and once in the context of hydro in Chapters 3 and 4, and its miniature version comes as a summary of the results on plant operation and valuation in Chapter 5. This produces some overlap and, after the study of pumped storage, that of hydro may well give an impression of *déjà lu*, but the two are never exactly the same. The parallel presentations of the two storage problems, and their subsequent summary, mean also that each chapter can be read independently. Indeed, it is possible to get a good idea of the whole work from Chapter 5 alone.

All the formal paragraphs (Theorems, Definitions, etc.) are numbered in the same sequence, which is reset in each Chapter C and Section S: thus Theorem C.S.P may be followed by Corollary C.S.P+1. Section numbering includes the chapter's number (Section C.S). Figures and Tables are numbered in two independent sequences, which are reset in each chapter, but not in a section: thus Figure C.F and Table C.T. There are also unnumbered Comments, some of which are gathered in bullet lists.

A. J. Wrobel London, England, November 2005

Chapter 1

Introduction

Since Boiteux's seminal article [12], a large amount of research has been done on the theory of peak-load pricing but, with the exception of a landmark study of hydro-thermal electricity generation by Koopmans [55], the significant body of theory is set up in the framework of discretised time and surplus maximisation.[1] The best and most comprehensive account of what has been achieved in that approach is to be found in the book by Crew and Kleindorfer [17], which is essentially up-to-date despite having been completed in 1979. Their analysis assumes constant returns to scale (c.r.t.s.) in electricity generation. Transmission, which has increasing returns to scale, is left out; its share of the total costs of supply is, however, relatively small.[2]

In this Thesis, returns to scale are also taken to be constant or decreasing, but this is used to put the problem in a different framework—that of general competitive equilibrium with continuous time. The analysis is also extended to include energy storage by giving the first realistic model of pumped storage and by recasting the hydro operation problem as one of profit maximisation. This is a setting that is relevant for a modern decentralised electricity supply industry, and it also allows a much simpler solution than does Koopmans's problem of hydro-thermal cost minimisation. A successful analysis of operation and plant valuation for the three technologies (thermal, pumped storage and hydro, in Chapters 2 to 4) means that the long-run equilibrium problem can be approached by building on the short-run solutions (Chapter 5). This is much easier than a direct long-run analysis, and it is also a practical approach in view of the importance of the short-run solution.

The model is deterministic, and so the equilibrium price is a pure time-of-use tariff (TOU tariff). A basic extension to the case of stochastic demand (assuming risk neutrality and symmetric information) requires little more than a re-interpretation of the time variable as time-and-event. This produces a weather-dependent tariff, e.g., an electricity price dependent on the current temperature as well as time.

The use of continuous time turns out to facilitate the treatment of the plant operation and

[1] In [79, 8.D.b], the problems of maximising the social surplus and the monopoly profit are set up in continuous time. Continuous time is also used for an econometric cost-benefit appraisal of peak-load pricing in [28].

[2] Although Crew and Kleindorfer mention the issue of i.r.t.s. in public utility pricing, they assume c.r.t.s. for the mathematical analysis [17, pp. 26, 37, 69, 171], except in brief passages on pricing subject to a profit constraint [17, pp. 16–17 and 59–60]. Their case study of electricity comprises only the thermal generation techniques [17, Chapter 10, p. 159], which have c.r.t.s. In hydro generation (which they exclude from their model of the technology), returns to scale are decreasing once account has been taken of the fixity of the river flows and the sites suitable for reservoirs (at the very least, the marginal cost of expanding these inputs is steeply increasing).

valuation problems. It also gives verisimilitude in describing commodity differentiation in physical flows of goods. As for the surplus concept with its well-known shortcomings, it is made unnecessary by the equilibrium framework: in the first-best theory, it can serve no purpose other than deriving the marginal-cost pricing principle as in [17, (2.8)–(2.9)], but the principle is of course a part of the competitive equilibrium concept.

The basic results that verify the model's consistency have been published in three papers [43], [45] and [47].[3] Referred to a number of times (starting with Section 2.2, in Chapter 2), these results establish: (i) the continuity of the demand map and the existence of a competitive equilibrium [47], (ii) the representation of the equilibrium price system by a density function, i.e., by a time-dependent price rate in $/kWh [43], and (iii) the continuity of the equilibrium price as a function of time [45]. The demand continuity result of [47] is a basis not only for an equilibrium existence proof but also for the sensitivity analysis that is necessary in any implementation of the equilibrium solution: it is essential to know that small deviations from the equilibrium price system will not result in large shifts of demand. It should also be noted that the density representation of prices in [43] is a result that adapts Bewley's framework to continuous-time pricing problems by relaxing his so-called "Exclusion Assumption" on the production sets [10, p. 524 and Theorem 3]. Applied to electricity pricing, the price-density result settles Boiteux's conjecture on the shifting-peak problem: his spread-out form of capacity charges obtains if brief interruptions of consumption would cause little loss to the users. In mathematical terms, this is the case when the consumers' utility functions and the industrial users' production functions are Mackey continuous. This is assumed in the Thesis when it comes to the general equilibrium analysis of electricity supply (Section 5.15, in Chapter 5). Under additional assumptions, the price density is a continuous function of commodity characteristics such as time [45]. A continuously varying time-of-use tariff (TOU tariff) has two uses in electricity pricing. First, it precludes demand jumps that would arise from discontinuous switches from one price rate to another. Second, in the problems of operating and valuing hydroelectric and pumped-storage plants (studied in Chapters 3 and Chapter 4), price continuity guarantees that their capacities (viz., the reservoir and the converter), the energy stocks, and in the case of hydro also the river flows, have well-defined marginal values.

The equilibrium existence and price continuity results of [43], [45] and [47] are set in the commodity space $L^{\infty}[0,T]$, which consists of all the essentially bounded functions on the time interval that represents one pricing cycle. This is the largest commodity space that can be used for cyclical continuous-time problems involving capacity costs or constraints. There are advantages,

[3] The three papers are not part of the Thesis, but are on its list of references as [43], [45] and [47], published in *Economic Theory* (vols 20 and 26) and *Banach Center Publications* (vol. 71). These papers were written on an equal basis with Anthony Horsley on grant R000232822 from the Economic and Social Research Council for the project "Applications of modern equilibrium theory".

spelt out later, to be had from using the smaller commodity space of continuous functions, $\mathcal{C}[0,T]$. But $L^\infty[0,T]$ is mathematically the more convenient setting for capacity pricing with interruptible demand, since it contains the 0-1 indicator function $1_{[0,T]\setminus E}$ that describes the users' switch-off response to a capacity charge concentrated on a set of small measure, E. Furthermore, being a rearrangement-invariant space, $L^\infty[0,T]$ is also an appropriate setting for formulating the weak symmetry-like conditions that underlie the price continuity result of [45].

In this Thesis, the commodity space is therefore taken to be $L^\infty[0,T]$, the space of essentially bounded functions on the time interval that represents one pricing cycle, and it is paired with $L^1[0,T]$, the price space of integrable functions (with the space of continuous functions, $\mathcal{C}[0,T]$, as a price subspace). This means that price singularities, which represent concentrated charges, are excluded in general equilibrium (on the assumption of interruptible demand). They are, however, included in the discussion of price systems and marginal costs for the three supply technologies (Chapters 2 to 4). This is because, although both Bewley's and later work on the L^∞-model has been preoccupied with excluding price singularities, these do actually have an essential role in continuous-time problems as capacity charges concentrated on pointed peaks (as opposed to capacity charges spread out as a density over a peak plateau). When singular prices *are* an essential term of the equilibrium price system, the task is not to exclude them but to give them a tractable mathematical representation. This cannot be done within the L^∞-model, but it can be achieved by restricting the commodity space to the space of continuous functions, $\mathcal{C}[0,T]$. Then an instantaneous capacity charge on a point peak takes the form of a Dirac measure; it is a charge in \$ per kW demanded at the peak instant, and it is additional to a price density (which is a price rate in \$/kWh). A price of this form can arise in equilibrium when some of the demand is uninterruptible, i.e., when the user's utility or production function is norm-continuous but not Mackey continuous: see [39] and Section 2.2 (in Chapter 2).[4]

In the context of storage and hydro plant valuation, continuity of the electricity price function turns out to guarantee that all the fixed inputs—viz., the river flows, the reservoirs and the generating equipment—have unique, fully definite marginal values. That is, the short-run profit is a differentiable function of these inputs (Section 3.9, in Chapter 3, and Section 4.9, in Chapter 4). In other words, the infinite-dimensional linear programme of plant valuation has a unique solution.

But other optimal values, such as the short-run and long-run joint costs of thermal electricity generation are manifestly nondifferentiable as functions of the output (as well as of the other arguments). Therefore, the marginal costs are formalised by using the subdifferential as a generalised, multi-valued derivative. An effective extension of the Wong-Viner Envelope Theorem for nondifferentiable costs is provided as one of the tools for the short-run approach (Section 5.9, in Chapter 5).

[4] The norm in question is the supremum norm on $\mathcal{C}[0,T]$, which, as a subspace of $L^\infty[0,T]$, carries also the Mackey topology (for the pairing of L^∞ with L^1).

It uses, in an essential way, the short-run profit function and the profit-imputed values of the fixed inputs.

Thus it is shown that infinite-dimensional equilibrium analysis can be much enhanced and made more applicable by exploiting the usually under-used mathematical resources of linear and convex programming and the subdifferential calculus. This is the wider programme of this work.

In summary, Chapter 2 sets up a continuous-time model of the thermal technology of electricity generation using the commodity space of essentially bounded functions. Since there are advantages to using the smaller commodity space of continuous functions, feasibility of this choice is also discussed. Explicit formulae are given for the optimal solution, values and marginal values of the problems of short-run or long-run cost minimisation and short-run profit maximisation (Section 2.3). Because the joint costs are convex but nondifferentiable, marginal costs are formalised as subdifferentials. As a function of the cyclical trajectory of output, the short-run cost is a convex integral functional. By applying rearrangement and majorisation theory, the long-run cost (net of the peak term) is shown to be a basic symmetric function, and its subdifferential is calculated by using the Hardy-Littlewood-Pólya Inequality. The peak term of the long-run cost is the supremum functional, and the known formula for its subdifferential is quoted. So is the formula for the cost-minimising generating system. Given a TOU electricity tariff, the short-run profit-maximising output and the capacity value are also spelt out, in Section 2.4, for use in the short-run approach to market equilibrium (in Chapter 5). The role of profit-imputed capacity values in extending the Wong-Viner Envelope Theorem to nondifferentiable costs is also described.

In Chapter 3, the duality method of linear and convex programming is applied to the problems of operation and rental valuation of facilities for conversion and storage of electricity (when it is priced by time of use). Both problems are approached by shadow-pricing the energy stock (which is a purely intermediate commodity), and the marginal values of the plant's capacities are expressed in terms of the stock's shadow price function ψ and the given TOU market price p for electricity (Section 3.9). In particular, the unit reservoir rent equals the total positive variation of ψ over the cycle. If p is a continuous function of time, then the short-run profit is shown to be a differentiable function of the capacities, which therefore have definite and separate marginal values, despite being perfect Allen-Hicks complements. (In the case of perfect conversion, ψ itself is unique if p is continuous.) The optimal storage policy is also given in terms of ψ and p (Section 3.8). The marginal capacity values are used to determine the optimum investment in storage plants (Section 3.11). Finally, the conditions which guarantee that the storage technology can be included in a continuous-time competitive equilibrium model of peak-load pricing with the space of essentially bounded functions as the commodity space are verified (Section 3.16).

Chapter 4 gives a parallel analysis for hydroelectric operation: the duality method of linear programming is applied to the problems of operation and rental valuation of a hydro plant (when electricity is priced by time of use in the cycle). Both problems are approached using time-dependent shadow-pricing of water; and if the given market price for electricity, p, is a continuous function of time, then the shadow price function for water, ψ, is shown to be unique (Section 4.9). The marginal values of the plant's capacities—defined as derivatives of the short-run profit—are expressed in terms of ψ (and p). In particular, the unit reservoir rent equals the total positive variation of ψ over the cycle. The profit-imputed values of the river flow and of the hydro capacities (reservoir and turbine) are therefore definite—unlike the corresponding values imputed from fuel savings for a mixed hydro-thermal system (as in the Koopmans's work). The optimal water storage policy is also given in terms of ψ and p (Section 4.8). The marginal capacity values are used to determine the optimum investment in hydro plants (Section 4.12).

In Chapter 5, the preceding studies of operation and valuation of the three plant types are first summarised and then used in applying the short-run approach to the long-run equilibrium pricing of electricity (Sections 5.13 and 5.14, and Section 5.15). Before its application, the short-run approach itself has to be developed. This is done in Sections 5.3 to 5.9 for an individual producer's optimum, and in Sections 5.10 and 5.11 for the general equilibrium model with a focus on the market supplied by a particular industry—such as the electricity supply industry.

Thus Chapter 5 gives a new formal framework for the theory of competitive equilibrium and its applications. The "short-run approach" means the calculation of long-run producer optima and general equilibria from the short-run solutions to the producer's profit maximisation programme and its dual. The marginal interpretation of the dual solution means that it can be used to value the capital and other fixed inputs, whose levels are then adjusted accordingly (where possible). But short-run profit can be a nondifferentiable function of the fixed quantities, and the short-run cost is nondifferentiable whenever there is a rigid capacity constraint. Nondifferentiability of the optimal value requires the introduction of nonsmooth calculus into equilibrium analysis, and subdifferential generalisations of smooth-calculus results of microeconomics are given, including the key Wong-Viner Envelope Theorem. This resolves long-standing discrepancies between "textbook theory" and industrial experience. The other tool employed to characterise long-run producer optima is a primal-dual pair of programmes. Both marginalist and programming characterisations of producer optima are given in a taxonomy of seventeen equivalent systems of conditions (with six systems spelt out in full detail). When the technology is described by production sets, the most useful system for the short-run approach is that using the short-run profit programme and its dual. This programme pair is employed to set up a formal framework for long-run general-equilibrium pricing of a range of commodities with joint costs of production. This gives a practical method that finds the short-run general equilibrium en route to the long-run equilibrium, exploiting the operating policies and plant valuations that must be determined anyway. These critical short-run solutions have relatively simple forms that can greatly ease the fixed-point problem of solving for equilibrium, as is shown on an electricity pricing example. Applicable criteria are given for the existence of the short-run solutions and for the absence of a duality gap. The general analysis is spelt out for technologies with conditionally fixed coefficients, a concept extending that of the fixed-coefficients production function to the case of multiple outputs. The short-run approach is applied to the peak-load pricing of electricity generated by thermal, hydro and pumped-storage plants. This gives, for the first time, a sound method of valuing the fixed assets—in this case, river flows and the sites suitable for reservoirs.

Chapter 2

COST MINIMISATION AND PROFIT MAXIMISATION FOR THERMAL ELECTRICITY GENERATION IN CONTINUOUS TIME

2.1 Introduction to Chapter 2

This is a formal account of the multi-station constant-coefficients model of thermal electricity generation technology in continuous time. Two sets of problems, set in the framework of infinite-dimensional commodity and price spaces, are discussed: (i) calculating the short-run and long-run costs and marginal costs, and determining the optimal generating system (i.e., the system that minimises the total production cost), and (ii) running the plants to maximise their operating profits and valuing the plants on the basis of these profits.

The cost calculations are of particular interest to a centralised public utility that aims to meet demand at a minimum operating cost, optimise its capital stock, and price its outputs at long-run marginal cost (LRMC). Through these policies the utility can achieve a long-run competitive equilibrium outcome. And even if the utility does not actually price its output at marginal cost, it is still interested in minimising its cost and in knowing its marginal costs.

The thermal technology of electricity generation is so simple that, except for the fixed-point problem of market clearance, all the problems of the cost approach have explicit solutions. These are presented in Section 2.3. Cost-minimising operation of a thermal system consists in switching the stations on in the order of increasing unit operating cost, a.k.a. the merit order. The system's SRC is additively separable over time and so the SRMC is described by the usual SRMC curve in the instantaneous quantity-price plane (Figure 2.1 and Theorem 2.3.1). The cost-minimising plant mix can be determined from the break-even points on the load-duration curve. This comes along with the calculation of the LRC for an output trajectory of a special form, viz., a unit output of an arbitrary duration per cycle (Figure 2.2), as in, e.g., [9], [60], [64] and [83]. This argument is extended by following Horsley [33, (3)] to give an integral formula for the LRC of any output trajectory by slicing the area between it and the time axis, in parallel to that axis, into infinitesimal outputs of varying durations. This is then used, as in [33, (5) and (6)], to derive a formula for the LRC that can be explicitly differentiated to calculate the LRMC; the heuristic calculation in [33, (6)] is fully formalised in [36, Section 5]. Here, this analysis is presented in Theorems 2.3.3, 2.3.6 and 2.3.7 and their variants (Theorems 2.3.9 and 2.3.10).

For mathematical rigour, these calculations require subdifferential calculus, since both the SRC and the LRC functions are nondifferentiable at the points of most interest. A continuous-time analysis requires, furthermore, a suitable function space as the commodity space for electricity outputs. Because of the capacity costs and constraints, this must be a space of bounded functions and, for the main part, the largest of such spaces—viz., the space $L^{\infty}[0,T]$ of all the essentially bounded functions on the time interval that represents one pricing cycle—is used here. An alternative formulation, which is also presented, uses the smaller commodity space of continuous functions, $\mathcal{C}[0,T]$. This choice is not without its difficulties in equilibrium analysis, but it allows both a more realistic representation of electricity flows and a better representation of singular charges such as the capacity charges concentrated on pointed peaks. This was first recognised in [33, (6)]. The variants of marginal cost calculations with $\mathcal{C}[0,T]$ as the commodity space are given in Theorems 2.3.8, 2.3.9 and 2.3.10.

The short-run profit (SRP) calculations are of interest to producers in a decentralised electricity supply industry, but they can also be useful to a centralised public utility as a basis for another way of arriving at the long-run competitive equilibrium, with two advantages. Unlike the policy of LRMC pricing, this method consists in calculating the short-run equilibrium and then adjusting the capacities until their profit-imputed unit values are equal to their prices. This is the short-run approach which is developed in Chapter 5 and in [46]. As is also pointed out there, the short-run approach to equilibrium can, in principle, be based on calculating the SRC instead of the SRP, but the cost approach is problematic when there are different kinds of plant with dissimilar technologies. This is because the generating system's minimum operating cost is, as a function of output, the infimal convolution of the individual plants' operating costs—which means that cost-minimising operation requires splitting the system's output optimally among the plants. This is known as optimal system despatch. Despatch of a purely thermal system is obvious from the plants' merit order—see (2.3.5)—but the problem becomes difficult when another plant type, such as hydro or pumped storage, is added. For a hydro-thermal system, its cost-minimising despatch problem is determined by Koopmans [55]; his solution is unavoidably very complicated. The profit approach by-passes the despatch problem because profit maximisation by plants using a *common* output price system results "automatically" in a cost-minimising allocation of the total output among the plants. The total output is, of course, determined in the process instead of being given. This means that the profit approach is of little use to a utility intent on minimising its costs but not on competitive equilibrium pricing. Such a utility has no choice but to tackle the formidable problem of cost-minimising despatch.

By contrast, the SRP calculations for hydro and pumped storage plants are relatively simple, as is shown in Chapters 3 and 4. And for a thermal station they are essentially trivial: given a time-of-use electricity tariff p, the profit-maximising output rate for a thermal plant of capacity k with a running cost w is k at a time t when $p(t) > w$ and 0 when $p(t) < w$ (when $p(t) = w$, it is anything between 0 and k). The operating profit per unit capacity is therefore $\int_0^T (p(t) - w)^+ \, dt$

(where the "+" means the nonnegative part). This is spelt out in Section 2.4, for use in the short-run approach to equilibrium with a generating technology containing thermal, hydro and pumped storage techniques (Chapter 5 and [46]).

2.2 The commodity and price spaces for electricity

Cyclical continuous-time problems that involve capacity costs or constraints, such as peak-load pricing of electricity, must be set up in a commodity space which consists entirely of bounded functions of time. An obvious choice is the space of continuous functions $\mathcal{C}[0,T]$—and an immediate modelling benefit is that its norm-dual, the space of all Borel measures $\mathcal{M}[0,T]$, is available as the price space. This can accommodate the instantaneous capacity charges that arise in the case of firm point peaks. However, as is well known, $\mathcal{C}[0,T]$ is not a dual Banach space, and equilibrium analysis with $\mathcal{C}[0,T]$ as the commodity space is hampered by the consequent lack of a vector topology that would make the unit ball compact. One way to get round this mathematical difficulty is to use the larger commodity space of essentially bounded functions, $L^{\infty}[0,T]$. Unlike $\mathcal{C}[0,T]$, the space $L^{\infty}[0,T]$ does have a norm-predual, which is $L^{1}[0,T]$, the space of Lebesgue integrable functions. Bewley [10] uses this first to give an equilibrium existence result with a price system $p^{\star}$ in the norm-dual $L^{\infty\star}$, and then to deduce the existence of an equilibrium price system in the subspace L^{1} under additional assumptions. This is done by showing that any singular part of $p^{\star}$ can be deleted without disturbing the equilibrium; hence the remaining density part, which belongs to L^{1}, is itself an equilibrium price. Some important cases allow the stronger conclusion that $p^{\star}$, the original price system, is itself a pure density function. As long as L^{∞} is the commodity space, the price density result is an indispensable part of the analysis because the singularities in $L^{\infty\star}$ are mathematically intractable and therefore unsuitable for describing prices. But, since the L^{1}-price functions obviously cannot represent the instantaneous capacity charges, some equilibria cannot be adequately described within the L^{∞}-model; this is its basic limitation.

As is shown in [43], Bewley's model can be adapted to peak-load pricing if the users' preferences and production functions are Mackey continuous. This assumption means that demand for the good in question is harmlessly *interruptible*, i.e., that brief interruptions of consumption flows cause only small losses to the users. In this case, the customers would switch off briefly rather than pay any concentrated or instantaneous charges. So, being ineffective, such charges cannot be part of an equilibrium tariff. This is the continuous-time interpretation of Bewley's argument that singular prices would "make an arbitrarily small set of commodities extraordinarily expensive", so that consumers "would prefer to trade them for cheaper ones" [10, p. 523]. In the context of peak-load pricing, it can be seen as Boiteux's solution to the "shifting-peak problem" [12, 3.4 and 3.3.3]. A concentration of the capacity charge during a short-lived peak can cause the peak to shift, but the incentive to shift demand may be removed by spreading the capacity charge over a flattened peak. In this type of equilibrium, the capacity charge is spread as a density over a peak plateau in the output trajectory. This is the only type of equilibrium that the price space $L^1[0,T]$ can accommodate.

Since Bewley's type of equilibrium obtains only under the restrictive assumption of interruptible demand, it is by no means always valid. With *uninterruptible* demand, it is a salient feature of the peak-load pricing problem that the demand trajectory can have a firm, pointed peak. In such a case, the peak capacity charge is levied wholly at the peak instant; it is then a charge for the *rate* of consumption at that instant, and *not* a charge per unit of the good. In the context of electricity pricing, this is a capacity charge in \$ per kW demanded at the peak instant—and it is additional to the marginal fuel charge, which is a price density, i.e., a price rate in \$/kWh. In other words, as is pointed out by Horsley [33, (6)], there is a charge per unit of *power* taken at peak, as well a charge per unit of *energy* at any time. Such a price system can be represented by the sum of a point measure and a measure with a density (with respect to the Lebesgue measure), but this requires restricting the commodity space to $\mathcal{C}[0,T]$ and pairing it with the price space $\mathcal{M}[0,T]$. This can be done from the start if time-continuity is taken to be a physical constraint on consumption bundles, as is assumed in [39]. If this is *not* assumed, then the commodity space has to be $L^\infty[0,T]$ to start with, and it is restricted to $\mathcal{C}[0,T]$ only in the end, after the equilibrium allocation has been shown to lie in $\mathcal{C}$.[1] The equilibrium price functional $p \in L^{\infty *}$ can then be restricted to $\mathcal{C}$, and its restriction can be represented by a measure $p_{\mathcal{C}} \in \mathcal{M}$. Like any measure on $[0,T]$, $p_{\mathcal{C}}$ is the sum of an absolutely continuous measure (which is identified with its density) and a measure that is singular (with respect to the Lebesgue measure). The two parts of $p_{\mathcal{C}}$ need not always correspond to the density and the singular parts of p as an element of $L^{\infty *}$, but they do in peak-load pricing when the equilibrium output has a peak of zero duration: this is spelt out at the end of this section.

Therefore, although a singular linear functional on L^∞ is unsuitable as a *final* mathematical representation of a price system, it is useful as a working representation of prices for two very different purposes: either to show that the equilibrium price system contains no singularity (i.e., lies in L^1), or to represent the price singularity in another way (as a measure) if it does arise in general equilibrium. For both reasons, singular terms are included in the formal discussion of marginal costs and prices.

The commodity space $L^\infty[0,T]$ is the vector space of all essentially bounded real-valued functions; a function y: $[0,T] \to \mathbb{R}$ is *essentially bounded*, with respect to the Lebesgue measure (w.r.t. meas), if y is bounded on $[0,T] \setminus N$ for some set N with $\operatorname{meas} N = 0$. Functions which are equal almost everywhere (a.e.)—i.e., differ only on a set of measure zero—are identified with each other. The

[1] This can be done by extending the argument of [45, Section 6], albeit under rather restrictive assumptions (in particular, those users capable of consuming discontinuous bundles must have additively separable utility or production functions).

space L^∞ is normed by the supremum norm

$$\|y\|_\infty := \operatorname{ess}\sup_{t\in[0,T]} |y(t)| := \inf_{N:\,\mathrm{meas}\,N=0}\ \sup_{t\in[0,T]\setminus N} |y(t)|\,.$$

The notation for the essential supremum of y is abbreviated to $\mathrm{EssSup}\,(y)$.

For a complete mathematical description of marginal costs and prices, the commodity space $L^\infty[0,T]$ must be paired with its *norm-dual* $L^{\infty *}$. This is larger than the norm-predual of L^∞, which is $L^1[0,T]$, the space of functions integrable w.r.t. meas. A function $p \in L^1[0,T]$ represents a time-of-use (TOU) electricity tariff that is a *price density*, i.e., a *time-dependent price rate* $p(t)$ in \$/kWh. But not all price systems have this form and, what is more, there are output trajectories with no marginal costs of the pure-density form: see Theorems 2.3.1 and 2.3.7. When $L^\infty[0,T]$ serves as the commodity space for a cyclically priced good such as electricity, a *general TOU tariff* is represented by a $p \in L^{\infty *}[0,T]$, and such a p can be identified with a finitely additive set function vanishing on the meas-null sets: the integral of a $y \in L^\infty$ w.r.t. such a set function defines a bounded linear functional on L^∞. See, e.g., [25, III.1–III.2 and IV.8.16] or [86, 2.3]. As an additive set function, a $p \in L^{\infty *}$ has the Hewitt-Yosida decomposition into $p_{\mathrm{CA}} + p_{\mathrm{FA}}$, the sum of its *countably additive* (c.a.) and *purely finitely additive* (p.f.a.) parts: see, e.g., [10, Appendix I: (26)–(27)], [25, III.7.8] or [86, 1.23 and 1.24].[2] The c.a. part of p is identified with its density w.r.t. meas, which exists by the Radon-Nikodym Theorem [25, III.10.2]—so it is a price function $p_{\mathrm{CA}} \in L^1[0,T]$. The p.f.a. part of p can be characterised as a singular element of $L^{\infty *}[0,T]$, i.e., p_{FA} is concentrated on a subset of $[0,T]$ with an arbitrarily small Lebesgue measure. Formally, a $p \in L^{\infty *}[0,T]$ is *concentrated on*, or *supported by*, a measurable set $A \subset [0,T]$ if $\langle p \,|\, y\rangle = \langle p \,|\, y 1_A\rangle$ for every $y \in L^\infty$, where 1_A denotes the 0-1 indicator of a set A (i.e., the function equal to 1 on A and to 0 outside A). A sequence of sets (A_j) is *evanescent* if $A_{j+1} \subseteq A_j$ for every j and $\mathrm{meas}\left(\bigcap_{j=1}^\infty A_j\right) = 0$, and p is called *singular* if there exists an evanescent sequence (A_j) such that p is concentrated on A_j for each j. A $p \in L^{\infty *}$ is singular if and only if it is p.f.a.: see [86, 3.1]. This gives p_{FA} the interpretation of an *extremely concentrated charge*.

However, a singular element of $L^{\infty *}[0,T]$ cannot be a fully satisfactory representation of a capacity charge on a pointed peak. This cannot be achieved with this price space: a linear functional representing such a charge should be concentrated on the exact output peaks, but if this is a set of Lebesgue measure zero, then it cannot support any nonzero functional on L^∞. With this commodity space, a capacity charge on a peak of zero duration can be formalised only as a singular element of $L^{\infty *}$ that is concentrated arbitrarily close to the peak, i.e., on the ϵ-near-peaks for every $\epsilon > 0$. Concentration on the exact output peaks cannot follow from this, and indeed it does not because ϵ cannot be set equal to 0 for lack of countable additivity (of the p.f.a. set function that defines the singular functional). As a result, this representation of a pointed-peak charge is not only awkward but also

[2] A p.f.a. set function is one that is lattice-disjoint from every c.a. one.

always nonunique: the EssSup functional on L^{∞} has multiple subgradients at *every* $y \in L^{\infty}$—unlike the Max functional on $\mathcal{C}$, which, at every $y \in \mathcal{C}$ with a single peak at t, has a unique subgradient and thus a Gateaux gradient (viz., the unit point measure at t): see Corollary 2.3.11. With $\mathcal{C}$ as the commodity space, nonuniqueness of the peak charges in the LRMC comes—as it should in a "clean" model—only from the nonuniqueness of the peak instants.

Since the integral w.r.t. a p.f.a. set function is one that lacks some basic properties [86], the symbol $\int$ is reserved here for integration w.r.t. measures, which are countably additive by definition (so the only measures in $L^{\infty*}$ are those having densities). The Hewitt-Yosida decomposition of the value of a flow $y \in L^{\infty}[0,T]$ at a TOU tariff $p \in L^{\infty*}[0,T]$ is therefore written as

$$\langle p \mid y \rangle = \int_0^T p_{\mathrm{CA}}(t)\, y(t)\, \mathrm{d}t + \langle p_{\mathrm{FA}} \mid y \rangle . \tag{2.2.1}$$

In summary, for all its shortcomings, the $L^{\infty*}$-representation of a concentrated capacity charge is useful as a means to one of two ends: either to exclude a price singularity (in the case of interruptible demand with a flattened peak and a pure density as the equilibrium price), or to re-represent it as a measure on $[0,T]$ by restricting the equilibrium price functional $p \in L^{\infty*}$ to the commodity subspace $\mathcal{C}[0,T]$. The measure $p_{\mathcal{C}}$ that represents this restriction has an absolutely continuous part and a singular part. For an *arbitrary* $p \in L^{\infty*}$, the two parts of $p_{\mathcal{C}}$ need not correspond to p_{CA} and p_{FA} because the restriction to $\mathcal{C}$ of a singular (p.f.a.) element of $L^{\infty*}$ can be an absolutely continuous measure, rather than a singular one [82]. But if p_{FA} is supported by an evanescent sequence of *closed* sets (A_j), then its restriction to $\mathcal{C}$ is represented by a singular measure, i.e., a measure concentrated on a set of Lebesgue measure zero (viz., $\bigcap_{j=1}^{\infty} A_j$). This measure is then the singular part of $p_{\mathcal{C}}$, and p_{CA} is the density of the absolutely continuous part of $p_{\mathcal{C}}$. This is the case when, for every constant $\epsilon > 0$, p_{FA} is concentrated on the set of ϵ-near-peaks $\{t : y(t) \geq \mathrm{Max}(y) - \epsilon\}$ of a *continuous* function y with an *exact* peak of zero duration, i.e., with $\mathrm{meas}\{t : y(t) = \mathrm{Max}(y)\} = 0$. The restriction of p_{FA} to $\mathcal{C}$ is then a singular measure concentrated on the latter set, i.e., on the exact peaks of y.

2.3 Thermal technology of electricity generation and its marginal costs

A *thermal technique* generates an output flow $y \in L^{\infty}_{+}[0,T]$ from two input quantities: k (in kW) of generating capacity, and v (in kWh) of fuel of the matching kind. Its long-run production set is the convex cone

$$\mathbb{Y}_{\mathrm{Th}} := \left\{ (y; -k, -v) \in L^{\infty}_{+} \times \mathbb{R}^2_{-} : y \leq k,\ \frac{1}{\eta} \int_0^T y(t)\, \mathrm{d}t \leq v,\ y \geq 0 \right\} \tag{2.3.1}$$

where the constant η is the *efficiency of energy conversion* (the ratio of electricity output to heat input); any startup or shutdown costs and delays are ignored for simplicity. The *unit fuel cost* w (in \$ per kWh of electricity output) is the fuel's price (in \$ per kWh of heat input) times the *heat rate*

$1/\eta$. Henceforth, it is taken to represent all of the unit running cost (a.k.a. operating or variable cost).[3]

There is a number of thermal techniques $\theta = 1, 2, \ldots, \Theta$. Each has the same structure (2.3.1), but it uses its own input commodities, viz., the capacity of type θ and the suitable type of fuel, ξ_θ.[4] Its production set, $\mathbb{Y}_\theta$, is formally $\mathbb{Y}_{\mathrm{Th}}$ embedded in the full commodity space by inserting zeros in the input-output bundle at all the positions other than θ, ξ_θ and the t's. The relevant quantities and prices are indicated with the subscript θ: technique θ generates an output flow y_θ from an input k_θ of generating capacity of type θ and from an input v_θ of fuel of type ξ_θ. Its unit fuel cost is its heat rate $1/\eta_\theta$ times its fuel's price $\widetilde{w}_{\xi_\theta}$. From here on, the *unit fuel cost* of plant type θ is denoted by

$$w_\theta := \widetilde{w}_{\xi_\theta}/\eta_\theta.$$

Without loss of generality, one can assume that the thermal techniques are numbered in the order of increasing unit operating cost (a.k.a. the merit order), i.e., that

$$w_1 < w_2 < \ldots < w_\Theta. \tag{2.3.2}$$

The instantaneous short-run cost per unit time (in \$/h) of generating an output rate y (in kW) from a *single* thermal plant of capacity k with a running cost w is $w\mathrm{y}$ if $0 \le \mathrm{y} \le k$, and $+\infty$ otherwise. The *one-station short-run cost* of a periodic output $y \in L^\infty[0,T]$ is therefore $w\int_0^T y(t)\,\mathrm{d}t$ if $0 \le y(t) \le k$ for a.e. t, and $+\infty$ otherwise. Thus the SRC function represents the capacity and nonnegativity constraints as well as the variable cost actually incurred.

A *thermal system's short-run cost*, i.e., the SRC of producing a total output $y_{\mathrm{To}} \in L^\infty[0,T]$ from a thermal system of plants with capacities

$$k = (k_1, \ldots, k_\Theta)$$

and unit running costs

$$w = (w_1, \ldots, w_\Theta)$$

is the convex integral functional

$$C_{\mathrm{SR}}(y_{\mathrm{To}}, k, w) = \int_0^T c_{\mathrm{SR}}(y_{\mathrm{To}}(t), k, w)\,\mathrm{d}t \tag{2.3.3}$$

where $c_{\mathrm{SR}}(\mathrm{y}_{\mathrm{To}}, k, w)$ is the *system's instantaneous short-run cost* per unit time when its total output rate is y_{To}. As a function of y_{To}, it is the infimal convolution of the individual plants' short-run costs, i.e.,

$$c_{\mathrm{SR}}(\mathrm{y}_{\mathrm{To}}, k, w) = \inf_{(\mathrm{y}_1,\ldots,\mathrm{y}_\Theta)} \left\{ \sum_{\theta=1}^{\Theta} w_\theta \mathrm{y}_\theta : 0 \le \mathrm{y}_\theta \le k_\theta \text{ for each } \theta,\ \mathrm{y}_{\mathrm{To}} = \sum_{\theta=1}^{\Theta} \mathrm{y}_\theta \right\} \tag{2.3.4}$$

[3] The other components of unit running cost (extra maintenance, etc.) can be accounted for by a levy on fuel.

[4] For example, if θ is a nuclear station then ξ_θ is uranium (whose quantity is measured as its energy content).

for $\mathrm{y_{To}} \in \left[0, \sum_{\theta=1}^{\Theta} k_\theta\right]$. Under (2.3.2), the infimum is attained at $\mathrm{y}_1 = \mathrm{y_{To}} \wedge k_1 := \min\{\mathrm{y_{To}}, k_1\}$, $\mathrm{y}_2 = (\mathrm{y_{To}} - k_1)^+ \wedge k_2$, etc., i.e., at

$$\mathrm{y}_\theta = k_\theta \wedge \left(\mathrm{y_{To}} - \sum_{\omega=1}^{\theta-1} k_\omega\right)^+ := \min\left\{k_\theta, \left(\mathrm{y_{To}} - \sum_{\omega=1}^{\theta-1} k_\omega\right)^+\right\} \quad \text{for } \theta = 1, \ldots, \Theta \tag{2.3.5}$$

where a^+ denotes the nonnegative part of a number a; the sequence of capacity cumulatives $\left(\sum_{\omega=1}^{\theta} k_\omega\right)_{\theta=0}^{\Theta}$ starts from $\sum_{\omega=1}^{0} k_\omega = 0$. A counterpart formula to (2.3.4) gives the system's SRC for the cycle as the infimal convolution of the individual plants' costs, i.e.,

$$C_{\mathrm{SR}}(y_{\mathrm{To}}, k, w) = \inf_{(y_1,\ldots,y_\Theta)} \left\{\sum_{\theta=1}^{\Theta} w_\theta \int_0^T y_\theta(t)\,\mathrm{d}t : \forall\theta\ k_\theta \geq y_\theta \geq 0,\ y_{\mathrm{To}} = \sum_{\theta=1}^{\Theta} y_\theta\right\}. \tag{2.3.6}$$

Notation The notation for the system's total output, y_{To}, is abbreviated to y—except when this might conflict with the presumption that $y = (y_1, y_2, \ldots)$, i.e., when the individual plants' outputs (y_θ) are also explicitly discussed, as in (2.3.6).

In terms of k and w, and with 1_A denoting the 0-1 indicator of a set A, the instantaneous SRC of an output rate $\mathrm{y} \in \mathbb{R}$ can be given as

$$\begin{aligned} c_{\mathrm{SR}}(\mathrm{y}, k, w) &:= \int_0^{\mathrm{y}} \sum_{\theta=1}^{\Theta} w_\theta 1_{\left[\sum_{\omega=1}^{\theta-1} k_\omega, \sum_{\omega=1}^{\theta} k_\omega\right]}(\mathrm{q})\,\mathrm{dq} \\ &= w_1 \mathrm{y} + \sum_{\theta=1}^{\Theta-1} (w_{\theta+1} - w_\theta)\left(\mathrm{y} - \sum_{\omega=1}^{\theta} k_\omega\right)^+ \end{aligned} \tag{2.3.7}$$

if $0 \leq \mathrm{y} \leq \sum_{\theta=1}^{\Theta} k_\theta$; otherwise $c_{\mathrm{SR}}(\mathrm{y}, k, w) = +\infty$. This is an increasing and convex (and piecewise linear) function of the scalar $\mathrm{y} \in \left[0, \sum_{\theta=1}^{\Theta} k_\theta\right]$, with $c_{\mathrm{SR}}(0) = 0$. See Figure 2.1a, which shows the case of a two-station system ($\Theta = 2$ with $w_1 < w_2$).

The integrand in (2.3.7), viz.,

$$\mathrm{y} \mapsto \sum_{\theta=1}^{\Theta} w_\theta 1_{\left[\sum_{\omega=1}^{\theta-1} k_\omega, \sum_{\omega=1}^{\theta} k_\omega\right]}(\mathrm{y}) \tag{2.3.8}$$

is the *system's marginal variable cost* as a function of the output rate y, i.e., this is the unit variable cost of the system's marginal station on line when the system load is y. Its graph is also known as the *capacity-incremental operating cost curve*: see, e.g., [9, Figure 5(a)]. With the jumps "filled in", it becomes, in the terminology of [70, 24.3], a complete nondecreasing curve: more precisely, it is the right-angled broken line consisting of (i) the Θ "horizontal" segments

$$\left[\sum_{\omega=1}^{\theta-1} k_\omega, \sum_{\omega=1}^{\theta} k_\omega\right] \times \{w_\theta\} \quad \text{for } \theta = 1, \ldots, \Theta$$

(with $k_0 := 0$) and (ii) the $\Theta + 1$ "vertical" segments

$$\left\{\sum_{\omega=1}^{\theta} k_\omega\right\} \times [w_\theta, w_{\theta+1}] \quad \text{for } \theta = 0, 1, \ldots, \Theta$$

(with $w_{\Theta+1} := +\infty$, and with $w_0 := -\infty$ since free disposal is not included). Known as the *short-run marginal cost curve* (SRMC curve) in the instantaneous quantity-price plane, it is the graph of the subdifferential correspondence

$$\mathrm{y} \mapsto \partial c_{\mathrm{SR}}(\mathrm{y}) = \left[\frac{\mathrm{d}c_{\mathrm{SR}}}{\mathrm{d}_{-}\mathrm{y}}, \frac{\mathrm{d}c_{\mathrm{SR}}}{\mathrm{d}_{+}\mathrm{y}}\right]. \tag{2.3.9}$$

The *left derivative* $\mathrm{d}c/\mathrm{d}_{-}\mathrm{y}$ and the *right derivative* $\mathrm{d}c/\mathrm{d}_{+}\mathrm{y}$ exist at each $\mathrm{y} \in \left[0, \sum_{\theta=1}^{\Theta} k_\theta\right]$ but they differ at $\mathrm{y} = \sum_{\omega=1}^{\theta-1} k_\omega$, where $\mathrm{d}c/\mathrm{d}_{-}\mathrm{y} \le w_{\theta-1}$ (with equality if $k_{\theta-1} > 0$) and $\mathrm{d}c/\mathrm{d}_{+}\mathrm{y} \ge w_\theta$ (with equality if $k_\theta > 0$), for each $\theta = 1, \ldots, \Theta$. In other words, when $k \gg 0$ (i.e., $k_\theta > 0$ for each $\theta = 1, \ldots, \Theta$),[5]

$$\partial_{\mathrm{y}} c_{\mathrm{SR}}(\mathrm{y}, k, w) = \begin{cases} (-\infty, w_1] & \text{if } \mathrm{y} = 0 \\ \{w_\theta\} & \text{if } \mathrm{y} \in \left(\sum_{\omega=1}^{\theta-1} k_\omega, \sum_{\omega=1}^{\theta} k_\omega\right) \\ [w_\theta, w_{\theta+1}] & \text{if } \mathrm{y} = \sum_{\omega=1}^{\theta} k_\omega \text{ and } 1 \le \theta \le \Theta - 1 \\ [w_\Theta, +\infty) & \text{if } \mathrm{y} = \sum_{\theta=1}^{\Theta} k_\theta \\ \emptyset & \text{if } \mathrm{y} > \sum_{\theta=1}^{\Theta} k_\theta \text{ or } \mathrm{y} < 0 \end{cases} . \tag{2.3.10}$$

Figure 2.1b shows the case of two station types (i.e., $\Theta = 2$).

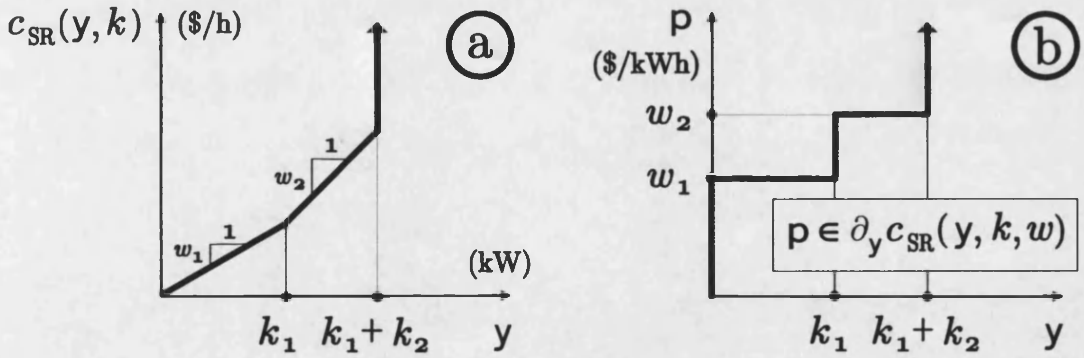

FIGURE 2.1. Thermal short-run costs: (a) instantaneous SRC as a function of the output rate y, for a two-station generating system $k = (k_1, k_2)$ with unit running costs $w = (w_1, w_2)$, and (b) the system's SRMC curve.

For the reasons given in Section 2.2, the following descriptions of the thermal SRMC and LRMC are not limited to densities and include singular charges. Offpeak, the SRMC, as a function of time t in the cycle, is simply a trajectory of the system's marginal variable cost, i.e., it is the function

[5] Obvious changes are needed in (2.3.10) when k is not strictly positive.

(2.3.8) evaluated at the time-varying output rate $\mathrm{y} = y(t)$. In addition to this, the density part of the SRMC includes a peak charge if the maximum output reaches the system's capacity and has a positive duration: the density part it is a trajectory of the instantaneous SRMC over time, i.e., it is an integrable selection from $\partial_y c_{\mathrm{SR}}$ evaluated at $\mathrm{y} = y(t)$. Since $\partial_y c_{\mathrm{SR}}$ is unbounded from above at $\mathrm{y} = \sum_{\theta=1}^{\Theta} k_\theta$, the density of the capacity charge can be arbitrarily high. The singular part of the SRMC is also a capacity charge with an indefinite total. (The capacity charge is determined in the short-run equilibrium of supply and demand, but it cannot be determined just by SRC calculations.) At $\mathrm{y} = 0$, negative charges are, in principle, possible because the output is constrained to be nonnegative, but these will not appear in general equilibrium with monotone consumer preferences or monotone production functions of industrial users. Formalised next, this description of the SRMC is a case of subdifferentiating a convex integral functional.

Theorem 2.3.1 (Thermal SRMC). *Assume that* $0 \le y \le \sum_{\theta=1}^{\Theta} k_\theta$ *and* $k > 0$ *(i.e.,* $k \ge 0$ *and* $k \ne 0$*, and so* $\sum_{\theta=1}^{\Theta} k_\theta > 0$*). Then* $p \in \partial_y C_{\mathrm{SR}}(y, k, w)$ *if and only if* $p \in L^{\infty *}[0, T]$ *and:*

1. $p_{\mathrm{CA}}(t) \in \partial_y c_{\mathrm{SR}}(y(t), k, w)$ *for almost every* $t \in [0, T]$.
2. $p_{\mathrm{FA}} = \kappa - \nu$ *for some* κ *and* ν *in* $L_+^{\infty *}[0, T]$ *such that, for every number* $\epsilon > 0$, κ *is concentrated on* $\{t : y(t) \ge \sum k - \epsilon\}$ *and* ν *is concentrated on* $\{t : y(t) \le \epsilon\}$.[6]

Proof. Apply the formula given in, e.g., [72, Corollary 1B] and [50, Section 4: Theorem 1]. □

Comment (short-run capacity charges): In Theorem 2.3.1, the density part of the short-run capacity charge is $(p_{\mathrm{CA}} - \max_\theta \{w_\theta : k_\theta > 0\})^+$ and the singular part is $p_{\mathrm{FA}}^+ = \kappa$. Both vanish if $\mathrm{EssSup}(y) < \sum_{\theta=1}^{\Theta} k_\theta$. The density part vanishes also when $\mathrm{EssSup}(y) = \sum_{\theta=1}^{\Theta} k_\theta$ but $\mathrm{meas}\left\{t : y(t) = \sum_{\theta=1}^{\Theta} k_\theta\right\} = 0$.

To calculate the thermal LRMC, the LRC is first expressed as a sum of two terms, in Formula (2.3.19). One term is the minimum cost of providing sufficient capacity: it equals the maximum output times the unit capital cost of the least capital-intensive type of station. Its subdifferential is the long-run peak charge, which has a definite total but an indeterminate distribution over the output peaks. (The distribution is determined in the long-run equilibrium of supply and demand, but it cannot be determined just by LRC calculations.) The peak charge need not be a pure density, i.e., it can include a singular charge; indeed, it is entirely singular if the peak has a zero duration. The other term of the LRMC is the marginal fuel cost of the optimal generating system; it is always a pure density. Formalised next, these concepts and results are taken from Horsley's work [33, Section 2], which is expanded here and in [36] by using rearrangement and majorisation theory.

[6] It follows that: (i) $\kappa = 0$ if $\mathrm{EssSup}(y) < \sum_{\theta=1}^{\Theta} k_\theta$, (ii) $\nu = 0$ if $\mathrm{EssInf}(y) > 0$, (iii) κ is lattice-disjoint from ν (i.e., $\kappa \wedge \nu = 0$), and (iv) both κ and ν are p.f.a. or, equivalently, singular.

Given a unit capacity cost r_θ (a.k.a. fixed cost, in $\$/\mathrm{kW}$) for each $\theta \in \Theta$—in addition to the unit running cost w_θ (a.k.a. variable cost, in $\$/\mathrm{kWh}$)—the *long-run cost* (LRC) of generating an output $y \in L^\infty[0,T]$ from the thermal technology is, by definition,

$$C_{\mathrm{LR}}(y,r,w) = \inf_k \{r \cdot k + C_{\mathrm{SR}}(y,k,w)\}. \tag{2.3.11}$$

Any k yielding the infimum is an *optimal thermal generating system* for the given output (i.e., a system that minimises the output's total cost). The set of all the cost-minimising systems is denoted by $\check{K}(y,r,w)$; when it *is* unique, the optimal thermal system is denoted by $\check{k}(y,r,w)$.

In terms of Legendre-Fenchel conjugacy, Formula (2.3.11) means that C_{LR} is, as a function of r, the *concave conjugate* of $-C_{\mathrm{SR}}$ as a function of k (with y and w fixed). This is, of course, a general relationship between the LRC and SRC for any technology whatsoever. To give specific formulae for the optimal thermal system and the thermal LRC, the special case of a unit output of any given duration, τ, per cycle—i.e., an output 1_A with $\operatorname{meas} A = \tau$—is dealt with first. An optimal plant type for generating such an output is a θ that minimises $r_\theta + \tau w_\theta$; the set of all such plant types is $\operatorname{ArgMin}_\theta(r_\theta + \tau w_\theta)$.[7] Except for a finite number of τ's, the optimal plant type $\check{\theta}(\tau)$ is unique, in which case its unit variable and fixed costs are denoted by

$$\check{w}(\tau) := w_{\check{\theta}(\tau)} \quad \text{and} \quad \check{r}(\tau) := r_{\check{\theta}(\tau)}. \tag{2.3.12}$$

The function $\check{w}$ is nonincreasing and piecewise constant on $(0,T]$. If the numbers

$$\check{\tau}_\theta(r,w) := \frac{r_\theta - r_{\theta+1}}{w_{\theta+1} - w_\theta} \quad \text{for } \theta = 1,\ldots,\Theta-1 \tag{2.3.13}$$

form a decreasing sequence in $(0,T)$, i.e., if

$$0 < \check{\tau}_{\Theta-1} < \ldots < \check{\tau}_2 < \check{\tau}_1 < T \tag{2.3.14}$$

(which can be ensured by discarding any redundant plant types from the technology), then these are the jump points of the function $\check{w}$, which equals w_θ on the interval $(\check{\tau}_\theta, \check{\tau}_{\theta-1})$, with $\check{\tau}_\Theta := 0$ and $\check{\tau}_0 := T$. In other words, $\check{\theta}(\tau) = \theta$ for $\tau \in (\check{\tau}_\theta, \check{\tau}_{\theta-1})$. Therefore, $\check{\tau}_\theta$ is called the *break-even load duration* (a.k.a. equilibrium duration) for the two "adjacent" types of station (θ and $\theta+1$). See Figure 2.2a, which shows the case of three station types (i.e., $\Theta = 3$); in the terminology of [60, Figure 14.3], the $\check{\tau}_\theta$'s are found by intersecting the "screening lines".

The (minimum) thermal LRC of a unit load of duration $\tau > 0$ (i.e., an output 1_A with $\operatorname{meas} A = \tau$) is

$$c_{\mathrm{LR}}(\tau,r,w) := \min_\theta (r_\theta + \tau w_\theta) = \check{r}(\tau) + \tau \check{w}(\tau). \tag{2.3.15}$$

For a fixed r and w, this is an increasing and concave (piecewise linear) function of τ, with

$$c_{\mathrm{LR}}(0+) = \operatorname{Min}(r) := \min_\theta r_\theta \tag{2.3.16}$$

[7] The ratio τ/T is a special case of the *load factor* (the ratio of the average to the maximum output rate).

but $c_{\mathrm{LR}}(0) := 0$. Also,

$$\frac{\mathrm{d}c_{\mathrm{LR}}}{\mathrm{d}\tau}(\tau) = \check{w}(\tau) \tag{2.3.17}$$

except for a finite set of τ's, which are exactly the $\check{\tau}_\theta$'s under (2.3.14). The graph of c_{LR}, also known as the *total cost–duration curve*, is shown in Figure 2.2a here (as well as in, e.g., [9, 61–65: Figure 7], [64, pp. 37–40: Figure 3-4] and [83, 6.2: Figure 6.1]). Figure 2.2b shows $\check{w}$ as a function of τ.

The optimal (cost-minimising) thermal generating system $\check{k}$ can be determined by referring the break-even durations ($\check{\tau}_\theta$) to the load-duration curve, as is done in Figure 2.2c here (and in, e.g., [9, 61–65: Figure 7], [64, pp. 37–40: Figure 3-4] and [83, 6.2: Figure 6.1]). The *load-duration curve* (LDC) is next introduced formally, as the graph of the decreasing rearrangement of the output trajectory.

Definition 2.3.2 (Monotone rearrangement). *Let $L^0[0,T]$ denote the space of all equivalence classes of measurable functions on $[0,T]$. The nonincreasing rearrangement $y_\downarrow$ of a $y \in L^0[0,T]$ is the nonincreasing function on $[0,T]$ with the same distribution, relative to the Lebesgue measure, as the distribution of y. That is, $y_\downarrow$ is nonincreasing and, for every Borel set $B \subset \mathbb{R}$,*

$$\operatorname{meas}\{\tau \in [0,T] : y_\downarrow(\tau) \in B\} = \operatorname{meas}\{t \in [0,T] : y(t) \in B\}.$$

For definiteness, $y_\downarrow$ is taken to be right-continuous on $[0,T)$, so $y_\downarrow(0) = \operatorname{EssSup}(y)$. Also, if $y \geq 0$ then $y_\downarrow(T) := 0$ (a useful convention).

In this notation, the optimal system $\check{k}$ contains $y_\downarrow(\check{\tau}_\theta) - y_\downarrow(\check{\tau}_{\theta-1})$ units (kWs) of plant of type θ (if, for the optimal system to be unique, $y_\downarrow$ is assumed to be continuous, at least at all the points of the sequence $\check{\tau}_{\Theta-1} < \ldots < \check{\tau}_2 < \check{\tau}_1$). Note that $y_\downarrow(\check{\tau}_\Theta) := y_\downarrow(0) = \operatorname{EssSup}(y)$ and that, to take account of the base load, $y_\downarrow(\check{\tau}_0) := y_\downarrow(T) := 0$ (not $\operatorname{EssInf}(y)$).

The LRC of any output $y \in L^\infty_+[0,T]$ is next given in terms of the LDC ($y_\downarrow$) by slicing the area between its graph and the time axis in parallel to that axis and adding up the LRCs of all the infinitesimal slices—i.e., by evaluating c_{LR} at the duration of each load level y, and then integrating along the load axis. After the substitution y $= y_\downarrow(\tau)$, the integral can be evaluated by parts; these key steps were first made in [33, (3) and (5)].

Theorem 2.3.3 (Thermal LRC and cost-minimising system). *For every $y \in L^\infty_+[0,T]$*

$$C_{\mathrm{LR}}(y,r,w) = -\int_{[0,T]} c_{\mathrm{LR}}(\tau,r,w)\,\mathrm{d}y_\downarrow(\tau) \tag{2.3.18}$$

$$= \operatorname{Min}(r)\operatorname{EssSup}(y) + \int_0^T \check{w}(\tau)\,y_\downarrow(\tau)\,\mathrm{d}\tau. \tag{2.3.19}$$

(For the purpose of Lebesgue-Stieltjes integration in (2.3.18), $y_\downarrow(0-) := \operatorname{EssSup}(y)$ and $y_\downarrow(T+) := 0$, i.e., $\mathrm{d}y_\downarrow\{0\} = 0$ and $\mathrm{d}y_\downarrow\{T\} = \operatorname{EssInf}(y)$.) Furthermore, a thermal system $k \in \mathbb{R}^\Theta$ minimises

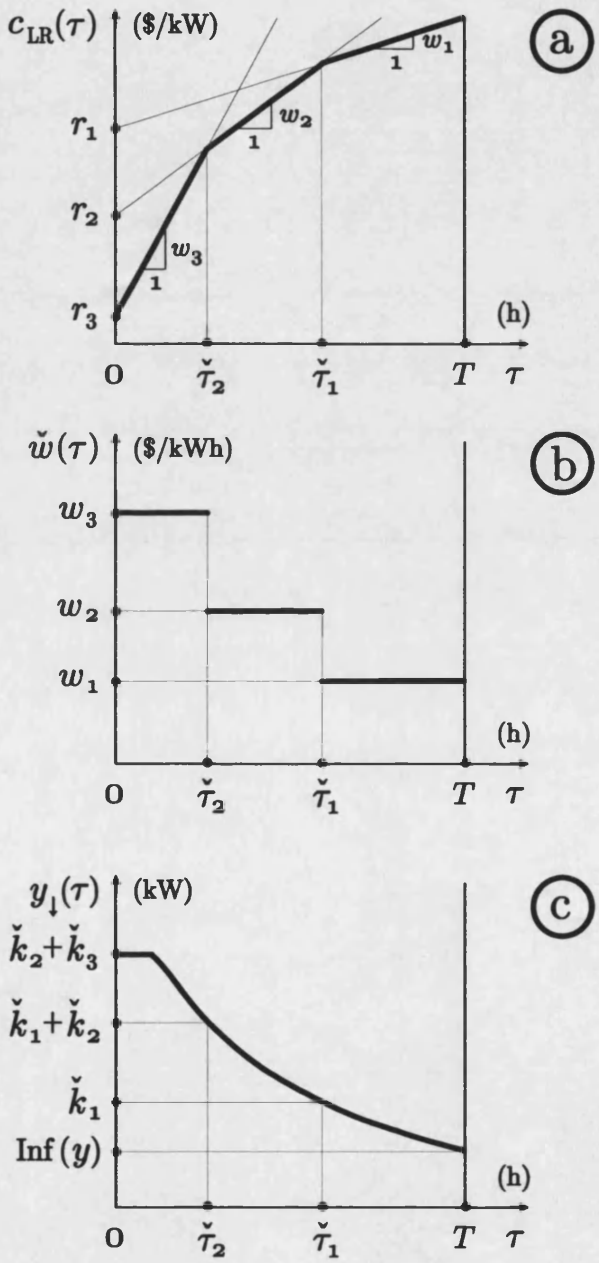

FIGURE 2.2. Thermal long-run costs and cost-minimising system: (a) LRC of a unit output as a function of its duration τ, for a three-station technology with unit capacity costs (r_1, r_2, r_3) and unit running costs (w_1, w_2, w_3), (b) unit running cost of the optimal station as a function of its load duration τ, and (c) load-duration curve and the optimal system $(\check{k}_1, \check{k}_2, \check{k}_3)$.

the total cost of y in (2.3.11)—i.e., $k \in \check{K}(y, r, w)$—if and only if

$$\sum_{\omega=1}^{\theta} k_\omega \in [y_\downarrow(\check{\tau}_\theta+), y_\downarrow(\check{\tau}_\theta-)] \quad \text{for each } \theta = 1, \ldots, \Theta. \tag{2.3.20}$$

Proof. Substitute the integral (2.3.7) for c_{SR} in (2.3.3), and apply Fubini's Theorem: with

$$\widetilde{\tau}_y(\mathrm{y}) := \operatorname{meas}\{t : y(t) \geq \mathrm{y}\} \tag{2.3.21}$$

denoting the total time when the output rate $y(t)$ exceeds any given level y, this gives

$$\begin{aligned} C_{\mathrm{SR}}(y, k, w) &= \int_0^T \int_0^{y(t)} \sum_{\theta=1}^{\Theta} w_\theta 1_{\left[\sum_{\omega=1}^{\theta-1} k_\omega, \sum_{\omega=1}^{\theta} k_\omega\right]}(\mathrm{y})\, \mathrm{dy}\, \mathrm{d}t \\ &= \sum_{\theta=1}^{\Theta} w_\theta \int_{\sum_{\omega=1}^{\theta-1} k_\omega}^{\sum_{\omega=1}^{\theta} k_\omega} \operatorname{meas}\{t : y(t) \geq \mathrm{y}\}\, \mathrm{dy} = \sum_{\theta=1}^{\Theta} w_\theta \int_{k_1+\ldots+k_{\theta-1}}^{k_1+\ldots+k_\theta} \widetilde{\tau}_y(\mathrm{y})\, \mathrm{dy} \end{aligned}$$

if $y \leq \sum_{\theta=1}^{\Theta} k_\theta$ (otherwise $C_{\mathrm{SR}}(y, k, w) = +\infty$). Substitute this expression for the SRC in the conjugacy formula for the LRC (2.3.11) to obtain

$$\begin{aligned} C_{\mathrm{LR}}(y, r, w) &= \inf_k \left\{ \sum_{\theta=1}^{\Theta} \left(r_\theta k_\theta + w_\theta \int_{\sum_{\omega=1}^{\theta-1} k_\omega}^{\sum_{\omega=1}^{\theta} k_\omega} \widetilde{\tau}_y(\mathrm{y})\, \mathrm{dy} \right) : \sum_{\theta=1}^{\Theta} k_\theta = \operatorname{EssSup}(y) \right\} \quad (2.3.22) \\ &\geq \int_0^{\operatorname{EssSup}(y)} c_{\mathrm{LR}}(\widetilde{\tau}_y(\mathrm{y}), r, w)\, \mathrm{dy} \end{aligned}$$

(since $r_\theta + w_\theta \tau \geq c_{\mathrm{LR}}(\tau, r, w)$ for every θ and τ). To show that the integral actually equals C_{LR}, take any k such that $\sum_{\omega=1}^{\theta} k_\omega \in [y_\downarrow(\check{\tau}_\theta+), y_\downarrow(\check{\tau}_\theta-)]$ for each $\theta = 1, \ldots, \Theta$: at such a k, the minimand on the r.h.s. of (2.3.22) is equal to the integral of c_{LR}. To see this, note first that for $\mathrm{y} > \sum_{\omega=1}^{\theta-1} k_\omega$ if $y(t) \geq \mathrm{y}$ then $y(t) > \sum_{\omega=1}^{\theta-1} k_\omega \geq y_\downarrow(\check{\tau}_{\theta-1}+)$ by the choice of k, and so

$$\widetilde{\tau}_y(\mathrm{y}) \leq \operatorname{meas}\{t : y(t) > y_\downarrow(\check{\tau}_{\theta-1}+)\} \leq \check{\tau}_{\theta-1}$$

(since $\operatorname{meas}\{t : y(t) > y_\downarrow(\tau+)\} \leq \tau$, with equality if $y_\downarrow$ has a jump at τ). Second, for $\mathrm{y} \leq \sum_{\omega=1}^{\theta} k_\omega$, if $y(t) \geq y_\downarrow(\check{\tau}_\theta-)$ then $y(t) \geq \sum_{\omega=1}^{\theta} k_\omega \geq \mathrm{y}$ (again by the choice of k), and so

$$\widetilde{\tau}_y(\mathrm{y}) \geq \operatorname{meas}\{t : y(t) \geq y_\downarrow(\check{\tau}_\theta-)\} \geq \check{\tau}_\theta$$

since $\operatorname{meas}\{t : y(t) \geq y_\downarrow(\tau+)\} \geq \tau$, with equality if $y_\downarrow$ has a jump at τ. For $\mathrm{y} \in \left(\sum_{\omega=1}^{\theta-1} k_\omega, \sum_{\omega=1}^{\theta} k_\omega\right]$ this shows that $\widetilde{\tau}_y(\mathrm{y}) \in [\check{\tau}_\theta, \check{\tau}_{\theta-1}]$ and hence $c_{\mathrm{LR}}(\widetilde{\tau}_y(\mathrm{y}), r, w) = r_\theta + w_\theta \tau$. By adding over θ, it follows that, at any k satisfying (2.3.20), and only at such a k, the minimand on the r.h.s. of (2.3.22) is equal to the integral of c_{LR}, i.e.,

$$C_{\mathrm{LR}}(y, r, w) = \int_0^{\operatorname{EssSup}(y)} c_{\mathrm{LR}}(\widetilde{\tau}_y(\mathrm{y}), r, w)\, \mathrm{dy}. \tag{2.3.23}$$

This proves the characterisation (2.3.20) of $\check{K}$ and provides a basic integral formula for C_{LR}. To transform the integral into (2.3.18), substitute $\mathrm{y} = y_\downarrow(\tau)$ in (2.3.23). Finally, use (2.3.16) and (2.3.17) to integrate (2.3.18) by parts and thus transform it into (2.3.19). $\square$

Comment: $\widetilde{\tau}_y(\mathrm{y}) = (y_\downarrow)^{-1}(\mathrm{y})$ for nearly every $\mathrm{y} \in [0, \mathrm{EssSup}(y)]$ when $y_\downarrow$ is inverted as a nonincreasing interval-valued correspondence from $[0, T]$ onto $[0, \mathrm{EssSup}(y)]$, i.e., when the LDC has its jumps filled in—so that, in particular, $(y_\downarrow)^{-1}(\mathrm{y}) = T$ for $\mathrm{y} \in [0, \mathrm{EssInf}(y)]$. The exceptional y's are those with $\mathrm{meas}\{t : y(t) = \mathrm{y}\} > 0$, i.e., any plateau levels of y.

Formula (2.3.18) can be used to calculate the thermal LRMCs, i.e., to subdifferentiate C_{LR} w.r.t. y. Since this is done term by term, denote

$$C_{\mathrm{LR}}^{\mathrm{N}}(y, r, w) := \int_0^T \breve{w}(\tau)\, y_\downarrow(\tau)\, \mathrm{d}\tau \quad \text{for } y \in L_+^\infty[0, T] \tag{2.3.24}$$

with $C_{\mathrm{LR}}^{\mathrm{N}}(y)$ equal to $+\infty$ (like C_{LR} and C_{SR}) if $y \not\geq 0$. This is the integral term of C_{LR} in the decomposition (2.3.18); it is the *long-run cost net of the peak capacity cost* (i.e., net of the supremum term).[8] Along with $C_{\mathrm{LR}}^{\mathrm{N}}$, it is useful to study also the function—denoted by $C_{\mathrm{Ex}}^{\mathrm{N}}$—that is defined by the same formula (2.3.24), but for *every* $y \in L^\infty$ (instead of L_+^∞); this is the simplest finite extension of $C_{\mathrm{LR}}^{\mathrm{N}}$ from L_+^∞ to all of L^∞. Being sublinear, i.e., subadditive and positively linearly homogeneous (p.l.h.), it is the support function of some convex set S, i.e., it equals

$$\delta^\#(y \mid S) := \sup_p \left\{ \int_0^T p(\tau)\, y(\tau)\, \mathrm{d}\tau : p \in S \right\}$$

for some $S \subset L^1[0, T]$; the superscript $\#$ means that this is the Fenchel-Legendre convex conjugate of $\delta(\cdot \mid S)$, the 0-∞ indicator function of the set S (0 on S and $+\infty$ outside of S). This set is next identified, in Proposition 2.3.5, as the set of all the functions on $[0, T]$ that are majorised by $\breve{w}$, in the sense of the Hardy-Littlewood-Pólya order $\prec_{\mathrm{HLP}}$ (abbreviated to $\prec$).

Definition 2.3.4 (Majorisation). *A function $p \in L^1[0, T]$ is weakly majorised by another integrable function f if $\int_0^\tau p_\downarrow(t)\, \mathrm{d}t \leq \int_0^\tau f_\downarrow(t)\, \mathrm{d}t$ for every $\tau \in [0, T]$; this relationship is denoted by $p \prec\prec f$.*[9] *If, in addition, equality holds for $\tau = T$ then f majorises p; this is written as $p \prec f$.*

The set of all the functions majorised by f is denoted by

$$\mathrm{maj}(f) := \left\{ p \in L^1[0, T] : p \prec f \right\}$$

and its subset consisting of all the functions equidistributed to f—a.k.a. the rearrangements of f, i.e., the functions on $[0, T]$ with the same distribution (w.r.t. meas) as that of f—is denoted by

$$\mathrm{eqd}(f) := \left\{ p : \mathrm{meas}\, p^{-1}(B) = \mathrm{meas}\, f^{-1}(B) \text{ for each Borel set } B \subset \mathbb{R} \right\}.$$

The set $\mathrm{maj}(f)$ is convex and weakly compact [77, Theorem 2], and it is equal to the closed convex hull of $\mathrm{eqd}(f)$, both for the L^1-norm and for the weak topology $\mathrm{w}(L^1, L^\infty)$: see, e.g., [14, 21.9],

[8] This is the minimum cost of providing sufficient capacity, *not* the capital cost of an optimal system.

[9] The order $\prec\prec$ is also known as the *lower* weak majorisation and denoted by $\prec_{\mathrm{w}}$ (to distinguish it from the upper weak majorisation $\prec^{\mathrm{w}}$) in, e.g., [61].

[19, 5.2] or [59, 15.6 (i)]. (The same is true even for the L^∞-norm when $f \in L^\infty$: see [19, 5.2].) A stronger result of Ryff [78, p. 1026] is that $\mathrm{eqd}(f)$ is the set of all the extreme points of $\mathrm{maj}(f)$, i.e.,

$$\mathrm{ext}\,\mathrm{maj}(f) = \mathrm{eqd}(f). \tag{2.3.25}$$

Proposition 2.3.5 (Luxemburg). *For each $f \in L^1[0,T]$,*[10]

$$\delta^{\#}(y \mid \mathrm{maj}(f)) = \int_0^T f_{\downarrow}(\tau)\, y_{\downarrow}(\tau)\,\mathrm{d}\tau \quad \text{for every } y \in L^\infty[0,T]. \tag{2.3.26}$$

So the finite extension of $C^{\mathrm{N}}_{\mathrm{LR}}$ defined by (2.3.24) for every $y \in L^\infty$ is

$$C^{\mathrm{N}}_{\mathrm{Ex}}(y) := \int_0^T \check{w}(\tau)\, y_{\downarrow}(\tau)\,\mathrm{d}\tau = \delta^{\#}(y \mid \mathrm{maj}(\check{w})) \tag{2.3.27}$$

i.e., it is the support function of the set of functions majorised by the nonincreasing function $\check{w}$ defined by (2.3.12).

Proof. This is based on the Hardy-Littlewood-Pólya Inequality, which is that

$$\int_0^T f(\tau)\, y(\tau)\,\mathrm{d}\tau \le \int_0^T f_{\downarrow}(\tau)\, y_{\downarrow}(\tau)\,\mathrm{d}\tau \tag{2.3.28}$$

(see, e.g., [14, 12.2], [19, 3.4], or [59, 8.2]). In other words,

$$\int_0^T p(\tau)\, y(\tau)\,\mathrm{d}\tau \le \int_0^T f_{\downarrow}(\tau)\, y_{\downarrow}(\tau)\,\mathrm{d}\tau \quad \text{for every } p \in \mathrm{eqd}(f)$$

and, given that $\mathrm{maj}(f) = \mathrm{cl}\,\mathrm{conv}\,\mathrm{eqd}(f)$, it follows that

$$\int_0^T p(\tau)\, y(\tau)\,\mathrm{d}\tau \le \int_0^T f_{\downarrow}(\tau)\, y_{\downarrow}(\tau)\,\mathrm{d}\tau \quad \text{for every } p \in \mathrm{maj}(f). \tag{2.3.29}$$

Finally, note that this upper bound (on $\int py\,\mathrm{d}\tau$) is attained at a suitable choice of p from $\mathrm{eqd}(f)$: take any Lebesgue measure-preserving map ρ: $[0,T] \to [0,T]$ such that $y = y_{\downarrow} \circ \rho$, and set $p := f_{\downarrow} \circ \rho$. Then

$$\int_0^T p(\tau)\, y(\tau)\,\mathrm{d}\tau = \int_0^T f_{\downarrow}(\rho(\tau))\, y_{\downarrow}(\rho(\tau))\,\mathrm{d}\tau = \int_0^T f_{\downarrow}(\tau)\, y_{\downarrow}(\tau)\,\mathrm{d}\tau.$$

(The required ρ exists by the Lorentz-Ryff Lemma: see, e.g., [14, 6.2], [19, 3.3] or [77, Lemma 1].) The proof of (2.3.26) is complete, and (2.3.27) follows because $\check{w}$ is a nonincreasing function on $[0,T]$. □

Comment (continuity of LRC as function of output): Since $\mathrm{maj}(f)$ is a $\mathrm{w}\left(L^1, L^\infty\right)$-compact set, its support function (defined on $L^\infty[0,T]$) is continuous for the Mackey topology $\mathrm{m}\left(L^\infty, L^1\right)$. Its continuity follows also from its representation (2.3.26) as the composition of the linear functional $\langle f_{\downarrow} \mid \cdot \rangle$ and the nonincreasing-rearrangement operator ($y \mapsto y_{\downarrow}$), since the latter is Mackey continuous on L^∞: see [34]. Since continuity of a convex function implies that all of its algebraic subgradients are

[10] This is in, e.g., [14, 13.8 (i)] and [59, 9.3].

continuous linear functionals—see, e.g., [32, 14B: Proof of Theorem]—the algebraic subdifferential $\partial^{\mathrm{a}} C^{\mathrm{N}}_{\mathrm{Ex}}(y)$ lies wholly in L^1. Similarly $\partial^{\mathrm{a}}\,\mathrm{EssSup}(y) \subset L^{\infty *}$, since EssSup is a norm-continuous function on L^{∞}. The two subdifferentials are calculated next (Theorems 2.3.6 and 2.3.7).

Comment ($C^{\mathrm{N}}_{\mathrm{LR}}$ as a "basic" function): A function $y \mapsto C(y)$ on a Lebesgue space $L^{\varrho}[0,T]$ is called *symmetric* a.k.a. *rearrangement-invariant* if $C(y) = C(y_{\downarrow})$, i.e., if its value depends only on the distribution of its argument. A function of the form (2.3.26) is the simplest example—and it has been called a *basic convex symmetric function* in [36] because every convex symmetric function is the supremum of a family of sums of such a function and a constant: see [14, 20.2] or [59, 13.4]. Thus such a function is to a general convex symmetric function as a linear function is to a convex one.

When S is a nonempty, convex and closed subset of a real vector space P that is paired with another space Y by a bilinear form $\langle \cdot | \cdot \rangle : P \times Y \to \mathbb{R}$, its support function can be subdifferentiated by the formula

$$\partial\delta^{\#}(y \mid S) = \operatorname*{argmax}_{S} \langle \cdot \,|\, y\rangle = \left\{p \in S : \langle p \,|\, y\rangle = \delta^{\#}(y \mid S)\right\} \tag{2.3.30}$$

which is given in, e.g., [70, 23.5.3] and [73, p. 36, lines 1–7]. This is a variant of Euler's Theorem on homogeneous functions. Applied to $S = \mathrm{maj}(f)$, as a subset of the space $P = L^1[0,T]$ paired with $Y = L^{\infty}[0,T]$, it gives

$$\partial\delta^{\#}(y \mid \mathrm{maj}(f)) = \left\{p \in L^1 : p \prec f,\ \int_0^T p(\tau)\, y(\tau)\,\mathrm{d}\tau = \int_0^T f_{\downarrow}(\tau)\, y_{\downarrow}(\tau)\,\mathrm{d}\tau\right\}. \tag{2.3.31}$$

This formula can be further spelt out by analysing the case of equality in (2.3.29). This can be done similarly to the case of equality in the Hardy-Littlewood-Pólya Inequality (2.3.28), which is dealt with by Day [18, 5.2 and 6.2]; it can also be deduced from Day's analysis and (2.3.25). When $y_{\downarrow}$ is *strictly* decreasing, the result is that the conditions $\int_0^T p(\tau)\, y(\tau)\,\mathrm{d}\tau = \int_0^T f_{\downarrow}(\tau)\, y_{\downarrow}(\tau)\,\mathrm{d}\tau$ and $p \prec f$ imply that p is actually equidistributed to f (i.e., $p_{\downarrow} = f_{\downarrow}$) and, furthermore, that $p = p_{\downarrow} \circ \rho$ for $\rho = (y_{\downarrow})^{-1} \circ y$—which is the unique Lebesgue measure-preserving map ρ such that $y = y_{\downarrow} \circ \rho$. When $y_{\downarrow}$ is *not* strictly decreasing, the set of its *plateau levels* is

$$\mathbb{P}_y := \{\mathrm{y} \in \mathbb{R} : \mathrm{meas}\{t \in [0,T] : y(t) = \mathrm{y}\} > 0\} \tag{2.3.32}$$

and, for a $p \prec f$, the equality $\int_0^T p(\tau)\, y(\tau)\,\mathrm{d}\tau = \int_0^T f_{\downarrow}(\tau)\, y_{\downarrow}(\tau)\,\mathrm{d}\tau$ holds if and only if:

$$p = f_{\downarrow} \circ (y_{\downarrow})^{-1} \circ y \quad \text{on } \{t \in [0,T] : y(t) \notin \mathbb{P}_y\} \tag{2.3.33}$$

$$f_{\downarrow} \text{ restricted to } \{t : y_{\downarrow}(t) = \mathrm{y}\} \text{ majorises the restriction of } p \text{ to } \{t : y(t) = \mathrm{y}\}. \tag{2.3.34}$$

Since $C^{\mathrm{N}}_{\mathrm{LR}}(y) = \delta^{\#}(y \mid \mathrm{maj}(\check{w}))$ if $y \geq 0$ (with $C^{\mathrm{N}}_{\mathrm{LR}}(y) = +\infty$ otherwise), a description of $\partial_y C^{\mathrm{N}}_{\mathrm{LR}}$ follows from (2.3.33) and (2.3.34).

Theorem 2.3.6 (Thermal LRMC net of peak charge). *For every $y \geq 0$, $p \in \partial_y C_{\mathrm{LR}}^{\mathrm{N}}(y, r, w)$ if and only if $p \in L^{\infty *}[0, T]$ and:*

1. *$p_{\mathrm{CA}}(t) = p_{\mathrm{N}} - \nu$ for some p_{N} and ν in $L_+^1[0, T]$ such that:*

   (a) $p_{\mathrm{N}} = \check{w} \circ (y_\downarrow)^{-1} \circ y$ on $\{t : y(t) \notin \mathbb{P}_y\}$.

   (b) $p_{\mathrm{N} \mid \{t : y(t) = \mathrm{y}\}} \prec \check{w}_{\mid \{t : y_\downarrow(t) = \mathrm{y}\}}$ for any $\mathrm{y} \in \mathbb{P}_y$ (i.e., for any plateau level of y).

   (c) ν vanishes almost everywhere outside of $\{t : y(t) = 0\}$.

2. *$p_{\mathrm{FA}} \leq 0$ and p_{FA} is concentrated on $\{t : y(t) \leq \epsilon\}$ for every number $\epsilon > 0$.*

Proof. Fix any r and w. Since

$$C_{\mathrm{LR}}^{\mathrm{N}}(\cdot, r, w) = C_{\mathrm{Ex}}^{\mathrm{N}}(\cdot, r, w) + \delta\left(\cdot \mid L_+^\infty\right) \tag{2.3.35}$$

and one of the terms ($C_{\mathrm{Ex}}^{\mathrm{N}}$) is continuous on L^∞ (for the norm topology and even for $\mathrm{m}\left(L^\infty, L^1\right)$), subdifferentiation is additive at every point where the other function ($\delta\left(\cdot \mid L_+^\infty\right)$) is finite, i.e., at every $y \geq 0$: see, e.g., [73, Theorem 20 (i) under (a)] or [80, 5.38 (b)]. In view of (2.3.27) and (2.3.31), a $p_{\mathrm{N}} \in \partial C_{\mathrm{Ex}}^{\mathrm{N}}(y) \subset L^1$ is fully characterised by (2.3.33) and (2.3.34) with $\check{w}$ in place of $f_\downarrow$—i.e., by Conditions 1a and 1b. And a $\lambda \in \partial\delta\left(y \mid L_+^\infty\right)$ if and only if $y \geq 0$, $\lambda \leq 0$ and $\langle \lambda \mid y \rangle = 0$—which translates into Conditions 1c and 2 on $\nu := -\lambda_{\mathrm{CA}}$ and $p_{\mathrm{FA}} = \lambda_{\mathrm{FA}}$. □

Comment (extreme subgradients of $C_{\mathrm{LR}}^{\mathrm{N}}$): Formulae (2.3.33) and (2.3.34) can be enhanced by describing the extreme points of $\partial\delta^\#(y \mid \mathrm{maj}(f))$. This can be done in terms of the measure-preserving maps ρ: $[0, T] \to [0, T]$ such that $y = y_\downarrow \circ \rho$. The set of all such maps—which are called the *ranking patterns* of y—is denoted by $\mathcal{R}(y)$. If $y_\downarrow$ decreases strictly (i.e., $\mathbb{P}_y = \emptyset$), then y has a unique pattern, viz.,

$$\rho_y(t) = (y_\downarrow)^{-1}(y(t)) = \mathrm{meas}\{\tau : y(\tau) \geq y(t)\}.$$

In other words, $\rho_y(t)/T$ is t's "percentage above"—the fraction of $[0, T]$ on which y is above its "current" value $y(t)$. Thus ρ_y ranks the points of $[0, T]$ by the value of y (hence its name, "the ranking pattern"). When $y_\downarrow$ is *not* strictly decreasing, $\rho \in \mathcal{R}(y)$ if and only if, for each $\mathrm{y} \in \mathbb{P}_y$ (i.e., for each plateau level of y):

$$\rho_{\mid \{t : y(t) = \mathrm{y}\}} \text{ is a measure preserving map of } \{t : y(t) = \mathrm{y}\} \text{ onto } \{t : y_\downarrow(t) = \mathrm{y}\}$$
$$\rho = (y_\downarrow)^{-1} \circ y \quad \text{on } \{t : y(t) \notin \mathbb{P}_y\}.$$

In these terms,

$$\mathrm{ext}\,\partial\delta^\#(y \mid \mathrm{maj}(f)) = \{f \circ \rho : \rho \in \mathcal{R}(y)\} \tag{2.3.36}$$

i.e., the extreme subgradients of the support function of $\mathrm{maj}(f)$ are those rearrangements of f which have a *common pattern* with y or, in other words, are *arranged similarly* to y in the sense of Day

[18, p. 932].[11] This is a result derived from (2.3.25) in [36, Theorem 3]. With $\check{w}$ in place of $f_{\downarrow}$, it gives $\operatorname{ext} \partial C^{\mathrm{N}}_{\mathrm{Ex}}(y)$—which equals $\operatorname{ext} \partial C^{\mathrm{N}}_{\mathrm{LR}}(y)$ if $y \gg 0$ (i.e., if $y(t) > 0$ for almost every t).

The other term of the LRMC is the peak charge, which is formally a subgradient of EssSup as a function on the space $L^{\infty}[0,T]$ paired with its norm-dual $L^{\infty *}[0,T]$. The following description of $\partial \operatorname{EssSup}(y)$ as a unit "mass" concentrated on the near-peaks of y is in, e.g., [24, Example 4.7].

Theorem 2.3.7 (LRMC peak charge). *For every* $y \in L^{\infty}$, $\gamma \in \partial \operatorname{EssSup}(y)$ *if and only if* $\gamma \in L^{\infty *}[0,T]$ *and:*

1. $\gamma \geq 0$ *and* $\|\gamma\|^{*}_{\infty} = 1$, *i.e.,* $\langle \gamma \mid 1_{[0,T]} \rangle = 1$.
2. γ *is concentrated on* $\{t : y(t) \geq \operatorname{EssSup}(y) - \epsilon\}$ *for every number* $\epsilon > 0$.

For a $\gamma \in L^{1}[0,T]$, *these conditions simplify to:* $\gamma \geq 0$, $\int_0^T \gamma(t)\,\mathrm{d}t = 1$, *and* γ *vanishes outside of the set* $\{t : y(t) = \operatorname{EssSup}(y)\}$.

Including the peak charge, the set of all LRMCs at a $y \geq 0$, in the price space $L^{\infty *}[0,T]$, is therefore

$$\partial_y C_{\mathrm{LR}}(y,r,w) = \partial_y C^{\mathrm{N}}_{\mathrm{LR}}(y,r,w) + \operatorname{Min}(r)\, \partial \operatorname{EssSup}(y)$$

which is the sum of the subdifferentials described in Theorems 2.3.6 and 2.3.7. (As in the Proof of Theorem 2.3.6, subdifferentiation is additive: see, e.g., [73, Theorem 20 (i) under (a)] or [80, 5.38 (b)].) It follows that if $\operatorname{Min}(r) > 0$ then an LRMC represented by a density exists at y if and only if y has a peak plateau, i.e., $L^1 \cap \partial_y C_{\mathrm{LR}}(y) \neq \emptyset$ if and only if $\operatorname{meas}\{t : y(t) = \operatorname{EssSup}(y)\} > 0$.

Comments (inclusion of free disposal in the cost function):

- The simplest finite extension of $C^{\mathrm{N}}_{\mathrm{LR}}$ (viz., $C^{\mathrm{N}}_{\mathrm{Ex}}$) is adequate as a tool for subdifferentiating it as the sum (2.3.35), but another finite extension is of additional interest. It is defined by

$$C^{\mathrm{N}\dagger}_{\mathrm{Ex}}(y,r,w) := C^{\mathrm{N}}_{\mathrm{LR}}(y^{+},r,w) = \int_0^T \check{w}(\tau)\, y^{+}_{\downarrow}(\tau)\,\mathrm{d}\tau \tag{2.3.37}$$

  where $y^{+} := \sup\{y, 0\}$ is the nonnegative part of y, and $y^{+}_{\downarrow}$ means $(y^{+})_{\downarrow} = (y_{\downarrow})^{+}$. With the peak term included, the extension of C_{LR} (from L^{∞}_{+} to all of L^{∞}) is

$$C^{\dagger}_{\mathrm{LR}}(y,r,w) := C_{\mathrm{LR}}(y^{+},r,w) = \int_0^T \check{w}(\tau)\, y^{+}_{\downarrow}(\tau)\,\mathrm{d}\tau + \operatorname{Min}(r) \operatorname{EssSup}(y^{+}) \tag{2.3.38}$$

  and this is the cost function that corresponds to the free-disposal hull of the production set: see [37].

[11] Day's definition of similarity of arrangement [18, p. 932] is equivalent to the existence of a common ranking pattern by [18, p. 939, 5.6]. See also [45] for a discussion and applications of arrangement similarity and useful weaker conditions.

- By adapting the Proof of Proposition 2.3.5, one can show that $C_{\mathrm{Ex}}^{\mathrm{N}\dagger}$ is the support function of the set of all the nonnegative functions on $[0,T]$ weakly majorised by $\check{w}$, which is denoted by

$$\mathrm{wmj}^{+}(\check{w}) := \{p : 0 \le p \prec\prec \check{w}\}.$$

So this function can be subdifferentiated by Formula (2.3.30), with the result that: $p_{\mathrm{N}} \in \partial C_{\mathrm{Ex}}^{\mathrm{N}\dagger}(y) = \partial\delta^{\#}(y \mid \mathrm{wmj}^{+}(\check{w}))$ if and only if (i) $p_{\mathrm{N}} = \check{w} \circ (y_{\downarrow})^{-1} \circ y$ on $\{t : y(t) > 0, y(t) \notin \mathbb{P}_y\}$, (ii) $p_{\mathrm{N}\,|\,\{t:\,y(t)=\mathrm{y}\}} \prec \check{w}_{|\,\{t:\,y_{\downarrow}(t)=\mathrm{y}\}}$ for any positive $\mathrm{y} \in \mathbb{P}_y$ (i.e., for any positive plateau level of y), (iii) $p_{\mathrm{N}\,|\,\{t:\,y(t)=0\}} \prec\prec \check{w}_{|\,\{t:\,y_{\downarrow}(t)=0\}}$ if $\mathrm{meas}\,\{t : y(t) = 0\} > 0$ (i.e., if $0 \in \mathbb{P}_y$), and (iv) $p_{\mathrm{N}}(t) = 0$ for any t with $y(t) < 0$. Furthermore, the extreme subgradients are characterised in [36, Theorem 4] by using the counterpart of (2.3.25) for the set $\mathrm{wmj}^{+}(f)$, which is given in [35].

Finally, variants of these results are given for the restrictions of the SRC and LRC functions from $L^{\infty}[0,T]$ to $\mathcal{C}[0,T]$; these are denoted by

$$C_{\mathrm{SR}|\mathcal{C}}(\cdot,k,w) := C_{\mathrm{SR}}(\cdot,k,w)_{|\mathcal{C}[0,T]}$$
$$C_{\mathrm{LR}|\mathcal{C}}(\cdot,r,w) := C_{\mathrm{LR}}(\cdot,r,w)_{|\mathcal{C}[0,T]}.$$

With the commodity space restricted to contain only the continuous functions, singular capacity charges have a simpler mathematical representation by a measure κ or γ that—unlike its counterpart in Theorem 2.3.1 or 2.3.7—is concentrated on the set of *exact* output peaks. Being a level set of a continuous function, this set is closed, and κ is concentrated on it if and only if the set contains the *support* a.k.a. carrier $\mathrm{supp}(\kappa)$, which is defined as the smallest *closed* set of full κ-measure. So this set can be used to describe the capacity charges when $\mathcal{C}[0,T]$ is the commodity space (it has no counterpart for $\kappa \in L^{\infty *}$).

A measure $p \in \mathcal{M}[0,T]$ has the Lebesgue decomposition into $p_{\mathrm{AC}} + p_{\mathrm{Si}}$, the sum of its *absolutely continuous* and *singular* parts, with respect to the Lebesgue measure: see, e.g., [25, III.4.14]. The singular part, p_{Si}, is a measure concentrated on a set of zero Lebesgue measure (not generally a closed one). The absolutely continuous part of p is identified with its density w.r.t. meas (which exists by the Radon-Nikodym Theorem)—so it is a price function $p_{\mathrm{AC}} \in L^{1}[0,T]$.

Theorem 2.3.8 (Thermal SRMC of continuous outputs). *Assume that $y \in \mathcal{C}[0,T]$, in addition to $0 \le y \le \sum_{\theta=1}^{\Theta} k_\theta$ and $k > 0$. Then $p \in \partial_y C_{\mathrm{SR}|\mathcal{C}}(y,k,w)$ if and only if $p \in \mathcal{M}[0,T]$ and:*

1. *$p_{\mathrm{AC}}(t) \in \partial_y c_{\mathrm{SR}}(y(t), k, w)$ for almost every $t \in [0,T]$.*

2. *$p_{\mathrm{Si}} = \kappa - \nu$ for some κ and ν in $\mathcal{M}_{+}[0,T]$ such that* $\mathrm{supp}(\kappa) \subseteq \left\{t : y(t) = \sum_{\theta=1}^{\Theta} k_\theta\right\}$ *and* $\mathrm{supp}(\nu) \subseteq \{t : y(t) = 0\}$.[12]

[12] It follows that: (i) $\kappa = 0$ if $\mathrm{Max}(y) < \sum_{\theta=1}^{\Theta} k_\theta$, (ii) $\nu = 0$ if $\mathrm{Min}(y) > 0$, (iii) $\mathrm{supp}(\kappa) \cap \mathrm{supp}(\nu) = \emptyset$ (so, *a fortiori*, $\kappa \wedge \nu = 0$), and (iv) both κ and ν are singular (w.r.t. meas).

Proof. This follows from Theorem 2.3.1: see, e.g., [72, Corollary 4B], or apply [50, Section 4: Theorem 2] to $\mathcal{C}$ as a subspace of L^{∞}. □

The restriction to $\mathcal{C}$ does not change the mathematical form of the LRMC net of the peak charge, except for simplifying the term that comes from the nonnegativity constraint, i.e., from the indicator term in (2.3.35). Formally, $\partial_y C^{\mathrm{N}}_{\mathrm{Ex}}$ stays exactly the same (as a subset of L^1), i.e., $\partial_y C^{\mathrm{N}}_{\mathrm{Ex}|\mathcal{C}} = \partial_y C^{\mathrm{N}}_{\mathrm{Ex}}$ at every $y \in \mathcal{C}[0,T]$. This is because: (i) $C^{\mathrm{N}}_{\mathrm{Ex}}$ (given by (2.3.27)) is m (L^{∞}, L^1)-continuous, and (ii) $\mathcal{C}[0,T]$ is m (L^{∞}, L^1)-dense in $L^{\infty}[0,T]$. Finally, a measure λ is in $\partial\delta(y \mid \mathcal{C}_+)$ if and only if $y \geq 0$, $\lambda \leq 0$ and $\int_{[0,T]} y(t)\,\lambda(\mathrm{d}t) = 0$. Spelt out in terms of $\nu := -\lambda_{\mathrm{AC}}$ and $p_{\mathrm{Si}} = \lambda_{\mathrm{FA}}$, this modification of the Proof of Theorem 2.3.6 gives the following variant.

Theorem 2.3.9 (LRMC net of peak charge for continuous outputs). *For every $y \geq 0$, $p \in \partial_y C^{\mathrm{N}}_{\mathrm{LR}|\mathcal{C}}(y,r,w)$ if and only if $p \in \mathcal{C}[0,T]$ and:*

1. *$p_{\mathrm{AC}}(t) = p_{\mathrm{N}} - \nu$ for some p_{N} and ν in $L^1_+[0,T]$ that meet Conditions 1a–1c of Theorem 2.3.6.*

2. *$p_{\mathrm{Si}} \leq 0$ and $\mathrm{supp}(p_{\mathrm{Si}}) \subseteq \{t : y(t) = 0\}$.*

The other term of the LRMC is the peak charge, which is formally a subgradient of Max as a function on the space $\mathcal{C}[0,T]$ paired with its norm-dual $\mathcal{M}[0,T]$. The following description of $\partial\,\mathrm{Max}(y)$ as a unit measure concentrated on the *exact* peaks of y is in, e.g., [24, Example 4.5] and [51, 4.5.2]. At a continuous y with a single peak, $\partial\,\mathrm{Max}(y)$ is therefore single-valued. (For comparison, $\partial\,\mathrm{EssSup}(y)$ is multi-valued at every $y \in L^{\infty}$, as is noted in, e.g., [85, 4.4.8].)

Theorem 2.3.10 (LRMC peak charge for continuous outputs). *For every $y \in \mathcal{C}[0,T]$, $\gamma \in \partial\,\mathrm{Max}(y)$ if and only if $\gamma \in \mathcal{M}[0,T]$ and:*

1. *$\gamma \geq 0$ and $\gamma[0,T] = 1$.*

2. *γ is concentrated on $\mathrm{ArgMax}(y)$, i.e.,*

$$\mathrm{supp}(\gamma) \subseteq \mathrm{ArgMax}(y) := \{t \in [0,T] : y(t) = \mathrm{Max}(y)\}.$$

Corollary 2.3.11 (Unique LRMC peak charge). *If $y \in \mathcal{C}[0,T]$ and $\mathrm{ArgMax}(y) = \{t\}$ then $\partial\,\mathrm{Max}(y) = \{\varepsilon_t\}$, i.e., it is the Dirac measure at t. So Max is Gateaux-differentiable at y (but it is not Fréchet differentiable).*

Proof. By Theorem 2.3.10, $\partial\,\mathrm{Max}(y)$ is the singleton, and Gateaux-differentiability at y follows: see, e.g., [32, 7E] or [51, 4.2.1: Example 1] or [80, 5.37]. □

Comment: That Max is not Fréchet-differentiable is noted in, e.g., [85, 4.4.4]. It is readily seen by considering an increment Δy, of unit norm, that equals 0 at t (the maximum point of y) but equals 1 on an interval which approaches t (as the increment varies). The first-order approximation to Max $(y + \epsilon \Delta y)$ is then Max $(y) + 0$, and the supremum of its error (over Δy) is ϵ—so it fails to be of a lower order than ϵ. This exemplifies the difference between Gateaux and Fréchet differentiability of convex functions on infinite-dimensional normed spaces. (For *convex* functions on $\mathbb{R}^n$, the two concepts are equivalent.)

2.4 Profit-maximising operation and valuation of a thermal plant

Given a TOU electricity tariff p, the profit-maximising operation of a thermal system with capacities and running costs

$$(k_1, \ldots, k_\Theta) \text{ and } (w_1, \ldots, w_\Theta)$$

is defined by the SRMC curve, since this is also the system's short-run supply curve: formally, the short-run instantaneous supply correspondence is the inverse of the instantaneous SRMC correspondence (2.3.9), so the two have the same graph (Figure 2.1b). Another way to obtain the system's supply correspondence is to sum, over θ, the supply correspondences of the individual plants, which are:

$$S(\mathsf{p}, k_\theta, w_\theta) := \begin{cases} \{0\} & \text{if } \mathsf{p} < w_\theta \\ [0, k_\theta] & \text{if } \mathsf{p} = w_\theta \\ \{k_\theta\} & \text{if } \mathsf{p} > w_\theta \end{cases} \quad \text{for } \theta = 1, \ldots, \Theta. \tag{2.4.1}$$

This means that, given a price function $p \in L^1[0, T]$, a profit-maximising output trajectory for plant θ is a selection from the correspondence

$$t \mapsto S(p(t), k_\theta, w_\theta) \tag{2.4.2}$$

and the system's profit-maximising output is obtained by adding up the plants' outputs over θ.

When the price system p lies in the larger price space $L^{\infty *}[0, T]$, there may be no profit-maximising output, but any optimal output remains optimal after replacing p by its density part p_{CA}—which narrows down the search for any profit maxima at p. This can be shown by a duality argument (Corollary 2.4.2). The dual programme can also be used to value the capacity, although the thermal technology is so simple that the marginal capacity value can be obtained by differentiating the short-run profit function directly. These results are formalised next. It is assumed that $p \geq 0$ (since this is usually the case in general equilibrium); recall that p has the Hewitt-Yosida decomposition (2.2.1) into $p_{\mathrm{CA}} + p_{\mathrm{FA}}$. For the rest of this section (except its final Comment), k and w are scalars (i.e., characteristics of a single plant, and not of a whole system of plants).

The linear programme of maximising the operating profit of a single thermal plant of capacity k with a unit running cost w is:

$$\text{Given } (p,k,w) \in L_{+}^{\infty*}[0,T] \times \mathbb{R}_{+} \times \mathbb{R}_{+} \tag{2.4.3}$$

$$\text{maximise } \int_0^T (p_{\mathrm{CA}}(t) - w)\, y(t)\,\mathrm{d}t + \langle p_{\mathrm{FA}} \,|\, y \rangle \text{ over } y \in L^{\infty}[0,T] \tag{2.4.4}$$

$$\text{subject to: } \quad 0 \le y(t) \le k \quad \text{for almost every } t. \tag{2.4.5}$$

Its optimal value is the (maximised) *one-station operating profit* $\Pi_{\mathrm{SR}}^{\mathrm{Th}}(p,k,w)$; its solution set is denoted by $\hat{Y}_{\mathrm{Th}}(p,k,w)$. When $p \in L^1$, this is the set of all selections from (2.4.2); if additionally $p(t) \ne w$ for a.e. t then the solution *is* unique, in which case it is denoted by the lowercase $\hat{y}_{\mathrm{Th}}(p,k,w)$. The profit-maximising output of a thermal system $(k_\theta)_{\theta=1}^{\Theta}$ is $\sum_\theta \hat{Y}_{\mathrm{Th}}(p,k_\theta,w_\theta)$, and its (maximum) total operating profit is $\sum_\theta \Pi_{\mathrm{SR}}^{\mathrm{Th}}(p,k_\theta,w_\theta)$.

The standard dual of a plant's operation LP is obtained—in the way described in, e.g., [73, Examples 1' and 4']—from the standard parameterisation of the primal constraints, which consists here in adding time-dependent increments $(\Delta k(t), \Delta n(t))$ to the constants $(k,0) \in \mathbb{R} \times \mathbb{R}$ in (2.4.5). Like the capacity increment Δk, the negative of the increment Δn to the zero floor for the output in (2.4.5) can be thought of as a resource increment. The increments $(\Delta k, -\Delta n) \in L^\infty \times L^\infty$ are paired with Lagrange multipliers $(\kappa,\nu) \in L^{\infty*} \times L^{\infty*}$. Thus, by considering a separate increment $\Delta k(t)$ for each t, one can impute an instantaneous value, $\kappa(t)$, to capacity services at any time t, if $p \in L^1[0,T]$. (When $p \in L^{\infty*} \setminus L^1$, this has to be formally rephrased as imputation of the value $\langle \kappa \,|\, 1_A \rangle$ to capacity services on any time subinterval $A \subset [0,T]$.) Similarly, $\nu(t)$ is the loss of profit from raising the output floor by a unit, at time t (i.e., from perturbing the constraint $y(t) \ge 0$ to $y(t) \ge 1$). The standard dual is therefore the following programme for the flow of capacity rent κ and for ν (the Lagrange multiplier for the nonnegativity constraint on y):

$$\text{Given } (p,k,w) \text{ as in (2.4.3)} \tag{2.4.6}$$

$$\text{minimise } k\left\langle \kappa \,|\, 1_{[0,T]} \right\rangle \text{ over } \kappa \in L^{\infty*}[0,T] \text{ and } \nu \in L^{\infty*}[0,T] \tag{2.4.7}$$

$$\text{subject to: } \quad \kappa \ge 0,\ \nu \ge 0 \tag{2.4.8}$$

$$p - w = \kappa - \nu. \tag{2.4.9}$$

As is spelt out next, the dual solution exists, and it is unique if $k > 0$, is (the term p_{FA}^{-} vanishes if $p \ge 0$, as can be assumed with free disposal).

Proposition 2.4.1 (Dual solution and optimality conditions). *Assume that* $p \ge 0$. *Then:*

1. For every $k > 0$,[13] *the dual programme of capacity value minimisation (2.4.6)–(2.4.9) has a*

[13] When $k = 0$, the only feasible output is $y = 0$. The primal and dual optimal values (the operating profit and the capacity value) are both zero, and every feasible (κ,ν) is a nonunique dual optimum (so (2.4.10)–(2.4.11) is a dual solution also then).

unique optimal solution, viz.,

$$\hat{\kappa}_{\mathrm{Th}}(p,w) = (p-w)^{+} = (p_{\mathrm{CA}} - w)^{+} + p_{\mathrm{FA}} \tag{2.4.10}$$

$$\hat{\nu}_{\mathrm{Th}}(p,w) = (p-w)^{-} = (p_{\mathrm{CA}} - w)^{-}. \tag{2.4.11}$$

The programme's optimal value—i.e., the thermal capacity value—is therefore

$$k \int_0^T (p_{\mathrm{CA}}(t) - w)^{+} \, \mathrm{d}t + k \left\langle p_{\mathrm{FA}} \mid 1_{[0,T]} \right\rangle \tag{2.4.12}$$

and it is equal to the plant's short-run profit $\Pi_{\mathrm{SR}}^{\mathrm{Th}}(p,k,w)$, *which is the optimal value of (2.4.3)–(2.4.5).*

2. *A* $y \in L^{\infty}$ *is an optimal solution to the primal (2.4.3)–(2.4.5) if and only if:*

   (a) *y is feasible, i.e.,* $0 \le y \le k$.

   (b) *(i)* $y = 0$ *almost everywhere on* $\{t : p_{\mathrm{CA}}(t) < w\}$, *(ii)* $y = k$ *almost everywhere on* $\{t : p_{\mathrm{CA}}(t) > w\}$, *and (iii) for every number* $\epsilon > 0$, p_{FA} *is concentrated on the set* $\{t : y(t) \ge k - \epsilon\}$.

Proof. Part 1 is nearly obvious: the dual constraints (2.4.8) and (2.4.8) mean that κ and ν must equal $(p-w)^{+} + \sigma$ and $(p-w)^{-} + \sigma$ for some $\sigma \ge 0$ (viz., for $\sigma = \kappa \wedge \nu$). Since $k > 0$, (2.4.7) is minimised if and *only* if $\sigma = 0$. And this translates into (2.4.10) and (2.4.11) because $p \ge 0$. So the dual value is (2.4.12). That this is also the primal value can be shown directly, but it also a case of a general result given in, e.g., [73, Theorem 17 (a)]: when $k > 0$, the primal constraints meet the generalised Slater's Condition of [73, (8.12)] with the supremum norm on the primal parameter space $L^{\infty} \times L^{\infty}$ (since L^{∞}_{+} has a nonempty interior).

For Part 2, apply the Kuhn-Tucker saddle-point characterisation of optima—given in, e.g., [73, Theorem 15 (e) and (f)]—which, as with any LP and its standard dual, translates into the conjunction of feasibility and complementary slackness. Here, primal feasibility is Condition 2b, and complementary slackness means that, for every number $\epsilon > 0$, $\hat{\kappa}_{\mathrm{Th}}(p,w)$ is concentrated on $\{t : y(t) \ge k - \epsilon\}$, and $\hat{\nu}_{\mathrm{Th}}(p,w)$ is concentrated on $\{t : y(t) \le \epsilon\}$. Since $\hat{\kappa}_{\mathrm{Th}}$ and $\hat{\nu}_{\mathrm{Th}}$ are given by (2.4.10) and (2.4.11), the concentration conditions translate into Condition 2b. □

As is shown next, it follows that the singular part, p_{FA}, can be removed from a price system supporting a plant's output. Although this lowers the plant's rent by the second term in (2.4.12), any optimal output continues to be so: investment may cease to be profitable, but this has no effect on the operation of existing plant.

Corollary 2.4.2. $\hat{Y}_{\mathrm{Th}}(p,k,w) \subseteq \hat{Y}_{\mathrm{Th}}(p_{\mathrm{CA}},k,w)$ *for every* $p \in L^{\infty *}[0,T]$, *i.e., if p supports a* $y \in L^{\infty}[0,T]$ *as a profit-maximising output of a thermal plant, then so does* p_{CA}.

Proof. Any $y \in \hat{Y}_{\mathrm{Th}}(p,k,w)$ and the dual solution $(\hat{\kappa}_{\mathrm{Th}}(p,w), \hat{\nu}_{\mathrm{Th}}(p,w))$—abbreviated to $(\hat{\kappa},\hat{\nu})$—meet the Kuhn-Tucker Conditions 2a and 2b of Proposition 2.4.1. It readily follows that, after p has been replaced by p_{CA}, the Kuhn-Tucker Conditions hold for the same y with $(\hat{\kappa}_{\mathrm{CA}}, \hat{\nu}_{\mathrm{CA}})$ instead of $(\hat{\kappa},\hat{\nu})$. This shows that $y \in \hat{Y}_{\mathrm{Th}}(p_{\mathrm{CA}},k,w)$ and, also, that $(\hat{\kappa}_{\mathrm{CA}}, \hat{\nu}_{\mathrm{CA}})$ is the dual solution (now that the output price is p_{CA}). □

An output that maximises the operating profit obviously exists if the price system is a pure density function (by Part 2 of Proposition 2.4.1 with $p_{\mathrm{FA}} = 0$). If the price does have a singular part, then an optimal output exists if (and only if) the singular charge comes at a time when the price rate of the density part is not less than the unit running cost (and is thus consistent with an output rate equal to capacity).

Corollary 2.4.3. *Assume that $k > 0$ and $p \geq 0$. Then: $\hat{Y}_{\mathrm{Th}}(p,k,w) \neq \emptyset$ if and only if p_{FA} is concentrated on $\{t : p_{\mathrm{CA}}(t) \geq w\}$.*

Proof. Since $p \geq 0$, $p_{\mathrm{FA}} = \hat{\kappa}_{\mathrm{FA}}$ by the formula for $\hat{\kappa} = \hat{\kappa}_{\mathrm{Th}}(p,w)$. Fix any positive $\epsilon < k$. If $y \in \hat{Y}_{\mathrm{Th}}(p,k,w)$ then, by Condition 2b of Proposition 2.4.1, $p_{\mathrm{FA}} = \hat{\kappa}_{\mathrm{FA}}$ is concentrated on $\{t : y(t) \geq k - \epsilon\}$ and *a fortiori* on $\{t : y(t) > 0\}$. And this set is contained in $\{t : p_{\mathrm{CA}} \geq w\}$ because $(p_{\mathrm{CA}} - w)^{-}$ equals $\hat{\nu}$, which vanishes outside $\{t : y(t) = 0\}$.[14]

For the converse, one profit-maximising output is

$$y(t) = \begin{cases} 0 & \text{if } p(t) < w \\ k & \text{if } p(t) \geq w \end{cases}$$

because, with p_{FA} nonnegative and concentrated on $\{t : p_{\mathrm{CA}}(t) \geq w\}$, it gives

$$\langle p \,|\, y \rangle = k \int_0^T (p_{\mathrm{CA}}(t) - w)^{+} \, \mathrm{d}t + \langle p_{\mathrm{FA}} \,|\, k \rangle = \Pi_{\mathrm{SR}}^{\mathrm{Th}}(p,k,w)$$

by (2.4.12). □

Comment: The SRP function can be used to extend the Wong-Viner Envelope Theorem to the case of convex but nondifferentiable cost functions—such as the thermal SRC and LRC. The naive extension is false: an SRMC of an optimal system $k = (k_\theta)_{\theta=1}^{\Theta}$ need not be an LRMC, i.e., when $p \in \partial_y C_{\mathrm{SR}}(y,k,w)$ and $k \in \check{K}(y,k,w)$, it does *not* follow that $p \in \partial_y C_{\mathrm{LR}}(y,r,w)$. This is readily seen with the thermal technology (even with the one-station technology).[15] But if $p \in \partial_y C_{\mathrm{SR}}(y,k,w)$ and $r \in \widehat{\partial}_k \Pi_{\mathrm{SR}}(p,k,w) := -\partial_k(-\Pi_{\mathrm{SR}})(p,k,w)$,[16] then it *does* follow that $p \in \partial_y C_{\mathrm{LR}}(y,r,w)$, and

[14] Without assuming that $p \geq 0$, the same argument shows that p_{FA}^{+} is concentrated on $\{t : p_{\mathrm{CA}}(t) \geq w\}$ and, similarly, that p_{FA}^{-} is concentrated on $\{t : p_{\mathrm{CA}}(t) \leq w\}$.

[15] It is the *reverse* inclusion between the cost subdifferentials that always holds (for any technology), i.e., if $k \in \check{K}(y,r,w)$ then $\partial_y C_{\mathrm{LR}}(y,r,w) \subseteq \partial_y C_{\mathrm{SR}}(y,k,w)$.

[16] This is the superdifferential of Π_{SR} as a concave function of k.

also that $k \in \check{K}(y, r, w)$. In other words, if the profit-imputed marginal values of the fixed inputs are equal to their rental prices, then an SRMC *is* an LRMC. When $p \in \partial_y C_{\mathrm{SR}}$, the valuation condition is stronger than cost-optimality of the fixed inputs (which is why it works); for the thermal technology (when $p \in L^1$ and $k \gg 0$), it reads: $r_\theta = \int_0^T (p(t) - w_\theta)^+ \, dt$ for each θ. This is a case of the extension of the Wong-Viner Theorem in Section 5.9 and in [46].

2.5 Conclusions from Chapter 2

A continuous-time model of peak-load pricing can be adequately set up in the commodity space of essentially bounded functions but the smaller commodity space of continuous functions affords a better representation of the instantaneous capacity charges that arise in the case of pointed peaks. These charges are terms of the marginal costs that come from the capacity cost or constraint in the long-run or short-run cost function. Since the costs are convex but nondifferentiable functions of the output bundle, the subdifferential must be used to formalise the concept of marginal cost. For the thermal technology of electricity generation, the cost functions can be expressed by formulae that can be subdifferentiated by using the Hardy-Littlewood-Pólya theory of rearrangements and majorisation. Thus a cost-based analysis of the supply side of the long-run competitive equilibrium problem is feasible with a purely thermal technology. This ceases to be the case once other technologies, such as hydro or pumped storage, are included in the model. But, as is shown elsewhere, the short-run profit maximisation problem for those types of plant is still tractable, and its counterpart for a thermal plant is very simple (as has been shown here). This provides a basis for the short-run profit approach to the long-run equilibrium.

Chapter 3

EFFICIENCY RENTS OF PUMPED-STORAGE PLANTS AND THEIR USES FOR OPERATION AND INVESTMENT DECISIONS

3.1 Introduction to Chapter 3

The problems of optimal operation and rental valuation of storage facilities for cyclically priced goods have been studied mainly in the context of hydroelectric generation by, among others, Koopmans [55] and Bauer et al. [7]. The corresponding questions for pumped storage of energy and other goods have received less attention, and existing models of such technologies lack verisimilitude.[1] To fill this gap, a realistic but tractable model of pumped storage (PS) is set up, and plant operation and valuation is analysed in the framework of short-run profit maximisation. Given a time-of-use (TOU) market price $p(t)$ for the good in question (say, electricity), an optimal TOU value $\psi(t)$ is imputed to the stock (of energy converted to a storable form). This essentially solves the operation problem: see (3.8.2). It therefore makes sense to value the plant's capacities by their marginal contributions to the maximum operating profit, $\Pi_{\mathrm{SR}}^{\mathrm{PS}}$; and these efficiency rents can be expressed in terms of p and ψ (Theorem 3.9.2). The rental values can serve as guides to investment (Section 3.11).

When the given tariff p is a continuous function of time, the stock's shadow price function ψ is unique, either literally or at least at the times which matter for capacity valuation. It follows that the capital inputs—viz., the reservoir and the converter or "pump-turbine"—have definite and separate marginal values, $\partial\Pi/\partial k_{\mathrm{St}}$ and $\partial\Pi/\partial k_{\mathrm{Co}}$. Their ratio gives a well-defined rate of substitution in product-value terms. This is a striking property because the inputs are also perfect complements—in the sense that no input substitution is possible after fixing the output bundle. That is, the conditional input demands for the storage and conversion capacities depend *only* on the trajectory of net output from storage, $y(t)$, over the cycle $[0, T]$.

That perfect complements can substitute for each other may seem paradoxical, and of course it would be impossible with a homogeneous, one-dimensional output good: in such a case the output from an input bundle k could only have the familiar fixed-coefficients form $\min\{k_1, k_2, \ldots\}$. But with a multi-dimensional, differentiated output good, perfect complementarity would imply fixed input

[1] The existing literature disregards one or both of the main factors in pumped storage, viz., the storage capacity cost and the conversion capacity cost. Pyatt [68, p. 752, (10)] assumes that there is no capacity constraint on the stock. Nguyen [65, pp. 242–243] excludes both types of capacity cost and concentrates on the running cost (which is of little importance in pumped storage). Gravelle [30] limits his treatment to a two-subperiod model which loses the distinction between the different kinds of storage costs.

proportions *only* if the output proportions were fixed—and they are not. With output proportions (as well as scale) allowed to vary, it is the output price system p that aggregates the output bundle y into a scalar, viz., the revenue; and, given a suitable p, substitution in revenue terms is possible. With multiple outputs, the inputs can be perfect complements without, like a nut and bolt, having to be used in a fixed proportion.

The problem of maximising the operating profit of a storage plant can be formulated as a linear programme (3.4.5)–(3.4.10). Its dual (3.5.1)–(3.5.6) is the problem of minimising the plant's value subject to a constraint which decomposes the given price $p(t)$ into a sum of the values of the plant's capital services (plus a constant λ). The dual can be reformulated as a problem of shadow pricing for the stock (3.7.7)–(3.7.9); this change of variables makes the analysis more transparent and leads to new insights.

The imputed capacity values are useful in planning investment, either as an expansion of existing plants or as a large-scale development of new sites. The investment problem is formulated and it is shown how to solve it for the optimal capacities by equating their marginal values to their marginal costs: see (3.11.1)–(3.11.2) and (3.11.6)–(3.11.7). It is worth noting that the marginal values are, explicitly or implicitly, essential for *any* profit-based appraisal of investment plans. Even a comparison of just two alternatives, k' and k'', requires the knowledge of $\Pi_{\mathrm{SR}}^{\mathrm{PS}}(p,k)$ for $k = k', k''$, but there is no explicit formula for $\Pi_{\mathrm{SR}}^{\mathrm{PS}}$ (except with the crudest of tariffs, such as the two-valued p of Example 3.15.1). By contrast, once the marginals $\nabla_k \Pi$ are known, the total profit can be evaluated as $\Pi(k) = \nabla_k \Pi \cdot k$ by Euler's Theorem.[2] This is what is done when Π is calculated as the dual value (since the dual solution is equal to the marginal value): see Section 3.9. And although Π could be evaluated as the primal value, the successful algorithms exploit duality and provide the dual solution along with the primal one.

For its general approach—viz., a continuous-time treatment of storage rents—this study takes inspiration from Koopmans' pioneering paper [55] on optimal water storage policies for a hydro-thermal electricity generating system. In all other respects, however, this work is different. One of its main purposes is to provide a flexible, general framework for dealing with a whole class of problems, whilst Koopmans' analysis is limited to hydroelectric storage—i.e., the storage of a given, natural inflow—and it does not readily extend to similar technologies such as pumped storage. Furthermore, the profit-imputed rental values are unique—unlike Koopmans' rents, which are typically nonunique as a result of being imputed from the saving on the (thermal) operating cost. Also, the dual to the profit maximisation programme is a simple and direct way of deriving the marginal values, whereas Koopmans' rents are given in terms of a complex operating solution: they do serve his main purpose—which is to verify the cost-optimality of the storage policy he constructs—but the nonuniqueness and complexity of the construction are obstacles to their use in practical investment

[2] This identity can also be used to divide the plant's total rent between the fixed inputs on marginalist principles.

analysis.

In the short-run cost-minimisation framework, a production technique with practically no operating cost, such as energy storage, can be studied only in conjunction with others that do have variable costs—such as the thermal fuel cost in Koopmans' problem. By contrast, the profit-maximum formulation allows such a technique to be analysed separately; and this approach is better suited to the more decentralised structure of today's utilities. Also, the switch from cost minimisation to profit maximisation is actually essential for removing the indeterminacy of marginal capacity values. This is because valuations of the storage plant's capacities depend on two time-of-use (TOU) prices, $p(t)$ for the marketed good and $\psi(t)$ for the stock. In the cost-minimum approach, both commodities must be shadow-priced, and both p and ψ can be to some extent indeterminate. But in competitive maximisation of the short-run profit, the good's price function p is treated as given. And a possible indeterminacy of ψ is excluded (at least at the times which matter) by a problem-specific argument, viz., an examination of the Lagrange multipliers for the capacity constraints (Lemma 3.13.1 with perfect conversion, and Lemma 3.13.3 with imperfect conversion).

Profit-imputed values of capital inputs are useful in investment calculations not only to privately-owned industry, but also to a publicly-owned (or regulated) utility which aims to meet the demand, price its outputs at long-run marginal cost and optimise its capital stock. The utility can achieve this by meeting the demand at short-run marginal cost and adjusting its capital inputs until their rental prices are equal to their marginal values. But these values must be imputed by the short-run profit, and not by the short-run cost if the latter is nondifferentiable (as is the case in peak-load pricing): see Chapter 5.[3]

Time-continuity is not just a natural assumption on the good's price p and the only one to guarantee uniqueness of the imputed capacity values:[4] it is also an assumption that is verified for the competitive equilibrium price in [45], where the price function is *proved* to be continuous for a class of problems that includes peak-load pricing with storage. The general equilibrium model is set up in a commodity space of bounded functions of time. In part, it is an application of Bewley's framework [10], which is adapted and extended in [49], [43] and [47]; some mathematical tools are provided in [34], [35], [36] and [42]. It is hoped that this will lead to an integration of hitherto largely separate economic, engineering and OR studies of topics such as peak-load pricing and energy storage.

Section 3.2 describes the technology. Formal analysis is preceded by heuristics, in Section 3.3. In Sections 3.4 and 3.5, the short-run profit maximum problem and its dual are set up as linear

[3] In particular, to extend the Wong-Viner Theorem to the case of nondifferentiable costs, it is insufficient just to maintain the usual assumption of fixed-input cost-optimality (i.e., total-cost minimisation). It must be strengthened to equality between the inputs' rental prices and their values imputed by the short-run profit (not cost).

[4] Discretisation of time is, however, necessary in solving the relevant programmes by standard numerical methods. In this context, uniqueness of the continuous-time solution ensures that the approximate solutions converge as the discretisation is refined.

programmes (LPs) that are doubly infinite: with continuous-time dating of commodities, the primal (3.4.5)–(3.4.10) contains a continuum of output variables y and also a continuum of dated capacity constraints (on the flow y and on the stock s). The primal and the dual are shown to be soluble, and their optimal solutions are characterised in Section 3.6. In Section 3.7, the dual LP is reformulated as an unconstrained convex programme (CP) for shadow pricing of stock. Sections 3.8 and 3.9 give formulae, in terms of an optimal stock price ψ, for the optimal output $\hat{y}$ and for the operating profit $\Pi_{\mathrm{SR}}^{\mathrm{PS}}$ and its derivatives w.r.t. the reservoir and conversion capacities. These marginal values are the basis for calculating the optimum investment, in Section 3.11. This uses the bounds on the marginals which are established in Section 3.10. This completes the core matter, which is followed by proofs. Proofs for Sections 3.5 to 3.8 are gathered in Section 3.12. Proofs for Section 3.9 are given in Section 3.13, along with the required auxiliary results.

The rest of this chapter consists of various supplements. In Section 3.14, the optimal output $\hat{y}$ is shown to be invariant under monotone transformations of the price function p, i.e., $\hat{y}$ depends on the ranking pattern of p but *not* on the distribution of p. Also, the dual (shadow-pricing) problem is reformulated by using a distance concept known from the Monge-Kantorovich mass transfer problem (a.k.a. the transportation problem). Section 3.15 presents a counterexample to the existence of $\nabla_k \Pi$ when the price p is a discontinuous step function (so that time is effectively a discrete variable). Section 3.16 verifies the conditions for including the storage technology in an equilibrium model with the commodity space of bounded functions, $L^\infty [0, T]$.

Table 3.1 summarises the notation.

3.2 Pumped-storage technology

Consider a cyclically priced good that, once put in storage, can be held at no running cost (or loss of stock), as long as the stock does not exceed the *reservoir's capacity*, k_{St}. The reservoir is charged and discharged with converters; the equipment is so called because the good itself is actually nonstorable (or too costly to store directly), and so it must first be converted into a storable *medium*. This is a purely intermediate commodity, useful only for reconversion to the original good. Examples include gas liquefaction and conversion of electricity to a storable form of energy: in both cases the running cost of storage is negligible. Each of these techniques is referred to as *pumped storage* (PS), irrespectively of the particular good (AC electricity, natural gas), the medium (DC electricity, potential or other energy; liquid gas) and the corresponding devices.[5]

A nonreversible charger or discharger is termed a *pump* or a *turbine* (Pu or Tu, respectively); this terminology originates from energy storage (ES). Some conversion processes involve considerable losses; and the "round-trip" technical efficiency is $\eta_{\mathrm{Ro}} := \eta_{\mathrm{Pu}} \eta_{\mathrm{Tu}} \leq 1$, where η_{Pu} and η_{Tu} are the

[5] AC/DC means alternating/direct current.

Role in programme pair	Variable	Notation	Unit
Price data (dual parameters)	electricity price at time t	$p(t)$	\$/kWh
Quantity data (primal parameters)	reservoir capacity energy-stock floor turbine capacity turbine's output floor pump capacity pump's output floor top-up of stock	$k_{\mathrm{St}}(t) = \text{const.}$ $n_{\mathrm{St}}(t) = 0$ $k_{\mathrm{Tu}}(t) = \text{const.}$ $n_{\mathrm{Tu}}(t) = 0$ $k_{\mathrm{Pu}}(t) = \text{const.}$ $n_{\mathrm{Pu}}(t) = 0$ $\zeta = 0$	kWh kWh kW kW kW kW kWh
Quantity decisions (primal variables)	turbine's output rate pump's output rate (at time t) energy stock at time 0 and T	$y_{\mathrm{Tu}}(t)$ $y_{\mathrm{Pu}}(t)$ s_0	kW kW kWh
Derived quantities	plant's net output rate at time t rate of outflow from reservoir at time t energy stock at time t	$y(t) := y_{\mathrm{Tu}}(t) - y_{\mathrm{Pu}}(t)$ $f(t) := \frac{y_{\mathrm{Tu}}(t)}{\eta_{\mathrm{Tu}}} - \eta_{\mathrm{Pu}} y_{\mathrm{Pu}}(t)$ (η_{Tu}, η_{Pu} are efficiencies) $s(t) := s_0 - \int_0^t f(\tau)\,\mathrm{d}\tau$	kW kW (100%) kWh
Shadow prices (dual decision variables, paired to primal parameters)	unit reservoir value on interval of length $\mathrm{d}t$ value of energy-stock floor (nonnegativity constraint) unit turbine value at time t value of turb.'s output floor (nonnegativity constraint) unit pump value at time t value of pump's output floor (nonnegativity constraint) energy-stock value at 0 and T	$\kappa_{\mathrm{St}}(\mathrm{d}t)$ $\nu_{\mathrm{St}}(\mathrm{d}t)$ $\kappa_{\mathrm{Tu}}(t)$ $\nu_{\mathrm{Tu}}(t)$ $\kappa_{\mathrm{Pu}}(t)$ $\nu_{\mathrm{Pu}}(t)$ λ	\$/kWh \$/kWh \$/kWh \$/kWh \$/kWh \$/kWh \$/kWh
Derived valuations	energy-stock value at time t total reservoir rent for whole cycle $[0,T]$ total turbine rent total pump rent	$\psi(t) = \lambda + (\kappa_{\mathrm{St}} - \nu_{\mathrm{St}})[0,t]$ $\kappa_{\mathrm{St}}[0,T] = \int_0^T \kappa_{\mathrm{St}}(\mathrm{d}t)$ $\int_0^T \kappa_{\mathrm{Tu}}(t)\,\mathrm{d}t$ $\int_0^T \kappa_{\mathrm{Pu}}(t)\,\mathrm{d}t$	\$/kWh \$ \$ \$

TABLE 3.1. Notation for Chapter 3. Some functions of time (k_{St}, etc.) are equated to "const.". This indicates that they are constants in the original, unperturbed programme, but are perturbed with time-varying increments ($\Delta k_{\mathrm{St}}(t)$, etc.) to interpret the time-dependent dual variables ($\Delta\kappa_{\mathrm{St}}$, etc.). The duality scheme (Section 5) similarly uses a nonzero increment $\Delta\zeta$ to $\zeta = 0$ (paired with the dual variable λ).

one-way rates of transformation, of the good into the medium and *vice versa.* Both transformations are taken to be instantaneous (although a constant lag can be readily taken into account). Also, both η_{Pu} and η_{Tu} are assumed to be constant. With the stock $s(t)$ measured in potential terms—i.e., as the amount of the good that it would yield after a perfect transformation—both η_{Pu} and η_{Tu} are dimensionless numbers between 0 and 1. In the case of perfect conversion $\eta_{\mathrm{Pu}} = \eta_{\mathrm{Tu}} = 1$ (i.e., $\eta_{\mathrm{Ro}} = 1$).

The *pump* or *turbine capacity*, k_{Pu} or k_{Tu}, is its maximum output rate (i.e., the rate of inflow to reservoir or outflow from plant). In other words, in unit time a unit pump can turn $1/\eta_{\mathrm{Pu}}$ units of the good into 1 unit of the storable medium; and a unit turbine can turn $1/\eta_{\mathrm{Tu}}$ units of the medium into 1 unit of the good.

It suffices to analyse the case of nonreversible equipment. The reversible case is readily deduced, but it is spelt out for completeness; and henceforth a *converter* (Co) means *reversible* equipment, capable of transforming both ways, though not necessarily at the same rate. A converter's capacity, k_{Co}, is by convention defined as the maximum output rate in the *charging* mode. A unit converter provides, then, a unit of charging capacity, whilst its *discharging* capacity is some $\beta > 0$. (In other words, a unit converter is operationally equivalent to a unit pump together with β units of turbine: in unit time a unit converter can either turn $1/\eta_{\mathrm{Pu}}$ units of the good into 1 unit of the storable medium, or turn β/η_{Tu} units of the medium into β units of the good.) The converter is termed *symmetric* if $\beta = 1$.

Energy storage techniques include pumped-water energy storage (PWES), in which electricity (the good) is used to pump water from the lower to the upper reservoir, and the accumulated potential energy (the medium) is reconverted by releasing the water through a turbine-generator. Compressed-air energy storage (CAES) is similar: air is pumped under pressure into a reservoir (such as an underground cavern), and it is later let out through a gas turbine. In both techniques the converter is usually a reversible pump-turbine, although nonreversible multi-stage pumps have also been used in high-head PWES plants. Another ES technique is the superconducting magnetic energy storage (SMES), in which AC electricity (the good) is converted by a reversible inverter into DC electricity (the medium), to be stored in a superconductive coil. There is also battery storage (of DC electricity) and inertial storage (of kinetic energy, in a flywheel). With each of these techniques, k_{St} can be measured in kWh (of the intermediate form of energy); both k_{Tu} and k_{Pu} can be measured in kW (of, respectively, electric and intermediate forms of power). In PWES and CAES, η_{Ro} is typically around 70% to 75% (i.e., 0.7 kWh of electricity is recovered from a kWh used up): see [60, p. 89]. In SMES, η_{Ro} is over 95%, with β close to 1: see [63].

Storage is studied here as a large-scale technique to be used for profit maximisation (or cost minimisation). For this purpose the Electricity Supply Industry (ESI) uses at present mainly pumped-water and compressed-air plants, but superconductive coils and batteries may also become economical for bulk storage. Another use of storage plants is as fast-response emergency backups for control of quality (frequency and voltage) when thermal generators break down or there is an unanticipated surge in demand. Start-up times of PWES and CAES plants are around 1 to 5 minutes (like those of hydro and gas turbines): see, e.g., [60, Table 8.2].[6] SMES coils and batteries are thousands of times faster to respond (switching from charging to discharging in 4 to 20 ms). This makes them unrivalled for elimination of brief outages and quality disturbances; and small or medium-scale SMES devices are used by both suppliers and users of electricity to ensure transmission stability and uninterruptible power supply to sensitive equipment: see [13] and [63]. These are important applications but, being specific to electrical engineering, they are left out of this analysis.

The rate of outflow of the good, from the plant's turbine to the market, is denoted by $y_{\mathrm{Tu}} \geq 0$; the inflow (from market to the plant's pump) is $y_{\mathrm{Pu}} \geq 0$. The storage plant's net output rate is therefore the signed, periodic function

$$y = y_{\mathrm{Tu}} - y_{\mathrm{Pu}}$$

defined on a time interval $[0, T]$ which represents one price cycle. The pair $(y_{\mathrm{Tu}}, y_{\mathrm{Pu}})$ is termed a *storage policy*. When $\eta_{\mathrm{Ro}} < 1$, it is convenient to allow an overlap of y_{Tu} and y_{Pu}. This is a purely formal trick that does not require simultaneous charging and discharging to be actually feasible, since this could never be optimal if the good's prices are positive (Lemma 3.8.2).

The nonnegative and nonpositive parts of the output, y^+ and y^-, represent the outflow of the good (from plant to market) and the inflow (from market to plant). Note that

$$y_{\mathrm{Tu}} - y^+ = y_{\mathrm{Pu}} - y^- = y_{\mathrm{Tu}} \wedge y_{\mathrm{Pu}} \geq 0$$

where $\wedge$ means the smaller of the two. The associated flows of the medium, from reservoir to turbine and from pump to reservoir, are

$$f_{\mathrm{Tu}} = \frac{y_{\mathrm{Tu}}}{\eta_{\mathrm{Tu}}} \quad \text{and} \quad f_{\mathrm{Pu}} = \eta_{\mathrm{Pu}} y_{\mathrm{Pu}}. \tag{3.2.1}$$

The signed outflow from the reservoir is therefore

$$f := f_{\mathrm{Tu}} - f_{\mathrm{Pu}} = \left(\frac{y^+}{\eta_{\mathrm{Tu}}} - \eta_{\mathrm{Pu}} y^-\right) + \left(\frac{1}{\eta_{\mathrm{Tu}}} - \eta_{\mathrm{Pu}}\right) (y_{\mathrm{Tu}} \wedge y_{\mathrm{Pu}}). \tag{3.2.2}$$

This shows that an "overlapping" policy $(y_{\mathrm{Tu}}, y_{\mathrm{Pu}})$ has the same effects as the corresponding *reduced* policy (y^+, y^-) together with the "spillage" represented by the last term in (3.2.2). Thus the overlap, $y_{\mathrm{Tu}} \wedge y_{\mathrm{Pu}}$, amounts to a limited form of free disposal; and it is allowed in the model in order to represent the efficient input-output bundles as the frontier of a convex production set. The frontier itself is not a convex set (except for the case of $\eta_{\mathrm{Ro}} = 1$), since the penultimate term in (3.2.2) is not linear in y: see Figure 3.1.

[6] The start-up times must be distinguished from the very much shorter loading times applicable to the generators already on line.

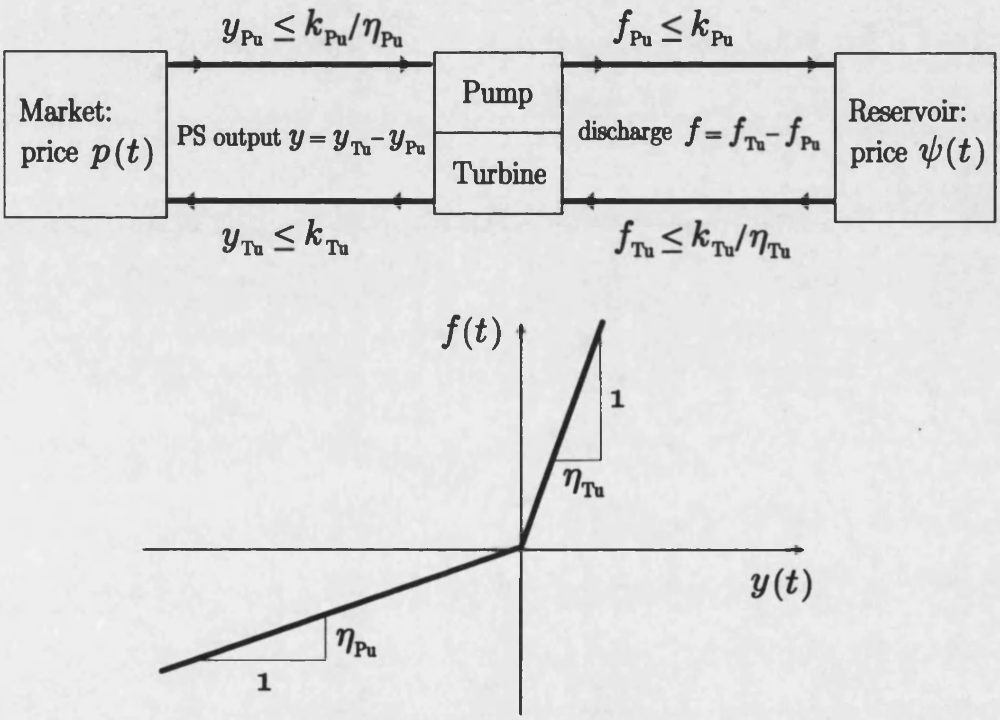

FIGURE 3.1. Conversion efficiencies $(\eta_{\mathrm{Pu}}, \eta_{\mathrm{Tu}})$ and the relationship between the storage plant's output rate $y(t)$ and the outflow from the reservoir $f(t)$.

The stock of medium, $s(t)$ at time t, is an absolutely continuous function on $[0,T]$ that satisfies the evolution equation $\dot{s} := \mathrm{d}s/\mathrm{d}t = -f$.[7] This can be restated as

$$s(t) = s(0) - \int_0^t f(\tau)\,\mathrm{d}\tau \tag{3.2.3}$$

and it follows that s is a Lipschitz function. This is because

$$k_{\mathrm{Tu}} \geq y_{\mathrm{Tu}} \geq 0 \geq -y_{\mathrm{Pu}} \geq -\frac{k_{\mathrm{Pu}}}{\eta_{\mathrm{Pu}}}$$

which shows that both y and f are bounded. That is, y and f belong to $L^{\infty}[0,T]$, which is the vector space of all essentially bounded functions, with functions equal almost everywhere (a.e.) being identified with each other. This space is normed by the supremum norm

$$\|y\|_{\infty} := \mathrm{EssSup}\,|y| = \operatorname*{ess\,sup}_{t\in[0,T]} |y(t)|.$$

The space of all continuous functions $\mathcal{C}[0,T]$, which contains the Lipschitz functions, is normed by the maximum norm

$$\|s\|_{\infty} = \mathrm{Max}\,|s| = \max_{t\in[0,T]} |s(t)|.$$

Its norm-dual $\mathcal{C}^*$, which serves as the price space for the services of storage capacity, is identified as the space of all (signed, finite) Borel measures $\mathcal{M}[0,T]$ by means of the bilinear form

$$\langle \mu \mid s \rangle := \int_{[0,T]} s(t)\,\mu(\mathrm{d}t)$$

for $s \in \mathcal{C}$ and $\mu \in \mathcal{M}$.

The available capacities are taken to equal the installed capacities, and therefore to be constant over the cycle. This does play a part in some of the main results, including the determinacy of rental values (Theorem 3.9.2).[8] However, to take full advantage of sensitivity analysis, the constant existing capacities k are perturbed with increments Δk which are (periodic) functions of time; this is further explained in Section 3.5. (The notation Δk, etc., is always to be interpreted as a single symbol meaning "an increment to k".)

On the assumption of constant capacities $k = (k_{\mathrm{St}}, k_{\mathrm{Tu}}, k_{\mathrm{Pu}})$, the long-run production set of the pumped-storage technique (with nonreversible equipment) is the convex cone

$$\mathbb{Y}_{\mathrm{PS}} := \Bigg\{ (y, -k) \in L^{\infty} \times \mathbb{R}^3_- : \exists\, (y_{\mathrm{Tu}}, y_{\mathrm{Pu}}) \in L^{\infty}_+[0,T] \times L^{\infty}_+[0,T] \tag{3.2.4}$$

$$y = y_{\mathrm{Tu}} - y_{\mathrm{Pu}},\ y_{\mathrm{Tu}} \leq k_{\mathrm{Tu}},\ \eta_{\mathrm{Pu}} y_{\mathrm{Pu}} \leq k_{\mathrm{Pu}}$$

$$\int_0^T \left(y_{\mathrm{Tu}}(t)/\eta_{\mathrm{Tu}} - \eta_{\mathrm{Pu}} y_{\mathrm{Pu}}(t) \right) \mathrm{d}t = 0 \quad \text{and}$$

[7] Since s is absolutely continuous, its derivative $\mathrm{d}s/\mathrm{d}t$ is well defined for almost every (a.e.) t. For these concepts see, e.g., [27], [76] or [84].

[8] The available capacities (i.e., the capacities in service) might generally vary because of maintenance schedules, etc.

$$\exists s_0 \in \mathbb{R}\ \forall t \quad 0 \leq s_0 - \int_0^t \left(y_{\mathrm{Tu}}(\tau)/\eta_{\mathrm{Tu}} - \eta_{\mathrm{Pu}} y_{\mathrm{Pu}}(\tau)\right) \mathrm{d}\tau \leq k_{\mathrm{St}} \Big\} .$$

This formulation imposes the periodicity constraint on the stock or, equivalently, the balance constraint on the flows to and from the reservoir ($\int_0^T f(t)\,\mathrm{d}t = 0$), but the stock level at the beginning or end of a cycle is taken to be a costless decision variable, s_0. In other words, when it is first commissioned, the reservoir comes charged up to any required level at no extra cost, but its periodic operation thereafter is taken to be a technological constraint.

3.3 Heuristics for valuation of stock and capacities

To start with, assume that not only the good's market price, $p(t)$, but also the stored medium's shadow price, $\psi(t)$, is known. Then the operating decisions can be decentralised within the storage plant, with the reservoir "buying" or "selling" the medium at the price $\psi(t)$ to or from the converter, which buys or sells the good at the market price $p(t)$ outside the plant. Short-run profit maximisation separates into two problems with obvious solutions, one for each kind of capacity. For simplicity, consider the case of a perfectly efficient, reversible and symmetric converter.[9] The maximum profits of the storage and the conversion capacities, $\Pi^{\mathrm{St}}(\psi, k_{\mathrm{St}})$ and $\Pi^{\mathrm{Co}}(p-\psi, k_{\mathrm{Co}})$, are both linear in k. A unit converter can earn the profit flow $(p-\psi)^-$ by putting the good into storage when $p(t) < \psi(t)$, as well as earning $(p-\psi)^+$ by taking the good out of storage when $p(t) > \psi(t)$. In both modes, profits are earned only at the times of full capacity utilisation, since the optimum output is $y(t) = \pm k_{\mathrm{Co}}$ whenever $\psi(t) \neq p(t)$: see Figures 3.2a and 3.2b.

In total over the cycle, the value of a unit converter is therefore

$$\frac{\Pi^{\mathrm{Co}}}{k_{\mathrm{Co}}} = \int_0^T |p(t) - \psi(t)|\,\mathrm{d}t.$$

As for the reservoir, a unit can earn a profit of $\psi(\overline{\tau}) - \psi(\underline{\tau})$ by buying stock at a time $\underline{\tau}$ and selling it at a later time $\overline{\tau}$ when $\psi(\overline{\tau}) > \psi(\underline{\tau})$. The value of a unit reservoir is therefore the sum of all shadow price rises in a cycle. In precise terms: if $\psi(T) \geq \psi(0)$, then

$$\frac{\Pi^{\mathrm{St}}}{k_{\mathrm{St}}} = \mathrm{Var}^+(\psi)$$

which denotes the *total positive variation* (a.k.a. upper variation) of ψ, i.e., the supremum of $\sum_m \left(\psi(\overline{\tau}_m) - \psi(\underline{\tau}_m)\right)^+$ over all finite sets of pairwise disjoint subintervals $(\underline{\tau}_m, \overline{\tau}_m)$ of $(0, T)$.[10]

If $\psi(T) < \psi(0)$, the reservoir should start the cycle full, and refill towards the end of the cycle. This brings an extra profit of $\psi(0) - \psi(T)$, so in general the unit rent is the *cyclic* positive variation

$$\mathrm{Var}_c^+(\psi) := \mathrm{Var}^+(\psi) + (\psi(0) - \psi(T))^+ . \tag{3.3.1}$$

[9] In this case the conversion constraints on an output y simplify to: $-k_{\mathrm{Co}} \leq y \leq k_{\mathrm{Co}}$; and stock evolution simplifies to: $\dot{s} = -y$.

[10] For a discussion of Var^+ see, e.g., [27, Section 8.1] or [84, Section 3.5].

It is later shown that actually $\psi(0) = \psi(T)$ if $p \in C[0,T]$ and $p(0) = p(T)$.

The maximum operating profit of the whole storage plant ($\Pi_{\mathrm{SR}}^{\mathrm{PS}}$) is, however, a function *not* of ψ but of the problem's parameters $(p, k_{\mathrm{St}}, k_{\mathrm{Co}})$ alone: ψ is an auxiliary function which must eventually be given in terms of $(p, k_{\mathrm{St}}, k_{\mathrm{Co}})$. Then $\partial \Pi_{\mathrm{SR}}^{\mathrm{PS}} / \partial k_{\mathrm{St}}$ and $\partial \Pi_{\mathrm{SR}}^{\mathrm{PS}} / \partial k_{\mathrm{Co}}$ can be obtained by substituting the correct ψ into the expressions $\mathrm{Var}_c^+(\psi)$ and $\int_0^T |p(t) - \psi(t)| \, \mathrm{d}t$.

The correct value, $\hat{\psi}$, is that stock price function which minimises the total value of the storage plant's fixed resources $(k_{\mathrm{St}}, k_{\mathrm{Co}})$. So, given a cyclic TOU tariff p, one can find $\hat{\psi}$ by unconstrained minimisation of

$$k_{\mathrm{St}} \, \mathrm{Var}_c^+(\psi) + k_{\mathrm{Co}} \int_0^T |p(t) - \psi(t)| \, \mathrm{d}t \tag{3.3.2}$$

over ψ, an arbitrary bounded-variation function on $(0, T)$.

The main feature of this programme is the trade-off between minimising the variation (which on its own would require setting ψ at a constant value) and minimising the integral (which on its own would require setting ψ equal to p). From this trade-off it is clear to what extent the local peaks of p should be "shaved off" and the troughs "filled in" to obtain the optimum shadow price $\hat{\psi}$—at least in the case of a piecewise strictly monotone market price p. The solution, presented graphically in Figure 3.2a, is determined by constancy intervals for $\hat{\psi}$, on each of which either $p(t) < \hat{\psi}$ throughout (around a trough of p) or $p(t) > \hat{\psi}$ throughout (around a local peak of p). Unless $k_{\mathrm{St}}/k_{\mathrm{Co}}$, the time needed to fully charge or discharge the reservoir, is relatively long, these intervals do not abut, and must all be of that length. This is the first-order condition (FOC) for the dual optimum: the increment in the minimand (3.3.2) that results from shifting the constant value of ψ up or down by an infinitesimal unit, on an interval of length τ, is $\pm(k_{\mathrm{St}} - k_{\mathrm{Co}}\tau)$. Equating this to zero gives the optimum as $\hat{\tau} = k_{\mathrm{St}}/k_{\mathrm{Co}}$, i.e., $k_{\mathrm{St}}/k_{\mathrm{Co}}$ is the common length of the intervals on which alternately $\hat{\psi} > p$ or $\hat{\psi} < p$.[11] This makes it feasible to produce the "bang-coast-bang" output (viz., $y(t) = \pm k_{\mathrm{Co}}$ when $\hat{\psi}(t) \neq p(t)$, with $y(t) = 0$ when $\hat{\psi}(t) = p(t)$): the reservoir goes alternately from empty to full and *vice versa* (Figures 3.2b and 3.2c). This is the optimal output.

The same marginal calculation for the dual problem also shows that an optimum function ψ can be nonunique if p is discontinuous. Suppose, for example, that p jumps at the beginning, and drops at the end, of an interval $A = (\underline{t}, \bar{t})$, of length $k_{\mathrm{St}}/k_{\mathrm{Co}}$, with

$$p(\underline{t}-) \vee p(\bar{t}+) < p(\underline{t}+) \wedge p(\bar{t}-) = \inf_{t \in A} p(t) \tag{3.3.3}$$

where $\vee$ and $\wedge$ mean the smaller and the larger of the two, and $p(t-)$ and $p(t+)$ denote the left and right limits at t. Just before $\underline{t}$ and just after $\bar{t}$, an optimal ψ equals p, i.e., $\psi(\underline{t}-) = p(\underline{t}-)$ and $\psi(\bar{t}+) = p(\bar{t}+)$. Inside A, $p > \psi = \text{const.}$; but an optimal constant value of ψ on A can

[11] Matters complicate when the ratio $k_{\mathrm{St}}/k_{\mathrm{Co}}$ is comparable to the times elapsed between the successive local peaks and troughs of p, so that the neighbouring constancy intervals of $\hat{\psi}$ start to abut; but a similar optimality rule applies to such clusters.

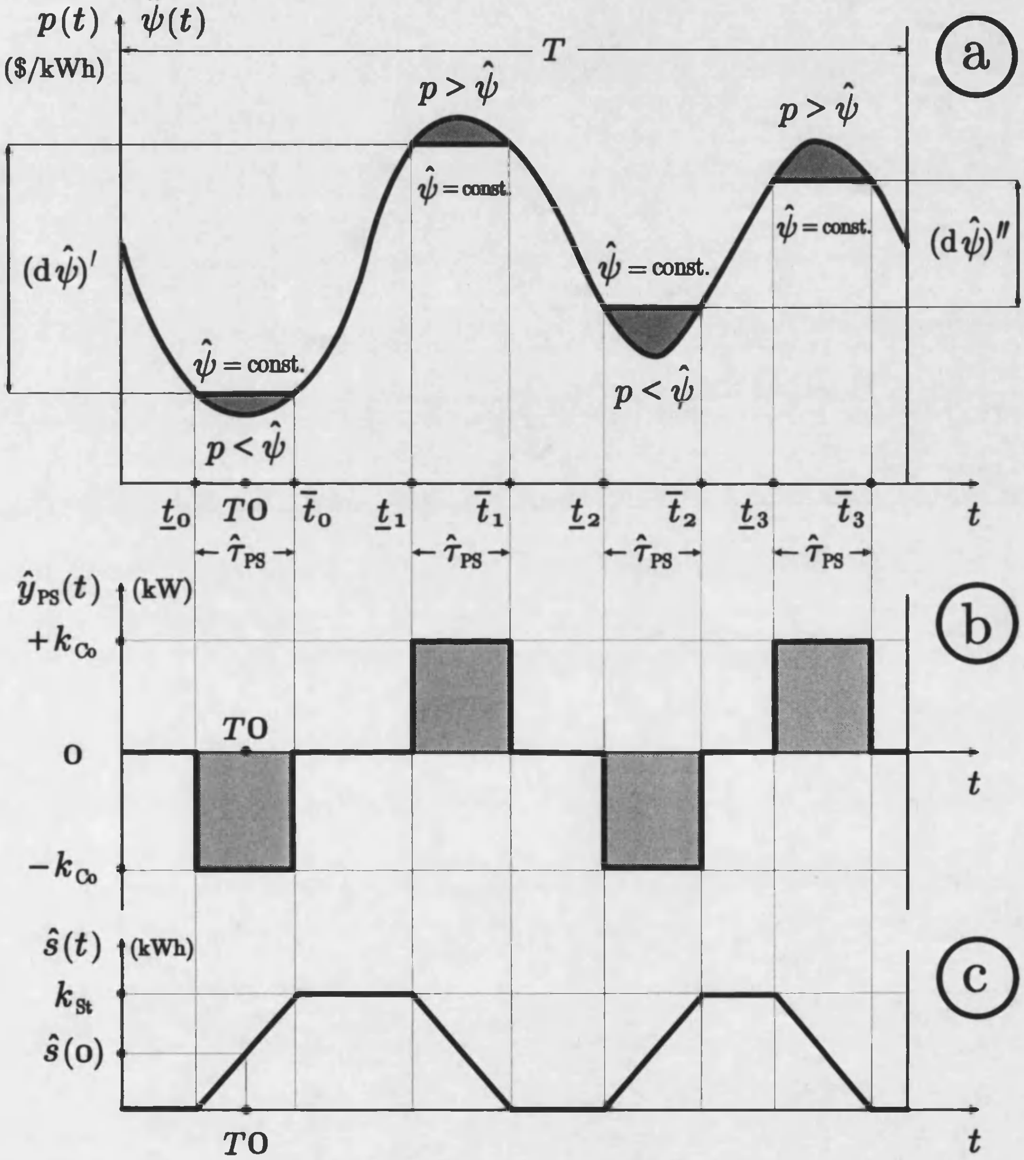

FIGURE 3.2. Trajectories of: (a) the optimal shadow price of stock $\hat{\psi}$, (b) the output of pumped-storage plant $\hat{y}_{\mathrm{PS}}$, and (c) the stock $\hat{s}$, in the case of a perfectly efficient and symmetric converter. Unit rent for storage capacity is $\mathrm{Var}_c^+\left(\hat{\psi}\right) = \left(\mathrm{d}\hat{\psi}\right)' + \left(\mathrm{d}\hat{\psi}\right)''$, the sum of rises of $\hat{\psi}$. Unit rent for conversion capacity is $\int_0^T \left|p(t) - \hat{\psi}(t)\right| \mathrm{d}t$, the sum of dark grey areas in (a). In (b), each of the light grey areas equals the reservoir's capacity k_{St}. By definition $\hat{\tau}_{\mathrm{PS}} = k_{\mathrm{St}}/k_{\mathrm{Co}}$.

be anywhere between the two unequal terms of (3.3.3): the jump and the drop of p create an "indifference zone" for $\psi_{|A}$. Figure 3.3 shows this when $p\left(\bar{t}+\right) \leq p\left(\underline{t}-\right) < p\left(\underline{t}+\right) \leq p\left(\bar{t}-\right)$ so $p\left(\underline{t}-\right) \leq \psi_{|A} \leq p\left(\underline{t}+\right)$. Different values from this range divide the same total rent differently between the two capacities. The jump $\mathrm{d}\psi\left\{\underline{t}\right\} := \psi\left(\underline{t}+\right) - \psi\left(\underline{t}-\right)$ can be any fraction of $p\left(\underline{t}+\right) - p\left(\underline{t}-\right)$, and it is an indeterminate contribution to the reservoir's unit rent. The interval's contribution to the converter's rent, $\int_A \left(p\left(t\right) - \psi\right) \mathrm{d}t$, is similarly indeterminate (since it depends on the arbitrary choice of $\psi\left(\underline{t}+\right)$, which fixes the constant value of ψ on A).

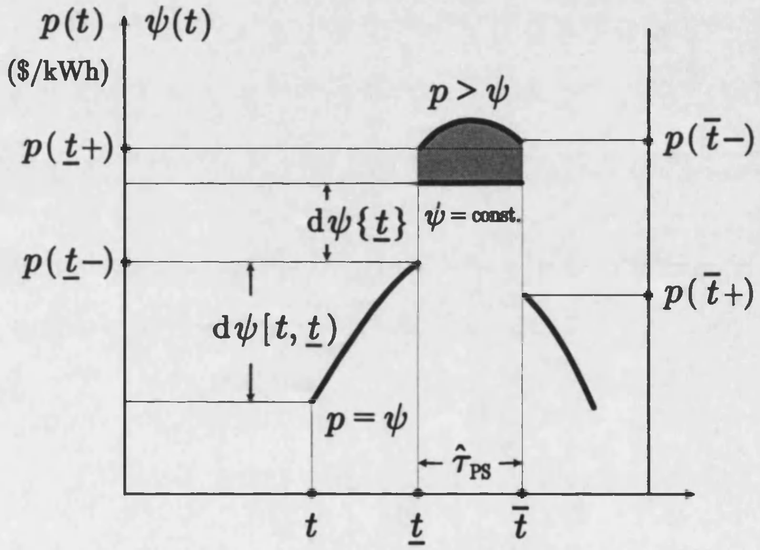

FIGURE 3.3. Indeterminacy of an optimal shadow price of stock ψ when the TOU price of good p has jump discontinuities (at instants differing by $\hat{\tau}_{\mathrm{PS}} = k_{\mathrm{St}}/k_{\mathrm{Co}}$). In the case shown, the constant value of ψ on $\left(\underline{t}, \bar{t}\right)$ can be set at any level between $p\left(\underline{t}-\right)$ and $p\left(\underline{t}+\right)$; so the jump of ψ at $\underline{t}$ is an indeterminate part of the reservoir's unit rent. The dark grey area represents $\int_{\underline{t}}^{\bar{t}} \left|p\left(t\right) - \psi\left(t\right)\right| \mathrm{d}t$, the interval's contribution to the converter's unit rent.

Conversely, given a continuous p, the optimum ψ is unique (Lemma 3.13.1). Therefore the gradient $\nabla_k \Pi$ exists; and this result extends to the case of imperfect conversion (Theorem 3.9.2).

3.4 The linear programme of plant operation

In terms of the production set (3.2.4), the problem of profit-maximising operation of a storage plant is

$$\text{Given } \left(p; k_{\mathrm{St}}, k_{\mathrm{Tu}}, k_{\mathrm{Pu}}\right) \in L_{+}^{\infty *}\left[0, T\right] \times \mathbb{R}_{+}^{3} \tag{3.4.1}$$

$$\text{maximise } \left\langle p \,|\, y \right\rangle \text{ over } y \in L^{\infty}\left[0, T\right] \tag{3.4.2}$$

subject to: $(y, -k_{\mathrm{St}}, -k_{\mathrm{Tu}}, -k_{\mathrm{Pu}}) \in \mathbb{Y}_{\mathrm{PS}}$ defined by (3.2.4). (3.4.3)

Notation The optimal value of (3.4.1)–(3.4.3) is the (maximum) operating profit of the storage plant, denoted by $\Pi^{\mathrm{PS}}_{\mathrm{SR}}(p,k)$. The (optimal) solution set is $\hat{Y}_{\mathrm{PS}}(p,k)$, abbreviated to $\hat{Y}$. The corresponding lowercase notation, $\hat{y}$, is used *only when the solution is known to be unique.*

The space $L^{\infty *}$ appearing in (3.4.1) is the norm-dual of L^{∞}. It contains L^1, the space of all functions integrable with respect to (w.r.t.) meas, the Lebesgue measure. Much of the analysis applies not only to a TOU tariff represented by a price function $p \in L^1[0,T]$ but also to a tariff represented by a $p \in L^{\infty *}[0,T]$. Such a p can be identified with a finitely additive set function vanishing on meas-null sets, since the integral of a $y \in L^{\infty}$ w.r.t. such a set function defines a bounded linear functional on L^{∞}: see, e.g., [25, III.1–III.2 and IV.8.16] or [86, 2.3]. As an additive set function, a $p \in L^{\infty *}$ has the Hewitt-Yosida decomposition into $p_{\mathrm{CA}} + p_{\mathrm{FA}}$, the sum of its countably additive (c.a.) and purely finitely additive (p.f.a.) parts: see, e.g., [10, Appendix I: (26)–(27)], [25, III.7.8] or [86, 1.23 and 1.24].[12] The c.a. part of p is identified with its density w.r.t. meas (which exists by the Radon-Nikodym Theorem); so it is a price function $p_{\mathrm{CA}} \in L^1[0,T]$. The p.f.a. part can be characterised as a singular element of $L^{\infty *}[0,T]$, i.e., p_{FA} is concentrated on a subset of $[0,T]$ with an arbitrarily small Lebesgue measure. (Formally, a $p \in L^{\infty *}$ is *concentrated on*, or *supported by*, a measurable set S if $\langle p \mid y \rangle = \langle p \mid y 1_S \rangle$ for every $y \in L^{\infty}$, where 1_S denotes the 0-1 indicator of S (equal to 1 on S and to 0 outside). A sequence of sets (S_m) is *evanescent* if $S_{m+1} \subseteq S_m$ for every m and $\mathrm{meas}\left(\bigcap_{m=1}^{\infty} S_m\right) = 0$; and p is called *singular* if there exists an evanescent sequence (S_m) such that p is concentrated on S_m for each m. A $p \in L^{\infty *}$ is singular if and only if it is p.f.a.: see [86, 3.1].) This gives p_{FA} the interpretation of an extremely concentrated charge. In the storage context it can arise as a turbine capacity charge (Remark 3.14.6).

The value of $y \in L^{\infty}$ at $p \in L^{\infty *}$ is

$$\langle p \mid y \rangle_{L^{\infty *}, L^{\infty}} = \int_0^T p_{\mathrm{CA}}(t)\, y(t)\, \mathrm{d}t + \langle p_{\mathrm{FA}} \mid y \rangle \tag{3.4.4}$$

which is abbreviated to $\langle p \mid y \rangle$. Although the last term in (3.4.4) is also an integral, it is one that lacks some basic properties; and the symbol $\int$ is reserved here for integration w.r.t. a measure, which is countably additive by definition. The only measures in $L^{\infty *}$ are those having densities, i.e., $L^{\infty *} \cap \mathcal{M} = L^1$.

By definition, a $p \in L^{\infty *}$ is (strictly) *positive as a linear functional* on L^{∞} if $\langle p \mid \cdot \rangle$ is positive on $L^{\infty}_{+} \setminus \{0\}$. This is the case if and only if $p_{\mathrm{FA}} \geq 0$ and $p_{\mathrm{CA}} > 0$ a.e. on $[0,T]$. The latter condition is also written as $p_{\mathrm{CA}} \gg 0$, or as $p_{\mathrm{CA}} \in L^1_{++}$. For the subspace $\mathcal{C}$, note that $p \in \mathcal{C}_{++}$ if and only if $\mathrm{Min}(p) > 0$.

[12] A p.f.a. set function is one that is lattice-disjoint from every c.a. one.

The plant operation problem is next formulated as an LP. With the constants k_{St}, k_{Tu} and k_{Pu} viewed as special cases of cyclically varying functions, this primal LP is:

$$\text{Given } (p; k_{\mathrm{St}}, k_{\mathrm{Tu}}, k_{\mathrm{Pu}}) \in L_+^{\infty *} \times \mathbb{R}_+^3 \subset L_+^{\infty *} \times (\mathcal{C}_+ \times L_+^{\infty} \times L_+^{\infty}) \text{ with } p_{\mathrm{CA}} \gg 0 \tag{3.4.5}$$

$$\text{maximise } \langle p \,|\, y_{\mathrm{Tu}} - y_{\mathrm{Pu}} \rangle \text{ over } (y_{\mathrm{Tu}}, y_{\mathrm{Pu}}) \in L^{\infty} \times L^{\infty} \text{ and } s_0 \in \mathbb{R} \tag{3.4.6}$$

$$\text{subject to: } \quad 0 \le y_{\mathrm{Tu}}(t) \le k_{\mathrm{Tu}} \quad \text{for a.e. } t \tag{3.4.7}$$

$$0 \le \eta_{\mathrm{Pu}} y_{\mathrm{Pu}}(t) \le k_{\mathrm{Pu}} \quad \text{for a.e. } t \tag{3.4.8}$$

$$\int_0^T f(t)\,\mathrm{d}t = 0 \tag{3.4.9}$$

$$0 \le s_0 - \int_0^t f(\tau)\,\mathrm{d}\tau \le k_{\mathrm{St}} \quad \text{for every } t \tag{3.4.10}$$

where $f := y_{\mathrm{Tu}}/\eta_{\mathrm{Tu}} - \eta_{\mathrm{Pu}} y_{\mathrm{Pu}}$ as per (3.2.1)–(3.2.2).

The two formulations of the operation problem are equivalent in the sense that y solves (3.4.1)–(3.4.3) if and only if $y = y_{\mathrm{Tu}} - y_{\mathrm{Pu}}$ for some $(y_{\mathrm{Tu}}, y_{\mathrm{Pu}}, s_0)$ that solves (3.4.5)–(3.4.10)—in which case $(y_{\mathrm{Tu}}, y_{\mathrm{Pu}})$ together with the specific value

$$\underline{s}_{0,y_{\mathrm{Tu}},y_{\mathrm{Pu}}} := \max_{t \in [0,T]} \int_0^t f(\tau)\,\mathrm{d}\tau = \max_{t \in [0,T]} \int_0^t \left(\frac{y_{\mathrm{Tu}}(\tau)}{\eta_{\mathrm{Tu}}} - \eta_{\mathrm{Pu}} y_{\mathrm{Pu}}(\tau) \right) \mathrm{d}\tau \tag{3.4.11}$$

is a solution: $\underline{s}_{0,y_{\mathrm{Tu}},y_{\mathrm{Pu}}}$ is the lowest initial stock required for $s(t)$ never to fall below 0. (Unless there is spare storage capacity, this is actually the only feasible value for s_0, given $(y_{\mathrm{Tu}}, y_{\mathrm{Pu}})$.) One can therefore restrict attention to points $(y_{\mathrm{Tu}}, y_{\mathrm{Pu}}, s_0)$ with $s_0 = \underline{s}_{0,y_{\mathrm{Tu}},y_{\mathrm{Pu}}}$; and so the stock trajectory associated with a storage policy $(y_{\mathrm{Tu}}, y_{\mathrm{Pu}})$ is

$$s(t) = \underline{s}_{0,y_{\mathrm{Tu}},y_{\mathrm{Pu}}} - \int_0^t \left(\frac{y_{\mathrm{Tu}}(\tau)}{\eta_{\mathrm{Tu}}} - \eta_{\mathrm{Pu}} y_{\mathrm{Pu}}(\tau) \right) \mathrm{d}\tau. \tag{3.4.12}$$

The dual programme, introduced next, serves the purposes of characterising optimal operation and calculating the marginal capacity values. To ensure that the problem is nontrivial—and that the dual is soluble—for the most part it is assumed from here on that $k \gg 0$, i.e., that[13]

$$k_{\mathrm{Tu}} > 0, \; k_{\mathrm{Pu}} > 0 \text{ and } k_{\mathrm{St}} > 0. \tag{3.4.13}$$

3.5 Capacity valuation as the dual linear programme

As is set out in, e.g., [73], the dual to a convex programme depends on the choice of perturbations for the primal parameters. A choice of admissible perturbations determines the structure of the

[13] A srtictly positive k means that the primal meets Slater's Condition. This standard constraint qualification for CPs is, in the infinite-dimensional case, useful with LPs as well. Without it, the primal and dual values may be different, or there may be no dual optimum. For example, if $p \in L^1$, $k_{\mathrm{Tu}} > 0$ and $k_{\mathrm{Pu}} > 0$ but $k_{\mathrm{St}} = 0$, then the primal and dual values are equal (viz., 0); but if additionally $\eta_{\mathrm{Ro}} = 1$, then the dual optimum exists only if $p \in \mathrm{BV}$ (in which case the optimal stock price is $\hat{\psi} = p$). See also [4, p. 31.].

dual variables (a.k.a. Lagrange multipliers) to be paired with the parameter increments. Therefore, the dual programme depends not only on the particular values of the primal parameters, but also on the vector space of parameter increments or perturbations. This "ambient" space for the given parameter point can be chosen to suit one's purpose.

In the case of (3.4.5)–(3.4.10), the programme contains a separate set of capacity constraints for each time t—so, by considering a separate increment $\Delta k(t)$ for each t, one can impute an instantaneous value, $\kappa(t)$, to capacity services at each time t. In other words, not only their total value, but also its distribution over the cycle can be determined. Even if the existing capacities k are actually constant, it is useful to consider the cyclically varying increments Δk because this gives a marginal interpretation to the time-dependent Lagrange multipliers for the capacity constraints: denoted by $\kappa = (\kappa_{\mathrm{St}}, \kappa_{\mathrm{Tu}}, \kappa_{\mathrm{Pu}})$, these are the unit values of the capacities' services at any particular time. As part of the "variation of constants", a varying increment $\Delta n_{\mathrm{St}}(t)$ to the zero floor for the stock in (3.4.10) is also considered, as are cyclically varying increments, $\Delta n_{\mathrm{Tu}}(t)$ and $\Delta n_{\mathrm{Pu}}(t)$, to the zero floors for the turbine and pump output rates in (3.4.7) and (3.4.8). This gives a marginal interpretation to the time-dependent Lagrange multipliers for the nonnegativity constraints: denoted by $\nu = (\nu_{\mathrm{St}}, \nu_{\mathrm{Tu}}, \nu_{\mathrm{Pu}})$, these are the unit values of lowering the "floors" at any time. Finally, a scalar $\Delta\zeta$ is an increment to the zero on the r.h.s. of (3.4.9); this can be thought of as the quantity of the medium taken to be available for topping up the reservoir between cycles. Its multiplier, a scalar λ, is the marginal value of stock at the beginning (or end) of cycle. All the multipliers (κ, ν and λ) are terms of the TOU price p in its decomposition (3.5.6) below, which is a part of the dual programme's constraints.

The short-run profit maximisation problem (3.4.5)–(3.4.10) is thus embedded in the family of perturbed programmes obtained by adding an arbitrary cyclically varying increment (Δk_{St}, Δn_{St}, Δk_{Tu}, Δn_{Tu}, Δk_{Pu}, Δn_{Pu}) and a scalar $\Delta\zeta \in \mathbb{R}$ to the particular parameter point consisting of the constants (k_{St}, 0, k_{Tu}, 0, k_{Pu}, 0) and 0. This perturbation is termed *refined*, to distinguish it from the coarser perturbation by constant increments.

The function spaces for the resource increments, already indicated in (3.4.5), are: $\mathcal{C}[0,T]$ for Δk_{St} and Δn_{St}, and $L^\infty[0,T]$ for Δk_{Tu}, Δn_{Tu}, Δk_{Pu} and Δn_{Pu}. These are paired with $\mathcal{M}[0,T]$ and $L^{\infty *}[0,T]$ as the shadow price (multiplier) spaces. With an infinite-dimensional parameter space such as L^∞, the dual programme depends also on the choice of the dual space—and L^∞ can be usefully paired with either $L^{\infty *}$ or L^1. But when $p \in L^1[0,T]$, the pairing of L^∞ with $L^{\infty *}$ is needed only in proving the dual's solubility: any optimal κ_{Tu} and κ_{Pu} are actually in L^1 (as are ν_{Tu} and ν_{Pu}).

In other words, the marginal value of the storage capacity services on an interval $A \subset [0,T]$ is given by a measure $\kappa_{\mathrm{St}}(A)$; this is the incremental operating profit from the availability on A of an extra unit of the reservoir. Another measure, $\nu_{\mathrm{St}}(A)$, gives the incremental profit from lowering the stock floor by a unit, on A. The marginal value of the turbine capacity services, on A, is the integral of a $\kappa_{\mathrm{Tu}} \in L^{\infty *}$; similarly, the marginal value of the pump capacity services, on A, is the integral of a $\kappa_{\mathrm{Pu}} \in L^1$. The value of lowering the conversion rate floor by a unit is is the integral of a $\nu_{\mathrm{Tu}} \in L^1$ for the turbine, and of a $\nu_{\mathrm{Pu}} \in L^{\infty *}$ for the pump.

Thus the complete shadow-price system (κ_{St}, ν_{St}; κ_{Tu}, ν_{Tu}; κ_{Pu}, ν_{Pu}; λ) values all the resource increments (Δk_{St}, $-\Delta n_{\mathrm{St}}$; Δk_{Tu}, $-\Delta n_{\mathrm{Tu}}$; Δk_{Pu}, $-\Delta n_{\mathrm{Pu}}$; $\Delta\zeta$). Of course, it also values the particular resource bundle (k_{St}, 0; k_{Tu}, 0; k_{Pu}, 0; 0) that represents the plant itself—and the dual to the operation programme (3.4.5)–(3.4.10) is to minimise the plant's value by an admissible choice of the shadow prices. The main dual constraints (3.5.6) are two decompositions of the good's price p into a signed sum of: the conversion capacity charge, the value of the conversion floor, and the shadow price of stock. There is one decomposition for the turbine and one for the pump. The stock price, later denoted by ψ, is the sum of: the initial price λ, the cumulative of reservoir capacity charges κ_{St}, and the cumulative of $-\nu_{\mathrm{St}}$; it is the middle sum in (3.5.6). This spelt out next.

Theorem 3.5.1 (Fixed-input value minimisation as the dual). *The dual of the linear programme (3.4.5)–(3.4.10), relative to the refined perturbation and the pairing of the parameter spaces C and L^∞ with $\mathcal{M}$ and $L^{\infty *}$ respectively, is:*

$$\text{Given } (p, k_{\mathrm{St}}, k_{\mathrm{Tu}}, k_{\mathrm{Pu}}) \text{ as in (3.4.5)} \tag{3.5.1}$$

$$\text{minimise } k_{\mathrm{St}} \int_{[0,T]} \kappa_{\mathrm{St}}(\mathrm{d}t) + k_{\mathrm{Tu}} \left\langle \kappa_{\mathrm{Tu}} \mid 1_{[0,T]} \right\rangle + k_{\mathrm{Pu}} \left\langle \kappa_{\mathrm{Pu}} \mid 1_{[0,T]} \right\rangle \tag{3.5.2}$$

$$\text{over } \lambda \in \mathbb{R} \text{ and } (\kappa_{\mathrm{St}}, \nu_{\mathrm{St}}; \kappa_{\mathrm{Tu}}, \nu_{\mathrm{Tu}}; \kappa_{\mathrm{Pu}}, \nu_{\mathrm{Pu}}) \in \mathcal{M}^2 \times (L^{\infty *})^2 \times (L^{\infty *})^2 \tag{3.5.3}$$

$$\text{subject to:} \quad (\kappa_{\mathrm{St}}, \nu_{\mathrm{St}}, \kappa_{\mathrm{Tu}}, \nu_{\mathrm{Tu}}, \kappa_{\mathrm{Pu}}, \nu_{\mathrm{Pu}}) \geq 0 \tag{3.5.4}$$

$$\kappa_{\mathrm{St}}[0,T] = \nu_{\mathrm{St}}[0,T] \tag{3.5.5}$$

$$\eta_{\mathrm{Tu}} p - \eta_{\mathrm{Tu}}(\kappa_{\mathrm{Tu}} - \nu_{\mathrm{Tu}}) = \lambda + (\kappa_{\mathrm{St}} - \nu_{\mathrm{St}})[0,\cdot] = \frac{p}{\eta_{\mathrm{Pu}}} + (\kappa_{\mathrm{Pu}} - \nu_{\mathrm{Pu}}). \tag{3.5.6}$$

Remark 3.5.2. *Under (3.4.13), any solution to (3.5.1)–(3.5.6) has the disjointness properties that*

$$\kappa_\phi \wedge \nu_\phi = 0 \quad \text{for } \phi = \mathrm{St}, \mathrm{Tu}, \mathrm{Pu} \quad \text{and} \quad \kappa_{\mathrm{St}}\{0,T\} \wedge \nu_{\mathrm{St}}\{0,T\} = 0 \tag{3.5.7}$$

i.e., it is not optimal for the dual variables to overlap and partly cancel each other out.[14]

Comments:

- Therefore the programme (3.5.1)–(3.5.6) can be reformulated in terms of the signed variables

$$\mu_\phi := \kappa_\phi - \nu_\phi \quad \text{for } \phi = \mathrm{St}, \mathrm{Tu}, \mathrm{Pu} \tag{3.5.8}$$

by replacing (κ_ϕ, ν_ϕ) with $\left(\mu_\phi^+, \mu_\phi^-\right)$. At an optimum, $\mu_{\mathrm{St}}\{0\}$ and $\mu_{\mathrm{St}}\{T\}$ do not have opposite signs.

[14] For $\phi = \mathrm{St}$, this means that κ_{St} and ν_{St} are disjoint as measures on the circle obtained from the interval $[0,T]$, and not only on $[0,T]$ itself.

- If $\eta_{\mathrm{Ro}} = 1$, then $\kappa_{\mathrm{Tu}} = \nu_{\mathrm{Pu}}$ and $\kappa_{\mathrm{Pu}} = \nu_{\mathrm{Tu}}$ from (3.5.6) and (3.5.7); so in this case

$$\kappa_{\mathrm{Tu}} \wedge \kappa_{\mathrm{Pu}} = 0. \tag{3.5.9}$$

- Since $p \geq 0$, one has $\kappa_{\mathrm{Tu}} \wedge \kappa_{\mathrm{Pu}} = 0$ also when $\eta_{\mathrm{Ro}} < 1$. (This is because, as can be seen by using (3.5.6) to expand $((1/\eta_{\mathrm{Pu}}) - \eta_{\mathrm{Tu}}) p$, if $\kappa_{\mathrm{Tu}} \wedge \kappa_{\mathrm{Pu}} > 0$ then $\kappa_{\mathrm{Tu}} \wedge \nu_{\mathrm{Tu}} > 0$ or $\kappa_{\mathrm{Pu}} \wedge \nu_{\mathrm{Pu}} > 0$, which contradicts (3.5.7).) As for ν, with $\eta_{\mathrm{Ro}} < 1$ it can be that $\nu_{\mathrm{Tu}} \wedge \nu_{\mathrm{Pu}} > 0$.

- By the Hewitt-Yosida decomposition, (3.5.6) can be restated as

$$\eta_{\mathrm{Tu}} \left(p_{\mathrm{CA}} - \kappa_{\mathrm{Tu}}^{\mathrm{CA}} + \nu_{\mathrm{Tu}}^{\mathrm{CA}}\right)(t) = \lambda + (\kappa_{\mathrm{St}} - \nu_{\mathrm{St}})\,[0,t] = \left(\frac{p_{\mathrm{CA}}}{\eta_{\mathrm{Pu}}} + \kappa_{\mathrm{Pu}}^{\mathrm{CA}} - \nu_{\mathrm{Pu}}^{\mathrm{CA}}\right)(t)$$

for a.e. t, with

$$\kappa_{\mathrm{Tu}}^{\mathrm{FA}} - \nu_{\mathrm{Tu}}^{\mathrm{FA}} = p_{\mathrm{FA}} = -\eta_{\mathrm{Pu}} \left(\kappa_{\mathrm{Pu}}^{\mathrm{FA}} - \nu_{\mathrm{Pu}}^{\mathrm{FA}}\right). \tag{3.5.10}$$

- Since $p_{\mathrm{FA}} \geq 0$, (3.5.7) and (3.5.10) give

$$\nu_{\mathrm{Tu}}^{\mathrm{FA}} = \kappa_{\mathrm{Pu}}^{\mathrm{FA}} = p_{\mathrm{FA}}^{-} = 0 \quad \text{and} \quad \kappa_{\mathrm{Tu}}^{\mathrm{FA}} = \eta_{\mathrm{Pu}} \nu_{\mathrm{Pu}}^{\mathrm{FA}} = p_{\mathrm{FA}}^{+} = p_{\mathrm{FA}}. \tag{3.5.11}$$

  So $\kappa_{\mathrm{Pu}} \in L^1$, and the third term of (3.5.2) can be rewritten as $k_{\mathrm{Pu}} \int_0^T \kappa_{\mathrm{Pu}}(t)\,\mathrm{d}t$. If $p \in L^1$, i.e., $p_{\mathrm{FA}} = 0$, then also $\kappa_{\mathrm{Tu}} \in L^1$; and in this case the second term of (3.5.2) can similarly be rewritten as $k_{\mathrm{Tu}} \int_0^T \kappa_{\mathrm{Tu}}(t)\,\mathrm{d}t$.

3.6 Conditions for optimal operation and valuation

The dual programme (3.5.1)–(3.5.6) has a solution, in which κ_{Pu} and ν_{Tu} are in L^1 by (3.5.11), whilst κ_{Tu} and ν_{Pu} are generally in $L^{\infty *}$ (and κ_{St} and ν_{St} are in $\mathcal{M}$). The primal and dual optima are characterised by the Kuhn-Tucker Conditions, which for LPs reduce to feasibility and complementary slackness. Spelt out next, these conditions are later used to determine plant operation in terms of the stock price, and to establish that this shadow price is unique (at least at the times which matter, and literally unique if $\eta_{\mathrm{Ro}} = 1$).

Proposition 3.6.1 (Dual solubility and optimality conditions). *Assume (3.4.13). Then:*

1. *The fixed-input value minimisation programme (3.5.1)–(3.5.6) has an optimal solution*

$$(\kappa_{\mathrm{St}}, \nu_{\mathrm{St}}, \kappa_{\mathrm{Tu}}, \nu_{\mathrm{Tu}}, \kappa_{\mathrm{Pu}}, \nu_{\mathrm{Pu}}, \lambda) \in \mathcal{M} \times \mathcal{M} \times L^{\infty *} \times L^1 \times L^1 \times L^{\infty *} \times \mathbb{R}.$$

   The programme's value is finite and equal to the short-run profit $\Pi_{\mathrm{SR}}^{\mathrm{PS}}(p, k_{\mathrm{St}}, k_{\mathrm{Tu}}, k_{\mathrm{Pu}})$, the optimal value of (3.4.5)–(3.4.10). Furthermore, if $p \in L^1[0,T]$, then also $\kappa_{\mathrm{Tu}} \in L^1$ and $\nu_{\mathrm{Pu}} \in L^1$ in every solution.

2. Points $\left(y_{\mathrm{Tu}}, y_{\mathrm{Pu}}, \underline{s}_{0,y_{\mathrm{Tu}},y_{\mathrm{Pu}}}\right) \in L^{\infty} \times L^{\infty} \times \mathbb{R}$ *and* $(\kappa_{\mathrm{St}}, \nu_{\mathrm{St}}, \kappa_{\mathrm{Tu}}, \nu_{\mathrm{Tu}}, \kappa_{\mathrm{Pu}}, \nu_{\mathrm{Pu}}, \lambda)$ *are optimal solutions to, respectively, the primal (3.4.5)–(3.4.10) and the dual (3.5.1)–(3.5.6) if and only if:*

   (a) $\left(y_{\mathrm{Tu}}, y_{\mathrm{Pu}}, \underline{s}_{0,y_{\mathrm{Tu}},y_{\mathrm{Pu}}}\right)$ *and* $(\kappa_{\mathrm{St}}, \nu_{\mathrm{St}}, \kappa_{\mathrm{Tu}}, \nu_{\mathrm{Tu}}, \kappa_{\mathrm{Pu}}, \nu_{\mathrm{Pu}}, \lambda)$ *are feasible, i.e., satisfy (3.4.7)–(3.4.10) and (3.5.4)–(3.5.6).*

   (b) The measure κ_{St} *is concentrated on* $\{t \in [0,T] : s(t) = k_{\mathrm{St}}\}$*, and* ν_{St} *is concentrated on* $\{t : s(t) = 0\}$*, where s is given by (3.4.11)–(3.4.12).*

   (c) For every number $\epsilon > 0$*,* $\kappa_{\mathrm{Tu}} \in L^{\infty *}$ *is concentrated on* $\{t : y_{\mathrm{Tu}}(t) \geq k_{\mathrm{Tu}} - \epsilon\}$*, and the function* $\nu_{\mathrm{Tu}} \in L^1$ *vanishes a.e. outside of* $\{t : y_{\mathrm{Tu}}(t) = 0\}$*. Similarly* $\kappa_{\mathrm{Pu}} \in L^1$ *vanishes a.e. outside of* $\{t : \eta_{\mathrm{Pu}} y_{\mathrm{Pu}}(t) = k_{\mathrm{Pu}}\}$*, and* $\nu_{\mathrm{Pu}} \in L^{\infty *}$ *is concentrated on* $\{t : y_{\mathrm{Pu}}(t) \leq \epsilon\}$ *for each* $\epsilon > 0$*. If* $p \in L^1$*, then also* $\kappa_{\mathrm{Tu}} \in L^1$ *and* $\nu_{\mathrm{Pu}} \in L^1$ *(and then these functions vanish a.e. outside the sets* $\{t : y_{\mathrm{Tu}}(t) = k_{\mathrm{Tu}}\}$ *and* $\{t : y_{\mathrm{Pu}}(t) = 0\}$*, respectively).*

The following reformulation of the dual problem extends its pricing interpretation to the valuation of stock.

3.7 Shadow pricing of stock as the dual problem

By the change of variables from $(\lambda, \kappa_{\mathrm{St}}(\mathrm{d}t), \nu_{\mathrm{St}}(\mathrm{d}t))$ to

$$\psi(t) = \lambda + (\kappa_{\mathrm{St}} - \nu_{\mathrm{St}})[0,t] \quad \text{for } t \in (0,T) \tag{3.7.1}$$

and by using the dual constraints (3.5.4)–(3.5.6) and the disjointness condition (3.5.7) to eliminate the other dual variables, the dual problem can be transformed into one of *unconstrained* minimisation over ψ, an arbitrary bounded-variation function on $(0,T)$.

Notation The space $\mathrm{BV}(0,T)$ consists of all functions ψ of bounded variation on $(0,T)$ with $\psi(t)$ lying between the left and right limits, $\psi(t-) = \lim_{\tau \nearrow t} \psi(\tau)$ and $\psi(t+) = \lim_{\tau \searrow t} \psi(\tau)$.[15] A $\psi \in \mathrm{BV}(0,T)$ is extended by continuity to $[0,T]$; i.e., $\psi(0) := \psi(0+)$ and $\psi(T) := \psi(T-)$. The *cyclic positive variation* of ψ is defined by (3.3.1).

If finite numbers $\psi(0-)$ and $\psi(T+)$ are additionally specified, then $\psi \in \mathrm{BV}[0-,T+]$; and such a ψ defines a measure on $[0,T]$ by

$$\mathrm{d}\psi[t',t''] := \psi(t''+) - \psi(t'-) \tag{3.7.2}$$

for $t' \leq t''$. The Lebesgue-Stieltjes integral of a function s w.r.t. the measure $(\mathrm{d}\psi)^+$ is written as $\int s\,(\mathrm{d}\psi)^+$ or $\int s(t)\,(\mathrm{d}\psi(t))^+$. When $\psi(0-) = \psi(T+)$, the usual variation norm of $(\mathrm{d}\psi)^+$ equals $\mathrm{Var}_{\mathrm{c}}^{+}(\psi)$.

[15] The one-sided limits exist at every t and are equal nearly everywhere (n.e.), i.e., everywhere except for a countable set. Specification of $\psi(t)$ between $\psi(t-)$ and $\psi(t+)$ is unnecessary.

Formulae (3.7.1) and

$$\psi(0-) = \psi(T+) = \lambda \tag{3.7.3}$$

define together a one-to-one map of the set of all those $(\lambda, \kappa_{\mathrm{St}}, \nu_{\mathrm{St}})$ satisfying (3.5.4), (3.5.5) and (3.5.7) onto the set of all those $\psi \in \mathrm{BV}[0-, T+]$ with $\psi(0-) = \psi(T+)$ lying between $\psi(0+)$ and $\psi(T-)$. The inverse map is given by (3.7.3) together with

$$\kappa_{\mathrm{St}} = (\mathrm{d}\psi)^{+} \quad \text{and} \quad \nu_{\mathrm{St}} = (\mathrm{d}\psi)^{-}. \tag{3.7.4}$$

As for the variables κ_{Tu}, ν_{Tu}, κ_{Pu} and ν_{Pu}, these can now be eliminated by using (3.5.4), (3.5.6) and (3.5.7) to express them as

$$(\kappa_{\mathrm{Tu}}, \kappa_{\mathrm{Pu}}) = \left(\left(p - \frac{\psi}{\eta_{\mathrm{Tu}}} \right)^{+}, \left(\frac{p}{\eta_{\mathrm{Pu}}} - \psi \right)^{-} \right) \tag{3.7.5}$$

$$(\nu_{\mathrm{Tu}}, \nu_{\mathrm{Pu}}) = \left(\left(p - \frac{\psi}{\eta_{\mathrm{Tu}}} \right)^{-}, \left(\frac{p}{\eta_{\mathrm{Pu}}} - \psi \right)^{+} \right). \tag{3.7.6}$$

Proposition 3.7.1 (Stock pricing as the dual). *Assume (3.4.13). Then the fixed-input value minimisation programme (3.5.1)–(3.5.6) is equivalent, through the change of variables, to the following convex programme:*

$$\text{Given } (p, k_{\mathrm{St}}, k_{\mathrm{Tu}}, k_{\mathrm{Pu}}) \text{ as in (3.4.5), minimise} \tag{3.7.7}$$

$$k_{\mathrm{St}} \operatorname{Var}_{\mathrm{c}}^{+}(\psi) + k_{\mathrm{Tu}} \left\langle \left(p - \frac{\psi}{\eta_{\mathrm{Tu}}} \right)^{+} \mid 1 \right\rangle + k_{\mathrm{Pu}} \int_0^T \left(\frac{p(t)}{\eta_{\mathrm{Pu}}} - \psi(t) \right)^{-} \mathrm{d}t \tag{3.7.8}$$

$$\text{over } \psi \in \mathrm{BV}(0, T). \tag{3.7.9}$$

Notation The (optimal) solution set for (3.7.7)–(3.7.9) is denoted by $\hat{\Psi}_{\mathrm{PS}}(p, k)$, abbreviated to $\hat{\Psi}$. Again, the corresponding lowercase notation, $\hat{\psi}$, is used *only when the dual solution is unique.*

The function ψ defined by (3.7.1) can be interpreted as the *shadow price of stock* at any time t. Heuristically, this follows from the marginal interpretations of κ, ν and λ (viz., that κ_{St}, as the multiplier for the upper reservoir constraint, represents the reservoir capacity value, with a similar interpretation of the multiplier ν_{St} for the lower constraint, whilst λ is the stock value at the beginning of cycle).[16]

It is this formulation of the dual that leads to the idea of obtaining $\hat{\psi}$ by "levelling off" the local extremes of p in the way described in Section 3.3. The insight can be developed into a specialised algorithm for the case of a piecewise monotone p. In this approach the dual is tackled first, in the

[16] The shadow-price interpretation of ψ can be formalised as a rigorous marginal-value result by introducing a (hypothetical) exogeneous inflow to the reservoir, $e \in L^{\infty}$, as an additional parameter with its own multiplier ψ. This means that (3.4.9) and (3.4.10) are perturbed by replacing y with $y - \Delta e$. Then (3.7.1) becomes a constraint of the dual problem, whose solution $\hat{\psi}_{\mathrm{PS}}$ equals $\nabla_e \Pi_{\mathrm{SR}}^{\mathrm{PS}}$ at $e = 0$ (in pumped storage). This is formally similar to the case of hydro, in which e is the river flow, and $\hat{\psi}_{\mathrm{H}}$ equals $\nabla_e \Pi_{\mathrm{SR}}^{\mathrm{H}}$ at the given, positive e: see Chapter 4.

CP form (3.7.7)–(3.7.9), with the primal solution found subsequently. For comparison, the simplex method for LPs finds both the primal and the dual solutions simultaneously.

3.8 Determination of optimal storage policy

A storage plant's operation problem is soluble for every $p \in L^1[0,T]$, though not for every $p \in L^{\infty *}$. The assumption that $p \in L^1$ (i.e., that $p_{\mathrm{FA}} = 0$) is maintained from here on until Remark 3.14.6.

Proposition 3.8.1 (Primal solubility). *For every $p \in L^1_{++}$ and $k = (k_{\mathrm{St}}, k_{\mathrm{Tu}}, k_{\mathrm{Pu}}) \geq 0$, the short-run profit-maximisation programme (3.4.5)–(3.4.10) has a solution $(y_{\mathrm{Tu}}, y_{\mathrm{Pu}}, s_0)$. Equivalently, the problem (3.4.1)–(3.4.3) has a solution, i.e., $\hat{Y}(p,k) \neq \emptyset$.*

When $\eta_{\mathrm{Ro}} < 1$, simultaneous charging and discharging would be counterproductive (if it were at all feasible). This is next stated formally (since it is used in proving the optimal-output formula).

Lemma 3.8.2 (Nonoptimality of conversion overlap). *Assume that $\eta_{\mathrm{Ro}} < 1$ and $p \in L^1_{++}$ (or, more generally, $p \in L^{\infty *}_{+}$ and $p_{\mathrm{CA}} \gg 0$). Then $y_{\mathrm{Tu}} \wedge y_{\mathrm{Pu}} = 0$ for every solution to the primal (3.4.5)–(3.4.10). So $f = y^+/\eta_{\mathrm{Tu}} - \eta_{\mathrm{Pu}} y^-$ from (3.2.2).*

Once the dual is solved, so that an optimal ψ is known, the operation problem largely reduces to maximisation of instantaneous profits (as Part 2c of Proposition 3.6.1 shows). At each t with $\eta_{\mathrm{Tu}} p(t) \neq \psi(t) \neq p(t)/\eta_{\mathrm{Pu}}$, the optimum output $y(t)$ is of the "bang-coast-bang control" type, either k_{Tu} or 0 or $-k_{\mathrm{Pu}}/\eta_{\mathrm{Pu}}$ (when $\eta_{\mathrm{Ro}} = 1$, this simplifies to a "bang-bang" y on $\{p \neq \psi\}$). Any remaining part of an optimal y is a "singular control", which arises at a time t when the instantaneous optimum is multi-valued because $\psi(t)$ equals $\eta_{\mathrm{Tu}} p(t)$ or $p(t)/\eta_{\mathrm{Pu}}$. This part can be determined on the assumption (3.8.1) that p has no plateau: this ensures that $\eta_{\mathrm{Tu}} p(t) = \psi(t)$ or $\psi(t) = p(t)/\eta_{\mathrm{Pu}}$ only when the reservoir is full or empty (respectively); and at those times the output rate must be 0. See Figure 3.2 (for $\eta_{\mathrm{Ro}} = 1$) and Figure 3.4 (for $\eta_{\mathrm{Ro}} < 1$).

Proposition 3.8.3 (PS output with plateau-less price). *In addition to $p \in L^1_{++}[0,T]$ and $k = (k_{\mathrm{St}}, k_{\mathrm{Tu}}, k_{\mathrm{Pu}}) \gg 0$, assume that p has no plateau, i.e., that*

$$\forall \mathrm{p} \in \mathbb{R}_+ \quad \mathrm{meas}\{t : p(t) = \mathrm{p}\} = 0. \tag{3.8.1}$$

If $y \in \hat{Y}(p,k)$ and $\psi \in \hat{\Psi}(p,k)$, i.e., y solves (3.4.1)–(3.4.3) and ψ solves (3.7.7)–(3.7.9), then

$$y(t) = \begin{cases} k_{\mathrm{Tu}} & \text{if } \eta_{\mathrm{Tu}} p(t) > \psi(t) \\ 0 & \text{if } \eta_{\mathrm{Tu}} p(t) \leq \psi(t) \leq p(t)/\eta_{\mathrm{Pu}} \\ -k_{\mathrm{Pu}}/\eta_{\mathrm{Pu}} & \text{if } p(t)/\eta_{\mathrm{Pu}} < \psi(t) \end{cases}. \tag{3.8.2}$$

So (3.4.1)–(3.4.3) has a unique solution $\hat{y}(p,k)$.

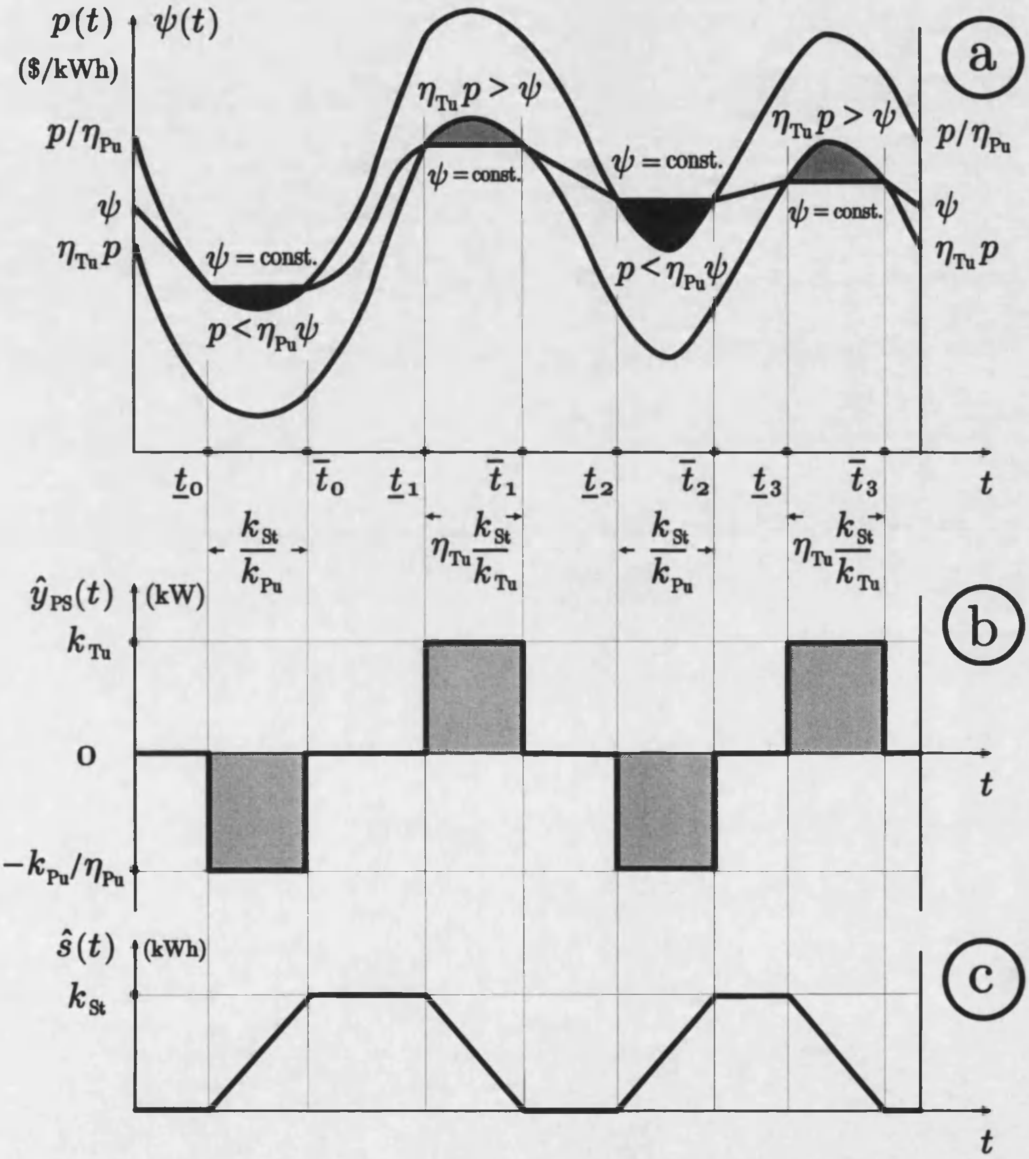

FIGURE 3.4. Trajectories of: (a) an optimal shadow price of stock ψ, (b) the output of pumped-storage plant $\hat{y}_{PS}$, and (c) the stock $\hat{s}$, in the case of imperfect conversion ($\eta_{Ro} = \eta_{Tu}\eta_{Pu} < 1$). Unit rent for storage capacity is $\mathrm{Var}_c^+(\psi)$, the sum of rises of ψ. Unit rent for the pump capacity is $\int_0^T (p(t)/\eta_{Pu} - \psi(t))^- \, dt$, the sum of black areas in (a). Unit rent for the turbine capacity is $\int_0^T (p(t) - \psi(t)/\eta_{Tu})^+ \, dt$, the sum of dark grey areas in (a) times $1/\eta_{Tu}$. In (b), each of the light grey areas equals the reservoir's capacity k_{St}.

3.9 Marginal capacity values in terms of stock price

By definition, $\Pi_{\mathrm{SR}}^{\mathrm{PS}}$ is the optimal value, $\max_y \langle p \mid y \rangle$, of the primal (operation) problem. Since the dual and primal values are equal (Proposition 3.6.1), a dual (stock-pricing) solution ψ gives Π as the total capacity value (the plant's total rent); and it has the advantage of giving the marginal values $\nabla_k \Pi$ as well.

Corollary 3.9.1 (Dual calculation of SR profit). *Assume that* $k = (k_{\mathrm{St}}, k_{\mathrm{Tu}}, k_{\mathrm{Pu}}) \gg 0$ *and* $p \in L^1_{++}$. *Then, for every* $\psi \in \hat{\Psi}(p,k)$,

$$\Pi_{\mathrm{SR}}^{\mathrm{PS}}(p,k) = k_{\mathrm{St}} \operatorname{Var}_{\mathrm{c}}^{+}(\psi) + k_{\mathrm{Tu}} \int_0^T \left(p(t) - \frac{\psi(t)}{\eta_{\mathrm{Tu}}} \right)^{+} \mathrm{d}t + k_{\mathrm{Pu}} \int_0^T \left(\frac{p(t)}{\eta_{\mathrm{Pu}}} - \psi(t) \right)^{-} \mathrm{d}t. \tag{3.9.1}$$

Furthermore, this sum equals

$$\int_0^T \psi(t) f(t)\,\mathrm{d}t + \int_0^T \left(p(t) - \frac{\psi(t)}{\eta_{\mathrm{Tu}}} \right) y^{+}(t)\,\mathrm{d}t + \int_0^T \left(\eta_{\mathrm{Pu}} \psi(t) - p(t) \right) y^{-}(t)\,\mathrm{d}t$$

term-by-term, for every $y \in \hat{Y}(p,k)$.[17]

Since Π is positively homogeneous of degree 1 (a.k.a. linearly homogeneous) in k,[18] Euler's Theorem shows that if Π is differentiable in k,[19] then

$$\Pi_{\mathrm{SR}}^{\mathrm{PS}}(p,k) = k_{\mathrm{St}} \frac{\partial \Pi_{\mathrm{SR}}^{\mathrm{PS}}}{\partial k_{\mathrm{St}}} + k_{\mathrm{Tu}} \frac{\partial \Pi_{\mathrm{SR}}^{\mathrm{PS}}}{\partial k_{\mathrm{Tu}}} + k_{\mathrm{Pu}} \frac{\partial \Pi_{\mathrm{SR}}^{\mathrm{PS}}}{\partial k_{\mathrm{Pu}}}. \tag{3.9.2}$$

A comparison with (3.9.1) suggests that if there is a *unique* optimal ψ, then the partial derivatives of Π do exist and equal the coefficients of k_{St}, k_{Tu} and k_{Pu} in (3.9.1); formally this follows from (3.7.1) and the marginal interpretation of κ_{St}, κ_{Tu} and κ_{Pu} (spelt out in the Proof of Theorem 3.9.2). And the optimal stock price ψ is indeed unique at the times which matter if p, the TOU price of the good, is continuous. The result extends to the case of a $p \in L^{\infty *}$, if p_{CA} is continuous (Remark 3.14.6).

Theorem 3.9.2 (Efficiency rents of a storage plant). *Assume that* $p \in \mathcal{C}_{++}[0,T]$. *Then the operating profit of a pumped-storage plant—i.e., the value of the primal problem (3.4.1)–(3.4.3)—is differentiable with respect to the capacities (of the reservoir and the conversion equipment), at every* $k = (k_{\mathrm{St}}, k_{\mathrm{Tu}}, k_{\mathrm{Pu}}) \gg 0$. *The derivatives defining the unit rents are given by the formulae*

$$\frac{\partial \Pi_{\mathrm{SR}}^{\mathrm{PS}}}{\partial k_{\mathrm{St}}}(p,k) = \operatorname{Var}_{\mathrm{c}}^{+}(\psi) \tag{3.9.3}$$

[17] This shows that the capacity values are equal to the capacities' profits—$\langle \psi \mid f \rangle$ for the reservoir, etc.—when the shadow price ψ is used to decentralise the operation within the plant (as is described in Section 3.3.2).

[18] That is, $\Pi(p, \alpha k) = \alpha \Pi(p,k)$ for every scalar $\alpha > 0$. Note also that $\hat{Y}$ and $\hat{\Psi}$ are positively homogeneous, in k, of degrees 1 and 0 respectively; i.e., $\hat{Y}(p, \alpha k) = \alpha \hat{Y}(p,k)$ and $\hat{\Psi}(p, \alpha k) = \hat{\Psi}(p,k)$ for $\alpha > 0$.

[19] If Π is nondifferentiable, then $\Pi(k) = r \cdot k$ for every $r \in \hat{\partial}_k \Pi$ (the superdifferential of Π as a concave function of k).

$$\frac{\partial \Pi_{\mathrm{SR}}^{\mathrm{PS}}}{\partial k_{\mathrm{Tu}}}(p,k) = \int_0^T \left(p(t) - \frac{\psi(t)}{\eta_{\mathrm{Tu}}}\right)^+ \mathrm{d}t \tag{3.9.4}$$

$$\frac{\partial \Pi_{\mathrm{SR}}^{\mathrm{PS}}}{\partial k_{\mathrm{Pu}}}(p,k) = \int_0^T \left(\frac{p(t)}{\eta_{\mathrm{Pu}}} - \psi(t)\right)^- \mathrm{d}t \tag{3.9.5}$$

in which ψ is any solution to the dual problem (3.7.7)–(3.7.9) of shadow pricing the stock. (The above values are the same for every $\psi \in \hat{\Psi}(p,k)$, and this set is nonempty by Proposition 3.6.1. If additionally $\eta_{\mathrm{Ro}} = 1$, then there is a unique dual solution, $\hat{\psi}(p,k)$.)

Comment: In the case of a *reversible* converter of capacity k_{Co} one obtains, by setting $k_{\mathrm{Tu}} = \beta k_{\mathrm{Co}}$ and $k_{\mathrm{Pu}} = k_{\mathrm{Co}}$ in (3.7.8) and by adding up the two integrals (3.9.4)–(3.9.5), that

$$\frac{\partial \Pi_{\mathrm{SR}}^{\mathrm{PS}}}{\partial k_{\mathrm{Co}}} = \int_0^T \left(\beta\left(p(t) - \frac{\psi(t)}{\eta_{\mathrm{Tu}}}\right)^+ + \left(\frac{p(t)}{\eta_{\mathrm{Pu}}} - \psi(t)\right)^-\right) \mathrm{d}t$$

where ψ is any solution to (3.7.7)–(3.7.9) with the above substitutions. The integral simplifies to $\int_0^T |p(t) - \psi(t)|\, \mathrm{d}t$ if the converter is symmetric and perfectly efficient (i.e., if $\beta = 1$ and $\eta_{\mathrm{Ro}} = 1$).

3.10 Bounds on marginal capacity values

For this and the next sections the conversion equipment is assumed to be reversible (but not necessarily perfectly efficient or symmetric). Recall from Section 3.2 that a unit converter provides a unit of pump capacity (with efficiency η_{Pu}) and β units of turbine capacity (with efficiency η_{Tu}).

Since $\Pi_{\mathrm{SR}}^{\mathrm{PS}}(p, k_{\mathrm{St}}, k_{\mathrm{Co}})$ is, by Proposition 3.7.1 (with $p \in L^1$), the minimum of

$$k_{\mathrm{St}} \operatorname{Var}_c^+(\psi) + k_{\mathrm{Co}} \int_0^T \left(\beta\left(p(t) - \frac{\psi(t)}{\eta_{\mathrm{Tu}}}\right)^+ + \left(\frac{p(t)}{\eta_{\mathrm{Pu}}} - \psi(t)\right)^-\right) \mathrm{d}t$$

over $\psi \in \mathrm{BV}$, an upper bound on Π/k_{Co} that depends only on p can be obtained by setting $\psi = \mathrm{const}$. Assume that $p \in \mathcal{C}$; then the best (minimising) constant value for ψ is unique, and it is the $\psi \in \mathbb{R}$ that satisfies (3.13.25)–(3.13.26). It is denoted by

$$\mathrm{gq}(p, \beta, \eta_{\mathrm{Tu}}, \eta_{\mathrm{Pu}})$$

since it generalises the quantile concept: when $\eta_{\mathrm{Pu}}\eta_{\mathrm{Tu}} = 1$, $\mathrm{gq}(p,\beta)$ is the lower quantile of p of order $\beta/(\beta+1)$.[20] From this and from (3.9.2)

$$\frac{\partial \Pi_{\mathrm{SR}}^{\mathrm{PS}}}{\partial k_{\mathrm{Co}}} \le \frac{\Pi_{\mathrm{SR}}^{\mathrm{PS}}}{k_{\mathrm{Co}}} \le \overline{r}_{\mathrm{Co}} \tag{3.10.1}$$

$$:= \int_0^T \left(\beta\left(p(t) - \frac{\mathrm{gq}(p,\beta,\eta)}{\eta_{\mathrm{Tu}}}\right)^+ + \left(\frac{p(t)}{\eta_{\mathrm{Pu}}} - \mathrm{gq}(p,\beta,\eta)\right)^-\right) \mathrm{d}t.$$

Choosing a ψ between $\eta_{\mathrm{Tu}} p$ and p/η_{Pu} shows similarly that

$$\frac{\partial \Pi_{\mathrm{SR}}^{\mathrm{PS}}}{\partial k_{\mathrm{St}}} \le \frac{\Pi_{\mathrm{SR}}^{\mathrm{PS}}}{k_{\mathrm{St}}} \le \overline{r}_{\mathrm{St}} := \inf_{\psi} \left\{ \operatorname{Var}_c^+(\psi) : \eta_{\mathrm{Tu}} p \le \psi \le \frac{p}{\eta_{\mathrm{Pu}}} \right\} \tag{3.10.2}$$

[20] For the quantile's definition, see (3.13.10) with $k_{\mathrm{Tu}} = \beta k_{\mathrm{Co}}$ and $k_{\mathrm{Pu}} = k_{\mathrm{Co}}$.

This bound is finite if, as is assumed henceforth, $\eta_{\mathrm{Ro}} < 1$ or $p \in \mathrm{BV}$ (in the latter case $\overline{r}_{\mathrm{St}} \leq \mathrm{Var}_{\mathrm{c}}^{+}(p) < +\infty$).

With p fixed, $\nabla_k \Pi$ is homogeneous of degree 0 in k, i.e., it depends only on the capacity ratio $\vartheta := k_{\mathrm{Co}}/k_{\mathrm{St}}$. As ϑ increases from 0 to $+\infty$, $\partial\Pi/\partial k_{\mathrm{St}}$ increases, whilst $\partial\Pi/\partial k_{\mathrm{Co}}$ decreases to 0 (in the limit as $\vartheta \nearrow +\infty$), since

$$\frac{k_{\mathrm{Co}}}{k_{\mathrm{St}}} \frac{\partial \Pi_{\mathrm{SR}}^{\mathrm{PS}}}{\partial k_{\mathrm{Co}}} \leq \frac{\Pi_{\mathrm{SR}}^{\mathrm{PS}}}{k_{\mathrm{St}}} \leq \overline{r}_{\mathrm{St}}$$

by (3.9.2) and the second inequality of (3.10.2). A similar argument using (3.10.1) shows that

$$\frac{k_{\mathrm{St}}}{k_{\mathrm{Co}}} \frac{\partial \Pi_{\mathrm{SR}}^{\mathrm{PS}}}{\partial k_{\mathrm{St}}} \leq \frac{\Pi_{\mathrm{SR}}^{\mathrm{PS}}}{k_{\mathrm{Co}}} \leq \overline{r}_{\mathrm{Co}}$$

so $\partial\Pi/\partial k_{\mathrm{St}}$ decreases to 0 as $\vartheta \searrow 0$. Furthermore, whilst $\partial\Pi/\partial k_{\mathrm{Co}}$ may never equal 0,[21] $\partial\Pi/\partial k_{\mathrm{St}}$ actually *is* 0 for small enough $\vartheta = k_{\mathrm{Co}}/k_{\mathrm{St}}$. This is obviously the case for $\vartheta < (\beta + \eta_{\mathrm{Tu}})/\beta T$: an extra unit of the reservoir is then useless because it is already so large that it cannot be fully charged and discharged in one cycle (since this takes $k_{\mathrm{St}}/k_{\mathrm{Co}}$ plus $\eta_{\mathrm{Tu}} k_{\mathrm{St}}/\beta k_{\mathrm{Co}}$, which exceeds T). The largest ϑ with $\partial\Pi/\partial k_{\mathrm{St}} = 0$ is denoted by $\underline{\vartheta}$. (It can be given explicitly in terms of p, β and η.) Note that $\partial\Pi/\partial k_{\mathrm{Co}}$ attains its upper bound at (and only at) $\vartheta \leq \underline{\vartheta}$.[22] See Figure 3.5a.

3.11 Optimum investment in storage plants

The marginal capacity values $\nabla_k \Pi_{\mathrm{SR}}^{\mathrm{PS}}$ can be used to determine the optimum investment into pumped storage on the basis of a given TOU tariff p and the supply costs of the two inputs, the reservoir and the (reversible) converter. The following formulation of the problem applies chiefly to energy storage techniques such as PWES and CAES, which utilise special geological features. The converter's unit cost, r_{Co}, can be reasonably regarded as constant, i.e., independent of the capacity k_{Co}. By contrast, in PWES or CAES the reservoir's marginal cost, r_{St}, typically increases with k_{St} because the most suitable parts of the site are developed first. In formal terms, on a potential site for a particular storage technique, a reservoir can be built at a cost which is a strictly convex and increasing function, G, of its capacity $k_{\mathrm{St}} \in [0, \overline{k}_{\mathrm{St}}]$, with $G(0) = 0$. Although G may be nondifferentiable, it has the one-sided, left and right derivatives, $\mathrm{d}G/\mathrm{d}_{-}k_{\mathrm{St}} \leq \mathrm{d}G/\mathrm{d}_{+}k_{\mathrm{St}}$. Where these differ, the subdifferential $\partial G = [\mathrm{d}G/\mathrm{d}_{-}k_{\mathrm{St}}, \mathrm{d}G/\mathrm{d}_{+}k_{\mathrm{St}}]$ is multi-valued; but this can be the case only on a countable subset of $(0, \overline{k}_{\mathrm{St}})$. In other words, the two-sided derivative $\mathrm{d}G/\mathrm{d}k_{\mathrm{St}}$ exists nearly everywhere; and its right or left limit equals $\mathrm{d}G/\mathrm{d}_{\pm}k_{\mathrm{St}}$, respectively. Also, $\partial G(0) = [0, (\mathrm{d}G/\mathrm{d}k_{\mathrm{St}})(0+)]$ and $\partial G(\overline{k}_{\mathrm{St}}) = [(\mathrm{d}G/\mathrm{d}k_{\mathrm{St}})(\overline{k}_{\mathrm{St}}-), +\infty)$. See Figure 3.5b.

The investment problem is:

$$\text{Given } (p, r_{\mathrm{Co}}) \in \mathcal{C}[0,T] \times \mathbb{R}_{++} \text{ (and given the function } G) \tag{3.11.1}$$

[21] If the local peaks and troughs of p are strict, and $\mathrm{Min}(p) < \eta_{\mathrm{Ro}} \mathrm{Max}(p)$ as in Remark 3.14.1, then an extra unit of converter is always useful because it allows conversion to be concentrated closer to the troughs and peaks.

[22] If $\vartheta \leq \underline{\vartheta}$, then $\partial\Pi/\partial k_{\mathrm{St}} = 0$; i.e., $\hat{\psi} = \mathrm{gq}(p, \beta, \eta)$, so equalities hold in (3.10.1).

$$\text{maximise } \Pi_{\mathrm{SR}}^{\mathrm{PS}}\left(p, k_{\mathrm{St}}, k_{\mathrm{Co}}\right) - G\left(k_{\mathrm{St}}\right) - r_{\mathrm{Co}} k_{\mathrm{Co}} \text{ over } \left(k_{\mathrm{St}}, k_{\mathrm{Co}}\right) \in \mathbb{R}_{+}^{2}. \tag{3.11.2}$$

It can be solved in two stages, first for the proportion $\vartheta := k_{\mathrm{Co}}/k_{\mathrm{St}}$, and then for the scale: since Π is positively linearly homogeneous in k,

$$\Pi_{\mathrm{SR}}^{\mathrm{PS}}\left(p, k_{\mathrm{St}}, k_{\mathrm{Co}}\right) - G\left(k_{\mathrm{St}}\right) - r_{\mathrm{Co}} k_{\mathrm{Co}} = k_{\mathrm{St}} \left(\Pi_{\mathrm{SR}}^{\mathrm{PS}}\left(p, 1, \frac{k_{\mathrm{Co}}}{k_{\mathrm{St}}}\right) - \frac{r_{\mathrm{Co}} k_{\mathrm{Co}}}{k_{\mathrm{St}}}\right) - G\left(k_{\mathrm{St}}\right) \tag{3.11.3}$$

for $k_{\mathrm{St}} > 0$; and—with p suppressed from the notation—the subproblem of maximising

$$\Pi_{\mathrm{SR}}^{\mathrm{PS}}\left(1, \vartheta\right) - r_{\mathrm{Co}} \vartheta \tag{3.11.4}$$

over $\vartheta \in \mathbb{R}_{+}$ can be solved first. Once a maximum point ϑ is known, it only remains to maximise

$$k_{\mathrm{St}} \left(\Pi_{\mathrm{SR}}^{\mathrm{PS}}\left(1, \vartheta\right) - r_{\mathrm{Co}} \vartheta\right) - G\left(k_{\mathrm{St}}\right) \tag{3.11.5}$$

over $k_{\mathrm{St}} \in \mathbb{R}_{+}$. The solution gives the other optimum capacity as $k_{\mathrm{Co}} = \vartheta k_{\mathrm{St}}$.

When $\max_{\vartheta} \left(\Pi\left(1, \vartheta\right) - r_{\mathrm{Co}} \vartheta\right) \leq \left(\mathrm{d}G/\mathrm{d}k_{\mathrm{St}}\right)(0+)$, the maximum of (3.11.5) is at $k_{\mathrm{St}} = 0$, and this means that the maximum of (3.11.2) is at $\left(k_{\mathrm{St}}, k_{\mathrm{Co}}\right) = (0, 0)$. Therefore a necessary condition for a nonzero solution to (3.11.2) is that r_{Co} is less than the $\overline{r}_{\mathrm{Co}}$ defined in (3.10.1). This is because, from Section 3.10, $\overline{r}_{\mathrm{Co}}$ is the maximum of $\partial \Pi/\partial k_{\mathrm{Co}}$, so $\max_{\vartheta} \left(\Pi\left(1, \vartheta\right) - r_{\mathrm{Co}} \vartheta\right) = 0$ if $r_{\mathrm{Co}} \geq \overline{r}_{\mathrm{Co}}$.

Given any $r_{\mathrm{Co}} < \overline{r}_{\mathrm{Co}}$, a ϑ maximising (3.11.4) can be found from the FOC[23]

$$\left.\frac{\partial \Pi_{\mathrm{SR}}^{\mathrm{PS}}}{\partial k_{\mathrm{Co}}}\right|_{\left(k_{\mathrm{St}}, k_{\mathrm{Co}}\right) = (1, \vartheta)} = r_{\mathrm{Co}}. \tag{3.11.6}$$

This has a solution because $\partial \Pi/\partial k_{\mathrm{Co}} \searrow 0$ as $\vartheta \nearrow +\infty$, at least if $\eta_{\mathrm{Ro}} < 1$ or $p \in \mathrm{BV}$: see Section 3.10. In general the maximum points of (3.11.4) form a (nonempty) subinterval of $(\underline{\vartheta}, +\infty)$, but if p has no plateau, then the solution is actually unique, in which case it is denoted by $\vartheta^{*}\left(r_{\mathrm{Co}}\right)$, as in Figure 3.5a.

Given an optimum ϑ, the k_{St} maximising (3.11.5) can be found from the condition $\Pi\left(1, \vartheta\right) - r_{\mathrm{Co}} \vartheta \in \partial G\left(k_{\mathrm{St}}\right)$, which is equivalent to

$$\left.\frac{\partial \Pi_{\mathrm{SR}}^{\mathrm{PS}}}{\partial k_{\mathrm{St}}}\right|_{\left(k_{\mathrm{St}}, k_{\mathrm{Co}}\right) = (1, \vartheta)} \in \partial G\left(k_{\mathrm{St}}\right) \tag{3.11.7}$$

by (3.11.6) and (3.9.2). Since G is strictly convex, the solution for k_{St} is unique: see Figure 3.5b.

In summary, given an $r_{\mathrm{Co}} < \overline{r}_{\mathrm{Co}}$ and a plateau-less continuous p of bounded variation, there is a unique optimum investment, $k_{\mathrm{St}}^{*}\left(G, r_{\mathrm{Co}}\right)$ and $k_{\mathrm{Co}}^{*}\left(G, r_{\mathrm{Co}}\right)$, which can be found by using $\nabla_{k} \Pi$: first (3.11.6) is solved to obtain $\vartheta^{*}\left(r_{\mathrm{Co}}\right)$, and then (3.11.7) with $\vartheta = \vartheta^{*}$ is solved to obtain k_{St}^{*} and hence also $k_{\mathrm{Co}}^{*} = \vartheta^{*} k_{\mathrm{St}}^{*}$.[24]

Comments:

[23] This can be solved numerically by, e.g., the secant method—which requires no more than the calculation of $\partial \Pi/\partial k_{\mathrm{Co}}$ at the successive approximations.

[24] The procedure is valid also when $k_{\mathrm{St}}^{*} = 0$ because ϑ^{*} is the *candidate* for the optimum capacity ratio, and it can be found without presupposing that the ratio is well defined (i.e., that $k_{\mathrm{St}}^{*} > 0$).

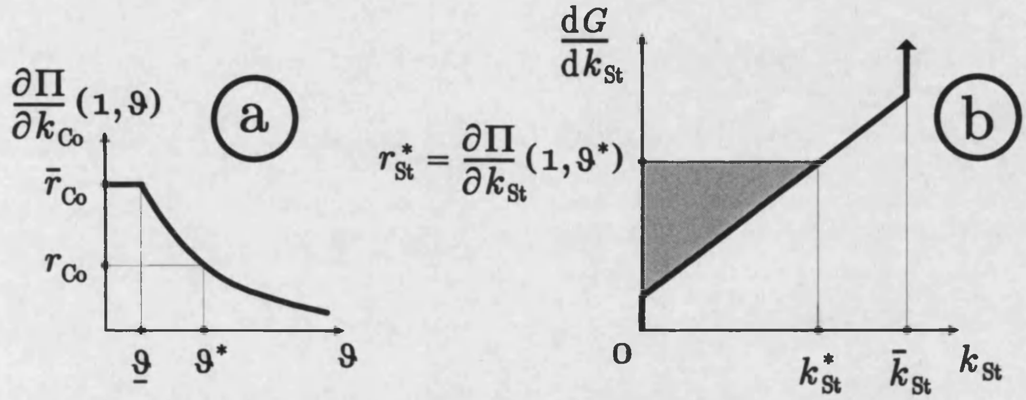

FIGURE 3.5. Optimal investment on a storage site: determination of (a) the capacity ratio ϑ^* (given r_{Co}), and (b) the reservoir's capacity k^*_{St} (and hence the converter's capacity $k^*_{\mathrm{Co}} = \vartheta^* k^*_{\mathrm{St}}$). The shaded area in (b) represents the site's rent.

- The maximum of (3.11.2) equals $r^*_{\mathrm{St}} k^*_{\mathrm{St}} - G\left(k^*_{\mathrm{St}}\right)$, where

$$r^*_{\mathrm{St}}\left(r_{\mathrm{Co}}\right) := \frac{\partial \Pi^{\mathrm{PS}}_{\mathrm{SR}}}{\partial k_{\mathrm{St}}}\left(1, \vartheta^*\left(r_{\mathrm{Co}}\right)\right) = \frac{\partial \Pi^{\mathrm{PS}}_{\mathrm{SR}}}{\partial k_{\mathrm{St}}}\left(k^*_{\mathrm{St}}\left(G, r_{\mathrm{Co}}\right), k^*_{\mathrm{Co}}\left(G, r_{\mathrm{Co}}\right)\right).$$

  Since $r^*_{\mathrm{St}} \in \partial G\left(k^*_{\mathrm{St}}\right)$, this is the price for storage capacity that would induce a price-taking owner of the site to build a reservoir of the optimum size k^*_{St}, to be optimally complemented by k^*_{Co} of the converter. In practice the site owner is likely to either build a complete plant himself or let the site for a rent to the highest-bidding entrepreneur. With perfect competition the entrepreneur's net profit is zero, i.e., the rent for the site is $r^*_{\mathrm{St}} k^*_{\mathrm{St}} - G\left(k^*_{\mathrm{St}}\right)$ per cycle (the shaded area in Figure 3.5b).

- The analysis obviously extends to any number of sites and techniques (for storing the same good with the tariff p). On all of the sites for a particular storage technique the optimum capacity ratio ϑ^* is the same, since it depends only on r_{Co}.

- The independence of ϑ^* on G gives a simple but useful comparative statics result: a fall in the marginal cost schedule to some $\mathrm{d}G'/\mathrm{d}k_{\mathrm{St}} \le \mathrm{d}G/\mathrm{d}k_{\mathrm{St}}$ changes the scale of optimum investment but *not* the optimal capacity ratio. So if the reservoir construction cost falls to G' after an investment on the basis of G, optimality can be restored by a proportional expansion of the existing plant. (This is usually feasible with sizeable projects, which are planned to be carried out in stages.)

3.12 Proofs for Sections 3.5 to 3.8

These proofs are largely routine applications of duality for optimisation in infinite-dimensional spaces, as expounded in, e.g., [73, Examples 4, 4', 4"] and [4, 3.3–3.7]. To put the primal constraints in the operator form required by this framework, define the integrals I_0 and I_T: $L^\infty[0,T] \to \mathcal{C}[0,T]$ by

$$(I_0 f)(t) := \int_0^t f(\tau)\,\mathrm{d}\tau \quad \text{and} \quad (I_T f)(t) := \int_t^T f(\tau)\,\mathrm{d}\tau. \tag{3.12.1}$$

The reservoir constraints (3.4.10) on $(y_{\mathrm{Tu}}, y_{\mathrm{Pu}}, s_0)$ can then be rewritten as

$$0 \le s_0 1_{[0,T]} - I_0 f \le k_{\mathrm{St}}. \tag{3.12.2}$$

A formula for the adjoint operation I_0^*: $\mathcal{M}[0,T] \to L^{\infty *}[0,T]$ is needed. (As for the embedding $\mathbb{R} \ni \mathrm{s} \mapsto \mathrm{s}1_{[0,T]} \in \mathcal{C}$, its adjoint is: $\mathcal{M} \ni \kappa \mapsto \langle \kappa \,|\, 1 \rangle = \kappa[0,T]$.)

Lemma 3.12.1. *The adjoints I_0^*, I_T^* map $\mathcal{M}[0,T]$ into* $\mathrm{BV}[0,T] \subset L^1[0,T]$. *They are given by*

$$(I_0^* \mu)(t) = \mu[t,T] \quad \text{and} \quad (I_T^* \mu)(t) = \mu[0,t] \quad \text{for a.e. } t,$$

for every $\mu \in \mathcal{M}$. If $\mu[0,T] = 0$, then $-I_0^*\mu = \mu[0,\cdot] = I_T^*\mu$.

Proof. The linear operation I_0: $L^\infty \to \mathcal{C}[0,T]$ is obviously norm-to-norm continuous, so its adjoint maps $\mathcal{M}[0,T]$ into $L^{\infty *}$. To calculate I_0^*, use Fubini's Theorem: for $\mu \in \mathcal{M}[0,T]$ and $f \in L^\infty$,

$$\langle f \,|\, I_0^*\mu \rangle := \langle I_0 f \,|\, \mu \rangle = \int_{[0,T]} \int_0^t f(\tau)\,\mathrm{d}\tau\,\mu(\mathrm{d}t) = \int_0^T f(\tau)\,\mu[\tau,T]\,\mathrm{d}\tau.$$

This means that $I_0^*\mu$ is represented by the function equal a.e. to $\mu[\cdot,T]$; so it belongs to $\mathrm{BV} \subset L^1$. A similar argument applies to I_T^*. To complete, note that $\mu[\cdot,T] = \mu(\cdot,T]$ a.e. (and actually n.e.). □

Remark 3.12.2. *The operations I_0, I_T: $L^\infty \to \mathcal{C}$ are $\mathrm{m}(L^\infty, L^1)$-to-$\|\cdot\|_\infty$ continuous, where $\mathrm{m}(L^\infty, L^1)$ is the Mackey topology on L^∞ for the duality with L^1.*

Proof. For I_0, this follows directly from the definition (3.12.1), used in conjunction with two facts: (i) that the 0-1 indicators $\{1_{[0,t]} : t \in [0,T]\}$ form a uniformly integrable subset of L^1, and (ii) that $\mathrm{m}(L^\infty, L^1)$ can be characterised as the topology of uniform convergence on uniformly integrable sets (since these are the same as the weak relative compacts of L^1, by the Dunford-Pettis Criterion). A similar argument using $1_{[t,T]}$ applies to I_T^*. □

Comment: That I_0^* and I_T^* map $\mathcal{M}$ into L^1 follows also directly, without any calculation, from their Mackey-to-norm continuity (i.e., from Remark 3.12.2).

Proof of Theorem 3.5.1 (Fixed-input value minimisation as the dual). Since (3.4.5)–(3.4.10) is an LP, it would suffice to apply results such as those of [4, 3.3 and 3.6–3.7]. However, to facilitate extensions and adaptations requiring convex but nonlinear models, this proof is couched in CP terms. The dual to a concave maximisation programme consists in minimising, over the dual variables (the Lagrange multipliers for the primal), the supremum of the Lagrange function over the primal decision variables: see, e.g., [73, (4.6) and (5.13)]. The "cone model" of [73, Example 4']—also expounded in, e.g., [16, 4.2] and [57, 7.9]—is applicable, since (3.12.2) and (3.4.7)–(3.4.8) represent the inequality constraints of the primal programme (3.4.5)–(3.4.10) by means of the nonnegative cones ($\mathcal{C}_+$ and L_+^∞) and convex constraint maps (which are actually linear). The dual variables here are the κ_{St}, ν_{St}; κ_{Tu}, ν_{Tu}, κ_{Pu}, ν_{Pu} and λ of (3.5.3); and these are paired with the parameter increments Δk_{St}, $-\Delta n_{\mathrm{St}}$, Δk_{Tu}, $-\Delta n_{\mathrm{Tu}}$, Δk_{Pu}, $-\Delta n_{\mathrm{Pu}}$ and $\Delta\zeta$ (as is discussed in Section 3.5).[25] The primal variables are $(y_{\mathrm{Tu}}, y_{\mathrm{Pu}}, s_0) \in L^\infty \times L^\infty \times \mathbb{R}$, and the Lagrange function (of primal and dual variables) is

$$\mathcal{L}(y_{\mathrm{Tu}}, y_{\mathrm{Pu}}, s_0; \kappa, \nu, \lambda) = \begin{cases} \Pi_{\mathrm{Exc}}(y_{\mathrm{Tu}}, y_{\mathrm{Pu}}, s_0; \kappa, \nu, \lambda) + V(\kappa) & \text{if } (\kappa, \nu) \geq 0 \\ +\infty & \text{if } (\kappa, \nu) \ngeq 0 \end{cases} \tag{3.12.3}$$

where

$$V := \langle \kappa_{\mathrm{St}} \mid k_{\mathrm{St}} \rangle_{\mathcal{M},\mathcal{C}} + \langle \kappa_{\mathrm{Tu}} \mid k_{\mathrm{Tu}} \rangle_{L^{\infty *}, L^\infty} + \langle \kappa_{\mathrm{Pu}} \mid k_{\mathrm{Pu}} \rangle_{L^{\infty *}, L^\infty} \tag{3.12.4}$$

and, with the notation (3.5.8) and with $f := y_{\mathrm{Tu}}/\eta_{\mathrm{Tu}} - \eta_{\mathrm{Pu}} y_{\mathrm{Pu}}$ as per (3.2.1)–(3.2.2),

$$\begin{aligned} \Pi_{\mathrm{Exc}} :=& \langle p \mid y_{\mathrm{Tu}} - y_{\mathrm{Pu}} \rangle - \langle \kappa_{\mathrm{St}} - \nu_{\mathrm{St}} \mid s_0 - I_0 f \rangle - \langle \kappa_{\mathrm{Tu}} - \nu_{\mathrm{Tu}} \mid y_{\mathrm{Tu}} \rangle \\ &- \eta_{\mathrm{Pu}} \langle \kappa_{\mathrm{Pu}} - \nu_{\mathrm{Pu}} \mid y_{\mathrm{Pu}} \rangle - \lambda \langle 1 \mid f \rangle \\ =& \langle p \mid y_{\mathrm{Tu}} - y_{\mathrm{Pu}} \rangle + \langle I_0^* \mu_{\mathrm{St}} \mid f \rangle - \langle \mu_{\mathrm{St}} \mid s_0 \rangle - \langle \mu_{\mathrm{Tu}} \mid y_{\mathrm{Tu}} \rangle \\ &- \eta_{\mathrm{Pu}} \langle \mu_{\mathrm{Pu}} \mid y_{\mathrm{Pu}} \rangle - \lambda \langle 1 \mid f \rangle \\ =& \langle p \mid y_{\mathrm{Tu}} - y_{\mathrm{Pu}} \rangle - \langle \lambda - \mu_{\mathrm{St}}(\cdot, T] \mid f \rangle - s_0 \mu_{\mathrm{St}}[0, T] \\ &- \langle \mu_{\mathrm{Tu}} \mid y_{\mathrm{Tu}} \rangle - \eta_{\mathrm{Pu}} \langle \mu_{\mathrm{Pu}} \mid y_{\mathrm{Pu}} \rangle \end{aligned} \tag{3.12.5}$$

since $I_0^* \mu_{\mathrm{St}} = \mu_{\mathrm{St}}(\cdot, T]$ by Lemma 3.12.1.

To calculate the dual minimand when $(\kappa, \nu) \geq 0$ (which is a dual constraint, since the minimand is $+\infty$ otherwise), note that

$$\sup_{(y_{\mathrm{Tu}}, y_{\mathrm{Pu}}, s_0)} \mathcal{L} = V + \sup_{(y_{\mathrm{Tu}}, y_{\mathrm{Pu}}, s_0)} \Pi_{\mathrm{Exc}},$$

since V is independent of $(y_{\mathrm{Tu}}, y_{\mathrm{Pu}}, s_0)$. By (3.12.5), Π_{Exc} is linear in these variables, so its supremum is either 0 or $+\infty$; and it is zero if and only if $\partial \Pi_{\mathrm{Exc}}/\partial s_0 = 0$ and $\nabla_{y_\phi} \Pi_{\mathrm{Exc}} = 0$ for $\phi = \mathrm{Tu}, \mathrm{Pu}$. These conditions are equivalent to (3.5.5)–(3.5.6). So the dual programme is: given (p, k), minimise the $V(\kappa, k)$ of (3.12.4) over $(\kappa, \nu) \geq 0$ and λ, subject to (3.5.5)–(3.5.6). □

[25] These parameter increments are what Rockafellar [73] calls "parameters". This is because, unlike [73], here the origin of the parameter vector space is not placed at the original parameter point, which is $(k_{\mathrm{St}}, 0; k_{\mathrm{Tu}}, 0; k_{\mathrm{Pu}}, 0; 0)$. This helps keep track of the the dual programme's dependence on the primal parameter point.

Comment: In (3.12.4)–(3.12.5), V is the value of the available resources k, priced at κ. And Π_{Exc} is, for an entrepreneur buying all the inputs, the excess profit (a.k.a. pure profit) from a storage policy $(y_{\mathrm{Tu}}, y_{\mathrm{Pu}})$ and the use of an initial stock s_0: the sum (3.12.5) defines Π_{Exc} as the total over the cycle of the revenue minus the cost of the resources needed at any time t. The resources in question are: the time-varying minimum requirements for the three capacities (priced at κ), for the three floors (for stock and for turbine and pump outputs, priced at ν), and for the required top-up (priced at λ). To see this, recall that $s = s_0 - I_0 f$ is the stock trajectory (since f is the flow from the reservoir).

Proof of Remark 3.5.2. If this were false, then the minimand's value could be decreased by replacing $(\kappa_{\mathrm{St}}, \nu_{\mathrm{St}}; \kappa_{\mathrm{Tu}}, \nu_{\mathrm{Tu}}; \kappa_{\mathrm{Pu}}, \nu_{\mathrm{Pu}})$ with $(\mu^+_{\mathrm{St}}, \mu^-_{\mathrm{St}}; \mu^+_{\mathrm{Tu}}, \mu^-_{\mathrm{Tu}}; \mu^+_{\mathrm{Pu}}, \mu^-_{\mathrm{Pu}})$ defined by (3.5.8). □

Proof of Proposition 3.6.1 (Dual solubility and optimality conditions). Like that of Theorem 3.5.1, this proof is put in CP terms. Since the nonnegative cones in the (primal) parameter spaces ($\mathcal{C}_+$ and L^∞_+) have nonempty interiors (for the supremum norm), the framework of [73, Examples 4, 4', 4"] is applicable. To verify the Generalised Slater's Condition of [73, (8.12)] for the primal constraints (3.4.7)–(3.4.10), it suffices to take $y_{\mathrm{Tu}} = \epsilon = \eta_{\mathrm{Ro}} y_{\mathrm{Pu}}$ (so that $f = 0$) with a sufficiently small constant $\epsilon > 0$, setting s_0 at any value strictly between 0 and k_{St}. So the dual has a solution, and the primal and dual values are equal (and finite): see, e.g., [73, Theorems 18 (a) and 17 (a)]. Furthermore, ν_{Tu} and κ_{Pu} belong to L^1 by (3.5.11). This proves Part 1.

For Part 2, apply the Kuhn-Tucker saddle-point characterisation of optima—given in, e.g., [73, Theorem 15 (e) and (f)]—to the primal (3.4.5)–(3.4.10) and its dual (3.5.1)–(3.5.6). This shows that $(y_{\mathrm{Tu}}, y_{\mathrm{Pu}}, s_0)$ and (κ, ν, λ) form a dual pair of solutions if and only if they maximise and minimise (respectively) the Lagrange function $\mathcal{L}$ given by (3.12.3). The minimum in question is characterised by: nonnegativity (3.5.4), primal feasibility (3.4.7)–(3.4.10) and complementary slackness, which translates here into Conditions 2b–2c. As for the maximum in question, it is characterised by the conditions $\partial \Pi_{\mathrm{Exc}} / \partial s_0 = 0$ and $\nabla_{y_\phi} \Pi_{\mathrm{Exc}} = 0$ for $\phi = \mathrm{Tu}, \mathrm{Pu}$, i.e., by (3.5.5)–(3.5.6). □

Comment: Existence of a dual optimum in the *norm*-dual spaces (κ_{St} and ν_{St} in $\mathcal{M} = \mathcal{C}^*$, and κ_{Tu}, ν_{Tu}, κ_{Pu} and ν_{Pu} in $L^{\infty *}$) comes automatically from (3.4.13), which ensures that the Generalised Slater's Condition of [73, (8.12)] holds with the norm topologies on the primal parameter spaces L^∞ and $\mathcal{C}$. The density representation (of the dual variables other than κ_{St} and ν_{St}) comes from the problem's structure and the assumptions on p: with $p \geq 0$, every optimal κ_{Pu} and ν_{Tu} is in L^1; and if $p \in L^1$ then every optimal κ_{Tu} and ν_{Pu} is also in L^1.

Proof of Proposition 3.7.1 (Stock pricing as the dual). This is a reformulation of Theorem 3.5.1: substitute the ψ given by (3.7.1) into (3.5.6), and note that, given ψ (and p), the best choices for κ_{St}, κ_{Tu} and κ_{Pu} are as in (3.7.4)–(3.7.5), because $k_{\mathrm{St}} > 0$, $k_{\mathrm{Tu}} > 0$ and $k_{\mathrm{Pu}} > 0$. This reduces the dual programme (3.5.1)–(3.5.6) to minimisation of

$$k_{\mathrm{St}} \int_{[0,T]} (\mathrm{d}\psi(t))^{+} + k_{\mathrm{Tu}} \left\langle \left(p - \frac{\psi}{\eta_{\mathrm{Tu}}}\right)^{+} \,\middle|\, 1 \right\rangle + k_{\mathrm{Pu}} \int_0^T \left(\psi(t) - \frac{p(t)}{\eta_{\mathrm{Pu}}}\right)^{+} \mathrm{d}t$$

over $\psi \in \mathrm{BV}\,[0-, T+]$, subject to $\psi(0-) = \psi(T+)$ lying between $\psi(0+)$ and $\psi(T-)$. Hence the first of the integrals equals the sum of $(\psi(0+) - \psi(T-))^{+}$ and $\int_{(0,T)} (\mathrm{d}\psi)^{+}$; and this sum is $\mathrm{Var}_{\mathrm{c}}^{+}(\psi)$. □

Proof of Proposition 3.8.1 (Primal solubility). With $p \in L^1$, the maximand of (3.4.6) is continuous for the weak* topology $\mathrm{w}(L^\infty, L^1)$. The feasible set is bounded: in $(y_{\mathrm{Tu}}, y_{\mathrm{Pu}})$ by (3.4.7)–(3.4.8), and in s_0 by (3.4.10) with, e.g., $t = 0$. So, being also weakly* closed, the feasible set is compact by the Banach-Alaoglu Theorem. And it is nonempty, since the point $(0,0,0)$ is feasible by assumption. So an optimum exists by Weierstrass's Extreme Value Theorem. □

Proof of Lemma 3.8.2 (Nonoptimality of conversion overlap). Assume (3.4.13), since otherwise the result holds trivially.[26] Take an $\epsilon > 0$ and a solution (κ, ν, λ) to the dual (3.5.1)–(3.5.6). By Part 2c of Proposition 3.6.1, $\nu_{\mathrm{Tu}} = 0 = \nu_{\mathrm{Pu}}$ on the set $S_\epsilon := \{t : (y_{\mathrm{Tu}} \wedge y_{\mathrm{Pu}})(t) \geq \epsilon\}$. So

$$0 \geq -(\eta_{\mathrm{Tu}}\kappa_{\mathrm{Tu}} + \kappa_{\mathrm{Pu}}) = \left(\frac{1}{\eta_{\mathrm{Pu}}} - \eta_{\mathrm{Tu}}\right) p \geq \left(\frac{1}{\eta_{\mathrm{Pu}}} - \eta_{\mathrm{Tu}}\right) p_{\mathrm{CA}} \quad \text{on } S_\epsilon$$

by (3.5.6) and because $p_{\mathrm{FA}} \geq 0$.[27] Since $p_{\mathrm{CA}} \gg 0$ and $\eta_{\mathrm{Ro}} < 1$, this implies that $\operatorname{meas} S_\epsilon = 0$. And this means that $y_{\mathrm{Tu}} \wedge y_{\mathrm{Pu}} = 0$ a.e. (since ϵ is arbitrary). □

At this stage, it is useful to introduce a notation for the sets of those times when the reservoir is empty or full or neither. The sets (which have already appeared in Condition 2b of Proposition 3.6.1) are:

$$E(f) := \{t \in [0,T] : s(t) = 0\} \tag{3.12.6}$$

$$F(f, k_{\mathrm{St}}) := \{t \in [0,T] : s(t) = k_{\mathrm{St}}\} \tag{3.12.7}$$

$$B(f, k_{\mathrm{St}}) := [0,T] \setminus (E \cup F) = \{t : 0 < s(t) < k_{\mathrm{St}}\} \tag{3.12.8}$$

where $f := y_{\mathrm{Tu}}/\eta_{\mathrm{Tu}} - \eta_{\mathrm{Pu}} y_{\mathrm{Pu}}$ for a storage policy $(y_{\mathrm{Tu}}, y_{\mathrm{Pu}})$ meeting the balance constraint $\int_0^T f(t)\,\mathrm{d}t = 0$, $s(t)$ is given by (3.4.11)–(3.4.12), and $k_{\mathrm{St}} \geq \mathrm{Max}(s)$. (When $\eta_{\mathrm{Ro}} = 1$, $f = y = y_{\mathrm{Tu}} - y_{\mathrm{Pu}}$.) Since $s(0) = s(T)$, 0 and T are either both in B, or both in E, or both in F.[28] From (3.4.11), $E \neq \emptyset$. Unless there is spare reservoir capacity, $F \neq \emptyset$ also; and then all three sets are nonempty. Their connected components are subintervals of $[0,T]$; and, being open, B is the union of a countable (finite or denumerable) sequence of intervals. Those *not* containing 0 or T are denoted by

$$A_m = \left(\underline{t}_m, \overline{t}_m\right) \neq \emptyset$$

[26] If k_{Tu} or k_{Pu} is 0, then $y_{\mathrm{Tu}} = 0 = y_{\mathrm{Pu}}$ is the only feasible point. If $k_{\mathrm{St}} = 0$ then $f = 0$, i.e., $y_{\mathrm{Tu}} = \eta_{\mathrm{Ro}} y_{\mathrm{Pu}}$ throughout; so the unique optimum is $y_{\mathrm{Tu}} = 0 = y_{\mathrm{Pu}}$.

[27] For p' and p'' in $L^{\infty *}$, $p' \leq p''$ on A means, by definition, that $\langle p' \mid y 1_A \rangle \leq \langle p'' \mid y 1_A \rangle$ for each $y \in L_{+}^{\infty}$.

[28] These cases do not really differ if $p \in \mathcal{C}$ and $p(0) = p(T)$.

for $m = 1, \ldots, M \leq \infty$, where $0 \leq \underline{t}_m < \bar{t}_m \leq T$. If $\{0, T\} \subset B$, then B additionally contains two subintervals whose union is

$$A_0 = (\underline{t}_0, T] \cup [0, \bar{t}_0)$$

for some $0 < \bar{t}_0 < \underline{t}_0 < T$. When $0, T \notin B$, set for completeness $\underline{t}_0 = T$ and $\bar{t}_0 = 0$, so that $A_0 = \emptyset$ in this case. In either case, $B = \bigcup_{m \geq 0} A_m$.

All these sets may be thought of as subsets of the circle that results from "gluing" 0 and T into a single point $T0$. Then $(A_m)_{m \geq 0}$ are the *component arcs* of B (a.k.a. B-arcs); A_0 is that arc which contains $T0$ (if $T0 \in B$); and $\underline{t}_m$ and $\bar{t}_m$ are the beginning and the end of arc A_m (w.r.t. the "clockwise" orientation).

The formula for the output y, in terms of any $\psi \in \hat{\Psi}$, is proved next. On the set $\{t : \eta_{\mathrm{Tu}} p(t) \neq \psi(t) \neq p(t) / \eta_{\mathrm{Pu}}\}$, the optimal y equals unambiguously k_{Tu} or 0 or $-k_{\mathrm{Pu}}/\eta_{\mathrm{Pu}}$. Uniqueness of y on $\{t : \eta_{\mathrm{Tu}} p(t) = \psi(t)\} \cup \{t : \psi(t) = p(t) / \eta_{\mathrm{Pu}}\}$ comes from the no-plateau assumption (3.8.1) on p: this ensures that

$$\{t : \eta_{\mathrm{Tu}} p(t) = \psi(t)\} \cup \left\{t : \psi(t) = \frac{p(t)}{\eta_{\mathrm{Pu}}}\right\} \subseteq E \cup F$$

up to a null set. And at each $t \in E \cup F$ one has $f(t) = -\dot{s}(t) = 0$ (and hence $y(t) = 0$), since s has an extremum at t.

Remark 3.12.3. *If* s: $[0, T] \to [0, 1]$ *is absolutely continuous, then* $\dot{s} = 0$ *almost everywhere on the set* $E := \{t \in [0, T] : s(t) = 0\}$.

Comment: By using Lebesgue's Metric Density Theorem, one can also show that the derivative of a Lipschitz function vanishes a.e. on a constancy set—i.e., if s: $[0, T] \to \mathbb{R}$ is a Lipschitz function, then $\dot{s} = 0$ a.e. on the set $E := \{t \in [0, T] : s(t) = 0\}$. The result is nontrivial unless the open set $[0, T] \setminus E$ consists of a *finite* sequence of intervals.

Proof of Proposition 3.8.3 (PS output with plateau-less price). Take any $y \in \hat{Y}$ (not yet known to be unique) and any $\psi \in \hat{\Psi}$ (which may be nonunique, unless $p \in \mathcal{C}$ and $\eta_{\mathrm{Ro}} = 1$). The first and the third lines of (3.8.2) follow from Part 2c of Proposition 3.6.1 with (3.5.6) and (3.7.1), which also show that $f = 0$ a.e. on $\{t : \eta_{\mathrm{Tu}} p(t) < \psi(t) < p(t) / \eta_{\mathrm{Pu}}\}$. It remains to show that $f = 0$ a.e. on the set

$$S := \left\{t : p(t) = \frac{\psi(t)}{\eta_{\mathrm{Tu}}}\right\} \cup \{t : p = \eta_{\mathrm{Pu}} \psi\}.$$

For each m, one has $\psi = \text{const.}$ on $A_m(f, k_{\mathrm{St}})$ by Part 2b of Proposition 3.6.1. So $\operatorname{meas}(S \cap A_m) = 0$ by (3.8.1), and hence $\operatorname{meas}(S \cap B(f, k_{\mathrm{St}})) = 0$ by countable additivity. This means that S is, up to a null set, contained in the set $F(f, k_{\mathrm{St}}) \cup E(f)$, on which $f = -\dot{s} = 0$ a.e. and hence $y = 0$ a.e. (by Remark 3.12.3), Lemma 3.8.2 and (3.2.2)). This completes the proof of (3.8.2). It follows that $\hat{Y}$ is a singleton, even when $\hat{\Psi}$ is not. (Given any $\psi \in \hat{\Psi}$, any y' and y'' from $\hat{Y}$ satisfy (3.8.2) and are therefore equal.) $\square$

3.13 Proofs for Section 3.9

Proof of Corollary 3.9.1 (Dual calculation of SR profit). Formula (3.9.1) follows from Propositions 3.6.1 and 3.7.1. To derive it term-by-term, use the optimality conditions (complementary slackness and feasibility) to expand $\langle p \,|\, y \rangle$:

$$\begin{aligned}
\Pi &:= \int_0^T p(t)\, y(t)\, \mathrm{d}t = \int_0^T \left(p(t)\, y(t) - \psi(t)\, f(t)\right) \mathrm{d}t + \int_0^T \psi(t)\, f(t)\, \mathrm{d}t \\
&= \int_0^T \left(p(t)\left(y^+(t) - y^-(t)\right) - \psi\left(\frac{y^+(t)}{\eta_{\mathrm{Tu}}} - \eta_{\mathrm{Pu}} y^-(t)\right)\right) \mathrm{d}t + \int_0^T \psi(t)\, f(t)\, \mathrm{d}t \\
&= \int_0^T \left(p(t) - \frac{\psi(t)}{\eta_{\mathrm{Tu}}}\right) y^+(t)\, \mathrm{d}t + \int_0^T \left(\eta_{\mathrm{Pu}} \psi(t) - p(t)\right) y^-(t)\, \mathrm{d}t + \int_0^T \psi(t)\, f(t)\, \mathrm{d}t \\
&= k_{\mathrm{Tu}} \int_0^T \left(p(t) - \frac{\psi(t)}{\eta_{\mathrm{Tu}}}\right)^+ \mathrm{d}t + k_{\mathrm{Pu}} \int_0^T \left(\frac{p(t)}{\eta_{\mathrm{Pu}}} - \psi(t)\right)^- \mathrm{d}t - \int_0^T \psi(t)\, \frac{\mathrm{d}s(t)}{\mathrm{d}t}\, \mathrm{d}t
\end{aligned}$$

integrating the last term by parts to obtain

$$\begin{aligned}
-\int_0^T \psi(t)\, \mathrm{d}s(t) &= -\left[\psi s\right]_{t=0-}^{t=T+} + \int_{[0,T]} s(t)\, \mathrm{d}\psi(t) = s(0)\left(\psi(0-) - \psi(T+)\right) + k_{\mathrm{St}} \int_{[0,T]} \left(\mathrm{d}\psi(t)\right)^+ \\
&= 0 + k_{\mathrm{St}} \operatorname{Var}_{\mathrm{c}}^+(\psi)
\end{aligned}$$

as required. □

Except for the shadow-price determinacy results (Lemmas 3.13.1 and 3.13.3 below), the derivation of Theorem 3.9.2 is a routine use of the marginal interpretation of the dual solution. Before a formal proof, it is worth retracing in the present context the familiar argument which establishes the derivative property of the value function when differentiability is taken for granted. With the dual minimand (3.7.8) denoted by $V(k, \psi)$, the r.h.s.'s of (3.9.3)–(3.9.5) are obviously the partial derivatives of V in k, evaluated at the dual optimum $\hat{\psi}(k)$. And the total derivatives, in k, of the dual value $V\left(k, \hat{\psi}(k)\right)$ are equal to the corresponding partial derivatives, since the partial derivative of V in ψ vanishes by the FOC for optimality of $\hat{\psi}$. To complete the calculation, note that the dual value equals the primal value $\Pi_{\mathrm{SR}}^{\mathrm{PS}}$ (if $k \gg 0$).[29] This is, indeed, the substance of the first step in the Proof of Theorem 3.9.2, except that a standard convex duality result is used instead of the above derivation "from first principles". This is necessary because a rigorous application of the chain rule would run into difficulties, since it would require the differentiability of $\hat{\psi}$ in k, and of V in ψ. This would make their composition $\Pi(k) = V\left(k, \hat{\psi}(k)\right)$ differentiable, but neither this nor even the uniqueness of an optimal ψ (i.e., the existence of $\hat{\psi}$) may be presupposed. Differentiability of Π must be proved—by using price continuity, since it is known to fail in general if $p \notin C$ (Example 3.15.1). Filling this gap requires lemmas on uniqueness of an optimal ψ. The cases of perfect and imperfect

[29]Conversely, the equality of short-run profit to the capacity value can be rederived from (3.9.3)–(3.9.5) by applying Euler's Theorem to Π as a homogeneous function of k.

conversion are separated, since they differ in the properties of ψ and y, and therefore require different arguments. It is only with $\eta_{\mathrm{Ro}} = 1$ that ψ is unique at all times.

Before a detailed proof that ψ is unique when $\eta_{\mathrm{Ro}} = 1$ (and $p \in C$), it is worth presenting the main ideas. The key principle is that a rent can be earned only at a time of full capacity utilisation. In the present context this means that $p - \psi$ can be nonzero only when the converter (taken to be symmetrically reversible) is working at full power (i.e., when $y(t) = \pm k_{\mathrm{Co}}$); and therefore $\psi(t)$ equals $p(t)$ whenever the reservoir is either full or empty (since $s(t) = 0$ or $s(t) = k_{\mathrm{St}}$ implies that $y(t) = -\dot{s}(t) = 0 \neq \pm k_{\mathrm{Co}}$). By the same principle, ψ can be rising or falling only when the reservoir is full or empty (respectively); so ψ stays constant on each open interval $(\underline{t}, \overline{t})$ during which the reservoir constraints are inactive (i.e., $0 < s < k_{\mathrm{St}}$). Together, these conditions determine the function ψ *almost* completely—except for the possibility of jumps or drops of ψ that may occur at the endpoints of a (closed) interval on which the reservoir is either full throughout or empty throughout.[30] Suppose, for example, that $\underline{t}$ is the end of an interval on which the reservoir is full. At that instant, ψ can jump but not drop; and the same is true of $p - \psi$ (since $p = \psi$ just before $\underline{t}$, and $p \geq \psi$ just after $\underline{t}$). So neither term, ψ or $p - \psi$, can jump at $\underline{t}$ if their sum (p) is continuous. This determines the constant value of ψ on $(\underline{t}, \overline{t})$ as $p(\underline{t})$; so ψ is unique.

Lemma 3.13.1 (Shadow-price uniqueness with perfect conversion). *Assume that $\eta_{\mathrm{Ro}} = 1$, $p \in C_{++}[0,T]$ and $k = (k_{\mathrm{St}}, k_{\mathrm{Tu}}, k_{\mathrm{Pu}}) \gg 0$. Then the dual (3.7.7)–(3.7.9) has a unique solution $\hat{\psi}$, which belongs to $C_{++}[0,T]$. If additionally $p(0) = p(T)$, then also $\hat{\psi}(0) = \hat{\psi}(T)$.*

Proof. Fix any primal solution $y \in \hat{Y}(p,k)$, which exists by Proposition 3.8.1 (though it may be nonunique). To show that there is just one dual solution, every dual solution $\psi \in \hat{\Psi}(p,k)$ will be expressed by the same formula in terms of y.[31]

In the case of $F(y, k_{\mathrm{St}}) \neq \emptyset$, which is dealt with first, the Kuhn-Tucker Conditions will be used to show that any $\psi \in \hat{\Psi}$ can be given, in terms of y, as

$$\psi(t) = p(t) \quad \text{for every } t \in (E \cup F)(y, k_{\mathrm{St}}) \setminus \{0, T\} \tag{3.13.1}$$

whereas on the m-th component A_m of $B(y, k_{\mathrm{St}})$, whose endpoints are $\underline{t}_m$ and $\overline{t}_m$, it is the constant

$$\psi(t) = \begin{cases} p(\underline{t}_m) & \text{if } \underline{t}_m \neq 0 \\ p(\overline{t}_m) & \text{if } \overline{t}_m \neq T \end{cases} \qquad \text{for every } t \in A_m(y, k_{\mathrm{St}}) \tag{3.13.2}$$

[30] To simplify, it is assumed here that the times when the reservoir is full form a set F that consists of a finite number of intervals (which may be single instants, as in Example 3.15.1). Although F can be more complex, this is only a technicality (dealt with in the Proof of Lemma 3.13.1).

[31] The basis for this strategy (used also in proving Proposition 3.8.3) is that every dual solution supports every primal solution; i.e., the set of Kuhn-Tucker (saddle) points for a dual pair of convex programmes is the Cartesian product (of the primal and dual solution sets): see Proposition 3.6.1.

for each $m \geq 0$. Since both E and F are nonempty, $A_m \neq (0,T)$, so at least one line of (3.13.2) applies; and when both do, they are consistent. So (3.13.1)–(3.13.2) fully determine ψ on $(0,T)$, and hence on $[0,T]$ because $\psi(0)$ and $\psi(T)$ are defined by continuity.

To use the optimality conditions as stated in Proposition 3.6.1—i.e., in terms of (κ,ν,λ) rather than ψ—recall from Section 3.7 that if a $\psi \in \mathrm{BV}(0,T)$ solves (3.7.7)–(3.7.9), then (3.5.1)–(3.5.6) is solved by: $(\kappa_{\mathrm{Tu}},\kappa_{\mathrm{Pu}}) = \left((p-\psi)^+,(p-\psi)^-\right)$, any λ between $\psi(0+)$ and $\psi(T-)$, and $(\kappa_{\mathrm{St}},\nu_{\mathrm{St}}) = (\mu_{\mathrm{St}}^+,\mu_{\mathrm{St}}^-)$, where $\mu_{\mathrm{St}} = \mathrm{d}\psi$ on $(0,T)$ with $\mu\{0\} = \psi(0+) - \lambda$ and $\mu\{T\} = \lambda - \psi(T-)$.

With $\eta_{\mathrm{Pu}} = \eta_{\mathrm{Tu}} = 1$, (3.5.6) gives

$$p = \lambda + (\kappa_{\mathrm{St}} - \nu_{\mathrm{St}})[0,\cdot] + (\kappa_{\mathrm{Tu}} - \kappa_{\mathrm{Pu}}) = \psi + (\kappa_{\mathrm{Tu}} - \kappa_{\mathrm{Pu}}) \quad \text{a.e.} \tag{3.13.3}$$

It suffices to show that, at every point of $(E \cup F) \setminus \{0,T\}$, ψ is continuous and equal to p: then (3.13.2) follows, since ψ is constant on each B-component A_m, and since $A_m \neq (0,T)$.

A discontinuity of ψ could only be a jump at a time when the reservoir is full, or a drop when it is empty. If $t \in F$ say, then, being full at t, the reservoir cannot be being discharged just before t or charged just after t.[32] *A fortiori*, the capacity charge κ_{Tu} must be zero just before t, and κ_{Pu} must be zero just after t. So $p - \psi = \kappa_{\mathrm{Tu}} - \kappa_{\mathrm{Pu}}$ is nonpositive just before t and nonnegative just after t, and hence $p - \psi$ cannot drop at a $t \in F$. This means that any discontinuous changes in ψ and $p - \psi$ are of the same sign and cannot cancel each other out. So ψ (and $p - \psi$) must be continuous if p is. And it follows (from the signs of $p - \psi$ before and after t) that $p(t) = \psi(t)$. The "upside down" version of this reasoning applies to $t \in E$.

Since κ_{Tu} and κ_{Pu} are classes of a.e. equal functions (rather than functions), this argument is formalised by using the essential limit concept—for which see, e.g., [20, IV.36–IV.37] or [81, II.9: p. 90]. It is also convenient to say that an inequality between functions (of t) holds *somewhere* on $A \subseteq [0,T]$ to mean that it holds on an $A' \subseteq A$ with $\operatorname{meas} A' > 0$ (i.e., it is *not* the case that the reverse inequality holds a.e. on A).

The storage policy $y_{\mathrm{Tu}} := y^+$ and $y_{\mathrm{Pu}} := y^-$, with the $\underline{s}_0$ of (3.4.11), solves (3.4.5)–(3.4.10).[33] Consider first a $t \in F \setminus \{0,T\}$. For every $\Delta t > 0$, it cannot be that $y > 0$ a.e. on $(t - \Delta t, t)$; i.e., somewhere on $(t - \Delta t, t)$ one has $y \leq 0$ so $y_{\mathrm{Tu}} = y^+ = 0 < k_{\mathrm{Tu}}$. Therefore $\kappa_{\mathrm{Tu}} = 0$ somewhere on $(t - \Delta t, t)$, by Part 2c of Proposition 3.6.1; and, as $\Delta t \to 0$, this shows that the lower left essential limit of κ_{Tu} at t is zero. Similarly, somewhere on $(t, t + \Delta t)$ one has $y \geq 0$ so $y_{\mathrm{Pu}} = y^- = 0 < k_{\mathrm{Pu}}$. Therefore $\kappa_{\mathrm{Pu}} = 0$ somewhere on $(t, t + \Delta t)$. This means that the lower right essential limit of κ_{Pu} at t is zero; i.e.,

$$\operatorname*{ess\,lim\,inf}_{\tau \searrow t} \kappa_{\mathrm{Pu}}(\tau) = 0 = \operatorname*{ess\,lim\,inf}_{\tau \nearrow t} \kappa_{\mathrm{Tu}}(\tau) \quad \text{for } t \in F \setminus \{0,T\}. \tag{3.13.4}$$

[32] This, by the way, is where the constancy of k_{St} over time is used.

[33] To show this formally, it suffices to take any optimal $(y_{\mathrm{Tu}}, y_{\mathrm{Pu}})$ with $y_{\mathrm{Tu}} - y_{\mathrm{Pu}} = y$ and note that (y^+, y^-) is also feasible (since $y^+ \leq y_{\mathrm{Tu}}$, $y^- \leq y_{\mathrm{Pu}}$ and, with $\eta_{\mathrm{Ro}} = 1$, the last term in (3.2.2) vanishes even if y_{Tu} and y_{Pu} do overlap).

Given (3.13.3) as well as continuity of p and nonnegativity of κ_{Pu} and κ_{Tu}, it follows from (3.13.4) that[34]

$$\begin{aligned} p(t) - \psi(t-) &= \operatorname{ess}\lim_{\tau\nearrow t} (\kappa_{\mathrm{Tu}} - \kappa_{\mathrm{Pu}})(\tau) \\ &= \operatorname{ess}\liminf_{\tau\nearrow t} \kappa_{\mathrm{Tu}}(\tau) - \operatorname{ess}\liminf_{\tau\nearrow t} \kappa_{\mathrm{Pu}}(\tau) \le 0 &\quad (3.13.5) \\ &\le \operatorname{ess}\liminf_{\tau\searrow t} \kappa_{\mathrm{Tu}}(\tau) - \operatorname{ess}\liminf_{\tau\searrow t} \kappa_{\mathrm{Pu}}(\tau) = \operatorname{ess}\lim_{\tau\searrow t} (\kappa_{\mathrm{Tu}} - \kappa_{\mathrm{Pu}})(\tau) &\quad (3.13.6) \\ &= p(t) - \psi(t+). \end{aligned}$$

Therefore $\psi(t-) \ge \psi(t+)$ from a comparison of the first and the last sums. But also, since $t \in F$,

$$\psi(t-) \le \psi(t+) \tag{3.13.7}$$

by Part 2b of Proposition 3.6.1; so all three inequalities of (3.13.5), (3.13.6) and (3.13.7) must actually hold as equalities. This shows that $\psi(t-) = \psi(t+) = p(t)$, i.e., the two-sided limit of ψ at t exists and equals $p(t)$. (Since it exists, it also equals $\psi(t)$ because $\psi(t)$ always lies between $\psi(t-)$ and $\psi(t+)$.) The same can be shown for $t \in E$ (by an "upside down" version of the preceding proof for $t \in F$); so

$$\psi(t) = \lim_{\tau\to t} \psi(\tau) = p(t) \quad \text{for } t \in (E \cup F) \setminus \{0, T\} \ne \emptyset. \tag{3.13.8}$$

Nonemptiness of this set follows from the assumption that $F \ne \emptyset$, since $E \ne \emptyset$ always, by (3.4.11).

By Part 2b of Proposition 3.6.1, ψ is constant on each A_m. This and (3.13.8) show that $\psi \in \mathcal{C}(0,T)$. (Equivalently $\psi \in \mathcal{C}[0,T]$, since $\psi(0) := \psi(0+)$ and $\psi(T) := \psi(T-)$.)

It remains to check that the proven properties of ψ imply (3.13.2). Since $E \cup F \not\subseteq \{0, T\}$, the set B consists of two or more nonempty components A_m. Each of these has at least one endpoint that is neither 0 nor T; i.e., $\underline{t}_m \ne 0$ or $\bar{t}_m \ne T$ ($\underline{t}_m \ne T$ and $\bar{t}_m \ne 0$ always). Say it is $\underline{t}_m$; then $\underline{t}_m \in (E \cup F) \setminus \{0, T\}$, since $\underline{t}_m \notin A_m$ (A_m is an open arc). So, by (3.13.8) and the constancy of ψ on A_m,

$$p(\underline{t}_m) = \psi(\underline{t}_m) = \psi(t) \quad \text{for every } t \in A_m. \tag{3.13.9}$$

If $T \ne \bar{t}_m$, then (3.13.9) holds with $\bar{t}_m$ in place of $\underline{t}_m$, by the same argument. This also shows that $p(\underline{t}_m) = p(\bar{t}_m)$ if both $\underline{t}_m \ne 0$ and $\bar{t}_m \ne T$. (All this applies to $m = 0$ as well, if $A_0 \ne \emptyset$. In this case ψ is additionally constant on $A_0 \supset \{0, T\}$; so $\psi(0) = \psi(T)$ even if $p(0) \ne p(T)$.) This completes the proof of (3.13.1)–(3.13.2) when $F \ne \emptyset$.

If $p(0) = p(T)$, then $\psi(0) = \psi(T)$ follows by virtually the same argument as that proving (3.13.8), with 0 and T thought of as a single point of the circle.

[34] This argument uses also the fact that $\liminf(A - B) \le \liminf A - \liminf B \le \limsup(A - B)$ whenever the middle term is well defined. It equals $\lim(A - B)$ if the latter exists, as here (although the inequalities suffice). The same holds with $\limsup A - \limsup B$ as the middle term.

Finally, consider the case of $F(y, k_{\mathrm{St}}) = \emptyset$, which is trivial in that the reservoir is never used to capacity, and it earns no rent. Formally, $\kappa_{\mathrm{St}} = \nu_{\mathrm{St}} = 0$ by Part 2b of Proposition 3.6.1 and (3.5.5); so ψ is a constant. Its uniqueness is readily shown: ψ minimises (3.7.8) over $\mathrm{BV}(0, T)$, so, *a fortiori*, it minimises (3.7.8) over $\mathbb{R}$. Since for $\psi \in \mathbb{R}$ the sum (3.7.8) simplifies to

$$k_{\mathrm{Tu}} \int_0^T (p(t) - \psi(t))^+ \, \mathrm{d}t + k_{\mathrm{Pu}} \int_0^T (\psi(t) - p(t))^+ \, \mathrm{d}t,$$

the minimum in question is characterised by the FOC

$$\operatorname{meas}\{t : p(t) < \psi\} \le T \frac{k_{\mathrm{Tu}}}{k_{\mathrm{Tu}} + k_{\mathrm{Pu}}} \le \operatorname{meas}\{t : p(t) \le \psi\} \tag{3.13.10}$$

which means that ψ is a lower quantile of order $k_{\mathrm{Tu}} / (k_{\mathrm{Tu}} + k_{\mathrm{Pu}})$ for the distribution of p w.r.t. meas/T. And the quantile is unique if $p \in \mathcal{C}[0, T]$, since the cumulative distribution function of p is then strictly increasing on the interval $(\mathrm{Min}(p), \mathrm{Max}(p))$. □

Comment: Although (3.13.4) suffices for the argument, both inf signs can be deleted, i.e., (3.13.4) can be strengthened to: $\kappa_{\mathrm{Tu}}(t-) = 0 = \kappa_{\mathrm{Pu}}(t+)$ with $\kappa_{\mathrm{Pu}}(t-) \ge 0$ and $\kappa_{\mathrm{Tu}}(t+) \ge 0$, for $t \in F \setminus \{0, T\}$, whenever $p(t\pm)$ exist.[35] This is because, by (3.5.9) and the continuity of $\kappa \mapsto \kappa^{\pm} \in \mathbb{R}_+$, the four limits exist and are equal to $(\kappa_{\mathrm{Tu}} - \kappa_{\mathrm{Pu}})^{\pm}(t\pm) = (p - \psi)^{\pm}(t\pm)$. All four limits are zero if p is continuous at t.

The case of imperfect conversion is dealt with next. With $\eta_{\mathrm{Ro}} < 1$, the restriction of an optimal ψ to $E \cup F$ lies between $\eta_{\mathrm{Tu}} p$ and p/η_{Pu} (Figure 3.4a), instead of being equal to p as in (3.13.1). This obviously makes ψ both nonunique and in general discontinuous, but *not* at those instants which matter for capacity valuation—as is shown in Lemma 3.13.3 below. And this case is simpler in some ways. For example, piecewise monotonicity of ψ is easier to establish because, when $\eta_{\mathrm{Ro}} < 1$, each of the sets B, E and F has only a *finite* number of connected components. This in turn follows from a lower bound on the length of any component of E or F that does not contain 0 or T and is, therefore, a closed subinterval $R = [\underline{t}, \overline{t}]$ of $[0, T]$: once the reservoir becomes, say, full, it must stay full for as long as it takes the price $p(t)$ to rise sufficiently to reverse the profitable action from that of charging, until $\underline{t}$, to discharging, from $\overline{t}$ on. The price $p(\overline{t})$ must be higher than $p(\underline{t})$ by a factor of at least $1/\eta_{\mathrm{Ro}} > 1$; so $p(\overline{t})$ must exceed $p(\underline{t})$ by at least the fraction $1/\eta_{\mathrm{Ro}} - 1$ of the lowest price, $\mathrm{Min}(p)$. And such a price rise takes a certain minimum time to come about, since a price jump is excluded by assumption.[36] A similar argument applies to a price fall while the reservoir is empty.

[35] The abbreviations $\kappa(t\pm)$ for the essential (one-sided) limits should not be mistaken for the ordinary limits of a particular variant of κ, in as much as the ordinary limits may be nonexistent.

[36] If $p(0) = p(T)$, this also applies to any component arc of F or E that contains the point $T0$ of the circle formed from $[0, T]$.

Lemma 3.13.2 (Minimum arc length with imperfect conversion). *Assume that* $\eta_{\mathrm{Ro}} < 1$, $p \in C_{++}[0,T]$ *and* $k = (k_{\mathrm{St}}, k_{\mathrm{Tu}}, k_{\mathrm{Pu}}) \gg 0$. *If* $R = [\underline{t}, \overline{t}]$ *is a component arc of the set* $(F \cup E)(f, k_{\mathrm{St}})$*—where* $f := (y^+/\eta_{\mathrm{Tu}}) - \eta_{\mathrm{Pu}} y^-$ *and* $y \in \hat{Y}(p,k)$, *i.e.,* y *solves (3.4.1)–(3.4.3)—and if* $0, T \notin R$, *then*

$$\left|p\left(\overline{t}\right) - p\left(\underline{t}\right)\right| \geq \left(\frac{1}{\eta_{\mathrm{Ro}}} - 1\right) \min_{t \in [0,T]} p(t) > 0. \tag{3.13.11}$$

So there exists a positive number δ *(which depends only on* p *and* η_{Ro}*), such that*

$$\operatorname{meas} R \geq \delta > 0 \tag{3.13.12}$$

for every component R *of* E *or* F *(except possibly one). Therefore* E, F *and* B *have only a finite number of components.*

Proof. Fix a solution, $\psi \in \hat{\Psi}$, to the dual (3.7.7)–(3.7.9). To apply Proposition 3.6.1, introduce the corresponding (κ, ν, λ) as in the Proof of Lemma 3.13.1 (except that here κ_{Tu}, κ_{Pu}, ν_{Tu} and ν_{Pu} are given by (3.7.5)–(3.7.6) with $\eta_{\mathrm{Ro}} < 1$).

By Lemma 3.8.2 (and the remark preceding (3.4.11)), the storage policy $y_{\mathrm{Tu}} := y^+$ and $y_{\mathrm{Pu}} := y^-$, with the $\underline{s}_0$ of (3.4.11), solves (3.4.5)–(3.4.10). Consider a component $R = [\underline{t}, \overline{t}]$ of F with $0 < \underline{t} \leq \overline{t} < T$. For every $\Delta t > 0$, it cannot be that $f \leq 0$ a.e. on $(\overline{t}, \overline{t} + \Delta t)$. So somewhere[37] on $(\overline{t}, \overline{t} + \Delta t)$ one has $f > 0$, i.e., $y_{\mathrm{Tu}} = y^+ > 0$ and therefore $\nu_{\mathrm{Tu}} = 0$ and $\eta_{\mathrm{Tu}} p = \psi + \eta_{\mathrm{Tu}} \kappa_{\mathrm{Tu}} \geq \psi$, by Part 2c of Proposition 3.6.1, (3.5.6) and (3.7.1). In the limit as $\Delta t \searrow 0$, this gives

$$\eta_{\mathrm{Tu}} p\left(\overline{t}\right) \geq \psi\left(\overline{t}+\right). \tag{3.13.13}$$

Similarly, somewhere on $(\underline{t} - \Delta t, \underline{t})$ one has $f < 0$, i.e., $y_{\mathrm{Pu}} = y^- > 0$ and so $\nu_{\mathrm{Pu}} = 0$ and $p/\eta_{\mathrm{Pu}} = \psi - \kappa_{\mathrm{Pu}} \leq \psi$. In the limit this gives

$$\frac{p\left(\underline{t}\right)}{\eta_{\mathrm{Pu}}} \leq \psi\left(\underline{t}-\right). \tag{3.13.14}$$

With $\psi\left(\overline{t}+\right) \geq \psi\left(\underline{t}-\right)$ by Part 2b of Proposition 3.6.1, (3.13.13)–(3.13.14) give

$$p\left(\overline{t}\right) - p\left(\underline{t}\right) \geq \left(\frac{1}{\eta_{\mathrm{Tu}} \eta_{\mathrm{Pu}}} - 1\right) p\left(\underline{t}\right)$$

and hence (3.13.11). To deduce (3.13.12), note first that $\operatorname{Min}(p) > 0$, since $p \in C_{++}[0,T]$. So, by the uniform continuity of p on $[0,T]$ with the usual metric (of $\mathbb{R}$, not the circle), choose a $\delta > 0$ such that $|t' - t''| \geq \delta$ whenever $|p(t') - p(t'')| \geq (1/\eta_{\mathrm{Ro}} - 1) \operatorname{Min}(p)$. Then $\operatorname{meas} R = \overline{t} - \underline{t} \geq \delta$ by (3.13.11). A similar argument applies when R is a component of E. $\square$

Lemma 3.13.3 (On stock price determinacy with imperfect conversion). *Assume that* $\eta_{\mathrm{Ro}} < 1$, $p \in C_{++}[0,T]$ *and* $k = (k_{\mathrm{St}}, k_{\mathrm{Tu}}, k_{\mathrm{Pu}}) \gg 0$. *If* $y \in \hat{Y}(p,k)$ *and* $\psi \in \hat{\Psi}(p,k)$*—i.e.,* y *solves (3.4.1)–(3.4.3) and* ψ *solves (3.7.7)–(3.7.9)—and if* $F(f, k_{\mathrm{St}}) \neq \emptyset$, *where* $f = y^+/\eta_{\mathrm{Tu}} - \eta_{\mathrm{Pu}} y^-$, *then*

$$\eta_{\mathrm{Tu}} p(t) \leq \psi(t) \leq \frac{p(t)}{\eta_{\mathrm{Pu}}} \quad \text{for every } t \in (E \cup F)(f, k_{\mathrm{St}}) \setminus \{0, T\}, \tag{3.13.15}$$

[37] The precise meaning of "somewhere" is defined in the Proof of Lemma 3.13.1, before (3.13.4).

whereas for $t \in A_m(f, k_{St})$*—i.e., on the m-th component of* $B(f, k_{St})$*—*ψ *is the constant*

$$\psi(t) = \begin{cases} \eta_{Tu} p(\underline{t}_m) & \text{if } 0 \neq \underline{t}_m \in F(f, k_{St}) \\ p(\underline{t}_m)/\eta_{Pu} & \text{if } 0 \neq \underline{t}_m \in E(f) \\ p(\bar{t}_m)/\eta_{Pu} & \text{if } T \neq \bar{t}_m \in F(f, k_{St}) \\ \eta_{Tu} p(\bar{t}_m) & \text{if } T \neq \bar{t}_m \in E(f) \end{cases} \tag{3.13.16}$$

for $m = 1, \ldots, M < \infty$*. (At least one line of (3.13.16) applies; and when two do, they are consistent.)*

Furthermore, ψ *is continuous at all the endpoints of B-arcs (or, equivalently, endpoints of F-arcs and E-arcs). In general* $\psi(0)$*.and* $\psi(T)$ *may differ, but if additionally* $p(0) = p(T)$*, and* 0 *or* T *is an endpoint of a B-arc, then also* $\psi(0) = \psi(T)$*.*

Given the constancy of ψ *on the B-arcs, an equivalent form of (3.13.16) is that for every F-arc* $[\underline{t}, \bar{t}]$ *that does not contain* 0 *or* T

$$\psi(\underline{t}) = \frac{p(\underline{t})}{\eta_{Pu}} \quad \text{and} \quad \psi(\bar{t}) = \eta_{Tu} p(\bar{t}), \tag{3.13.17}$$

whilst for every E-arc $[\underline{t}, \bar{t}]$ *that does not contain* 0 *or* T

$$\psi(\underline{t}) = \eta_{Tu} p(\underline{t}) \quad \text{and} \quad \psi(\bar{t}) = \frac{p(\bar{t})}{\eta_{Pu}}. \tag{3.13.18}$$

In the case of an F-arc or E-arc $[\underline{t}, T] \cup [0, \bar{t}]$ *containing* $\{0, T\}$*, where* $0 \le \bar{t} < \underline{t} \le T$*, this applies also to* $\underline{t}$ *if* $\underline{t} \neq T$*, and to* $\bar{t}$ *if* $\bar{t} \neq 0$*.*[38]

Proof. As in the Proof of Lemma 3.13.2, take a $\psi \in \hat{\Psi}$, and introduce the corresponding (κ, ν, λ) to apply Proposition 3.6.1. Formulae (3.13.15) and (3.13.17)–(3.13.18) will be proved. (The equivalence of (3.13.17)–(3.13.18) to (3.13.16) follows from the fact that if the beginning of a B-arc is in F, then it is the end of an F-arc, etc.)

On any F-arc $R = [\underline{t}, \bar{t}] \subset (0, T)$ one has $s = \text{const.}$ so $f = 0$ a.e., and so (3.13.15) holds a.e. on R by (3.5.6) and Part 2c of Proposition 3.6.1. Actually (3.13.15) holds for every $t \in \operatorname{int} R = (\underline{t}, \bar{t})$.[39]

Since $\operatorname{int} R \neq \emptyset$ by (3.13.12), the inequalities (3.13.15) for $t \in \operatorname{int} R$ (together with Part 2b of Proposition 3.6.1) give, by passing to the limit as $t \to \underline{t}$ or as $t \to \bar{t}$, that

$$\eta_{Tu} p(\bar{t}) \le \psi(\bar{t}-) \le \psi(\bar{t}+) \tag{3.13.19}$$

$$\frac{p(\underline{t})}{\eta_{Pu}} \ge \psi(\underline{t}+) \ge \psi(\underline{t}-). \tag{3.13.20}$$

[38] These conditions fully characterise a dual solution ψ.

[39] This is because if ψ is a function of bounded variation, p' and p'' are continuous and $p' \le \psi \le p''$ a.e. on an open interval I, then $p' \le \psi \le p''$ everywhere on I. (This is so for every l.s.c. p', u.s.c. p'' and regulated ψ with $\psi(t)$ between $\psi(t-)$ and $\psi(t+)$ for each t.)

Given the reverse inequalities (3.13.13)–(3.13.14), all of (3.13.19)–(3.13.20) must actually hold as four equalities. This proves (3.13.17) and the continuity of ψ at both $\underline{t}$ and $\bar{t}$; so (3.13.15) holds also at the endpoints of R. See Figure 3.4a.

For an E-arc $R = [\underline{t}, \bar{t}]$, a similar argument establishes (3.13.15) and (3.13.18). □

Proof of Theorem 3.9.2 (Efficiency rents of a storage plant). The first, routine step is to identify the dual variables as marginal values of the primal parameters, with the marginal values formalised as supergradients (of the primal value, a concave function of the parameters): see, e.g., [73, Theorem 16: (b) and (a), with Theorem 15: (e) and (f)] or [51, 7.3: Theorem 1']. This is applied in such a way as to give the marginal interpretation to the optimal κ and ν themselves, rather than only to their totals over the cycle, although the formulae to be proved are for the total values. Therefore the short-run profit is considered as a function, $\widetilde{\Pi}^{\mathrm{PS}}_{\mathrm{SR}}$, of all the quantity parameters

$$(k_{\mathrm{St}}, n_{\mathrm{St}}, k_{\mathrm{Tu}}, n_{\mathrm{Tu}}, k_{\mathrm{Pu}}, n_{\mathrm{Pu}}, \zeta) \in \mathcal{C} \times \mathcal{C} \times L^{\infty} \times L^{\infty} \times L^{\infty} \times L^{\infty} \times \mathbb{R}$$

discussed in Section 3.5. It is an extension of the optimal value of the programme (3.4.5)–(3.4.10), i.e.,

$$\Pi^{\mathrm{PS}}_{\mathrm{SR}} (p; k_{\mathrm{St}}, k_{\mathrm{Tu}}, k_{\mathrm{Pu}}) = \widetilde{\Pi}^{\mathrm{PS}}_{\mathrm{SR}} (p; k_{\mathrm{St}}, 0; k_{\mathrm{Tu}}, 0; k_{\mathrm{Pu}}, 0; 0) \quad \text{for } (k_{\mathrm{St}}, k_{\mathrm{Tu}}, k_{\mathrm{Pu}}) \in \mathbb{R}^3,$$

where the scalars are identified with constant functions on $[0, T]$. In this setting, the result giving the marginal values of the primal parameters is[40]

$$\widehat{\partial}_{k_{\mathrm{St}}, n_{\mathrm{St}}, k_{\mathrm{Tu}}, n_{\mathrm{Tu}}, k_{\mathrm{Pu}}, n_{\mathrm{Pu}}, \zeta} \widetilde{\Pi}^{\mathrm{PS}}_{\mathrm{SR}} = \{(\kappa_{\mathrm{St}}, -\nu_{\mathrm{St}}, \kappa_{\mathrm{Tu}}, -\nu_{\mathrm{Tu}}, \kappa_{\mathrm{Pu}}, -\nu_{\mathrm{Pu}}, \lambda) : (\kappa, \nu, \lambda) \text{ meet Conditions 2a, 2b and 2c of Proposition 3.6.1}\} \quad (3.13.21)$$

where $\widehat{\partial}$ denotes the superdifferential (a.k.a. the supergradient set) for a concave function (so $\widehat{\partial}\Pi = -\partial(-\Pi)$, where ∂ is the subdifferential). For differentiation of $\Pi^{\mathrm{PS}}_{\mathrm{SR}}$, with respect to the *constant* capacities, it follows from (3.13.21) that

$$\begin{aligned} \widehat{\partial}_{k_{\mathrm{St}}, k_{\mathrm{Tu}}, k_{\mathrm{Pu}}} \Pi^{\mathrm{PS}}_{\mathrm{SR}} &= \left\{ \left(\int_{[0,T]} \kappa_{\mathrm{St}} (\mathrm{d}t), \int_0^T \kappa_{\mathrm{Tu}} (t)\, \mathrm{d}t, \int_0^T \kappa_{\mathrm{Pu}} (t)\, \mathrm{d}t \right) : \exists \nu\ \exists \lambda\ (\kappa, -\nu, \lambda) \in \widehat{\partial}_{k,n,\zeta} \widetilde{\Pi}^{\mathrm{PS}}_{\mathrm{SR}} \right\} \\ &= \left\{ \left(\mathrm{Var}^{+}_{\mathrm{c}} (\psi), \int_0^T \left(p(t) - \frac{\psi(t)}{\eta_{\mathrm{Tu}}} \right)^{+} \mathrm{d}t, \int_0^T \left(\psi(t) - \frac{p(t)}{\eta_{\mathrm{Pu}}} \right)^{+} \mathrm{d}t \right) : \psi \in \hat{\Psi} (p, k_{\mathrm{St}}, k_{\mathrm{Tu}}, k_{\mathrm{Pu}}) \right\} \end{aligned} \quad (3.13.22)$$

by using (3.7.5) and substituting $\kappa_{\mathrm{St}} = (\mathrm{d}\psi)^{+}$.

[40]The corresponding result for the marginal values of the *dual* parameters is that $\partial_p \Pi^{\mathrm{PS}}_{\mathrm{SR}} = \hat{Y}$, which is a case of Hotelling's Lemma.

It remains to use the preceding lemmas to show that the set in question is actually a singleton; i.e., that the triple in (3.13.22), which consists of the r.h.s.'s of (3.9.3)–(3.9.5), is the same for every $\psi \in \hat{\Psi}$. In the case of *perfect conversion* ($\eta_{\mathrm{Ro}} = 1$), this is obvious because the set $\hat{\Psi}$ itself is a singleton (Lemma 3.13.1).

Now consider the case of *imperfect conversion* ($\eta_{\mathrm{Ro}} < 1$). Fix any $y \in \hat{Y}$, a solution to (3.4.1)–(3.4.3), which exists by Proposition 3.8.1. The storage policy $y_{\mathrm{Tu}} := y^{+}$ and $y_{\mathrm{Pu}} := y^{-}$, with the $\underline{s}_0$ of (3.4.11), solves (3.4.5)–(3.4.10).[41] The flow from the reservoir is $f = y^{+}/\eta_{\mathrm{Tu}} - \eta_{\mathrm{Pu}} y^{-}$. Although the three values in (3.13.22) involve a nonunique $\psi \in \hat{\Psi}$, each value is unique because, as is shown below, it can be expressed in terms of only p, f and k_{St}.

By Lemma 3.13.2, the sets E, F and B have a finite number of component arcs; and by Part 2b of Proposition 3.6.1 and (3.5.5) with (3.7.1) and (3.7.2),

$$\mathrm{Var}_{\mathrm{c}}^{+}(\psi) = \mathrm{Var}^{+}(\psi) \vee \mathrm{Var}^{-}(\psi)$$

where

$$\mathrm{Var}^{+}(\psi) = \sum_{R} \{\mathrm{d}\psi(R) : R \text{ is an } F\text{-arc and } 0, T \notin R\} \tag{3.13.23}$$

$$\mathrm{Var}^{-}(\psi) = \sum_{R} \{\mathrm{d}\psi(R) : R \text{ is an } E\text{-arc and } 0, T \notin R\}. \tag{3.13.24}$$

To express $\mathrm{Var}_{\mathrm{c}}^{+}(\psi)$ in terms of p, f and k_{St}, use Lemma 3.13.3 (on the assumption that $F(f, k_{\mathrm{St}}) \neq \emptyset$) to substitute

$$\begin{aligned}
\mathrm{d}\psi(R) &= \eta_{\mathrm{Tu}} p(\bar{t}) - \frac{p(\underline{t})}{\eta_{\mathrm{Pu}}} \quad \text{for an } F\text{-arc } R = [\underline{t}, \bar{t}] \\
\mathrm{d}\psi(R) &= \frac{p(\bar{t})}{\eta_{\mathrm{Pu}}} - \eta_{\mathrm{Tu}} p(\underline{t}) \quad \text{for an } E\text{-arc } R = [\underline{t}, \bar{t}]
\end{aligned}$$

into (3.13.23)–(3.13.24).

The unit turbine and pump rents are

$$\int_0^T \left(p(t) - \frac{\psi(t)}{\eta_{\mathrm{Tu}}}\right)^{+} \mathrm{d}t = \sum_m \int_{A_m} \left(p(t) - \frac{\psi(t)}{\eta_{\mathrm{Tu}}}\right)^{+} \mathrm{d}t$$

$$\int_0^T \left(\psi(t) - \frac{p(t)}{\eta_{\mathrm{Pu}}}\right)^{+} \mathrm{d}t = \sum_m \int_{A_m} \left(\psi(t) - \frac{p(t)}{\eta_{\mathrm{Pu}}}\right)^{+} \mathrm{d}t$$

and these integrals can be expressed in terms of p, f and k_{St} by substituting for ψ the value given by (3.13.16), for each B-arc A_m. This completes the proof for the case of $F \neq \emptyset$, as in Figure 3.4a.

The case of $F = \emptyset$ is, again, trivial: ψ is a constant, and the reservoir's rent is zero. If $\mathrm{Min}(p)/\eta_{\mathrm{Pu}} \geq \eta_{\mathrm{Tu}} \mathrm{Max}(p)$, then the turbine and pump rents are also zero (with ψ nonunique if the inequality is strict). If $\mathrm{Min}(p) < \eta_{\mathrm{Ro}} \mathrm{Max}(p)$, then the minimum of (3.7.8) over $\psi \in \mathbb{R}$ is

[41] As in the Proof of Lemma 3.13.2, this follows from Lemma 3.8.2 and the remark preceding (3.4.11).

characterised by the following FOC, which is (3.13.10) generalised to $\eta_{\mathrm{Ro}} \leq 1$:

$$-\frac{k_{\mathrm{Tu}}}{\eta_{\mathrm{Tu}}} \operatorname{meas}\{t : \eta_{\mathrm{Tu}} p(t) \geq \psi\} + k_{\mathrm{Pu}} \operatorname{meas}\left\{t : \frac{p(t)}{\eta_{\mathrm{Pu}}} < \psi\right\} \leq 0 \tag{3.13.25}$$

$$\leq -\frac{k_{\mathrm{Tu}}}{\eta_{\mathrm{Tu}}} \operatorname{meas}\{t : \eta_{\mathrm{Tu}} p(t) > \psi\} + k_{\mathrm{Pu}} \operatorname{meas}\left\{t : \frac{p(t)}{\eta_{\mathrm{Pu}}} \leq \psi\right\}. \tag{3.13.26}$$

To show that a ψ meeting these conditions is unique, note that if such a ψ is replaced with a $\psi' > \psi$, then the sum in (3.13.25) cannot remain nonpositive because it increases by more than the previous difference between the two sums in (3.13.25)–(3.13.26). This is because the c.d.f. of p (w.r.t. meas) is strictly increasing (since $p \in \mathcal{C}$). In the case of a $\psi' < \psi$, the sum in (3.13.26) similarly decreases and ceases to be nonnegative. □

Comment: Some weaker results on the relationship of an optimal ψ to p are simpler to establish than (3.13.1)–(3.13.2) or (3.13.15)–(3.13.16), but such results are so weak as to be of little use by themselves. For example:

1. When the number of B-arcs is finite, the inclusion $\psi \in [\eta_{\mathrm{Tu}} p, p/\eta_{\mathrm{Pu}}]$ a.e. on $F \cup E$ can be shown by the first argument of the Proof of Lemma 3.13.3—viz., that $\hat{s} = \text{const.}$ and so $\hat{y} = 0$ a.e. on each F-arc or E-arc R, so $\psi \in [\eta_{\mathrm{Tu}} p, p/\eta_{\mathrm{Pu}}]$ a.e. on R (and everywhere on $\operatorname{int} R$ if p is continuous, in which case it follows that $\psi \in [\eta_{\mathrm{Tu}} p, p/\eta_{\mathrm{Pu}}]$ on $F \cup E$, except possibly at the endpoints of F-arcs and E-arcs, the number of which is finite). But capacity valuation requires also the values of ψ on the B-arcs, and this necessitates the additional arguments in the Proofs of Lemmas 3.13.1–3.13.3.

2. By using Remark 3.12.3, the inclusion $\psi \in [\eta_{\mathrm{Tu}} p, p/\eta_{\mathrm{Pu}}]$ a.e. on $F \cup E$ can be shown for every $p \in L^1_{++}$. But this may even be vacuous ($F \cup E$ may be a null set, unless both $p \in \mathcal{C}$ and $\eta_{\mathrm{Ro}} < 1$); and the stronger results (3.13.1) and (3.13.15) do depend on the continuity of p.

3.14 Miscellaneous remarks

A storage plant is profitable to operate if (and only if) the lowest-to-highest price ratio is less than η_{Ro} (the round-trip technical efficiency).

Remark 3.14.1 (Nonzero output from storage). *Given any* $p \in L^1_+[0,T]$ *and* $k = (k_{\mathrm{St}}, k_{\mathrm{Tu}}, k_{\mathrm{Pu}}) \gg 0$, *the condition* $\operatorname{EssInf}(p) < \eta_{\mathrm{Ro}} \operatorname{EssSup}(p)$ *is necessary and sufficient for* $0 \notin \hat{Y}(p,k)$. *(When* $\eta_{\mathrm{Ro}} = 1$, *this simplifies to:* $0 \notin \hat{Y}(p,k)$ *unless* p *is a constant.)*

With perfect efficiency ($\eta_{\mathrm{Ro}} = 1$), the storage plant's optimal output is invariant under monotone transformations of the price function p.

Remark 3.14.2 (Output invariance under monotone price transformation). *Assume that* $\eta_{\mathrm{Ro}} = 1$, $k = (k_{\mathrm{St}}, k_{\mathrm{Tu}}, k_{\mathrm{Pu}}) \gg 0$, *and* $p \in L^1[0,T]$. *If* ι *is a strictly increasing (real-valued) function on* $p[0,T]$ *such that* $\iota \circ p \in L^1$, *then*[42]

$$\hat{Y}(\iota \circ p, k) = \hat{Y}(p,k) \quad \text{and} \quad \hat{\Psi}(\iota \circ p, k) = \iota \circ \hat{\Psi}(p,k) \tag{3.14.1}$$

where $\iota \circ \Psi := \{\iota \circ \psi : \psi \in \Psi\}$.

A general continuous function of bounded variation may fail to be monotone on any interval: see, e.g. [27, 8.1: Exercise 1]. But this is not so with the optimal ψ.

Remark 3.14.3. *With* $k_{\mathrm{St}} > 0$, *every shadow price* $\psi \in \hat{\Psi}(p,k)$*—i.e., every solution to (3.7.7)–(3.7.9)—is always piecewise monotone (also when* p *is not and* $\eta_{\mathrm{Ro}} = 1$*).*

When the good's price p is of bounded variation (e.g., when it is piecewise monotone), the shadow pricing problem can be reformulated as minimisation of a weighted sum of two distances: the variation norm of ψ and the Kantorovich-Rubinshtein-Vassershtein (KRV) distance of ψ from p, with the time circle as the underlying metric space.

Remark 3.14.4 (Reformulation of the dual). *When* $p \in \mathrm{BV}(0,T)$, *the dual problem (3.3.2)—which is (3.7.7)–(3.7.9) with* $\eta_{\mathrm{Ro}} = 1$ *and* $k_{\mathrm{Pu}} = k_{\mathrm{Tu}} = k_{\mathrm{Co}}$ *assumed for simplicity—can be restated as:*

$$\text{Given } (p, k_{\mathrm{St}}, k_{\mathrm{Co}}) \tag{3.14.2}$$

$$\text{minimise } \frac{1}{2} k_{\mathrm{St}} \|\mu\|_{\mathrm{Var}} + k_{\mathrm{Co}} \left\| \mu - (\mathrm{d}p)^{\mathrm{N}} \right\|_{\mathrm{KRV}} \tag{3.14.3}$$

$$\text{over } \mu \in \mathcal{M}^{\mathrm{cN}} \tag{3.14.4}$$

where $\mathcal{M}^{\mathrm{cN}}$ *is the space of (signed) null measures on the circle. (It is a subspace of* $\mathcal{M}^{\mathrm{c}}$, *the space of all measures on the circle, which is the norm-dual of* C^{c}*.)*

Comment: The reformulation leads to an alternative proof of the dual's solubility. This is based directly on Weierstrass's Theorem and makes no reference to the primal (unlike the earlier Proof of Proposition 3.6.1, which relies on Slater's Condition for the primal). Note first that the range of μ in (3.14.4) can be restricted to the closed ball, centred at 0, of radius $\left\|(\mathrm{d}p)^{\mathrm{N}}\right\|_{\mathrm{Var}}$ in the variation norm. This is because if $\|\mu\|_{\mathrm{Var}} > \left\|(\mathrm{d}p)^{\mathrm{N}}\right\|_{\mathrm{Var}}$ (and $(k_{\mathrm{St}}, k_{\mathrm{Co}}) > 0$) then the minimand's value at μ is strictly larger than it is at $(\mathrm{d}p)^{\mathrm{N}}$. On the ball, as on any $\|\cdot\|_{\mathrm{Var}}$-bounded subset of $\mathcal{M}^{\mathrm{cN}}$, the KRV norm topology is equivalent to the weak* topology $\mathrm{w}(\mathcal{M}^{\mathrm{c}}, C^{\mathrm{c}})$: see, e.g., [53, VIII.4: Theorem 3]. The ball is weakly* compact. Furthermore, $\|\cdot\|_{\mathrm{Var}}$ is weakly* lower semicontinuous (l.s.c.), on the

[42] Since a $p \in L^1$ is defined only up to a null set, $p[0,T]$ means here the essential range of p, i.e., the smallest closed set whose inverse image under p has full Lebesgue measure. For $p \in C$, this is the usual range of p.

whole of $\mathcal{M}^c$; and the KRV norm is weakly* continuous on each $\|\cdot\|_{\mathrm{Var}}$-bounded set (by the aforementioned equivalence). Therefore the problem (3.14.2)–(3.14.4) is effectively one of minimising an l.s.c. function on a compact set (so an optimum point exists).

As is stated next, the reservoir and the (reversible) converter are Wicksell technical complements in product-value terms; this is always the case with constants returns to scale and just two inputs. Implications for comparative statics of investment are spelt out in [40].

Remark 3.14.5. *The $\Pi_{\mathrm{SR}}^{\mathrm{PS}}$ is a supermodular function of $k = (k_{\mathrm{St}}, k_{\mathrm{Co}})$, i.e., $\Pi(k') + \Pi(k'') \leq \Pi(k' \wedge k'') + \Pi(k' \vee k'')$ for each k' and k'' in $\mathbb{R}_+^2$. (This means that $\partial^2\Pi/\partial k_{\mathrm{St}}\partial k_{\mathrm{Co}} \geq 0$ whenever the mixed second partial derivative exists.)*

The assumption needed for $\nabla_k \Pi_{\mathrm{SR}}^{\mathrm{PS}}$ to exist is next weakened to: $p_{\mathrm{CA}} \in \mathcal{C}$. That is, the density part of p is required to be continuous on $[0, T]$, but p may also have a nonzero p.f.a. part (in the Hewitt-Yosida decomposition (3.4.4)). If $p_{\mathrm{FA}} > 0$, it can be interpreted as the "concentrated" part of turbine capacity charges (since $p_{\mathrm{FA}} = \kappa_{\mathrm{Tu}}^{\mathrm{FA}}$ at every dual optimum by (3.5.11)). Unless demand for the flow in question is interruptible, such a charge can arise in general equilibrium, and it has a tractable mathematical representation by a singular measure (such as a point measure) if the consumption and output rates are continuous over time: see [39, Example 3.1]. Out of equilibrium, the presence of a nonzero p_{FA} can result in nonexistence of an optimum y for the primal (3.4.1)–(3.4.3): see Case (b) in Part 4 below. Except for this, the preceding analysis extends *mutatis mutandis* to the case of a $p \in L_{++}^{\infty *}$, as is spelt out next.

Remark 3.14.6 (Concentrated charges). *For every $p \in L_+^{\infty *}$ with $p_{\mathrm{CA}} \gg 0$:*

1. *The dual problem of stock pricing is (3.7.8)–(3.7.9) with p_{CA} instead of p and with $k_{\mathrm{Tu}} \|p_{\mathrm{FA}}\|_\infty^*$ added to the minimand (3.7.8).[43] Since the extra term is a constant (i.e., is independent of ψ), its addition does not change the solution set, i.e., $\hat{\Psi}(p, k) = \hat{\Psi}(p_{\mathrm{CA}}, k)$.*

2. *Formulae (3.9.3)–(3.9.5), which give $\nabla_k \Pi_{\mathrm{SR}}^{\mathrm{PS}}$ in terms of p and $\hat{\psi}$, hold with p replaced by p_{CA} on their r.h.s.'s and with $\|p_{\mathrm{FA}}\|_\infty^* = \langle p_{\mathrm{FA}} \mid 1 \rangle$ added to the r.h.s. of (3.9.4).[44]*

3. *The Kuhn-Tucker Conditions 2a–2c of Proposition 3.6.1 imply the same but with p_{CA} in place of p and with $\kappa_{\mathrm{Tu}}^{\mathrm{CA}}$ in place of κ_{Tu}. (The converse is obviously false.) It follows that, for*

[43] This also points to cases of the primal value being strictly less—it is never greater—than the dual value. This is when $p_{\mathrm{FA}} > 0$, $p_{\mathrm{CA}} \in \mathrm{BV}$ and $k_{\mathrm{Tu}} > 0$ but $k_{\mathrm{St}} = 0$: the short-run profit is then 0, but the capacity value (as found from the dual) is $k_{\mathrm{Tu}} \|p_{\mathrm{FA}}\|_\infty^* > 0$, since a dual solution is any $\psi \in [\eta_{\mathrm{Tu}} p_{\mathrm{CA}}, p_{\mathrm{CA}}/\eta_{\mathrm{Pu}}]$. (Similarly, if $k_{\mathrm{Pu}} = 0$ instead of k_{St} being zero, then the short-run profit is again 0, but the capacity value is at least $k_{\mathrm{Tu}} \|p_{\mathrm{FA}}\|_\infty^*$.) When $k \gg 0$, the primal and the dual values are of course equal.

[44] To see this, note that, by Part 1 and the equality of the dual and primal values (when $k_{\mathrm{PS}} \gg 0$), $\Pi_{\mathrm{SR}}^{\mathrm{PS}}(p) - \Pi_{\mathrm{SR}}^{\mathrm{PS}}(p_{\mathrm{CA}}) = k_{\mathrm{Tu}} \|p_{\mathrm{FA}}\|_\infty^*$.

$k = (k_{\mathrm{St}}, k_{\mathrm{Tu}}, k_{\mathrm{Pu}}) \gg 0$,

$$\hat{Y}(p,k) \subseteq \hat{Y}(p_{\mathrm{CA}}, k)$$

i.e., if p supports y as a short-run profit maximum, then so does p_{CA} (or, put formally, if y solves (3.4.1)–(3.4.3), then it also solves (3.4.1)–(3.4.3) with p_{CA} in place of p). So the conclusions about any optimal output y, such as (3.8.2), hold also with p_{CA} in place of p. Such results are of course vacuous when, at p, there is no optimal y.

4. *The timing of p_{FA} matters for the existence of an optimal output. Consider the cases in which a $p_{\mathrm{FA}} > 0$ is concentrated on each neighbourhood of: either (a) a peak $\bar{t}$, or (b) a trough $\underline{t}$, of a piecewise monotone $p_{\mathrm{CA}} \in C^{\mathrm{c}}$. With $\eta_{\mathrm{Ro}} = 1$ for simplicity (and $k \gg 0$), one has $\hat{y}(p_{\mathrm{CA}}) = k_{\mathrm{Tu}}$ around $\bar{t}$, and $\hat{y}(p_{\mathrm{CA}}) = -k_{\mathrm{Pu}}$ around $\underline{t}$. At $p = p_{\mathrm{CA}} + p_{\mathrm{FA}}$, one has $\hat{y}(p) = \hat{y}(p_{\mathrm{CA}})$ in Case (a), but $\hat{Y}(p) = \emptyset$ in Case (b).*[45]

3.15 Indeterminacy of marginal values with discrete time

As is next shown by means of a two-period model (with $\beta = 1$ and $\eta_{\mathrm{Ro}} = 1$), discretisation of time can make $\Pi_{\mathrm{SR}}^{\mathrm{PS}}(p,k)$ nondifferentiable in k. This is because it forces p to be piecewise constant and thus discontinuous; and the optimal ψ's are nonunique if p has a jump paired with a drop at two instants which differ exactly by $k_{\mathrm{St}}/k_{\mathrm{Co}}$—which is *always* the case in the two-period model (unless there is spare capacity of one kind or the other). In the following example the cycle is divided into subperiods of lengths d and $T-d$. Then $\Pi(k)$ is proportional to $\min\{k_{\mathrm{St}}, \delta k_{\mathrm{Co}}\}$, where $\delta := \min\{d, T-d\}$. The only efficient capacity ratio is $k_{\mathrm{St}}/k_{\mathrm{Co}} = \delta$; and Π is obviously nondifferentiable at such a k.

The form which $\Pi(k)$ has in the two-period model may create the false impression that storage is a fixed-coefficients technology—but actually even the two-period framework (with a varying d) reveals that this is not so: although, given a two-valued tariff p, there is just one efficient capacity ratio δ, it is not determined by the technology alone because it depends on the price duration d (while being independent of the two price levels in p). This is why the example is not limited to the case of $d = T/2$, although it is this case that is shown in Figure 3.6.

Example 3.15.1 (Indeterminacy of marginal values with discontinuous price). *The short-run profit function of the pumped storage technique (3.2.4) can be nondifferentiable in $(k_{\mathrm{St}}, k_{\mathrm{Co}})$. To see this,*

[45] In heuristic terms, this is because in Case (b) the extra price term requires a brief switch from charging to discharging around $\underline{t}$—the briefer the better, so no storage policy is best. (The same idea leads to an example of nonexistence of a consumer optimum with Mackey-continuous preferences but $p \notin L^1$: see [47].) For a formal proof, compare the increments to $\Pi_{\mathrm{SR}}^{\mathrm{PS}}$ and to the value of the output $\hat{y}(p_{\mathrm{CA}})$ that result from adding p_{FA}: in Case (a) both increments are equal (to $k_{\mathrm{Tu}}\|p_{\mathrm{FA}}\| > 0$), so $\hat{y}(p_{\mathrm{CA}})$ remains optimal, i.e., $\hat{y}(p) = \hat{y}(p_{\mathrm{CA}})$. But in Case (b), the profit increases by $k_{\mathrm{Tu}}\|p_{\mathrm{FA}}\|$ by Part 2, whereas the output's value decreases by $k_{\mathrm{Pu}}\|p_{\mathrm{FA}}\|$; so there is no optimum at p (since $\hat{y}(p_{\mathrm{CA}})$ is the only possibility, by Part 3).

take any numbers $\overline{\mathrm{p}} > \underline{\mathrm{p}} \geq 0$ *and* $d \in (0, T)$*, and set a piecewise constant tariff*

$$p(t) := \begin{cases} \underline{\mathrm{p}} & \text{if } t < d \\ \overline{\mathrm{p}} & \text{if } t \geq d \end{cases}.$$

Then, for a storage plant with capacities $(k_{\mathrm{St}}, k_{\mathrm{Co}})$*, a profit-maximising output is*[46]

$$y(t) = \begin{cases} -(k_{\mathrm{St}} \wedge \delta k_{\mathrm{Co}})/d & \text{if } t < d \\ (k_{\mathrm{St}} \wedge \delta k_{\mathrm{Co}})/(T-d) & \text{if } t \geq d \end{cases}$$

where $\delta := d \wedge (T - d)$*. So*

$$\frac{1}{\overline{\mathrm{p}} - \underline{\mathrm{p}}} \Pi_{\mathrm{SR}}^{\mathrm{PS}}(p, k_{\mathrm{St}}, k_{\mathrm{Co}}) = k_{\mathrm{St}} \wedge \delta k_{\mathrm{Co}} := \min\{k_{\mathrm{St}}, \delta k_{\mathrm{Co}}\}. \tag{3.15.1}$$

Therefore $\Pi_{\mathrm{SR}}^{\mathrm{PS}}$ *is nondifferentiable in* $(k_{\mathrm{St}}, k_{\mathrm{Co}})$ *whenever* $k_{\mathrm{St}} = \delta k_{\mathrm{Co}}$.

Comment: With (p, k) as above, an optimum ψ is nonunique; and it is almost completely indeterminate if $d = T/2$: in this case it is any two-valued function

$$\psi(t) = \begin{cases} \underline{\psi} & \text{if } t < T/2 \\ \overline{\psi} & \text{if } t \geq T/2 \end{cases} \tag{3.15.2}$$

subject only to the obvious conditions, viz.,[47]

$$\underline{\mathrm{p}} \leq \underline{\psi} \leq \overline{\psi} \leq \overline{\mathrm{p}}. \tag{3.15.3}$$

3.16 Conditions for equilibrium in commodity space of bounded functions

To ensure that the storage technology can be included in an Arrow-Debreu model of general equilibrium with $L^{\infty}[0, T]$ and $L^{1}[0, T]$ as the commodity and price spaces, two conditions have to be verified. The first one is needed for the existence of a price system in the larger price space $L^{\infty *}$.

Lemma 3.16.1. *The set* $\mathbb{Y}_{\mathrm{PS}}$ *is* $\mathrm{w}(L^{\infty}, L^{1})$*-closed.*

Proof. By the Krein-Smulian Theorem (for which see, e.g., [32, 18E: Corollary 2]), it suffices to show that $\mathbb{Y}_{\mathrm{PS}}$ is closed for the bounded weak* topology of L^{∞}. Since the bounds on k_{Tu} and k_{Pu} bound y

[46]This y implements the policy of carrying over, from the low-price period to the high-price period, as much stock as the capacity constraints allow, viz., $\min\{k_{\mathrm{St}}, \delta k_{\mathrm{Co}}\}$. It is optimal independently of the two price levels, as long as $\overline{\mathrm{p}} > \underline{\mathrm{p}}$. (Also, it is the only *two-valued* optimal output function; but in the class of all functions it is the unique optimum if and only if $d = T/2$ and $k_{\mathrm{St}} \geq k_{\mathrm{Co}} T/2$.)

[47]If $d \neq T/2$ then d replaces $T/2$ in (3.15.2); but additionally $\overline{\psi} = \overline{\mathrm{p}}$ if $d < T/2$, and similarly $\underline{\mathrm{p}} = \underline{\psi}$ if $d > T/2$.

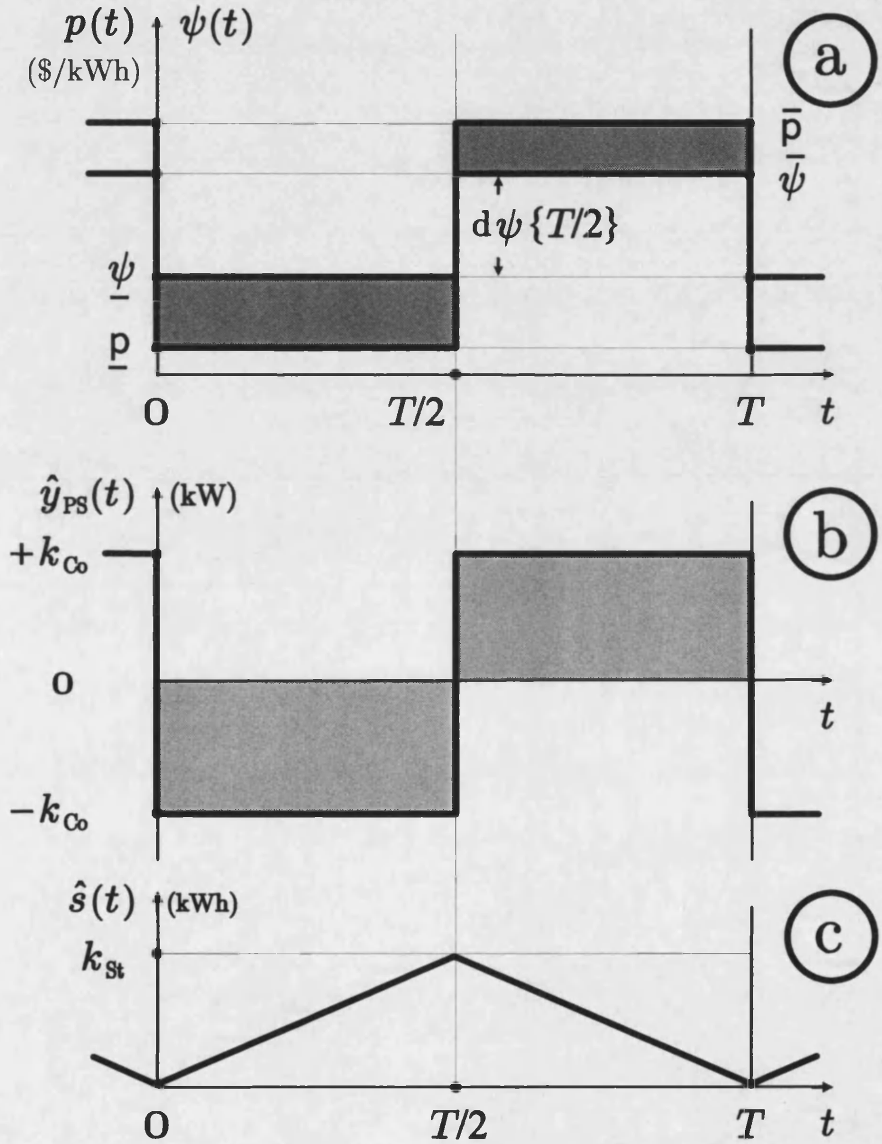

FIGURE 3.6. Indeterminacy of an optimal shadow price of stock ψ (in the case of two equal subperiods). The jump $\overline{\psi} - \underline{\psi}$, which equals the reservoir's unit rent, can take any value not exceeding the jump of p. The total dark grey area in (a) represents the converter's unit rent. In (b), each of the light grey areas equals the reservoir's capacity k_{St} (since $k_{St}/k_{Co} = T/2$).

as well, it suffices to establish that, for each $\overline{k} = (\overline{k}_{\mathrm{St}}, \overline{k}_{\mathrm{Pu}}, \overline{k}_{\mathrm{Tu}}) \in \mathbb{R}^3_+$, the set $\mathbb{Y}_{\mathrm{PS}} \cap \{(y, -k) : k \leq \overline{k}\}$ is weakly* compact. This set is the image, $\pi(S)$, of the set S of all those points $(y_{\mathrm{Tu}}, y_{\mathrm{Pu}}, -k; s_0)$ with $k \leq \overline{k}$ that meet Conditions (3.4.7)–(3.4.10), under the projection map π that sends such a point to $(y_{\mathrm{Tu}} - y_{\mathrm{Pu}}, -k)$. And $\pi(S)$ is weakly* compact because π is weak*-to-weak* continuous, and because S is weakly* compact by the Banach-Alaoglu Theorem. □

To ensure the existence of a price system in the price space L^1, one needs to verify the Exclusion Condition of [43].[48] This is facilitated by the use of an input requirement function.

To work out the storage capacity requirement as a function of f, the signed outflow from the reservoir, the initial stock s_0 is eliminated from the reservoir constraints (3.12.2) on f, by setting s_0 at its lowest possible value (3.4.11). The time-continuum of reservoir constraints is thus summarised as: $\mathrm{Max}(I_0 f) - \mathrm{Min}(I_0 f) \leq k_{\mathrm{St}}$, with I_0 defined by (3.12.1). So the capacity requirement is

$$\check{k}_{\mathrm{St}}(f) = \mathrm{Max}(I_0 f) - \mathrm{Min}(I_0 f) = \mathrm{Max}(I_0 f) + \mathrm{Max}(I_T f) \tag{3.16.1}$$

since $\int f = 0$ implies that $I_0 f = -I_T f$. The requirement function itself is defined only on the subspace $L^\infty_0 = \{f \in L^\infty : \int f = 0\}$, but the above rule defines a finite, convex extension to the whole of L^∞.

Lemma 3.16.2. *The set* $\mathbb{Y}_{\mathrm{PS}}$ *meets the Exclusion Condition of [43].*

Proof. This follows from Mackey continuity of $\check{k}_{\mathrm{St}}$. (The upper semicontinuity is what is relevant here.) To see this, take any $(p, r) \in L^{\infty *} \times \mathbb{R}^3$ and an evanescent sequence of measurable sets $S_m \subset [0, T]$ supporting p_{FA} (so $\mathrm{meas}\, S_m \to 0$ as $m \to \infty$). Take any $(y, -k) \in \mathbb{Y}_{\mathrm{PS}}$; i.e., there exist $(y_{\mathrm{Tu}}, y_{\mathrm{Pu}}) \geq 0$ such that $y = y_{\mathrm{Tu}} - y_{\mathrm{Pu}}$ and

$$y_{\mathrm{Tu}} \leq k_{\mathrm{Tu}},\ \eta_{\mathrm{Pu}} y_{\mathrm{Pu}} \leq k_{\mathrm{Pu}},\ \int_0^T f(t)\,\mathrm{d}t = 0, \text{ and } \check{k}_{\mathrm{St}}(f) \leq k_{\mathrm{St}} \tag{3.16.2}$$

where $f := y_{\mathrm{Tu}}/\eta_{\mathrm{Tu}} - \eta_{\mathrm{Pu}} y_{\mathrm{Pu}}$. As can readily be shown, there exists a sequence $Z_m \supseteq S_m$ with $\mathrm{meas}\, Z_m \to 0$ and $\int_{Z_m} f(t)\,\mathrm{d}t = 0$. Define $y^m_{\mathrm{Pu}} = y_{\mathrm{Pu}} 1_{[0,T] \setminus Z_m}$ and $y^m_{\mathrm{Tu}} = y_{\mathrm{Tu}} 1_{[0,T] \setminus Z_m}$; then

$$y^m := y^m_{\mathrm{Tu}} - y^m_{\mathrm{Pu}} = y 1_{[0,T] \setminus Z_m} \quad \text{and} \quad f^m := \frac{y^m_{\mathrm{Tu}}}{\eta_{\mathrm{Tu}}} - \eta_{\mathrm{Pu}} y^m_{\mathrm{Pu}} = f 1_{[0,T] \setminus Z_m}$$

where 1_A denotes the 0-1 indicator of a set A. Define also

$$k^m_{\mathrm{St}} := k_{\mathrm{St}} - \check{k}_{\mathrm{St}}(f) + \check{k}_{\mathrm{St}}(f^m).$$

Then $\int_0^T f^m = 0$; and $\check{k}_{\mathrm{St}}(f^m) \leq k^m_{\mathrm{St}}$ (from the definitions and the inequality $\check{k}_{\mathrm{St}}(f) \leq k_{\mathrm{St}}$). Also, $y^m_{\mathrm{Pu}} \leq y_{\mathrm{Pu}} \leq k_{\mathrm{Pu}}/\eta_{\mathrm{Pu}}$ and $y^m_{\mathrm{Tu}} \leq y_{\mathrm{Tu}} \leq k_{\mathrm{Tu}}$. As $m \to \infty$, one has $f^m \to f$ in $\mathrm{m}(L^\infty, L^1)$ and, therefore, $k^m_{\mathrm{St}} \to k_{\mathrm{St}}$ by (3.16.1) and Remark 3.12.2. Put together, this shows that the sequence

$$(y^m, -k^m) = (y^m, -k^m_{\mathrm{St}}, -k_{\mathrm{Pu}}, -k_{\mathrm{Tu}}) \in \mathbb{Y}_{\mathrm{PS}}$$

[48] This is less restrictive than Bewley's Exclusion Assumption in [10].

has the required properties, viz.,

$$\begin{aligned}\langle (p,r)_{\mathrm{FA}} \mid (y^m, -k^m)\rangle &= \langle p_{\mathrm{FA}} \mid y^m \rangle = 0 \\ \langle (p,r)_{\mathrm{CA}} \mid (y^m, -k^m) - (y, -k)\rangle &= \langle p_{\mathrm{CA}} \mid y^m - y\rangle - \langle k^m_{\mathrm{St}} - k_{\mathrm{St}} \mid r_{\mathrm{St}}\rangle \to 0\end{aligned}$$

as $m \to \infty$. □

It follows that pure density prices obtain in a general equilibrium model of peak-load pricing with storage if the users' utility and production functions are Mackey continuous: see [43]. For the case of electricity supplied from thermal generation and pumped storage, this means that if the demand for electricity is interruptible (i.e., a brief interruption causes only a small loss of utility or output), then the equilibrium TOU tariff is a time-varying rate in \$/kWh (with no instantaneous charges in \$/kW).

3.17 Conclusions from Chapter 3

This analysis gives, for the first time, a sound basis for valuation and optimal operation of existing pumped-storage plants, as well as for investment decisions. This model of the technology distinguishes the different types of capacity within a storage plant, viz., the reservoir and the converter. Their marginal contributions to the operating profit turn out to be well defined, at least when the given TOU price is continuous over the cycle. These values can be calculated by solving a linear programme (or an equivalent convex programme). It has also been shown how to use the marginal values to determine the optimum investment into storage. The framework is flexible and can deal with similar storage problems: for example, that of hydroelectric generation studied in Chapter 4 (which, unlike pumped storage, is a case of storing an exogenous inflow).

Chapter 4

EFFICIENCY RENTS OF HYDROELECTRIC STORAGE PLANTS AND THEIR USES FOR OPERATION AND INVESTMENT DECISIONS

4.1 Introduction to Chapter 4

In view of the economic significance of hydroelectric generation to many countries, the scale of the investment it can entail, and the planning and operational difficulties it presents, it is unsurprising that it is the subject of so much study by engineers, operations researchers and economists. A common economic understanding should instruct this work and, indeed, in his 1957 article [55], and again in his 1975 Nobel Lecture [56, pp. 262–263], Koopmans pointed to the efficiency rents of the fixed inputs (river flow, reservoir and turbine) as the elements that can underpin the various approaches. In fact, the models of different researchers have remained largely separate,[1] and this is because of the technical obstacles faced by economists in taking their part of the project forward—for although Koopmans's work is much cited, it has never been used in practice or, until now, followed up in theory.

Koopmans's operation problem is recast here as one of competitive profit maximisation, which is the relevant setting for modern decentralised electricity supply industries. Several advances in mathematical economics inform the solution, and Koopmans's continuous-time formulation can now be handled as part of a general equilibrium problem in an infinite-dimensional commodity space. The framework used is the adaptation of Bewley's equilibrium model [43] that has been developed to investigate Boiteux's conjectures on the peak-load pricing of electricity [12, 3.4 and 3.3.3]. Koopmans's scheme, like Boiteux's, is marginalist, and both encounter the problem of nondifferentiability of joint cost functions. Subdifferentials are employed to describe multi-valued derivatives and generalisations of the smooth-calculus results that economists commonly use are worked out, including a subdifferential version of the Wong-Viner Envelope Theorem on the equality of short-run and long-run marginal costs (Chapter 5).[2] The short-run approach to long-run general equilibrium devised in

[1] In [41, Sections 5 and 6], Koopmans's analysis is discussed in more detail, and other work on cost minimisation for a hydro-thermal system is reviewed, including [26, Chapters 5 and 6], [52] and [64]. See also the overview in [3, pp. 277–282]. A more recent operational study of hydro in [7], [29] and [67] is set up as a profit maximisation problem for a single hydro plant, but it concentrates entirely on operation and does not address the economic questions of valuation and investment (it makes no use of the dual solution ψ, except as a tool for deriving the primal operating solution, and does not point to the interpretation of ψ as the marginal value of water).

[2] See also Chapter 2 or [36] for subdifferentiation of symmetric functions, such as the thermal generation cost (as a function of the output bundle).

Chapter 5 is the wider conceptual setting for this study of hydro. This is because a key element of the short-run approach is the profit-based valuation of capital inputs.

Koopmans undertakes the task of minimising the operating cost of an entire electricity supply system by constructing a water storage plan for the hydro-plant operation that minimises the fuel cost incurred by the thermal generating plant in producing a given output of electricity. From this operating solution he imputes time-of-use (TOU) values both to electricity (p) and to water (ψ), and thence the two hydro capacities, viz., the reservoir and the generator. These shadow prices enable him to verify that his water storage plan is optimal. His objectives are of particular interest to a centralised utility (with a predominantly thermal system) that seeks efficient utilisation of its plant and needs to calculate the marginal costs of electricity in its system. However, he adds greatly to his difficulties by setting out to infer all the values associated with the hydro plant (the shadow prices of electricity, water and the hydro capacities) in one fell swoop. As is shown in Chapter 5, it is much simpler to split the complex problem of simultaneous valuation of both outputs and fixed inputs into subproblems, one of which entails short-run profit calculations—even when cost minimisation is the explicit operational objective.[3] Furthermore, Koopmans's method has little or no place in today's largely deregulated and decentralised supply industry in which each plant aims to maximise its own profit (as opposed to participating in the collective cost-minimising operation of a system of plants).

Having profit maximisation as the optimising principle not only allows us to address the problems of a decentralised supply industry (as well as those of a centralised utility), but also it facilitates a full and simple solution. In particular: (i) profit-maximising hydro operation and the dual problem of valuation can be handled (in the case of a constant hydrostatic head) by means of linear programmes (LPs), rather than the convex programmes (CPs) needed for Koopmans's cost minimisation; (ii) one can depict simply the solutions to the operation and valuation LPs, which is not possible with Koopmans's operation CP or its dual; and (iii) profit-imputed values of the hydro capacities and the river flow (i.e., their marginal contributions to the operating profit) turn out to be fully determinate—unlike Koopmans's cost-imputed values of the hydro inputs (i.e., the marginal savings on the thermal

[3]In terms of the sub- or super-differential, ∂ or $\widehat{\partial}$—a generalised, multi-valued derivative of a convex or concave function—the split calculation uses the rule: $(p, -r) \in \partial_{y,k} C_{\mathrm{SR}}(y, k)$ if and only if both $p \in \partial_y C_{\mathrm{SR}}(y, k)$ and $r \in \widehat{\partial}_k \Pi_{\mathrm{SR}}(p, k)$, where C_{SR} is the operating a.k.a. short-run cost as a function of the output bundle y and fixed-input bundle k, and Π_{SR} is the operating profit as a function of k and the output price system p. If the joint marginal values are nonunique (i.e., $\partial_{y,k} C_{\mathrm{SR}}$ is multi-valued because C_{SR} is nondifferentiable) then, for a $p \in \partial_y C_{\mathrm{SR}}(y, k)$, the set $\widehat{\partial}_k \Pi_{\mathrm{SR}}(p, k)$ is generally a *proper* subset of $-\partial_k C_{\mathrm{SR}}(y, k)$, and it may even be a singleton (in which case the ordinary gradient vector $\nabla_k \Pi_{\mathrm{SR}}$ exists): indeed, this is so in Theorem 4.9.3. That is why $\partial_{y,k} C_{\mathrm{SR}}$ does *not* factorise into the Cartesian product of $\partial_y C_{\mathrm{SR}}$ and $\partial_k C_{\mathrm{SR}}$, and why its calculation in terms of partial subdifferentials requires the function Π_{SR} (which is, by definition, a partial convex conjugate of C_{SR}). Like all results on marginal values of optimisation programmes, the splitting of $\partial_{y,k} C_{\mathrm{SR}}$ can be reformulated in terms of programme solutions (in particular, any r from $\widehat{\partial}_k \Pi_{\mathrm{SR}}$ can be obtained from the fixed-input valuation programme that is dual to the profit-maximising operation programme). This is also spelt out in Chapter 5.

operating cost).

In this setup, the TOU electricity value $p(t)$ is a *given* market price; it is *not* an imputed shadow price as in Koopmans's analysis. Given p, an optimal TOU value, $\psi(t)$, is imputed to water (or, more precisely, to the water's potential energy). This essentially solves the operation problem (Section 4.8): the hydro plant is operated just like a thermal plant with a time-varying "fuel" price $\psi(t)$. It makes sense, then, to value a hydro plant's capacities by their marginal contributions to the maximum operating profit, a.k.a. short-run profit $\Pi^{\mathrm{H}}_{\mathrm{SR}}$. These rents can be expressed in terms of the electricity and water values p and ψ (Theorem 4.9.3). For a hydro plant with a constant head, the shadow price for water (ψ) can be determined from a linear programme (4.5.1)–(4.5.7) that is dual to the LP of profit-maximising operation (4.4.4)–(4.4.8).[4] By using the dual constraints to eliminate the dual variables other than ψ, the dual is reformulated as a convex but *unconstrained* programme of shadow pricing the water (4.7.5)–(4.7.7). This leads to a simple characterisation of the solution: the optimal ψ is obtained from p by "shaving off" the local peaks of p and "filling in" its troughs, and the optimal output (y) follows from this (Section 4.3, Figures 4.1a and 4.1b).

This last insight also makes it easier to identify a critical case of the dual solution: the imputed TOU value of water (the function ψ) is unique if the given TOU electricity price (p) is a continuous function of time (Lemma 4.9.2). It then follows that the capital inputs (reservoir and turbine) also have definite marginal values ($\partial\Pi^{\mathrm{H}}_{\mathrm{SR}}/\partial k_{\mathrm{St}}$ and $\partial\Pi^{\mathrm{H}}_{\mathrm{SR}}/\partial k_{\mathrm{Tu}}$). This is *not* so in Koopmans's cost-minimum framework because he has to value both flows (electricity and water), and the shadow-price pair (p and ψ) is typically indeterminate (although for each continuous p there is just one ψ). With competitive profit maximisation, the output price (p) is unique simply because it is treated as given, and although a fixed-input's shadow price (ψ) might still be indeterminate, this possibility can be excluded by a problem-specific argument (which in this case consists in examining the structure of Lagrange multipliers for the capacity constraints). This is a major advantage of the profit approach.

The imputed values of the hydro capacities and the river flow are useful in making investment decisions, whether to expand an existing plant or to develop a new hydro site. This is an end Koopmans envisaged for his cost-imputed values,[5] but their nonuniqueness causes complications because, for example, it means that the incremental value of investment becomes nonadditive (superadditive) when two or more hydro inputs are being varied. Such calculations are made much simpler by using the profit-imputed values: being unique, they can be simply equated to the corresponding marginal costs of investment to determine the optimal capacities. Also, the dual LP (or the equivalent unconstrained CP) gives a simple and direct way of imputing the values. By contrast, Koopmans's values are derived from a tortuous operating solution. They do serve his immediate purpose—to verify the

[4] The dual is the problem of minimising the plant's value subject to the constraints that decompose the given price $p(t)$ into the sum of values of the plant's capital services (κ) and the Lagrange multipliers (ν) for the nonnegativity constraints on water stock and electricity output (plus a constant, λ).

[5] See [55, pp. 194, 200, 225–226].

cost-optimality of his storage plan—but the nonuniqueness of his values is an obstacle to their use in practical investment planning.

Time-continuity is not just a natural assumption on the electricity price p and the only one to guarantee uniqueness of the imputed values of water and of the hydro capacities: it is also an assumption that is verified for competitive equilibrium in the commodity space of bounded functions in [45], where the price function is *proved* to be continuous for a class of problems that includes peak-load pricing with storage.[6]

Section 4.2 describes the hydro technology. Formal analysis (with proofs in the Appendix) is preceded by heuristics, in Section 4.3. In Sections 4.4 and 4.5, the short-run profit maximum problem and its dual are set up as linear programmes (LPs) which are doubly infinite: with continuous-time dating of commodities, the primal (4.4.4)–(4.4.8) contains continua of output and input variables (electricity y and river flow e) as well as continua of dated capacity constraints (on the electricity flow y and on the water stock s). The primal and the dual are shown to be soluble, and their (optimal) solutions are described in Section 4.6. In Section 4.7, the dual LP is reformulated as an unconstrained CP of shadow-pricing the water. In Sections 4.8 and 4.9, the optimal water price ($\hat{\psi}$) is shown to be unique if the given electricity price p is continuous over time; and formulae are given, in terms of p and $\hat{\psi}$, for the optimal output ($\hat{y}$) and for the profit derivatives that represent the marginal values of the reservoir and the turbine ($\partial\Pi^{\mathrm{H}}_{\mathrm{SR}}/\partial k_{\mathrm{St}}$ and $\partial\Pi^{\mathrm{H}}_{\mathrm{SR}}/\partial k_{\mathrm{Tu}}$). This completes the core matter, which is followed by proofs (along with the required auxiliary results) in Section 4.10.

The rest of this chapter consists of various supplements. Section 4.11 indicates the changes needed when the policy of pure "coasting" ($y = e$) is infeasible, i.e., when $e \not\leq k_{\mathrm{Tu}}\cdot$; the hydro plant operation may then necessitate spillage. Section 4.12 sketches the use of marginal capacity values as a basis for calculating the optimum investment. Section 4.13 presents a counterexample to the existence of $\nabla_{k,e}\Pi$ when the price p is a discontinuous step function (so that time is effectively a discrete variable). Section 4.14 deals with the possibility that the price system (p) may, in general, contain a concentrated charge (such as an instantaneous capacity charge in \$/kW), in addition to a price density function (which is a time-dependent price rate in \$/kWh).

Table 4.1 summarises the notation.

In [48], the analysis, and especially the valuation method, is extended to the case of a variable head. This requires the use of a controlled differential equation, and the optimisation problem becomes nonlinear (although it remains convex). Another reason for presenting that case separately is that the "hydro" technology has other interpretations, in which there is no equivalent of head variability. For example, the model herein is applicable to water supply (when priced by TOU), as well as to other natural energy flows (e.g., geothermal or tidal).

[6] The case of thermal electricity generation with pumped storage is fully worked out in [45]; the case of hydro-thermal generation is similar.

Role in programme pair	Variable	Notation	Unit
Price data (dual parameters)	electricity price at time t	$p(t)$	\$/kWh
Quantity data (primal parameters)	reservoir capacity	$k_{\mathrm{St}}(t) = \mathrm{const.}$	kWh
	water-stock floor	$n_{\mathrm{St}}(t) = 0$	kWh
	turbine capacity	$k_{\mathrm{Tu}}(t) = \mathrm{const.}$	kW
	electr. output floor	$n_{\mathrm{Tu}}(t) = 0$	kW
	river inflow rate at t	$e(t)$	kW
	top-up of stock	$\zeta = 0$	kWh
Quantity decisions (primal variables)	hydroelectric output (water discharge rate) at time t	$y(t)$	kW
	water stock at time 0 and T	s_0	kWh
Derived quantities	rate of outflow from reservoir at time t	$f(t) := y(t) - e(t)$	kW
	water stock at time t	$s(t) := s_0 - \int_0^t f(\tau)\,\mathrm{d}\tau$	kWh
Shadow prices (dual decision variables, paired to primal parameters)	unit reservoir value on interval length $\mathrm{d}t$	$\kappa_{\mathrm{St}}(\mathrm{d}t)$	\$/kWh
	value of stock floor (nonnegativity constraint)	$\nu_{\mathrm{St}}(\mathrm{d}t)$	\$/kWh
	unit turbine value at time t	$\kappa_{\mathrm{Tu}}(t)$	\$/kWh
	value of output floor (nonnegativity constraint)	$\nu_{\mathrm{Tu}}(t)$	\$/kWh
	water value at time t	$\psi(t)$	\$/kWh
	water value at 0 and T	λ	\$/kWh
Derived valuations	total reservoir rent for whole cycle $[0,T]$	$\kappa_{\mathrm{St}}[0,T] = \int_0^T \kappa_{\mathrm{St}}(\mathrm{d}t)$	\$
	total turbine rent	$\int_0^T \kappa_{\mathrm{Tu}}(t)\,\mathrm{d}t$	\$

TABLE 4.1. Notation for Chapter 4. Some functions of time (k_{St}, etc.) are equated to "const.". This indicates that they are constants in the original, unperturbed programme, but are perturbed with time-varying increments ($\Delta k_{\mathrm{St}}(t)$, etc.) to interpret the time-dependent dual variables ($\Delta\kappa_{\mathrm{St}}$, etc.). The duality scheme (Section 5) similarly uses a nonzero increment $\Delta\zeta$ to $\zeta = 0$ (paired with the dual variable λ). NB: $\psi(t) = \lambda + (\kappa_{\mathrm{St}} - \nu_{\mathrm{St}})[0,t]$ by a constraint of the dual (valuation) LP. Also, Section 11 uses an extra primal variable σ to denote spillage.

4.2 Hydro technology

Hydro generation produces electricity, a nonstorable good with a cyclical demand and price, from a storable input of water.[7] It is assumed that a water stock, up to the *reservoir's capacity*, k_{St}, can be held at no running cost (or loss of stock). Water is stored at a height, called the *head*, which determines its potential energy. This is converted in penstocks to kinetic energy, and then to electric energy by a turbine-generator (or "turbine" for brevity). The effective head is assumed to be fixed. Therefore the energy stock $s(t)$ is always in a constant proportion to the stored water volume, and it can be referred to as "water". Similarly, the rate of river inflow, $e(t)$, can be measured in terms of power (instead of volume per unit time).

The turbine-generator's technical efficiency is also taken to be constant.[8] The water stock can therefore be measured as the output it actually yields on conversion (i.e., in kWh of electric energy). The *turbine capacity*, k_{Tu}, is its maximum output rate (in kW of electric power), i.e., in unit time a unit turbine can convert a unit of stock into a unit of output.

The river inflow e is taken to be known with certainty. It varies periodically over time; and a common cycle for the water inflow and the output price is represented by an interval $[0, T]$ of the real line $\mathbb{R}$. The cycle is generally a year because of seasonal variation.[9]

The inflow function e is usually continuous, but it suffices to assume that e is assume that e is bounded. That is, e belongs to $L^{\infty}[0, T]$, which is the vector space of all essentially bounded functions, with functions equal almost everywhere (a.e.) being identified with each other. This space is normed by the supremum norm

$$\|e\|_{\infty} := \mathrm{EssSup}\,|e| = \operatorname*{ess\,sup}_{t\in[0,T]} |e(t)|.$$

The hydro plant's output rate is also a periodic function, y. A *storage policy* generally consists of an output rate $y(t) \geq 0$ and a spillage rate $\sigma(t) \geq 0$ for each t. However, except in Section 4.11, spillage is excluded by the assumption that $k_{\mathrm{Tu}} \geq e$. This makes it feasible for the plant to "coast", i.e., to generate at the rate equal to the inflow rate $e(t)$. It also means that all the incentive to use the reservoir comes from a time-dependent output price: if p were a constant, the plant might as well coast all the time.

[7] The model applies also to other forms of natural energy, e.g., geothermal energy. It can be adapted to the case of tidal energy, although this requires changing the assumption that, when $s(t) > 0$, the output rate $y(t)$ is constrained only by k_{Tu} and is, therefore, independent of the inflow rate $e(t)$. And it applies also to the supply of other goods, such as water and natural gas (when priced by TOU). In the case of water supply, $e(t)$ is the rainfall collected in reservoirs; its conversion to the consumable good consists in water purification and pumping to users.

[8] In reality the equipment is not perfectly divisible, and a turbine's efficiency varies with the load, reaching 90% to 95% at full load. At one-quarter load it goes down to 80–85% for movable-blade types, or 60–70% for fixed-blade types. The generator's efficiency is 90–95%.

[9] The weekly cycle is also considered, e.g., in [29].

The net outflow from the reservoir is the signed function

$$f = y - e + \sigma \tag{4.2.1}$$

and the stock, $s(t)$ at time t, is an absolutely continuous function on $[0,T]$ that satisfies the evolution equation $\dot{s} := \mathrm{d}s/\mathrm{d}t = -f$. This can be restated as

$$s(t) - s(0) = -\int_0^t f(\tau)\,\mathrm{d}\tau := \int_0^t (-y+e-\sigma)(\tau)\,\mathrm{d}\tau. \tag{4.2.2}$$

So s is actually a Lipschitz function, since $k_{\mathrm{Tu}} \geq y \geq 0$ a.e., and since both e and σ are also bounded (by assumption).

The space of all continuous functions $\mathcal{C}[0,T]$, which contains the Lipschitz functions, is normed by the maximum norm

$$\|s\|_\infty = \mathrm{Max}\,|s| = \max_{t\in[0,T]} |s(t)|.$$

Its norm-dual $\mathcal{C}^*$, which serves as the price space for the services of storage capacity, is identified as the space of all (signed, finite) Borel measures $\mathcal{M}[0,T]$ by means of the bilinear form

$$\langle \mu \,|\, s\rangle := \int_{[0,T]} s(t)\,\mu(\mathrm{d}t)$$

for $s \in \mathcal{C}$ and $\mu \in \mathcal{M}$.

The available capacities are taken to equal the installed capacities, and therefore to be constant over the cycle. This does play a part in some of the main results, including the determinacy of rental values (Lemma 4.9.2 and Theorem 4.9.3). However, to take full advantage of sensitivity analysis, the constant existing capacities k are perturbed with increments Δk which are periodic functions of time; this is further explained in Section 4.5. (The notation Δk, etc., is always to be interpreted as a single symbol meaning "an increment to k".)

On the assumption of constant capacities $k = (k_{\mathrm{St}}, k_{\mathrm{Tu}})$, the long-run production set of the hydro technique is the convex cone

$$\begin{aligned}\mathbb{Y}_{\mathrm{H}} := \Bigg\{ &(y, -k, -e) \in L_+^\infty[0,T] \times \mathbb{R}_-^2 \times L_-^\infty[0,T] : 0 \leq y \leq k_{\mathrm{Tu}} \quad \text{and} \\ &\exists \sigma \in [0,e] \left(\int_0^T (y(t) - e(t) + \sigma(t))\,\mathrm{d}t = 0 \quad \text{and} \right. \\ &\left. \exists s_0 \in \mathbb{R}\ \forall t \quad 0 \leq s_0 - \int_0^t (y(\tau) - e(\tau) + \sigma(\tau))\,\mathrm{d}\tau \leq k_{\mathrm{St}} \right) \Bigg\}.\end{aligned} \tag{4.2.3}$$

This formulation imposes a periodicity constraint on the stock $s(T) = s(0)$ or, equivalently, a balance constraint on the flows to and from the reservoir ($\int_0^T f(t)\,\mathrm{d}t = 0$), but the stock level at the beginning or end of a cycle is taken to be a costless decision variable. In other words, when it is first commissioned, the reservoir comes filled up to any required level at no extra cost, but its periodic operation thereafter is taken to be a technological constraint. As for the constraint $\sigma \leq e$, it is never binding (see Section 4.11), but it is realistic, and it simplifies a proof that $\mathbb{Y}_{\mathrm{H}}$ is weakly* closed (Lemma 4.14.2).

4.3 Heuristics for valuation of water and capacities

To start with, assume that not only the market price of electricity, $p(t)$, but also the shadow price of water, $\psi(t)$, is known.[10] Then the operating decisions can be decentralised within the hydro plant, with the reservoir "buying" water at the price $\psi(t)$ from the river and "selling" it to the turbine, which in turn sells the generated electricity at the market price $p(t)$ outside the plant. Short-run profit maximisation separates into problems with obvious solutions, one for each kind of capacity. The maximum profits of the reservoir and the turbine, $\Pi^{\mathrm{St}}(\psi, k_{\mathrm{St}})$ and $\Pi^{\mathrm{Tu}}(p-\psi, k_{\mathrm{Tu}})$, are both linear in k. A unit turbine can earn the profit flow $(p-\psi)^+$, which is the nonnegative part of $p-\psi$, by generating when $p(t) > \psi(t)$. The profit is earned only at the times of full capacity utilisation, since the optimum output is $y(t) = k_{\mathrm{Tu}}$ when $p(t) > \psi(t)$: see Figures 4.1a and 4.1b.

In total over the cycle, the value of a unit turbine is therefore

$$\frac{\Pi^{\mathrm{Tu}}}{k_{\mathrm{Tu}}} = \int_0^T (p(t) - \psi(t))^+ \, \mathrm{d}t.$$

As for the reservoir, a unit can earn a profit of $\psi(\overline{\tau}) - \psi(\underline{\tau})$ by buying stock at time $\underline{\tau}$ and selling it at a later time $\overline{\tau}$ when $\psi(\overline{\tau}) > \psi(\underline{\tau})$. The value of a unit reservoir is therefore the sum of all shadow price rises in a cycle. In precise terms: if $\psi(T) \geq \psi(0)$, then

$$\frac{\Pi^{\mathrm{St}}}{k_{\mathrm{St}}} = \mathrm{Var}^+(\psi)$$

which denotes the *total positive variation* (a.k.a. upper variation) of ψ, i.e., the supremum of $\sum_m (\psi(\overline{\tau}_m) - \psi(\underline{\tau}_m))^+$ over all finite sets of pairwise disjoint subintervals $(\underline{\tau}_m, \overline{\tau}_m)$ of $(0,T)$. (For a discussion of Var^+ see, e.g., [27, Section 8.1].)

If $\psi(T) < \psi(0)$, the reservoir should start the cycle full, and refill towards the end of the cycle. This brings an extra profit of $\psi(0) - \psi(T)$, so in general the unit rent is the *cyclic* positive variation

$$\mathrm{Var}_{\mathrm{c}}^+(\psi) := \mathrm{Var}^+(\psi) + (\psi(0) - \psi(T))^+. \tag{4.3.1}$$

Later it is shown that actually $\psi(0) = \psi(T)$ if $p(0) = p(T)$ and $p \in \mathcal{C}[0,T]$.

The maximum operating profit of the whole hydro plant ($\Pi_{\mathrm{SR}}^{\mathrm{H}}$) is, however, a function *not* of ψ but of the problem's parameters (p, k, e) alone: ψ is an auxiliary function which must eventually be given in terms of (p, k, e). Then $\partial \Pi_{\mathrm{SR}}^{\mathrm{H}} / \partial k_{\mathrm{St}}$ and $\partial \Pi_{\mathrm{SR}}^{\mathrm{H}} / \partial k_{\mathrm{Tu}}$ can be obtained by substituting the correct ψ into the expressions $\mathrm{Var}_{\mathrm{c}}^+(\psi)$ and $\int_0^T (p(t) - \psi(t))^+ \, \mathrm{d}t$.

The correct value, $\hat{\psi}$, is that water price function which minimises the value of the hydro plant's fixed resources (k, e). So, given a TOU electricity tariff p, one can find $\hat{\psi}$ by unconstrained minimi-

[10] When ψ is formally introduced, as the Lagrange multiplier paired with the parameter e, it is by definition the price of the inflowing water. However, it must equal the price of water stored for hydro generation because, by assumption, there is no alternative use. This is why the inflow's price cannot exceed that of the stock. The reverse inequality is obvious.

sation of

$$k_{\mathrm{St}}\,\mathrm{Var}_{\mathrm{c}}^{+}(\psi)+k_{\mathrm{Tu}}\int_{0}^{T}(p(t)-\psi(t))^{+}\,\mathrm{d}t+\int_{0}^{T}\psi(t)\,e(t)\,\mathrm{d}t \tag{4.3.2}$$

over ψ, an arbitrary bounded-variation function on $(0,T)$.

In the case of $k_{\mathrm{Tu}} > e(t) > 0$ for every t, the sum of the two integrals in (4.3.2) has a minimum at (and only at) $\psi = p$.[11] Therefore the main feature of this programme is the trade-off between minimising the variation (which on its own would require setting ψ at a constant value) and minimising the integrals (which on its own would require setting ψ equal to p). From this trade-off it is clear to what extent the local peaks of p should be "shaved off" and the troughs "filled in" to obtain the optimum shadow price $\hat{\psi}$—at least in the case that the market price p is piecewise strictly monotone and $k_{\mathrm{Tu}} > e > 0$ at all times. (An extension dispensing with the upper bound on e is sketched in Section 4.11.) The solution, presented graphically in Figure 4.1a, is determined by constancy intervals of $\hat{\psi}$, on each of which $p(t)-\hat{\psi}$ has a constant sign. If $k_{\mathrm{St}}/\operatorname{Min}(e)$ and $k_{\mathrm{St}}/(k_{\mathrm{Tu}}-\operatorname{Max}(e))$, upper bounds on the times needed to fill up and to empty the reservoir, are sufficiently short, then the constancy intervals do not abut. Around a trough of p, there is an interval $(\underline{t},\bar{t})$ characterised by

$$k_{\mathrm{St}}=\int_{\underline{t}}^{\bar{t}}e(t)\,\mathrm{d}t \tag{4.3.3}$$

on which $p(t) < \hat{\psi}$ throughout. Around a local peak of p, there is an interval $(\underline{t},\bar{t})$ characterised by

$$k_{\mathrm{St}}=\int_{\underline{t}}^{\bar{t}}(k_{\mathrm{Tu}}-e(t))\,\mathrm{d}t \tag{4.3.4}$$

on which $p(t) > \hat{\psi}$ throughout. These are the first-order conditions (FOCs) for the dual optimum: (4.3.3) or (4.3.4) is obtained by equating to zero the increment in the minimand (4.3.2) that results from shifting the constant value of ψ by an infinitesimal unit, on an interval around a peak or a trough of p.[12] These conditions make it feasible to produce the "bang-coast-bang" output (viz., $y(t) = k_{\mathrm{Tu}}$ when $\hat{\psi}(t) < p(t)$, $y(t) = e(t)$ when $\hat{\psi}(t) = p(t)$, and $y(t) = 0$ when $\hat{\psi}(t) > p(t)$): the reservoir goes alternately from empty to full and *vice versa* (Figures 4.1b and 4.1c). This is the optimal output.

The same marginal calculation for the dual problem also shows that an optimum ψ can be nonunique if p is discontinuous. Suppose, for example, that p jumps at the beginning, and drops at

[11] This is proved by subdifferentiating, w.r.t. ψ, the two terms $V_{\mathrm{Tu}}(\psi) := k_{\mathrm{Tu}}\int_0^T(p-\psi)^{+}\,\mathrm{d}t$ and $V_{\mathrm{Ri}}(\psi) := \int_0^T \psi e\,\mathrm{d}t$. For a rigorous proof, consider $V = V_{\mathrm{Tu}} + V_{\mathrm{Ri}}$ as a convex and $\|\cdot\|_1$-continuous function on $L^1[0,T]$. It has a minimum at a ψ if and only if $0 \in \partial V(\psi) \subset L^{\infty}$ (i.e., the zero function belongs to the subdifferential, a.k.a. the set of all subgradients, of V at ψ). And $g \in \partial V_{\mathrm{Tu}}(\psi)$ if and only if: $g = k_{\mathrm{Tu}}$ a.e. on $\{t : \psi < p\}$, $k_{\mathrm{Tu}} \geq g \geq 0$ a.e. on $\{t : \psi = p\}$, and $g = 0$ a.e. on $\{t : \psi > p\}$. Also, $\nabla V_{\mathrm{Ri}} = e$. Since $k_{\mathrm{Tu}} > e > 0$ a.e., it follows that $0 \in \partial V(\psi)$ if and only if $\psi = p$ a.e.

[12] Matters complicate when, for relatively large k_{St}, the neighbouring intervals of water collection and of discharge abut; but a similar optimality rule applies to such clusters.

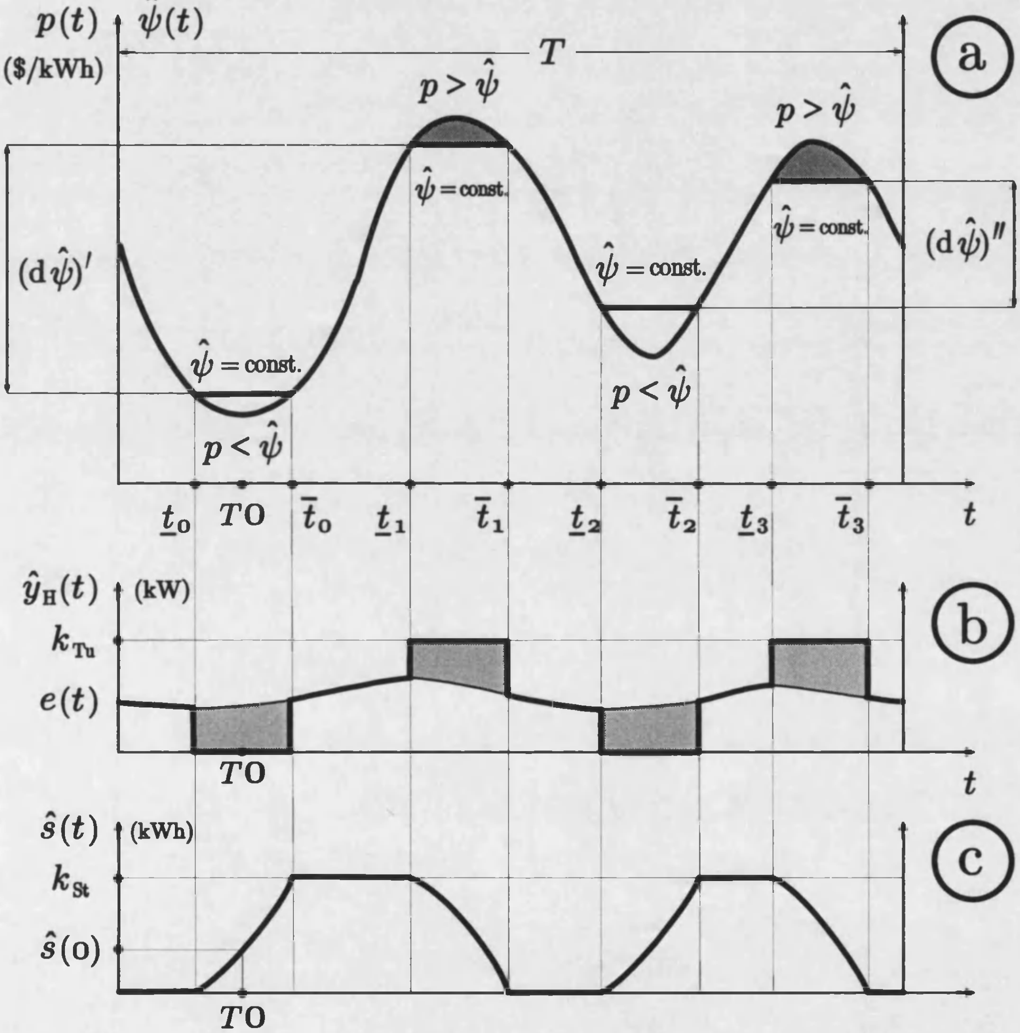

FIGURE 4.1. Trajectories of: (a) shadow price for water $\hat{\psi}$, (b) profit-maximising hydro output $\hat{y}_{\mathrm{H}}$, (c) water stock. Unit rent for storage capacity is $\mathrm{Var}_{\mathrm{c}}^{+}(\hat{\psi}) = (\mathrm{d}\hat{\psi})' + (\mathrm{d}\hat{\psi})''$, the sum of rises of $\hat{\psi}$. Unit rent for turbine capacity is $\int_0^T \left(p(t) - \hat{\psi}(t)\right)^+ \mathrm{d}t$, the sum of dark grey areas in (a). In (b), each of the light grey areas equals the reservoir's capacity k_{St}. When $\hat{y}_{\mathrm{H}}(t) \neq e(t)$ in (b), the thin line is the inflow trajectory e, and the thick line is $\hat{y}_{\mathrm{H}}$.

the end, of an interval $A = (\underline{t}, \bar{t})$ meeting (4.3.4) and the condition

$$\max\left\{p\left(\underline{t}-\right), p\left(\bar{t}+\right)\right\} < \min\left\{p\left(\underline{t}+\right), p\left(\bar{t}-\right)\right\} = \inf_{t \in A} p\left(t\right). \tag{4.3.5}$$

Just before $\underline{t}$ and just after $\bar{t}$, an optimal ψ equals p, i.e., $\psi\left(\underline{t}-\right) = p\left(\underline{t}-\right)$ and $\psi\left(\bar{t}+\right) = p\left(\bar{t}+\right)$. Inside A, $p > \psi = \text{const.}$; but an optimal constant value of ψ on A can be anywhere between the two unequal terms of (4.3.5): the jump and the drop of p create an "indifference zone" for $\psi_{|A}$. Figure 4.2 shows this when $p\left(\bar{t}+\right) \leq p\left(\underline{t}-\right) < p\left(\underline{t}+\right) \leq p\left(\bar{t}-\right)$ so $p\left(\underline{t}-\right) \leq \psi_{|A} \leq p\left(\underline{t}+\right)$. Different values from this range divide the same total rent differently between the three fixed inputs: the jump $\mathrm{d}\psi\left\{\underline{t}\right\} := \psi\left(\underline{t}+\right) - \psi\left(\underline{t}-\right)$ can be any fraction of $p\left(\underline{t}+\right) - p\left(\underline{t}-\right)$, and it is an indeterminate contribution to the reservoir's unit rent. The interval's contribution to the turbine's rent, $\int_A \left(p\left(t\right) - \psi\right) \mathrm{d}t$, is similarly indeterminate (since it depends on the arbitrary choice of $\psi\left(\underline{t}+\right)$, which fixes the constant value of ψ on A). And the indeterminate $\psi_{|A}$ itself is the river's unit rent, on A.[13]

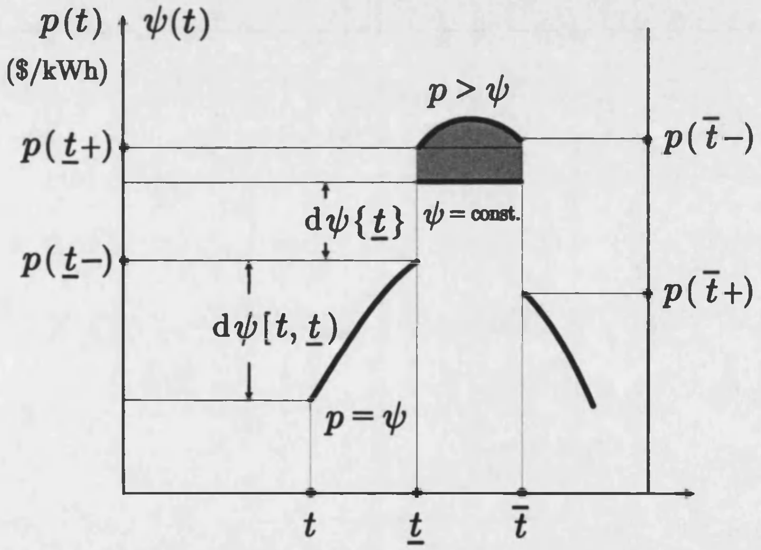

FIGURE 4.2. Indeterminacy of an optimal shadow price of water ψ when the TOU price of good p is discontinuous. In the case shown, the constant value of ψ on $(\underline{t}, \bar{t})$ can be set at any level between $p\left(\underline{t}-\right)$ and $p\left(\underline{t}+\right)$; so the jump of ψ at $\underline{t}$ is an indeterminate part of the reservoir's unit rent. The dark grey area represents $\int_{\underline{t}}^{\bar{t}} \left(p\left(t\right) - \psi\left(t\right)\right)^{+} \mathrm{d}t$, the interval's contribution to the turbine's unit rent.

Conversely, given a continuous p, there is a unique optimum, $\hat{\psi}$ (Lemma 4.9.2). Therefore the gradient $\nabla_{k,e}\Pi$ exists, and $\nabla_e \Pi = \hat{\psi}$ (Theorem 4.9.3). The directional derivative of Π w.r.t. the

[13] The case of p dropping at the beginning, and jumping at the end, of an interval $A = (\underline{t}, \bar{t})$ that meets Condition (4.3.3) is similar, except that the turbine's rent on A is zero (since $p < \psi_{|A}$).

capacities and the inflow is then a linear function of their increments; i.e.,

$$\mathrm{D}\Pi^{\mathrm{H}}_{\mathrm{SR}}\left(\Delta k_{\mathrm{St}}, \Delta k_{\mathrm{Tu}}, \Delta e\right) = \frac{\partial \Pi^{\mathrm{H}}_{\mathrm{SR}}}{\partial k_{\mathrm{St}}} \Delta k_{\mathrm{St}} + \frac{\partial \Pi^{\mathrm{H}}_{\mathrm{SR}}}{\partial k_{\mathrm{Tu}}} \Delta k_{\mathrm{Tu}} + \left\langle \nabla_e \Pi^{\mathrm{H}}_{\mathrm{SR}} \,|\, \Delta e \right\rangle \tag{4.3.6}$$

$$= \Delta k_{\mathrm{St}} \operatorname{Var}^{+}_{\mathrm{c}}\left(\hat{\psi}\right) + \Delta k_{\mathrm{Tu}} \int_0^T \left(p(t) - \hat{\psi}(t)\right)^+ \mathrm{d}t + \int_0^T \hat{\psi}(t)\, \Delta e(t)\, \mathrm{d}t$$

(with all the derivatives and $\hat{\psi}$ evaluated at the given k and e). So the *profit*-imputed value of investment is (jointly) additive in all the increments—unlike Koopmans's *cost*-imputed incremental value (which must be calculated from a multi-valued subdifferential $\partial_{k,e} C$ of the short-run cost, instead of the single-valued gradient $\nabla_{k,e}\Pi$).

4.4 The linear programme of profit-maximising plant operation

For a hydro plant with storage and turbine capacities $k = (k_{\mathrm{St}}, k_{\mathrm{Tu}})$, and with a river inflow e, the operation problem is to maximise the value of electricity output y, at a given TOU price represented by an integrable function p on $[0, T]$, subject to the technological constraints in (4.2.3), i.e.,

$$\text{Given } (p; k_{\mathrm{St}}, k_{\mathrm{Tu}}; e) \in L^1[0,T] \times \mathbb{R}^2 \times L^\infty[0,T] \tag{4.4.1}$$

$$\text{maximise } \langle p \,|\, y \rangle \text{ over } y \in L^\infty[0,T] \tag{4.4.2}$$

$$\text{subject to } (y, -k_{\mathrm{St}}, -k_{\mathrm{Tu}}, -e) \in \mathbb{Y}_{\mathrm{H}} \text{ defined by (4.2.3).} \tag{4.4.3}$$

The plant operation problem is next formulated as an LP by expanding the technological constraint (4.4.3) into turbine constraints on the output rate, and reservoir constraints and a balance constraint on the water stock. To exclude spillage (i.e., make it unnecessary and unprofitable), it is assumed that the river inflow rate never exceeds the turbine capacity (i.e., that $e \le k_{\mathrm{Tu}}$ until Section 4.11), and that the electricity price is strictly positive at all times (i.e., $p \in L^1_{++}[0,T]$). With the constants k_{St} and k_{Tu} viewed as special cases of cyclically varying functions, the primal LP of plant operation is:

$$\text{Given } (p; k_{\mathrm{St}}, k_{\mathrm{Tu}}; e) \in L^1_{++} \times \mathbb{R}^2_+ \times L^\infty_+ \subset L^1_{++} \times \left(\mathcal{C}_+ \times L^\infty_+\right) \times L^\infty_+ \text{ with } k_{\mathrm{Tu}} \ge e \tag{4.4.4}$$

$$\text{maximise } \langle p \,|\, y \rangle \text{ over } y \in L^\infty[0,T] \text{ and } s_0 \in \mathbb{R} \tag{4.4.5}$$

$$\text{subject to: } 0 \le y(t) \le k_{\mathrm{Tu}} \quad \text{for a.e. } t \tag{4.4.6}$$

$$\int_0^T f(t)\, \mathrm{d}t = 0 \tag{4.4.7}$$

$$0 \le s_0 - \int_0^t f(\tau)\, \mathrm{d}\tau \le k_{\mathrm{St}} \quad \text{for every } t \tag{4.4.8}$$

where $f := y - e$, as per (4.2.1) with $\sigma = 0$.

Notation The optimal value of (4.4.1)–(4.4.3) or (4.4.4)–(4.4.8 is the (maximum) operating profit of the hydro plant, denoted by $\Pi^{\mathrm{H}}_{\mathrm{SR}}(p, k, e)$. The (optimal) solution set of (4.4.1)–(4.4.3) is

$\hat{Y}_{\mathrm{H}}(p,k,e)$, abbreviated to $\hat{Y}$. The corresponding lowercase notation $\hat{y}$ is used *only when the solution is known to be unique.* Also, the space L^1 appearing in (4.4.1) consists of all functions integrable with respect to (w.r.t.) meas, the Lebesgue measure. The integral $\int_0^T p(t)\, y(t)\, \mathrm{d}t$ is also written as $\langle p \,|\, y\rangle$. The condition $p > 0$ a.e. on $[0,T]$ is also written as $p \gg 0$, or as $p \in L^1_{++}$.

The two formulations of the operation problem are equivalent in the sense that y solves (4.4.1)–(4.4.3) if and only if y together with some s_0 solves (4.4.4)–(4.4.8)—in which case y together with the specific value

$$\underline{s}_{0,y} := \max_{t\in[0,T]} \left(\int_0^t f(\tau)\,\mathrm{d}\tau\right) = \max_{t\in[0,T]} \left(\int_0^t (y(\tau) - e(\tau))\,\mathrm{d}\tau\right) \tag{4.4.9}$$

is a solution: $\underline{s}_{0,y}$ is the lowest initial stock required for $s(t)$ never to fall below 0. (Unless there is spare storage capacity, this is actually the only feasible value for s_0, given y.) One can therefore restrict attention to points (y, s_0) with $s_0 = \underline{s}_{0,y}$; and so the stock trajectory associated with a hydro output y is

$$s(t) = \underline{s}_{0,y} - \int_0^t (y(\tau) - e(\tau))\,\mathrm{d}\tau. \tag{4.4.10}$$

The dual programme, introduced next, serves the purposes of characterising optimal operation and calculating the marginal values of the capacities and the inflow. To ensure that the dual has a solution of the kind sketched in Section 4.3, for the most part it is assumed from here on that

$$k_{\mathrm{Tu}} > \mathrm{EssSup}\,(e) \geq \mathrm{EssInf}\,(e) > 0 \quad \text{and} \quad k_{\mathrm{St}} > 0. \tag{4.4.11}$$

This means that the "pure coasting" policy (i.e., $y = e$ with $\sigma = 0$) is feasible and, furthermore, that it verifies Slater's Condition for the primal.[14]

4.5 Fixed-input valuation as the dual linear programme

As is set out in, e.g., [73], the dual to a convex programme depends on the choice of perturbations for the primal parameters. A choice of admissible perturbations determines the structure of the dual variables (a.k.a. Lagrange multipliers) to be paired with the parameter increments. Therefore, the dual programme depends not only on the particular values of the primal parameters, but also on the vector space of parameter increments or perturbations. This "ambient" space for the given parameter point can be chosen to suit one's purpose.

[14] This standard constraint qualification for CPs is, in the infinite-dimensional case, useful with LPs as well. Without it, the primal and dual values may be different, or there may be no dual optimum. For example, if $p \in L^1$, $k_{\mathrm{Tu}} > e > 0$ but $k_{\mathrm{St}} = 0$, then the primal and dual values are the same (viz., $\int pe\,\mathrm{d}t$), but an (exact) dual optimum exists only if $p \in \mathrm{BV}$ (in which case the optimal stock price is $\hat{\psi} = p$). See also [4, p. 31.].

In the case of (4.4.4)–(4.4.8), the programme contains a separate set of capacity constraints for each time t—so, by considering a separate increment $\Delta k(t)$ for each t, one can impute an instantaneous value, $\kappa(t)$, to capacity services at each time t. In other words, not only their total value, but also its distribution over the cycle can be determined. Even if the existing capacities k are actually constant, it is useful to consider the cyclically varying increments Δk because this gives a marginal interpretation to the time-dependent Lagrange multipliers for the capacity constraints: denoted by $\kappa = (\kappa_{\mathrm{St}}, \kappa_{\mathrm{Tu}})$, these are the unit values of the capacities' services at any particular time. As part of the "variation of constants", a varying increment $\Delta n_{\mathrm{St}}(t)$ to the zero floor for the water stock in (4.4.8) is also considered, as is a varying increment $\Delta n_{\mathrm{Tu}}(t)$ to the zero floor for the turbine output rate in (4.4.6). This gives a marginal interpretation to the time-dependent Lagrange multipliers for the nonnegativity constraints: denoted by $\nu = (\nu_{\mathrm{St}}, \nu_{\mathrm{Tu}})$, these are the unit values of lowering the "floors" at any time. Finally, a scalar $\Delta\zeta$ is an increment to the zero on the r.h.s. of (4.4.7); this can be thought of as the quantity of water taken to be available for topping up the reservoir between cycles. Its multiplier, a scalar λ, is the marginal value of water at the beginning (or end) of cycle. All the multipliers (κ, $\nu$and λ) are terms of the TOU electricity price p in its decomposition (4.5.6)–(4.5.7) below, which is a part of the dual programme's constraints.

The short-run profit maximisation problem (4.4.4)–(4.4.8) is thus embedded in the family of perturbed programmes obtained by adding an arbitrary cyclically varying increment (Δk_{St}, Δn_{St}, Δk_{Tu}, Δn_{Tu}, Δe) and a scalar $\Delta\zeta \in \mathbb{R}$ to the particular parameter point consisting of the constants (k_{St}, 0, k_{Tu}, 0, e) and 0. The function spaces for the resource increments, indicated already in (4.4.4), are: $\mathcal{C}[0,T]$ for Δk_{St} and Δn_{St}, and $L^{\infty}[0,T]$ for Δk_{Tu} and Δn_{Tu}. These are paired with $\mathcal{M}[0,T]$ and $L^{1}[0,T]$ as spaces for the shadow prices, i.e., Lagrange multipliers. (The pairing of L^{∞} with its norm-dual $L^{\infty *}$, instead of the smaller space L^{1}, is also needed, but only in proving the dual's solubility: both κ_{Tu} and ν_{Tu} are actually in L^{1}, although $\kappa_{\mathrm{Tu}} \in L^{1}$ only because $p \in L^{1}$ instead of $L^{\infty *}$.)

In other words, the marginal value of the storage capacity services on an interval $A \subset [0,T]$ is given by a measure $\kappa_{\mathrm{St}}(A)$; this is the incremental operating profit from the availability on A of an extra unit of the reservoir. Another measure, $\nu_{\mathrm{St}}(A)$, gives the incremental profit from lowering the stock floor by a unit, on A. The marginal value of the turbine capacity services, on A, is the Lebesgue integral of a function $\kappa_{\mathrm{Tu}} \in L^{1}$. The value of lowering the turbine output floor by a unit is the integral of another function, $\nu_{\mathrm{Tu}} \in L^{1}$.

Thus the complete shadow-price system (κ_{St}, ν_{St}; κ_{Tu}, ν_{Tu}; ψ, λ) values all the resource increments (Δk_{St}, $-\Delta n_{\mathrm{St}}$; Δk_{Tu}, $-\Delta n_{\mathrm{Tu}}$; Δe, $\Delta\zeta$). Of course, it also values the particular resource bundle (k_{St}, 0; k_{Tu}, 0; e, 0) that represents the plant itself—and the dual to the operation programme (4.4.4)–(4.4.8) is to minimise the plant's value by an admissible choice of the shadow prices. The main dual constraints (4.5.6)–(4.5.7) decompose the electricity price p into the sum of: the turbine capacity charge κ_{Tu}, minus the value of the output floor ν_{Tu}, and the shadow price of water ψ. The water price is the sum of: the initial price λ, the cumulative of reservoir capacity charges κ_{St}, and the cumulative of $-\nu_{\mathrm{St}}$. This spelt out next.

Theorem 4.5.1 (Fixed-input value minimisation as the dual). *The dual of the linear programme (4.4.4)–(4.4.8), relative to the specified perturbation and the pairing of the parameter spaces $\mathcal{C}$ and L^∞ with $\mathcal{M}$ and L^1 respectively, is:*

$$\text{Given } (p, k_{\mathrm{St}}, k_{\mathrm{Tu}}, e) \text{ as in (4.4.4)} \tag{4.5.1}$$

$$\text{minimise } k_{\mathrm{St}} \int_{[0,T]} \kappa_{\mathrm{St}}(\mathrm{d}t) + k_{\mathrm{Tu}} \int_0^T \kappa_{\mathrm{Tu}}(\mathrm{t})\,\mathrm{d}t + \int_0^T \psi(\mathrm{t})\, e(\mathrm{t})\,\mathrm{d}t \tag{4.5.2}$$

$$\text{over } \lambda \in \mathbb{R},\ \psi \in L^1 \text{ and } (\kappa_{\mathrm{St}}, \nu_{\mathrm{St}}; \kappa_{\mathrm{Tu}}, \nu_{\mathrm{Tu}}) \in \mathcal{M}[0,T] \times \mathcal{M}[0,T] \times L^1 \times L^1 \tag{4.5.3}$$

$$\text{subject to: } (\kappa_{\mathrm{St}}, \nu_{\mathrm{St}}; \kappa_{\mathrm{Tu}}, \nu_{\mathrm{Tu}}) \geq 0 \tag{4.5.4}$$

$$\kappa_{\mathrm{St}}[0,T] = \nu_{\mathrm{St}}[0,T] \tag{4.5.5}$$

$$p = \psi + \kappa_{\mathrm{Tu}} - \nu_{\mathrm{Tu}} \tag{4.5.6}$$

$$\psi = \lambda + (\kappa_{\mathrm{St}} - \nu_{\mathrm{St}})[0,\cdot]. \tag{4.5.7}$$

Remark 4.5.2. *Under (4.4.11), any solution to (4.5.1)–(4.5.7) has the disjointness properties that*

$$\kappa_\phi \wedge \nu_\phi = 0 \quad \text{for } \phi = \mathrm{Tu}, \mathrm{St} \quad \text{and} \quad \kappa_{\mathrm{St}}\{0,T\} \wedge \nu_{\mathrm{St}}\{0,T\} = 0 \tag{4.5.8}$$

i.e., it is not optimal for the dual variables to overlap and partly cancel each other out.[15]

4.6 Conditions for optimal operation and valuation

The dual programme (4.5.1)–(4.5.7) has a solution, (in which $\psi \in \mathrm{BV}(0,T)$ by (4.5.7) and ν_{Tu} and κ_{Tu} are in L^1 because $p \in L^1$, whilst κ_{St} and ν_{St} are in $\mathcal{M}$). The primal and dual optima are characterised by the Kuhn-Tucker Conditions, which for LPs reduce to feasibility and complementary slackness. Spelt out next, these conditions are later used to determine plant operation in terms of the water price, and to establish that this shadow price is unique.

Proposition 4.6.1 (Dual solubility and optimality conditions). *Assume (4.4.11). Then:*

1. *The fixed-input value minimisation programme (4.5.1)–(4.5.7) has an (optimal) solution*

$$(\kappa_{\mathrm{St}}, \nu_{\mathrm{St}}; \kappa_{\mathrm{Tu}}, \nu_{\mathrm{Tu}}; \psi, \lambda) \in \mathcal{M}[0,T] \times \mathcal{M}[0,T] \times L^1 \times L^1 \times \mathrm{BV}(0,T) \times \mathbb{R}.$$

   The programme's value is finite and equal to the short-run profit $\Pi^{\mathrm{H}}_{\mathrm{SR}}(p, k_{\mathrm{St}}, k_{\mathrm{Tu}}, e)$, the optimal value of (4.4.4)–(4.4.8).

[15] For $\phi = \mathrm{St}$, this means that κ_{St} and ν_{St} are disjoint as measures on the circle obtained from the interval $[0,T]$, and not only on $[0,T]$ itself.

2. *Points* $(y, \underline{s}_{0,y}) \in L^\infty \times \mathbb{R}$ *and* $(\kappa_{\mathrm{St}}, \nu_{\mathrm{St}}; \kappa_{\mathrm{Tu}}, \nu_{\mathrm{Tu}}; \psi, \lambda)$ *are optimal solutions to, respectively, the primal (4.4.4)–(4.4.8) and the dual (4.5.1)–(4.5.7) if and only if:*

   (a) $(y, \underline{s}_{0,y})$ *and* $(\kappa_{\mathrm{St}}, \nu_{\mathrm{St}}; \kappa_{\mathrm{Tu}}, \nu_{\mathrm{Tu}}; \psi, \lambda)$ *are feasible, i.e., satisfy (4.4.6)–(4.4.8) and (4.5.4)–(4.5.7).*

   (b) *The measure* κ_{St} *is concentrated on* $\{t \in [0,T] : s(t) = k_{\mathrm{St}}\}$, *whilst* ν_{St} *is concentrated on* $\{t : s(t) = 0\}$, *where* s *is given by (4.4.9)–(4.4.10).*

   (c) *The function* κ_{Tu} *vanishes a.e. outside of* $\{t : y(t) = k_{\mathrm{Tu}}\}$, *whilst* ν_{Tu} *vanishes outside of* $\{t : y(t) = 0\}$.

4.7 Shadow pricing of water as the dual problem

The dual problem can be transformed into one of *unconstrained* minimisation over $\psi \in \mathrm{BV}(0,T)$ by using the dual constraints (4.5.5)–(4.5.7) and the disjointness condition (4.5.8) to eliminate the other dual variables (κ_{St}, ν_{St}; κ_{Tu}, ν_{Tu}; λ).

Notation The space $\mathrm{BV}(0,T)$ consists of all functions ψ of bounded variation on $(0,T)$ with $\psi(t)$ lying between the left and right limits, $\psi(t-) = \lim_{\tau \nearrow t} \psi(\tau)$ and $\psi(t+) = \lim_{\tau \searrow t} \psi(\tau)$.[16] A $\psi \in \mathrm{BV}(0,T)$ is extended by continuity to $[0,T]$; i.e., $\psi(0) := \psi(0+)$ and $\psi(T) := \psi(T-)$. The *cyclic positive variation* of ψ is defined by (4.3.1).

If finite numbers $\psi(0-)$ and $\psi(T+)$ are additionally specified, then $\psi \in \mathrm{BV}[0-, T+]$; and such a ψ defines a measure on $[0,T]$ by

$$\mathrm{d}\psi\,[t', t''] := \psi(t''+) - \psi(t'-) \tag{4.7.1}$$

for $t' \le t''$. The Lebesgue-Stieltjes integral of a function s w.r.t. the measure $(\mathrm{d}\psi)^+$ is written as $\int s\,(\mathrm{d}\psi)^+$ or $\int s(t)\,(\mathrm{d}\psi(t))^+$. When $\psi(0-) = \psi(T+)$, the usual variation norm of $(\mathrm{d}\psi)^+$ equals $\mathrm{Var}_c^+(\psi)$.

It is convenient to set

$$\psi(0-) = \psi(T+) = \lambda \tag{4.7.2}$$

so that, from (4.5.7) and (4.5.5),

$$\kappa_{\mathrm{St}} - \nu_{\mathrm{St}} = \mathrm{d}\psi \quad \text{on } [0,T].$$

From this, (4.5.4) and (4.5.8),

$$\kappa_{\mathrm{St}} = (\mathrm{d}\psi)^+ \quad \text{and} \quad \nu_{\mathrm{St}} = (\mathrm{d}\psi)^- \tag{4.7.3}$$

[16] The one-sided limits exist at every t and are equal nearly everywhere (n.e.), i.e., everywhere except for a countable set. Specification of $\psi(t)$ between $\psi(t-)$ and $\psi(t+)$ is unnecessary.

$$\psi(0+) \wedge \psi(T-) \leq \lambda \leq \psi(0+) \vee \psi(T-).$$

i.e., λ lies between $\psi(0+)$ and $\psi(T-)$. All choices of λ from this range are equally good, i.e., contribute the same to (4.5.2). Lastly, from (4.5.4), (4.5.6) and (4.5.8),

$$\kappa_{\mathrm{Tu}} = (p - \psi)^{+} \quad \text{and} \quad \nu_{\mathrm{Tu}} = (p - \psi)^{-}. \tag{4.7.4}$$

Proposition 4.7.1 (Shadow pricing of water as the dual). *Assume (4.4.11). The fixed-input value minimisation programme (4.5.1)–(4.5.7) is then equivalent to the following convex programme:*

$$\text{Given } (p, k_{\mathrm{St}}, k_{\mathrm{Tu}}, e) \text{ as in (4.4.4)}, \tag{4.7.5}$$

$$\text{minimise } k_{\mathrm{St}} \operatorname{Var}_{\mathrm{c}}^{+}(\psi) + k_{\mathrm{Tu}} \int_0^T (p - \psi)^{+} \, \mathrm{d}t + \int_0^T \psi(t)\, e(t) \, \mathrm{d}t \tag{4.7.6}$$

$$\text{over } \psi \in \mathrm{BV}(0, T). \tag{4.7.7}$$

Notation The solution set for (4.7.5)–(4.7.7) is denoted by $\hat{\Psi}_{\mathrm{H}}(p, k, e)$, abbreviated to $\hat{\Psi}$. Again, the corresponding lowercase notation, $\hat{\psi}$, is used *only when the dual solution is unique.*

It is this formulation of the dual that leads to the idea of obtaining $\hat{\psi}$ by "levelling off" the local extremes of p in the way described in Section 4.3. The insight can be developed into a specialised algorithm for the case of a piecewise monotone p. In this approach the dual is tackled first, in the CP form (4.7.5)–(4.7.7), with the primal solution found subsequently. (For comparison, the simplex and other methods for LPs find both solutions simultaneously.)

4.8 Determination of hydro output

The plant operation problem is soluble for every $p \in L^1[0, T]$.

Proposition 4.8.1 (Primal solubility). *Assume that $k_{\mathrm{Tu}} \geq e \geq 0$. If $p \in L^1$, then the short-run profit-maximisation programme (4.4.4)–(4.4.8) has an (optimal) solution (y, s_0). It follows that the problem (4.4.1)–(4.4.3) has a solution, i.e., $\hat{Y}(p, k, e) \neq \emptyset$.*

Once the dual is solved, so that an optimal ψ is known, the operation problem largely reduces to maximisation of instantaneous profits (as Part 2c of Proposition 4.6.1 shows). At each t with $p(t) \neq \psi(t)$, the optimum output $y(t)$ is of the "bang-bang control" type, either k_{Tu} or 0. Any remaining part of an optimal y is a "singular control", which arises at a time t when the instantaneous optimum is multi-valued because $\psi(t) = p(t)$. This part can be determined on the assumption (4.8.1) that p has no plateau: this ensures that $p(t) = \psi(t)$ only when the reservoir is either empty or full; and at those times the output rate must equal $e(t)$. See Figure 4.1.

Proposition 4.8.2 (Hydro output with plateau-less price). *Assume, in addition to (4.4.11) and $p \in L^1_{++}[0,T]$, that p has no plateau, i.e., that*

$$\forall \mathrm{p} \in \mathbb{R}_+ \ \operatorname{meas}\{t : p(t) = \mathrm{p}\} = 0. \tag{4.8.1}$$

If $y \in \hat{Y}(p,k,e)$ and $\psi \in \hat{\Psi}(p,k,e)$, i.e., y solves (4.4.1)–(4.4.3) and ψ solves (4.7.5)–(4.7.7), then

$$y(t) = \begin{cases} k_{\mathrm{Tu}} & \text{if } p(t) > \psi(t) \\ e(t) & \text{if } p(t) = \psi(t) \\ 0 & \text{if } p(t) < \psi(t) \end{cases} . \tag{4.8.2}$$

So (4.4.1)–(4.4.3) has a unique solution $\hat{y}(p,k,e)$.[17]

4.9 Marginal capacity values in terms of water price

By definition, $\Pi^{\mathrm{H}}_{\mathrm{SR}}$ is the optimal value, $\max_y \langle p \,|\, y \rangle$, of the primal (operation) problem. Since the dual and primal values are equal (Proposition 4.6.1), a dual (water-pricing) solution ψ gives Π as the total fixed-input value (the plant's total rent on the capacities and the river inflow); and it has the advantage of giving the marginal values $\nabla_{k,e}\Pi$ as well.

Corollary 4.9.1 (Dual calculation of SR profit). *Assume (4.4.11). Then, for every $\psi \in \hat{\Psi}(p,k,e)$,*

$$\Pi^{\mathrm{H}}_{\mathrm{SR}}(p,k,e) = k_{\mathrm{St}} \operatorname{Var}^+_c(\psi) + k_{\mathrm{Tu}} \int_0^T (p(t) - \psi(t))^+ \,\mathrm{d}t + \int_0^T \psi(t)\, e(t) \,\mathrm{d}t \tag{4.9.1}$$

Furthermore, this sum equals

$$\int_0^T \psi(t)(y(t) - e(t)) \,\mathrm{d}t + \int_0^T (p(t) - \psi(t))\, y(t) \,\mathrm{d}t + \int_0^T \psi(t)\, e(t) \,\mathrm{d}t$$

term-by-term, for every $y \in \hat{Y}(p,k,e)$.[18]

Since Π is positively homogeneous of degree 1 (a.k.a. linearly homogeneous) in (k,e),[19] Euler's Theorem shows that if Π is differentiable in k,[20] then

$$\Pi^{\mathrm{H}}_{\mathrm{SR}}(p,k_{\mathrm{H}},e) = k_{\mathrm{St}} \frac{\partial \Pi^{\mathrm{H}}_{\mathrm{SR}}}{\partial k_{\mathrm{St}}} + k_{\mathrm{Tu}} \frac{\partial \Pi^{\mathrm{H}}_{\mathrm{SR}}}{\partial k_{\mathrm{Tu}}} + \left\langle \nabla_e \Pi^{\mathrm{H}}_{\mathrm{SR}} \,\middle|\, e \right\rangle . \tag{4.9.2}$$

[17] Since y is fully determined in terms of any optimal ψ (so y is unique even though ψ may be nonunique unless $p \in C$).

[18] This shows that the values of the fixed inputs are equal to their profits—$\langle \psi \,|\, y - e \rangle$ for the reservoir, $\langle p - \psi \,|\, y \rangle$ for the turbine, and $\langle \psi \,|\, e \rangle$ for the river—when the shadow price ψ is used to decentralise the operation within the plant (as is described in Section 4.3).

[19] That is, $\Pi(p;\alpha k,\alpha e) = \alpha \Pi(p;k,e)$ for every scalar $\alpha > 0$. Note also that $\hat{Y}$ and $\hat{\Psi}$ are positively homogeneous, in (k,e), of degrees 1 and 0 respectively; i.e., $\hat{Y}(p;\alpha k,\alpha e) = \alpha \hat{Y}(p;k,e)$ and $\hat{\Psi}(p;\alpha k,\alpha e) = \hat{\Psi}(p;k,e)$ for $\alpha > 0$.

[20] If Π is nondifferentiable, then $\Pi(k,e) = r \cdot k + \langle \psi \,|\, e \rangle$ for every $(r,\psi) \in \partial_{k,e}\Pi$ (the superdifferential of Π as a concave function of (k,e)).

A comparison with (4.9.1) suggests that if there is a *unique* optimal ψ, then the partial derivatives of Π do exist and equal the coefficients of k_{St}, k_{Tu} and e in (4.9.1); formally this follows from (4.5.7) and the marginal interpretation of κ_{St}, κ_{Tu} and ψ (spelt out in the Proof of Theorem 4.9.3). And the optimal stock price ψ is indeed unique if p, the TOU price of the good, is continuous over time.

Lemma 4.9.2 (Water price uniqueness and continuity). *In addition to (4.4.11), assume that* $p \in \mathcal{C}_{++}[0,T]$. *Then the dual (4.7.5)–(4.7.7) has a unique (optimal) solution* $\hat{\psi}$, *which belongs to* $\mathcal{C}_{++}[0,T]$. *If additionally* $p(0) = p(T)$, *then also* $\hat{\psi}(0) = \hat{\psi}(T)$.

Theorem 4.9.3 (Efficiency rents of a hydro plant). *Assume that* $p \in \mathcal{C}_{++}[0,T]$. *Then the operating profit of a hydro plant—i.e., the value of the primal problem (4.4.1)–(4.4.3)—is differentiable with respect to the water inflow function (e) and the capacities (of the reservoir and the turbine,* $k = (k_{\mathrm{St}}, k_{\mathrm{Tu}})$*), at every* (k,e) *satisfying (4.4.11). The derivatives defining the unit rents are given by the formulae*

$$\frac{\partial \Pi^{\mathrm{H}}_{\mathrm{SR}}}{\partial k_{\mathrm{St}}}(p,k,e) = \mathrm{Var}^{+}_{c}\left(\hat{\psi}(p,k,e)\right) \tag{4.9.3}$$

$$\frac{\partial \Pi^{\mathrm{H}}_{\mathrm{SR}}}{\partial k_{\mathrm{Tu}}}(p,k,e) = \int_0^T \left(p(t) - \hat{\psi}(p,k,e)(t)\right)^+ \mathrm{d}t \tag{4.9.4}$$

$$\nabla_e \Pi^{\mathrm{H}}_{\mathrm{SR}}(p,k,e) = \hat{\psi} \tag{4.9.5}$$

in which $\hat{\psi}$ *is the unique solution to the dual problem (4.7.5)–(4.7.7) of water pricing.*

4.10 Proofs for Sections 4.5 to 4.9

Except for the shadow-price uniqueness result (Lemma 4.9.2), the proofs are mostly routine applications of duality for optimisation in infinite-dimensional spaces, as expounded in, e.g., [73, Examples 4, 4', 4"] and [4, 3.3–3.7]. To put the primal constraints in the required operator form, define the integrals I_0 and I_T: $L^\infty[0,T] \to \mathcal{C}[0,T]$ by

$$(I_0 f)(t) := \int_0^t f(\tau)\,\mathrm{d}\tau \quad \text{and} \quad (I_T f)(t) := \int_t^T f(\tau)\,\mathrm{d}\tau. \tag{4.10.1}$$

The reservoir constraints (4.4.8) on (y, s_0) can then be rewritten as

$$0 \le s_0 1_{[0,T]} - I_0 f \le k_{\mathrm{St}}. \tag{4.10.2}$$

A formula for the adjoint operation I_0^*: $\mathcal{M}[0,T] \to L^{\infty *}[0,T]$ is needed. (As for the embedding $\mathbb{R} \ni s_0 \mapsto s_0 1_{[0,T]} \in \mathcal{C}$, its adjoint is: $\mathcal{M} \ni \kappa \mapsto \langle \kappa \mid 1 \rangle = \kappa[0,T]$.)

Lemma 4.10.1. *The adjoints* I_0^*, I_T^* *map* $\mathcal{M}[0,T]$ *into* $\mathrm{BV}[0,T] \subset L^1[0,T]$; *and they are given by*

$$(I_0^* \mu)(t) = \mu[t,T] \quad \text{and} \quad (I_T^* \mu)(t) = \mu[0,t] \quad \text{for a.e. } t, \tag{4.10.3}$$

for every $\mu \in \mathcal{M}[0,T]$. *If* $\mu[0,T] = 0$, *then* $-I_0^* \mu = \mu[0,\cdot] = I_T^* \mu$.

Proof. This follows from Fubini's Theorem: see the Proof of Lemma 3.12.1. □

Remark 4.10.2. *The operations I_0, I_T: $L^\infty \to \mathcal{C}$ are $\mathrm{m}\left(L^\infty, L^1\right)$-to-$\|\cdot\|_\infty$ continuous, where $\mathrm{m}\left(L^\infty, L^1\right)$ is the Mackey topology on L^∞ for the duality with L^1.*

Proof. For I_0, this follows from (4.10.1), the weak relative compactness of $\left\{1_{[0,t]} : t \in [0,T]\right\} \subset L^1$, and the characterisation of $\mathrm{m}\left(L^\infty, L^1\right)$ as the topology of uniform convergence on weak compacts of L^1. The case of I_T^* is similar. For more detail, see the Proof of Remark 3.12.2. □

Proof of Theorem 4.5.1 (Fixed-input value minimisation as the dual). Since (4.4.4)–(4.4.8) is an LP, it would suffice to apply results such as those of [4, 3.3 and 3.6–3.7]. However, to facilitate extensions requiring nonlinear models, this proof is couched in CP terms. The dual to a concave maximisation programme consists in minimising, over the dual variables (the Lagrange multipliers for the primal), the supremum of the Lagrange function over the primal decision variables: see, e.g., [73, (4.6) and (5.13)]. The "cone model" of [73, Example 4'] is applicable, since (4.10.2) and (4.4.6)–(4.4.7) represent the inequality constraints of the primal programme (4.4.4)–(4.4.8) by means of the nonnegative cones ($\mathcal{C}_+$ and L^∞_+) and convex constraint maps (which are actually linear).

The dual variables here are the κ_{St}, ν_{St}; κ_{Tu}, ν_{Tu}, ψ and λ of (4.5.3); and these are paired with the parameter increments Δk_{St}, $-\Delta n_{\mathrm{St}}$, Δk_{Tu}, $-\Delta n_{\mathrm{Tu}}$, Δe and $\Delta\zeta$ (as is discussed in Section 4.5). The primal variables are $(y, s_0) \in L^\infty \times \mathbb{R}$, and the Lagrange function is

$$\mathcal{L}(y, s_0; \kappa, \nu, \psi, \lambda) = \begin{cases} \Pi_{\mathrm{Exc}}(y, s_0; \kappa, \nu, \lambda) + V(\kappa, \psi) & \text{if } (\kappa, \nu) \geq 0 \text{ and } \\ & \psi = \lambda - I_0^*(\kappa_{\mathrm{St}} - \nu_{\mathrm{St}}) \\ +\infty & \text{otherwise} \end{cases} \tag{4.10.4}$$

where

$$V := \langle \kappa_{\mathrm{St}} \,|\, k_{\mathrm{St}} \rangle_{\mathcal{M},\mathcal{C}} + \langle \kappa_{\mathrm{Tu}} \,|\, k_{\mathrm{Tu}} \rangle_{L^{\infty*},L^\infty} + \langle \psi \,|\, e \rangle_{L^1,L^\infty} \tag{4.10.5}$$

and, with the notation

$$\mu_{\mathrm{St}} := \kappa_{\mathrm{St}} - \nu_{\mathrm{St}} \quad \text{and} \quad \mu_{\mathrm{Tu}} := \kappa_{\mathrm{Tu}} - \nu_{\mathrm{Tu}} \tag{4.10.6}$$

one has

$$\begin{aligned} \Pi_{\mathrm{Exc}} &:= \langle p - \mu_{\mathrm{Tu}} - \lambda + I_0^* \mu_{\mathrm{St}} \,|\, y \rangle - \langle \mu_{\mathrm{St}} \,|\, s_0 \rangle \\ &= \langle p - \mu_{\mathrm{Tu}} - \lambda + \mu_{\mathrm{St}}(\cdot, T] \,|\, y \rangle - s_0 \mu_{\mathrm{St}}[0, T] \end{aligned} \tag{4.10.7}$$

since $I_0^* \mu_{\mathrm{St}} = \mu_{\mathrm{St}}(\cdot, T]$ by Lemma 4.10.1.

Formulae (4.10.4)–(4.10.7) are interpreted below. But first, to complete the calculation of the dual minimand when $(\kappa, \nu) \geq 0$ and

$$\psi = \lambda - I_0^* \mu_{\mathrm{St}} \tag{4.10.8}$$

(which are dual constraints, since the minimand is $+\infty$ otherwise), note that

$$\sup_{y,s_0} \mathcal{L} = V + \sup_{y,s_0} \Pi_{\mathrm{Exc}} \tag{4.10.9}$$

since V is independent of (y, s_0). By (4.10.7), Π_{Exc} is linear in these variables, so its supremum is either 0 or $+\infty$; and it is zero if and only if $\partial \Pi_{\mathrm{Exc}} / \partial s_0 = 0$ and $\nabla_y \Pi_{\mathrm{Exc}} = 0$. These conditions are equivalent to the conjunction of (4.5.5) and

$$p = \lambda + \mu_{\mathrm{St}} [0, \cdot] + \mu_{\mathrm{Tu}}. \tag{4.10.10}$$

In view of (4.5.5) and Lemma 4.10.1, (4.10.10) with (4.10.8) are the same as (4.5.6)–(4.5.7). So the dual programme is: given $(p; k, e)$, minimise the $V(\kappa, \psi; k, e)$ of (4.10.5) over $(\kappa, \nu) \geq 0$, ψ and λ subject to (4.10.5)–(4.5.7). □

Comments:

- In (4.10.5)–(4.10.7), V is the value of the *available* resources (k, e), priced at (κ, ψ).
- For an entrepreneur buying all the inputs, Π_{Exc} is the excess profit (a.k.a. pure profit) from an output y and the use of an inflow e and an initial stock s_0. To see this, recall from (4.10.4) that $0 = \langle \lambda - \psi - I_0^* \mu_{\mathrm{St}} \,|\, e \rangle$, add this to (4.10.7) and use the identities $f(t) = y(t) - e(t)$ and $s(t) = s_0 - I_0 f(t)$ to obtain that

$$\Pi_{\mathrm{Exc}} = \langle p \,|\, y \rangle - \langle \kappa_{\mathrm{Tu}} - \nu_{\mathrm{Tu}} \,|\, y \rangle - \langle \kappa_{\mathrm{St}} - \nu_{\mathrm{St}} \,|\, s \rangle - \lambda \langle 1 \,|\, f \rangle - \langle \psi \,|\, e \rangle . \tag{4.10.11}$$

  This sum is the total over the cycle of the revenue from sales to the market minus the cost of all the resources needed at each time t. The resources in question are: the time-varying minimum requirements for the turbine and reservoir capacities (priced at κ), the floors for generation and stock (priced at ν), the required top-up (priced at λ), and the river inflow (priced at ψ). The last term in (4.10.11) can be rewritten as $\int_0^T \psi(t) e(t) \, \mathrm{d}t$, since $\psi \in L^1$ by (4.5.7).
- By adding and subtracting the value of internal sales (of the outflow y from reservoir to turbine, priced at ψ), (4.10.11) can be restated as

$$\Pi_{\mathrm{Exc}} = \langle p \,|\, y \rangle - \langle \mu_{\mathrm{Tu}} \,|\, y \rangle - \langle \psi \,|\, y \rangle + \langle \psi \,|\, y - e \rangle - \langle \mu_{\mathrm{St}} \,|\, s_0 - I_0 (y - e) \rangle - \langle \lambda \,|\, y - e \rangle .$$

  This gives Π_{Exc} as the sum of pure profits from the two parts of the plant: the first three terms add up to the excess profit from generation alone, whilst the other three terms add up to the excess profit from storage. The latter sum is equal to the appreciation of s_0 over the cycle because, with $\lambda - \psi = I_0^* \mu_{\mathrm{St}}$ and $f := y - e$ as per (4.2.1),

$$\langle \psi \,|\, f \rangle - \langle \lambda \,|\, f \rangle - \langle \mu_{\mathrm{St}} \,|\, s_0 - I_0 (f) \rangle = - \langle I_0^* \mu_{\mathrm{St}} \,|\, f \rangle - \langle \mu_{\mathrm{St}} \,|\, s_0 - I_0 (f) \rangle = -s_0 \langle \mu_{\mathrm{St}} \,|\, 1 \rangle .$$

Proof of Remark 4.5.2. If this were false, then the minimand's value could be decreased by replacing $(\kappa_{\rm St}, \nu_{\rm St}; \kappa_{\rm Tu}, \nu_{\rm Tu})$ with $(\mu^+_{\rm St}, \mu^-_{\rm St}; \mu^+_{\rm Tu}, \mu^-_{\rm Tu})$ given by (4.10.6). □

Proof of Proposition 4.6.1 (Dual solubility and optimality conditions). Like that of Theorem 4.5.1, this proof is put in CP terms. Consider first the dual problem with $L^{\infty *}$, instead of L^1, as the range for ψ, $\kappa_{\rm Tu}$ and $\nu_{\rm Tu}$ in (4.5.3). Since the nonnegative cones in the (primal) parameter spaces ($\mathcal{C}_+$ and L^∞_+) have nonempty interiors (for the supremum norm), the framework of [73, Examples 4, 4', 4"] is applicable. To verify the Generalised Slater's Condition of [73, (8.12)] for the primal constraints (4.4.6)–(4.4.8), it suffices to take $y = e$ (so that $f = 0$), setting s_0 at any value strictly between 0 and $k_{\rm St}$. So the dual has a (proper) solution, and the primal and dual values are equal (and finite): see, e.g., [73, Theorems 18 (a) and 17 (a)].

To complete the proof of Part 1, it remains to show that ψ, $\kappa_{\rm Tu}$ and $\nu_{\rm Tu}$ are in L^1. For ψ this is obvious from (4.5.7). Next, from the Hewitt-Yosida decomposition of (4.5.6) one has $\kappa^{\rm FA}_{\rm Tu} - \nu^{\rm FA}_{\rm Tu} = p_{\rm FA} = 0$, where $p_{\rm FA}$ means the purely finitely additive part of p: see, e.g., [10, Appendix I: (26)–(27)]. Given (4.5.8), this means that $\kappa^{\rm FA}_{\rm Tu} = 0 = \nu^{\rm FA}_{\rm Tu}$, as required. (That $\nu^{\rm FA}_{\rm Tu} = 0$ follows also from $p \geq 0$ alone: (4.5.8) and the Hewitt-Yosida decomposition of (4.5.6) give $\nu^{\rm FA}_{\rm Tu} = p^-_{\rm FA} = 0$, as well as $\kappa^{\rm FA}_{\rm Tu} = p^+_{\rm FA} = p_{\rm FA}$.)

For Part 2, apply the Kuhn-Tucker saddle-point characterisation of optima—given in, e.g., [73, Theorem 1 (e) and (f)]—to the primal (4.4.4)–(4.4.8) and its dual (4.5.1)–(4.5.7). This shows that (y, s_0) and $(\kappa, \nu, \psi, \lambda)$ is a dual pair of solutions if and only if they maximise and minimise (respectively) the Lagrange function $\mathcal{L}$ given by (4.10.4). The minimum in question is characterised by: nonnegativity (4.5.4) and compatibility (4.5.7) of dual variables, primal feasibility (4.4.6)–(4.4.8) and complementary slackness, which translates here into Conditions 2b and 2c. As for the maximum in question, it is characterised by the conditions $\partial \Pi_{\rm Exc}/\partial s_0 = 0$ and $\nabla_y \Pi_{\rm Exc} = 0$, i.e., by (4.5.5)–(4.5.6). □

Comment: Existence of a dual optimum in the *norm*-dual spaces ($\kappa_{\rm St}$ and $\nu_{\rm St}$ in $\mathcal{M} = \mathcal{C}^*$, and $\kappa_{\rm Tu}$, $\nu_{\rm Tu}$ and ψ in $L^{\infty *}$) comes automatically from (4.4.11), which ensures that the Generalised Slater's Condition of [73, (8.12)] holds with the norm topologies of the primal parameter spaces L^∞ and $\mathcal{C}$. The density representation (of the dual variables other than $\kappa_{\rm St}$ and $\nu_{\rm St}$) comes from the problem's structure and the assumptions on p: by the constraint (4.5.7), $\psi \in {\rm BV} \subset L^1$; with $p \geq 0$, every optimal $\nu_{\rm Tu}$ is in L^1; and if $p \in L^1$ then every optimal $\kappa_{\rm Tu}$ is also in L^1.

Proof of Proposition 4.7.1 (Shadow pricing of water as the dual). This is a reformulation of Theorem 4.5.1: substitute the ψ given by (4.5.7) into (4.5.6), and note that, given any ψ (and p), the best choice for $\kappa_{\rm St}$ and $\kappa_{\rm Tu}$ is as in (4.7.3)–(4.7.4), because $k_{\rm St} > 0$ and $k_{\rm Tu} > 0$. This reduces the dual programme (4.5.1)–(4.5.7) to minimisation of

$$k_{\mathrm{St}} \int_{[0,T]} (\mathrm{d}\psi(t))^{+} + k_{\mathrm{Tu}} \int_0^T \kappa_{\mathrm{Tu}}(t)\,\mathrm{d}t + \int_0^T \psi(t)\,e(t)\,\mathrm{d}t$$

over $\psi \in \mathrm{BV}[0-, T+]$, subject to $\psi(0-) = \psi(T+)$ lying between $\psi(0+)$ and $\psi(T-)$. Hence the first of the integrals equals the sum of $(\psi(0+) - \psi(T-))^{+}$ and $\int_{(0,T)} (\mathrm{d}\psi)^{+}$; and this sum is $\mathrm{Var}_{\mathrm{c}}^{+}(\psi)$. □

Proof of Proposition 4.8.1 (Primal solubility). With $p \in L^1$, the maximand of (4.4.5) is continuous for the weak* topology $\mathrm{w}(L^\infty, L^1)$. The feasible set is bounded: in y by (4.4.6), and in s_0 by (4.4.8) with, e.g., $t = 0$. So, being also weakly* closed, the feasible set is compact by the Banach-Alaoglu Theorem. And it is nonempty, since the point $(y, s_0) = (e, 0)$ is feasible by assumption. So an optimum exists by Weierstrass's Extreme Value Theorem. □

At this stage, it is useful to introduce a notation for the sets of those times when the reservoir is empty or full or neither, given a hydro output y meeting the balance constraint $\int_0^T f(t)\,\mathrm{d}t = 0$. These sets (which have already appeared in Condition 2b of Proposition 4.6.1) are:

$$E(f) := \{t \in [0,T] : s(t) = 0\} \tag{4.10.12}$$

$$F(f, k_{\mathrm{St}}) := \{t \in [0,T] : s(t) = k_{\mathrm{St}}\} \tag{4.10.13}$$

$$B(f, k_{\mathrm{St}}) := [0,T] \setminus (E \cup F) = \{t : 0 < s(t) < k_{\mathrm{St}}\} \tag{4.10.14}$$

where $s(t)$ is given by (4.4.9)–(4.4.10) in terms of $f := y - e$, and $k_{\mathrm{St}} \geq \mathrm{Max}(s)$. Since $s(0) = s(T)$, 0 and T are either both in B, or both in E, or both in F. From (4.4.9), $E \neq \emptyset$. Unless there is spare reservoir capacity, $F \neq \emptyset$ also; and then all three sets are nonempty. Their connected components are subintervals of $[0,T]$; and, being open, B is the union of a countable (finite or denumerable) sequence of intervals. Those *not* containing 0 or T are denoted by

$$A_m = (\underline{t}_m, \bar{t}_m) \neq \emptyset$$

for $m = 1, \ldots, M \leq \infty$, where $0 \leq \underline{t}_m < \bar{t}_m \leq T$. If $\{0, T\} \subseteq B$, then B additionally contains two subintervals whose union is

$$A_0 = (\underline{t}_0, T] \cup [0, \bar{t}_0)$$

for some $0 < \bar{t}_0 < \underline{t}_0 < T$. When $0, T \notin B$, set for completeness $\underline{t}_0 = T$ and $\bar{t}_0 = 0$, so that $A_0 = \emptyset$ in this case. In either case $B = \bigcup_{m \geq 0} A_m$.

All these sets may be thought of as subsets of the circle that results from "gluing" 0 and T into a single point $T0$. Then $(A_m)_{m \geq 0}$ are the *component arcs* of B (a.k.a. B-arcs); A_0 is that arc which contains $T0$ (if $T0 \in B$); and $\underline{t}_m$ and $\bar{t}_m$ are the beginning and the end of arc A_m (w.r.t. the "clockwise" orientation).

The formula for the output y, in terms of any $\psi \in \hat{\Psi}$, is proved next. On $\{t : p \neq \psi\}$, the optimal y equals unambiguously k_{Tu} or 0. Uniqueness of y on $\{p = \psi\}$ comes from the no-plateau assumption

(4.8.1) on p: this ensures that $\{p = \psi\} \subseteq E \cup F$, up to a null set. And at each $t \in E \cup F$ one has $f(t) = -\dot{s}(t) = 0$ (and hence $y(t) = e(t)$), since s has an extremum at t.

Remark 4.10.3. *If s: $[0, T] \to [0, 1]$ is absolutely continuous, then $\dot{s} = 0$ almost everywhere on the set $E := \{t \in [0, T] : s(t) = 0\}$.*

Proof of Proposition 4.8.2 (Hydro output with plateau-less price). Take any $y \in \hat{Y}$ (not yet known to be unique) and any $\psi \in \hat{\Psi}$ (which may be nonunique, unless $p \in \mathcal{C}$). The first and the third lines of (4.8.2) follow from Part 2c of Proposition 4.6.1 with (4.5.6)–(4.5.7). It remains to show that $y = e$ a.e. on $S := \{t : p = \psi\}$. For each m, one has $\psi = \text{const.}$ on $A_m(f, k_{\text{St}})$ by Part 2b of Proposition 4.6.1. Therefore $\text{meas}(S \cap A_m) = 0$ by (4.8.1), and hence $\text{meas}(S \cap B(f, k_{\text{St}})) = 0$ by countable additivity. This means that S is, up to a null set, contained in the set $F(f, k_{\text{St}}) \cup E(f)$, on which $y - e = -\dot{s} = 0$ a.e. (by Remark 4.10.3). This completes the proof of (4.8.2). It follows that $\hat{Y}$ is a singleton, even when $\hat{\Psi}$ is not. (Given any $\psi \in \hat{\Psi}$, any y' and y'' from $\hat{Y}$ satisfy (4.8.2) and are therefore equal.) $\square$

Proof of Corollary 4.9.1 (Dual calculation of SR profit). Formula (4.9.1) follows from Propositions 4.6.1 and 4.7.1. To derive it term-by-term, use the optimality conditions (complementary slackness and feasibility) to expand $\langle p \mid y \rangle$:

$$\begin{aligned}\Pi &:= \int_0^T p(t)\, y(t)\, \mathrm{d}t = \int_0^T (p(t) - \psi(t))\, y(t)\, \mathrm{d}t + \int_0^T \psi(t)\, e(t)\, \mathrm{d}t + \int_0^T \psi(t)\,(y(t) - e(t))\, \mathrm{d}t \\ &= k_{\text{Tu}} \int_0^T (p(t) - \psi(t))^+\, \mathrm{d}t + \int_0^T \psi(t)\, e(t)\, \mathrm{d}t - \int_0^T \psi(t)\, \frac{\mathrm{d}s(t)}{\mathrm{d}t}\, \mathrm{d}t\end{aligned}$$

integrating the last term by parts to obtain

$$\begin{aligned}-\int_0^T \psi(t)\, \mathrm{d}s(t) &= -[\psi s]_{t=0-}^{t=T+} + \int_{[0,T]} s(t)\, \mathrm{d}\psi(t) = s(0)\,(\psi(0-) - \psi(T+)) + k_{\text{St}} \int_{[0,T]} (\mathrm{d}\psi(t))^+ \\ &= 0 + k_{\text{St}}\, \mathrm{Var}_{\mathrm{c}}^{+}(\psi)\end{aligned}$$

as required. $\square$

Before a detailed proof of Lemma 4.9.2, it is worth presenting the main ideas. The key principle is that equipment can earn a rent only at a time of full capacity utilisation. In the present context this means that p can exceed ψ only when the turbine is working at full power (i.e., when $y(t) = k_{\text{Tu}}$). Similarly, ψ can exceed p only when the turbine is off (i.e., when $y(t) = 0$). Therefore $\psi(t)$ equals $p(t)$ when the reservoir is either full or empty (since $s(t) = 0$ or $s(t) = k_{\text{St}}$ implies that $y(t) = -\dot{s}(t) + e(t) = e(t)$, which lies *strictly* between 0 and k_{Tu} by assumption). By the same principle, ψ can be rising or falling only when the reservoir is full or empty (respectively); so ψ stays constant on each open interval $(\underline{t}, \overline{t})$ during which the reservoir constraints are inactive (i.e., $0 < s(t) < k_{\text{St}}$). Together, these conditions determine the function ψ *almost* completely—except for the possibility of jumps or drops of ψ that may occur at endpoints of a (closed) interval on which the reservoir is either full throughout or empty throughout.[21] Suppose, for example, that $\underline{t}$ is the end of an interval on which the reservoir is full. At that instant, ψ can jump but not drop; and the same is true of $p - \psi$ (since $p = \psi$ just before $\underline{t}$, and $p \geq \psi$ just after $\underline{t}$). So neither term, ψ or $p - \psi$, can jump at $\underline{t}$ if their sum (p) is continuous. This determines the constant value of ψ on $(\underline{t}, \bar{t})$ as $p(\underline{t})$; so ψ is unique.

Proof of Lemma 4.9.2 (Water price uniqueness and continuity). Fix any primal solution $y \in \hat{Y}(p, k, e)$, which exists by Proposition 4.6.1 (though it may be nonunique). To show that there is just one dual solution, every dual solution $\psi \in \hat{\Psi}(p, k, e)$ will be expressed by the same formula in terms of y.[22]

In the case of $F(y, k_{\mathrm{St}}) \neq \emptyset$, which is dealt with first, the Kuhn-Tucker Conditions will be used to show that any $\psi \in \hat{\Psi}(p, k, e)$ can be given, in terms of y, as

$$\psi(t) = p(t) \quad \text{for every } t \in (E \cup F)(f, k_{\mathrm{St}}) \setminus \{0, T\} \tag{4.10.15}$$

whereas on the m-th component A_m of $B(f, k_{\mathrm{St}})$, whose endpoints are $\underline{t}_m$ and $\bar{t}_m$, it is the constant

$$\psi(t) = \begin{cases} p(\underline{t}_m) & \text{if } \underline{t}_m \neq 0 \\ p(\bar{t}_m) & \text{if } \bar{t}_m \neq T \end{cases} \qquad \text{for every } t \in A_m(f, k_{\mathrm{St}}) \tag{4.10.16}$$

for each $m \geq 0$. Since both E and F are nonempty, $A_m \neq (0, T)$, so at least one line of (4.10.16) applies; and when both do, they are consistent. So (4.10.15)–(4.10.16) fully determine ψ on $(0, T)$, and hence on $[0, T]$ because $\psi(0)$ and $\psi(T)$ are defined by continuity.

To use the optimality conditions as stated in Proposition 4.6.1—i.e., in terms of $(\kappa, \nu, \psi, \lambda)$ rather than ψ alone—recall from Section 4.7 that if a $\psi \in \mathrm{BV}(0, T)$ solves (4.7.5)–(4.7.7), then (4.5.1)–(4.5.7) is solved by: the same ψ, $(\kappa_{\mathrm{Tu}}, \nu_{\mathrm{Tu}}) = \left((p - \psi)^+, (p - \psi)^-\right)$, any λ between $\psi(0+)$ and $\psi(T-)$ and $(\kappa_{\mathrm{St}}, \nu_{\mathrm{St}}) = \left(\mu_{\mathrm{St}}^+, \mu_{\mathrm{St}}^-\right)$, where $\mu_{\mathrm{St}} = \mathrm{d}\psi$ on $(0, T)$ with $\mu\{0\} = \psi(0+) - \lambda$ and $\mu\{T\} = \lambda - \psi(T-)$.

By (4.5.6)–(4.5.7),

$$p = \psi + \kappa_{\mathrm{Tu}} - \nu_{\mathrm{Tu}} = \lambda + (\kappa_{\mathrm{St}} - \nu_{\mathrm{St}})[0, \cdot] + \kappa_{\mathrm{Tu}} - \nu_{\mathrm{Tu}} \quad \text{a.e.} \tag{4.10.17}$$

It suffices to show that, at every point of $(E \cup F) \setminus \{0, T\}$, ψ is continuous and equal to p: then (4.10.16) follows, since ψ is constant on each B-component A_m, and since $A_m \neq (0, T)$.

[21] To simplify, it is assumed here that the times when the reservoir is full form a set F that consists of a finite number of intervals (which may be single instants, as in Example 4.13.1). Although F can be more complex, this is only a technicality (dealt with in the Proof of Lemma 4.9.2).

[22] The basis for this strategy (used also in proving Proposition 4.8.2) is that every dual solution supports every primal solution; i.e., the set of Kuhn-Tucker (saddle) points for a dual pair of convex programmes is the Cartesian product (of the primal and dual solution sets): see Proposition 4.6.1.

A discontinuity of ψ could only be a jump at a time when the reservoir is full, or a drop when it is empty. If $t \in F$ say, then, being full at t, the reservoir cannot be being emptied just before t.[23] That is, just before t the outflow y cannot exceed the inflow e, which, by assumption, is smaller than k_{Tu}. *A fortiori*, the capacity charge κ_{Tu} must be zero just before t. Similarly, just after a $t \in F$ the reservoir cannot be being filled, i.e., y cannot be less than e, which is positive by assumption; and so ν_{Tu} must be zero just after t. So $p - \psi = \kappa_{\mathrm{Tu}} - \nu_{\mathrm{Tu}}$ is nonpositive just before t and nonnegative just after t, and hence $p - \psi$ cannot drop at a $t \in F$. This means that any discontinuous changes in ψ and $p - \psi$ are of the same sign and cannot cancel each other out. So ψ (and $p - \psi$) must be continuous if p is. And it follows (from the signs of $p - \psi$ before and after t) that $p(t) = \psi(t)$. The "upside down" version of this reasoning applies to $t \in E$.

Since κ_{Tu} and ν_{Tu} are equivalence classes, this argument is formalised by using the essential limit concept—for which see, e.g., [20, IV.36–IV.37] or [81, II.9: p. 90]. It is also convenient to say that an inequality between functions holds *somewhere* on $A \subseteq [0,T]$ to mean that it holds on an $A' \subseteq A$ with $\operatorname{meas} A' > 0$ (i.e., it is *not* the case that the reverse inequality holds a.e. on A).

Recall from Section 4.4 that y with the $\underline{s}_{0,y}$ of (4.4.9) solve (4.4.4)–(4.4.8). Consider first a $t \in F \setminus \{0,T\}$. For every $\Delta t > 0$, it cannot be that $f > 0$ a.e. on $(t - \Delta t, t)$; i.e., somewhere on $(t - \Delta t, t)$ one has $y \leq e < k_{\mathrm{Tu}}$. Therefore $\kappa_{\mathrm{Tu}} = 0$ somewhere on $(t - \Delta t, t)$, by Part 2c of Proposition 4.6.1; and, as $\Delta t \to 0$, this shows that the lower left essential limit of κ_{Tu} at t is zero. Similarly, somewhere on $(t, t + \Delta t)$ one has $f \geq 0$, i.e., $y \geq e > 0$. Therefore $\nu_{\mathrm{Tu}} = 0$ somewhere on $(t, t + \Delta t)$. This means that the lower right essential limit of ν_{Tu} at t is zero; i.e.,

$$\operatorname{ess}\liminf_{\tau \searrow t} \nu_{\mathrm{Tu}}(\tau) = 0 = \operatorname{ess}\liminf_{\tau \nearrow t} \kappa_{\mathrm{Tu}}(\tau) \quad \text{for } t \in F \setminus \{0,T\}. \tag{4.10.18}$$

Given (4.10.17) as well as continuity of p and nonnegativity of κ_{Tu} and ν_{Tu}, it follows from (4.10.18) that[24]

$$\begin{aligned}
p(t) - \psi(t-) &= \operatorname{ess}\lim_{\tau \nearrow t} (\kappa_{\mathrm{Tu}} - \nu_{\mathrm{Tu}})(\tau) \\
&= \operatorname{ess}\liminf_{\tau \nearrow t} \kappa_{\mathrm{Tu}}(\tau) - \operatorname{ess}\liminf_{\tau \nearrow t} \nu_{\mathrm{Tu}}(\tau) \leq 0 && (4.10.19) \\
&\leq \operatorname{ess}\liminf_{\tau \searrow t} \kappa_{\mathrm{Tu}}(\tau) - \operatorname{ess}\liminf_{\tau \searrow t} \nu_{\mathrm{Tu}}(\tau) = \operatorname{ess}\lim_{\tau \searrow t} (\kappa_{\mathrm{Tu}} - \nu_{\mathrm{Tu}})(\tau) && (4.10.20) \\
&= p(t) - \psi(t+).
\end{aligned}$$

Therefore $\psi(t-) \geq \psi(t+)$ from a comparison of the first and the last sums. But also, since $t \in F$,

$$\psi(t-) \leq \psi(t+) \tag{4.10.21}$$

[23] This, by the way, is where the constancy of k_{St} over time is used.

[24] This argument uses also the fact that $\liminf(A - B) \leq \liminf A - \liminf B \leq \limsup(A - B)$ whenever the middle term is well defined. It equals $\lim(A - B)$ if the latter exists, as here (although the inequalities suffice). The same holds with $\limsup A - \limsup B$ as the middle term.

by Part 2b of Proposition 4.6.1; so all three inequalities (4.10.19), (4.10.20) and (4.10.21) must actually hold as equalities. This shows that $\psi(t-) = \psi(t+) = p(t)$, i.e., the two-sided limit of ψ at t exists and equals $p(t)$. (Since it exists, it also equals $\psi(t)$ because $\psi(t)$ always lies between $\psi(t-)$ and $\psi(t+)$.) The same can be shown for $t \in E$ (by the "upside down" version of the proof for $t \in F$); so

$$\psi(t) = \lim_{\tau \to t} \psi(\tau) = p(t) \quad \text{for } t \in (E \cup F) \setminus \{0, T\} \neq \emptyset. \tag{4.10.22}$$

Nonemptiness of this set follows from the assumption that $F \neq \emptyset$, since $E \neq \emptyset$ always, by (4.4.9).

By Part 2b of Proposition 4.6.1, ψ is constant on each A_m. This and (4.10.22) show that $\psi \in C(0,T)$. (Equivalently $\psi \in C[0,T]$, since $\psi(0) := \psi(0+)$ and $\psi(T) := \psi(T-)$.)

It remains to show that the proven properties of ψ imply (4.10.16). Since $E \cup F \not\subseteq \{0,T\}$, the set B consists of two or more nonempty components A_m. Each of these has at least one endpoint that is neither 0 nor T; i.e., $\underline{t}_m \neq 0$ or $\bar{t}_m \neq T$ ($\underline{t}_m \neq T$ and $\bar{t}_m \neq 0$ always). Say it is $\underline{t}_m$; then $\underline{t}_m \in (E \cup F) \setminus \{0,T\}$, since $\underline{t}_m \notin A_m$ (A_m is an open arc). So, by (4.10.22) and the constancy of ψ on A_m,

$$p(\underline{t}_m) = \psi(\underline{t}_m) = \psi(t) \quad \text{for every } t \in A_m. \tag{4.10.23}$$

If $T \neq \bar{t}_m$, then (4.10.23) holds with $\bar{t}_m$ in place of $\underline{t}_m$, by the same argument. This also shows that $p(\underline{t}_m) = p(\bar{t}_m)$ if both $\underline{t}_m \neq 0$ and $\bar{t}_m \neq T$. (All this applies to $m = 0$ as well, if $A_0 \neq \emptyset$. In this case ψ is additionally constant on $A_0 \supset \{0,T\}$; so $\psi(0) = \psi(T)$ even if $p(0) \neq p(T)$.) This completes the proof of (4.10.15)–(4.10.16) when $F \neq \emptyset$.

If $p(0) = p(T)$, then $\psi(0) = \psi(T)$ follows by virtually the same argument as that proving (4.10.22), with 0 and T thought of as a single point of the circle.

Finally, consider the case of $F(f, k_{\mathrm{St}}) = \emptyset$, which is trivial in that the reservoir is never used to capacity, and it earns no rent. Formally, $\kappa_{\mathrm{St}} = \nu_{\mathrm{St}} = 0$ by Part 2b of Proposition 4.6.1 and (4.5.5); so ψ is a constant. Its uniqueness is readily shown: ψ minimises (4.7.6) over $\mathrm{BV}(0,T)$, so, *a fortiori*, it minimises (4.7.6) over $\mathbb{R}$. Since for $\psi \in \mathbb{R}$ the sum (4.7.6) simplifies to

$$k_{\mathrm{Tu}} \int_0^T (p(t) - \psi)^+ \, \mathrm{d}t + \psi \int_0^T e(t) \, \mathrm{d}t$$

the minimum in question is characterised by the FOC

$$\operatorname{meas}\{t : p(t) > \psi\} \leq \frac{1}{k_{\mathrm{Tu}}} \int_0^T e(t) \, \mathrm{d}t \leq \operatorname{meas}\{t : p(t) \geq \psi\} \tag{4.10.24}$$

which means that ψ is an upper quantile of order $(1/Tk_{\mathrm{Tu}}) \int_0^T e(t) \, \mathrm{d}t$ for the distribution of p with respect to meas/T.[25] And the quantile is unique if $p \in C[0,T]$, since the cumulative distribution function of p is then strictly increasing on the interval $(\operatorname{Min}(p), \operatorname{Max}(p))$. □

[25] Note that $0 < \int_0^T e(t) \, \mathrm{d}t < Tk_{\mathrm{Tu}}$ by (4.4.11).

Comment: Although (4.10.18) suffices for the argument, both inf signs can be deleted, i.e., (4.10.18) can be strengthened to: $\kappa_{\mathrm{Tu}}(t-) = 0 = \nu_{\mathrm{Tu}}(t+)$ with $\nu_{\mathrm{Tu}}(t-) \geq 0$ and $\kappa_{\mathrm{Tu}}(t+) \geq 0$, for $t \in F \backslash \{0, T\}$, whenever $p(t\pm)$ exist.[26] This is because, by (4.5.8) and the continuity of $\kappa \mapsto \kappa^{\pm} \in \mathbb{R}_{+}$, the four limits exist and are equal to $(\kappa_{\mathrm{Tu}} - \nu_{\mathrm{Tu}})^{\pm}(t\pm) = (p - \psi)^{\pm}(t\pm)$. All four limits are zero if p is continuous at t.

Given Lemma 4.9.2, Theorem 4.9.3 is a routine case of the marginal interpretation of the dual solution. Before a formal proof, it is worth retracing in the present context the familiar argument which establishes the derivative property of the value function when differentiability is taken for granted. With the dual minimand (4.7.6) denoted by $V(k, e, \psi)$, the r.h.s.'s of (4.9.3)–(4.9.5) are obviously the partial derivatives of V in (k, e) evaluated at the dual optimum $\hat{\psi}(k, e)$. And the total derivatives, in (k, e), of the dual value $V\left(k, e, \hat{\psi}(k, e)\right)$ are equal to the corresponding partial derivatives, since the partial derivative of V in ψ vanishes by the FOC for the optimality of $\hat{\psi}$. To complete the calculation, note that the dual value equals the primal value $\Pi_{\mathrm{SR}}^{\mathrm{H}}$.[27] This is, indeed, the substance of the first step in the Proof of Theorem 4.9.3, except that a standard convex duality result is used instead of the above derivation "from first principles". This is necessary because a rigorous application of the chain rule would run into difficulties, since it would require the differentiability of $\hat{\psi}$ in (k, e), and of V in ψ. This would make their composition $\Pi(k, e) = V\left(k, e, \hat{\psi}(k, e)\right)$ differentiable, but neither this nor even the uniqueness of an optimal ψ (i.e., the existence of $\hat{\psi}$) may be presupposed. Rather, these properties must be derived—by using price continuity, since they are known to fail in general if $p \notin \mathcal{C}$ (Example 4.13.1). This gap is filled by Lemma 4.9.2.

Proof of Theorem 4.9.3 (Efficiency rents of a hydro plant). The first, routine, step is to identify the dual variables as marginal values of the primal parameters, with the marginal values formalised as supergradients (of the primal value, a concave function of the parameters): see, e.g., [73, Theorem 16: (b) and (a), with Theorem 15: (e) and (f)] or [51, 7.3: Theorem 1']. This is applied in such a way as to give the marginal interpretation to the optimal κ and ν themselves, rather than only to their totals over the cycle, although the formulae to be proved are for the total values. Therefore the short-run profit is considered as a function, $\widetilde{\Pi}_{\mathrm{SR}}^{\mathrm{H}}$, of all the quantity parameters

$$(\Delta k_{\mathrm{St}}, \Delta n_{\mathrm{St}}; \Delta k_{\mathrm{Tu}}, \Delta n_{\mathrm{Tu}}; \Delta e, \Delta \zeta) \in \mathcal{C} \times \mathcal{C} \times L^{\infty} \times L^{\infty} \times L^{\infty} \times \mathbb{R}$$

discussed in Section 4.5. It is an extension of the optimal value of the programme (4.4.4)–(4.4.8), i.e.,

$$\Pi_{\mathrm{SR}}^{\mathrm{H}}(p; k_{\mathrm{St}}, k_{\mathrm{Tu}}, e) = \widetilde{\Pi}_{\mathrm{SR}}^{\mathrm{H}}(p; k_{\mathrm{St}}, 0; k_{\mathrm{Tu}}, 0; e, 0) \quad \text{for } (k_{\mathrm{St}}, k_{\mathrm{Tu}}) \in \mathbb{R}^2$$

[26] The abbreviations $\kappa(t\pm)$ for the essential (one-sided) limits should not be mistaken for the ordinary limits of a particular variant of κ, in as much as the ordinary limits may be nonexistent.

[27] Conversely, the equality of short-run profit to the fixed-input value can be rederived from (4.9.3)–(4.9.5) by applying Euler's Theorem to Π as a jointly homogeneous function of (k, e).

where the scalars are identified with constant functions on $[0,T]$. In this setting, the result giving the marginal values of the primal parameters is

$$\widehat{\partial}_{k_{\mathrm{St}},n_{\mathrm{St}},k_{\mathrm{Tu}},n_{\mathrm{Tu}},e,\zeta}\widetilde{\Pi}^{\mathrm{H}}_{\mathrm{SR}} = \{(\kappa_{\mathrm{St}}, -\nu_{\mathrm{St}}, \kappa_{\mathrm{Tu}}, -\nu_{\mathrm{Tu}}, \psi, \lambda) : (\kappa, \nu, \psi, \lambda) \text{ meet Conditions 2a, 2b and 2c of Proposition 4.6.1}\} \quad (4.10.25)$$

where $\widehat{\partial}$ denotes the superdifferential (a.k.a. the supergradient set) for a concave function (so $\widehat{\partial}\Pi = -\partial(-\Pi)$, where ∂ is the subdifferential). For differentiation of $\Pi^{\mathrm{H}}_{\mathrm{SR}}$, with respect to the *constant* capacities and the cyclically varying inflow, it follows from (4.10.25) that

$$\begin{aligned}\widehat{\partial}_{k_{\mathrm{St}},k_{\mathrm{Tu}},e}\Pi^{\mathrm{H}}_{\mathrm{SR}} &= \left\{\left(\int_{[0,T]}\kappa_{\mathrm{St}}(\mathrm{d}t), \int_0^T \kappa_{\mathrm{Tu}}(t)\,\mathrm{d}t, \psi\right) : \exists\nu\ \exists\lambda\ (\kappa, -\nu, \psi, \lambda) \in \widehat{\partial}_{k,n,e,\zeta}\widetilde{\Pi}\right\} \\ &= \left\{\left(\mathrm{Var}^+_{\mathrm{c}}(\psi), \int_0^T (p-\psi)^+\,\mathrm{d}t, \psi\right) : \psi \in \hat{\Psi}(p; k_{\mathrm{St}}, k_{\mathrm{Tu}}, e)\right\} \end{aligned} \quad (4.10.26)$$

by using (4.7.4) and substituting $\kappa_{\mathrm{St}} = (\mathrm{d}\psi)^+$. When $p \in \mathcal{C}$, the set $\hat{\Psi}$ in (4.10.26) is actually a singleton by Lemma 4.9.2, and hence so is $\widehat{\partial}_{k,e}\Pi^{\mathrm{H}}_{\mathrm{SR}}(p; k, e)$. □

Comments: Some weaker results on the relationship of an optimal ψ to p are simpler to establish than (4.10.15)–(4.10.16), but such results are so weak as to be of little use by themselves. For example:

- When the number of B-arcs is finite, the equality $\psi = p$ a.e. on $F \cup E$ can be shown by the argument that $\hat{s} =$ const. and so $\hat{y} = e$ a.e. on each F-arc or E-arc R, so $\psi = p$ a.e. on R (and everywhere on $\operatorname{int} R$ if p is continuous, in which case it follows that $\psi = p$ on $F \cup E$, except possibly at the endpoints of F-arcs and E-arcs, the number of which is finite). But capacity valuation requires also the values of ψ on the B-arcs, and this necessitates the additional arguments in the Proof of Lemma 4.9.2.
- By using Remark 4.10.3, the equality $\psi = p$ a.e. on $F \cup E$ can be shown for every $p \in L^1_{++}$. But this may even be vacuous ($F \cup E$ may be a null set); and the stronger result (4.10.16) does depend on the continuity of p.

4.11 Case of infeasible coasting

With spillage assumed feasible as in (4.2.3), one can drop the condition that $e \le k_{\mathrm{Tu}}$. But with $e \not\le k_{\mathrm{Tu}}$, i.e., with coasting no longer feasible, an optimal water price ψ need not be continuous or unique (despite the continuity of the electricity price p).

For this extension, the primal problem (4.4.4)–(4.4.8) is modified by adding the spillage term, $\sigma \in L^\infty$, to the net outflow f, as in (4.2.1). The extra variable is constrained as in (4.2.3), i.e.,

$0 \le \sigma \le e$. There is, however, no real need for an extra Lagrange multiplier for the constraint $\sigma \ge 0$ because such a multiplier would turn out to be identical to ψ (at the dual optimum). The multiplier must be nonnegative; i.e., the constraint $\psi \ge 0$ must be adjoined to the dual (4.7.5)–(4.7.7).[28] The multiplier for the constraint $\sigma \le e$ turns out to be zero: the primal value is the same with or without this constraint.[29] This means that free disposal of water is effectively unlimited, as in [55, 1.4a].[30] Finally, an extra slackness condition, that $\psi = 0$ a.e. on $\{t : \sigma(t) > 0\}$, is adjoined to Part 2c of Proposition 4.6.1.

In the extended framework, one can formally prove that an optimal storage policy involves no spillage if $k_{\mathrm{Tu}} \ge e$ and $p \in L^1_{++}$. This can be shown either by establishing that $\psi \gg 0$, or directly as follows. Suppose contrarily that $\sigma > 0$ on a neighbourhood of some t. If $y(t) < k_{\mathrm{Tu}}(t)$ then the output can be increased around t, so (y, σ) is not optimal. If $y(t) = k_{\mathrm{Tu}}(t)$ then $\dot{s}(t) = (-y + e - \sigma)(t) \le 0 - \sigma(t) < 0$, i.e., the stock is falling around t, and so there is room to store a unit being spilt, to release it at the nearest opportunity (which will come, since $\sigma \ne 0$ implies that $y(\tau) < e(\tau) \le k_{\mathrm{Tu}}$ for some τ). Again, this shows that (y, σ) is not optimal. And although this argument handles y, e and σ as though they were continuous functions (rather than elements of L^∞), it can be made rigorous by choosing t to be a density point of the set $\{t : y(t) < k_{\mathrm{Tu}}\}$ or $\{t : y(t) = k_{\mathrm{Tu}}\}$, respectively.[31]

With $\mathrm{EssInf}(e) > 0$ (but without assuming that $e \le k_{\mathrm{Tu}}$), the modified primal and dual problems remain feasible, and the Kuhn-Tucker characterisation of optimality (Proposition 4.6.1) continues to hold.[32] If the inflow exceeds the turbine's capacity only on a relatively short interval, spillage is still avoided. Consider an inflow increment $(k_{\mathrm{Tu}} - e) + \Delta e$ on an interval $[\underline{t}, \bar{t}]$ on which the reservoir is full in the original solution, one which corresponds to an inflow $e < k_{\mathrm{Tu}}$. To make room for the excess inflow, an extra amount $\Delta E = \int_{\underline{t}}^{\bar{t}} \Delta e(t)\,\mathrm{d}t$ of water is discharged immediately before $\underline{t}$, with the turbine operating at full capacity to sell the extra output at best prices, as close to $p(\underline{t})$ as possible. This solution is supported by the stock price ψ that "freezes" when the discharge starts and stays constant until $\bar{t}$, when it jumps back to the original price trajectory (so ψ is discontinuous at $\bar{t}$). As ΔE increases, so the discharge period preceding $[\underline{t}, \bar{t}]$ starts earlier. Here, it is assume that it does not merge with an earlier water collection period (during which $p < \psi$) before ΔE reaches k_{St}.[33] In the borderline case of $\Delta E = k_{\mathrm{St}}$, the reservoir becomes empty at $\underline{t}$ and full again at $\bar{t}$.

[28]The constraint $\psi \ge 0$ is superfluous when $e < k_{\mathrm{Tu}}$ because in this case every solution, ψ, to (4.7.5)–(4.7.7) is nonnegative anyway.

[29]When $p \in L^1_+$, there is an optimum policy with $\sigma(t) \le (e(t) - k_{\mathrm{Tu}})^+ < e(t)$.

[30]In reality, the spillage rate is constrained—quite apart from the considerations of flood control, etc.—by spillway capacity (unless this is exceeded by $e - y$ at a time when the reservoir is full, in which case it automatically overflows "from the top").

[31]For the concept of a density point, see, e.g., [27, (5.8)].

[32]Verification of Slater's Condition now requires a different choice of a feasible policy, viz., any (y, σ) with $y + \sigma = e$ and $k_{\mathrm{Tu}} - \epsilon \ge y \ge \epsilon$ and $\sigma \ge \epsilon$, for some number $\epsilon > 0$.

[33]If the two do merge, then the two constant values of ψ become one value, which decreases as ΔE continues to

The no-spillage solution is still feasible, but only just; and the water price on $[\underline{t}, \bar{t}]$ is an arbitrary constant between 0 and $\psi(\underline{t})$.[34] In this case ψ is nonunique (and discontinuous). If ΔE is further increased (keeping $\underline{t}$ and $\bar{t}$ fixed), then a total of $\Delta E - k_{\mathrm{St}}$ must be spilt on $[\underline{t}, \bar{t}]$. This can be done in any way, but ψ is unique (though it is discontinuous at $\underline{t}$ and $\bar{t}$), since $\psi = 0$ on $[\underline{t}, \bar{t}]$.

4.12 Optimum investment in a hydro plant

The marginal capacity values $\nabla_k \Pi^{\mathrm{H}}_{\mathrm{SR}}$ can be used to determine the optimum investment into a hydro plant on the basis of a given TOU electricity tariff p, a given river inflow e and the supply costs of the two capital inputs, the reservoir and the turbine. The programme formulated next can be thought of in two ways: either as the case of small investment which will not significantly change the existing tariff, or as a subproblem of a general equilibrium system which determines the tariff as well. The turbine's unit cost, r_{Tu}, can be reasonably regarded as constant, i.e., independent of the capacity k_{Tu}. By contrast, the reservoir's marginal cost, r_{St}, typically increases with k_{St} because the most suitable parts of the site are developed first. In formal terms, on a potential hydro site, a reservoir can be built at a cost which is a convex and increasing function, G, of its capacity $k_{\mathrm{St}} \in [0, \bar{k}_{\mathrm{St}}]$, with $G(0) = 0$.

The investment problem is:

$$\text{Given } (p, e, r_{\mathrm{Tu}}) \in \mathcal{C}[0,T] \times L^{\infty}[0,T] \times \mathbb{R}_{++} \text{ (and given the function } G) \tag{4.12.1}$$

$$\text{maximise } \Pi^{\mathrm{H}}_{\mathrm{SR}}(p, k_{\mathrm{St}}, k_{\mathrm{Tu}}, e) - G(k_{\mathrm{St}}) - r_{\mathrm{Tu}} k_{\mathrm{Tu}} \text{ over } (k_{\mathrm{St}}, k_{\mathrm{Tu}}) \in \mathbb{R}^2_{+} \tag{4.12.2}$$

and the FOCs for an interior solution are:

$$\frac{\partial \Pi^{\mathrm{H}}_{\mathrm{SR}}}{\partial k_{\mathrm{St}}}(k_{\mathrm{St}}, k_{\mathrm{Tu}}) = \frac{\mathrm{d}G}{\mathrm{d}k_{\mathrm{St}}}(k_{\mathrm{St}}) \tag{4.12.3}$$

$$\frac{\partial \Pi^{\mathrm{H}}_{\mathrm{SR}}}{\partial k_{\mathrm{Tu}}}(k_{\mathrm{St}}, k_{\mathrm{Tu}}) = r_{\mathrm{Tu}}. \tag{4.12.4}$$

This system can be solved numerically by, e.g., a quasi-Newton method. A similar application to investment in pumped storage is presented in more detail in Chapter 3.

4.13 Indeterminacy of marginal values with discrete time

As is next shown by means of a two-period model, discretisation of time can make $\Pi^{\mathrm{H}}_{\mathrm{SR}}(p, k, e)$ nondifferentiable in k and e. This is because it forces p to be piecewise constant and thus discontinuous. In the following example, the inflow rate is assumed to be a constant e; and the cycle is divided into subperiods of lengths d and $T - d$. Then $\Pi(k, \mathrm{e})$ has a term which is proportional to

increase (and the water collection period shrinks).

[34] This indeterminacy is noted in [55, p. 226: last paragraph].

$\min\{k_{\mathrm{St}}, (T-d)(k_{\mathrm{Tu}}-\mathrm{e}), d\mathrm{e}\}$. It follows that, given e, an efficient choice of capacities must satisfy $k_{\mathrm{St}}/(k_{\mathrm{Tu}}-\mathrm{e}) = T-d$; and Π is obviously nondifferentiable at such a point (k,e). The case of $d=T/2$ is shown in Figure 4.3. With a varying d, even this two-period framework reveals that the ratio $k_{\mathrm{St}}/(k_{\mathrm{Tu}}-\mathrm{e})$ depends on the TOU price p, through the price duration d (although it is independent of the two price levels in p). This is why the example is not limited to the case of $d=T/2$, although it is this case that is shown in Figure 4.3.

Example 4.13.1 (Indeterminacy of marginal values with discontinuous price). *The short-run profit function of the hydro technique (4.2.3) can be nondifferentiable in (k,e). To see this, take any numbers $\overline{\mathrm{p}} > \underline{\mathrm{p}} \geq 0$ and e with $k_{\mathrm{Tu}} > \mathrm{e} > 0$, and any $d \in (0,T)$. Set a piecewise constant tariff*

$$p(t) := \begin{cases} \underline{\mathrm{p}} & \text{if } t < d \\ \overline{\mathrm{p}} & \text{if } t \geq d \end{cases}$$

and specify a constant inflow $e(t) = \mathrm{e}$ for $t \in [0,T]$. Then, for a hydro plant with capacities $k = (k_{\mathrm{St}}, k_{\mathrm{Tu}})$, a profit-maximising output is[35]

$$y(t) = \begin{cases} \mathrm{y}' := \mathrm{e} - \epsilon & \text{if } t < d \\ \mathrm{y}'' := \mathrm{e} + \epsilon\frac{d}{T-d} & \text{if } t \geq d \end{cases}$$

where

$$\epsilon := \frac{k_{\mathrm{St}}}{d} \wedge \frac{T-d}{d}(k_{\mathrm{Tu}}-\mathrm{e}) \wedge \mathrm{e} := \min\left\{\frac{k_{\mathrm{St}}}{d}, \frac{T-d}{d}(k_{\mathrm{Tu}}-\mathrm{e}), \mathrm{e}\right\}. \tag{4.13.1}$$

So

$$\Pi^{\mathrm{H}}_{\mathrm{SR}}(p,k,\mathrm{e}) = d\underline{\mathrm{p}}\mathrm{e} + (T-d)\overline{\mathrm{p}}\mathrm{e} + d(\overline{\mathrm{p}}-\underline{\mathrm{p}})\epsilon. \tag{4.13.2}$$

Therefore $\Pi^{\mathrm{H}}_{\mathrm{SR}}$ is nondifferentiable in (k,e) whenever the minimum defining ϵ in (4.13.1) is attained at more than one of the three terms. (See Figure 4.3 for the case of $d=T/2$ and $2k_{\mathrm{St}}/T = k_{\mathrm{Tu}}-\mathrm{e} < \mathrm{e}$.)

Comments:

- The superdifferential (the set of supergradients) of $\mathrm{Min}(k) := \min_{\phi\in\{1,2,3\}} k_\phi$, as a concave function of k, is

$$\hat{\partial}_k \mathrm{Min}(k) = \left\{(r_\phi)_{\phi=1}^3 \geq 0 : \sum_{\phi=1}^{3} r_\phi = 1 \text{ and } \forall\phi \ (r_\phi = 0 \text{ if } \exists\phi'\ k_{\phi'} < k_\phi)\right\} \tag{4.13.3}$$

[35] This y implements the policy of carrying over, from the low-price period to the high-price period, as much water as the constraints allow, viz., $\min\{k_{\mathrm{St}}, (T-d)(k_{\mathrm{Tu}}-\mathrm{e}), d\mathrm{e}\}$. It is optimal independently of the two price levels, as long as $\overline{\mathrm{p}} > \underline{\mathrm{p}}$. (Also, it is the only *two-valued* optimal output function; but in the class of all functions it is the unique optimum if and only if $k_{\mathrm{St}} \geq (T-d)(k_{\mathrm{Tu}}-\mathrm{e}) = d\mathrm{e}$.)

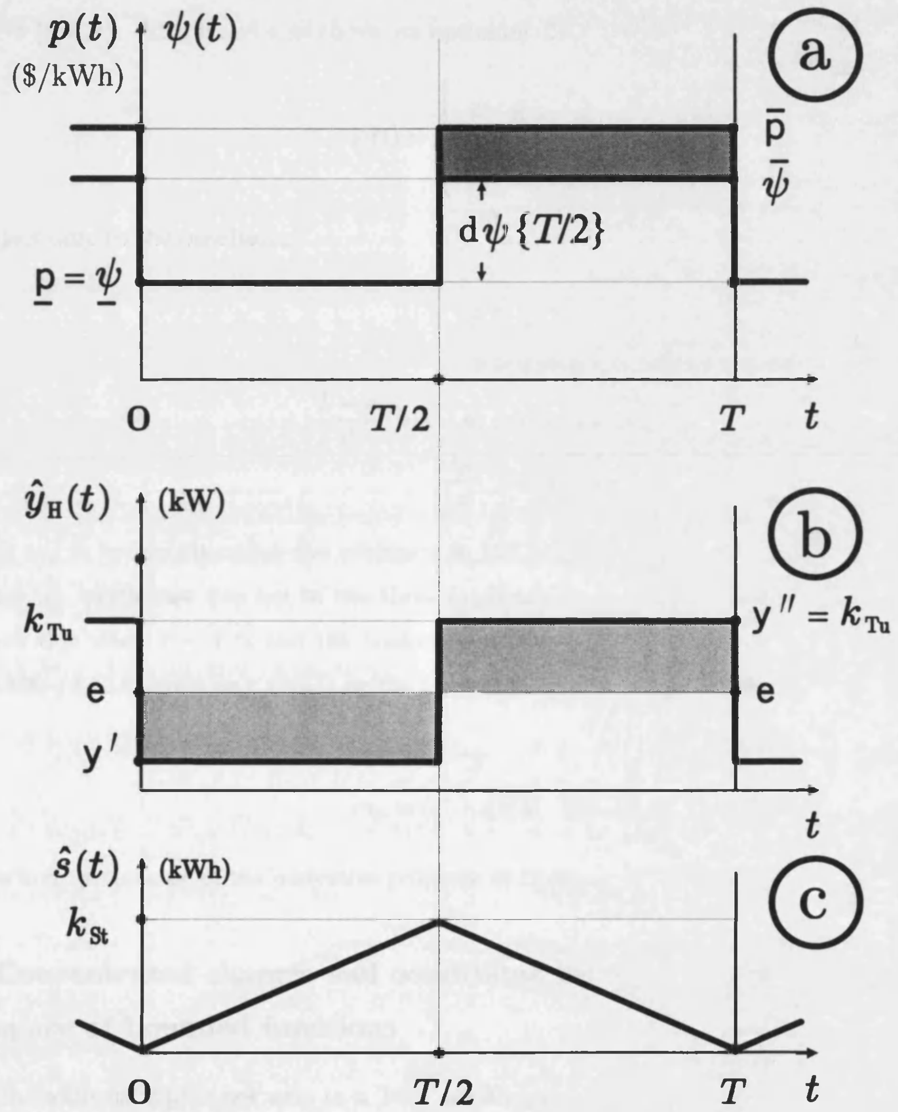

FIGURE 4.3. Indeterminacy of an optimal shadow price of water ψ (in the case of two equal subperiods). The jump $\overline{\psi} - \underline{\psi}$, which equals the reservoir's unit rent, can take any value not exceeding the jump of p. The dark grey area in (a) represents the turbine's unit rent. In (b), each of the light grey areas equals the reservoir's capacity k_{St}.

and hence, for this example, $\widehat{\partial}_{k,e}\Pi$ can be worked out directly from (4.13.1)–(4.13.2) by the chain rule.

- Each supergradient in $\widehat{\partial}_{k,e}\Pi$ can also be obtained from an optimal shadow price of water $\psi \in \hat{\Psi}(p,k,e)$. With p and e as above, an optimum shadow price is any

$$\psi(t) = \begin{cases} \underline{\psi} & \text{if } t < d \\ \overline{\psi} & \text{if } t \geq d \end{cases} \tag{4.13.4}$$

subject only to the conditions

$$\underline{\mathrm{p}} \leq \underline{\psi} \leq \overline{\psi} \leq \overline{\mathrm{p}} \tag{4.13.5}$$

$$\mathrm{e} > \epsilon \Rightarrow \underline{\mathrm{p}} = \underline{\psi} \tag{4.13.6}$$

$$\frac{T-d}{d}(k_{\mathrm{Tu}} - \mathrm{e}) > \epsilon \Rightarrow \overline{\psi} = \overline{\mathrm{p}} \tag{4.13.7}$$

$$\frac{k_{\mathrm{St}}}{d} > \epsilon \Rightarrow \underline{\psi} = \overline{\psi}. \tag{4.13.8}$$

Such a ψ is nonunique unless the minimum in (4.13.1) is attained at exactly one of the three terms (in which case two out of the three implications (4.13.6)–(4.13.8) apply). Figure 4.3 shows this when $d = T/2$ and the minimum is $2k_{\mathrm{St}}/T = k_{\mathrm{Tu}} - \mathrm{e} < \mathrm{e}$. Every ψ satisfying (4.13.4)–(4.13.8) gives an $r \in \widehat{\partial}_k\Pi$ by the formulae

$$r_{\mathrm{St}} = \overline{\psi} - \underline{\psi} \tag{4.13.9}$$

$$r_{\mathrm{Tu}} = (T-d)\left(\overline{\mathrm{p}} - \overline{\psi}\right). \tag{4.13.10}$$

This is a special case of the derivative property of Π stated in (4.10.26).

4.14 Concentrated charges and conditions for equilibrium in commodity space of bounded functions

Much of the analysis applies not only to a TOU tariff represented by a price function $p \in L^1[0,T]$ but also to a tariff represented by a p in $L^{\infty *}[0,T]$, the norm-dual of L^∞. Such a p can be identified with a finitely additive set function vanishing on meas-null sets, since the integral of a $y \in L^\infty$ w.r.t. such a set function defines a bounded linear functional on L^∞: see, e.g., [86, 2.3]. As an additive set function, a $p \in L^{\infty *}$ has the Hewitt-Yosida decomposition into $p_{\mathrm{CA}} + p_{\mathrm{FA}}$, the sum of its countably additive (c.a.) and purely finitely additive (p.f.a.) parts: see, e.g., [10, Appendix I: (26)–(27)] or [86, 1.23 and 1.24].[36] The c.a. part of p is identified with its density w.r.t. meas (which exists by the Radon-Nikodym Theorem); so it is a price function $p_{\mathrm{CA}} \in L^1[0,T]$. The p.f.a. part of p can be

[36] A p.f.a. set function is one that is lattice-disjoint from every c.a. one.

characterised as a singular element of $L^{\infty *}[0,T]$, i.e., p_{FA} is concentrated on a subset of $[0,T]$ with an arbitrarily small Lebesgue measure. (Formally, a $p \in L^{\infty *}$ is *concentrated on*, or *supported by*, a measurable set S if $\langle p \,|\, y \rangle = \langle p \,|\, y 1_S \rangle$ for every $y \in L^{\infty}$. A sequence of sets (S_m) is *evanescent* if $S_{m+1} \subseteq S_m$ for every m and $\mathrm{meas}(\bigcap_{m=1}^{\infty} S_m) = 0$; and p is called *singular* if there exists an evanescent sequence (S_m) such that p is concentrated on S_m for each m. A $p \in L^{\infty *}$ is singular if and only if it is p.f.a.: see [86, 3.1].) This gives p_{FA} the interpretation of an extremely concentrated charge. In the hydro context it can arise as a turbine capacity charge (Remark 4.14.1).

The assumption needed for $\nabla_{k,e}\Pi^{\mathrm{H}}_{\mathrm{SR}}$ to exist is next weakened to: $p_{\mathrm{CA}} \in \mathcal{C}$. That is, the density part of p is required to be continuous on $[0,T]$, but p may also have a nonzero p.f.a. part. If $p_{\mathrm{FA}} > 0$, it can be interpreted as the "concentrated" part of turbine capacity charges (since $p_{\mathrm{FA}} = \kappa^{\mathrm{FA}}_{\mathrm{Tu}}$ at every dual optimum, from (4.5.6) and (4.5.8)). Unless demand for electricity is interruptible, such a charge can arise in general equilibrium, and it has a tractable mathematical representation by a singular measure (such as a point measure) if the consumption and output rates are continuous over time: see [39, Example 3.1]. Out of equilibrium, the presence of a nonzero p_{FA} can result in nonexistence of an optimum solution y to the primal (4.4.1)–(4.4.3): see Case (b) in Part 4 below. Except for this, the preceding analysis extends *mutatis mutandis* to the case of a $p \in L^{\infty *}_{++}$, as is spelt out next.

Remark 4.14.1 (Concentrated charges). *For every $p \in L^{\infty *}_{+}$ with $p_{\mathrm{CA}} \gg 0$:*

1. *The dual problem of water pricing is (4.7.6)–(4.7.7) with p_{CA} instead of p and with $k_{\mathrm{Tu}} \|p_{\mathrm{FA}}\|^{*}_{\infty}$ added to the minimand (4.7.6).*[37] *Since the extra term is a constant (i.e., is independent of ψ), its addition does not change the solution set, i.e., $\hat{\Psi}(p,k,e) = \hat{\Psi}(p_{\mathrm{CA}},k,e)$.*

2. *Formulae (4.9.3)–(4.9.5), which give $\nabla_{k,e}\Pi^{\mathrm{H}}_{\mathrm{SR}}$ in terms of p and $\hat{\psi}$, hold with p replaced by p_{CA} on their r.h.s.'s and with $\|p_{\mathrm{FA}}\|^{*}_{\infty} = \langle p_{\mathrm{FA}} \,|\, 1 \rangle$ added to the r.h.s. of (4.9.4).*[38]

3. *The Kuhn-Tucker Conditions 2a–2c of Proposition 4.6.1 imply the same but with p_{CA} in place of p and with $\kappa^{\mathrm{CA}}_{\mathrm{Tu}}$ in place of κ_{Tu}. (The converse is obviously false.) It follows that, under Slater's Condition (4.4.11),*

$$\hat{Y}(p,k,e) \subseteq \hat{Y}(p_{\mathrm{CA}},k,e)$$

i.e., if p supports y as a short-run profit maximum, then so does p_{CA} (or, put formally, if y solves (4.4.2)–(4.4.3), then it also solves (4.4.2)–(4.4.3) with p_{CA} in place of p). So the

[37] This also points to a case of the primal value being strictly less—it is never greater—than the dual value. This is when $p_{\mathrm{FA}} > 0$, $p_{\mathrm{CA}} \in \mathrm{BV}$ and $k_{\mathrm{Tu}} > \mathrm{Sup}(e) \geq \mathrm{Inf}(e) > 0$ but $k_{\mathrm{St}} = 0$: the short-run profit is then $\langle p \,|\, e \rangle$, but the fixed-input value (as found from the dual) is $\int_0^T p_{\mathrm{CA}} e \, dt + k_{\mathrm{Tu}} \|p_{\mathrm{FA}}\|^{*}_{\infty} > \langle p \,|\, e \rangle$, since the dual solution is $\hat{\psi} = p_{\mathrm{CA}}$, and since $k_{\mathrm{Tu}} > \mathrm{Sup}(e)$. When $k_{\mathrm{St}} > 0$, the primal and dual values are of course equal.

[38] To see this, note that, by Part 1 and the equality of the dual and primal values, $\Pi^{\mathrm{H}}_{\mathrm{SR}}(p) - \Pi^{\mathrm{H}}_{\mathrm{SR}}(p_{\mathrm{CA}}) = k_{\mathrm{Tu}} \|p_{\mathrm{FA}}\|^{*}_{\infty}$.

conclusions about any optimal output y, such as (4.8.2), hold also with p_{CA} *in place of* p. *Such results are of course vacuous when, at* p, *there is no optimal* y.

4. *The timing of a* $p_{FA} > 0$ *matters for the existence of an optimal output. Consider the cases in which such a term is concentrated on each neighbourhood of: either (a) a peak* $\bar{t}$, *or (b) a trough* $\underline{t}$, *of a piecewise monotone* $p_{CA} \in C^{c}_{++}$. *Under Slater's Condition (4.4.11), one has* $\hat{y}(p_{CA}) = k_{Tu}$ *around* $\bar{t}$, *and* $\hat{y}(p_{CA}) = 0$ *around* $\underline{t}$. *At* $p = p_{CA} + p_{FA}$, *one has* $\hat{y}(p) = \hat{y}(p_{CA})$ *in Case (a), but* $\hat{Y}(p) = \emptyset$ *in Case (b).*[39]

To ensure that the hydro technology can be included in an Arrow-Debreu model of general equilibrium with $L^{\infty}[0,T]$ and $L^{1}[0,T]$ as the commodity and price spaces, two conditions have to be verified. The first one is needed for the existence of a price system in the larger price space $L^{\infty *}$.

Lemma 4.14.2. *The set* $\mathbb{Y}_{H}$ *is* $\mathrm{w}(L^{\infty}, L^{1})$*-closed.*

Proof. By the Krein-Smulian Theorem (for which see, e.g., [32, 18E]), it suffices to show that $\mathbb{Y}_{H}$ is closed for the bounded weak* topology of L^{∞}. Since the bound on k_{Tu} bounds y as well, it suffices to establish that, for each $\bar{k} = (\bar{k}_{St}, \bar{k}_{Tu}) \in \mathbb{R}^{2}_{+}$ and $\bar{e} \in L^{\infty}_{+}$, the set

$$\mathbb{Y}_{H} \cap \{(y, -k, -e) : k \leq \bar{k}, e \leq \bar{e}\}$$

is weakly* compact. This set is the image, $\pi(S)$, of the set S of all those points $(y, -k, -e; s_0, \sigma)$ that meet the conditions: $k \leq \bar{k}$, $e \leq \bar{e}$, $\sigma \in [0, e]$ and (4.4.6)–(4.4.8) with $f = y - e + \sigma$, under the projection map π that sends such a point to $(y, -k, -e)$. And $\pi(S)$ is weakly* compact because π is weak*-to-weak* continuous, and because S is weakly* compact by the Banach-Alaoglu Theorem. □

To ensure the existence of a price system in the smaller price space L^{1}, one needs to verify the Exclusion Condition of [43].[40] This is facilitated by the use of an input requirement function.

Lemma 4.14.3. *The set* $\mathbb{Y}_{H}$ *meets the Exclusion Condition of [43].*

Proof. This follows from Mackey continuity[41] of the function $\check{k}_{St}\colon L^{\infty} \to \mathbb{R}$ defined by

$$\check{k}_{St}(f) := \mathrm{Max}(I_0 f) + \mathrm{Max}(I_T f) \tag{4.14.1}$$

[39] In heuristic terms, this is because in Case (b) the extra price term requires a brief switch from filling ($y = 0$) to emptying ($y = k_{Tu}$) around $\underline{t}$—the briefer the better, so no storage policy is best. For a formal proof, compare the increments to Π^{H}_{SR} and to the value of the output $\hat{y}(p_{CA})$ that result from adding p_{FA}: in Case (a) both are equal (to $k_{Tu}\|p_{FA}\| > 0$), so $\hat{y}(p_{CA})$ remains optimal, i.e., $\hat{y}(p) = \hat{y}(p_{CA})$. But in Case (b), the profit increases by $k_{Tu}\|p_{FA}\| > 0$ by Part 2, whereas the output's value stays the same; so there is no optimum at p (since $\hat{y}(p_{CA})$ is the only possibility, by Part 3).

[40] This is less restrictive than Bewley's Exclusion Assumption in [10].

[41] Upper semicontinuity is what is relevant here.

which is the *storage capacity requirement* (when f is the net outflow from the reservoir): see (3.16.1) for a derivation. To verify the Exclusion Condition, take any $(p, r, \psi) \in L^{\infty *} \times \mathbb{R}^2 \times L^{\infty *}$ and an evanescent sequence of measurable sets $S_m \subset [0, T]$ supporting both p_{FA} and ψ_{FA} (so $\operatorname{meas} S_m \to 0$ as $m \to \infty$). Take any $(y, -k, -e) \in \mathbb{Y}_{\mathrm{H}}$; i.e., $y \in [0, k_{\mathrm{Tu}}]$ and there exists a $\sigma \in [0, e]$ such that

$$\int_0^T f(t)\,\mathrm{d}t = 0 \quad \text{and} \quad \check{k}_{\mathrm{St}}(f) \le k_{\mathrm{St}}$$

where $f := y - e + \sigma$. As can readily be shown, there is a sequence $Z_m \supseteq S_m$ with $\operatorname{meas} Z_m \to 0$ and $\int_{Z_m} f(t)\,\mathrm{d}t = 0$. Define $y^m := y 1_{[0,T] \setminus Z_m}$ and $e^m := e 1_{[0,T] \setminus Z_m}$ and $\sigma^m := \sigma 1_{[0,T] \setminus Z_m}$; then

$$f^m := y^m - e^m + \sigma^m = f 1_{[0,T] \setminus Z_m}$$

where 1_A denotes the 0-1 indicator of a set A. Define also

$$k_{\mathrm{St}}^m := k_{\mathrm{St}} - \check{k}_{\mathrm{St}}(f) + \check{k}_{\mathrm{St}}(f^m).$$

Then $\int_0^T f^m = 0$; and $\check{k}_{\mathrm{St}}(f^m) \le k_{\mathrm{St}}^m$ (from the definitions and the inequality $\check{k}_{\mathrm{St}}(f) \le k_{\mathrm{St}}$). Also, $0 \le y^m \le y \le k_{\mathrm{Tu}}$ and $0 \le \sigma^m \le e^m$. As $m \to \infty$, one has $f^m \to f$ in $\mathrm{m}\left(L^\infty, L^1\right)$ and, therefore, $k_{\mathrm{St}}^m \to k_{\mathrm{St}}$ by (4.14.1) and Remark 4.10.2. Put together, this shows that the sequence

$$(y^m, -k^m, -e) = (y^m, -k_{\mathrm{St}}^m, -k_{\mathrm{Tu}}, -e) \in \mathbb{Y}_{\mathrm{H}}$$

has the required properties, viz.,

$$\langle (y^m, -k^m, -e) \mid (p, r, \psi)_{\mathrm{FA}} \rangle = \langle y^m \mid p_{\mathrm{FA}} \rangle - \langle e^m \mid \psi_{\mathrm{FA}} \rangle = 0$$

and

$$\begin{aligned} &\langle (y^m, -k^m, -e) - (y, -k, -e) \mid (p, r, \psi)_{\mathrm{CA}} \rangle \\ &\qquad = \langle y^m - y \mid p_{\mathrm{CA}} \rangle - \langle k_{\mathrm{St}}^m - k_{\mathrm{St}} \mid r_{\mathrm{St}} \rangle - \langle e^m - e \mid \psi_{\mathrm{CA}} \rangle \to 0 \end{aligned}$$

as $m \to \infty$. □

It follows that pure density prices obtain in a general equilibrium model of electricity pricing with hydro-thermal generation, if the users' utility and production functions are Mackey continuous: see [43]. This means that if the demand for electricity is interruptible (i.e., a brief interruption causes only a small loss of utility or output), then the equilibrium TOU tariff is a time-varying rate in \$/kWh (with no instantaneous charges in \$/kW).[42]

[42] The interruptibility assumption is unnecessary for the density representation of water prices: that $\psi_{\mathrm{FA}} = 0$ follows from $\psi \in \partial_e \Pi_{\mathrm{SR}}^{\mathrm{H}} \subset \mathrm{BV} \subset L^1$.

4.15 Conclusions from Chapter 4

This analysis shows how to operate a hydro plant to maximise its profit, how to value the plant's capacities and its river flow on this basis, and how to use these valuations in investment decisions. As well as being better suited to the more decentralised structure of today's utilities, short-run profit-maximisation for an individual hydro plant turns out to be a much simpler problem to solve than that of cost minimisation for a whole hydro-thermal system. When a hydro plant is operated to maximise profit, the hydro inputs (including the water inflow) have well defined marginal values, at least if the given TOU price for electricity is continuous over the cycle. The marginal capacity values and the TOU water value can be calculated by solving a linear programme (or an equivalent convex but unconstrained programme). These values can be used to determine the optimum levels of investment on a hydro site.

The "constant-head" model of the hydro technology has other interpretations as well: the analysis and its valuation method are applicable to other natural energy flows (e.g., geothermal or tidal), and also to water supply (when priced by TOU). An extension to the variable-head case (which requires convex control theory) is given in [48]. Extension to the case of stochastic river inflow is a subject for future work. This would especially enhance the model's application to water supply (as well as to the original hydro problem).

Chapter 5

THE SHORT-RUN APPROACH TO LONG-RUN EQUILIBRIUM: A GENERAL THEORY WITH APPLICATIONS TO ELECTRICITY PRICING

5.1 Introduction to Chapter 5

This is a new formal framework for the theory of competitive equilibrium and its applications. The "short-run approach" is a scheme for calculating long-run producer optima and general equilibria by building on short-run solutions to the producer's profit maximisation problem, in which the capital inputs and natural resources are treated as fixed. These fixed inputs are valued at their marginal contributions to the operating profits and, where possible, their levels are then adjusted accordingly.[1] Since short-run profit is a concave but generally nondifferentiable function of the fixed inputs, their marginal values are defined as nonunique supergradient vectors. Also, they usually have to be obtained as solutions to the dual programme of fixed-input valuation because there is rarely an explicit formula for the operating profit. Thus the key property of the dual solution is its marginal interpretation, but this requires the use of a generalised, multi-valued derivative—viz., the subdifferential—because an optimal-value function, such as profit or cost, is commonly nondifferentiable.

Differential calculus is essential for applications, but it has been purged from geometric treatments of the Arrow-Debreu model, which are limited to equilibrium existence and Pareto optimality results. The use of subgradients rehabilitates calculus as a rigorous method for equilibrium theory. The mathematical tools employed here—convex programmes and subdifferentials—make it possible to reformulate some basic microeconomic results. In addition to the known subdifferential versions of the Shephard-Hotelling Lemmas, which are stated in [46], a subdifferential version of the Wong-Viner Envelope Theorem is devised here for the short-run approach especially (Section 5.9). As well as facilitating the economic analysis, this also resolves some long-standing discrepancies between "textbook theory" and industrial experience.[2]

These methods are used to set up a framework for the general-equilibrium pricing of multiple outputs with joint production costs. This is applied to the pricing, operation and investment prob-

[1] When carried out by iterations, the calculations might also be seen as modelling the real processes of price and quantity adjustments.

[2] The usual theory of differentiable convex functions is, of course, included in subdifferential calculus as a special case. Furthermore, the subgradient concept can also be used to prove—rather than assume—that a convex function is differentiable by showing that it has a unique subgradient. This method is used in Chapters 3 and 4.

lems of an electricity supply industry with a technology that can include hydroelectric generation and pumped storage of energy, in addition to thermal generation (Sections 5.13 to 5.15). This application draws on the much simpler case of purely thermal generation (Section 5.2) and on the studies of operation and valuation of hydroelectric and pumped-storage plants in Chapters 3 and 4. Those results are summarised and then "fed into" the short-run approach.

The short-run approach starts with fixing the producer's capacities k and optimising the variable quantities, viz., the outputs y and the variable inputs v. For a competitive, price-taking producer, the optimum quantities, $\hat{y}$ and $\hat{v}$, depend on their given prices, p and w, as well as on k.[3] The primal solution ($\hat{y}$ and $\hat{v}$) is associated with the dual solution $\hat{r}$, which gives the unit values of the fixed inputs (with $\hat{r} \cdot k$ as their total value); the optima are, for the moment, taken to be unique for simplicity. When the goal is limited to finding the producer's long-run profit maxima, it can be achieved by part-inverting the short-run solution map of (p, k, w) to $(y, -v; r)$ so that the prices (p, r, w) are mapped to the quantities $(y, -k, -v)$. This is done by solving the equation $\hat{r}(p, k, w) = r$ for k and substituting any solution into $\hat{y}(p, k, w)$ and $\hat{v}(p, k, w)$ to complete a long-run profit-maximising input-output bundle. Such a bundle may be unique but only up to scale if the returns to scale are constant (making $\hat{r}(p, k, w)$ homogeneous of degree zero in k).

Even within the confines of the producer problem, this approach saves effort by building on the short-run solutions that have to be found anyway: the problems of plant operation and plant valuation are of central practical interest and always have to be tackled by producers. But the short-run approach is even more important as a practical method for calculating market equilibria. For this, with the input prices r and w taken as fixed for simplicity, the short-run profit-maximising supply $\hat{y}(p, k, w)$ is equated to the demand for the products $\hat{x}(p)$ to determine the short-run equilibrium output prices $p^{\star}_{\mathrm{SR}}(k, w)$. The capacity values $\hat{r}(p, k, w)$, evaluated at the equilibrium prices $p = p^{\star}_{\mathrm{SR}}(k, w)$ with the given k and w, are only *then* equated to the given capacity prices r to determine the long-run equilibrium capacities $k^{\star}(r, w)$—by solving for k the equation $\hat{r}(p^{\star}_{\mathrm{SR}}(k, w), k, w) = r$. And hence, by substituting $k^{\star}(r, w)$ for k in the short-run equilibrium solution, the long-run equilibrium output prices and quantities are determined.[4] In other words, the determination of investment is postponed until *after* the equilibrium in the product markets has been found: the producer's long-run problem is split into two—that of operation and that of investment—and the equilibrium problem is "inserted" in-between. Since the operating solutions usually have relatively simple forms, doing things in this order can greatly ease the fixed-point problem of solving for equilibrium: indeed, the problem can even be elementary if approached in this way (Section 5.2). Furthermore,

[3] From Section 5.4 on, short-run cost minimisation is split off as a subprogramme, whose solution is $\check{v}(y, k, w)$. In these terms, $\hat{v}(p, k, w) = \check{v}(\hat{y}(p, k, w), k, w)$.

[4] The short-run approach to equilibrium might also be based on short-run cost minimisation, in which not only the capital inputs (k) but also the outputs (y) are kept fixed and are shadow-priced in the dual problem, but such cost-based calculations are usually much more complicated than those using profit maximisation: see Section 5.10.

unlike the optimal investment of the pure producer problem, the equilibrium investment $k^\star$ has a definite scale (determined by demand for the products). Put another way: $\hat{r}\left(p^\star_{\mathrm{SR}}\left(k,w\right),k,w\right)$, the value to be equated to r, is *not* homogeneous of degree zero in k like $\hat{r}\left(p,k,w\right)$. Thus one can keep mostly to single-valued maps, and avoid dealing with multi-valued correspondences—even when the returns to scale are constant. And finally, like the short-run producer optimum, the short-run general equilibrium is of interest in itself.

The core of this exposition is a framework for the short-run approach to the long-run general-equilibrium pricing of a range of commodities with joint costs of production (Sections 5.10 and 5.11). This is applied to the peak-load pricing of electricity generated by a variety of techniques (Sections 5.13 to 5.15); a greatly simplified version of this problem serves also as an introductory example (Section 5.2). Between this example and the core matter, the profit and cost optimisation programmes and their duals are introduced (Sections 5.3 and 5.5), and the required characterisations of long-run producer optima are given (Sections 5.4, 5.6 and 5.7). Several other characterisations are sketched in Sections 5.8 and 5.9; these are spelt out in [46]. Also in [46], the characterisations of producer optima are complemented by criteria for the existence of optimum quantities and shadow prices for the short-run profit maximisation and cost minimisation problems, and for the equality of total values of the variable quantities and the fixed quantities, i.e., for the absence of a gap between the primal and dual solutions.[5] Two short appendices (A and B) provide contextual examples of mathematical complications, one possible but exceptional (a duality gap), the other typical (a nonfactorable joint subdifferential of a nondifferentiable bivariate convex function). Appendix C gives the required results of convex calculus (with one innovation, viz., Lemma C.7.2 on subdifferential sections).

First, for a simple but instructive introduction to the short-run approach to long-run equilibrium, Boiteux's treatment of the simplest peak-load pricing problem is rehearsed; this is the problem of pricing the services of a homogeneous capacity that produces a nonstorable good with cyclic demands (such as electricity). A direct calculation of the long-run equilibrium poses a fixed-point problem, but, with cross-price independent demands, the short-run equilibrium can be obtained by the elementary method of intersecting the supply and demand curves for each time instant separately. At each time t, the short-run equilibrium output price $p^\star_{\mathrm{SR}}\left(t\right)$ is the sum of the unit operating cost w and a capacity charge $\kappa^\star_{\mathrm{SR}}\left(t\right) \geq 0$ that is nonzero only at times of full capacity utilisation, i.e., when the output rate $y^\star_{\mathrm{SR}}\left(t\right)$ equals the given capacity k. Finally, the long-run equilibrium is found by adjusting the capacity k so that its unit cost r equals its unit value, defined as the unit operating profit, which equals the total capacity charge over the cycle, $\int_0^T \kappa^\star_{\mathrm{SR}}\left(t\right) \mathrm{d}t = \int_0^T \left(p^\star_{\mathrm{SR}}\left(t\right) - w\right) \mathrm{d}t$. This solution is given by Boiteux with discretised time [12, 3.2–3.3].[6] Its continuous-time version is

[5] Furthermore, the concept of technologies with conditionally fixed coefficients is introduced in [46], and the general analysis is specialised to this class.

[6] Boiteux's work is also presented by Drèze [23, pp. 10–16], but the short-run character of the approach is more

given in Section 5.2.

Boiteux's idea is developed into a frame for the analysis of investment and pricing by an industry that supplies a range of commodities—such as a good differentiated over time, locations or events (Sections 5.10 and 5.11). In Sections 5.13 to 5.15, this is applied to augment the rudimentary one-station model to a continuous-time equilibrium model of electricity pricing with a diverse technology, including energy storage and hydro as well as thermal generation. Such a plant mix makes supply cross-price dependent, even in the short run (i.e., with the capacities fixed). Demand, too, is allowed to be cross-price dependent.

The setting up of the short-run approach to pricing and investment (Sections 5.10 and 5.11) is the most novel part of this study. Unlike the characterisation and existence results about producer optima, this is not fully formalised into theorems: it is assumed, rather than proved, that the short-run equilibrium is unique, and it is merely noted that its existence cannot be guaranteed unless the fixed capacities are all positive (i.e., unless $k \gg 0$).[7] The question of a general method of computing short-run market equilibria is only touched upon, in Figure 5.3, where the use of Walrasian tatonnement is suggested.[8] And no qualitative properties of the long-run condition $\hat{r}\left(p^{\star}_{\mathrm{SR}}\left(k, w\right), k, w\right) = r$, as an equation for the investment k, are established.[9] But it *is* shown that the *SRP programme-based system*, consisting of Conditions (5.11.11)–(5.11.15) together with (5.11.18)–(5.11.19), is a full characterisation of long-run market equilibrium. Furthermore, it is clear already from the introductory example of Section 5.2 that the short-run approach can greatly simplify the problem of solving for long-run equilibrium (as well as finding the short-run equilibrium on the way). It is apparent that the approach is worth applying not only to the case of electricity but also to the supply of other time-differentiated commodities (such as water, natural gas, etc.). The questions of uniqueness, stability and iterative computation of equilibria, though important, are not specific to the short-run approach; also, they have been much studied and are well understood (at least for finite-dimensional commodity spaces). The central and distinctive quantitative elements of the approach are valuation and operation of plants, and these are problems that have been fully solved for the various types of plant in the electricity supply industry (see Section 5.14 and its references). The priorities in developing the approach are: (i) to analyse the valuation and operation problems for other technologies and industries, and (ii) to compute numerical solutions from real data by using, at least to

evident from the original [12, 3.2–3.3] because Boiteux discusses the short-run equilibrium first, before using it as part of the long-run equilibrium system. Drèze mentions the short-run equilibrium on its own only as an afterthought [23, p. 16].

[7] This is not an unacceptable condition, but some capacities could of course be zero in long-run equilibrium. The long-run model meets the usual adequacy assumption, as does the short-run model with positive capacities, and so the existence of an equilibrium follows from results such as Bewley's [10, Theorem 1], which is amplified in [49, Section 3] and [47] by a proof using the continuity of demand in prices.

[8] As is well known, this process does not always converge, but there are other iterative methods.

[9] More generally, this is not an ordinary equation but an inclusion, viz., (5.11.18).

start with, the standard methods (viz., linear programming for producer optima and tatonnement for market equilibria). It would seem sensible to address the theoretical questions of uniqueness and stability in the light of future computational experience (in which more elaborate iterative methods could be employed if necessary). These questions are potentially important for practice as well as for completing the theory, but they are not priorities for this study, and are left for further research.

Each of the characterisations of long-run producer optima is either an *optimisation system* or a *differential system*, i.e., it is a set of conditions formulated in terms of either the marginal optimal values or the optimal solutions to a primal-dual pair of programmes (although one can also mix the two kinds of condition in one system). Though equivalent, the various systems are not equally usable, and the best choice of system depends on one's purpose as well as on the available mathematical description of the technology. In the application to electricity pricing with non-thermal as well as thermal generation, the technology is given by production sets (rather than profit or cost functions), and so the best tool for the short-run approach is the system using the programme of maximising the short-run profit (SRP), together with the dual programme for shadow-pricing the fixed inputs. For each *individual* plant type,[10] the problem of minimising the short-run cost (SRC) is typically easy (if it arises at all); therefore, it can be split off as a subprogramme (of profit maximisation). The resulting *split SRP optimisation system* serves as the basis of the present framework for the short-run approach to pricing and industrial investment (Section 5.11). Because of its importance to the application, this system is introduced as soon as possible, in Section 5.4—not only before the differential systems (Sections 5.7, 5.8 and 5.9), but also before the other optimisation systems (Sections 5.6 and 5.9), and even before a discussion of the dual programmes (in Section 5.5).

Of the differential systems, the first one to be presented formally, in Section 5.7, is that which generalises Boiteux's original set of conditions, limited though it is to technologies that are simple enough to allow explicit formulae not only for the SRC function but also for the SRP function. Another differential system, introduced informally in Section 5.2 and formally in Section 5.9, has the same mathematical form but uses the LRC instead of the SRP function (with the variables suitably switched). The two systems' equivalence extends the Wong-Viner Envelope Theorem (on the equality of SRMC and LRMC) to convex technologies with nondifferentiable cost functions by Formula (5.9.1)—and this is the result outlined earlier in Section 5.2 (where it is exemplified by an account of Boiteux's short-run approach to the simple peak-load pricing problem). The extension is made possible by using the subdifferential (a.k.a. the subgradient set) as a generalised, multi-valued derivative. This is necessary because the joint-cost functions may lose differentiability at crucial

[10] By contrast, SRC minimisation for a *system* of plants can be difficult because it involves allocating the system's given output among the plants. Its complexity shows in, e.g., the case of a hydro-thermal electricity-generating system [55]. The present decentralised approach avoids having to deal directly with the formidable problem of minimising the entire system's cost: see the Comments containing Formulae (5.10.3) and (5.10.4).

points. For example, in the simplest peak-load pricing problem,. the long-run cost is nondifferentiable at every output bundle with multiple global peaks because, although the total capacity charge is determinate (being equal to r, the given rental price of capacity), its distribution over the peaks cannot be determined purely by cost calculations. And, far from being exceptional, multiple peaks forming an output plateau do arise in equilibrium as a solution to the shifting-peak problem, as is shown in [43] under appropriate assumptions about demand.[11] The short-run marginal cost is even less determinate: whenever the output rate reaches full capacity, an SRMC exceeds the unit operating cost w by an arbitrary amount κ—which makes the capacity charge indeterminate in its total as well as in its distribution. This is an example of the inclusion between the subdifferentials of the two costs, as functions of the output bundle: the set of SRMCs is larger than the set of LRMCs when the capital inputs are at an optimum (i.e., minimise the total cost). It then takes a stronger condition to ensure that a particular SRMC is actually an LRMC. What is needed is the equality of rental prices to the *profit*-imputed values of the fixed inputs (which are the fixed inputs' marginal contributions to the operating profit). This equality is the required generalisation of Boiteux's long-run optimum condition, which, for his one-station technology, equates the capacity price r to the unit operating profit $\int \kappa \, \mathrm{d}t = \int (p(t) - w) \, \mathrm{d}t$ [12, 3.3, and Appendix: 12]. The valuations *must* be based on increments to the operating profit: it is generally ineffective to try to value capacity increments by any reductions in the operating cost. The one-station example shows just how futile such an attempt can be: excess capacity does not reduce the operating cost at all, but any capacity shortage makes the required output infeasible. This leaves the capacity value completely indeterminate by SRC calculations—in contrast to the definite value $\int (p(t) - w) \, \mathrm{d}t$ obtained by calculating the SRP. Only with differentiable costs is the SRC as good as the SRP for the purpose of capital-input valuation.

The extension of the Wong-Viner Envelope Theorem uses the SRP function and thus achieves for any convex technology what Boiteux [12, 1.1–1.2 and 3.2–3.3] in effect does with the very simple but nondifferentiable cost functions of his problem—which are spelt out here in (5.2.5) and (5.2.6). He realises that there is something wrong with the supposed equality of SRMC and LRMC [12, 1.1.4 and 1.2.2]. As he puts it,

> "It seems practically out of the question that these costs should be equal; it is difficult to imagine, for instance, how the marginal cost of operating a thermal power station could become high enough to equal the development cost (including plant) of the thermal energy [its long-term marginal cost]. The paradox is due to the fact that most industrial plants

[11] This shows how mistaken is the widespread but unexamined view that nondifferentiabilities of convex functions are of little consequence: the very points which are a *priori* exceptional turn out to be the rule rather than the exception in equilibrium. Also, it is only on finite-dimensional spaces that convex functions are "generically smooth" or, more precisely, twice differentiable almost everywhere with respect to the Lebesgue measure (Alexandroff's Theorem). On an infinite-dimensional space, a convex function can be nondifferentiable everywhere.

are in reality very 'rigid'. ...

There is no... question of equating the development cost to the cost of overloading the plant, since any such overloading is precluded by the assumption of rigidity. ...The more usual types of plant have some slight flexibility in the region of their limit capacities... but... any 'overloading' which might be contemplated in practice would never be sufficient to equate its cost with the development cost; hence the paradox referred to above."

Its resolution starts with his

"new notion which will play an essential part in 'peak-load pricing': for output equal to maximum, the differential cost [the SRMC] is indeterminate: it may be equal to, or less or greater than the development cost [the LRMC]."

In the language of subdifferentials, Boiteux's "new notion"—that the LRMC is just one of many SRMCs—is a case of the afore-mentioned general inclusion between the LRMCs and SRMCs, which is usually strict: $\partial_y C_{\mathrm{LR}}(y,r) \subsetneq \partial_y C_{\mathrm{SR}}(y,k)$ when $r \in -\partial_k C_{\mathrm{SR}}(y,k)$, i.e., when the bundle of capital inputs k minimises the total cost of an output bundle y, given their prices r (and given also the variable-input prices w, which, being kept fixed, are suppressed from the notation). For differentiable costs, this reduces to the Wong-Viner equality of gradient vectors: $\nabla_y C_{\mathrm{LR}} = \nabla_y C_{\mathrm{SR}}$ (when the capital inputs are at an optimum). But for nondifferentiable costs, all it shows is that each LRMC is an SRMC—which is the *reverse* of what is required for the short-run approach. The way out of this difficulty is to bring in the SRP function, Π_{SR}, and require that the given prices for the capital inputs are equal to their profit-imputed values, i.e., that $r = \nabla_k \Pi_{\mathrm{SR}}(p,k)$ or, should the gradient not exist, that $r \in \widehat{\partial}_k \Pi_{\mathrm{SR}}$ (which is the superdifferential a.k.a. the supergradient set). This condition is stronger than cost-optimality of the fixed inputs when the output price system p is an SRMC, i.e., if $p \in \partial_y C_{\mathrm{SR}}(y,k)$ then $\widehat{\partial}_k \Pi_{\mathrm{SR}}(p,k) \subseteq -\partial_k C_{\mathrm{SR}}(y,k)$, generally with a strict inclusion (indeed, $\nabla_k \Pi_{\mathrm{SR}}$ can exist even when $\nabla_k C_{\mathrm{SR}}$ does not, in which case $\nabla_k \Pi_{\mathrm{SR}} \in -\partial_k C_{\mathrm{SR}}$). And the new condition—that $r \in \widehat{\partial}_k \Pi_{\mathrm{SR}}(p,k)$—is no stronger than it need be: it is just strong enough to do the job and guarantee that if $p \in \partial_y C_{\mathrm{SR}}(y,k)$ then $p \in \partial_y C_{\mathrm{LR}}(y,r)$.

Thus this analysis of the relationship between SRMC and LRMC bears out, amplifies and develops Boiteux's ideas, which, at the time, he allowed, with a hint of exasperation, were "false in the theoretical general case, but more or less true of ordinary industrial plant". Both cases are accommodated here: the industrial reality of fixed coefficients and rigid capacities as well as the unrealistic textbook supposition of smooth costs. By bridging the gap between the inadequate existing theory and its intended applications, an end is put to its disturbing and unnecessary divorce from reality. This allows peak-load pricing to be put, for the first time, on a sound and rigorous theoretical basis (Sections 5.13 to 5.15).

From this perspective, Boiteux's long-run optimum condition, that $r = \int (p(t) - w)\,\mathrm{d}t$, should be viewed as a special case, for the one-station technology, of the equation $r = \nabla_k \Pi_{\mathrm{SR}}$. But staying within the confines of this particular example, Boiteux interprets his condition merely as recovery of the total cost of production, including the capital cost [12, 3.4.2: (2) and Conclusions: 4]. This is correct, but only in the case of a single capital input, and it cannot provide a basis for dealing with a production technique that uses a number of interdependent capital inputs.[12] In such a case, the generalisation of Boiteux's long-run optimum condition is stronger than capital-cost recovery: i.e., under constant returns to scale, if $r \in \widehat{\partial}_k \Pi_{\mathrm{SR}}$ (or $r = \nabla_k \Pi_{\mathrm{SR}}$), then $r \cdot k = \Pi_{\mathrm{SR}}$, but not *vice versa* (though the converse *is* of course true when k is a positive scalar). To think purely in terms of marginal costs and cost recovery is a dead end: with multiple capital inputs, cost recovery is *not* sufficient to guarantee that a short-run equilibrium is also a long-run equilibrium or, equivalently, that an SRMC tariff is also an LRMC tariff. The SRP function with its marginals (derivatives w.r.t. k), or the SRP programme with the dual solution, have to be brought into the short-run approach. This is done here for the first time.

In mathematical terms, the Extended Wong-Viner Theorem (5.9.1) comes from what is called the Subdifferential Sections Lemma (SSL), which gives the joint subdifferential of a bivariate convex function ($\partial_{y,k} C$) in terms of one of *its* partial subdifferentials ($\partial_y C$) and a partial superdifferential, $\widehat{\partial}_k \Pi(p, k)$, of the relevant partial conjugate (which is a saddle function): see Lemma C.7.2 in Appendix C. This is applied, twice, to either the SRP or the LRC as a saddle function obtained by partial conjugacy from the SRC, which is a jointly convex function (C) of the output bundle y and the fixed-input bundle k, with the variable-input prices w kept fixed (Section 5.9). The SSL can be regarded as a direct precursor of a well-known result of convex calculus, viz., the Partial Inversion Rule (PIR), which relates the partial sub/super-differentials of a saddle function ($\partial_p \Pi$ and $\widehat{\partial}_k \Pi$) to the joint subdifferential of its bivariate convex "parent" function ($\partial_{y,k} C$): see Lemmas C.7.3 and C.7.5 (whose proofs derive the PIR from the SSL). One well-known application of this fundamental principle is the equivalence of two optimality conditions, viz., the parametric version of Fermat's Rule and the Kuhn-Tucker characterisation of primal and dual optima as a saddle-point of the Lagrange function: see, e.g., [74, 11.39 (d) and 11.50]. Another well-known use of the PIR establishes the equivalence of Hamiltonian and Lagrangian systems in convex variational calculus; when the Lagrange integrand is nondifferentiable, this usefully splits the Euler-Lagrange differential inclusion

[12] Capital inputs are called *independent* if the SRP function (Π_{SR}) is linear in the capital-input bundle $k = (k_1, k_2, \ldots)$; an example is the multi-station technology of thermal electricity generation. Such a technology effectively separates into a number of production techniques with a single capital input each, and Boiteux's analysis applies readily: to ensure that the short-run equilibrium is also a long-run one, it suffices to require cost recovery for each production technique θ with $k_\theta > 0$, although one must also remember to check that any unused production technique (one with $k_\theta = 0$) cannot be profitable (e.g., that $r_\theta \geq \int (p(t) - w_\theta)\,dt$ for any unused type of thermal station).

(a generalised equation system) into the pair of Hamiltonian differential inclusions, and it may even transform the inclusion into ordinary equations because the Hamiltonian can be differentiable even when the Lagrangian is not: see, e.g., [73, (10.38) and (10.40)], [71, Theorem 6] or [5, 4.8.2].[13] The present use of the PIR or the SSL relates the marginal optimal values for a programme to those of a subprogramme, to put it in general terms. In the specific context of extending the Wong-Viner Theorem, SRC minimisation is a subprogramme both of SRP maximisation and of LRC minimisation; their optimal values are $C_{\mathrm{SR}}(y,k)$, $\Pi_{\mathrm{SR}}(p,k)$ and $C_{\mathrm{LR}}(y,r)$, respectively. This is a new use of what is, in Rockafellar's words, "a striking relationship...at the heart of programming theory" [69, p. 604].[14]

Like all optimisation, economic theory has to deal with the nondifferentiability of optimal values that commonly arises even when the programmes' objective and constraint functions are all smooth. This has led to the eschewing of marginal concepts in rigorous equilibrium analysis, but any need for this disappeared with the advent of nonsmooth calculus. Of course, in using generalised derivatives such as the subdifferential, one cannot expect to transcribe familiar theorems from the smooth to the subdifferentiable case simply by replacing the ordinary single gradients with multi-valued subdifferentials—proper subdifferential calculus must be applied. This not only extends the scope for marginal analysis, but also leads to a rethinking and reinterpretation that can give a new economic content to well-known results. The Wong-Viner Theorem is a case in point: a useful extension depends on recasting its fixed-input optimality assumption in terms of profit-based valuations (i.e., on restating the optimality of fixed inputs as equality of their rental prices to their marginal contributions to the operating profit). After this reformulation of optimality in terms of SRP marginals—but not before—the "smooth" version of the theorem can be transcribed to the case of subdifferentiable costs (by replacing each ∇ with a ∂). Without this preparatory step, all extension attempts are doomed: a direct transcription of the original Wong-Viner equality of SRMC and LRMC to the subdifferentiable case is plainly false, and although it can be changed to a true inclusion without bringing in the SRP function, that kind of result fails to attain the goal of identifying an SRMC as an LRMC.[15]

One well-known optimality condition is conspicuous by its absence from this analysis. The Lagrangian saddle-point condition of Kuhn and Tucker is central to the duality theory of convex

[13] To distinguish the two quite different meanings of the word "Lagrangian", it is occasionally expanded into either "Lagrange function" (in the multiplier method of optimisation) or "Lagrange integrand" (in the calculus of variations only).

[14] One half of this argument (the application of the SSL to the saddle function Π_{SR} as a partial conjugate of the bivariate convex function C_{SR} to prove the first equivalence in (5.9.1)) is given already in Section 5.8, at the bottom of Table 5.2.

[15] Without involving Π_{SR}, the inclusion ($\partial_y C_{\mathrm{LR}} \subseteq \partial_y C_{\mathrm{SR}}$) can be improved only by making it more precise but no more useful: $\partial_y C_{\mathrm{SR}}(y,k)$ can be shown to equal the union of $\partial_y C_{\mathrm{LR}}(y,r)$ over $r \in -\partial_k C_{\mathrm{SR}}(y,k)$, i.e., over all those fixed-input price systems r for which k is an optimal fixed-input bundle for the output bundle y (given also the omitted variable-input price system w). See the Comments at the end of Appendix B.

programmes (CPs)—and it is used in the studies of hydro and energy storage (Chapters 3 and 4) which serve the short-run framework's application to electricity supply in Sections 5.13 to 5.15—but the Kuhn-Tucker system is not used here. Instead, for a general analysis with an abstract production cone, the Complementarity Conditions on the price system and the input-output bundle (5.3.5) are preferred. This system is a case of what is called the FFE Conditions, which consist of primal feasibility, dual feasibility and equality of the primal and dual objectives (at the feasible points in question). The FFE Conditions form an effective system whenever the dual programme can be worked out from the primal explicitly. This is so with the profit and cost problems because they become linear programmes (LPs) once the production cone is represented by linear inequalities. For an LP, the FFE system is linear in the primal and dual variables jointly—unlike the Kuhn-Tucker system (which is nonlinear because of the quadratic term in the complementary slackness condition): compare (5.5.3) with (5.5.2). And a linear system (i.e., a system of linear equalities and inequalities) is much simpler to deal with: in particular, it can be solved numerically by the simplex method (or another LP algorithm). The problem's size is smaller, though, when the method is applied directly to the relevant LP (or to its dual), rather than to its FFE system.[16] Either way, there is no need for the Kuhn-Tucker system in solving the SRP programmes with their fixed-input valuation duals—although it is instrumental in proving uniqueness of their solutions, in Chapters 3 and 4.

In the LP formulation of a profit or cost programme, the fixed quantities are primal parameters but need not be the same as the standard "right-hand side" parameters—and so their shadow prices, which are the dual variables, need not be identical to the standard dual variables. Yet the usual theory of linear programming works with the standard parameterisation, and it is the *standard* dual solution that the simplex method provides along with the primal solution. But, as is shown in Section 5.12, this is not much of a complication because any other dual variables can be expressed in terms of the standard dual variables, i.e., in terms of the usual Lagrange multipliers for constraints. This is used in valuing the fixed inputs for electricity generation (Section 5.14). The principle has also a counterpart beyond the linear or convex duality framework: it is the Generalised Envelope Theorem for smooth optimisation, whereby the marginal values of all parameters, including any nonstandard ones, are equal to the corresponding partial derivatives of the ordinary Lagrangian—and are thus expressed in terms of the constraints' multipliers. See [1, (10.8)] or [79, 1.F.b].

This exposition of producer optimum pauses for "stock-taking" in Section 5.8. In particular, Tables 5.1 and 5.2 summarise the various characterisations of the long-run optimum, though not their "mirror images" which result from a formal substitution of the LRC for the SRP. These tables record also the methods employed to transform these systems into one another. This shows a unity: the same methods are applied to systems of the same type, even though the exposition gives special places to the two systems of importance for the application of the short-run approach, viz., the split

[16] For a count of variables and constraints, see a Comment in [46].

SRP optimisation system of Section 5.4 and the SRC-P saddle differential system of Section 5.7. The latter system's "mirror image", the L-SRC saddle differential system of Section 5.9, is also directly involved in applications when its conditions of LRMC pricing and LRC minimisation serve as the definition of long-run optimum—as is often the case in public utility pricing, including Boiteux's work and the account of it in Section 5.2. The other fourteen systems are not used here, but any can be the best tool, for the short-run approach as for other purposes, if the technology is described most simply in the system's own terms. In particular, one should not be trapped by the language into thinking that a system using the LRC programme or function is somehow fundamentally unsuitable for the short-run approach.

Section 5.8 ends by noting that some of the systems—including the two "special" ones—can be partitioned into a *short-run subsystem* (which characterises SRP maxima) and a supplementary condition that generalises Boiteux's long-run optimum condition and requires that investment be at a profit maximum. A complete formalisation of all the duality-based systems is carried out in [46], where the systems' equivalence are cast as rigorous results with proofs.

Notation is explained when first used, but it is also listed below in several categories. Later, Table 5.3 shows the correspondence of notation between the general duality scheme of Sections 5.5 and 5.12 and its application to electricity supply.[17]

List of Notation for Chapter 5

Profit and cost optimisation and shadow-pricing programmes: parameters and decision variables, solutions, optimal values and marginal values

$y \in Y$ an output bundle, in a space Y

$k \in K$ a fixed-input bundle, in a space K

$v \in V$ a variable-input bundle, in a space V

$p \in P$ an output price system, in a space P

$r \in R$ a fixed-input price system, in a space R

$w \in W$ a variable-input price system, in a space W

$\mathit{\Delta} y$, $\mathit{\Delta} k$, etc. increments to y, k, etc. ($\mathit{\Delta}$ differs from the upright Δ)

$\mathbb{Y}$ a production set (in the commodity space $Y \times K \times V$)

A, B and C matrices or linear operations, esp. such that $(y, -k, -v) \in \mathbb{Y}$ if and only if $Ay - Bk - Cv \leq 0$

A^{T} the transpose of a matrix A

[17]Note the two different uses of the symbols s and σ: in Sections 5.5 and 5.12, these mean the standard parameters and dual variables, but in Section 5.13 they mean the energy stock and water spillage. Also, the n_θ, n_{St} and n_{Tu} of Section 5.14 are lower constraint parameters (whose original, unperturbed values are zeros). In Sections 5.11 and 5.15, n means an input of the numeraire.

$\delta(\cdot \mid \mathbb{Y})$ the 0-∞ indicator function of the set $\mathbb{Y}$ (equal to 0 on $\mathbb{Y}$)

$\mathbb{Y}^{\circ}$ the polar cone of $\mathbb{Y}$ (a cone in $P \times R \times W$ when $\mathbb{Y}$ is a cone in $Y \times K \times V$)

$\mathbb{Y}^{\circ}_{p,w}$ the polar cone's section through (p, w)

$\mathcal{G}'$ and $\mathcal{G}''$ respectively, the sets of generators and of spanning vectors of $\mathbb{Y}^{\circ}$, when $\mathbb{Y}$ is a polyhedral cone in a finite-dimensional space

Π_{LR} the maximum long-run profit, a function of (p, r, w)

Π_{SR} the maximum short-run a.k.a. operating profit, a function of (p, k, w)

C_{LR} the minimum long-run cost, a function of (y, r, w)

C_{SR} the minimum short-run cost, a function of (y, k, w)

∂C the subdifferential of a convex function C

$\widehat{\partial}\Pi$ the superdifferential of a concave function Π

$\nabla\Pi$ the (Gateaux) gradient vector of a function Π

$\partial/\partial k$ partial differentiation with respect to a scalar variable k

$\check{V}(y, k, v)$ the set of all variable-input bundles that minimise the short-run cost

$\check{v}(y, k, v)$ *the* variable-input bundle such as above (i.e., minimising the short-run cost), if it is unique

$\hat{Y}(p, k, w)$ the set of all output bundles that maximise the short-run profit (i.e., maximise the function $\langle p \mid \cdot \rangle - C_{\mathrm{SR}}(\cdot, k, w)$)

$\hat{y}(p, k, w)$ *the* output bundle such as above (i.e., maximising the function $\langle p \mid \cdot \rangle - C_{\mathrm{SR}}(\cdot, k, w)$), if it is unique

$\hat{K}(p, r, w)$ the set of all fixed-input bundles that maximise the long-run profit

$\hat{k}(p, r, w)$ *the* fixed-input bundle such as above (i.e., maximising the long-run profit), if it is unique (under decreasing returns to scale)

$\underline{C}_{\mathrm{SR}}(y, k, w)$ the maximum, over shadow prices, of output value less fixed-input value (and less Π_{LR} when $\mathbb{Y}$ is not a cone)

$\underline{C}_{\mathrm{LR}}(y, r, w)$ the maximum, over shadow prices, of output value (less Π_{LR} when $\mathbb{Y}$ is not a cone)

$\overline{\Pi}_{\mathrm{SR}}(p, k, w)$ the minimum, over shadow prices, of total fixed-input value (plus Π_{LR} when $\mathbb{Y}$ is not a cone)

$\hat{R}(p, k, w)$ the set of all fixed-input price systems that minimise the total fixed-input value (plus Π_{LR} when $\mathbb{Y}$ is not a cone)

$\hat{r}(p, k, w)$ *the* fixed-input price system such as above (i.e., minimising the total fixed-input value), if it is unique

$\check{P}(y, k, w)$ the set of all output price systems that maximise the output value less fixed-input value, $\langle \cdot \mid y \rangle - \overline{\Pi}_{\mathrm{SR}}(\cdot, k, w)$, less Π_{LR} when $\mathbb{Y}$ is not a cone

$\check{p}(y, k, w)$ *the* output price system such as above (i.e., maximising $\langle \cdot \mid y \rangle - \overline{\Pi}_{\mathrm{SR}}(\cdot, k, w)$), if it is unique

s vector of the standard primal parameters for a convex or linear programme (paired to its equality and inequality constraints)

σ vector of the standard dual variables (Lagrange multipliers of the constraints) for a convex or linear programme

$\hat{\Sigma}(p, s)$ the set of all the standard dual solutions (Lagrange multiplier systems), when the primal is a linear programme with s as its primal parameters and $\langle p \mid \cdot \rangle$ as its linear objective function

$\hat{\sigma}(p, s)$ *the* standard dual solution such as above, if it is unique

$\mathcal{L}$ the Lagrangian (the Lagrange function of the primal and dual variables and parameters)

Characteristics of the Supply Industry

θ a production technique of the Supply Industry

Φ_θ the set of fixed inputs of production technique θ

Ξ_θ the set of variable inputs of production technique θ

$\mathbb{Y}_\theta$ the production set of technique θ, a cone in $Y \times \mathbb{R}^{\Phi_\theta} \times \mathbb{R}^{\Xi_\theta}$

ξ a variable input, with a price w_ξ

ϕ a fixed input, with a price r_ϕ

Φ^{F} the set of fixed inputs with *given* prices r^{F}

Φ^{E} the set of fixed inputs with prices r^{E} to be determined in long-run equilibrium

G_ϕ the supply cost of an equilibrium-priced input $\phi \in \Phi^{\mathrm{E}}$, a function of the supplied quantity q_ϕ (or k_ϕ)

Characteristics of consumer and factor demands (from Industrial User)

F production function of the Industrial User—a function of inputs: n of the numeraire and z of the differentiated good (e.g., electricity)

U_h consumer h's utility, a function of consumptions: φ of the Industrial User's product, m of the numeraire and x of the differentiated good (e.g., electricity)

$u(t, \mathrm{x})$ the consumer's instantaneous utility from the consumption rate x at time t (when U is additively separable)

m_h^{En} consumer h's initial endowment of the numeraire

$\varsigma_{h\phi}$ consumer h's share of profit Π_ϕ from the supply of input $\phi \in \Phi^{\mathrm{E}}$

$\varsigma_{h\,\mathrm{IU}}$ consumer h's share in the Industrial User's profit, Π_{IU}

$\varpi_{h\theta}$ consumer h's share in the operating profit from production technique θ of the Supply Industry

$B(p, \varrho, M)$ consumer's budget set when his income is M, the differentiated good (electricity) price is p and the Industrial User's product price is ϱ

$\hat{M}_{\mathrm{SR}\,h}(p; r^{\mathrm{E}}, r^{\mathrm{F}}; w, \varrho \mid k)$ consumer's income in the short run

$\hat{M}_{\mathrm{LR}\,h}\,(p, r^{\mathrm{E}}, \varrho)$ consumer's income in the long run (Supply Industry's pure profit is zero)

$\hat{x}_h\,(p, \varrho; M)$ consumer h's demand for the differentiated good (electricity) when its price is p, the Industrial User's product price is ϱ, and the income is M

$\hat{\varphi}_h\,(p, \varrho; M)$ consumer h's demand for the Industrial User's product when its price is ϱ, the differentiated good's (electricity) price is p, and the consumer's income is M

$\hat{z}\,(p, \varrho)$ the Industrial User's factor demand for the differentiated good (electricity)

$\hat{n}\,(p, \varrho)$ the Industrial User's factor demand for the numeraire

Short-run general-equilibrium prices and quantities

$p^{\star}_{\mathrm{SR}}$, $\varrho^{\star}_{\mathrm{SR}}$ prices for the differentiated good (electricity) and for the IU's product

$y^{\star}_{\mathrm{SR}\,\theta}$ output of the differentiated good (electricity) by production technique θ

$v^{\star}_{\mathrm{SR}\,\theta}$ variable input into production technique θ

$x^{\star}_{\mathrm{SR}\,h}$, $z^{\star}_{\mathrm{SR}}$ consumer demand and factor demand for the differentiated good (electricity)

$m^{\star}_{\mathrm{SR}\,h}$, $n^{\star}_{\mathrm{SR}}$ consumer demand and factor demand for the numeraire

$\varphi^{\star}_{\mathrm{SR}}$ the Industrial User's output

Long-run general-equilibrium prices and quantities

w the given prices of the Supply Industry's variable inputs

r^{F} the *given* rental prices of the Supply Industry's fixed-priced capital inputs

r^{E} rental prices of the Supply Industry's equilibrium-priced capital inputs—to be determined in long-run equilibrium

$r^{\star}$ the equilibrium prices of the equilibrium-priced inputs (i.e., the equilibrium value of r^{E})

$k^{\star}_{\theta}$ equilibrium capacities of producer θ in the Supply Industry

$p^{\star}_{\mathrm{LR}}$, $y^{\star}_{\mathrm{LR}\,\theta}$, etc. equilibrium prices and quantities—as above, but for the long-run equilibrium

Electricity generation (all techniques)

$p\,(t)$ electricity price at time t (in \$/kWh), i.e., p is a time-of-use tariff

$y\,(t)$ rate of electricity output from a plant, at time t (in kW)

$D_t\,(\mathrm{p})$ cross-price independent demand for electricity (in kW) at time t, if the current price is p

Thermal generation

$S\,(\mathrm{p})$ in the short run, the cross-price independent rate of supply (in kW) of thermally generated electricity, if the current price is p

$c_{\mathrm{SR}}\,(\mathrm{y})$ the instantaneous short-run thermal cost per unit time (in \$/kWh), if the current output rate is y (in kW); the common graph of the correspondences S and ∂c_{SR} is the thermal SRMC curve

θ a type of thermal plant

ξ_θ fuel type used by plant type θ

v fuel input of a thermal plant (in kWh of heat)

η technical efficiency of a thermal plant, i.e., $1/\eta$ is the heat rate

w unit running cost of a thermal plant (in \$ per kWh of electricity output), equal to the price of fuel (in \$ per kWh of heat) times the heat rate

k a thermal generating capacity (in kW)

$\kappa(t)$ unit value of a thermal generating capacity at time t, per unit time (in \$/kWh)

$r = \int_0^T \kappa(t)\,\mathrm{d}t$ unit value of a thermal generating capacity in total for the cycle (in \$/kW)

$\gamma(t) = \kappa(t) / \int_0^T \kappa(t)\,\mathrm{d}t$ density, at time t, of the distribution of capacity charges over the cycle, i.e., a function representing a subgradient of the convex functional EssSup on $L^\infty[0,T]$

r^{F} the *given* rental price of a thermal generating capacity (in \$/kW)

$\nu(t)$ unit value of nonnegativity constraint on the output of a thermal plant at time t, per unit time (in \$/kWh)

$\hat{Y}_{\mathrm{Th}}(p,k,w)$ the set of all the electricity output bundles that maximize the operating profit of a thermal plant of capacity k with a unit running cost w, when the electricity tariff is p

$\hat{y}_{\mathrm{Th}}(p,k,w)$ *the* electricity output bundle such as above (i.e., the one maximizing the storage plant's operating profit), if it is unique

$y_\theta^\star(t)$ the general-equilibrium rate of electricity output from the thermal plant of type θ at time t (in kW)

Pumped storage

k_{St} the plant's storage a.k.a. reservoir capacity (in kWh)

$\kappa_{\mathrm{St}}(\mathrm{d}t)$ unit value of storage capacity on a time interval of length $\mathrm{d}t$ (in \$/kWh)

$r_{\mathrm{St}} = \int_0^T \kappa_{\mathrm{St}}(\mathrm{d}t)$ unit value of storage capacity in total for the cycle (in \$/kWh)

$r_{\mathrm{St}}^\star$ the (long-run) equilibrium rental price of storage capacity (in \$/kWh)

$G(k_{\mathrm{St}})$ the supply cost of k_{St} of storage capacity

$\nu_{\mathrm{St}}(\mathrm{d}t)$ unit value of nonnegativity constraint on energy stock on an interval of length $\mathrm{d}t$ (in \$/kWh)

k_{Co} the plant's conversion capacity (in kW)

$\kappa_{\mathrm{Pu}}(t)$ unit value of converter's pump capacity at time t, per unit time (in \$/kWh)

$\kappa_{\mathrm{Tu}}(t)$ unit value of converter's turbine capacity at time t, per unit time (in \$/kWh)

$\kappa_{\mathrm{Co}}(t) = \kappa_{\mathrm{Pu}}(t) + \kappa_{\mathrm{Tu}}(t)$ unit value of converter's capacity at time t, per unit time (in \$/kWh)

$r_{\mathrm{Co}} = \int_0^T \kappa_{\mathrm{Co}}(t)\,\mathrm{d}t$ unit value of conversion capacity in total for the cycle (in \$/kW)

$r_{\mathrm{Co}}^{\mathrm{F}}$ the *given* rental price of conversion capacity (in \$/kW)

$\hat{Y}_{\mathrm{PS}}(p; k_{\mathrm{St}}, k_{\mathrm{Co}})$ the set of all the electricity output bundles that maximise the operating profit of a pumped-storage plant with capacities $(k_{\mathrm{St}}, k_{\mathrm{Co}})$, when the electricity tariff is p

$\hat{y}_{\mathrm{PS}}(p; k_{\mathrm{St}}, k_{\mathrm{Co}})$ *the* electricity output bundle such as above (i.e., the one maximising the storage plant's operating profit), if it is unique

$y^{\star}_{\mathrm{PS}}(t)$ the general-equilibrium rate of electricity output from the pumped-storage plant at time t (in kW)

s_0 energy stock at time 0 and T (in kWh)

λ unit value of energy stock at time 0 and T (in \$/kWh)

$s(t)$ energy stock at time t (in kWh)

$\varsigma_{h\,\mathrm{St}}$ household h's share of profit from supplying the storage capacity (i.e., share of the rent for the storage site)

$\psi(t)$ unit value of energy stock at time t (in \$/kWh)

$\hat{\psi}_{\mathrm{PS}}(p; k_{\mathrm{St}}, k_{\mathrm{Co}})$ *the* imputed time-of-use unit value (shadow price) of energy stock, if the value is unique (as a function of time)

$\hat{\Psi}_{\mathrm{PS}}(p; k_{\mathrm{St}}, k_{\mathrm{Co}})$ the set of all the imputed time-of-use values of energy stock (shadow-price functions for energy stock) in a pumped-storage plant with capacities $(k_{\mathrm{St}}, k_{\mathrm{Co}})$, when the electricity tariff is p

Hydro

k_{St} the plant's storage a.k.a. reservoir capacity (in kWh)

$\kappa_{\mathrm{St}}(\mathrm{d}t)$ unit value of storage capacity on a time interval of length $\mathrm{d}t$ (in \$/kWh)

$r_{\mathrm{St}} = \int_0^T \kappa_{\mathrm{St}}(\mathrm{d}t)$ unit value of storage capacity in total for the cycle (in \$/kWh)

$r^{\star}_{\mathrm{St}}$ the (long-run) equilibrium rental price of storage capacity (in \$/kWh)

$G(k_{\mathrm{St}})$ the supply cost of reservoir of capacity k_{St}

$\nu_{\mathrm{St}}(\mathrm{d}t)$ unit value of nonnegativity constraint on water stock on an interval of length $\mathrm{d}t$ (in \$/kWh)

k_{Tu} the plant's turbine-generator capacity (in kW)

$\kappa_{\mathrm{Tu}}(t)$ unit value of turbine capacity at time t, per unit time (in \$/kWh)

$r_{\mathrm{Tu}} = \int_0^T \kappa_{\mathrm{Tu}}(t)\,\mathrm{d}t$ as the unit value of turbine capacity in total for the cycle (in \$/kW)

$r^{\mathrm{F}}_{\mathrm{Tu}}$ the *given* rental price of turbine capacity (in \$/kW)

$\nu_{\mathrm{Tu}}(t)$ unit value of nonnegativity constraint on turbine's output at time t, per unit time (in \$/kWh)

$e(t)$ rate of river flow at time t (in kW)

$\hat{Y}_{\mathrm{H}}(p; k_{\mathrm{St}}, k_{\mathrm{Tu}}; e)$ the set of all the electricity output bundles that maximise the operating profit of a hydro plant with capacities $(k_{\mathrm{St}}, k_{\mathrm{Tu}})$ and river inflow function e when the electricity tariff is p

$\hat{y}_{\mathrm{H}}(p; k_{\mathrm{St}}, k_{\mathrm{Tu}}; e)$ *the* electricity output bundle such as above (i.e., the one maximising the hydro plant's operating profit), if it is unique

$y^{\star}_{\mathrm{H}}(t)$ the general-equilibrium rate of electricity output from the hydro plant at time t (in kW)

$\sigma(t)$ rate of spillage from the reservoir at time t (in kW)

s_0 water stock at time 0 and T (in kWh)

λ unit value of water stock at time 0 and T (in \$/kWh)

$s(t)$ water stock at time t (in kWh)

$\psi(t)$ unit value of water stock at time t (in \$/kWh)

$\hat{\psi}_{\mathrm{H}}(p; k_{\mathrm{St}}, k_{\mathrm{Tu}}; e)$ *the* imputed time-of-use unit value of water (shadow price of water), if the value is unique (as a function of time)

$\hat{\Psi}_{\mathrm{H}}(p; k_{\mathrm{St}}, k_{\mathrm{Tu}}; e)$ the set of all the imputed time-of-use water values (shadow water-price functions) in a hydro plant with capacities $(k_{\mathrm{St}}, k_{\mathrm{Tu}})$ and river inflow function e, when the electricity tariff is p

$\varsigma_{h\,\mathrm{St}}$ household h's share of profit from supplying the reservoir capacity (i.e., share of the rent for the hydro site)

Specific vector spaces, norms and functionals

meas the Lebesgue measure, on an interval $[0,T]$ of the real line $\mathbb{R}$

$L^1[0,T]$ the space of meas-integrable real-valued functions on $[0,T]$

$L^\infty[0,T]$ the space of essentially bounded real-valued functions on $[0,T]$

$\mathrm{EssSup}(y) = \mathrm{ess\,sup}_{t\in[0,T]}\, y(t)$ the essential supremum of a $y \in L^\infty[0,T]$

$\|y\|_\infty := \mathrm{EssSup}\,|y|$ the supremum norm on L^∞

$\mathcal{C}[0,T]$ the space of continuous real-valued functions on $[0,T]$

$\mathcal{M}[0,T]$ the space of Borel measures on $[0,T]$

$\int_{[0,T]} s(t)\,\mu(\mathrm{d}t)$ the integral of a continuous function s with respect to a measure μ

$\mathrm{BV}(0,T)$ the space of functions of bounded variation on $(0,T)$

$\mathrm{Var}^+(\psi)$ the total positive variation (upper variation) of a $\psi \in \mathrm{BV}(0,T)$

$\mathrm{Var}^+_{\mathrm{c}}(\psi) := \mathrm{Var}^+(\psi) + (\psi(0) - \psi(T))^+$ the cyclic positive variation of ψ

Norms and topologies on vector spaces, dual spaces, order and nonnegativity, scalar product

Y^* the norm-dual of a Banach space $(Y, \|\cdot\|)$

$\|\cdot\|^*$ the dual norm on Y^*

Y' the Banach predual of $(Y, \|\cdot\|)$, when Y is a dual Banach space

$\|\cdot\|'$ the predual norm on Y'

Y_+, Y^*_+ and Y'_+ the nonnegative cones in Y, Y^* and Y' (when these are Banach lattices), e.g., L^∞_+ and L^1_+ are the nonnegative cones in L^∞ and L^1

y^+ and y^- the nonnegative and nonpositive parts of a $y \in Y$ (when Y is a vector lattice)

$k \gg 0$ means that k is a strictly positive vector (in a lattice paired with another one); here, used only with a finite-dimensional k

$\langle\cdot\,|\,\cdot\rangle$ a bilinear form (scalar product) on the Cartesian product, $P \times Y$, of two vector spaces (when $P = \mathbb{R}^n = Y$, $p \cdot y$ is an alternative notation for the scalar product $\langle p\,|\,y\rangle := p^{\mathrm{T}}y$, where y is a column vector and p^{T} is a row of the same, finite dimension n)

$\mathrm{w}\,(Y, P)$ the weak topology on a vector space Y for its pairing with another vector space P (e.g., with Y^* or Y' when Y is a dual Banach space)

$\mathrm{m}\,(Y, P)$ the Mackey topology on Y for its pairing with P (e.g., with $P = Y^*$ or with $P = Y'$ when Y is a dual Banach space)

w^* and m^* abbreviations for $\mathrm{w}\,(P^*, P)$ and $\mathrm{m}\,(P^*, P)$, the weak* and the Mackey topologies on the norm-dual of a Banach space P

$\mathrm{cl}_{Y,\mathcal{T}}\, Z$ the closure of a set Z relative to a (larger) set Y with a topology $\mathcal{T}$

$\mathrm{int}_{Y,\mathcal{T}}\, Z$ the interior of a set Z relative to a (larger) set Y with a topology $\mathcal{T}$

Y^{a} the algebraic dual of a vector space Y

$\mathcal{T}_{\mathrm{SLC}} = \mathrm{m}\,(Y, Y^{\mathrm{a}})$ the strongest locally convex topology on a vector space Y

Sets derived from a set in a vector space

$\mathrm{cone}\, Z$ the cone generated by a subset, Z, of a vector space (i.e., the smallest cone containing Z)

$\mathrm{conv}\, Z$ the convex hull of a subset, Z, of a vector space (i.e., the smallest convex set containing Z)

$\mathrm{ext}\, Z$ the set of all the extreme points of a subset, Z, of a vector space

$\mathrm{span}\, Z$ the linear span of a subset, Z, of a vector space

$\mathrm{N}\,(y \mid Z) = \partial\delta\,(y \mid Z)$ the outward normal cone to a convex set Z at a point $y \in Z$ (a cone in the dual space)

Sets and functions derived from functions or operations on a vector space

$\mathrm{argmax}_Z\, f$ the set of all maximum points of an extended-real-valued function f on a set Z

$\mathrm{dom}\, C$ the effective domain of a convex extended-real-valued function C

$\widehat{\mathrm{dom}}\Pi$ the effective domain of a concave extended-real-valued function Π

$\mathrm{epi}\, C$ the epigraph of a convex extended-real-valued function C (on a vector space)

$\mathrm{lsc}\, C$ the lower semicontinuous envelope of C (the greatest l.s.c. minorant of C)

$\mathrm{usc}\, \Pi$ the upper semicontinuous envelope of Π (the least u.s.c. majorant of Π)

$C^{\#}$ the Fenchel-Legendre convex conjugate (of a convex function C)

$\Pi_{\#}$ the concave conjugate (of a concave function Π)

$C^{\#_{1,2}}$, etc. the partial conjugate, of a multi-variate function, w.r.t. all the variables shown (here, w.r.t. the first and the second variables together)

$C' \bigtriangleup C''$ the infimal convolution of convex functions, C' and C''

Other notation

card Φ the number of elements in a (finite) set Φ

$\emptyset$ the empty set

1_A the 0-1 indicator function of a set A (equal to 1 on A)

lim inf, lim sup respectively, the lower and the upper limits (of a real-valued) function)

$\mathbb{R}$ the real line

5.2 Peak-load pricing with cross-price independent demands

The short-run approach to solving for long-run general equilibrium is next illustrated with the example of pricing, over the demand cycle, the services of a homogeneous productive capacity with a unit capital cost r and a unit running cost w. The technology can be interpreted as, e.g., electricity generation from a single type of thermal station with a fuel cost w (in \$/kWh) and a capacity cost r (in \$/kW) per period. The cycle is represented by a continuous time interval $[0, T]$. Demand for the time-differentiated, nonstorable product, $D_t(\mathsf{p})$, is assumed to depend only on the time t and the current price p. As a result, the short-run equilibrium can be found separately at each instant t, by intersecting the demand and supply curves in the price-quantity plane. This is because, with this technology, short-run supply is cross-price independent: given a capacity k, the supply is

$$S(\mathsf{p}, k, w) = \begin{cases} 0 & \text{for } \mathsf{p} < w \\ [0, k] & \text{for } \mathsf{p} = w \\ k & \text{for } \mathsf{p} > w \end{cases} \tag{5.2.1}$$

where p is the current price. That is, given a time-of-use (TOU) tariff p (i.e., given a price $p(t)$ at each time t), the set of profit-maximising output trajectories, $\hat{Y}(p, k, w)$, consists of selections from the correspondence $t \mapsto S(p(t), k, w)$. When $D_t(w) > k$, the short-run equilibrium TOU price, $p^\star_{\mathrm{SR}}(t, k, w)$, exceeds w by whatever is required to bring the demand down to k (Figure 5.1a). The total premium over the cycle is the unit operating profit, which in the long run should equal the unit capacity cost r—i.e., the long-run equilibrium capacity, $k^\star(r, w)$, can be determined by solving for k the equation

$$r = \int_0^T (p^\star_{\mathrm{SR}}(t, k, w) - w)^+ \mathrm{d}t \tag{5.2.2}$$

(i.e., by equating to r the shaded area in Figure 5.1b), where $\pi^+ = \max\{\pi, 0\}$ is the nonnegative part of π. Put into the short-run equilibrium price function, the equilibrium capacity gives the long-run equilibrium price

$$p^\star_{\mathrm{LR}}(t; r, w) = p^\star_{\mathrm{SR}}(t, k^\star(r, w), w). \tag{5.2.3}$$

An obvious advantage of this method is that the short-run equilibrium is of interest in itself. Also, the short-run calculations can be very simple, as in this example. For comparison, to calculate the long-run equilibrium directly requires timing the capacity charges so that they are borne entirely by the resulting demand peaks—i.e., it requires finding a density function $\gamma \geq 0$ such that

$$\int_0^T \gamma(t)\,\mathrm{d}t = 1 \quad \text{and if } \gamma(t) > 0 \text{ then } y(t) = \sup_\tau y(\tau) \qquad (5.2.4)$$
$$\text{where:} \quad y(t) = D_t(p(t)) \text{ and } \quad p(t) = w + r\gamma(t).$$

This poses a fixed-point problem that, unlike the short-run approach, is not much simplified by cross-price independence of demands.[18]

Since the operating profit is $\Pi_{\mathrm{SR}}(p, k, w) = k \int_0^T (p(t) - w)^+ \mathrm{d}t$, the break-even condition (5.2.2) can be rewritten as $r = \partial \Pi_{\mathrm{SR}} / \partial k$, i.e., it can be viewed as equating the capital input's price to its profit-imputed marginal value. This is, with any convex technology, the first-order necessary and sufficient condition for a profit-maximising choice of investment k: together with a choice of output y that maximises the short-run profit (SRP), such a choice of k maximises the long-run profit (LRP), and thus turns the short-run equilibrium into the long-run equilibrium.

Furthermore, with any technology and any number of capital inputs, $r = \nabla_k \Pi_{\mathrm{SR}}$ if and only if r is the unique solution to the dual of the SRP maximisation programme (and there is no duality gap): this is the derivative property of the optimal value Π_{SR} as a function of the primal parameter k. This identity is useful when, with a more complex technology, the SRP programme has to be solved by a duality method, i.e., together with its the dual. It means that the dual solution $\hat{r}(p, k, w)$, evaluated at the short-run equilibrium output price system $p^\star_{\mathrm{SR}}(k, w)$, can be equated to the capital inputs' *given* prices r to determine their long-run equilibrium quantities $k^\star$.

When the producer is a public utility, competitive profit maximisation usually takes the form of marginal-cost pricing. In this context, the equality $r = \partial \Pi_{\mathrm{SR}} / \partial k$, or $r = \nabla_k \Pi_{\mathrm{SR}}$ when there is more than one type of capacity, guarantees that an SRMC price system is actually an LRMC. The result applies to any convex technology—even when the costs are nondifferentiable, and the marginal cost has to be defined by using the subdifferential as a generalised, multi-valued derivative. This is so in the above example of capacity pricing, since the long-run cost

$$C_{\mathrm{LR}}(y(\cdot), r, w) = w \int_0^T y(t)\,\mathrm{d}t + r \sup_{t \in [0,T]} y(t) \qquad (5.2.5)$$

is nondifferentiable if the output y has multiple peaks: indeed, for every γ satisfying (5.2.4), the function $p = w + r\gamma$ represents a subgradient of C_{LR} with respect to y (w.r.t. y). And multiple peaks are more the rule than the exception in equilibrium (note the peak output plateau in Figure 5.1d here, and see [43] for an extension to the case of cross-price dependent demands). Similarly, the

[18] In terms of the subdifferential, ∂C, of the long-run cost (5.2.5) as a function of output, the fixed-point problem is to find a function p such that $p \in \partial C_{\mathrm{LR}}(D(p))$, where $D(p)(t) = D_t(p(t))$ if demands are cross-price independent.

short-run cost

$$C_{\mathrm{SR}}(y(\cdot), k, w) = \begin{cases} w \int_0^T y(t)\,\mathrm{d}t & \text{if } 0 \le y \le k \\ +\infty & \text{otherwise} \end{cases} \tag{5.2.6}$$

is nondifferentiable if $\sup_t y(t) = k$. In Figure 5.1a, the nondifferentiability shows in the (infinite) vertical interval $[w, +\infty)$ that represents the multi-valued instantaneous SRMC at $\mathrm{y} = k$.[19] In Figure 5.1c, it shows as a kink, at $\mathrm{y} = k$, in the graph of the instantaneous cost function

$$c_{\mathrm{SR}}(\mathrm{y}) = \begin{cases} w\mathrm{y} & \text{if } 0 \le \mathrm{y} \le k \\ +\infty & \text{otherwise} \end{cases} \tag{5.2.7}$$

(which gives $C_{\mathrm{SR}}(y)$ as $\int_0^T c_{\mathrm{SR}}(y(t))\,\mathrm{d}t$, so that a TOU price p is an SRMC at y if and only if $p(t)$ is an instantaneous SRMC at $y(t)$ for each t). With this technology, C_{SR} is therefore nondifferentiable whenever k is the cost-minimising capital input for the required output y: cost-optimality of k means merely that it provides just enough capacity, i.e., that $k = \mathrm{Sup}(y)$. Since this condition does not even involve the capital-input price r, it obviously cannot ensure that an SRMC price system p is an LRMC. To guarantee this, one must strengthen it to the condition that $r = \int_0^T (p(t) - w)^+ \mathrm{d}t$ in this example or, generally, that $r = \nabla_k \Pi_{\mathrm{SR}}$ (or that r belongs to the supergradient set $\widehat{\partial}_k \Pi_{\mathrm{SR}}(p, k, w)$, should Π_{SR} be nondifferentiable in k).[20] The capital's cost-optimality would suffice for the SRMC to be the LRMC if the costs were differentiable; this is the Wong-Viner Envelope Theorem. The preceding remarks show how to reformulate it to free it from differentiability assumptions. This is detailed in Section 5.9.

Cross-price independent demand arises from price-taking optimisation by consumers and industrial users with additively separable utility and production functions. In this case, the short-run equilibrium prices can readily be given in terms of the marginal utility of the differentiated good (and its productivity if there are industrial users). For the simplest illustration, all demand is assumed to come from a single household maximising the utility function

$$U(x(\cdot), m) = m + \int_0^T u(t, x(t))\,\mathrm{d}t$$

over $x(\cdot) \ge 0$ and $m \ge 0$, the consumptions of the nonstorable good and the numeraire, subject to the budget constraint

$$m + \int_0^T p(t)\, x(t)\,\mathrm{d}t \le M$$

where M is the income and $p(\cdot)$ is a TOU price in terms of the numeraire (which represents all the other goods and thus closes the model). For each t, the instantaneous utility $u(t, \mathrm{x})$ is taken to

[19]The SRMC and the short-run supply correspondences are inverse to each other, i.e., have the same graph: in Figure 5.1a, the broken line is both the supply curve and the SRMC curve.

[20]This condition ($r = \nabla_k \Pi_{\mathrm{SR}}$) is stronger than cost-optimality of the fixed inputs when p is an SRMC.

be a strictly concave, increasing and differentiable function of the consumption rate $\mathsf{x} \in \mathbb{R}_+$, with $(\partial u/\partial \mathsf{x})(t,0) > w$ (to ensure that, in a short-run equilibrium with $k > 0$, consumption is positive at every t). The income consists of an endowment of the numeraire (m^{En}) and the pure profit from electricity sales, i.e.,

$$M = m^{\mathrm{En}} + k \int_0^T (p(t) - w)^+ \, \mathrm{d}t - rk.$$

To guarantee a positive demand for the numeraire, assume that $m^{\mathrm{En}} > (Tw + r)k$. Then, at any time t, demand (for the good) depends only on the current price $p(t)$, and it is determined from the equation

$$\frac{\partial u}{\partial \mathsf{x}}(t, x(t)) = p(t).$$

In other words, $D_t(\mathsf{p}) = ((\partial u/\partial \mathsf{x})(t,\cdot))^{-1}(\mathsf{p})$. When $w < (\partial u/\partial \mathsf{x})(t,k)$, this value of $\partial u/\partial \mathsf{x}$ is the price needed to equate demand to k. So the short-run equilibrium price is

$$p^{\star}_{\mathrm{SR}}(t,k,w) = w + \left(\frac{\partial u}{\partial \mathsf{x}}(t,k) - w\right)^+. \tag{5.2.8}$$

By (5.2.2) and (5.2.3), the long-run equilibrium capacity $k^{\star}(r,w)$ is determined from

$$r = \int_0^T \left(\frac{\partial u}{\partial \mathsf{x}}(t,k) - w\right)^+ \mathrm{d}t$$

and the long-run equilibrium price is, in terms of $k^{\star}$,

$$p^{\star}_{\mathrm{LR}}(t,r,w) = w + \left(\frac{\partial u}{\partial \mathsf{x}}(t,k^{\star}(r,w)) - w\right)^+. \tag{5.2.9}$$

5.3 Cost and profit as values of programmes with quantity decisions

Costs and profits of a price-taking producer are, by definition, the optimal values of programmes with quantities as decision variables. With several variables, it can be much easier to solve the mathematical problem in stages, by fixing some variables and dealing with the resulting subproblem first. The subproblem may also be of independent interest, especially if it corresponds to a stage in a practical implementation of a complete solution. In production, the decision on plant operation (with fixed investment) corresponds to short-run profit maximisation as a subproblem of long-run profit maximisation: although operation is usually planned along with investment, the producer is still free to make operating decisions after constructing the plant. In other words, his final choices of the outputs y and the variable inputs v are made only after fixing the capital inputs k. Such a multi-stage problem can be solved in the reverse order: this means that the decisions to be implemented last are determined first but are made contingent on the decisions to be implemented earlier, and the complete solution is put together by back substitution. For the producer, this means first choosing y and v to maximise short-run profit, given an *arbitrary* k as well as the prices, p and w, for the

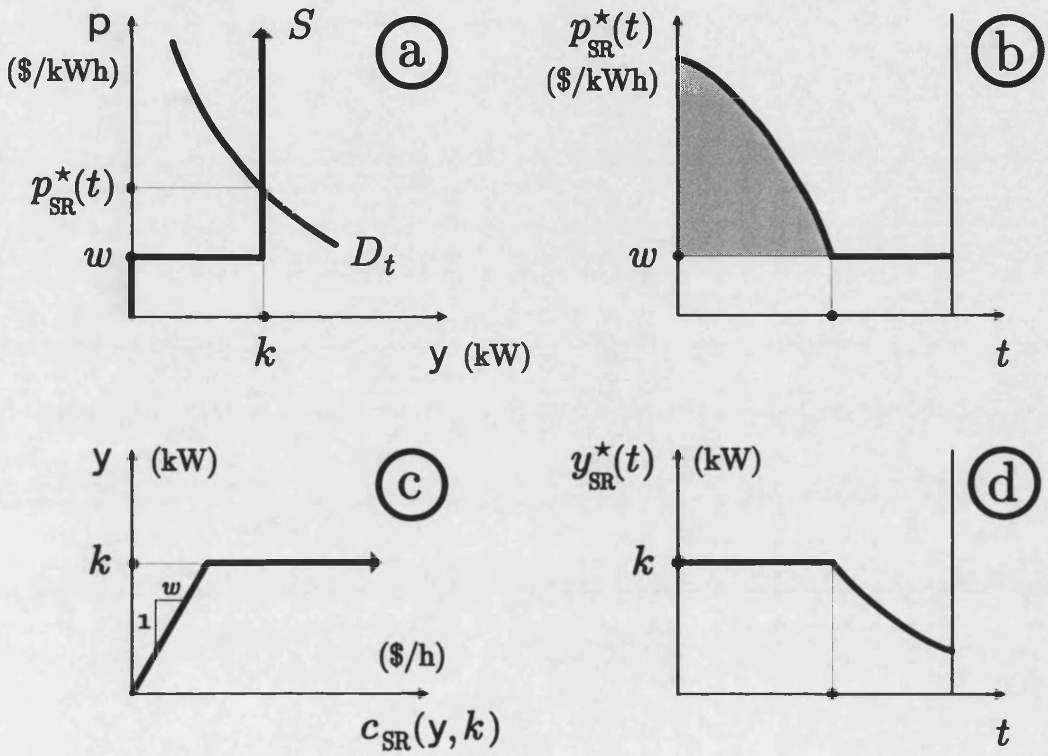

FIGURE 5.1. Short-run approach to long-run equilibrium of supply and (cross-price independent) demand for thermally generated electricity: (a) determination of the short-run equilibrium price and output for each instant t, given a capacity k; (b) and (d) trajectories of the short-run equilibrium price and output; (c) the short-run cost curve. When k is such that the shaded area in (b) equals r, the short-run equilibrium is the long-run equilibrium.

variable commodities. Even within the confines of the purely periodic (or static) problems considered here, this approach has a couple of analytical advantages. First, in addition to being of independent interest, the short-run equilibrium (given k) can be much easier to find than the long-run equilibrium, as in Section 5.2. Second, when there is a number of technologies, the short-run equilibrium is usually much easier to find by solving the profit maximisation programmes (to determine the total short-run supply and equate it to demand) than by solving the duals of cost minimisation programmes (to determine the SRMCs, which would have to be equated both to one another and to the inverse demand). This *profit approach* is simpler than the cost approach in two ways, viz., by giving unique solutions to the producer problem with its dual, and by reducing the number of unknowns in the subsequent equilibrium problem: see Section 5.10.

A third advantage of the short-run approach emerges only when the framework, unlike this one, takes account of non-periodic demand and price uncertainty. The prices for the variable commodities

(p, w), or their probability distribution in a stochastic model, will change in unforeseen ways between the planning and the completion of plants, and will also keep shifting thereafter. As a result, both the plant mix and the design of individual plants will become suboptimal. But whether a plant is optimal or not, it should be optimally operated, and a solution to this problem is part of the short-run approach.

It is the above considerations that make short-run profit maximisation the subproblem of central interest to us. It, too, may be solved in two stages, though this time the order in which the decision variables (y and v) are determined is only a matter of convenience: it is usually best to start with the simpler subproblem. Here, it is assumed that short-run cost minimisation (finding v given k and y) is easier than revenue maximisation (finding y given k and v). The solution sequence (first v, then y and finally k) corresponds to a chain of three problems: (i) the "small" one of short-run cost minimisation (with k and y as data, v as a decision), (ii) an "intermediate" problem of short-run profit maximisation (with k as a datum,and y and v as decisions), and (iii) the "large" problem of long-run profit maximisation (with k, y and v as decision variables).

A fourth problem, another intermediate one, is that of long-run cost minimisation (with y as a datum, k and v as decision variables). It is in terms of this problem and its value function that public utilities usually formulate their welfare-promoting principles of meeting the demand at a minimum operating cost, optimising their capital stocks, and pricing their outputs at LRMC. Together, these policies result in long-run profit maximisation and competitive equilibrium in the products' markets. Although the separate aims are stated in terms of long-run costs (as LRMC pricing and LRC minimisation), their combination is best achieved through short-run calculations—for the reasons outlined above and detailed in Section 5.10.

Each of the four problems, when formulated as one of optimisation constrained by a convex *production set* $\mathbb{Y}$, has a linear objective function. This has several implications. One is that each problem (SRC or LRC minimisation, or SRP or LRP maximisation) can be formulated as a linear programme (LP), by representing $\mathbb{Y}$ as the intersection of a finite or infinite set of half-spaces; this is discussed further in Section 5.12. What matters for now is that in passing to a subproblem, once a decision variable has become a datum (like k in passing from long to short run), the corresponding term of the linear optimand $(r \cdot k)$ can be dropped, since it is fixed. Its coefficient (r) can then be removed from the subproblem's data (which include k).

The *commodity spaces* for outputs, fixed inputs and variable inputs are denoted by Y, K and V, respectively. These are paired with *price spaces* P, R and W by bilinear forms (a.k.a. scalar products) denoted by $\langle p \,|\, y \rangle$, etc.; the alternative notation $p \cdot y$ is employed to mean $p^{\mathrm{T}} y$ when both P and Y are equal to the finite-dimensional space $\mathbb{R}^n$ (where p^{T} is the row vector obtained by transposing a column p). Unless specified, the range of a decision variable (say y) is the whole space (Y).

With p, r and w denoting the prices for outputs, fixed inputs and variable inputs (y, k and v, respectively), the *long-run profit maximisation* programme is:

$$\text{Given } (p,r,w)\,,\ \text{maximise } \langle p\,|\,y\rangle - \langle r\,|\,k\rangle - \langle w\,|\,v\rangle \ \text{ over } (y,k,v) \tag{5.3.1}$$

$$\text{subject to } (y,-k,-v) \in \mathbb{Y}. \tag{5.3.2}$$

Its optimal value, the maximum LRP as a function of the data, is denoted by $\Pi_{\mathrm{LR}}(p,r,w)$. By definition, (y,k,v) solves (5.3.1)–(5.3.2) if and only if

$$(y,-k,-v) \in \mathbb{Y} \text{ and } \langle p,r,w\,|\,y,-k,-v\rangle = \Pi_{\mathrm{LR}}(p,r,w)\,. \tag{5.3.3}$$

In the central case of constant returns to scale (c.r.t.s.), the production set $\mathbb{Y}$ is a cone, and Π_{LR} is the 0-∞ indicator of the *polar cone*

$$\mathbb{Y}^{\circ} = \{(p,r,w) : \forall\,(y,-k,-v) \in \mathbb{Y}\ \ \langle p\,|\,y\rangle - \langle r\,|\,k\rangle - \langle w\,|\,v\rangle \le 0\} \tag{5.3.4}$$

i.e., $\Pi_{\mathrm{LR}}(p,r,w)$ is 0 if $(p,r,w) \in \mathbb{Y}^{\circ}$, and it is $+\infty$ otherwise. Condition (5.3.3) is then equivalent to the conjunction of technological feasibility, price consistency and break-even conditions, which make up the *Complementarity Conditions*

$$(y,-k,-v) \in \mathbb{Y},\ (p,r,w) \in \mathbb{Y}^{\circ} \text{ and } \langle p,r,w\,|\,y,-k,-v\rangle = 0. \tag{5.3.5}$$

One subprogramme of (5.3.1)–(5.3.2) is short-run profit maximisation, i.e.,

$$\text{Given } (p,k,w)\,,\ \text{maximise } \langle p\,|\,y\rangle - \langle w\,|\,v\rangle \ \text{ over } (y,v) \tag{5.3.6}$$

$$\text{subject to } (y,-k,-v) \in \mathbb{Y}. \tag{5.3.7}$$

Its optimal value is $\Pi_{\mathrm{SR}}(p,k,w)$, the maximum SRP.

Another subprogramme of (5.3.1)–(5.3.2) is long-run cost minimisation, i.e.,

$$\text{Given } (y,r,w)\,,\ \text{minimise } \langle r\,|\,k\rangle + \langle w\,|\,v\rangle \ \text{ over } (k,v) \tag{5.3.8}$$

$$\text{subject to } (y,-k,-v) \in \mathbb{Y}. \tag{5.3.9}$$

Its optimal value is $C_{\mathrm{LR}}(y,r,w)$, the minimum LRC.

The common subprogramme (of all three of the above) is short-run cost minimisation, i.e.,

$$\text{Given } (y,k,w)\,,\ \text{minimise } \langle w\,|\,v\rangle \ \text{ over } v \tag{5.3.10}$$

$$\text{subject to } (y,-k,-v) \in \mathbb{Y}. \tag{5.3.11}$$

Its optimal value is $C_{\mathrm{SR}}(y,k,w)$, the minimum SRC.

Partial conjugacy relationships between the value functions (Π_{LR}, Π_{SR}, C_{LR}, C_{SR}) are summarised in the following diagram:

$$\boxed{\begin{array}{ccccccc}
 & & & w & & & \\
 & & & \Pi_{\mathrm{LR}} & & & \\
 & & r & & p & & \\
 & \nearrow & & & & \nwarrow & \\
 & k & & & & y & \\
w \quad \Pi_{\mathrm{SR}} & & & & & & C_{\mathrm{LR}} \quad w \\
 & p & & & & r & \\
 & \nwarrow & & & & \nearrow & \\
 & & y & & k & & \\
 & & & C_{\mathrm{SR}} & & & \\
 & & & w & & &
\end{array}} \tag{5.3.12}$$

For example, the arrow from the y next to C_{SR} to the p next to Π_{SR} indicates that Π_{SR} is, as a function of p, the Fenchel-Legendre convex conjugate of C_{SR} as a function of y, with (k, w) fixed; i.e., by definition,

$$\Pi_{\mathrm{SR}}(p, k, w) = \sup_y \{\langle p \mid y \rangle - C_{\mathrm{SR}}(y, k, w)\}. \tag{5.3.13}$$

Similarly, $-\Pi_{\mathrm{LR}}$ is, as a function of r, the concave conjugate of Π_{SR} as a function of k, with (p, w) fixed; i.e.,

$$\Pi_{\mathrm{LR}}(p, r, w) = \sup_k \{\Pi_{\mathrm{SR}}(p, k, w) - \langle r \mid k \rangle\}. \tag{5.3.14}$$

The right half of the diagram (5.3.12) represents similar links between C_{LR} and C_{SR} or Π_{LR}. Details such as the signs and convexity or concavity are omitted.

As is spelt out next, those y and k which yield the suprema in (5.3.13) and (5.3.14) are parts of an input-output bundle that maximises the long-run profit.

5.4 The split SRP optimisation system: a primal-dual system for the short-run approach

A joint programme for two or more decision variables can be split by optimising in stages: first over a subset of the variables (keeping the rest fixed), then over the other variables (the optimand comprising the value function from the first stage) to obtain the complete solution by back substitution. The method can be applied to solve the LRP maximisation programme (5.3.1)–(5.3.2) for (y, k, v) by:

1. first minimising $\langle w \mid v \rangle$ over v (subject to $(y, -k, -v) \in \mathbb{Y}$) to find the solution set $\check{V}(y, k, w)$, or *the* solution $\check{v}(y, k, w)$ if it is indeed unique, and the minimum value $C_{\mathrm{SR}}(y, k, w)$, which is $\langle w \mid \check{v} \rangle$;

2. then maximising $\langle p \mid y \rangle - C_{\mathrm{SR}}(y, k, w)$ over y to find the solution set $\hat{Y}(p, k, w)$, or *the* solution $\hat{y}(p, k, w)$ if it is unique, and the maximum value $\Pi_{\mathrm{SR}}(p, k, w)$, which is $\langle p \mid \hat{y} \rangle - C_{\mathrm{SR}}(\hat{y})$;

3. and finally, maximising $\Pi_{\mathrm{SR}}(p,k,w) - \langle r \mid k \rangle$ over k to find the solution set $\hat{K}(p,r,w)$, or *the* solution $\hat{k}(p,r,w)$, should it be unique (which it obviously cannot be if returns to scale are constant, in the long run).

Every complete solution can then be given, in terms of p, r and w, as a triple $(y,-k,-v)$ such that: $k \in \hat{K}(p,r,w)$, $y \in \hat{Y}(p,k,w)$ and $v \in \check{V}(y,k,w)$. With decreasing returns to scale, if the solution is unique, it is the triple: $\hat{k}(p,r,w)$, $\hat{y}\left(p,\hat{k}(p,r,w),w\right)$ and $\check{v}\left(\hat{y}\left(p,\hat{k}(p,r,w),w\right),\hat{k}(p,r,w),w\right)$.

In other words, the LRP programme (5.3.1)–(5.3.2) for (y,k,v) can be reduced to an investment programme, for k alone, by first solving the SRP programme (5.3.6)–(5.3.7) for (y,v) and substituting its optimal value (Π_{SR}) for the term $\langle p \mid y \rangle - \langle w \mid v \rangle$ in (5.3.1). The SRP programme for (y,v) can, in turn, be reduced to a programme for y alone by solving the SRC programme (5.3.10)–(5.3.11) and substituting its value (C_{SR}) for the term $\langle w \mid v \rangle$ in (5.3.6).

So an input-output bundle $(y,-k,-v)$ maximises long-run profit at prices (p,r,w) if and only if both

$$k \text{ maximises } \Pi_{\mathrm{SR}}(p,\cdot,w) - \langle r \mid \cdot \rangle \text{ on } K \text{ (given } p, r \text{ and } w) \tag{5.4.1}$$

and the bundle $(y,-v)$ maximises short-run profit (given k) at prices (p,w) or, equivalently,

$$y \text{ maximises } \langle p \mid \cdot \rangle - C_{\mathrm{SR}}(\cdot,k,w) \text{ on } Y \text{ (given } p, k \text{ and } w) \tag{5.4.2}$$

$$v \text{ minimises } \langle w \mid \cdot \rangle \text{ on } \{v \in V : (y,-k,-v) \in \mathbb{Y}\} \text{ (given } y, k \text{ and } w). \tag{5.4.3}$$

The system (5.4.1)–(5.4.3) is called the *split LRP optimisation system.* Its SRC subprogramme for v in (5.4.3) is taken to be readily soluble. By contrast, the reduced SRP programme for y in (5.4.2) may require a duality approach. This consists in pricing the constraining parameters, and in solving the dual programme of valuation together with the primal (when there is no duality gap). For the SRP programme as the primal, this means valuing the fixed inputs k: a dual solution (with no gap) is a shadow-price system r such that

$$r \text{ minimises } \langle \cdot \mid k \rangle + \Pi_{\mathrm{LR}}(p,\cdot,w) \text{ on } R \text{ (given } p, k \text{ and } w) \tag{5.4.4}$$

$$\text{and the minimum value, } \langle r \mid k \rangle + \Pi_{\mathrm{LR}}(p,r,w)\text{, equals } \Pi_{\mathrm{SR}}(p,k,w). \tag{5.4.5}$$

Under c.r.t.s., Conditions (5.4.4) and (5.4.5) become

$$r \text{ minimises } \langle \cdot \mid k \rangle \text{ on } \{r \in R : (p,r,w) \in \mathbb{Y}^{\circ}\} \text{ (given } p, k \text{ and } w) \tag{5.4.6}$$

$$\text{and the minimum value, } \langle r \mid k \rangle\text{, equals } \Pi_{\mathrm{SR}}(p,k,w). \tag{5.4.7}$$

The duality scheme that produces the programme in (5.4.6) or (5.4.4) as the dual to SRP maximisation is set out in Section 5.5.

As well as helping solve the operation problem in (5.4.2), the dual solution can be used to check the investment for optimality, i.e., (5.4.1) is equivalent to (5.4.4)–(5.4.5). Formally, this follows from the definitional conjugacy relationship (5.3.14) between Π_{SR} and Π_{LR} (as functions of k and r) by using the first-order condition (C.5.5) and the Inversion Rule (C.6.2), given in Appendix C. The system (5.4.2)–(5.4.5) is therefore equivalent to (5.4.1)–(5.4.3), and hence also to LRP maximisation (5.3.3), and to Complementarity (5.3.5) under c.r.t.s. It is, however, put entirely in terms of solutions to the SRP programme for (y, v) and its dual for r, with the primal split into the SRC programme (for v) and the reduced SRP programme (for y). Therefore, (5.4.2)–(5.4.5) is called the *split SRP optimisation system.* It is likely to be the best basis for the short-run approach when the technology is specified by means of a production set. Alternative systems are presented in Sections 5.6, 5.7, 5.8 and 5.9.

5.5 Cost and profit as values of programmes with price decisions

Unless there are duality gaps, short-run and long-run cost and profit are also the optimal values of programmes that are dual to those of Section 5.3. The scheme producing the duals is an application of the usual duality framework for convex programmes (CPs), expounded in, e.g., [73] and [57, Chapter 7]. However, this scheme starts not from a single programme but from a family of programmes that depend on a set of data, whose particular values complete the programme's specification. One way to perturb the programme is simply to add an increment to its data point, thus "shifting" it within the given family. Some, possibly all, of the scheme's primal perturbations are therefore increments to some—though typically not all—of the data. The same goes for the dual perturbations.

Before applying the duality scheme to the profit and cost programmes, it is discussed briefly and illustrated in the framework of linear programming. A central idea is that the dual programme depends on the choice of perturbations of the primal programme; different perturbation schemes produce different duals. Theoretical expositions usually start from a programme without any data variables whose increments might serve as primal perturbations: say, $f(y)$ is to be maximised over y subject to $G(y) \leq 0$. In such a case, any perturbations must first be introduced, and the standard choice is to add $\epsilon = (\epsilon_1, \epsilon_2, \ldots)$ to the zeros on the r.h.s.'s, thus perturbing the original constraints $G(y) \leq 0$ to $G(y) \leq \epsilon$. The original programme has no data other than the functions f and G themselves, and the increments Δf and ΔG (which change the programme to maximisation of $(f + \Delta f)(y)$ over y subject to $(G + \Delta G)(y) \leq 0$) could never serve as primal perturbations—not even if they were taken to be linear, i.e., if f and G were a vector and a matrix of coefficients of the primal variables, $y = (y_1, y_2, \ldots)$. This is because the perturbed constrained maximand must be *jointly* concave in the decision variables *and* the perturbations,[21] but the bilinear form $f \cdot y$ is not concave (or convex) in f and y jointly.[22]

[21] This is equivalent to joint convexity of the constrained minimand, which is the sum of the minimand and the 0-∞ indicator function of the constraint set. In [73] it is called "the minimand" for brevity.

[22] After a linear change of variables, it becomes a saddle function: $4f \cdot y = (f+y) \cdot (f+y) - (f-y) \cdot (f-y)$ is

But in applications, increments to some of the programme's data can commonly serve as primal perturbations. Here, those data are called the *intrinsic primal parameters*; some or all of the other data will turn out to be dual parameters. For example, in SRP maximisation (5.3.6)–(5.3.7), the fixed-input bundle k is a primal parameter because, since the production set $\mathbb{Y}$ is convex, the constrained maximand is a concave function of (y, k, v): it is

$$\langle p \mid y \rangle - \langle w \mid v \rangle - \delta \left(y, -k, -v \mid \mathbb{Y} \right)$$

where $\delta(\cdot, \cdot, \cdot \mid \mathbb{Y})$ denotes the 0-∞ indicator of $\mathbb{Y}$ (i.e., it equals 0 on $\mathbb{Y}$ and $+\infty$ outside of $\mathbb{Y}$). By contrast, the coefficient (say, p) of a primal variable (y) is *not* a primal parameter (i.e., its increment Δp cannot be a primal perturbation) because the bilinear form $\langle p \mid y \rangle$ is *not jointly* concave in p and y. For these reasons, all of the quantity data, but no price data, are primal parameters for the profit or cost optimisation programmes of Section 5.3. As for the production set, it cannot itself serve as a parameter because convex sets do not form a vector space to begin with. However, once the technological constraint $(y, -k, -v) \in \mathbb{Y}$ has been represented in the form $Ay - Bk - Cv \leq 0$ (under c.r.t.s.), the matrices or, more generally, the linear operations A, B and C are vectorial data. But none can be a primal parameter, for lack of joint convexity of Ay in A and y, etc. Nor can A, B or C be a dual parameter (for a similar reason). Such data variables, which are neither primal nor dual parameters, and hence play no role in the duality scheme, will be called *tertial* parameters.

It can be analytically useful, or indeed necessary, to introduce other primal perturbations, i.e., perturbations that are not increments to any of the data (which are listed after "Given" in the original programme). This amounts to introducing additional parameters, which are called *extrinsic*; their original, unperturbed values can be set as zeros, as in [73]. When the constraint set is represented by a system of inequalities and equalities, the standard "right-hand side" parameters are always available for this purpose (unless they are all intrinsic, but this is so only when the r.h.s. of each constraint is a separate datum of the programme and can therefore be varied independently of the other r.h. sides). Section 5.12 shows how to relate the marginal effects of any other, "nonstandard" perturbations to those of the standard ones—i.e., how to express any "nonstandard" dual variables in terms of the usual Lagrange multipliers for the constraints. This is useful in the problems of plant operation and valuation, including those that arise in peak-load pricing (Section 5.14).[23]

convex in $f + y$ and concave in $f - y$.

[23] In this as in other contexts, it can be convenient to think of extrinsic perturbations either as complementing the intrinsic perturbations (which are increments to the fixed inputs) by varying some aspects of the technology (such as nonnegativity constraints), or as replacing the intrinsic perturbations with finer, more varied increments (to the fixed inputs). For example, the time-constant capacity k_θ in (5.14.3) is an intrinsic primal parameter. The corresponding perturbation is a constant increment Δk_θ, and this can be refined to a time-varying increment $\Delta k_\theta(\cdot)$. This perturbation (Δk_θ or $\Delta k_\theta(\cdot)$) is complemented by the increment $\Delta n_\theta(\cdot)$ to the zero floor for the output rate $y_\theta(\cdot)$ in (5.14.3). The same goes for all the occurrences of Δk and Δn in the context of pumped storage and hydro, where $\Delta\zeta$ is another complementary extrinsic perturbation.

Once a primal perturbation scheme has been fully defined, the framework is completed automatically (except for the choice of topologies and the continuous-dual spaces in the infinite-dimensional case): dual decision variables are introduced and paired to the specified primal perturbations (both the intrinsic and any extrinsic ones). The corresponding dual match is set up in reverse: to be paired with the primal variables, dual perturbations are introduced. The perturbed dual minimand—a function of the dual variables, the dual perturbations and the data of the original, primal programme—is defined in the usual way (as in [73, (4.17)] but with the primal problem reoriented to maximisation). When all the primal perturbation are intrinsic, the resulting dual programme is called the *intrinsic dual*.

Some or possibly all of the dual perturbations may turn out to perturb the dual programme just like increments to some of the data—which are thus identified as the *intrinsic dual parameters*. Any other dual perturbations are called extrinsic, and these can be thought of as increments to *extrinsic dual parameters* (whose original, unperturbed values are set as zeros). However, in the profit or cost programmes, all the dual parameters are price data (and are therefore intrinsic).

In the *reduced* formulations of the profit or cost problems, some of the price data are *not* dual parameters because the corresponding quantities have been solved for in the reduction process, and have thus ceased to be decision variables: e.g., the variable-input price w is *not* a dual parameter of the reduced SRP programme in (5.4.2) because the corresponding input bundle v has been found in SRC minimisation (and so it is no longer a decision variable). But in the *full* (i.e., non-reduced) formulations, all the price data are dual parameters, and thus the programme's data (other than the technology itself) are partitioned into the primal parameters (the quantity data) and dual parameters (the price data).

The primal and dual optimal values can differ at some "degenerate" parameter points (see Appendix A), but such *duality gaps* are exceptional, and they do not occur when the primal or dual value is semicontinuous in, respectively, the primal or dual parameters: see, e.g., [73, Theorem 15] or [57, 7.3.2]. (In [46], this result is spelt out for the SRP, LRC and SRC problems, and it is complemented by sufficient criteria for semicontinuity or continuity of profit and cost as functions of the quantities.) Note that both optimal values, primal and dual, depend on the data, which are the same for both programmes. So, in this scheme, each of the optimal values (primal and dual) is a function of both the primal and the dual parameters), and it can have two varieties of continuity and differentiability properties:

- Properties of Type One are those of the primal value with respect to the primal parameters, and of the dual value w.r.t. the dual parameters.
- Properties of Type Two are those of the primal value w.r.t. the dual parameters, and of the dual value w.r.t. the primal parameters.

This distinction cannot be articulated when, as in [73] and [57], the primal and dual values are considered only as functions of either the primal or the dual parameters, respectively.

Comments (parameters and their marginal values, dual programme and FFE Conditions, the Lagrangian and Kuhn-Tucker Conditions for LPs):

- Let the primal linear programme be: Given any vectors p and s (and a matrix A), maximise $p \cdot y$ over y subject to $Ay \leq s$. Here, the only intrinsic primal parameter is the standard parameter s. There is no obviously useful candidate for an extrinsic primal parameter, and if none is introduced, then the dual is the *standard dual* LP: Given p and s (and A), minimise $\sigma \cdot s$ over $\sigma \geq 0$ subject to $A^{\mathrm{T}}\sigma = p$, where A^{T} is the transpose of A.[24] The only dual parameter is p. If both programmes have unique solutions, $\hat{y}(s,p,A)$ and $\hat{\sigma}(s,p,A)$, with equal values $\mathcal{V}(s,p,A)$ $:= p \cdot \hat{y} = \hat{\sigma} \cdot s =: \overline{\mathcal{V}}(s,p,A)$, then the marginal values of all the parameters, including the tertial (non-primal, non-dual) parameter A, exist as ordinary derivatives. Namely: (i) $\nabla_s \mathcal{V} = \nabla_s \overline{\mathcal{V}}$ $= \hat{\sigma}$, (ii) $\nabla_p \mathcal{V} = \nabla_p \overline{\mathcal{V}} = \hat{y}$, and (iii) $\nabla_A \mathcal{V} = \nabla_A \overline{\mathcal{V}} = -\hat{\sigma} \otimes \hat{y} = -\hat{\sigma}\hat{y}^{\mathrm{T}}$ (the matrix product of a column and a row, in this order, i.e., the tensor product), where ∇_A is arranged in a matrix like A (i.e., $\partial \mathcal{V} / \partial_{A_{ij}} = -\hat{\sigma}_i \hat{y}_j$ for each i and j). The first two formulae (for $\nabla_s \mathcal{V}$ and $\nabla_p \mathcal{V}$) are cases of a general derivative property of the optimal value in convex programming: see, e.g., [73, Theorem 16: (b) and (a)] or [51, 7.3: Theorem 1']. Heuristically, the third formula follows from each of the first two by comparing the marginal effect of A with that of either s or p on the constraints (primal or dual). It can also be proved formally by applying the Generalised Envelope Theorem for smooth optimisation [1, (10.8)],[25] whereby each marginal value ($\nabla_s \mathcal{V}$, $\nabla_p \mathcal{V}$ and $\nabla_A \mathcal{V}$) is equal to the corresponding partial derivative of the Lagrangian, which is here

$$\mathcal{L}(y,\sigma;p,s;A) := \begin{cases} p \cdot y + \sigma^{\mathrm{T}}(s - Ay) & \text{if } \sigma \geq 0 \\ +\infty & \text{if } \sigma \ngeq 0 \end{cases} . \tag{5.5.1}$$

- The Kuhn-Tucker Conditions form the system

$$\sigma \geq 0,\ Ay \leq s,\ \sigma^{\mathrm{T}}(Ay - s) = 0 \quad \text{and} \quad p^{\mathrm{T}} = \sigma^{\mathrm{T}} A \tag{5.5.2}$$

  which, because of the quadratic term $\sigma^{\mathrm{T}} Ay$, is nonlinear in the decision variables (y and σ).

- But the FFE Conditions (primal feasibility, dual feasibility and equality of the primal and dual objectives) form the system

$$Ay \leq s,\ \sigma \geq 0,\ p^{\mathrm{T}} = \sigma^{\mathrm{T}} A \quad \text{and} \quad p \cdot y = \sigma \cdot s \tag{5.5.3}$$

[24] The dual constraint must be changed to $A^{\mathrm{T}}\sigma \geq p$ if $y \geq 0$ is adjoined as another primal constraint. (In that case, the primal LP may be interpreted as, e.g., revenue maximisation given a resource bundle s, an output-price system p and a Leontief technology defined by an input-coefficient matrix A.)

[25] Without a proof of value differentiability, the Generalised Envelope Theorem is also given in, e.g., [79, 1.F.b].

which *is* linear (in y and σ). This makes it easier to solve than the Kuhn-Tucker system (5.5.2). For an LP, the FFE system is effective because the dual programme can be worked out from the primal explicitly.

- For a general CP, the dual cannot be given explicitly (i.e., without leaving an unevaluated extremum in the formula for the dual constrained objective function in terms of the Lagrangian).[26] That is why the Kuhn-Tucker system is better as a general solution method than the FFE system, although the latter is simpler in some specific cases (such as linear programming). The FFE system requires forming the dual from the primal to start with, but the Kuhn-Tucker system requires only the Lagrangian. It offers a workable method of solving the programme pair, and this matters more than an explicit expression for the dual programme. However, as with an LP, the FFE system can be simpler with a specific CP that has an explicit dual.

The duality scheme is next applied to all four of the profit and cost programmes of Section 5.3; the one of most importance for the applications given here is the dual to SRP maximisation. The duals are shown to consist in shadow-pricing the given quantities, so their subprogramme relationship is the reverse of that between the primals: the more quantities that are fixed, the more commodities there are to shadow-price. (In other words, the fewer primal variables, the more primal parameters, and hence the more dual variables.) For this reason, the duals are listed, below, in the reverse order to that of the primals (listed in Section 5.3). See also Figure 5.2, in which the large single arrows point from primal programmes to their subprogrammes, and the double arrows point from the dual programmes to their subprogrammes. Each of the four middle boxes gives the data for the pair of programmes represented by the two adjacent boxes (the outer box for the primal and the inner box for the dual); the data are partitioned into the primal parameters (the given quantities) and the dual parameters (the given prices). There are no other parameters in this scheme (i.e., it has no extrinsic parameters).

In the SRC minimisation programme (5.3.10)–(5.3.11), only y and k can serve as primal parameters;[27] and perturbation by both increments, Δy and Δk, yields the following dual programme for shadow-pricing both the outputs and the fixed inputs:

$$\text{Given } (y, k, w), \text{ maximise } \langle p \mid y \rangle - \langle r \mid k \rangle - \Pi_{\mathrm{LR}}(p, r, w) \text{ over } (p, r). \tag{5.5.4}$$

Its optimal value is denoted by $\underline{C}_{\mathrm{SR}}(y, k, w) \leq C_{\mathrm{SR}}(y, k, w)$, with equality when $C_{\mathrm{SR}}(\cdot, \cdot, w)$ is finite and lower semicontinuous (l.s.c.) at (y, k). The dual parameter is w.

[26]The standard dual to the ordinary CP of maximising a concave function $f(y)$ over y subject to $G(y) \leq s$ (where G_1, G_2, etc., are convex functions) is to minimise $\sup_y \mathcal{L}(y, \sigma) := \sup_y (f(y) + \sigma \cdot (s - G(y)))$ over $\sigma \geq 0$ (the standard dual variables, which are the Lagrange multipliers for the primal constraints): see, e.g., [73, (5.1)]. And $\sup_y \mathcal{L}$ (the Lagrangian's supremum over the primal variables) cannot be evaluated without assuming a specific form for f and G (the primal objective and constraint functions).

[27]Since the minimand $\langle w \mid v \rangle$ is not jointly convex in (w, v), w cannot serve as a primal parameter.

In the LRC minimisation programme (5.3.8)–(5.3.9), only y can serve as a primal parameter; and perturbation by the increment Δy yields the following dual programme for shadow-pricing the outputs:

$$\text{Given } (y, r, w), \text{ maximise } \langle p \mid y \rangle - \Pi_{\text{LR}}(p, r, w) \text{ over } p. \tag{5.5.5}$$

Its optimal value is denoted by $\underline{C}_{\text{LR}}(y, r, w) \leq C_{\text{LR}}(y, r, w)$, with equality when $C_{\text{LR}}(\cdot, r, w)$ is finite and l.s.c. at y. The dual parameters are r and w.

In the SRP maximisation programme (5.3.6)–(5.3.7), only k can serve as a primal parameter; and perturbation by the increment Δk yields the following dual programme for shadow-pricing the fixed inputs:

$$\text{Given } (p, k, w), \text{ minimise } \langle r \mid k \rangle + \Pi_{\text{LR}}(p, r, w) \text{ over } r. \tag{5.5.6}$$

Its optimal value is denoted by $\overline{\Pi}_{\text{SR}}(p, k, w) \geq \Pi_{\text{SR}}(p, k, w)$, with equality when with equality when $\Pi_{\text{SR}}(p, \cdot, w)$ is finite and upper semicontinuous (u.s.c.) at k.[28] The dual parameters are p and w.

The same programme for r—viz., (5.5.6) or (5.5.13)–(5.5.14) under c.r.t.s.—is also the dual of the *reduced* SRP programme in (5.4.2), again with k as the primal parameter. That is, the reduced and the full primal programmes have the same primal parameters and the same dual programme. Of course, the duality relationships cannot be exactly the same because the dual parameterisations are different: as has already been pointed out, the reduced primal programme has fewer variables, and hence fewer dual parameters, than the full programme (all of whose data are its primal and dual parameters). Since both programmes have the same data, the reduced one has therefore a datum that is neither a primal nor a dual parameter. In the case of the reduced SRP programme in (5.4.2), w is such a datum: the only primal parameter is k, and the only dual parameter is p (since y is the only primal variable). For comparison, in the full SRP programme (5.3.6)–(5.3.7) both p and w are dual parameters (paired to the primal variables y and v).[29]

The LRP maximisation programme (5.3.1)–(5.3.2) is, in this context, unusual because all its data (p, r and w) are dual parameters: no datum can serve as a primal parameter. This means that the intrinsic dual has no decision variable; formally, it is: given (p, r, w), minimise $\Pi_{\text{LR}}(p, r, w)$. Having no variable, the dual minimand is a constant, and it equals the primal value (Π_{LR}): since the dual is trivial, there can be no question of a duality gap in this case.

By contrast, the other programme pairs *can* have duality gaps, especially when the spaces are infinite-dimensional. But even then a gap can appear only at an exceptional data point: the primal and dual values are always equal under the generalised Slater's Condition of [73, (8.12)] or the

[28] As the notation indicates, $\overline{\Pi}$ and $\underline{C}$ are thought of mainly as dual expressions for Π and C (although duality of programmes is fully symmetric).

[29] A similar remark applies to the full and the reduced shadow-pricing programmes, (5.5.4) for (p, r) and that in (5.6.7) for p alone. Both are parameterised by w and have the same dual, viz., the SRC programme (5.3.10)–(5.3.11). All the vector data (y, k, w) are primal or dual parameters of the full programme (5.5.4) programme. But the datum k is not a primal or dual parameter of the reduced programme in (5.6.7).

compactness-and-continuity conditions of [73, Example 4' after (5.13)] and [73, Theorem 18' (d) or (e)]. In the problem of profit-maximising operation of a plant with capacity constraints (and no other fixed inputs), Slater's Condition requires only that the capacities be strictly positive, i.e., that $k \gg 0$; in other words, it is always met unless the plant k lacks a component: see [46].

Partial conjugacy relationships between the dual value functions ($\underline{C}_{\mathrm{SR}}$, $\underline{C}_{\mathrm{LR}}$, $\overline{\Pi}_{\mathrm{SR}}$, and $\overline{\Pi}_{\mathrm{LR}} = \Pi_{\mathrm{LR}}$) can be summarised in a diagram like (5.3.12) but with the arrows reversed and with bars added to the symbols Π and C):

$$\boxed{\begin{array}{ccccccc} & & & w & & & \\ & & & \Pi_{\mathrm{LR}} & & & \\ & & r & & p & & \\ & \swarrow & & & & \searrow & \\ & k & & & & y & \\ w\ \ \overline{\Pi}_{\mathrm{SR}} & & & & & & \underline{C}_{\mathrm{LR}}\ \ w \\ & p & & & & r & \\ & \searrow & & & & \swarrow & \\ & & y & & k & & \\ & & & \underline{C}_{\mathrm{SR}} & & & \\ & & & w & & & \end{array}} . \tag{5.5.7}$$

For example, the arrow from the p next to $\overline{\Pi}_{\mathrm{SR}}$ to the y next to $\underline{C}_{\mathrm{SR}}$ indicates that $\underline{C}_{\mathrm{SR}}$ is, as a function of y, the convex conjugate of $\overline{\Pi}_{\mathrm{SR}}$ as a function of p (with k and w fixed): i.e., by definition,

$$\underline{C}_{\mathrm{SR}}(y,k,w) = \sup_p \left\{ \langle p \mid y \rangle - \overline{\Pi}_{\mathrm{SR}}(p,k,w) \right\}. \tag{5.5.8}$$

In any specific case, formation of the primal-dual programme pair requires formulae for both $\mathbb{Y}$ and Π_{LR}. When the technology is given by a production set ($\mathbb{Y}$), this requires working out its support function (Π_{LR}). The task simplifies under c.r.t.s.: Π_{LR} is then $\delta(\cdot \mid \mathbb{Y}^\circ)$, the 0-$\infty$ indicator of the production cone's polar (5.3.4). In other words, $\mathbb{Y}^\circ$ is the implicit dual constraint set and, by making the constraint explicit, the dual programmes can be cast in the same form as the primals. For each primal, the general form of the dual is specialised to the case of c.r.t.s. in the same way, viz., by adjoining the constraint $(p,r,w) \in \mathbb{Y}^\circ$ and deleting the now-vanishing term Π_{LR} from (5.5.4), etc. So the dual programme is to impute optimal values to the given quantities by pricing them in a way consistent with the other, given prices, i.e., so that the entire price system lies in $\mathbb{Y}^\circ$.

Spelt out, under c.r.t.s., the dual to SRC minimisation is the following programme of maximising the output value less fixed-input value (OFIV) by shadow-pricing both the outputs and the fixed inputs:

$$\text{Given } (y,k,w), \text{ maximise } \langle p \mid y \rangle - \langle r \mid k \rangle \text{ over } (p,r) \tag{5.5.9}$$

$$\text{subject to } (p,r,w) \in \mathbb{Y}^\circ. \tag{5.5.10}$$

The dual to LRC minimisation is (with c.r.t.s.) the following programme of maximising the output value (OV) by shadow-pricing the outputs:

$$\text{Given } (y, r, w)\,, \text{ maximise } \langle p \,|\, y \rangle \text{ over } p \tag{5.5.11}$$

$$\text{subject to } (p, r, w) \in \mathbb{Y}^{\circ}. \tag{5.5.12}$$

The dual to SRP maximisation is (under c.r.t.s.) the following programme of minimising the total fixed-input value (FIV) by shadow-pricing the fixed inputs:

$$\text{Given } (p, k, w)\,, \text{ minimise } \langle r \,|\, k \rangle \text{ over } r \tag{5.5.13}$$

$$\text{subject to } (p, r, w) \in \mathbb{Y}^{\circ}. \tag{5.5.14}$$

The dual to LRP maximisation has no decision variable, and, with c.r.t.s., it may be thought of as a *price consistency check*: its value is 0 if $(p, r, w) \in \mathbb{Y}^{\circ}$, and $+\infty$ otherwise. Formally, the dual is:

$$\text{Given } (p, r, w)\,, \text{ minimise } 0 \text{ subject to } (p, r, w) \in \mathbb{Y}^{\circ}. \tag{5.5.15}$$

Thus, with c.r.t.s., the dual objectives are "automatic", and formation of the dual programmes boils down to working out $\mathbb{Y}^{\circ}$ from a specific cone $\mathbb{Y}$. One framework for this is provided in Section 5.12.

Like the primals, the dual programmes are henceforth named after their objectives, OFIV, OV and FIV. Strictly speaking, this terminology fits only the case of c.r.t.s. for the long run (i.e., the case of a production cone). But it is used also when c.r.t.s. are not assumed (e.g., in Figure 5.2, Section 5.6 and Tables 5.1 and 5.2).

5.6 The SRP and SRC optimisation systems

The use of the conjugacy (5.3.14) between the SRP and the LRP gives a characterisation of the profit-maximising investment in terms of its imputed values, i.e., it reformulates the investment-optimality condition (5.4.1) as the valuation condition (5.4.4). The valuation programme in (5.4.4) is subsequently obtained as the dual (5.5.6), or (5.5.13)–(5.5.14) under c.r.t.s., to the short-run profit maximisation programme (5.3.6)–(5.3.7), which appears in (5.4.2)–(5.4.3) in a split form. Thus the use of conjugacy produces the system (5.4.2)–(5.4.5) of optimality conditions on y, v and r; and the use of duality shows that this system means that (y, v) and r form a pair of solutions to the SRP programme and its dual.[30] Similar arguments lead to characterisations of optimality in terms of the

[30]These arguments exploit the subprogramme concept as well as that of duality, i.e., Π_{SR} is viewed in two ways: (i) as the value of a subprogramme, and (ii) as the primal value. Both contexts give rise to the conjugacy between Π_{SR} and Π_{LR}—and that is why there are two ways of deriving the valuation programme in (5.4.4). In detail, since Π_{SR} is the value of the subprogramme of LRP maximisation obtained by fixing k, its (concave) conjugate w.r.t. k is $-\Pi_{LR}$ as a function of r: this is (5.3.14). It follows, by (C.5.5) and (C.6.2), that k solves the "conjugacy programme" in (5.4.1) if and only if r solves the "reverse" one in (5.4.4) and (5.4.5) holds. The same programme for r can be derived independently as the dual to SRP maximisation parameterised by k, as is done in Proposition 5.16.1 (which

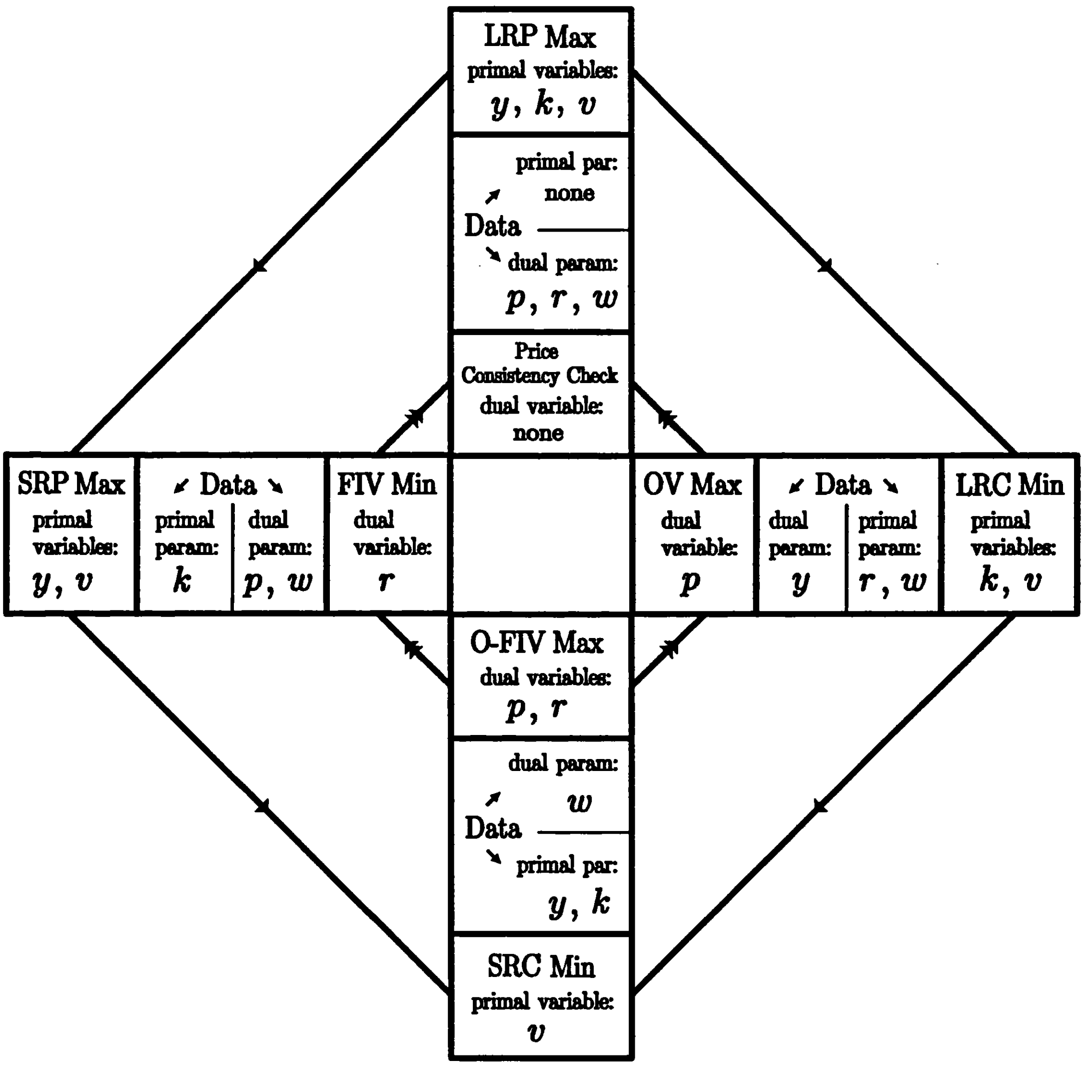

FIGURE 5.2. Decision variables and parameters for primal programmes (optimisation of long-run profit, short-run profit, long-run cost, short-run cost) and for dual programmes (price consistency check, optimisation of: fixed-input value, output value, output value less fixed-input value). In each programme pair, the same prices and quantities—(p, y) for outputs, (r, k) for fixed inputs, and (w, v) for variable inputs—are differently partitioned into decision variables and data (which are subdivided into primal and dual parameters). Arrows lead from programmes to subprogrammes.

SRC programme with its dual, i.e., each of the following two systems of conditions is equivalent to maximisation of long-run profit at prices (p, r, w) by an input-output bundle $(y, -k, -v)$.

The **SRP optimisation system:** $(y, -v)$ maximises the short-run profit at prices (p, w), and r minimises the value of the fixed-input k (plus maximum LRP if r.t.s. are decreasing), and the two optimal values are equal (i.e., under c.r.t.s., maximum SRP equals minimum FIV). Formally:

$$(y, v) \text{ solves the primal SRP programme (5.3.6)–(5.3.7), given } (p, k, w)\,. \tag{5.6.1}$$

$$r \text{ solves the dual (5.5.6), which is (5.5.13)–(5.5.14) under c.r.t.s., given } (p, k, w)\,. \tag{5.6.2}$$

$$\overline{\Pi}_{\mathrm{SR}}(p, k, w) = \Pi_{\mathrm{SR}}(p, k, w)\,. \tag{5.6.3}$$

The **SRC optimisation system:** v minimises the short-run cost at price w, and (p, r) maximises the value of output y less that of fixed-input k (and less maximum LRP under d.r.t.s.), and the two optimal values are equal (i.e., under c.r.t.s., minimum SRC equals maximum OFIV). Formally:

$$(p, r) \text{ solves the dual (5.5.4), a.k.a. (5.5.9)–(5.5.10) under c.r.t.s., given } (y, k, w)\,. \tag{5.6.4}$$

$$v \text{ solves the primal SRC programme (5.3.10)–(5.3.11), given } (y, k, w)\,. \tag{5.6.5}$$

$$\underline{C}_{\mathrm{SR}}(y, k, w) = C_{\mathrm{SR}}(y, k, w)\,. \tag{5.6.6}$$

Additionally, one can split the joint programme for two decision variables: just as (5.3.6)–(5.3.7) has been split into (5.4.2) and (5.4.3), so the joint programme (5.5.4) for (p, r) can be replaced by two programmes for p and r separately. Condition (5.6.4) is therefore equivalent to:[31]

$$p \text{ maximises } \langle \cdot \mid y \rangle - \overline{\Pi}_{\mathrm{SR}}(\cdot, k, w) \text{ on } P \text{ (given } y, k \text{ and } w) \tag{5.6.7}$$

$$r \text{ solves (5.5.6), given } (p, k, w)\,. \tag{5.6.8}$$

Thus the joint shadow-pricing programme (5.5.4) for (p, r) is reduced to an output-pricing programme, for p alone, by first solving the fixed-input shadow-pricing programme (5.5.6) for r and substituting its optimal value ($\overline{\Pi}_{\mathrm{SR}}$) for the term $\langle r \mid k \rangle + \Pi_{\mathrm{LR}}(p, r, w)$ in (5.5.4). In other words, two-stage solving means in this case:

1. first minimising $\langle r \mid k \rangle + \Pi_{\mathrm{LR}}(p, r, w)$ over r (or, under c.r.t.s., minimising $\langle r \mid k \rangle$ over r subject to $(p, r, w) \in \mathbb{Y}^{\circ}$) to find the solution set $\hat{R}(p, k, w)$, or *the* solution $\hat{r}(p, k, w)$ if it is indeed unique, and the minimum value $\overline{\Pi}_{\mathrm{SR}}(p, k, w)$, which is $\langle \hat{r} \mid k \rangle$;

also identifies p and w as the dual parameters). Alternatively, it can be identified as the dual by using the conjugacy between Π_{SR} and Π_{LR}: it is a foundation of duality for CPs that the (concave) conjugate of the primal maximum value (as a function of the primal parameter) plus the primal parameter times the dual variable (here, $\Pi_{\mathrm{LR}}(r) + \langle r \mid k \rangle$) is the dual minimand. See, e.g., [73, Theorem 7], which here must be applied to the function $\Delta k \mapsto \Pi_{\mathrm{SR}}(k + \Delta k)$ as Rockafellar's primal value (his is a function of the parameter increment, rather than of the parameter point like ours, and this shifts the argument by k and adds the term $\langle r \mid k \rangle$ to the conjugate).

[31] The maximum value in (5.6.7) is $\underline{C}_{\mathrm{SR}}(y, k, w)$, by the definitions of $\overline{\Pi}_{\mathrm{SR}}$ and $\underline{C}_{\mathrm{SR}}$ as the optimal values of (5.5.6) and (5.5.4).

2. then maximising $\langle p \,|\, y\rangle - \overline{\Pi}_{\mathrm{SR}}(p,k,w)$ over p to find the solution set $\check{P}(y,k,w)$, or *the* solution $\check{p}(y,k,w)$, should it be unique. This gives every complete solution (in terms of y, k and w) as a (p,r) such that $p \in \check{P}(y,k,w)$ and $r \in \hat{R}(p,k,w)$. Should the solution be unique, it is the pair $\check{p}(y,k,w)$ and $\hat{r}(\check{p}(y,k,w),k,w)$.

The two systems (5.6.1)–(5.6.3) and (5.6.4)–(5.6.6) are called the *SRP* and *SRC optimisation systems* because each is put entirely in terms of solutions to the named programme and its dual. Each system contains a joint programme, which can be split to produce the corresponding *split optimisation system.*

The **split SRP optimisation system** is (5.4.2)–(5.4.5).

The **split SRC optimisation system** is (5.6.5)–(5.6.8).

The first of these, the split SRP system (5.4.2)–(5.4.5), has been introduced before the programme for r in (5.4.4) could be formally identified as the dual of the SRP programme (in Section 5.5). In (5.6.2), the same programme is referred to as the dual. So the split SRP optimisation system can now be restated as the conjunction of (5.4.2)–(5.4.3) and (5.6.2)–(5.6.3).

Comments (on the equivalence and structure of the SRP and SRC optimisation systems):

- Another proof of equivalence, to LRP maximisation, of the two systems (5.6.1)–(5.6.3) and (5.6.4)–(5.6.6) follows from a general inequality between the values of a programme pair (taking for granted that (5.5.4) and (5.5.6) are indeed the relevant duals, as is stated and proved in Sections 5.5 and 5.16). What is to be shown is that each of the two systems is equivalent to (5.3.3), or to the Complementarity Conditions (5.3.5) in the case of c.r.t.s. For each programme pair, (5.3.3) or (5.3.5) means: (i) primal feasibility, of either (y,v) or v, (ii) dual feasibility, of either r or (p,r), and (iii) equality of the primal maximand to the dual minimand, at the two points in question. So it suffices to note that these *FFE Conditions* (which have already appeared as (5.5.3) in the LP context) fully characterise a pair of solutions with equal values. And this is because the primal maximand never exceeds the dual minimand (at feasible points).

- Thus the data (p,r,w) and the solution $(y,-k,-v)$ of the LRP programme (5.3.1)–(5.3.2) can be permuted to form the data and solutions to the SRP or SRC subprogramme with its dual (when there is no duality gap). In either case, a pair of solutions gives three of the six variables—one from each of the three price-quantity pairs (viz., (p,y) for outputs, (r,k) for fixed inputs, and (w,v) for variable inputs)—in terms of the other three (which are parameters, not decision variables).

5.7 The SRC-P saddle differential system: a partial subdifferential system for the short-run approach

In convex programming, optimality is fully expressed by the first-order condition. Furthermore, by combining the FOC with the Inversion Rule for the derivative of a conjugate function, the optimal solution can be interpreted as a marginal value. This *derivative property* of the optimal-value function extends to the case of nonunique solutions. The value is then nondifferentiable in the ordinary way, but it has a generalised, multi-valued derivative. For a convex function, this is the *subdifferential* (a.k.a. the subgradient set), defined by (C.3.1) and denoted by ∂. The *superdifferential* of a concave function, denoted here by $\widehat{\partial}$, is defined by (C.5.4). Each of the functions Π_{SR}, C_{SR} and C_{LR} is either convex or concave jointly in two of its three variables, and it is concave or convex in the other variable. For example, $\Pi_{\mathrm{SR}}(p,k,w)$ is jointly convex in (p,w), and concave in k (as is $\overline{\Pi}_{\mathrm{SR}}$).

The split LRP optimisation system (5.4.1)–(5.4.3) is thus transformed into the partial subdifferential system that consists of the FOCs for (5.4.1) and (5.4.2) and of the derivative property of C_{SR} as the optimal value of (5.4.3). This gives the **SRC-P saddle differential system**

$$r \in \widehat{\partial}_k \Pi_{\mathrm{SR}}(p,k,w) \tag{5.7.1}$$

$$p \in \partial_y C_{\mathrm{SR}}(y,k,w) \tag{5.7.2}$$

$$v \in \widehat{\partial}_w C_{\mathrm{SR}}(y,k,w). \tag{5.7.3}$$

It is called the *SRC-P saddle differential system* because it uses $\partial_y C_{\mathrm{SR}}$ and $\widehat{\partial}_w C_{\mathrm{SR}}$, the partial sub/super-differentials of C_{SR} as a saddle (convex-concave) function of (y,w), in addition to using $\widehat{\partial}_k \Pi_{\mathrm{SR}}$. A similar use of C_{SR}, as a saddle function of (k,w), arises later in the L-SRC system (5.9.8)–(5.9.10): the affices "L" and "P" in these names stand for "long-run" and "profit".

Comments (use of a differential condition to absorb a no-gap condition):

- The system (5.7.1)–(5.7.3) can be derived also from the split SRP optimisation system (5.4.2)–(5.4.5). The FOC for (5.4.2) and the derivative property of C_{SR} as the value function for (5.4.3) are used just as before. But, instead of the FOC for (5.4.1), this time the third condition is the derivative property of $\overline{\Pi}_{\mathrm{SR}}$ as the value function for (5.4.4) or (5.5.6), i.e., that $r \in \widehat{\partial}_k \overline{\Pi}_{\mathrm{SR}}(p,k,w)$. Taken together, this and (5.4.5) mean exactly that $r \in \widehat{\partial}_k \Pi_{\mathrm{SR}}$, since (5.4.5) means that $\overline{\Pi}_{\mathrm{SR}} = \Pi_{\mathrm{SR}}$, at (p,k,w).

- The last argument is a case of absorbing a no-gap condition in a subdifferential condition by changing the derivative from Type Two (here, $\widehat{\partial}_k \overline{\Pi}_{\mathrm{SR}}$) to Type One ($\widehat{\partial}_k \Pi_{\mathrm{SR}}$). This is done by changing the value function either from dual to primal (if the parameter in question is primal like the k here), or *vice versa*. The optimal solution is *always* equal to the marginal value of the programme being solved; this is a derivative of Type Two. It is actually of Type One—i.e., it is the marginal value of the programme *dual* to that being solved—if there is no duality gap. But if there is a gap, the Type One derivative does not exist. In the above case of fixed-input valuation, the set of solutions, for r, of (5.4.4) or (5.5.6) is always identical to $\widehat{\partial}_k \overline{\Pi}_{\mathrm{SR}}$ (which is a derivative of Type Two). It equals $\widehat{\partial}_k \Pi$ (a derivative of Type One) if $\Pi_{\mathrm{SR}} = \overline{\Pi}_{\mathrm{SR}}$ at the given (p, k, w). But if $\Pi_{\mathrm{SR}} \neq \overline{\Pi}_{\mathrm{SR}}$ then $\widehat{\partial}_k \Pi = \emptyset$ (the empty set); so if $r \in \widehat{\partial}_k \Pi_{\mathrm{SR}}$ then $\Pi_{\mathrm{SR}} = \overline{\Pi}_{\mathrm{SR}}$ (at the given p, k and w).

5.8 Summary of systems characterising a long-run producer optimum

The same arguments—viz., the derivative property of the optimal value function and the FOC—can be applied to the other optimisation systems of Section 5.6 to produce another five differential systems. These are given in Tables 5.1 and 5.2 along with the five systems already introduced. (The five new systems are: the SRP and FIV subdifferential systems obtained from the SRP optimisation system, the SRC and OFIV subdifferential systems obtained from the SRC optimisation system, and the FIV saddle differential system obtained from the split SRC optimisation system.) As Tables 5.1 and 5.2 also indicate, the partial subdifferential systems can also be derived from systems with joint subdifferentials, by applying the Subdifferential Sections Lemma (SSL, i.e., Lemma C.7.2 in Appendix C) or the Partial Inversion Rule or its dual variant (PIR and DPIR, i.e., Corollaries C.7.3 and C.7.5). Details are given in [46].

Thus Tables 5.1 and 5.2 summarise ten duality-based systems and proofs of their equivalence. (Since the top right entry of the one table is identical to the bottom right of the other, the twelve entries include two repetitions.) In Section 5.9, another system is spelt out, and six more are mentioned. All seven of those use the LRC programme and its dual or their value functions; they are mirror images of the systems shown in the two tables, from which they can be obtained by replacing $\Pi_{\mathrm{SR}}(p, k)$ with $C_{\mathrm{LR}}(y, r)$ and changing signs where needed.[32] In other words, Tables 5.1 and 5.2 deal explicitly with the values and programmes in the left halves of the conjugacy diagrams (5.3.12) and (5.5.7), but the analysis applies equally to the right halves.

In differential systems, the Type One derivatives that exclude duality gaps are identified. In optimisation systems, the various duals are referred to as "optimisation of the fixed quantities' value", although this name fits only the case of c.r.t.s. (which need not be assumed). The constraint sets ($\mathbb{Y}$ and $\mathbb{Y}^{\circ}$, under c.r.t.s.) are not shown.

Comment (partition into a short-run subsystem and a supplementary condition): Seven of the ten systems in Tables 5.1 and 5.2—all except for the three that appear on the left in Table 5.2—contain a condition on r and (p, k, w) that is either exactly or at least nearly equivalent to k being a profit-

[32] The three systems on the left in Table 5.2 do not yield new ones (when Π_{SR} is replaced by C_{LR}) simply because they do not involve Π_{SR} at all. So there are not ten but seven of the "mirror images".

SRP subdiff. sys.		SRC-P saddle diff. sys. (5.7.1)–(5.7.3)
$(y, -v) \in \partial_{p,w} \Pi_{\mathrm{SR}}$ (Type Two) $r \in \widehat{\partial}_k \Pi_{\mathrm{SR}}$ (Type One)	Dual Part. Inv. Rule $\Longleftrightarrow$	$p \in \partial_y C_{\mathrm{SR}}$ $v \in \widehat{\partial}_w C_{\mathrm{SR}}$ $r \in \widehat{\partial}_k \Pi_{\mathrm{SR}}$ (Type One)
$\Updownarrow$ Deriv. Prop. of Opt. Val. (twice) Absorption of No-Gap Cond.		First-Order Condition $\Updownarrow$ Deriv. Prop. of Opt. Val. (twice) Absorption of No-Gap Cond.
SRP opt. sys. (5.6.1)–(5.6.3)		split SRP opt. sys. (5.4.2)–(5.4.5)
(y, v) maxi'es short-run profit r minimises fixed-input value $\overline{\Pi}_{\mathrm{SR}} = \Pi_{\mathrm{SR}}$ at (p, k, w)	Two-stage solving $\Longleftrightarrow$	y maximises revenue less C_{SR} v minimises short-run cost r minimises fixed-input value $\overline{\Pi}_{\mathrm{SR}} = \Pi_{\mathrm{SR}}$ at (p, k, w)
$\Updownarrow$ Deriv. Prop. of Opt. Val. (twice) Absorption of No-Gap Cond.		
FIV subdiff. sys.		FIV saddle diff. sys.
$(y, -v) \in \partial_{p,w} \overline{\Pi}_{\mathrm{SR}}$ (Type One) $r \in \widehat{\partial}_k \overline{\Pi}_{\mathrm{SR}}$ (Type Two)	Subdiff. Sect. Lem. $\Longleftrightarrow$	$y \in \partial_p \overline{\Pi}_{\mathrm{SR}}$ $v \in \widehat{\partial}_w \underline{C}_{\mathrm{SR}}$ (Type One) $r \in \widehat{\partial}_k \overline{\Pi}_{\mathrm{SR}}$

TABLE 5.1. The SRP optimisation system with its split form, and four derived differential systems (of which three follow directly by the DP and FOC, and one indirectly by using additionally the SSL).

maximising investment at prices (p, r, w), i.e., to (5.4.1). The condition in question is: $r \in \widehat{\partial}_k \Pi_{\mathrm{SR}}$, or $r \in \widehat{\partial}_k \overline{\Pi}_{\mathrm{SR}}$, or "$r$ minimises FIV". Together, the system's other conditions—on p, y, w, v and k—are then essentially equivalent to (5.4.2)–(5.4.3), i.e., to $(y, -v)$ being a short-run profit-maximising input-output bundle at prices (p, w), given capital inputs k. This *short-run subsystem* is to be solved for v and either y or p—given w and either p or y, as well as k. The remaining *supplementary condition* involves r and essentially means that investment is at a profit maximum. This partition of a system is examined in detail in [46].

5.9 Extended Wong-Viner Theorem and other transcriptions from SRP to LRC

The preceding analysis can be re-applied to SRC minimisation as a subprogramme of LRC minimisation, instead of SRP maximisation. As part of this, the Subdifferential Sections Lemma (Lemma C.7.2) can be applied to C_{SR} as the bivariate convex "parent" function of the saddle function C_{LR}, instead of the saddle function Π_{SR} as is done at the bottom of Table 5.2. That is, both Π_{SR} and C_{LR} can be viewed as partial conjugates of C_{SR}. This shows that, with w fixed and suppressed

OFIV subdiff. sys.		FIV saddle diff. sys.
$v \in \widehat{\partial}_w \underline{C}_{\mathrm{SR}}$ (Type One) $(p, -r) \in \partial_{y,k} \underline{C}_{\mathrm{SR}}$ (Type Two)	Dual Part. Inv. Rule $\Longleftrightarrow$	$y \in \partial_p \overline{\Pi}_{\mathrm{SR}}$ $v \in \widehat{\partial}_w \underline{C}_{\mathrm{SR}}$ (Type One) $r \in \widehat{\partial}_k \overline{\Pi}_{\mathrm{SR}}$
$\Updownarrow$ Deriv. Prop. of Opt. Val. (twice) Absorption of No-Gap Cond.		First-Order Condition $\Updownarrow$ Deriv. Prop. of Opt. Val. (twice) Absorption of No-Gap Cond.
SRC opt. sys. (5.6.4)–(5.6.6)		split SRC opt. sys. (5.6.5)–(5.6.8)
v minimises short-run cost (p, r) maxs rev. − fix.-inp. val. $\underline{C}_{\mathrm{SR}} = C_{\mathrm{SR}}$ at (y, k, w)	Two-stage solving $\Longleftrightarrow$	p maximises revenue less $\overline{\Pi}_{\mathrm{SR}}$ v minimises short-run cost r minimises fixed-input value $\underline{C}_{\mathrm{SR}} = C_{\mathrm{SR}}$ at (y, k, w)
$\Updownarrow$ Deriv. Prop. of Opt. Val. (twice) Absorption of No-Gap Cond.		
SRC subdiff. sys.		SRC-P saddle diff. sys. (5.7.1)–(5.7.3)
$v \in \widehat{\partial}_w C_{\mathrm{SR}}$ (Type Two) $(p, -r) \in \partial_{y,k} C_{\mathrm{SR}}$ (Type One)	Subdiff. Sect. Lem. $\Longleftrightarrow$	$p \in \partial_y C_{\mathrm{SR}}$ $v \in \widehat{\partial}_w C_{\mathrm{SR}}$ $r \in \widehat{\partial}_k \Pi_{\mathrm{SR}}$ (Type One)

TABLE 5.2. The SRC optimisation system with its split form, and four derived differential systems (of which three follow directly by the DP and FOC, and one indirectly by using additionally the SSL).

from the notation,

$$\left.\begin{array}{r} p \in \partial_y C_{\mathrm{SR}}(y,k) \\ \\ r \in \widehat{\partial}_k \Pi_{\mathrm{SR}}(p,k) \end{array}\right\} \Leftrightarrow (p,-r) \in \partial_{y,k} C_{\mathrm{SR}}(y,k) \Leftrightarrow \left\{\begin{array}{l} p \in \partial_y C_{\mathrm{LR}}(y,r) \\ \\ r \in -\partial_k C_{\mathrm{SR}}(y,k) \end{array}\right. . \tag{5.9.1}$$

This is the *Extended Wong-Viner Theorem.* Note that the condition that $r \in -\partial_k C_{\mathrm{SR}}$ is the FOC for k to yield the infimum in the definitional formula

$$C_{\mathrm{LR}}(y,r,w) = \inf_k \{\langle r \mid k \rangle + C_{\mathrm{SR}}(y,k,w)\} \tag{5.9.2}$$

(which means that C_{LR} is, as a function of r, the concave conjugate of $-C_{\mathrm{SR}}$ as a function of k, with y and w fixed).

For comparison, the usual Wong-Viner Envelope Theorem for differentiable costs gives

$$\left.\begin{array}{r} p = \nabla_y C_{\mathrm{SR}}(y,k) \\ \\ r = -\nabla_k C_{\mathrm{SR}}(y,k) \text{ i.e., } k \text{ yields the inf in (5.9.2)} \end{array}\right\} \Rightarrow p = \nabla_y C_{\mathrm{LR}}(y,r). \tag{5.9.3}$$

Comparisons with the two "outer" systems in (5.9.1) show that their equivalence is indeed an extension of (5.9.3). This is because

$$\widehat{\partial}_k \Pi_{\mathrm{SR}}(p,k) \subseteq -\partial_k C_{\mathrm{SR}}(y,k) \quad \text{when } p \in \partial_y C_{\mathrm{SR}}(y,k) \tag{5.9.4}$$

i.e., when y yields the supremum in (5.3.13).[33] In the differentiable case, the inclusion (5.9.4) reduces to the equality $\nabla_k \Pi_{\mathrm{SR}} = -\nabla_k C_{\mathrm{SR}}$ (when $p = \nabla_y C_{\mathrm{SR}}$), and thus (5.9.1) becomes:

$$\text{if } r = -\nabla_k C_{\mathrm{SR}}(y,k) \text{ then } (p = \nabla_y C_{\mathrm{SR}}(y,k) \Leftrightarrow p = \nabla_y C_{\mathrm{LR}}(y,r)) \tag{5.9.5}$$

which is the usual Wong-Viner Theorem.

Comment (failure of naive extension): The Wong-Viner Theorem *cannot* be extended to the general, subdifferentiable case simply by transcribing the ∇'s to ∂'s in (5.9.5) or (5.9.3) because, even when $r \in -\partial_k C_{\mathrm{SR}}(y,k)$,

$$p \in \partial_y C_{\mathrm{SR}}(y,k) \not\Rightarrow p \in \partial_y C_{\mathrm{LR}}(y,r). \tag{5.9.6}$$

It is the *reverse* inclusion that always holds, i.e.,

$$\text{if } r \in -\partial_k C_{\mathrm{SR}}(y,k) \text{ then } \partial_y C_{\mathrm{LR}}(y,r) \subseteq \partial_y C_{\mathrm{SR}}(y,k) \tag{5.9.7}$$

but the inclusion is generally strict (i.e., $\partial_y C_{\mathrm{LR}} \neq \partial_y C_{\mathrm{SR}}$).[34] The extension (5.9.1) succeeds because it strengthens the insufficient condition $r \in -\partial_k C_{\mathrm{SR}}$ in (5.9.6) to $r \in \widehat{\partial}_k \Pi_{\mathrm{SR}}$ (this is stronger because the inclusion in (5.9.4) is usually strict, when C_{SR} is nondifferentiable). The peak-load pricing example of Section 5.2 provides a simple, yet extreme, illustration: when $r > 0$, that $r \in -\partial_k C_{\mathrm{SR}}(y,k,w)$ says merely that $k = \sup_t y(t)$; it says nothing about p or r. For comparison, the condition $r = \partial \Pi_{\mathrm{SR}} / \partial k = \int (p(t) - w)^+ \, \mathrm{d}t$ links p to r (and w), as well as implying that $\mathrm{Sup}(y) = k$ (if $p \in \partial_y C_{\mathrm{SR}}(y,k,w)$, i.e., if: $y(t) = k$ when $p(t) > w$, and $y(t) = 0$ when $p(t) < w$). It is therefore a much stronger condition, and it helps determine p in terms of r (and w and y). That it is strong enough to ensure that $p \in \partial_y C_{\mathrm{SR}}(y,k,w) \Rightarrow p \in \partial_y C_{\mathrm{LR}}(y,r,w)$ can also, in that example, be checked by calculating both subdifferentials explicitly.

It follows from (5.9.1) that LRP maximisation, being equivalent to (5.7.1)–(5.7.3), is also equivalent to the **L-SRC saddle differential system**

$$p \in \partial_y C_{\mathrm{LR}}(y,r,w) \tag{5.9.8}$$

$$r \in -\partial_k C_{\mathrm{SR}}(y,k,w) \tag{5.9.9}$$

$$v \in \widehat{\partial}_w C_{\mathrm{SR}}(y,k,w). \tag{5.9.10}$$

It is called the *L-SRC saddle differential system* because it uses $\partial_k C_{\mathrm{SR}}$ and $\widehat{\partial}_w C_{\mathrm{SR}}$, the partial sub/super-differentials of C_{SR} as a saddle (convex-concave) function of (k,w), in addition to using

[33] The inclusion (5.9.4) follows directly from (5.3.13) by Remark C.7.4 (applied to the saddle function Π_{SR} as a partial conjugate of C_{SR}).

[34] The inclusion (5.9.7) follows directly from (5.9.2) by Remark C.7.4 (applied to the saddle function C_{LR} as a partial conjugate of C_{SR}).

$\partial_y C_{\mathrm{LR}}$. It is the "mirror image" of the SRC-P saddle differential system (5.7.1)–(5.7.3), so it can be obtained by re-applying the same arguments (with LRC instead of SRP). It can also be derived from the SRC subdifferential system (which is shown in Table 5.2), by using the second equivalence in (5.9.1).

When the producer is a public utility, LRMC pricing and LRC minimisation—i.e., Conditions (5.9.8) to (5.9.10)—are often taken as the definition of a long-run producer optimum. If the SRC function is simpler than the LRC function (as is usually the case), and the SRP function is also simple, then the Extended Wong-Viner Theorem (5.9.1) can facilitate the short-run approach by characterising optimality in terms of the SRC and SRP functions. This has been used in the introductory peak-load pricing example of Section 5.2). In that problem, the cost-minimising inputs were obvious, but the question was how to ensure, by a simple condition put in terms of a short-run value function, that an SRMC output price was actually an LRMC price, i.e., that it met (5.9.8). This was achieved by employing the special case (5.2.2) of (5.7.1), i.e., of the condition that $r \in \widehat{\partial}_k \Pi_{\mathrm{SR}}$. Thus the argument was a case of the Extended Wong-Viner Theorem or, in other words, of the equivalence of (5.7.1)–(5.7.3) to (5.9.8)–(5.9.10).

Like (5.7.1)–(5.7.3), the other optimisation and differential systems shown on the right in Table 5.2, and all those in Table 5.1, can also be transcribed into equivalent characterisations of a long-run producer optimum by replacing the SRP with the LRC.[35] The three systems shown on the left in Table 5.1 transcribe into:

The **LRC subdifferential system** (which is a transcription of the SRP subdifferential system).

LRC optimisation system (which is a transcription of the SRP optimisation system (5.6.1)–(5.6.3)).

OV subdifferential system (which is a transcription of the FIV subdifferential system).

And, just as (5.7.1)–(5.7.3) transcribes into (5.9.8)–(5.9.10), so the other three systems shown on the right in Tables 5.1 and 5.2 transcribe into:

The **split LRC optimisation system** (which is a transcription of the split SRP optimisation system (5.4.2)–(5.4.5)).

The **OV saddle differential system** (which is a transcription of the FIV saddle differential system).

The **reverse-split SRC optimisation system**, which is so called to distinguish it from the split SRC optimisation system (5.6.5)–(5.6.8), of which it is a transcription. (The two systems differ only in the order in which p and r are optimised when the joint programme (5.5.4) is split into two stages.)

[35] In detail, this is done by swapping p with $-r$ and y with k, and by replacing the function $(p, k) \mapsto \Pi_{\mathrm{SR}}(p, k)$ with the function $(y, -r) \mapsto C_{\mathrm{LR}}(y, r)$.

5.10 Outline of the short-run approach to long-run general equilibrium

The preceding characterisations of long-run producer optima can serve various purposes; ours is the short-run approach to long-run general equilibrium (LRGE). This means that the capital inputs k are kept fixed at the stage of calculating the equilibrium in the products' market. The variable-input prices w are assumed to be fixed throughout this analysis (although this is not at all essential, and w might instead be determined in equilibrium just like the output prices p). This leaves two alternative ways to handle the supply side of the short-run general equilibrium (SRGE) problem, and hence two varieties of the short-run approach to long-run producer optimum and general equilibrium:

1. In the *short-run profit approach*, the output and variable-input quantities $\hat{y}$ and $\check{v}$, and the fixed-input values $\hat{r}$, are derived from any given p, k and w (usually by solving the SRP problem (5.3.6)–(5.3.7) and its dual (5.5.6) or (5.5.13)–(5.5.14) under c.r.t.s.). The supply $\hat{y}(p,k,w)$ is then equated to demand $\hat{x}(p)$ to determine the short-run equilibrium price system $p^{\star}_{\mathrm{SR}}(k)$, which depends also on w. This stage corresponds to the inner loop in Figure 5.3, if an iterative method is used to solve the demand-supply equation for p.[36] The capital inputs' marginal values $\hat{r}(p^{\star}_{\mathrm{SR}}(k,w),k,w)$, imputed at the short-run equilibrium prices, are then equated to their *given*, fixed rental prices r^{F} to determine, by solving for k, the (long-run) equilibrium capacities $k^{\star}(r^{\mathrm{F}},w)$. This also gives the long-run equilibrium price system $p^{\star}_{\mathrm{LR}}(r^{\mathrm{F}},w) = p^{\star}_{\mathrm{SR}}(k^{\star}(r^{\mathrm{F}},w),w)$. This stage corresponds to the outer loop in Figure 5.3, if an iterative method is used to solve the price-value equation for k.

2. In the *short-run cost approach*, the variable-input quantities $\check{v}$, and the shadow prices for outputs and fixed inputs—i.e., a typically nonunique $p \in \check{P}(y,k,w)$ with the associated, typically unique $\hat{r}(p,k,w)$—are derived from any given y, k and w (usually by solving the SRC problem (5.3.10)–(5.3.11) and its dual (5.5.4) or (5.5.9)–(5.5.10) under c.r.t.s.). To find the short-run equilibrium, inverse demand is then required to equal one of the typically nonunique output price systems that solve the short-run output-pricing programme in (5.6.7). This a subprogramme of (5.5.4); its solution set $\check{P}(y,k,w)$ consists essentially of SRMCs (see (5.10.3) for details). Finally, the long-run equilibrium capacities, and hence also the output prices, are found just as in the profit approach.

In principle, the duality theory of convex programming can be brought to bear however the commodities are divided into "variable" quantities with given prices and "fixed", unpriced quantities: in studying the producer optimum, the roles of prices and quantities are formally symmetric. At an abstract level, therefore, there is no reason to prefer any particular programme pair or the associated

[36] In finding $p^{\star}_{\mathrm{SR}}$ by Walrasian tatonnement, a manageable difficulty arises from discontinuity of supply when it is only an upper hemicontinuous correspondence (as in Figure 5.1a). With a continuous (single-valued) demand map, this is not much of a complication.

functional representation of the technology (by Π_{SR}, C_{LR} or C_{SR}, etc.). But the classification of commodities as "fixed in the short run" is not arbitrary and nominal but mostly real and objective: these are capital goods and natural resources. Their quantities (k) must be taken as known throughout the short-run analysis. Additionally, some of those quantities to be determined in the SRGE, such as the outputs (y), might also be taken as known at the earlier stage of finding the short-run producer optimum and the shadow prices: this would mean solving the SRC programme (for v) with its dual (for p and r). But this is disadvantageous analytically because, when the capital inputs (k) impose capacity constraints on a cyclic output (y), it results in dual solutions so indeterminate that they form an unbounded set: if not only r but also p are unknowns, then almost nothing can be said about capacity charges (which are terms of p, and give r as their total over the cycle). Another disadvantage of the SRC approach, which emerges only at the equilibrium stage, is that it entails working with the inverse supply maps ($\check{P}_\theta$) and "equating" each of these to inverse demand to find the SRGE output bundle ($y^\star_{\mathrm{SR}\,\theta}$) of each individual producer θ—from the inclusion a.k.a. "generalised equation" (5.10.2) below. This is usually much harder than simply adding up all the direct supply maps ($\hat{Y}_\theta$), equating their sum to demand, and solving (5.10.2) for the single market price system ($p^\star_{\mathrm{SR}}$)—which is what the SRP approach requires. In addition, unlike the multi-valued inverse supply map ($\check{P}_\theta$), the direct supply may well be a single-valued map ($\hat{y}_\theta$), in which case the relevant inclusion (5.10.2) is an ordinary equation.

In summary, it is better not to fix any more quantities than is necessary—and this means using the SRP rather than the SRC approach. The profit approach is likely to be more workable because it has two advantages over the cost approach: (i) determinacy of solutions to the short-run producer problem and its dual, and (ii) reduction of the number of unknowns in the subsequent equilibrium problem. Both are detailed next.

The first advantage is simply the convenience of dealing mostly with single-valued maps rather than multi-valued correspondences. Solutions for (p, r) to the dual (5.5.4) of the SRC problem are typically nonunique: indeed, the set of optimal (p, r)'s is unbounded because, in pure SRC calculations, the capacity premium is completely indeterminate (except when it vanishes because there is excess capacity). But the r associated with a particular p may well be unique, and so may y and v (as has been tacitly assumed by using the notation $\hat{r}$ and $\hat{y}$ in describing the short-run approach). That is, solutions for r and (y, v) to the SRP problem (5.3.6)–(5.3.7) and its dual (5.5.6) can both be expected to be unique or, at the very least, to form bounded sets. This can be illustrated with an elementary but instructive example. Suppose for simplicity that there is no variable input, and that $\mathbb{Y}$ is a cone. A long-run producer optimum is then described by the Complementarity Conditions (5.3.5), i.e.,

$$(y, -k) \in \mathbb{Y},\ (p, r) \in \mathbb{Y}^\circ \text{ and } \langle p \,|\, y \rangle = \langle r \,|\, k \rangle\,.$$

In the profit approach (given p and k), both inclusions are useful in solving this system for y and r. But in the cost approach (given y and k), the first inclusion restricts only the data—so, when it is met, it is of no help at all in solving for p and r. The simplest example is $\mathbb{Y} = \{(y, -k) \in \mathbb{R}^2 : y = k\}$; then $\mathbb{Y}^\circ = \{(p, r) \in \mathbb{R}^2 : p = r\}$. In the cost approach the level of (p, r) is indeterminate, but in the profit approach both solutions are unique, viz., $(\hat{y}, \hat{r}) = (k, p)$.[37] This principle is also borne out by more significant and complex examples such as peak-load pricing with storage, in which the optimum $\hat{r}(p, k, w)$ or $\hat{y}(p, k, w)$ is shown to be unique if the TOU tariff p is, respectively, a continuous or plateau-less function of time: see Chapters 3 and 4, or Section 5.14 in this chapter.

The second, and more significant, advantage of the SRP approach over the SRC approach emerges, at the equilibrium stage, whenever there is a number of producers, with technologies $\mathbb{Y}_\theta$ for $\theta \in \Theta$. In the profit approach, the short-run equilibrium is found by equating the demand $\hat{x}(p)$ to the profit-maximising total output $\sum_\theta \hat{y}_\theta (p, k_\theta, w)$ and solving for p; when the optimal output is nonunique, one solves for p the inclusion

$$\hat{x}(p) \in \sum_\theta \hat{Y}_\theta (p, k_\theta, w) \tag{5.10.1}$$

where $\hat{Y}_\theta$ is the solution set for the reduced SRP programme in (5.3.13) and (5.4.2). For comparison, the cost approach requires solving, for the output bundles (y_θ), the inclusion

$$\widetilde{p}\left(\sum_\theta y_\theta\right) \in \bigcap_\theta \check{P}_\theta (y_\theta, k_\theta, w) \tag{5.10.2}$$

where $\widetilde{p}$ is the inverse demand map and $\check{P}_\theta (y_\theta, k_\theta, w)$ is the solution set for the short-run output-pricing programme in (5.6.7), i.e., $\check{P}_\theta$ is essentially $\partial_y C^\theta_{\mathrm{SR}}$, the multi-valued SRMC of an individual plant. This route is likely to be more difficult because, with multiple producers, it means having to solve for a number of variables (y_θ) instead of the single variable p, as well as having to intersect the price sets $(\check{P}_\theta)$ to start with. And these are large, unbounded sets if the fixed inputs impose capacity constraints.

Comments (the relative complexity of the cost approach):

- It is not even easy just to identify all those output allocations (y_θ) with $\bigcap_\theta \check{P}_\theta \neq \emptyset$ in (5.10.2), since this involves splitting the industry's total output among the plants in a cost-minimising way, which can be a difficult problem (known as optimal system despatch in the context of

[37] When there are variable inputs whose cost-minimising quantities $\check{v}$ are known functions of the data (y, k, w), the condition $(y, -k, -v) \in \mathbb{Y}$ in (5.3.5) boils down to $(y, -k, -\check{v}(y, k, w)) \in \mathbb{Y}$, which is again a pure restriction on the data with no information about the unknowns p and r. Of course, the profit approach would have a similar comparative weakness in the condition $(p, r, w) \in \mathbb{Y}^\circ$ if the fixed-input values $\hat{r}$ were easily calculated functions of the data (p, k, w). But the programme that is taken here to be readily soluble, without using duality, is the SRC programme for v, and not the dual of the SRP programme for r.

electricity generation). To see this in detail, note that[38]

$$\text{if } \underline{C}^{\theta}_{\mathrm{SR}} = C^{\theta}_{\mathrm{SR}} \text{ at } (y_\theta, k_\theta, w) \text{ then } \check{P}_\theta \subseteq \partial_y C^{\theta}_{\mathrm{SR}} \text{ at } (y_\theta, k_\theta, w) . \tag{5.10.3}$$

So $\bigcap_\theta C^{\theta}_{\mathrm{SR}}$ is nonempty if $\bigcap_\theta \check{P}_\theta$ is. Furthermore, the industry's SRC as a function of its total output y_{To} is

$$\inf_{(y_\theta)_{\theta\in\Theta}} \left\{ \sum_{\theta\in\Theta} C^{\theta}_{\mathrm{SR}} (y_\theta, k_\theta, w) : \sum_{\theta\in\Theta} y_\theta = y_{\mathrm{To}} \right\} \tag{5.10.4}$$

i.e., it is the infimal convolution of the individual plants' operating cost functions $C^{\theta}_{\mathrm{SR}} (\cdot, k_\theta, w)$, abbreviated to C^θ. With $\triangle$ denoting the convolution operator, one has $p \in \bigcap_\theta \partial C^\theta (y_\theta)$ if and only if both $p \in \partial \left(\triangle_\theta C^\theta\right) \left(\sum_\theta y_\theta\right)$ and $\left(\triangle_\theta C^\theta\right) \left(\sum_\theta y_\theta\right) = \sum_\theta C^\theta (y_\theta)$: see, e.g., [57, 6.6.3 and 6.6.4]. The "only if" part shows that if $\bigcap_\theta \partial C^\theta (y_\theta) \neq \emptyset$, then (y_θ) is a cost-minimising split of the industry's total output $\sum_\theta y_\theta$ among the plants with the given capacities (k_θ) and technologies $(\mathbb{Y}_\theta)$. This means that competitive profit maximisation, by the choice of outputs (y_θ) at a *common* output price p, leads to such an optimal allocation of the total output.

- Thus the decentralised, plant-by-plant derivation of the industry's total output (given a common output price p) by-passes the problem of the cost-minimising allocation of any *given* total output y_{To}, which is usually much more complex than the individual profit-maximising operation problems. For example, cost-minimising despatch of a hydro-thermal electricity-generating system necessitates a CP with no simple form for either the primal or the dual solution: see the policy construction in [55, pp. 201–219]. By contrast, profit-maximising operation of a hydro plant (or a storage plant) is an LP whose solution has a relatively simple structure: see Chapters 3 and 4, or Section 5.14 in this chapter.

The above description of either variety, SRP or SRC, of the short-run approach assumes the use of either the SRP or the SRC optimisation system (or its split form). Of the optimisation systems, this is the one directly suited to the purpose; and when the technology is given by a production set (as in an engineering specification), there may be no tractable formulae for the value functions, and hence no usable alternative among differential systems. A differential system is likely to be useful only when each of the profit or cost functions it uses is either easy to calculate (by solving the relevant programme), or is simply given as a definition of the technology (as in econometric uses of duality). These remarks are expanded in [46].

[38] This is proved in [46] from a version of the dual to the short-run Shephard's Lemma that is limited to output prices.

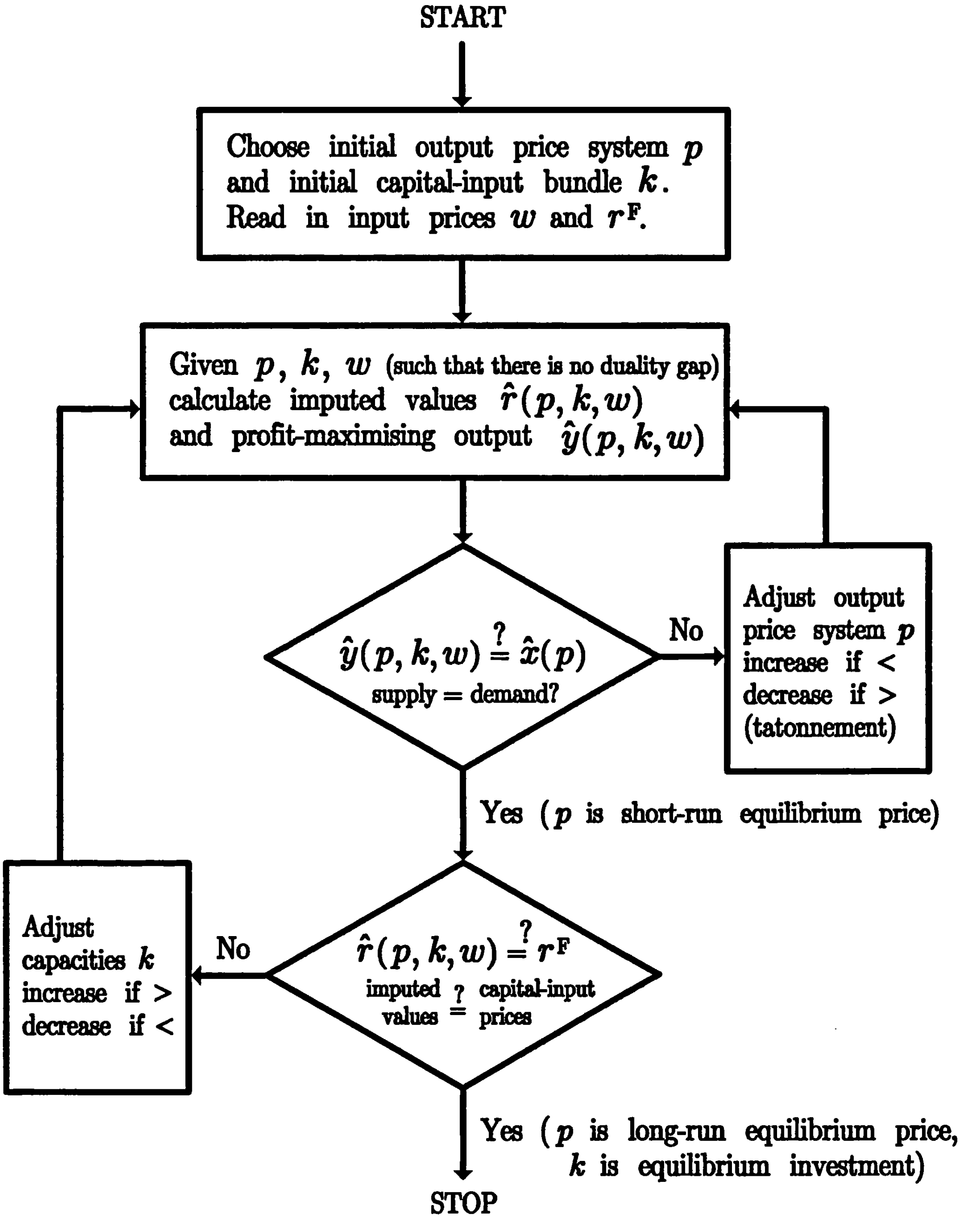

FIGURE 5.3. Flow chart for iterative implementation of SR profit approach to LR general equilibrium. For simplicity, all demand for outputs is taken to be consumer demand that is independent of profit income, and all input prices are fixed (in numeraire terms). Absence of duality gap and existence of optima ($\hat{r}$, $\hat{y}$) can be ensured by using the results of [46].

5.11 A framework for the short-run profit approach to long-run general equilibrium

The equilibrium framework set out next is designed to price a range of commodities with joint costs of production. The product range can be a single good differentiated over commodity characteristics, such as time. Such a differentiated good is usually produced by a variety of techniques; this is so in the motivating application to the peak-load pricing of electricity (Sections 5.13 to 5.15).

To concentrate on the issues of investment and pricing for the differentiated output of a particular Supply Industry (SI), the equilibrium model is simplified by aggregating commodities on the basis of some fixed relative prices. As a result, there are just two consumption goods apart from the differentiated good—viz., the numeraire (measured in $) and a produced final good which is a homogeneous composite representing those commodities whose production requires an input of the differentiated good. The prices for most of the SI's inputs, including all the variable inputs, are also assumed to be given. But, to keep the equilibrium capacities (and the variable inputs) as explicit entries of the equilibrium allocation, these inputs have *not* been aggregated with the numeraire (despite their fixed prices).

The Supply Industry's technology consists of a finite number of production techniques, each of which uses a different set of input commodities to produce the same set of output commodities. For each technique $\theta \in \Theta$, its sets of the fixed and the variable inputs are denoted by Φ_θ and Ξ_θ; and its long-run production set is taken to be a convex cone

$$\mathbb{Y}_\theta \subset Y \times \mathbb{R}^{\Phi_\theta} \times \mathbb{R}^{\Xi_\theta}. \tag{5.11.1}$$

Thus $\mathbb{Y}_\theta$ lies in a space that depends on θ. To be formally regarded as a subset of the full commodity space, $\mathbb{Y}_\theta$ must be embedded in it as $\mathbb{Y}_\theta \times \{(0,0,\ldots)\}$, i.e., by inserting zeros in the input-output bundle at the other positions.

Investment in technique θ is denoted by $k_\theta \in \mathbb{R}^{\Phi_\theta}$; so the SI's total investment in fixed input ϕ is

$$q_\phi = \sum_{\theta:\, \phi \in \Phi_\theta} k_{\theta\phi} \quad \text{for } \phi \in \Phi_\Theta := \bigcup_{\theta \in \Theta} \Phi_\theta \tag{5.11.2}$$

which is the SI's set of fixed inputs. When the sets Φ_θ are pairwise disjoint, the sum in (5.11.2) reduces to a single term (for each ϕ), and the notation can be simplified: see (5.11.20), etc.

The set of all the fixed inputs of the SI, Φ_Θ, is partitioned into two subsets: Φ_Θ^{F} consisting of those with given prices, and Φ_Θ^{E} consisting of those whose prices are determined only in long-run equilibrium. For a particular technique θ, its set of fixed inputs Φ_θ is thus partitioned into two subsets

$$\Phi_\theta^{\mathrm{E}} := \Phi_\Theta^{\mathrm{E}} \cap \Phi_\theta \quad \text{and} \quad \Phi_\theta^{\mathrm{F}} := \Phi_\Theta^{\mathrm{F}} \cap \Phi_\theta.$$

An input $\phi \in \Phi_\Theta^{\mathrm{F}} = \bigcup_{\theta \in \Theta} \Phi_\theta^{\mathrm{F}}$ is supplied at a fixed unit cost r_ϕ^{F} (in terms of the numeraire), so its total supply cost is linear. By contrast, the total supply cost of an input $\phi \in \Phi_\Theta^{\mathrm{E}} = \bigcup_{\theta \in \Theta} \Phi_\theta^{\mathrm{E}}$ is given by a convex function, G_ϕ, of the supplied quantity q_ϕ. Typically, G_ϕ is a strictly convex and increasing, finite function on an interval $[0, \overline{q}_\phi]$, with $G_\phi(0) = 0$. But the case of an input in a fixed supply $\overline{q}_\phi$ (without free disposal) is captured by setting $G_\phi(q_\phi)$ equal to 0 for $q_\phi = \overline{q}_\phi$ and to $+\infty$ otherwise (in which case the equilibrium condition that $r_\phi \in \partial G_\phi(q_\phi)$ means merely that $q_\phi = \overline{q}_\phi$). For examples in the electricity supply industry (ESI), see Chapters 3 and 4, or Section 5.15 in this chapter.

This classification of inputs will not always be clear-cut, but as a rough rule, for an industry supplying a good with a cyclical demand, its fixed inputs are those which cannot be adjusted within a demand cycle because of the cost and the time it takes. For example, there is usually no question of adjusting plant capacity to demand even if the cycle is as long as a year. Variable inputs are those which can be adjusted quickly, at negligible cost, to the time-varying output rate $y_\theta(t)$. For example, fuel inputs are assumed to be instantaneously adjustable in the model of thermal electricity generation: see (5.13.1). The variable inputs are regarded as having fixed prices (w_ξ), e.g., by reason of being internationally traded. Likewise, a typical fix-priced capital input $\phi \in \Phi_\Theta^{\mathrm{F}}$ is internationally traded equipment, and its rental price r_ϕ^{F} is the annuity consisting of interest on the purchase price and depreciation.[39] By contrast, an equilibrium-priced capital input $\phi \in \Phi_\Theta^{\mathrm{E}}$—whose rental price r_ϕ^{E} is determined only in long-run equilibrium—is typically a factor which can only be supplied locally and at an increasing marginal cost, as a result of the fixity of some assets required for its supply (such as special sites or other natural resources). Constancy of returns to scale for the SI's technology need not extend to its input supply, and in the application to peak-load pricing with storage the reservoir capacity has an increasing marginal cost (Section 5.15).

For simplicity, all input demand for the SI's products is taken to come from a single Industrial User (IU), who produces a final good from inputs of the differentiated good and the numeraire. The user's production function F: $Y_+ \times \mathbb{R}_+ \to \mathbb{R}$, assumed to be strictly concave and increasing, defines his production set

$$\mathbb{Y}_{\mathrm{IU}} = \{(-z; \varphi, -n) \in Y_- \times \mathbb{R} \times \mathbb{R}_- : F(z, n) \geq \varphi\} \tag{5.11.3}$$

where Y_+ is a convex cone that is P-closed (i.e., closed for some, and hence for every, locally convex topology on Y that yields P as the continuous dual space). When, as in superdifferentiation at the algebraic boundary points (non-core points) of $Y_+ \times \mathbb{R}_+$, the function F must be regarded as defined on the whole space $Y \times \mathbb{R}$, it is extended by setting its value to $-\infty$ outside of $Y_+ \times \mathbb{R}_+$.[40]

[39] Formally, the fixed prices r^{F} and w are built into the standard competitive equilibrium model by introducing a linear production set equal to the hyperplane perpendicular to the vector $(r^{\mathrm{F}}, w, 1)$ and passing through the origin in the space of the supplier's fix-priced inputs and the numeraire.

[40] This matters in calculating $\widehat{\partial} F$ at a point that belongs to $Y_+ \times \mathbb{R}_+$ but *not* to its core (a.k.a. algebraic interior). To spell this out, assume that F, as a function on its effective domain $Y_+ \times \mathbb{R}_+$, has a Mackey continuous concave extension F^{Ex} defined on all of $Y \times \mathbb{R}$. Then $\widehat{\partial} F = \widehat{\partial} F^{\mathrm{Ex}}$ at any core points of $Y_+ \times \mathbb{R}_+$, but in general $\widehat{\partial} F(z, n) = \widehat{\partial} F^{\mathrm{Ex}}(z, n) + \{(\mu, \nu) \in P_+ \times \mathbb{R}_+ : \langle \mu \mid z \rangle + \nu n = 0\}$ because $F = F^{\mathrm{Ex}} - \delta(\cdot \mid Y_+ \times \mathbb{R}_+)$.

A complete commodity bundle, then, consists of: (i) the produced differentiated good, (ii) the Supply Industry's fixed and variable inputs, (iii) the Industrial User's product, and (iv) the numeraire. The quantities are always listed in this order; but those which are irrelevant in a particular context (and can be set equal to zero) are omitted for brevity, as in (5.11.1) and (5.11.3). A consumption bundle consists of quantities of the differentiated good, the IU's product and the numeraire; so it may be written as $(x; \varphi, m) \in Y \times \mathbb{R}^2$. A matching consumer price system is $(p; \varrho, 1) \in P \times \mathbb{R}^2$—whilst a complete price system is

$$\left(p; r^{\mathrm{E}}, r^{\mathrm{F}}; w, \varrho, 1\right) = \left(p; \left(r_\phi^{\mathrm{E}}\right)_{\phi \in \Phi_\Theta^{\mathrm{E}}}, \left(r_\phi^{\mathrm{F}}\right)_{\phi \in \Phi_\Theta^{\mathrm{F}}}; \left(w_\xi\right)_{\xi \in \Xi_\Theta}, \varrho, 1\right)$$

(where $\Xi_\Theta := \bigcup_{\theta \in \Theta} \Xi_\theta$). There is a finite set, Ho, of households; and for each $h \in \mathrm{Ho}$ its utility is a concave nondecreasing function U_h on the consumption set $Y_+ \times \mathbb{R}_+^2$. It is assumed to be nonsatiated in each of the two homogeneous goods (the IU's product and the numeraire), i.e., $U_h(x; \varphi, m)$ is increasing in φ and in m; this guarantees that both prices are positive in equilibrium. Each household's initial endowment is a quantity $m_h^{\mathrm{En}} > 0$ of the numeraire only; and its share of profit from the supply of input $\phi \in \Phi_\Theta^{\mathrm{E}}$ is $\varsigma_{h\phi} \geq 0$, with $\sum_h \varsigma_{h\phi} = 1$. Similarly, $\varsigma_{h\,\mathrm{IU}}$ denotes household h's share in the User Industry's profit.

The Supply Industry's profit is zero in long-run equilibrium (because of c.r.t.s.), but an exact short-run analysis requires specifying the households' shares in the operating profits from the SI's plants—since the profit Π_{SR}^θ in (5.11.10) is only approximately offset by the liabilities $r_{|\theta}^{\mathrm{EF}} \cdot k_\theta$, which represents plant depreciation and interest (on the debt from which the plant is assumed to have been financed). A *plant* is specified by its type θ and by its capacities (or, more generally, its quantities of the fixed inputs) $k_{\theta\phi}$, for $\phi \in \Phi_\theta$. All plants of a particular type θ are assumed to have the *same capacity ratios* $(k_{\theta 1} : k_{\theta 2} : \ldots)$; with c.r.t.s., this amounts to assuming that there is at most one plant of each type. Though this is rarely so in a real industry which has evolved over time, the condition *is* met in long-run equilibrium, the calculation of which is the main use for the short-run model. It makes sense, then, to speak of *profit shares in a technique*: denoted by $\varpi_{h\theta}$ (with $\sum_h \varpi_{h\theta} = 1$), household h's share in the operating profit from technique θ is

$$\varpi_{h\theta} := \sum_i \beta_{hi} \alpha_{i\theta}$$

where β_{hi} is h's share in producer i, and $\alpha_{i\theta}$ is i's share in the plant of type θ. (In other words, one can assume that all plants of a type are wholly owned by one and the same producer.)

Notation The restriction, to Ξ_θ, of a $w \colon \Xi_\Theta \to \mathbb{R}$ is $w_{|\Xi_\theta}$, abbreviated to $w_{|\theta}$. Similarly, $r_{|\theta}^{\mathrm{E}}$ and $r_{|\theta}^{\mathrm{F}}$ mean the restrictions to Φ_θ^{E} and to Φ_θ^{F} of an $r^{\mathrm{E}} \colon \Phi_\Theta^{\mathrm{E}} \to \mathbb{R}$ and an $r^{\mathrm{F}} \colon \Phi_\Theta^{\mathrm{F}} \to \mathbb{R}$, respectively. The pair $\left(r^{\mathrm{E}}, r^{\mathrm{F}}\right)$ defines a case-function on $\Phi_\Theta := \Phi_\Theta^{\mathrm{E}} \cup \Phi_\Theta^{\mathrm{F}}$; it is occasionally denoted by r^{EF} for brevity.

By definition, given price systems $\left(r^{\mathrm{F}}, w\right)$ for the fix-priced capital inputs and the variable inputs, a *long-run competitive equilibrium* consists of:

- a system of prices $(p^\star, r^\star, \varrho^\star) \in P_+ \times \mathbb{R}_+^{\Phi_\Theta^{\mathrm{E}}} \times \mathbb{R}_{++}$ (all in terms of the numeraire) for: the Supply Industry's differentiated output good, the equilibrium-priced capital inputs, and the Industrial User's product
- an allocation made up of:
  - a consumption bundle $(x_h^\star, \varphi_h^\star, m_h^\star) \in Y \times \mathbb{R} \times \mathbb{R}$ for each household h
  - an input-output bundle of the Industrial User $(-z^\star, F(z^\star, n^\star), -n^\star) \in Y \times \mathbb{R} \times \mathbb{R}$
  - input-output bundles of the Supply Industry, $(y_\theta^\star, -k_\theta^\star, -v_\theta^\star) \in Y \times \mathbb{R}^{\Phi_\theta} \times \mathbb{R}^{\Xi_\theta}$ for each technique θ

that meet the following definitional conditions:

1. *Producer optimum in Supply Industry*: For each θ,

$$(y_\theta^\star, -k_\theta^\star, -v_\theta^\star) \in \mathbb{Y}_\theta \quad \text{and} \quad \left(p^\star, \left(r_{|\theta}^\star, r_{|\theta}^{\mathrm{F}}\right), w_{|\theta}\right) \in \mathbb{Y}_\theta^\circ \tag{5.11.4}$$

$$\langle p^\star \mid y_\theta^\star \rangle = \left(r_{|\theta}^\star, r_{|\theta}^{\mathrm{F}}\right) \cdot k_\theta^\star + w_{|\theta} \cdot v_\theta^\star \tag{5.11.5}$$

i.e., the equilibrium quantities and prices meet the Complementarity Conditions (5.3.5), or any of the preceding equivalent systems of conditions. In other words, $(y_\theta^\star, -k_\theta^\star, -v_\theta^\star)$ maximises (to zero) the long-run profit at prices $\left(p^\star, \left(r_{|\theta}^\star, r_{|\theta}^{\mathrm{F}}\right), w_{|\theta}\right)$.

2. *Producer optimum in User Industry*: $(p^\star, 1) \in \varrho^\star \widehat{\partial} F(z^\star, n^\star)$.

3. *Consumer utility maximisation*: For each h, $(x_h^\star, \varphi_h^\star, m_h^\star)$ maximises U_h on the budget set $B\left(p^\star, \varrho^\star, \hat{M}_{\mathrm{LR}\,h}(p^\star, r^\star, \varrho^\star)\right)$, where

$$B(p, \varrho, M) := \{(x, \varphi, m) \geq 0 : \langle p \mid x \rangle + \varrho\varphi + m \leq M\} \tag{5.11.6}$$

$$\Pi_\phi(r_\phi) := \sup_{q_\phi} (r_\phi q_\phi - G_\phi(q_\phi)) \quad \text{for } r_\phi \in \mathbb{R} \tag{5.11.7}$$

$$\Pi_{\mathrm{IU}}(p, \varrho) := \sup_{z,n} (\varrho F(z, n) - \langle p \mid z \rangle - n) \tag{5.11.8}$$

$$\hat{M}_{\mathrm{LR}\,h}\left(p, r^{\mathrm{E}}, \varrho\right) := m_h^{\mathrm{En}} + \varsigma_{h\,\mathrm{IU}} \Pi_{\mathrm{IU}}(p, \varrho) + \sum_{\phi \in \Phi_\Theta^{\mathrm{E}}} \varsigma_{h\phi} \Pi_\phi\left(r_\phi^{\mathrm{E}}\right). \tag{5.11.9}$$

4. *Market clearance*: $\sum_\theta y_\theta^\star = z^\star + \sum_h x_h^\star$ and $F(z^\star, n^\star) = \sum_h \varphi_h^\star$.

5. *MC pricing of SI's fixed inputs*: $r_\phi^\star \in \partial G_\phi\left(\sum_\theta k_{\theta\phi}^\star\right)$ for each $\phi \in \Phi_\Theta^{\mathrm{E}}$.[41]

[41] The subdifferential ∂G_ϕ is an interval if the left and right derivatives of G_ϕ differ; this can be the case only on a countable subset of $(0, \bar{q}_\phi)$. Also, $\partial G_\phi(0) = \left[0, (\mathrm{d}G_\phi/\mathrm{d}q_\phi)(0+)\right]$ and $\partial G_\phi(\bar{q}_\phi) = \left[(\mathrm{d}G_\phi/\mathrm{d}q_\phi)(\bar{q}_\phi -), +\infty\right)$.

Comment: This is an instance of the usual equilibrium concept, except for being specialised to the case of nonzero prices ($\varrho^\star$ and 1) for the two composite goods (in particular, the above characterisation of the IU's profit maximum, Condition 2, relies on the positivity of the output price $\varrho^\star$). The usual definition captures also the case of zero prices, but this cannot arise here because of the nonsatiation assumptions. In other words, price positivity is actually a property of an equilibrium (and not part of the concept itself).

The short-run profit approach to solving this system starts by fixing the SI's capital inputs $(k_\theta)_{\theta\in\Theta}$. Given these quantities as well as prices (p, w) for the SI's variable commodities, a suitably chosen system characterising the long-run producer optimum is then solved for: the plants' outputs y_θ, their variable inputs v_θ and the values, r_θ, imputed to the fixed inputs in the plant of each type θ. The optimal outputs $\hat{y}_\theta\left(p, k_\theta, w_{|\theta}\right)$ are then equated to demand to find the short-run equilibrium price system $p^\star_{\mathrm{SR}}$, which depends on the k_θ's.[42] Finally, to determine the capacities k_θ, and the prices r^{E} of any equilibrium-priced capital inputs, the imputed value $\hat{r}_{\theta\phi}\left(p, k_\theta, w_{|\theta}\right)$ is equated either to the given price r^{F}_ϕ (for $\phi \in \Phi^{\mathrm{F}}_\Theta$) or to the marginal supply cost $\mathrm{d}G_\phi/\mathrm{d}q_\phi$ at $q_\phi = \sum_\theta k_{\theta\phi}$ (for $\phi \in \Phi^{\mathrm{E}}_\Theta$). As part of this long-run equilibrium condition, if any input ϕ is used by two or more plant types θ' and θ'', i.e., $\phi \in \Phi_{\theta'} \cap \Phi_{\theta''}$, then its values imputed in the different uses, $\hat{r}_{\theta'\phi}$ and $\hat{r}_{\theta''\phi}$, are required to be equal. (In a short-run equilibrium, the values of the same capital input commodity in different uses may of course differ.) If done by iteration, the search for $p^\star_{\mathrm{SR}}$ corresponds to the inner loop in Figure 5.3, and the search for $k^\star_\theta$ corresponds to the outer loop in Figure 5.3.

Since the SI's technology is specified by production sets (rather than profit or cost functions), this approach generally uses, for a characterisation of long-run producer optimum, the SRP optimisation system (5.6.1)–(5.6.3) or its split form, which, with c.r.t.s., consists of (5.4.2)–(5.4.3) and (5.4.6)–(5.4.7). The split form can be convenient when the SRC programme is readily solved. The cases in which other systems may be equally workable are pointed to at the end of Section 5.10.

The two stages of calculating the long-run equilibrium are next described in detail. The first stage is to find the short-run equilibrium, given plants with arbitrary capacities $k = (k_\theta)_{\theta\in\Theta}$, and given arbitrary prices r^{E}, which complement the fixed prices r^{F} to a full capital-input price system $r^{\mathrm{EF}} = \left(r^{\mathrm{E}}, r^{\mathrm{F}}\right)$. At this stage, r^{EF} matters only in calculating the total short-run income, which is

$$\hat{M}_{\mathrm{SR}\,h}\left(p; r^{\mathrm{E}}, r^{\mathrm{F}}; w, \varrho \,|\, k\right) := m^{\mathrm{En}}_h + \sum_{\theta\in\Theta} \varpi_{h\theta}\left(\Pi^\theta_{\mathrm{SR}}\left(p, k_\theta, w_{|\theta}\right) - r^{\mathrm{EF}}_{|\theta}\cdot k_\theta\right)$$
$$+ \sum_{\phi\in\Phi^{\mathrm{E}}_\Theta} \varsigma_{h\phi}\left(r^{\mathrm{E}}_\phi \sum_{\theta:\,\phi\in\Phi^{\mathrm{E}}_\theta} k_{\theta\phi} - G_\phi\left(\sum_{\theta:\,\phi\in\Phi^{\mathrm{E}}_\theta} k_{\theta\phi}\right)\right) + \varsigma_{h\,\mathrm{IU}}\Pi_{\mathrm{IU}}\left(p, \varrho\right). \quad (5.11.10)$$

[42] The corresponding input demand, $\check{v}_\theta\left(\hat{y}_\theta\left(p, k_\theta, w_{|\theta}\right), k_\theta, w_{|\theta}\right)$, would similarly have to be equated to input supply, had the supply not been taken to be perfectly elastic (i.e., if the input prices w were not fixed, and had to be determined).

Comment (on the composition of income in the short and long runs): The exact expression for the short-run income (5.11.10) can be approximated by simpler ones. The first sum over θ in (5.11.10) represents pure-profit income from the SI, and the sum over ϕ is the profit income from supplying any equilibrium-priced inputs to the SI. In the long run, these profits are competitively maximised over k_θ and, as a result, the SI's profit is zero.[43] The profit incomes from input supply usually remain positive in the long run, and their sum over ϕ is a term of $\hat{M}_{\mathrm{LR}\,h}$ in (5.11.9). For the purpose of calculating the long-run equilibrium by the short-run approach, one can therefore replace $\hat{M}_{\mathrm{SR}\,h}$ by the simpler expression $\hat{M}_{\mathrm{LR}\,h}$ in the *short*-run consumer problem (5.11.14). This would make the short-run consumer demand map identical to the long-run one. (The short-run equilibria so calculated would differ from the exact ones, but not by very much unless the short-run problem's capacities were far from long-run equilibrium.) Also, since the profit from input supply is likely to be relatively small in practice, it may be acceptable to disregard it in calculating consumer demand (thus taking the consumer's income to be $m_h^{\mathrm{En}} + \varsigma_{h\,\mathrm{IU}}\Pi_{\mathrm{IU}}$, instead of $\hat{M}_{\mathrm{SR}\,h}$ or $\hat{M}_{\mathrm{LR}\,h}$).

Given a $k = (k_\theta)_{\theta\in\Theta}$ as well as r^{E}, r^{F} and w, the short-run general equilibrium (SRGE) system to be solved consists of the following conditions on the other variables (viz., prices p paired with quantities y_θ, x_h and z, price ϱ paired with quantity φ_h, quantities v_θ, and amounts of numeraire m_h and n):

$$y_\theta \text{ maximises SRP, i.e., satisfies (5.4.2), for each } \theta \tag{5.11.11}$$

$$v_\theta \text{ minimises SRC, i.e., satisfies (5.4.3), for each } \theta \tag{5.11.12}$$

$$(p,1) \in \varrho\hat{\partial}F(z,n) \tag{5.11.13}$$

$$(x_h, \varphi_h, m_h) \text{ maximises } U_h \text{ on } B\left(p, \varrho, \hat{M}_{\mathrm{SR}\,h}\left(p, r^{\mathrm{EF}}, w, \varrho \,|\, k\right)\right) \tag{5.11.14}$$

$$\sum_{\theta\in\Theta} y_\theta = z + \sum_{h\in\mathrm{Ho}} x_h \quad \text{and} \quad F(z,n) = \sum_{h\in\mathrm{Ho}} \varphi_h. \tag{5.11.15}$$

The short-run equilibrium system (5.11.11)–(5.11.15) can be solved in steps:

1. It is taken to be easiest to start by solving the SRC programme in (5.4.3) to determine the short-run conditional demand of each plant type θ for its variable inputs. For a technology with conditionally fixed technical coefficients, the conditional input demand $\check{v}_\theta(y_\theta)$ depends only on the plant's output y_θ. In general, it depends also on the fixed inputs k_θ and the variable-input prices $w_{|\theta}$.

2. Since C_{SR}^θ is now a known function of $(y_\theta, k_\theta, w_{|\theta})$—equal to $w_{|\theta} \cdot \check{v}_\theta$ if the SRC programme is feasible, and to $+\infty$ if not—the reduced SRP programme in (5.4.2) can be solved next; it is

[43] Formally, this is because in long-run equilibrium $r_{|\theta}^{\mathrm{EF}} = \hat{r}_\theta$ as per (5.11.18), and because $(p, \hat{r}_\theta, w_{|\theta}) \in \Upsilon_\theta^\circ$ for each θ by the dual constraint on r_θ. For the same reason, in calculating the long-run equilibrium one can restrict attention, already at the short-run stage, to those r^{EF}'s with $\left(p, r_{|\theta}^{\mathrm{EF}}, w_{|\theta}\right) \in \Upsilon_\theta^\circ$ for each θ.

an LP if $\breve{v}_\theta$ is linear in y_θ.[44] It generally has a multi-valued solution set, $\hat{Y}_\theta\left(p, k_\theta, w_{|\theta}\right)$.

3. Consumer demands are found as functions $(\hat{x}_h, \hat{\varphi}_h)$ of $(p, \varrho; M)$, and the known value of $\Pi^\theta_{\mathrm{SR}}\left(p, k_\theta, w_{|\theta}\right)$—viz., $\langle p \mid y_\theta \rangle - C^\theta_{\mathrm{SR}}\left(y_\theta\right)$ for any $y_\theta \in \hat{Y}_\theta$—is used to calculate $\hat{M}_{\mathrm{SR}\,h}$ as per (5.11.10). Factor demands (of the User Industry) are found as functions $(\hat{z}, \hat{n})$ of $(p, \varrho) \in P_+ \times \mathbb{R}_{++}$, from (5.11.13).[45]

4. Finally, the system

$$\hat{z}\left(p, \varrho\right) + \sum_{h \in \mathrm{Ho}} \hat{x}_h \left(p, \varrho; \hat{M}_{\mathrm{SR}\,h}\left(p; r^{\mathrm{E}}, r^{\mathrm{F}}; w, \varrho \mid k\right)\right) \in \sum_{\theta \in \Theta} \hat{Y}_\theta\left(p, k_\theta, w_{|\theta}\right) \qquad (5.11.16)$$

$$\sum_{h \in \mathrm{Ho}} \hat{\varphi}_h \left(p, \varrho; \hat{M}_{\mathrm{SR}\,h}\left(p; r^{\mathrm{E}}, r^{\mathrm{F}}; w, \varrho \mid k\right)\right) = F\left(\hat{z}\left(p, \varrho\right), \hat{n}\left(p, \varrho\right)\right) \qquad (5.11.17)$$

is solved for p and ϱ.

This gives the short-run equilibrium prices, $p^\star_{\mathrm{SR}}$ (for the Supply Industry's differentiated output good) and $\varrho^\star_{\mathrm{SR}}$ (for the Industrial User's product). It also gives, by back substitution, the short-run equilibrium quantities, viz.: (i) the outputs and demands for the differentiated good, with $\sum_\theta y^\star_{\mathrm{SR}\,\theta} = z^\star_{\mathrm{SR}} + \sum_h x^\star_{\mathrm{SR}\,h}$, (ii) the Supply Industry's variable inputs $v^\star_{\mathrm{SR}\,\theta}$, (iii) the User Industry's output $\varphi^\star_{\mathrm{SR}}$ and input $n^\star_{\mathrm{SR}}$, and (iv) consumption of the numeraire $\sum_h m^\star_{\mathrm{SR}\,h}$. Generally, all of these are functions of the short-run equilibrium problem's data k and r^{E} (as well as depending on the fixed prices r^{F} and w).[46]

The second stage is to determine the long-run equilibrium, i.e., the equilibrium capacities and the prices of any equilibrium-priced capital inputs (i.e., those in Φ^{E}_Θ). Optimality of investment k_θ in each technique is achieved by satisfying the rest of the split SRP optimisation system, viz., (5.4.6)–(5.4.7). For this, the solution set $\hat{R}_\theta\left(p, k_\theta, w_{|\theta}\right)$ of the FIV minimisation programme (5.5.13)–(5.5.14) with $\mathbb{Y}_\theta$ in place of $\mathbb{Y}$, or the solution $\hat{r}_\theta$ if it is unique, is calculated at $p = p^\star_{\mathrm{SR}}\left(k, r^{\mathrm{EF}}, w\right)$. Actually, $\hat{r}_\theta$ will usually have already been found as the dual solution in the process of solving the SRP programme for y_θ by a duality method, i.e., as a by-product of Step 2 in solving (5.11.11)–(5.11.15). Finally, the system of long-run equilibrium conditions

$$\left(r^{\mathrm{E}}_{|\theta}, r^{\mathrm{F}}_{|\theta}\right) \in \hat{R}_\theta\left(p^\star_{\mathrm{SR}}\left(k; r^{\mathrm{EF}}; w\right), k_\theta, w_{|\theta}\right) \text{ i.e., } r^{\mathrm{EF}}_{|\theta} \text{ satisfies (5.4.6) for each } \theta \in \Theta \qquad (5.11.18)$$

$$r^{\mathrm{E}}_\phi \in \partial G_\phi \left(\sum_{\theta:\, \phi \in \Phi_\theta} k_{\theta\phi} \right) \quad \text{for each } \phi \in \Phi^{\mathrm{E}}_\Theta \qquad (5.11.19)$$

is solved for $k = \left(k_\theta\right)_{\theta \in \Theta}$ and r^{E} (given r^{F} and w).[47] Any solution $(k^\star, r^\star)$ is a part of a long-run equilibrium—provided that there is no duality gap between the SRP programme and its dual

[44] For example, in thermal electricity generation, $\breve{v}_\theta\left(y_\theta\right) = \int y_\theta\left(t\right) \mathrm{d}t$ and so (5.14.1)–(5.14.3) is an LP.

[45] This step is independent of the preceding derivation of short-run supply.

[46] For simplicity, the short-run equilibrium is assumed to be unique.

[47] As a basic check, note that the number of "generalised equations" in this system (each d-dimensional vector inclusion counting as d "equations") is the same as the number of unknowns (viz., $\sum_{\theta \in \Theta} \mathrm{card}\, \Phi_\theta + \mathrm{card}\, \Phi^{\mathrm{E}}_\Theta$).

(5.5.13)–(5.5.14) for any θ (i.e., if (5.4.7) or equivalently (5.11.5) holds). The rest of the long-run equilibrium follows by substituting $k^\star$ and $r^\star$ into the short-run equilibrium solution. In particular, in long-run equilibrium, consumer and factor demands for the differentiated good, its total output and its price system are:

$$\sum_h x^\star_{\mathrm{LR},h} = \sum_h x^\star_{\mathrm{SR},h}\left(k^\star; r^\star, r^{\mathrm{F}}; w\right)$$
$$z^\star_{\mathrm{LR}} = z^\star_{\mathrm{SR}}\left(k^\star; r^\star, r^{\mathrm{F}}; w\right)$$
$$\sum_\theta y^\star_{\mathrm{LR}\,\theta} = \sum_\theta y^\star_{\mathrm{SR}\,\theta}\left(k^\star; r^\star, r^{\mathrm{F}}; w\right)$$
$$p^\star_{\mathrm{LR}} = p^\star_{\mathrm{SR}}\left(k^\star; r^\star, r^{\mathrm{F}}; w\right).$$

The SRGE system (5.11.11)–(5.11.15) together with the long-run conditions (5.11.18)–(5.11.19) can be called the *SRP programme-based LRGE system.*

Comments:

- The SRGE system simplifies when there is no income effect on the differentiated good (i.e., when $\hat{x}_h$ is independent of M, in the relevant range): the solution $\left(p^\star_{\mathrm{SR}}, \varrho^\star_{\mathrm{SR}}\right)$ to (5.11.16)–(5.11.17) is then independent of r^{EF}, as in Section 5.2.

- A production technique can usually be identified by its set of fixed inputs, i.e., $\Phi_{\theta'} \neq \Phi_{\theta''}$ for $\theta' \neq \theta''$. Under the stronger assumption that different techniques use disjoint sets of fixed inputs, i.e., that

$$\Phi_{\theta'} \cap \Phi_{\theta''} = \emptyset \quad \text{for } \theta' \neq \theta'', \tag{5.11.20}$$

the SI's total investment in fixed input ϕ is simply $k_{\theta\phi}$ for the one θ such that $\Phi_\theta \ni \phi$. In other words, it is the case-function (of ϕ) defined, piecewise, as equal to the function k_θ on each Φ_θ. Thus it can be identified with $k = (k_\theta)_{\theta\in\Theta}$ itself. So, under (5.11.20), the total investment can be denoted by k: $\Phi_\Theta \to \mathbb{R}$. The investment in technique θ is then the restriction of k to Φ_θ, which is denoted by $k_{|\Phi_\theta}$, abbreviated to $k_{|\theta}$. The investment in fixed input ϕ is k_ϕ (i.e., $q_\phi = k_\phi$ in this case). This is so in the model of the ESI's technology (Section 5.13).

- Assume that: (i) the techniques use disjoint sets of capital inputs, i.e., (5.11.20) holds, (ii) each input-cost, $G_\phi\left(k_\phi\right)$, is a differentiable function of $k_\phi \in \mathbb{R}_{++}$, and (iii) a unique shadow price system $\hat{r}_\theta\left(p, k_{|\theta}, w_{|\theta}\right)$ exists at every $k \gg 0$ and every p in a subspace of P that is known to contain $p^\star_{\mathrm{SR}}$. (As is shown in [45] for a class of problems that includes peak-load pricing with storage, this is so for the space of continuous real-valued functions $\mathcal{C}\left[0,T\right]$, as a price subspace of $P = L^1\left[0,T\right]$.) If a long-run equilibrium with $k^\star \gg 0$ is sought, then Conditions (5.11.18)–(5.11.19) on k reduce to the following equations for k (a strictly positive vector in $\mathbb{R}^{\Phi_\Theta}$):

$$\hat{r}_{\theta\phi}\left(p_{\mathrm{SR}}^{\star}\left(k;\left\{\frac{\mathrm{d}G_{\phi}}{\mathrm{d}k_{\phi}}\left(k_{\phi}\right)\right\}_{\phi\in\Phi_{\Theta}^{\mathrm{E}}},r^{\mathrm{F}};w\right),k_{|\theta},w_{|\theta}\right)$$

$$=\begin{cases}\frac{\mathrm{d}G_{\phi}}{\mathrm{d}k_{\phi}}\left(k_{\phi}\right) & \text{if } \phi\in\Phi_{\theta}^{\mathrm{E}}\\ r_{\phi}^{\mathrm{F}} & \text{if } \phi\in\Phi_{\theta}^{\mathrm{F}}\end{cases} \tag{5.11.21}$$

for each θ and $\phi \in \Phi_{\theta}$.

- This investment problem has a partial-equilibrium version in which a *given* p replaces the $p_{\mathrm{SR}}^{\star}$ in the system (5.11.21), for a particular production technique θ. It is studied in [40], and in Section 3.11 (Chapter 3) for the case of pumped storage.

- All of the SI's inputs have been assumed to be homogeneous goods, but in some cases an input is a differentiated good. If it is also an equilibrium-priced fixed input, then its supply cost G_{ϕ} is a joint-cost function of the commodity bundle, q_{ϕ}. The short-run approach readily accommodates such inputs (the only difference is that ∂G_{ϕ} is not an interval of $\mathbb{R}$, but a convex subset of the relevant price space). An example is the river flow $e \in L^{\infty}[0,T]$ for hydroelectric generation in Theorem 5.15.2, but in that case Condition (5.11.19) imposes no restriction on the water price function ψ because e is fixed (even in the long run).

5.12 Duality for linear programmes with nonstandard parameters in constraints

Once the production set $\mathbb{Y}$ has been represented as an intersection of half-spaces, each of the profit or cost programmes of Section 5.3 becomes an LP, i.e., a programme of optimising a linear function subject to linear inequality or equality constraints. It is a parametric LP, with the fixed quantities k as its primal parameters (Section 5.5). The fixed quantities need not, of course, be the standard "right-hand side" parameters. But the marginal effects of any nonstandard parameters can be expressed in terms of those of the standard parameters, i.e., in terms of the standard dual solution $\hat{\sigma}$, which consists of the usual Lagrange multipliers for the constraints. This is done in (5.12.12) below.

To start with, this formula is given for the case of a finite LP, i.e., an LP with finite numbers of decision variables, parameters and constraints. The focus is on the SRP programme of a production technique with c.r.t.s. To simplify the notation, it is assumed that there is no variable input (i.e., $\Xi = \emptyset$). As well as being met literally by some techniques (e.g., the storage and hydro techniques of Section 5.13), the assumption is not at all restrictive because the output bundle y can always be reinterpreted as the bundle of all the variable commodities (i.e., outputs and variable inputs).

For now, $\mathbb{Y}$ is therefore a polyhedral cone in the finite-dimensional space $Y \times K = \mathbb{R}^T \times \mathbb{R}^\Phi$, where T and Φ are the sets of output and fixed-input commodities. Its polar, $\mathbb{Y}^\circ$, is a finitely generated convex cone in the price space $P \times R = \mathbb{R}^T \times \mathbb{R}^\Phi$. It can be represented as the sum of a linear subspace spanned by a finite set $\mathcal{G}''$ and a line-free convex cone generated by a finite set $\mathcal{G}'$, i.e.,

$$\mathbb{Y}^\circ = \operatorname{cone} \operatorname{conv} \mathcal{G}' + \operatorname{span} \mathcal{G}''$$

for some positively independent, finite set $\mathcal{G}'$ and another finite set $\mathcal{G}''$ (which can be chosen to be linearly independent). The generators $\mathcal{G}'$ and the spanning vectors $\mathcal{G}''$ can serve as the rows of partitioned matrices $[A' \, B']$ and $[A'' \, B'']$ that give[48]

$$\mathbb{Y} = \left\{ (y, -k) \in \mathbb{R}^T \times \mathbb{R}^\Phi : A'y - B'k \leq 0 \text{ and } A''y - B''k = 0 \right\}. \tag{5.12.1}$$

The primal LP (of short-run profit maximisation) is: given $(p, k) \in \mathbb{R}^T \times \mathbb{R}^\Phi$,

$$\text{maximise } p \cdot y \text{ over } y \in \mathbb{R}^T \tag{5.12.2}$$

$$\text{subject to: } \quad A'y \leq B'k \tag{5.12.3}$$

$$A''y = B''k. \tag{5.12.4}$$

Its optimal value is $\Pi_{\mathrm{SR}}(p, k)$, abbreviated to $\Pi(p, k)$. As in Section 5.5, the vector k is called the *intrinsic* primal parameter, and its increment Δk is an *intrinsic perturbation* of (5.12.2)–(5.12.4).

The corresponding *standard parametric LP* has primal parameters s' and s'', ranging over $\mathbb{R}^{\mathcal{G}'}$ and $\mathbb{R}^{\mathcal{G}''}$, in place of the $B'k$ and $B''k$ in (5.12.3)–(5.12.4). Its optimal value is the *standard primal value*, denoted by $\widetilde{\Pi}(p, s)$, where $s = (s', s'')$. So by definition, for every (p, k),

$$\Pi(p, k) = \widetilde{\Pi}(p, Bk) \quad \text{where } B := \begin{bmatrix} B' \\ B'' \end{bmatrix}. \tag{5.12.5}$$

The *standard perturbation* consists in relaxing (or tightening) the inequality constraints by adding an arbitrary vector $\Delta s = (\Delta s', \Delta s'') \in \mathbb{R}^{\mathcal{G}'} \times \mathbb{R}^{\mathcal{G}''}$ to the r.h.s. of (5.12.3)–(5.12.4), i.e., it uses a separate scalar increment for each constraint. This produces the *standard dual* of (5.12.2)–(5.12.4), which is: given the same $(p, k) \in \mathbb{R}^T \times \mathbb{R}^\Phi$,

$$\text{minimise } \quad \sigma^{\mathrm{T}} Bk = \sigma'^{\mathrm{T}} B'k + \sigma''^{\mathrm{T}} B''k \quad \text{over } \sigma = (\sigma', \sigma'') \in \mathbb{R}^{\mathcal{G}'} \times \mathbb{R}^{\mathcal{G}''} \tag{5.12.6}$$

$$\text{subject to: } \quad \sigma' \geq 0 \tag{5.12.7}$$

$$p = A^{\mathrm{T}} \sigma := A'^{\mathrm{T}} \sigma' + A''^{\mathrm{T}} \sigma'' \tag{5.12.8}$$

[48] Formally, A' and B' are the $\mathcal{G}' \times T$ and $\mathcal{G}' \times \Phi$ matrices with entries $A'_{gt} = g_t$ and $B'_{g\phi} = g_\phi$ for $t \in T$, $\phi \in \Phi$ and $g \in \mathcal{G}'$ (and the same goes for $''$ instead of $'$). This is Farkas's Lemma: see, e.g., [16, 2.2.6], [70, 22.3.1], [74, 6.45] or [80, 4.19].

where $\cdot^{\mathrm{T}}$ denotes transposition. The variable σ is paired with $\varDelta s$ (not $\varDelta k$)—this is the dual of the standard primal LP, which is parametrised by s. It is only *after* forming the dual that Bk is substituted for s to give $\sigma^{\mathrm{T}} Bk$ in (5.12.6). The *standard dual value*, denoted by $\overline{\overline{\Pi}}(p,s)$, is the optimal value of the LP (5.12.6)–(5.12.8) with s instead of Bk, i.e., *before* the substitution. Its solution, the *standard dual solution*, is denoted by $\hat{\sigma}(p,s)$ when it is unique; in general, the solutions form a set $\hat{\Sigma}(p,s)$. The solution set of (5.12.6)–(5.12.8) is therefore $\hat{\Sigma}(p,Bk)$; when unique, the solution is $\hat{\sigma}(p,Bk)$. Its value is $\overline{\overline{\Pi}}(p,Bk)$. This is always equal to the fixed-input value as calculated from (5.5.13)–(5.5.14), i.e.,[49]

$$\overline{\overline{\Pi}}(p,Bk) = \overline{\Pi}(p,k) \quad \text{for every } (p,k). \tag{5.12.9}$$

In other words, the standard dual LP has the same value as the intrinsic dual; here, the two duals are (5.12.6)–(5.12.8) and (5.5.13)–(5.5.14). For their solution sets, $\hat{\Sigma}$ and $\hat{R}$, it follows that

$$\begin{aligned} \hat{R}(p,k) = \widehat{\partial}_k \overline{\Pi}(p,k) &= B^{\mathrm{T}}\, \widehat{\partial}_s \overline{\overline{\Pi}}(p,s)\Big|_{s=Bk} = B^{\mathrm{T}} \hat{\Sigma}(p,Bk) \\ &:= \left\{ B^{\mathrm{T}}\sigma : \sigma \in \hat{\Sigma}(p,Bk) \right\} \end{aligned} \tag{5.12.10}$$

by applying the Chain Rule to (5.12.9),[50] and by using (twice) the identity of the dual solution and the marginal value of Type Two (which is noted at the end of Section 5.7). Thus the intrinsic dual solution ($\hat{R}$) is expressed as the linear image of the standard dual solution ($\hat{\Sigma}$) under the adjoint ($B^{\mathrm{T}}\cdot$) of the operation that maps the intrinsic to the standard primal parameters ($s = Bk$).

When $\Pi = \overline{\Pi}$ at (p,k), the marginal value is actually of Type One, i.e.,

$$\widehat{\partial}_k \Pi(p,k) = \widehat{\partial}_k \overline{\Pi}(p,k) = B^{\mathrm{T}} \hat{\Sigma}(p,Bk). \tag{5.12.11}$$

This always applies to *finite* LPs because their primal and dual values are equal, unless both programmes are infeasible (in which case their values are oppositely infinite).[51] If additionally the dual solution is unique, then

$$\nabla_k \Pi(p,k) = B^{\mathrm{T}} \hat{\sigma}(p,Bk). \tag{5.12.12}$$

This gives the marginal values of the generally nonstandard intrinsic parameters (k) in terms of the *standard* dual solution ($\hat{\sigma}$).

[49] The identity (5.12.9) reduces to (5.12.5) when the primal and dual values are equal, i.e., when $\overline{\Pi} = \overline{\overline{\Pi}}$ and $\Pi = \overline{\Pi}$ at (p,k). This always applies to (feasible) finite LPs, but not always to infinite LPs. To prove (5.12.9) without relying on the absence of a duality gap, note that the constraint $(p,r) \in \mathbb{Y}^{\circ}$ in (5.5.14) means here that $A^{\mathrm{T}}\sigma = p$ and $B^{\mathrm{T}}\sigma = r$ for some $\sigma = (\sigma', \sigma'')$ with $\sigma' \geq 0$ (since the rows of $[A\, B]$ generate or span $\mathbb{Y}^{\circ}$). So the change of variable from r to σ transforms (5.5.13)–(5.5.14) into (5.12.6)–(5.12.8). This argument extends to infinite LPs (and it applies also when there is a duality gap).

[50] For the Chain Rule for subdifferentials, see, e.g., [5, 4.3.6 a], [51, 4.2: Theorem 2], [70, 23.9] or [73, Theorem 19].

[51] See, e.g., [15, 5.1 and 9.1] or [73, Example 1', p. 24] for proofs based on the simplex algorithm or on polyhedral convexity, respectively. This is *not* so with a pair of *infinite* LPs: both can be feasible without having the same value (i.e., the primal and dual values can both be finite but different). Appendix A gives an example.

Comment (on standard and intrinsic perturbations): If B were the unit matrix I, the two perturbation schemes would obviously be the same (and $\varDelta s$ could be renamed to $\varDelta k$). This *would* be so if the short-run production constraints corresponded, one-to-one, to the fixed inputs, i.e., if $\mathbb{Y}$ were defined by a system of inequalities (or equalities) of the form $(Ay)_\phi \le k_\phi$, one for each $\phi \in \Phi$. But such a correspondence generally fails to exist, for three reasons. First, two fixed inputs may appear in one constraint (say $a \cdot y \le k_1 + k_2$). Second, a constraint may involve only the outputs ($a \cdot y \le 0$, e.g., $y_t \ge 0$). Third, each fixed quantity k_ϕ may impose more than one constraint on y (say $(Ay)_1 \le k_\phi$, $(Ay)_2 \le k_\phi, \ldots$). Indeed, this is so whenever k_ϕ is a capacity: staying constant over a time period, it is a scalar but it imposes as many inequality constraints as there are time instants (e.g., $y_t \le k_\phi$ for each t).[52] In such a case, B is a 0-1 matrix whose unit entries appear just once in a row, but more than once in a column. When additionally k is a scalar, B is the single column $\begin{bmatrix} 1 & \ldots & 1 \end{bmatrix}^{\mathrm{T}}$; and an intrinsic perturbation of the constraint system $Ay \le \begin{bmatrix} k & k & \ldots \end{bmatrix}^{\mathrm{T}}$ relaxes all the constraints by the same amount, to $Ay \le \begin{bmatrix} k + \varDelta k & k + \varDelta k & \ldots \end{bmatrix}^{\mathrm{T}}$. By contrast, a standard perturbation relaxes each constraint by a different amount, to $Ay \le \begin{bmatrix} k + \varDelta s_1 & k + \varDelta s_2 & \ldots \end{bmatrix}^{\mathrm{T}}$. In this sense, the standard perturbation scheme is the finest; and, with this B, the intrinsic perturbation scheme is the coarsest. Once the scalar k is identified with the vector $(k, k, \ldots)$, the standard value function $\tilde{\Pi}(p, \cdot)$ becomes an extension of the intrinsic value function $\Pi(p, \cdot)$ from the subspace of constant tuples to all of $\mathbb{R}^{\mathcal{G}'} \times \mathbb{R}^{\mathcal{G}''}$ (with $\mathcal{G}''$ empty if there is no equality constraint), and the intrinsic dual solution (a scalar) is simply the total sum of the standard dual solution, i.e., $\hat{r} = \begin{bmatrix} 1 & \ldots & 1 \end{bmatrix} \hat{\sigma}$ $= \hat{\sigma}_1 + \hat{\sigma}_2 + \ldots$. In other words, the scalar parameter's marginal value is the sum of the marginal values of relaxing all the constraints in which it appears. This arises in the peak-load pricing application: the total capacity values are the integrals of the rent flows over the period, in (5.14.10), (5.14.23)–(5.14.24), and (5.14.44)–(5.14.45). Also, since $\tilde{\Pi}$ is an extension of Π, it can be convenient to use the same letter k as the second variable of both functions (i.e., to use k instead of the s in $\tilde{\Pi}(p, s)$), provided that it is always made clear whether k is a scalar or a vector. This is done in the context of hydro and energy storage (where s signifies the water or energy stock and is not a parameter).

A variant of the standard dual is obtained by including the intrinsic dual variable r paired with k, which is then constrained to equal $B^{\mathrm{T}}\sigma$; thus r is wholly dependent on σ. The objective, $\sigma^{\mathrm{T}} B k$, may then be rewritten concisely as $r \cdot k$. This produces the following LP: given $(p, k) \in \mathbb{R}^{T} \times \mathbb{R}^{\Phi}$,

$$\text{minimise } r \cdot k \text{ over } r \in \mathbb{R}^{\Phi} \text{ and } \sigma = (\sigma', \sigma'') \in \mathbb{R}^{\mathcal{G}'} \times \mathbb{R}^{\mathcal{G}''} \tag{5.12.13}$$

[52] Also, a nonnegativity constraint on k_ϕ makes it appear a second time even if it imposes just one constraint on y (i.e., $0 \le k_\phi$ in addition to $a \cdot y \le k_\phi$ for some $a \ne 0$).

$$\text{subject to: } \sigma' \geq 0,\ p = A^{\mathrm{T}}\sigma \text{ and } r = B^{\mathrm{T}}\sigma. \tag{5.12.14}$$

This may be called the *inclusive* standard dual—an LP for both r and σ. It is the dual that derives from *simultaneous* standard and intrinsic perturbations, i.e., from perturbing Bk on the r.h.s. of (5.12.2)–(5.12.4) to $\Delta s + B(k + \Delta k)$. Its solution gives both sets of marginal values explicitly ($\hat{\sigma}$ and $\hat{r}$), but it is in substance equivalent to the standard dual solution $\hat{\sigma}$ (since $\hat{r} = B^{\mathrm{T}}\hat{\sigma}$). It can be more convenient to use a *partly inclusive* form of the standard dual, which includes only some of the intrinsic dual variables, leaving out those coordinates of r which correspond to "the simplest" columns of B—e.g., to the columns with 0-1 entries as in the Comment above. For example, the programme of valuing the hydro inputs (5.14.37)–(5.14.43) includes the TOU shadow price of water ψ but not the total capacity values r_{St} and r_{Tu}, which are simply the totals of the standard dual variables κ_{St} and κ_{Tu}.

Finally, expressing the intrinsic dual variables (r) in terms of the standard ones (σ) can be extended to infinite LPs. This requires using suitable cones in infinite-dimensional spaces of variables and parameters to formulate infinite systems of constraints on, generally, an infinity of variables. Such a framework is provided in, e.g., [16, 4.2], [57, 7.9] and [73, Examples 4, 4', 4"]. The assumptions made here to adapt it are not the weakest possible; they are selected for simplicity and adequacy to the applications (Section 5.14). The output and fixed-input spaces, Y and K, are now taken to be general Banach spaces, i.e., complete normed spaces (instead of $\mathbb{R}^T$ and $\mathbb{R}^\Phi$). The norm-duals, Y^* and K^*, serve as the corresponding price spaces, P and R. For the primal programme of SRP maximisation, Y is the primal-variable space paired with the dual parameter space P, and K is the primal-parameter space paired with the dual-variable space R. The production cone is given by (5.12.1) in terms of two norm-to-norm continuous linear operations: (i) $A'\colon Y \to L$ and $B'\colon K \to L$, whose common codomain L is a Banach lattice (with a vector order $\leq$ and the corresponding nonnegative cone L_+), and (ii) $A''\colon Y \to X$ and $B''\colon K \to X$, whose codomain X is a Banach space. The spaces L and X replace $\mathbb{R}^{\mathcal{G}'}$ and $\mathbb{R}^{\mathcal{G}''}$ as the spaces for standard perturbations ($\Delta s'$, $\Delta s''$). Their norm-duals, L^* and X^*, serve as the spaces for standard dual variables (σ', σ''). It is best to keep L and X small, but obviously L must contain the ranges of both A' on Y and B' on K (and similarly X must contain both $A''Y$ and $B''K$).

As for the choice of topologies, this must be consistent with the pairing of spaces. Furthermore, the norm topology has to be put on the primal parameter space L if the generalised Slater's Condition of [73, (8.12)] is to be met for the SRP programme (5.12.2)–(5.12.4), i.e., if a y is to exist such that $A'y - B'k \in -\operatorname{int}(L_+)$ and $A''y - B''k = 0_X$. Topologies on Y, K, L and X must make the maximand u.s.c. and the constraint relations closed; here, this means making $\langle p \mid \cdot \rangle$, A and B continuous. So the norm topologies on Y (the primal-variable space) and on K, L and X (the primal-parameter spaces) will do. On the dual-variable spaces K^*, L^* and X^*, the weak* topologies will do. On Y^*

(the dual parameter space), the Mackey topology $\mathrm{m}(Y^*, Y)$ is the best choice if continuity of the dual value function is sought. When Y has a Banach predual Y', it can also be useful to pair Y with Y' as a dual parameter space that is generally smaller than Y^*; the restriction of $\mathrm{m}(Y^*, Y)$ to Y' is the norm topology of Y'. The pairing of Y with Y' is adequate when $p \in Y'$, but *not* when $p \in Y^* \setminus Y'$.

There are at least two sources for the linear operations A and B that describe $\mathbb{Y}$ by (5.12.1). First, such a formula may be the original definition of $\mathbb{Y}$—in which case A and B can simply be read off. This is so in the application to the ESI: the production sets (5.13.1), (5.13.5) or (5.13.9) are all of the form (5.12.1).[53]

Second, A' and B' (with no A'' or B'', i.e., with the zero space as X) can also be constructed from a weakly* compact convex base, Δ, for $\mathbb{Y}^\circ$, which exists if and only if $\mathbb{Y}$ is solid (i.e:, has a nonempty interior) for the norm on $Y \times K$: see, e.g., [4, Theorem 3.16]. An interior point $\left(y^{\mathrm{S}}, -k^{\mathrm{S}}\right)$ defines the base

$$\Delta := \left\{(p,r) \in \mathbb{Y}^\circ : \left\langle p \,|\, y^{\mathrm{S}} \right\rangle - \left\langle r \,|\, k^{\mathrm{S}} \right\rangle = -1 \right\}. \tag{5.12.15}$$

Such a Δ can serve as a replacement for the finite set $\mathcal{G}'$ that generates $\mathbb{Y}^\circ$ when $\mathbb{Y}$ is a solid polyhedral cone in a finite-dimensional space. The Banach lattice of all weakly* continuous functions on Δ, denoted by $\mathcal{C}(\Delta)$, replaces $\mathbb{R}^{\mathcal{G}'}$ and serves as the codomain (L) for the operations A' and B'. These are specified by[54]

$$(A'y - B'k)(p,r) := \langle p \,|\, y \rangle - \langle r \,|\, k \rangle \quad \text{for } (p,r) \in \Delta. \tag{5.12.16}$$

So $\mathcal{C}(\Delta)$ is the space of standard perturbations, and the space of standard dual variables (the constraints' multipliers) is the space of all finite Borel measures $\mathcal{M}(\Delta) = \mathcal{C}^*(\Delta)$ by Riesz's Representation Theorem. Some points of Δ are convex combinations of others. This redundancy can be lessened by replacing Δ with any *closed*, and hence compact, subset $\mathcal{G}'$ such that $\mathrm{cl\,conv}\,\mathcal{G}' = \Delta$. When the set of extreme points $\mathrm{ext}\,\Delta$ is closed, it is the best choice of $\mathcal{G}'$ (and all the redundancy is thus removed). But generally $\mathrm{ext}\,\Delta$ need not be closed, even if Δ is finite-dimensional.

[53] The output space is $Y = L^\infty[0,T]$, which has a predual $Y' = L^1[0,T]$. The fixed-input space K depends on the technique: it is either $\mathbb{R}$ for a thermal technique, or $\mathbb{R}^2$ for pumped storage, or $\mathbb{R}^2 \times L^\infty[0,T]$ for hydro. As for L (the space of standard perturbations of inequality constraints), it is either $L^\infty[0,T]$ or its Cartesian product with $\mathcal{C}[0,T]$ when, in the case of an energy storage technique, there are reservoir constraints in addition to generation constraints. And the balance constraint of a storage techniques has $\mathbb{R}$ as X (the space of standard perturbations of the equality constraint).

[54] The construction can be extended to the case that $\mathbb{Y}$ is relatively solid, i.e., has a nonempty interior in the linear subspace $\mathbb{Y} - \mathbb{Y}$ (assumed to be closed in $Y \times K$); the polar $\mathbb{Y}^\circ$ is then the sum of the annihilator $(\mathbb{Y} - \mathbb{Y})^\perp$ and a cone with a compact base Δ.

5.13 Technologies for electricity generation and energy storage

The rudimentary peak-load pricing example of Section 5.2 is next developed into a continuous-time equilibrium model of electricity pricing. This requires a fuller description of the industry's technology to start with. A typical electricity supply industry uses a combination of thermal generation, hydro, pumped energy storage, and other techniques. A thermal plant can be classified by fuel type as, e.g., nuclear, coal-, oil- or gas-fired. A hydro plant can be classified by head height as high-, medium-, or low-head. A pumped-storage plant can be classified by its medium for energy storage as, e.g., a pumped-water or compressed-air plant (PWES or CAES plant), a superconducting magnetic coil (SMES plant) or a battery. Each type can be further subdivided by the relevant design characteristics, which all affect the plant's unit input costs as well as its technical performance parameters (such as response time and efficiency of energy conversion). But the structure of feasible input-output bundles is nearly the same for all the techniques within each of the three main types (thermal, hydro and pumped storage). To simplify these technology structures, some of the cost complexities and technical imperfections are ignored:

1. A thermal plant is assumed to have a constant technical efficiency η, i.e., a constant heat rate (both incremental and average) of $1/\eta$.[55] So the plant has a constant unit running cost w (in \$/kWh, say) over the entire load range from zero to the plant's capacity.[56]

2. A hydro plant is assumed to have a constant head, and a turbine-generator of a constant technical efficiency.[57]

3. In a pumped-storage plant, the energy converter is taken to be perfectly efficient and symmetrically reversible (i.e., capable of converting both ways, and at the same rate).[58]

4. All plant types are assumed to have no startup or shutdown costs or delays.[59]

[55] A steam plant's efficiency is the product of the boiler's and turbine-generator's efficiencies, which is about $0.85 \times 0.45 \approx 38\%$ (i.e., the heat rate is about $1/0.38 \times 3600$ kJ/kWh≈ 9500 kJ/kWh).

[56] In reality, the minimum operating load is 10% to 25% of the maximum, and the incremental rate rises with load by up to 5% to 15%. Also, there is a no-load heat input (which is a sunk operating cost per unit time of being on line). See, e.g., [60, Figures 8.2 and 8.3, and Table 8.3].

[57] In reality, a turbine's efficiency varies with the load (from about 85% to 95% for movable-blade types, or 70% to 95% for fixed-blade types). Also, a plant's head varies with the water stock. The variation tends to be larger in lower-head plants, but it much depends on the particular plant: e.g., with a typical medium head (say about 150 m), the variation is 3% of the maximum in some plants, but over 30% in others. For a variable-head plant, the operation and valuation problems are studied in [48].

[58] In reality, the round-trip conversion efficiency η_{Ro} is close to 1 in SMES (over 95%). In PWES and CAES, η_{Ro} is around 70% to 75% (i.e., 0.7 kWh of electricity is recovered from every kWh used up). The case of $\eta_{\mathrm{Ro}} < 1$ is included in the model of pumped storage in Chapter 3, as are the cases of converter asymmetry or nonreversibility (although reversibility is usual, some high-head PWES plants do use nonreversible multi-stage pumps).

[59] In reality, startup times range from nearly zero for some energy storage plants (SMES coils and batteries can

5. Like operation, investment is assumed to be divisible.

Some of these conditions—viz., perfect conversion in pumped storage and constant head in hydro—are imposed purely to simplify this presentation, and can be removed by using the results of Chapter 3 and [48]. As for indivisibility, it does not loom large in large-scale systems (nor does the sunk operating cost of a thermal plant, i.e., the no-load fuel cost of its being on line). Also, the model can be extended to include transmission costs and constraints.

The one restriction that *cannot* be relaxed without changing some of the model's mathematical foundations is the assumption of immediate startup at no cost. This condition means that the thermal operating cost is additively separable over time; it also means that both short-run and long-run thermal generation costs are symmetric (a.k.a. rearrangement-invariant) functions of the output trajectory over the cycle. These properties are fundamental to the integral formulae for the short-run and long-run thermal costs,[60] and hence also to the method of calculating the long-run marginal cost of thermal generation in Chapter 2 and in [36]. The symmetry property, and its weaker variants for other techniques, underlies also the time-continuity result for the equilibrium price function [45]. And price continuity is what guarantees that the two capacities of a pumped-storage plant (viz., the reservoir and the energy converter) have well-defined and separate profit-imputed marginal values, despite their "perfect complementarity": see Chapter 3. In the case of a hydro plant, it also guarantees that the river flows have well-defined marginal values (as do the reservoir and turbine capacities): see Chapter 4.

But the assumption of no startup costs can be rather less distorting than it may seem. This is because the slow-starting plants tend to have low unit running costs, and the quick-starting plants tend to have high unit running costs. To minimise the operating cost, one allocates the base load to the lowest-cost plants, and the near-peak loads to the highest-cost plants. Thus the slowest starters end up serving mainly the constant load levels (the base load), and the quickest starters end up serving the most intermittent load levels (the near-peaks)—even if the differences in startup times are disregarded in the despatch policy.

The complete generating technology consists, then, of the various thermal, hydro and pumped-storage techniques, which form three sets: Θ_{Th}, Θ_{H} and Θ_{PS}. However, what is considered here is a smaller model with a number of thermal techniques and just one other, viz., either a pumped-storage technique or a hydro technique. So the single non-thermal technique can be denoted simply by PS or H, and the set of thermal techniques by $\{1, 2, \ldots, \Theta\}$, where Θ means the number of thermal

switch from charging to discharging in 4 to 20 miliseconds), through a few minutes (1–10 min) for other storage plants (PWES or CAES) as well as gas turbines and hydro plants, to hours for nuclear or fossil (coal, oil, gas) steam-plants (whose long startup times must of course be distinguished from the very much shorter loading times applicable to the spinning reserves): see, e.g., [60, Table 8.2] and [63] or [13].

[60] For a one-station technology, the thermal SRC and LRC are given by (5.2.5) and (5.2.6). The formulae are extended to a multi-station technology in Chapter 2.

techniques. In other words, the ESI's set of techniques is henceforth either $\{1, 2, \ldots, \Theta; \mathrm{PS}\}$ or $\{1, 2, \ldots, \Theta; \mathrm{H}\}$. It plays the role of the abstract set Θ of Sections 5.10 and 5.11.

The output space Y is here $L^\infty[0,T]$, which is the vector space of all essentially bounded real-valued functions on the interval $[0,T]$ that represents the cycle. Functions equal almost everywhere, w.r.t. the Lebesgue measure (meas), are identified with one another. With the usual order $\leq$ and the supremum norm

$$\|y\|_\infty := \mathrm{EssSup}\,|y| = \operatorname*{ess\ sup}_{t\in[0,T]} |y(t)|$$

L^∞ is a dual Banach lattice.[61] Its Banach predual is $L^1[0,T]$, the space of all integrable functions. When, as here, it serves as the price space P, a TOU electricity price is a density function, i.e., a time-dependent rate $p(t)$ in $\$/\mathrm{kWh}$. The price space $L^1[0,T]$ is sufficient in the case of interruptible demand because capacity charges are then spread out over a flattened peak: see [43]. A larger price space is needed to accommodate the instantaneous capacity charge that arises in the case of a firm, pointed peak.[62]

A *thermal technique* generates an output flow $y \in L^\infty_+[0,T]$ from two input quantities: k (in kW) of generating capacity, and v (in kWh) of fuel of the matching kind. Its long-run production set is the convex cone

$$\mathbb{Y}_{\mathrm{Th}} := \left\{ (y; -k, -v) \in L^\infty_+ \times \mathbb{R}^2_- : y \leq k,\ \frac{1}{\eta}\int_0^T y(t)\,\mathrm{d}t \leq v,\ y \geq 0 \right\} \tag{5.13.1}$$

where the constant η is the *efficiency of energy conversion* (the ratio of electricity output to heat input). The *unit fuel cost* w (in \$ per kWh of electricity output) is the fuel's price (in \$ per kWh of heat input) times the *heat rate* $1/\eta$. Henceforth, it is taken to represent all of the unit running cost (a.k.a. operating or variable cost).

There is a number of thermal techniques $\theta = 1, 2, \ldots, \Theta$. Each has the same structure (5.13.1), but it uses its own input commodities, viz., the capacity of type θ and the suitable type of fuel, ξ_θ: in terms of (5.11.1), $\Phi_\theta = \{\theta\}$ and $\Xi_\theta = \{\xi_\theta\}$. Its production set, $\mathbb{Y}_\theta$, is formally $\mathbb{Y}_{\mathrm{Th}}$ embedded in the full commodity space by inserting zeros in the input-output bundle at all the positions other than θ, ξ_θ and the t's (as in Section 5.11). The relevant quantities and prices are indicated with the

[61] For Banach-lattice theory, see, e.g., [2, Chapter 4], [11, XV.12] and [53, Chapter X].

[62] An instantaneous charge can be represented by a point measure; in the context of electricity pricing, this is a capacity charge in $\$$ per kW of *power* taken at the peak instant, and it is additional to the marginal fuel charge, which is a price density in $\$$ per kWh of *energy* at any time. A general singular measure can be interpreted as a concentrated charge. As is pointed out in [43, Sections 1 and 2], the Banach dual $L^{\infty *}$ can be useful in arriving at such a price representation when the equilibrium allocation lies actually in the space of continuous functions $C[0,T] \subset L^\infty[0,T]$. This is because the restriction, to C, of a linear functional $p \in L^{\infty *}$ has the Riesz representation by a (countably additive) measure $p_C \in \mathcal{M} = C^*$, which can have a singular part as well as a density part. The failure of $L^{\infty *}$ itself to have a tractable mathematical form is thus side-stepped without restricting the analysis to the case of price densities. (The alternative of working entirely within C and $\mathcal{M}$ as the commodity and price spaces is suitable when *all* demand is uninterruptible [39]. When all demand is harmlessly interruptible, the equilibrium price is a density [43].)

subscript θ: technique θ generates an output flow y_θ from an input k_θ of generating capacity of type θ and from an input v_θ of fuel of type ξ_θ. Its unit fuel cost is its heat rate $1/\eta_\theta$ times its *fuel's price* $\widetilde{w}_{\xi_\theta}$. The *unit fuel cost* of plant type θ is denoted by[63]

$$w_\theta := \widetilde{w}_{\xi_\theta}/\eta_\theta.$$

Without loss of generality, one can assume that the thermal techniques are numbered in the order of increasing unit operating cost (a.k.a. the merit order), i.e., that

$$w_1 < w_2 < \ldots < w_\Theta. \tag{5.13.2}$$

This condition is preserved under small changes in w.

Comment: Thermal generation is a technique with conditionally fixed coefficients, i.e., its conditional input demands depend on the output bundle y, but not on the input prices. The capacity and fuel requirement functions of technique θ are:

$$\check{k}_\theta(y) = \mathrm{EssSup}(y) := \operatorname{ess}\sup_{t\in[0,T]} y(t) \tag{5.13.3}$$

$$\check{v}_\theta(y) = \frac{1}{\eta_\theta}\int_0^T y(t)\,\mathrm{d}t \tag{5.13.4}$$

for $y \in L^\infty_+[0,T]$. For a general treatment of technologies with conditionally fixed coefficients, see [46].

Pumped storage produces a signed output flow $y \in L^\infty[0,T]$ from the inputs of storage capacity k_{St} (in kWh) and conversion capacity k_{Co} (in kW). Energy is moved in and out of the reservoir with a converter, which is taken to be perfectly efficient and symmetrically reversible: this means that, in unit time, a unit converter can either turn a unit of electricity into a unit of the storable energy, or *vice versa*. So the output from storage, $y = y^+ - y^-$, equals the rate of energy flow from the reservoir, $-\dot{s} = -\mathrm{d}s/\mathrm{d}t$ (where $s(t)$ is the energy stock at time t). Energy can be held in storage at no running cost (or loss of stock). The long-run production set is, therefore, the convex cone

$$\mathbb{Y}_{\mathrm{PS}} := \left\{ (y; -k_{\mathrm{St}}, -k_{\mathrm{Co}}) \in L^\infty[0,T]\times\mathbb{R}^2_- : |y| \le k_{\mathrm{Co}},\ \int_0^T y(t)\,\mathrm{d}t = 0 \right.$$
$$\left. \text{and } \exists s_0 \in \mathbb{R}\ \forall t\in[0,T]\ 0 \le s_0 - \int_0^t y(\tau)\,\mathrm{d}\tau \le k_{\mathrm{St}} \right\}. \tag{5.13.5}$$

Comment: This is also a technique with conditionally fixed coefficients. The requirements for storage capacity and conversion capacity, when the (signed) output from storage is y with $\int_0^T y(t)\,\mathrm{d}t$

[63] Note that each w_θ can be interpreted as the price of fuel of kind ξ_θ if different types of plants use different fuels (i.e., if $\xi_{\theta'} \neq \xi_{\theta''}$ for $\theta' \neq \theta''$). Each fuel can then be unambiguously measured in kWh of generated electricity instead of being measured as the heat input (and such measurement redefines the plant's efficiency as 1, thus equating its unit fuel cost to the price of its fuel).

$= 0$, are:

$$\check{k}_{\mathrm{St}}(y) = \max_{t\in[0,T]} \int_0^t y(t)\,\mathrm{d}t + \max_{t\in[0,T]} \int_t^T y(t)\,\mathrm{d}t \tag{5.13.6}$$

$$\check{k}_{\mathrm{Co}}(y) = \|y\|_\infty = \operatorname{ess}\sup_{t\in[0,T]} |y(t)|\,. \tag{5.13.7}$$

In these terms, $(y, -k_{\mathrm{St}}, -k_{\mathrm{Co}}) \in \mathbb{Y}_{\mathrm{PS}}$ if and only if:

$$\int_0^T y(t)\,\mathrm{d}t = 0,\ \check{k}_{\mathrm{St}}(y) \le k_{\mathrm{St}} \text{ and } \check{k}_{\mathrm{Co}}(y) \le k_{\mathrm{Co}}. \tag{5.13.8}$$

Formula (5.13.6) is derived in Chapter 3.

Hydro generation produces an output flow $y \in L_+^\infty[0,T]$ from the inputs of storage capacity k_{St} (in kWh), turbine capacity k_{Tu} (in kW) and river flow $e \in L_+^\infty[0,T]$, whose rate $e(t)$ can also be measured in units of power (instead of volume per unit time). This is because the height at which water flows in and is stored, called the head, is taken to be constant. So the potential energy of water is in a constant proportion to its volume, and the energy can be referred to as "water". Since the turbine-generator's efficiency η_{Tu} is also taken to be constant, water can be measured as the output it actually yields on conversion (i.e., in kWh of electric energy). This redefines η_{Tu} as 1, i.e., in unit time, a unit turbine can convert a unit of stock into a unit of output.

A hydroelectric water storage policy generally consists of an output $y(t) \ge 0$ and a spillage $\sigma(t) \ge 0$. The resulting net outflow from the reservoir is $-\dot{s} = y - e + \sigma$ (where $s(t)$ is the water stock at time t, and $e(t)$ is the rate of river flow). Water can be held in storage at no running cost (or loss of stock). The long-run hydro production set is, therefore, the convex cone

$$\begin{aligned}\mathbb{Y}_{\mathrm{H}} := \Bigg\{ &(y; -k_{\mathrm{St}}, k_{\mathrm{Tu}}; -e) \in L_+^\infty[0,T] \times \mathbb{R}_-^2 \times L_-^\infty[0,T] : 0 \le y \le k_{\mathrm{Tu}} \\ &\text{and } \exists \sigma \in [0,e] \left(\int_0^T (y(t) - e(t) + \sigma(t))\,\mathrm{d}t = 0 \text{ and} \right. \\ &\left. \exists s_0 \in \mathbb{R}\ \forall t \quad 0 \le s_0 - \int_0^t (y(\tau) - e(\tau) + \sigma(\tau))\,\mathrm{d}\tau \le k_{\mathrm{St}} \right) \Bigg\}. \end{aligned} \tag{5.13.9}$$

Comments (on hydro and pumped storage):

- If $k_{\mathrm{Tu}} \ge e$ then there is no need for spillage and, furthermore, it is feasible for the hydro plant to "coast", i.e., to generate at the rate $y(t) = e(t)$. In this case, all the incentive to use the reservoir comes from a time-dependent output price: if p were a constant, the plant might as well coast all the time.

- In both pumped storage and hydro generation, the flows to and from the reservoir are required to balance over the cycle ($\int_0^T \dot{s}(t)\,\mathrm{d}t = 0$), i.e., the stock must be a periodic function of time. But its level at the beginning or end of a cycle is taken to be a costless decision variable, s_0. In other words, when it is first commissioned, the reservoir comes filled up to any required level at no extra cost, but its periodic operation thereafter is taken to be a technological constraint.

- In some ways, the hydro technology is analytically similar to pumped storage. But, unlike pumped storage, hydro is not a technique with conditionally fixed coefficients: although the conditional input demand for the turbine depends only on the output (it is $\check{k}_{\mathrm{Tu}}(y) = \mathrm{EssSup}(y)$), various combinations of an inflow function and a reservoir capacity can yield the same hydro output y (e.g., any e with $\int_0^T e(t)\,\mathrm{d}t = \int_0^T y(t)\,\mathrm{d}t$ and a high enough k_{St} will do).

5.14 Operation and valuation of electric power plants

For each of the plant types described in Section 5.13, the problem of profit-maximising operation can be formulated as a doubly infinite linear programme for the output rate $y(t)$ at each time t (in kW), given a TOU electricity price rate $p(t)$ for each time t (in $\$$/kWh).

For a (single) *thermal* plant of capacity k with a unit running cost w, the operation LP (reduced by working out the short-run cost as $w \int_0^T y\,\mathrm{d}t$) is:

$$\text{Given } (p, k, w) \in L^1[0,T] \times \mathbb{R}_+ \times \mathbb{R}_+ \tag{5.14.1}$$

$$\text{maximize} \int_0^T (p(t) - w)\, y(t)\,\mathrm{d}t \text{ over } y \in L^\infty[0,T] \tag{5.14.2}$$

$$\text{subject to:} \quad 0 \le y(t) \le k \text{for a.e. } t. \tag{5.14.3}$$

Every optimal output is given by

$$y(t) \in \begin{cases} \{0\} & \text{for } p(t) < w \\ [0,k] & \text{for } p(t) = w \\ \{k\} & \text{for } p(t) > w \end{cases} \tag{5.14.4}$$

i.e., measurable functions satisfying (5.14.4) form the solution set $\hat{Y}_{\mathrm{Th}}(p,k,w)$. So the thermal plant's operating profit is $\Pi_{\mathrm{SR}}^{\mathrm{Th}}(p,k,w) = k \int_0^T (p(t) - w)^+\,\mathrm{d}t$, and its unit rental value (in $\$$/kW) is

$$\hat{r}_{\mathrm{Th}}(p,k,w) = \frac{\partial \Pi_{\mathrm{SR}}^{\mathrm{Th}}}{\partial k}(p,k,w) = \int_0^T (p(t) - w)^+\,\mathrm{d}t \quad \text{if } k > 0. \tag{5.14.5}$$

Differentiation is the simplest way to value a unit of thermal capacity because the operation problem is so simple that its solution and value function can be calculated directly (i.e., without using a duality method). Of course, $\hat{r}_{\mathrm{Th}}$ can also be calculated by solving the dual problem of capacity valuation. The standard dual of the operation LP is the following programme for the flow of rent κ (whose total for the cycle is r), with ν as the Lagrange multiplier for the nonnegativity constraint on y in (5.14.3):

$$\text{Given } (p,k,w) \text{ as in (5.14.1)} \tag{5.14.6}$$

$$\text{minimize } k \int_0^T \kappa(t)\,\mathrm{d}t \quad \text{over } \kappa \in L^1[0,T] \text{ and } \nu \in L^1[0,T] \tag{5.14.7}$$

$$\text{subject to:} \quad \kappa \geq 0, \ \nu \geq 0 \tag{5.14.8}$$

$$p(t) - w = \kappa(t) - \nu(t) \quad \text{for a.e. } t. \tag{5.14.9}$$

The standard dual's inclusive form, introduced in (5.12.13)–(5.12.14), has also the dependent decision variable

$$r = \int_0^T \kappa(t)\,\mathrm{d}t \tag{5.14.10}$$

which is the thermal plant's unit rental value. The standard dual solution, unique if $k > 0$, is

$$\hat{\kappa}_{\mathrm{Th}}(p,w) = (p-w)^+ \quad \text{and} \quad \hat{\nu}_{\mathrm{Th}}(p,w) = (p-w)^- \tag{5.14.11}$$

and hence, again,

$$\hat{r}_{\mathrm{Th}} = \int_0^T \hat{\kappa}_{\mathrm{Th}}(t)\,\mathrm{d}t = \int_0^T (p(t)-w)^+\,\mathrm{d}t.$$

Comments (comparison of standard and intrinsic duals of the thermal plant operation programme):

- The standard perturbation of the primal LP (5.14.1)–(5.14.3), which produces the dual LP (5.14.6)–(5.14.9), consists in adding cyclically varying increments $(\Delta k(t), \Delta n(t))$ to the constants $(k, 0) \in \mathbb{R} \times \mathbb{R}$ in (5.14.3). The resource increments, $(\Delta k, -\Delta n) \in L^\infty \times L^\infty$, are paired with Lagrange multipliers $(\kappa, \nu) \in L^1 \times L^1$.

- By giving the unit rent's distribution over time, κ—rather than only its total for the cycle, r—the standard dual LP is the "fine" form of the valuation problem (in the sense of the first Comment in Section 5.12, with the integral $\kappa \mapsto \int \kappa(t)\,\mathrm{d}t$ as the adjoint operation $\sigma \mapsto B^{\mathrm{T}}\sigma$). The "coarse" form of valuation is a case of the intrinsic dual (5.5.13)–(5.5.14). It can be formulated as an LP for the single variable r by using the input requirement functions (5.13.3) and (5.13.4). See [46] for details.

- In terms of the general duality scheme (Sections 5.5 and 5.12), r is the intrinsic dual variable. Correspondence of notation between that scheme and its applications to the ESI is spelt out in Table 5.3.

For a *pumped-storage* plant with capacities $(k_{\mathrm{St}}, k_{\mathrm{Co}})$, the operation LP is:

$$\text{Given } (p; k_{\mathrm{St}}, k_{\mathrm{Co}}) \in L^1[0,T] \times \mathbb{R}_+ \times \mathbb{R}_+ \tag{5.14.12}$$

$$\text{maximise} \int_0^T p(t)\,y(t)\,\mathrm{d}t \text{ over } y \in L^\infty[0,T] \text{ and } s_0 \in \mathbb{R} \tag{5.14.13}$$

$$\text{subject to:} \quad -k_{\mathrm{Co}} \leq y(t) \leq k_{\mathrm{Co}} \quad \text{for a.e. } t \tag{5.14.14}$$

	intrinsic primal param.	intrinsic dual vbles	std primal param.	std dual vbles	relationship between intrinsic and standard dual vbles
GS	k [vect]	r [vect]	s [vect]	σ [vect]	$r = B^{\mathrm{T}}\sigma$
Th	k [scal]	r [scal]	$k\,(\cdot)$ $n\,(\cdot)$	$\kappa\,(\cdot)$ $\nu\,(\cdot)$	$r = \int \kappa \, \mathrm{d}t$
PS	k_{St} [scal] k_{Co} [scal]	r_{St} [scal] r_{Co} [scal]	$k_{\mathrm{St}}\,(\cdot)$ $n_{\mathrm{St}}\,(\cdot)$ $k_{\mathrm{Tu}}\,(\cdot)$ $k_{\mathrm{Pu}}\,(\cdot)$ ζ	$\kappa_{\mathrm{St}}\,(\mathrm{d}\cdot)$ $\nu_{\mathrm{St}}\,(\mathrm{d}\cdot)$ $\kappa_{\mathrm{Tu}}\,(\cdot)$ $\kappa_{\mathrm{Pu}}\,(\cdot)$ λ	$r_{\mathrm{St}} = \int \kappa_{\mathrm{St}}\,(\mathrm{d}t)$ $r_{\mathrm{Co}} = \int (\kappa_{\mathrm{Tu}} + \kappa_{\mathrm{Pu}})\,\mathrm{d}t$
Hy	k_{St} [scal] k_{Tu} [scal] $e\,(\cdot)$	r_{St} [scal] r_{Tu} [scal] $\psi\,(\cdot)$	$k_{\mathrm{St}}\,(\cdot)$ $n_{\mathrm{St}}\,(\cdot)$ $k_{\mathrm{Tu}}\,(\cdot)$ $n_{\mathrm{Tu}}\,(\cdot)$ ζ	$\kappa_{\mathrm{St}}\,(\mathrm{d}\cdot)$ $\nu_{\mathrm{St}}\,(\mathrm{d}\cdot)$ $\kappa_{\mathrm{Tu}}\,(\cdot)$ $\nu_{\mathrm{Tu}}\,(\cdot)$ λ	$r_{\mathrm{St}} = \int \kappa_{\mathrm{St}}\,(\mathrm{d}t)$ $r_{\mathrm{Tu}} = \int \kappa_{\mathrm{Tu}}\,\mathrm{d}t$ $\psi\,(t) = \lambda + (\kappa_{\mathrm{St}} - \nu_{\mathrm{St}})\,[0,t]$

TABLE 5.3. Correspondence of notation between the general duality scheme (Sections 5.5 and 5.15) and its applications to the ESI (Section 5.17). The abbreviations read: (i) in the leftmost column: GS = general scheme, Th = thermal generation, PS = pumped storage, Hy = hydro generation; (ii) elsewhere: St = storage reservoir, Co = converter, Pu/Tu = pump/turbine (two working modes of a reversible PS converter), Tu = hydro turbine. Functions of time are marked with a $(\cdot)$, and measures on the time interval are marked with a $(\mathrm{d}\cdot)$. In the general scheme, s and σ mean the standard parameters and Lagrange multipliers. But in the context of storage (both PS and Hy), s means the energy stock (and σ means spillage in Hy). Also, the intrinsic parameters and dual variables of the general scheme, r and k, correspond to (r, ψ) and (k, e) in the hydro problem.

$$\int_0^T y(t)\,\mathrm{d}t = 0 \tag{5.14.15}$$

$$0 \leq s_0 - \int_0^t y(\tau)\,\mathrm{d}\tau \leq k_{\mathrm{St}} \quad \text{for every } t. \tag{5.14.16}$$

Unlike the case of $\Pi_{\mathrm{SR}}^{\mathrm{Th}}$, there is no explicit formula for $\Pi_{\mathrm{SR}}^{\mathrm{PS}}(p, k_{\mathrm{St}}, k_{\mathrm{Co}})$, the operating profit of a pumped-storage plant; and both operation and rental valuation of a storage plant are best approached through the dual problem of capacity valuation. The standard dual of the operation LP is the following programme for: (i) the flow of reservoir's rent κ_{St}, and (ii) the flow of converter's rents $\kappa_{\mathrm{Co}} = \kappa_{\mathrm{Pu}} + \kappa_{\mathrm{Tu}}$, which it earns in its two modes of work, viz., charging the reservoir as a "pump" and discharging it as a "turbine". Their totals for the cycle are the unit rental values: (i) of the reservoir r_{St} (in $\$/$kWh), and (ii) of the converter r_{Co} (in $\$/$kW). The dual variables κ_{Pu} and κ_{Tu} range over $L^1[0,T]$, like the κ in (5.14.7). The space for κ_{St} is $\mathcal{M}[0,T]$, the space of Borel measures on $[0,T]$, which is the norm-dual of the space of continuous functions $\mathcal{C}[0,T]$. This is also the space for the multiplier ν_{St} for the nonnegativity constraint in (5.14.16). The multiplier for the balance constraint (5.14.15) is a scalar λ. So the LP of capacity valuation for a pumped-storage plant is:

$$\text{Given } (p; k_{\mathrm{St}}, k_{\mathrm{Co}}) \text{ as in (5.14.12)} \tag{5.14.17}$$

$$\text{minimise } k_{\mathrm{St}} \int_{[0,T]} \kappa_{\mathrm{St}}(\mathrm{d}t) + k_{\mathrm{Co}} \int_0^T (\kappa_{\mathrm{Tu}} + \kappa_{\mathrm{Pu}})(t)\,\mathrm{d}t \tag{5.14.18}$$

$$\text{over } \lambda \in \mathbb{R} \text{ and } (\kappa_{\mathrm{St}}, \nu_{\mathrm{St}}, \kappa_{\mathrm{Pu}}, \kappa_{\mathrm{Tu}}) \in \mathcal{M} \times \mathcal{M} \times L^1 \times L^1 \tag{5.14.19}$$

$$\text{subject to: } \quad (\kappa_{\mathrm{St}}, \nu_{\mathrm{St}}, \kappa_{\mathrm{Pu}}, \kappa_{\mathrm{Tu}}) \geq 0 \tag{5.14.20}$$

$$\kappa_{\mathrm{St}}[0,T] = \nu_{\mathrm{St}}[0,T] \tag{5.14.21}$$

$$p(t) = \lambda + (\kappa_{\mathrm{St}} - \nu_{\mathrm{St}})[0,t] + \kappa_{\mathrm{Tu}}(t) - \kappa_{\mathrm{Pu}}(t) \quad \text{for a.e. } t. \tag{5.14.22}$$

The standard dual's inclusive form has also the dependent decision variables

$$r_{\mathrm{St}} = \int_0^T \kappa_{\mathrm{St}}(\mathrm{d}t) = \kappa_{\mathrm{St}}[0,T] \tag{5.14.23}$$

$$r_{\mathrm{Co}} = \int_0^T (\kappa_{\mathrm{Pu}}(t) + \kappa_{\mathrm{Tu}}(t))\,\mathrm{d}t \tag{5.14.24}$$

as per the last constraint of (5.12.14).

Comments (comparison of standard and intrinsic duals of the pumped-storage plant operation programme):

- The standard perturbation of the primal LP (5.14.12)–(5.14.16), which produces the dual LP (5.14.17)–(5.14.22), uses cyclically varying increments $(\Delta k_{\mathrm{St}}(t), \Delta n_{\mathrm{St}}(t))$ to the constants (k_{St}, 0) in (5.14.16). It also uses two separate increments $(\Delta k_{\mathrm{Pu}}(t), \Delta k_{\mathrm{Tu}}(t))$ to the two occurrences of k_{Co} in (5.14.14)—i.e., (5.14.14) is perturbed to:

$$-k_{\mathrm{Co}} - \Delta k_{\mathrm{Pu}}(t) \leq y \leq k_{\mathrm{Co}} + \Delta k_{\mathrm{Tu}}(t).$$

Additionally, a scalar $\Delta\zeta$ is used as an increment to the 0 on the r.h.s. of (5.14.15). The resource increments $\Delta k_{\mathrm{St}} \in \mathcal{C}$, $-\Delta n_{\mathrm{St}} \in \mathcal{C}$, $\Delta k_{\mathrm{Tu}} \in L^{\infty}$, $\Delta k_{\mathrm{Pu}} \in L^{\infty}$ and $\Delta\zeta \in \mathbb{R}$ are paired with the Lagrange multipliers $\kappa_{\mathrm{St}} \in \mathcal{M}$, $\nu_{\mathrm{St}} \in \mathcal{M}$, $\kappa_{\mathrm{Tu}} \in L^1$, $\kappa_{\mathrm{Pu}} \in L^1$ and $\lambda \in \mathbb{R}$. This perturbation scheme is described in detail in Chapter 3.

- By giving the distributions of unit rents over time (and over the two conversion modes), κ_{St} and $\kappa_{\mathrm{Pu}} + \kappa_{\mathrm{Tu}}$—rather than only their totals for the cycle, r_{St} and r_{Co}—the standard dual LP is the "fine" form of the valuation problem (in the sense of the first Comment in Section 5.12). The "coarse" form of valuation is a case of the intrinsic dual (5.5.13)–(5.5.14). It can be formulated as a semi-infinite LP for the variables r_{St} and r_{Co}, with an infinity of constraints, by using the input requirement functions (5.13.6)–(5.13.7). See [46] for details.

The storage-plant valuation LP (5.14.17)–(5.14.22) can be transformed into an *unconstrained* convex programme by changing the variables from λ, $\kappa_{\mathrm{St}}(\mathrm{d}t)$ and $\nu_{\mathrm{St}}(\mathrm{d}t)$ to

$$\psi(t) = \lambda + (\kappa_{\mathrm{St}} - \nu_{\mathrm{St}})[0,t] \quad \text{for } t \in (0,T) \tag{5.14.25}$$

and by substituting $(p-\psi)^+$ and $(p-\psi)^-$ for κ_{Tu} and κ_{Pu} to eliminate these variables: see Chapter 3 for details. The new continuum of variables, ψ, is a function of bounded variation that can be interpreted as the TOU marginal value of the energy stock, i.e., its *shadow price*.

Notation The space $\mathrm{BV}(0,T)$ consists of all functions ψ of bounded variation on $(0,T)$ with $\psi(t)$ lying between the left and right limits, $\psi(t-) = \lim_{\tau\nearrow t}\psi(\tau)$ and $\psi(t+) = \lim_{\tau\searrow t}\psi(\tau)$.[64] A $\psi \in \mathrm{BV}(0,T)$ is extended by continuity to $[0,T]$; i.e., $\psi(0) := \psi(0+)$ and $\psi(T) := \psi(T-)$. The *cyclic positive variation* of ψ is

$$\mathrm{Var}_{\mathrm{c}}^{+}(\psi) := \mathrm{Var}^{+}(\psi) + (\psi(0) - \psi(T))^{+} \tag{5.14.26}$$

where $\mathrm{Var}^{+}(\psi)$ is the total *positive variation* (a.k.a. upper variation) of ψ, i.e., the supremum of $\sum_m (\psi(\overline{\tau}_m) - \psi(\underline{\tau}_m))^{+}$ over all finite sets of pairwise disjoint subintervals $(\underline{\tau}_m, \overline{\tau}_m)$ of $(0,T)$: see, e.g., [27, Section 8.1] for details.[65]

In these terms, the capacity valuation problem (for a pumped-storage plant) becomes the following programme for shadow-pricing the energy stock:

$$\text{Given } (p; k_{\mathrm{St}}, k_{\mathrm{Co}}) \in L^1[0,T] \times \mathbb{R}_+ \times \mathbb{R}_+ \tag{5.14.27}$$

$$\text{minimise } k_{\mathrm{St}} \mathrm{Var}_{\mathrm{c}}^{+}(\psi) + k_{\mathrm{Co}} \int_0^T |p(t) - \psi(t)|\, \mathrm{d}t \quad \text{over } \psi \in \mathrm{BV}(0,T). \tag{5.14.28}$$

[64] The one-sided limits exist at every t and are equal nearly everywhere (n.e.), i.e., everywhere except for a countable set. Specification of $\psi(t)$ between $\psi(t-)$ and $\psi(t+)$ is unnecessary.

[65] The other term, $(\psi(0) - \psi(T))^{+}$, represents a possible jump of ψ at the instant separating two consecutive cycles.

Its main feature is the trade-off between minimising the variation (which on its own would require setting ψ at a constant value) and minimising the integral (which on its own would require setting ψ equal to p). This trade-off is what determines the extent to which local peaks of p should be "shaved off" and the troughs "filled in" to obtain the optimum shadow price function $\hat{\psi}_{\mathrm{PS}}(p, k_{\mathrm{St}}, k_{\mathrm{Co}})$, at least in the case of a piecewise strictly monotone p. The solution, shown in Figure 5.4a, is determined by constancy intervals for $\hat{\psi}_{\mathrm{PS}}$ around a local peak or trough of p (as a function of t). Unless $k_{\mathrm{St}}/k_{\mathrm{Co}}$ is relatively long, these intervals do not abut, and must all be of that length.[66] The optimal output has the "bang-coast-bang" form

$$\hat{y}_{\mathrm{PS}}(t; p, k_{\mathrm{St}}, k_{\mathrm{Co}}) = k_{\mathrm{Co}} \operatorname{sgn}\left(p(t) - \hat{\psi}_{\mathrm{PS}}(t; p, k_{\mathrm{St}}, k_{\mathrm{Co}})\right) \qquad (5.14.29)$$

i.e., $\hat{y}_{\mathrm{PS}}(t)$ equals k_{Co}, 0 or $-k_{\mathrm{Co}}$ if, respectively, $p(t) > \hat{\psi}_{\mathrm{PS}}(t)$, $p(t) = \hat{\psi}_{\mathrm{PS}}(t)$ or $p(t) < \hat{\psi}_{\mathrm{PS}}(t)$: see Figure 5.4b. The lowercase notation, $\hat{y}_{\mathrm{PS}}$ or $\hat{\psi}_{\mathrm{PS}}$, is used *only when the solution is unique*. In general, the solution sets for (5.14.12)–(5.14.16) and (5.14.27)–(5.14.28) are denoted by $\hat{Y}_{\mathrm{PS}}(p; k_{\mathrm{St}}, k_{\mathrm{Co}})$ and $\hat{\Psi}_{\mathrm{PS}}(p; k_{\mathrm{St}}, k_{\mathrm{Co}})$. More precisely, $y \in \hat{Y}_{\mathrm{PS}}$ means that y together with $s_0 = \max_t \int_0^t y(\tau)\,\mathrm{d}\tau$ (which the lowest initial stock needed for the stock $s_0 - \int_0^t y(\tau)\,\mathrm{d}\tau$ never to fall below 0) solves (5.14.12)–(5.14.16).

The stock-pricing programme (5.14.27)–(5.14.28) has a solution for every $k_{\mathrm{St}} > 0$ and $k_{\mathrm{Co}} > 0$.[67] If p is continuous, i.e., $p \in \mathcal{C}[0, T]$, then there is a *unique* solution $\hat{\psi}_{\mathrm{PS}}(p; k_{\mathrm{St}}, k_{\mathrm{Co}})$. It follows that the plant's operating profit $\Pi_{\mathrm{SR}}^{\mathrm{PS}}$ is differentiable in $(k_{\mathrm{St}}, k_{\mathrm{Co}})$; equivalently, with this technology the programme (5.5.13)–(5.5.14) has a unique solution $\hat{r}$. In terms of $\hat{\psi}_{\mathrm{PS}}$, the unit rental values of the reservoir and the converter (in \$/kWh and \$/kW, respectively) are:

$$\frac{\partial \Pi_{\mathrm{SR}}^{\mathrm{PS}}}{\partial k_{\mathrm{St}}} = \hat{r}_{\mathrm{St}}(p, k_{\mathrm{St}}, k_{\mathrm{Co}}) = \int_0^T \hat{\kappa}_{\mathrm{St}}(\mathrm{d}t) = \mathrm{Var}_{\mathrm{c}}^{+}\left(\hat{\psi}_{\mathrm{PS}}(p; k_{\mathrm{St}}, k_{\mathrm{Co}})\right) \qquad (5.14.30)$$

$$\frac{\partial \Pi_{\mathrm{SR}}^{\mathrm{PS}}}{\partial k_{\mathrm{Co}}} = \hat{r}_{\mathrm{Co}}(p, k_{\mathrm{St}}, k_{\mathrm{Co}}) = \int_0^T (\hat{\kappa}_{\mathrm{Pu}} + \hat{\kappa}_{\mathrm{Tu}})(t)\,\mathrm{d}t = \int_0^T \left|p(t) - \hat{\psi}_{\mathrm{PS}}(t)\right| \mathrm{d}t. \qquad (5.14.31)$$

For proofs, see Chapter 3.

As for the operation problem (5.14.12)–(5.14.16), it has a solution for any $p \in L^1[0, T]$ and every $(k_{\mathrm{St}}, k_{\mathrm{Co}}) \geq 0$. If p has no plateau (i.e., $\mathrm{meas}\{t : p(t) = \mathrm{p}\} = 0$ for every $\mathrm{p} \in \mathbb{R}$), then there is a unique solution $\hat{y}_{\mathrm{PS}}(p; k_{\mathrm{St}}, k_{\mathrm{Co}})$. It is given either by (5.14.29) itself (if $(k_{\mathrm{St}}, k_{\mathrm{Co}}) \gg 0$ and $p \in \mathcal{C}$), or by (5.14.29) with any $\psi \in \hat{\Psi}_{\mathrm{PS}}$ instead of $\hat{\psi}_{\mathrm{PS}}$ (if $(k_{\mathrm{St}}, k_{\mathrm{Co}}) \gg 0$ but $p \notin \mathcal{C}$). For proofs, see Chapter 3.

Comments (interpretation of ψ, and assumptions on p in the pumped-storage problem):

[66] Matters complicate when the ratio $k_{\mathrm{St}}/k_{\mathrm{Co}}$ is comparable to the durations between the successive local peaks and troughs of p, so that the neighbouring constancy intervals of $\hat{\psi}_{\mathrm{PS}}$ start to abut; but a similar optimality rule applies to such clusters.

[67] When $k_{\mathrm{St}} > 0$ but $k_{\mathrm{Co}} = 0$, any constant ψ is a solution. When $k_{\mathrm{Co}} > 0$ but $k_{\mathrm{St}} = 0$, a solution exists if and only if $p \in \mathrm{BV}$, in which case it is unique, viz., $\psi = p$.

- $\psi(t)$ has the interpretation of the shadow price of energy stock at time t. Heuristically, this follows from (5.14.25) and the marginal interpretations of κ, ν and λ, which are that: (i) κ_{St}, as the multiplier for the upper reservoir constraint, represents the reservoir capacity value, (ii) the multiplier ν_{St} has a similar interpretation for the lower reservoir constraint, and (iii) λ is the stock value at the beginning of cycle.

- This interpretation of ψ can be formalised as a rigorous marginal-value result by introducing a hypothetical inflow to the reservoir, $e \in L^{\infty}$, as a primal parameter with its own dual variable ψ. This means that (5.14.15) and (5.14.16) are perturbed by replacing y with $y - \Delta e$. Then (5.14.25) becomes a constraint of the dual problem, whose solution $\hat{\psi}_{\mathrm{PS}}$ equals $\nabla_e \Pi_{\mathrm{SR}}^{\mathrm{PS}}$ at $e = 0$. (This is formally similar to the hydro case (5.14.51), in which e is the river flow, and $\hat{\psi}_{\mathrm{H}}$ equals $\nabla_e \Pi_{\mathrm{SR}}^{\mathrm{H}}$ at the given, positive e.)

- Time-continuity of the electricity tariff p, which guarantees the uniqueness and time-continuity of the optimal price for energy stock $\hat{\psi}_{\mathrm{PS}}(p; k_{\mathrm{St}}, k_{\mathrm{Co}})$, is acceptable as an assumption for operation and valuation of storage plants because it can be verified for the general competitive equilibrium: see [45].

- Unlike price continuity, the no-plateau condition on the tariff p is rather questionable: it cannot hold in an equilibrium with continuous *quantity* trajectories (since it leads to the unique optimum $\hat{y}_{\mathrm{PS}}$, which is a discontinuous function of t because it takes only the three values $\pm k_{\mathrm{Co}}$ and 0, as per (5.14.29)).[68] Such an equilibrium is made possible only by the presence of intervals on which an optimal y can gradually change from 0 to $\pm k_{\mathrm{Co}}$ because $p = \psi = $ const. But all this means is that, with a price system consistent with output continuity, the storage operation problem is not fully solved by stock pricing alone.

For a *hydro* plant with capacities $(k_{\mathrm{St}}, k_{\mathrm{Tu}})$ and an inflow $e(t) \le k_{\mathrm{Tu}}$ (for a.e. t), the operation LP is:

$$\text{Given } (p; k_{\mathrm{St}}, k_{\mathrm{Tu}}; e) \in L_{+}^{1}[0,T] \times \mathbb{R}_{+} \times \mathbb{R}_{+} \times L_{+}^{\infty}[0,T] \text{ with } k_{\mathrm{Tu}} \ge e \tag{5.14.32}$$

$$\text{maximise } \int_0^T p(t)\, y(t)\, \mathrm{d}t \text{ over } y \in L^{\infty}[0,T] \text{ and } s_0 \in \mathbb{R} \tag{5.14.33}$$

$$\text{subject to: } \quad 0 \le y(t) \le k_{\mathrm{Tu}} \quad \text{for a.e. } t \tag{5.14.34}$$

$$\int_0^T (y(t) - e(t))\, \mathrm{d}t = 0 \tag{5.14.35}$$

$$0 \le s_0 - \int_0^t (y(\tau) - e(\tau))\, \mathrm{d}\tau \le k_{\mathrm{St}} \quad \text{for every } t. \tag{5.14.36}$$

[68] Furthermore, a time-continuous optimal output from storage cannot be unique (unless $k_{\mathrm{St}} = 0$ or $k_{\mathrm{Co}} = 0$).

As with pumped storage, there is no explicit formula for the hydro plant's operating profit $\Pi^{\mathrm{H}}_{\mathrm{SR}}(p; k_{\mathrm{St}}, k_{\mathrm{Tu}}; e)$, and both operation and rental valuation of a hydro plant are best approached through the dual problem of fixed-input valuation, which is an LP for: (i) the flow of reservoir's unit rent κ_{St}, (ii) the flow of turbine's unit rent κ_{Tu}, and (iii) the river's unit rent, i.e., the shadow price of water ψ. By including ψ but not r among the dual variables, this is a partly inclusive form of the standard dual LP. The fully inclusive form has also r_{St} and r_{Tu}, the rental values of the reservoir (in $\$/\mathrm{kWh}$) and of the turbine (in $\$/\mathrm{kW}$), but these are simply the totals of κ_{St} and κ_{Tu} for the cycle. The dual variable κ_{Tu} ranges over $L^1[0,T]$, and the space for κ_{St} is the space of measures $\mathcal{M}[0,T]$, as in pumped storage. The space for ψ can be $L^1[0,T]$ formally, but actually ψ is constrained to $\mathrm{BV}(0,T)$ by (5.14.43). The multipliers for the nonnegativity constraints in (5.14.34) and (5.14.36) are $\nu_{\mathrm{Tu}} \in L^1[0,T]$ and $\nu_{\mathrm{St}} \in \mathcal{M}[0,T]$. The multiplier for the balance constraint (5.14.35) is a scalar λ. So the LP of fixed-input valuation for a hydro plant is:

$$\text{Given } (p; k_{\mathrm{St}}, k_{\mathrm{Tu}}; e) \text{ as in (5.14.32)} \tag{5.14.37}$$

$$\text{minimise } k_{\mathrm{St}} \int_{[0,T]} \kappa_{\mathrm{St}}(\mathrm{d}t) + k_{\mathrm{Tu}} \int_0^T \kappa_{\mathrm{Tu}}(t)\,\mathrm{d}t + \int_0^T \psi(t)\, e(t)\,\mathrm{d}t \tag{5.14.38}$$

$$\text{over } \lambda \in \mathbb{R},\ \psi \in L^1[0,T] \text{ and } (\kappa_{\mathrm{St}}, \nu_{\mathrm{St}}; \kappa_{\mathrm{Tu}}, \nu_{\mathrm{Tu}}) \in \mathcal{M} \times \mathcal{M} \times L^1 \times L^1 \tag{5.14.39}$$

$$\text{subject to: } (\kappa_{\mathrm{St}}, \nu_{\mathrm{St}}; \kappa_{\mathrm{Tu}}, \nu_{\mathrm{Tu}}) \geq 0 \tag{5.14.40}$$

$$\kappa_{\mathrm{St}}[0,T] = \nu_{\mathrm{St}}[0,T] \tag{5.14.41}$$

$$p(t) = \psi(t) + \kappa_{\mathrm{Tu}}(t) - \nu_{\mathrm{Tu}}(t) \quad \text{for a.e. } t \tag{5.14.42}$$

$$\psi(t) = \lambda + (\kappa_{\mathrm{St}} - \nu_{\mathrm{St}})[0,t] \quad \text{for a.e. } t. \tag{5.14.43}$$

The dual's fully inclusive form has also the remaining dependent decision variables

$$r_{\mathrm{St}} = \int_0^T \kappa_{\mathrm{St}}(\mathrm{d}t) \tag{5.14.44}$$

$$r_{\mathrm{Tu}} = \int_0^T \kappa_{\mathrm{Tu}}(t)\,\mathrm{d}t. \tag{5.14.45}$$

Comments (comparison of the partly inclusive standard, standard, and intrinsic duals of the hydro plant operation programme):

- The perturbation that produces (5.14.37)–(5.14.43) as the dual of (5.14.32)–(5.14.36) includes an increment $\Delta e(t)$ in addition to the standard perturbation (which uses cyclically varying increments $(\Delta k_{\mathrm{St}}(t), \Delta n_{\mathrm{St}}(t); \Delta k_{\mathrm{Tu}}(t), \Delta n_{\mathrm{Tu}}(t))$ to the constants $(k_{\mathrm{St}}, 0; k_{\mathrm{Tu}}, 0)$ in (5.14.36) and (5.14.34), as well as a scalar $\Delta\zeta$ as an increment to the 0 on the r.h.s. of (5.14.35)). The resource increments $\Delta e \in L^\infty$, $\Delta k_{\mathrm{St}} \in \mathcal{C}$, $-\Delta n_{\mathrm{St}} \in \mathcal{C}$, $\Delta k_{\mathrm{Tu}} \in L^\infty$, $-\Delta n_{\mathrm{Tu}} \in L^\infty$ and $\Delta\zeta \in \mathbb{R}$ are paired with the dual variables $\psi \in L^1$, $\kappa_{\mathrm{St}} \in \mathcal{M}$, $\nu_{\mathrm{St}} \in \mathcal{M}$, $\kappa_{\mathrm{Tu}} \in L^1$, $\nu_{\mathrm{Tu}} \in L^1$ and $\lambda \in \mathbb{R}$. This perturbation scheme is described in detail in Chapter 4.

- Though it is more transparent to have an explicit dual variable for each parameter, the nonstandard dual variable ψ (paired with e) can be eliminated by replacing it in (5.14.38) and (5.14.42) with its equivalent in terms of the standard dual variables (5.14.43). This reduces the valuation LP (5.14.37)–(5.14.43) to the standard dual of the hydro operation LP (5.14.32)–(5.14.36), i.e., to the dual arising from the same perturbation as above but without Δe.

- By giving the distributions of unit rents over time, κ_{St} and κ_{Tu}—rather than only their totals for the cycle, r_{St} and r_{Tu}—the above dual LP is the "fine" form of the valuation problem. The "coarse" form of valuation is a case of the intrinsic dual (5.5.13)–(5.5.14); it is a programme for r_{St}, r_{Tu} and ψ.

The hydro-plant valuation LP (5.14.37)–(5.14.43) can be transformed into an *unconstrained* convex programme for the water price ψ by using the constraints (5.14.42) and (5.14.43) to substitute: $(p-\psi)^+$ and $(p-\psi)^-$ for κ_{Tu} and ν_{Tu}, $(\mathrm{d}\psi)^+$ and $(\mathrm{d}\psi)^-$ for κ_{St} and ν_{St}, and any number between $\psi(0+)$ and $\psi(T-)$ for λ: see Chapter 4 for details. In these terms, the fixed-input valuation problem (for a hydro plant) becomes:

$$\text{Given } (p; k_{\mathrm{St}}, k_{\mathrm{Tu}}; e) \in L^1_+[0,T] \times \mathbb{R}_+ \times \mathbb{R}_+ \times L^\infty_+[0,T] \text{ with } k_{\mathrm{Tu}} \geq e \tag{5.14.46}$$

$$\text{minimise } k_{\mathrm{St}} \operatorname{Var}^+_{\mathrm{c}}(\psi) + k_{\mathrm{Tu}} \int_0^T (p(t) - \psi(t))^+ \,\mathrm{d}t + \int_0^T \psi(t)\, e(t)\,\mathrm{d}t \tag{5.14.47}$$

$$\text{over } \psi \in \mathrm{BV}(0,T). \tag{5.14.48}$$

Recall that $\operatorname{Var}^+_{\mathrm{c}}(\psi)$, defined by (5.14.26), is the total of all rises of ψ over the cycle.

If $k_{\mathrm{Tu}} > e(t) > 0$ for every t, then the sum of the two integrals in (5.14.47) has a minimum at (and only at) $\psi = p$. Therefore, the programme's main feature is the trade-off between minimising the variation (which on its own would require setting ψ at a constant value) and minimising the sum of integrals (which on its own would require setting ψ equal to p). This trade-off is what determines the extent to which the local peaks of p should be "shaved off" and the troughs "filled in" to obtain the optimum shadow price function $\hat{\psi}_{\mathrm{H}}(p; k_{\mathrm{St}}, k_{\mathrm{Tu}}; e)$, at least in the case that p is piecewise strictly monotone and $k_{\mathrm{Tu}} > e > 0$ at all times. The solution is determined by constancy intervals for $\hat{\psi}_{\mathrm{H}}$. If $k_{\mathrm{St}}/\operatorname{Sup}(e)$ and $k_{\mathrm{St}}/(k_{\mathrm{Tu}} - \operatorname{Inf}(e))$, which are upper bounds on the times needed to fill up and to empty the reservoir, are sufficiently short, then the constancy intervals do not abut. Around a trough of p there is an interval $(\underline{t}, \overline{t})$ characterised by $\int_{\underline{t}}^{\overline{t}} e(t)\,\mathrm{d}t = k_{\mathrm{St}}$, on which $p(t) < \hat{\psi}_{\mathrm{H}}$ throughout. Around a local peak of p there is an interval $(\underline{t}, \overline{t})$ characterised by $\int_{\underline{t}}^{\overline{t}} (k_{\mathrm{Tu}} - e(t))\,\mathrm{d}t = k_{\mathrm{St}}$ on which

$p(t) > \hat{\psi}_{\mathrm{H}}$ throughout. The optimal output has the "bang-coast-bang" form

$$\hat{y}_{\mathrm{H}}(t; p, k_{\mathrm{St}}, k_{\mathrm{Tu}}, e) = \begin{cases} k_{\mathrm{Tu}} & \text{if } p(t) > \hat{\psi}_{\mathrm{H}}(t; p, k_{\mathrm{St}}, k_{\mathrm{Tu}}, e) \\ e(t) & \text{if } p(t) = \hat{\psi}_{\mathrm{H}}(t; p, k_{\mathrm{St}}, k_{\mathrm{Tu}}, e) \\ 0 & \text{if } p(t) < \hat{\psi}_{\mathrm{H}}(t; p, k_{\mathrm{St}}, k_{\mathrm{Tu}}, e) \end{cases}. \tag{5.14.49}$$

The lowercase notation, $\hat{y}_{\mathrm{H}}$ or $\hat{\psi}_{\mathrm{H}}$, is used *only when the solution is unique.* In general, the solution sets for (5.14.32)–(5.14.36) and (5.14.46)–(5.14.48) are denoted by $\hat{Y}_{\mathrm{H}}(p; k_{\mathrm{St}}, k_{\mathrm{Tu}}; e)$ and $\hat{\Psi}_{\mathrm{H}}(p; k_{\mathrm{St}}, k_{\mathrm{Tu}}; e)$.

The shadow-pricing programme (5.14.46)–(5.14.48) has a solution if

$$k_{\mathrm{St}} > 0 \quad \text{and} \quad k_{\mathrm{Tu}} > \mathrm{EssSup}(e) \geq \mathrm{EssInf}(e) > 0. \tag{5.14.50}$$

If additionally p is continuous, i.e., $p \in \mathcal{C}_{+}[0, T]$, then there is a *unique* solution

$$\hat{\psi}_{\mathrm{H}}(p; k_{\mathrm{St}}, k_{\mathrm{Tu}}; e) = \nabla_e \Pi^{\mathrm{H}}_{\mathrm{SR}}(p; k_{\mathrm{St}}, k_{\mathrm{Tu}}; e). \tag{5.14.51}$$

This is the TOU price of water (unit value of the river flow). It follows that the plant's operating profit $\Pi^{\mathrm{H}}_{\mathrm{SR}}$ is also differentiable in $(k_{\mathrm{St}}, k_{\mathrm{Tu}})$. In terms of $\hat{\psi}_{\mathrm{H}}$, the unit rental values of the reservoir and the turbine (in \$/kWh and \$/kW, respectively) are:

$$\hat{r}_{\mathrm{St}}(p; k_{\mathrm{St}}, k_{\mathrm{Tu}}; e) = \frac{\partial \Pi^{\mathrm{H}}_{\mathrm{SR}}}{\partial k_{\mathrm{St}}} = \mathrm{Var}^{+}_{\mathrm{c}}\left(\hat{\psi}_{\mathrm{H}}(p; k_{\mathrm{St}}, k_{\mathrm{Tu}}; e)\right) \tag{5.14.52}$$

$$\hat{r}_{\mathrm{Tu}}(p; k_{\mathrm{St}}, k_{\mathrm{Tu}}; e) = \frac{\partial \Pi^{\mathrm{H}}_{\mathrm{SR}}}{\partial k_{\mathrm{Tu}}} = \int_0^T \left(p(t) - \hat{\psi}_{\mathrm{H}}(t)\right)^{+} \mathrm{d}t. \tag{5.14.53}$$

For proofs, see Chapter 4.

As for the operation problem (5.14.32)–(5.14.36), it has a solution for any $p \in L^1_{+}[0, T]$ and every $(k_{\mathrm{St}}, k_{\mathrm{Tu}}) \geq 0$ and $e \leq k_{\mathrm{Tu}}$. If p has no plateau (i.e., $\mathrm{meas}\{t : p(t) = \mathsf{p}\} = 0$ for every $\mathsf{p} \in \mathbb{R}$), then there is a unique solution $\hat{y}_{\mathrm{H}}(p; k_{\mathrm{St}}, k_{\mathrm{Tu}}; e)$. It is given either by (5.14.49) itself (if (5.14.50) holds and $p \in \mathcal{C}$), or by (5.14.29) with any $\psi \in \hat{\Psi}_{\mathrm{H}}$ instead of $\hat{\psi}_{\mathrm{H}}$ (if (5.14.50) holds but $p \notin \mathcal{C}$). For proofs, see Chapter 4.

Comments (on assumptions on p and properties of water value ψ in the hydro problem):

- As in the case of thermal generation with pumped storage, time-continuity of the electricity tariff p, which guarantees uniqueness and continuity of the optimal water price ψ, can be verified for the general competitive equilibrium with hydro-thermal generation. The much less important condition that p have no plateau is, again, questionable: it cannot hold in an equilibrium with continuous *quantity* trajectories (since it leads to the unique optimum $\hat{y}_{\mathrm{H}}$, which is, under (5.14.50), a discontinuous function of t because it takes only the values k_{Tu}, $e(t)$, and 0, as per (5.14.49)).

- When $e \not\leq k_{\mathrm{Tu}}$ (i.e., when the policy of pure "coasting", $y = e$ with no spillage, is infeasible), the hydro operation and valuation LPs must be modified in the way indicated in Section 4.11. This complicates the solution, and an optimal water price ψ need *not* then be unique or continuous over time (despite the continuity of the electricity price p).

Comments (on choice of space for dual variables):

- For "automatic" proofs of the dual LPs' solubility, which are based on Slater's Condition, the dual-variable spaces must be the norm-duals of the corresponding primal perturbation spaces (L^∞ and $\mathcal{C}$). This means using $L^{\infty *}$, instead of L^1, as the space for each of the dual variables paired to those primal perturbations that range over L^∞ (viz., for κ and ν in (5.14.6)–(5.14.9), for κ_{Tu} and κ_{Pu} in (5.14.17)–(5.14.22), and for ψ, κ_{Tu} and ν_{Tu} in (5.14.37)–(5.14.43))—just as $\mathcal{M} = \mathcal{C}^*$ serves as the space for the dual variables paired to perturbations that range over $\mathcal{C}$ (viz., for κ_{St} and ν_{St}). This is because, like $\mathcal{C}_+$, the nonnegative cone L^∞_+ has a nonempty norm-interior, and so the positivity of capacities k, $(k_{\mathrm{St}}, k_{\mathrm{Co}})$ or $(k_{\mathrm{St}}, k_{\mathrm{Tu}})$, together with (5.14.50) for the hydro plant, imply that Slater's Condition, as generalised in [73, (8.12)] to infinite-dimensional inequality constraints, holds with the supremum norm topology on the primal parameter spaces L^∞ and $\mathcal{C}$. This ensures the existence of a dual optimum in the norm-dual spaces (i.e., of $\hat{\kappa}_{\mathrm{Th}}$ and $\hat{\nu}_{\mathrm{Th}}$ in $L^{\infty *}$, $\hat{\kappa}_{\mathrm{Tu}}$ and $\hat{\kappa}_{\mathrm{Pu}}$ in $L^{\infty *}$, $\hat{\kappa}_{\mathrm{St}}$ and $\hat{\nu}_{\mathrm{St}}$ in $\mathcal{M}$, and of $\hat{\psi}$, $\hat{\kappa}_{\mathrm{Tu}}$ and $\hat{\nu}_{\mathrm{Tu}}$ in $L^{\infty *}$). Density representation of these dual variables (other than $\hat{\kappa}_{\mathrm{St}}$ and $\hat{\nu}_{\mathrm{St}}$) comes from the problem's structure *and* the assumption that p is a density: since $p \in L^1$, every optimal κ and ν (for a thermal plant) is actually in L^1 by (5.14.11), as is every optimal κ_{Tu} and κ_{Pu} (for a storage plant), and every optimal κ_{Tu} and ν_{Tu} (for a hydro plant). And every feasible ψ is in BV $\subset L^1$ by (5.14.43). This is what justifies the use of L^1 (rather than $L^{\infty *}$) in the above formulations of the dual LPs (when $p \in L^1$).

- In the more general case of a $p \in L^{\infty *}$, the generating capacities' optimal rent flows, $\hat{\kappa}_{\mathrm{Th}}$ and $\hat{\kappa}_{\mathrm{Tu}}$, are in $L^{\infty *}$ (although the corresponding $\hat{\nu}_{\mathrm{Th}}$ and $\hat{\kappa}_{\mathrm{Pu}}$ or $\hat{\nu}_{\mathrm{Tu}}$ are in L^1 because $p \geq 0$). Also, when $p \in L^{\infty *}$, the degenerate case of zero storage capacity (with a positive conversion capacity) provides an example of a duality gap (Appendix A).

5.15 Peak-load pricing of electricity with pumped storage or hydro generation

The introductory application of the short-run approach to electricity pricing, in Section 5.2, is made simple by cross-price independence of short-run supply and the assumed cross-price independence of demand. In such a case, the short-run general equilibrium (SRGE) can be found separately for

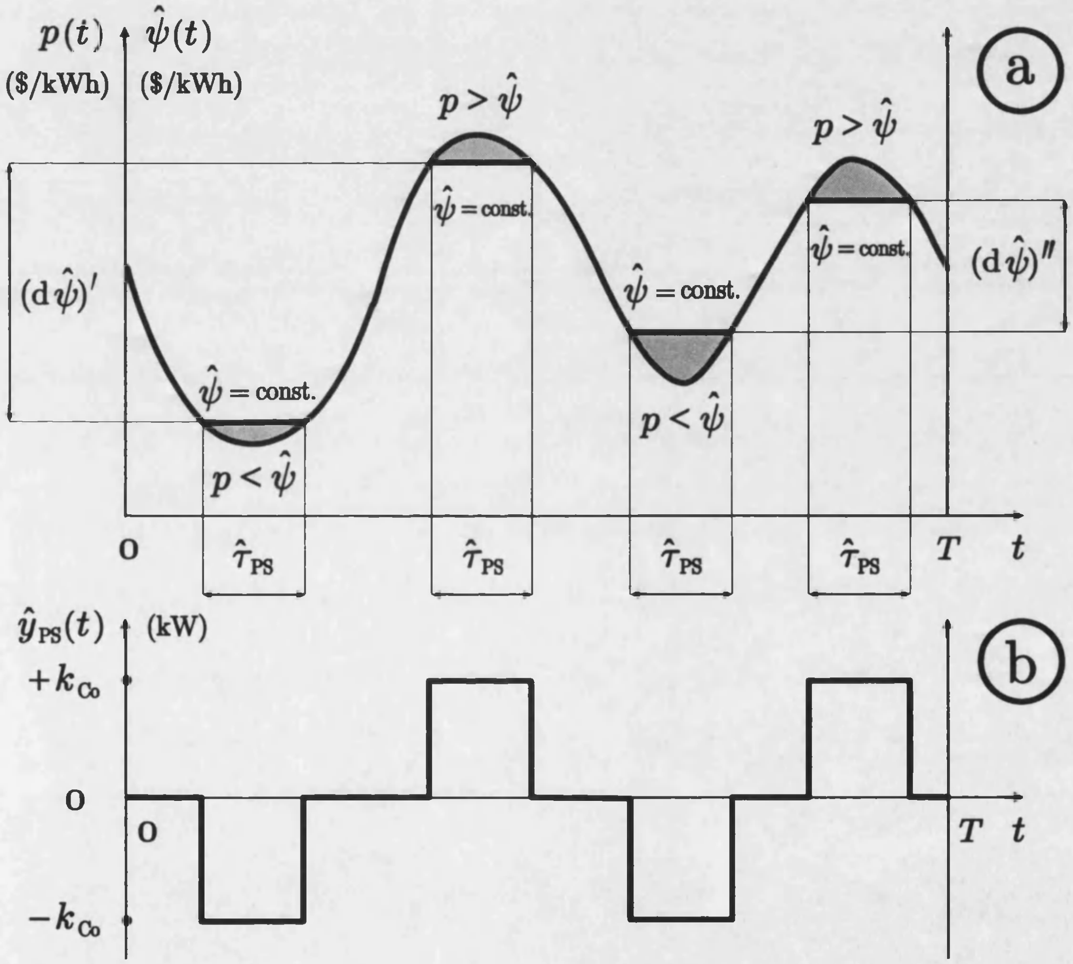

FIGURE 5.4. Trajectories of: (a) shadow price of stock $\hat{\psi}$, and (b) output of pumped-storage plant (optimum storage policy) $\hat{y}_{PS}$ in Section 5.14, and in Theorem 5.15.1. Unit rent for storage capacity is $\mathrm{Var}_c^+\left(\hat{\psi}\right) = \left(\mathrm{d}\hat{\psi}\right)' + \left(\mathrm{d}\hat{\psi}\right)''$, the sum of rises of $\hat{\psi}$. Unit rent for conversion capacity is $\int_0^T \left|p\left(t\right) - \hat{\psi}\left(t\right)\right| \mathrm{d}t$, the sum of grey areas. By definition, $\hat{\tau}_{PS} = k_{St}/k_{Co}$.

each time instant (by intersecting the demand and supply curves). It is equally simple to calculate the unit operating profit, and use it as an imputed capacity value to work out the long-run general equilibrium (LRGE).

That analysis is now extended to apply to cross-price dependent demand and to include storage or hydro plants, whose profit-maximising output is also cross-price dependent. Though the resulting general equilibrium problem cannot be solved by explicit formulae, the short-run approach does make it tractable: first, short-run supply can be determined by solving the plant operation LPs; then an iterative procedure (such as Walrasian tatonnement) can be used to find the short-run equilibrium; and finally plant valuations, obtained from dual LP solutions, can be used to find the long-run equilibrium by another iteration (as is indicated in Figure 5.3). A system of equilibrium conditions required for this approach is obtained by placing the operation and valuation results for the ESI's plants into the SRP programme-based LRGE system, (5.11.11)–(5.11.15) with (5.11.18)–(5.11.19). This is first done for an electricity supply technology that combines thermal generation with pumped storage.

Except for the storage capacity, all the ESI's inputs are taken to have *fixed* prices: $\left(r_1^{\mathrm{F}}, \ldots, r_\Theta^{\mathrm{F}}\right)$ for the thermal generating capacities, $(w_1, \ldots, w_\Theta)$ for the corresponding fuels, and $r_{\mathrm{Co}}^{\mathrm{F}}$ for the storage plant's converter. There is a location where an energy reservoir of capacity k_{St} can be constructed at a cost $G\left(k_{\mathrm{St}}\right)$. Usually, the marginal cost is increasing, i.e., the construction cost is a strictly convex and increasing function, G: $\left[0, \bar{k}_{\mathrm{St}}\right] \rightarrow \mathbb{R}_+$ with $G(0) = 0$. (This is especially so with the PWES and CAES techniques, which utilise special geological features.) In the terminology of Section 5.11, the reservoir is the single equilibrium-priced capital input; all the others have fixed prices. Formally, $\Phi_{\mathrm{PS}}^{\mathrm{E}} = \{\mathrm{St}\}$, $\Phi_{\mathrm{PS}}^{\mathrm{F}} = \{\mathrm{Co}\}$, and $\Phi_\theta^{\mathrm{F}} = \Phi_\theta = \{\theta\}$ for each $\theta \in \Theta$ (the set of thermal plant types).

All input demand for electricity is taken to come from a single Industrial User, who produces a final good from inputs of electricity and the numeraire, z and n. His production function, $(z, n) \mapsto F(z, n)$, is assumed to be strictly concave and increasing, and Mackey continuous, i.e., $\mathrm{m}\left(L^\infty \times \mathbb{R}, L^1 \times \mathbb{R}\right)$-continuous on $L_+^\infty[0,T] \times \mathbb{R}_+$. One example is the additively separable form for $F(\cdot, n)$, i.e., the integral functional $F(z,n) = \int_0^T f(t, z(t), n)\, \mathrm{d}t$, where f meets the conditions of [10, p. 535].[69]

A complete commodity bundle consists, then, of electricity (differentiated over time), the ESI's inputs (viz., the thermal capacities, the fuels, and the storage and conversion capacities), the produced final good and the numeraire. These quantities and their prices are always listed in this order, but those which are irrelevant in a particular context are omitted (as in Section 5.11). So a complete price system is $(p; (r_\theta); (w_\theta); r_{\mathrm{St}}, r_{\mathrm{Co}}; \varrho, 1)$, but a consumer price system is just $(p; \varrho, 1)$

[69]That is, the function $t \mapsto f(t, z, n)$ is integrable on $[0, T]$ for each $(z, n) \in \mathbb{R}_+^2$, and the function $(z, n) \mapsto f(t, z, n)$ is concave, increasing and continuous on $\mathbb{R}_+^2$, with $f(t, 0, 0) = 0$ for every $t \in [0, T]$. For a short proof of the Mackey continuity of F, see [42].

$\in L^1[0,T] \times \mathbb{R}^2$—since a consumption bundle consists of electricity, the produced final good and the numeraire, denoted by $(x; \varphi, m) \in L^\infty[0,T] \times \mathbb{R}^2$. The utility function, U_h for household h, is also assumed to be Mackey continuous, i.e., $\mathrm{m}\left(L^\infty \times \mathbb{R}^2, L^1 \times \mathbb{R}^2\right)$-continuous on the consumption set $L_+^\infty[0,T] \times \mathbb{R}_+^2$. Each household's initial endowment is a quantity of the numeraire $m_h^{\mathrm{En}} > 0$. The household's share in the User Industry's profit is $\varsigma_{h\,\mathrm{IU}}$, and its share of profit from supplying the storage capacity is $\varsigma_{h\,\mathrm{St}}$.

By feeding the programming results summarised in Section 5.14 into the framework of Section 5.11, long-run equilibrium is next characterised by optimality of the ESI's investments in addition to the SRGE system, which is either (5.15.4)–(5.15.9) for pumped storage or (5.15.14)–(5.15.19) for hydro-thermal generation. For simplicity, it is assumed that all the equilibrium capacities are positive, i.e., that each type of plant is built (in general, some plant types might not be built because of their costs).

Theorem 5.15.1 (Characterisation of long-run equilibrium with pumped storage). *Assume that the ESI's technology consists of thermal generation techniques (Θ) and a pumped storage technique. Then a price system made up of:*

- *a time-continuous electricity tariff $p^\star \in \mathcal{C}[0,T]$*
- *a rental price for storage capacity $r_{\mathrm{St}}^\star$*
- *a price $\varrho^\star > 0$ for the produced final good*
- *the given prices for fuels and the generating capacities (viz., r_θ^{F} for thermal capacity of type θ and w_θ for its fuel, and r_{Co} for the converter capacity)*

and an allocation made up of:

- *an output $y_\theta^\star \in L_+^\infty[0,T]$ from the thermal plant of type θ with*
  - *a capacity $k_\theta^\star > 0$*
  - *a fuel input $v_\theta^\star$ (for each θ)*
- *an output $y_{\mathrm{PS}}^\star \in L^\infty[0,T]$ from a pumped-storage plant with*
  - *a storage capacity $k_{\mathrm{St}}^\star > 0$*
  - *a conversion capacity $k_{\mathrm{Co}}^\star > 0$*
- *a consumption bundle $(x_h^\star, \varphi_h^\star, m_h^\star) \in L_+^\infty[0,T] \times \mathbb{R}_+ \times \mathbb{R}_+$ for each household h*
- *an input-output bundle of the User Industry $(-z^\star, F(z^\star, n^\star), -n^\star) \in L_-^\infty[0,T] \times \mathbb{R}_+ \times \mathbb{R}_-$*

form a long-run competitive equilibrium if and only if:

1. (a) *(Equality of ESI's capital-input prices to profit-imputed marginal values) For each* $\theta = 1, \ldots, \Theta$

$$r_\theta^{\mathrm{F}} = \int_0^T (p^\star(t) - w_\theta)^+ \, \mathrm{d}t \tag{5.15.1}$$

$$r_{\mathrm{St}}^\star = \mathrm{Var}_{\mathrm{c}}^+ (\psi^\star) \tag{5.15.2}$$

$$r_{\mathrm{Co}} = \int_0^T |p^\star(t) - \psi^\star(t)| \, \mathrm{d}t \tag{5.15.3}$$

where $\psi^\star := \hat{\psi}_{\mathrm{PS}}(p^\star, k_{\mathrm{St}}^\star, k_{\mathrm{Co}}^\star)$ *is the optimal price of energy stock, i.e., the unique solution to the programme (5.14.27)-(5.14.28) with* $(p^\star; k_{\mathrm{St}}^\star, k_{\mathrm{Co}}^\star)$ *as data.*[70]

(b) *(Operating profit maximisation by ESI) For each thermal plant type* θ *(whose heat rate is* $1/\eta_\theta$*)*

$$y_\theta^\star(t) \in \begin{cases} \{0\} & \text{if } p^\star(t) < w_\theta \\ [0, k_\theta^\star] & \text{if } p^\star(t) = w_\theta \\ \{k_\theta^\star\} & \text{if } p^\star(t) > w_\theta \end{cases} \quad \text{for a.e. } t \tag{5.15.4}$$

$$v_\theta^\star = \frac{1}{\eta_\theta} \int_0^T y_\theta^\star(t) \, \mathrm{d}t. \tag{5.15.5}$$

And, with $(p^\star; k_{\mathrm{St}}^\star, k_{\mathrm{Co}}^\star)$ *as the data,*

$$y_{\mathrm{PS}}^\star \text{ solves the linear programme (5.14.12) to (5.14.16)} \tag{5.15.6}$$

(which implies that the output from pumped storage is $y_{\mathrm{PS}}^\star(t) = k_{\mathrm{Co}}$ *when* $p^\star(t) > \psi^\star(t)$ *and* $y_{\mathrm{PS}}^\star(t) = -k_{\mathrm{Co}}$ *when* $p^\star(t) < \psi^\star(t)$*).*

2. *(Profit maximisation by User Industry)*[71]

$$(p^\star, 1) \in \varrho^\star \widehat{\partial} F(z^\star, n^\star). \tag{5.15.7}$$

3. *(Consumer utility maximisation) For each* h, $(x_h^\star, \varphi_h^\star, m_h^\star)$ *maximises* U_h *on the budget set*

$$\left\{ (x, \varphi, m) \geq 0 : \int_0^T p^\star(t) \, x(t) \, \mathrm{d}t + \varrho^\star \varphi + m \leq \hat{M}_h(p^\star, r_{\mathrm{St}}^\star, \varrho^\star) \right\}$$

[70] Since $p^\star \in C[0, T]$, the optimal ψ is indeed unique (Theorem 3.9.2, in Section 3.9).

[71] Since F is taken to be $-\infty$ outside of $L_+^\infty \times \mathbb{R}_+$, $\widehat{\partial} F$ contains a term arising from this nonnegativity constraint. To spell this out, assume that F, as a function on its effective domain $L_+^\infty \times \mathbb{R}_+$, has a Mackey continuous, concave and Gateaux differentiable extension F^{Ex} defined on all of $L^\infty \times \mathbb{R}$. Then (5.15.7) means that $(z^\star, n^\star) \geq 0$ and $(1/\varrho^\star) p^\star = \nabla_z F^{\mathrm{Ex}}(z^\star, n^\star) + \mu$ and $1/\varrho^\star = (\partial F^{\mathrm{Ex}}/\partial n)(z^\star, n^\star) + \nu$ for some $\mu \in L_+^1$ vanishing a.e. on the set $\{t : z^\star(t) > 0\}$, with $\nu = 0$ if $n^\star > 0$. (If $p^\star$ were in $L^{\infty\star}$ but not in L^1 then μ would be an element of $L_+^{\infty\star}$ concentrated on $\{t : z^\star(t) \leq \epsilon\}$ for each $\epsilon > 0$.)

where

$$\hat{M}_h(p, r_{\mathrm{St}}, \varrho) = m_h^{\mathrm{En}} + \varsigma_{h\,\mathrm{St}} \sup_{k_{\mathrm{St}}} (r_{\mathrm{St}} k_{\mathrm{St}} - G(k_{\mathrm{St}}))$$

$$+ \varsigma_{h\,\mathrm{IU}} \sup_{z,n} \left(\varrho F(z,n) - \int_0^T p(t)\, z(t)\, \mathrm{d}t - n \right). \quad (5.15.8)$$

4. (Market clearance)

$$y_{\mathrm{PS}}^\star + \sum_\theta y_\theta^\star = z^\star + \sum_h x_h^\star \quad \text{and} \quad F(z^\star, n^\star) = \sum_h \varphi_h^\star. \quad (5.15.9)$$

5. (MC pricing of storage capacity)

$$r_{\mathrm{St}}^\star \in \partial G(k_{\mathrm{St}}^\star). \quad (5.15.10)$$

Proof. Given the results of Section 5.14, this is a formality—except for verifying the absence of a duality gap. Note first that Conditions 2 to 5 of the theorem are simply specialisations, to the ESI case, of the corresponding parts of the definition of a long-run equilibrium (Section 5.11). What has to be shown is the equivalence of the theorem's Condition 1 (optimal operation and valuation of the ESI's plants) to the definition's Condition 1 (LRP maximisation). As a general principle, this has been established in Section 5.4 and restated in Section 5.6 (by taking account of Section 5.5). Its substance is that, in the long run, competitive profit maximisation is equivalent—as a system of conditions on both quantities and prices—to the conjunction of: (i) maximisation of the operating profit (short-run profit), which includes minimisation of the operating cost, (ii) minimisation of the fixed-input value by shadow pricing (which is identified as the dual programme), and (iii) equality of the maximum SRP to the minimum FIV (absence of a duality gap). For each of the ESI's plants, the SRP and FIV programmes are spelt out in Section 5.14, and it remains only to show that their values are equal. (In formal terms, (5.11.4)–(5.11.5) is (5.3.5) at equilibrium prices, which, as is noted before the Comment in Section 5.6, is equivalent to the conjunction of (5.4.2)–(5.4.3), (5.6.2) and (5.6.3). And, for the ESI's technology, (5.4.2)–(5.4.3) and (5.6.2) can be put as (5.15.4)–(5.15.6) and (5.15.1)–(5.15.3). It remains only to prove (5.6.3) for each of the ESI's plants.)

To this end, note first that the thermal operation LP (5.14.1)–(5.14.3) and its dual (5.14.6)–(5.14.9) always have the same value: with θ in place of Th, the common value of both LPs is $k_\theta \int_0^T (p(t) - w_\theta)^+ \,\mathrm{d}t$ for every (p, k_θ, w_θ), by (5.14.4) and by (5.14.5) or (5.14.10). For pumped storage, however, the equality of values of the operation LP (5.14.12)–(5.14.16) and its dual—in the form of either the standard dual LP (5.14.17)–(5.14.22) or the equivalent CP (5.14.27)–(5.14.28)—relies on the properties of its data in the general equilibrium, $(p^\star; k_{\mathrm{St}}^\star, k_{\mathrm{Co}}^\star)$. It can be proved in two ways because it follows from *either* of two assumptions: that $(k_{\mathrm{St}}^\star, k_{\mathrm{Co}}^\star) \gg 0$ and that $p^\star \in L^1[0,T]$. Strict positivity of the fixed-input bundle $(k_{\mathrm{St}}, k_{\mathrm{Co}})$ is a case of the generalised Slater's Condition for infinite-dimensional inequality constraints, formulated in [73, (8.12)]. *A fortiori*, it is a case of Slater's

Condition for generalised perturbed CPs, formulated in [73, Theorem 18 (a)]. So it guarantees the continuity of $\Pi_{\mathrm{SR}}^{\mathrm{PS}}(p,\cdot)$ on a neighbourhood of $(k_{\mathrm{St}}, k_{\mathrm{Co}})$, for every $p \in L^{\infty *}$. The other, alternative proof derives upper semicontinuity of $\Pi_{\mathrm{SR}}^{\mathrm{PS}}(p,\cdot)$ from the assumption that $p \in L^1$. This is a case of a price system in the predual of the commodity space: here, L^1 is the Banach predual of $L^\infty[0,T]$. The maximand $\langle p \mid \cdot \rangle$ is therefore continuous for the weak* topology $\mathrm{w}\left(L^\infty, L^1\right)$, and one can show that the maximum value, $\Pi_{\mathrm{SR}}^{\mathrm{PS}}(p,\cdot)$, is u.s.c. by exploiting the weak*-compactness of the short-run production set

$$\{y \in L^\infty[0,T] : (y; -k_{\mathrm{St}}, -k_{\mathrm{Co}}) \in \mathbb{Y}_{\mathrm{PS}}\} \subseteq \{y \in L^\infty : |y| \leq k_{\mathrm{Co}}\}$$

where $\mathbb{Y}_{\mathrm{PS}}$ is given by (5.13.5); formally, Berge's Maximum Theorem applies [8, VI.3: Theorem 2]. (A stronger result can be obtained by applying the dual-value continuity criterion of [73, Theorem 18' (e)]: this shows that the convex function $\overline{\Pi}_{\mathrm{SR}}^{\mathrm{PS}}(\cdot, k_{\mathrm{St}}, k_{\mathrm{Co}})$ is norm-continuous on L^1, which implies that the concave function $\Pi_{\mathrm{SR}}^{\mathrm{PS}}(p,\cdot,\cdot)$ is upper semicontinuous at $(k_{\mathrm{St}}, k_{\mathrm{Co}})$ for each $p \in L^1$.) Finally, the equality $\overline{\Pi}_{\mathrm{SR}}^{\mathrm{PS}} = \Pi_{\mathrm{SR}}^{\mathrm{PS}}$ at $(p; k_{\mathrm{St}}, k_{\mathrm{Co}})$ follows from upper semicontinuity, and *a fortiori* from continuity, of $\Pi_{\mathrm{SR}}^{\mathrm{PS}}(p,\cdot,\cdot)$ at $(k_{\mathrm{St}}, k_{\mathrm{Co}})$. Since $p^\star \in L^1$ and $(k_{\mathrm{St}}^\star, k_{\mathrm{Co}}^\star) \gg 0$, either method applies to this data point. □

A similar result is next presented for hydroelectric generation (H) instead of pumped storage (PS). The thermal technology remains the same, and its inputs have fixed prices, $\left(r_1^{\mathrm{F}}, \ldots, r_\Theta^{\mathrm{F}}\right)$ and $(w_1, \ldots, w_\Theta)$. The hydro turbine also has a fixed price, $r_{\mathrm{Tu}}^{\mathrm{F}}$. There is a river with a single location where a dam can be constructed to create a water reservoir of a capacity k_{St}, at a cost $G(k_{\mathrm{St}})$. The river has a fixed, periodic flow, $e(t)$ at time $t \in [0,T]$, which (on the assumption of a constant head) means a given energy inflow.[72] Its price, $\psi(t)$ at time t, is determined in long-run equilibrium. The river's total rent is $\int_0^T \psi e \, \mathrm{d}t$, and household h's share of the rent is $\varsigma_{h\,\mathrm{Ri}}$. Its share of profit from supplying the storage capacity is $\varsigma_{h\,\mathrm{St}}$. As before, there is a single Industrial User of electricity (whose production function is F), and the household's share in his profit is $\varsigma_{h\,\mathrm{IU}}$.

Theorem 5.15.2 (Characterisation of long-run equilibrium with hydro-thermal generation). *Assume that the ESI's technology consists of thermal generation techniques (Θ) and a hydroelectric technique. Then a price system made up of:*

- *a time-continuous electricity tariff $p^\star \in C[0,T]$*
- *a rental price for the hydro reservoir capacity $r_{\mathrm{St}}^\star$*
- *a price $\varrho^\star$ for the produced final good*

[72] More generally, it might be possible to improve the watershed to obtain a river flow e at a cost $G_{\mathrm{Ri}}(e)$, a convex function of e. The case of a fixed, unimprovable river flow $\overline{e}$ can be obtained by setting $G_{\mathrm{Ri}}(e)$ equal to 0 for $e = \overline{e}$ and $+\infty$ otherwise.

- *the given prices for fuels and the generating capacities (viz.,* r_θ^{F} *for thermal capacity of type* θ *and* w_θ *for its fuel, and* $r_{\mathrm{Tu}}^{\mathrm{F}}$ *for the turbine capacity)*

and an allocation made up of:

- *an output* $y_\theta^\star \in L_+^\infty[0,T]$ *from the thermal plant of type* θ *with*
  - *a capacity* $k_\theta^\star > 0$
  - *a fuel input* $v_\theta^\star$ *(for each* θ*)*
- *an output* $y_{\mathrm{H}}^\star \in L^\infty[0,T]$ *from a hydro plant with*
  - *reservoir and turbine capacities* $k_{\mathrm{St}}^\star > 0$ *and* $k_{\mathrm{Tu}}^\star > 0$
  - *the given river flow* $e \in L_+^\infty[0,T]$*, which is assumed to meet Condition (5.14.50)*[73]
- *a consumption bundle* $(x_h^\star, \varphi_h^\star, m_h^\star) \in L_+^\infty[0,T] \times \mathbb{R}_+ \times \mathbb{R}_+$ *for each household* h
- *an input-output bundle of the User Industry* $(-z^\star, F(z^\star, n^\star), -n^\star) \in L_-^\infty[0,T] \times \mathbb{R}_+ \times \mathbb{R}_-$

form a long-run competitive equilibrium if and only if:

1. (a) *(Equality of ESI's capital-input prices to their profit-imputed marginal values) For each* $\theta = 1, \ldots, \Theta$

$$r_\theta^{\mathrm{F}} = \int_0^T (p^\star(t) - w_\theta)^+ \, \mathrm{d}t \tag{5.15.11}$$

$$r_{\mathrm{St}}^\star = \mathrm{Var}_c^+(\psi^\star) \tag{5.15.12}$$

$$r_{\mathrm{Tu}}^{\mathrm{F}} = \int_0^T (p^\star(t) - \psi^\star(t))^+ \, \mathrm{d}t \tag{5.15.13}$$

(b) *(Operating profit maximisation by ESI) For each thermal plant type* θ *(whose heat rate isɔn to* $1/\eta_\theta$*)*

(b) *(Operating profit maximisation by ESI) For each thermal plant type* θ *(whose heat rate is* $1/\eta_\theta$*)*

$$y_\theta^\star(t) \in \begin{cases} \{0\} & \text{if } p^\star(t) < w_\theta \\ [0, k_\theta^\star] & \text{if } p^\star(t) = w_\theta \\ \{k_\theta^\star\} & \text{if } p^\star(t) > w_\theta \end{cases} \quad \text{for a.e. } t \tag{5.15.14}$$

$$v_\theta^\star = \frac{1}{\eta_\theta} \int_0^T y_\theta^\star(t) \, \mathrm{d}t \tag{5.15.15}$$

[73] The assumption can be dropped, but this complicates the problem and, as a result, an optimal water price function need not be unique or continuous: see Chapter 4.

and, with $(p^\star; k^\star_{\mathrm{St}}, k^\star_{\mathrm{Tu}}; e)$ *as the data,*

$$y^\star_{\mathrm{H}} \text{ solves the linear programme (5.14.32) to (5.14.36)} \tag{5.15.16}$$

(which implies that the hydro output is $y^\star_{\mathrm{H}}(t) = k_{\mathrm{Tu}}$ *when* $p^\star(t) > \psi^\star(t)$ *and* $y^\star_{\mathrm{H}}(t) = 0$ *when* $p^\star(t) < \psi^\star(t)$*).*

2. *(Profit maximisation by User Industry)*

$$(p^\star, 1) \in \varrho^\star \widehat{\partial} F(z^\star, n^\star). \tag{5.15.17}$$

3. *(Consumer utility maximisation) For each* h, $(x^\star_h, \varphi^\star_h, m^\star_h)$ *maximises* U_h *on the budget set*

$$\left\{(x, \varphi, m) \geq 0 : \int_0^T p^\star(t)\, x(t)\, \mathrm{d}t + \varrho^\star \varphi + m \leq \hat{M}_h(p^\star, r^\star_{\mathrm{St}}, \psi^\star, \varrho^\star)\right\}$$

where

$$\hat{M}_h(p, r_{\mathrm{St}}, \varrho, \psi) = m_h^{\mathrm{En}} + \varsigma_{h\,\mathrm{St}}\left(\sup_{k_{\mathrm{St}}}(r_{\mathrm{St}} k_{\mathrm{St}} - G(k_{\mathrm{St}}))\right)$$
$$+ \varsigma_{h\,\mathrm{IU}}\left(\sup_{z,n}\left(\varrho F(z,n) - \int_0^T p(t)\, z(t)\, \mathrm{d}t - n\right)\right) + \varsigma_{h\,\mathrm{Ri}} \int_0^T \psi(t)\, e(t)\, \mathrm{d}t. \tag{5.15.18}$$

4. *(Market clearance)*

$$y^\star_{\mathrm{H}} + \sum_\theta y^\star_\theta = z^\star + \sum_h x^\star_h \quad \text{and} \quad F(z^\star, n^\star) = \sum_h \varphi^\star_h. \tag{5.15.19}$$

5. *(MC pricing of reservoir capacity)*

$$r^\star_{\mathrm{St}} \in \partial G(k^\star_{\mathrm{St}}). \tag{5.15.20}$$

Proof. This is proved like Theorem 5.15.1 (taking into account the last Comment in Section 5.11). □

Remark 5.15.3 (Value of site for reservoir). *The rental value of the hydro or storage site is* $r^\star_{\mathrm{St}} k^\star_{\mathrm{St}} - G(k^\star_{\mathrm{St}})$ *per cycle (the reservoir's value less its construction cost).*

Comments (multiple sites): A similar analysis applies when there is a number of storage sites (or hydro sites) with different development costs, G_l for location l. Reservoir capacity is then a good differentiated by its location, and so is the river flow in the case of hydro. Therefore, some of the long-run equilibrium prices and quantities may depend on l:

- Consider first the case of pumped storage. Since $(\partial \Pi^{\mathrm{PS}}_{\mathrm{SR}} / \partial k_{\mathrm{Co}}) \left(k^\star_{\mathrm{St},l}, k^\star_{\mathrm{Co},l}\right)$ equals $r^{\mathrm{F}}_{\mathrm{Co}}$, which is independent of l, and since the derivative is homogeneous of degree 0 in $(k_{\mathrm{St}}, k_{\mathrm{Co}})$, the equilibrium capacity ratio $k^\star_{\mathrm{St},l} : k^\star_{\mathrm{Co},l}$ is independent of l. Therefore, the equilibrium price of storage capacity $r^\star_{\mathrm{St}}$ is also the same for each l (since it equals $\partial \Pi^{\mathrm{PS}}_{\mathrm{SR}}/\partial k_{\mathrm{St}}$, which is also homogeneous of degree 0). And this is so because the production technique has just one input whose supply cost depends on the location. The plant's size, however, does depend on l, since $k^\star_{\mathrm{St},l}$ meets the condition $r^\star_{\mathrm{St}} \in \partial G_l \left(k^\star_{\mathrm{St},l}\right)$. The site's rent, $r^\star_{\mathrm{St}} k^\star_{\mathrm{St},l} - G_l \left(k^\star_{\mathrm{St},l}\right)$, also depends on l.

- In hydro generation, both the reservoir construction cost function G_l and the fixed river flow e_l depend on the location l. So, in hydro, the equilibrium capacity ratio $k^\star_{\mathrm{St},l}/k^\star_{\mathrm{Tu},l}$, the price of reservoir capacity $r^\star_{\mathrm{St},l}$ and the shadow price of water $\psi^\star_l$ do all depend on l. (So do the reservoir's size $k^\star_{\mathrm{St},l}$ and the site's rent $r^\star_{\mathrm{St}} k^\star_{\mathrm{St},l} - G_l \left(k^\star_{\mathrm{St},l}\right)$.)

Comment (optimum of thermal output in terms of SRMC): Competitive profit maximization by the thermal plants can be reformulated as SRMC pricing by the thermal generating system, i.e., by using the system's instantaneous SRMC curve. With a finite number of plant types, Θ, this curve is actually a "right-angled" broken line:[74] under (5.13.2), it consists of (i) the Θ "horizontal" segments

$$[k_1 + \ldots + k_{\theta-1}, k_1 + \ldots + k_\theta] \times \{w_\theta\} \quad \text{for } \theta = 1, \ldots, \Theta$$

(with $k_0 := 0$) and (ii) the $\Theta + 1$ "vertical" segments

$$\{k_1 + \ldots + k_\theta\} \times [w_\theta, w_{\theta+1}] \quad \text{for } \theta = 0, 1, \ldots, \Theta$$

(with $w_{\Theta+1} := +\infty$, and with $w_0 := -\infty$ unless free disposal is included). Formally, Condition (5.15.4) or (5.15.14) for each θ is equivalent to:

$$p^\star(t) \in \partial_{\mathrm{y}} c_{\mathrm{SR}} \left(\sum_\theta y^\star_\theta(t); k^\star_1, \ldots, k^\star_\Theta; w_1, \ldots, w_\Theta \right) \quad \text{for a.e. } t$$

where c_{SR} is the *thermal system's instantaneous short-run cost* per unit time. With 1_A denoting the 0-1 indicator of a set A (1 on A and 0 outside), the system's instantaneous SRC of generating at a rate y can be given as

$$\begin{aligned} c_{\mathrm{SR}}(\mathrm{y}; (k_\theta), (w_\theta)) &:= \int_0^{\mathrm{y}} \sum_{\theta=1}^{\Theta} w_\theta 1_{[k_1+\ldots+k_{\theta-1}, k_1+\ldots+k_\theta]}(\mathrm{q})\, \mathrm{dq} \qquad (5.15.21) \\ &= w_1 \mathrm{y} + \sum_{\theta=1}^{\Theta-1} (w_{\theta+1} - w_\theta)(\mathrm{y} - (k_1 + \ldots + k_\theta))^+ \end{aligned}$$

if $0 \le \mathrm{y} \le \sum_{\theta=1}^{\Theta} k_\theta$ (otherwise $c_{\mathrm{SR}} = +\infty$). This is an increasing and convex (and piecewise linear) function of the output rate $\mathrm{y} \in \left[0, \sum_{\theta=1}^{\Theta} k_\theta\right]$, with $c_{\mathrm{SR}}(0) = 0$. The SRMC curve is the graph of

[74] In a model with a "continuum" of plant types, the SRMC curve is a general "complete nondecreasing curve", in the terminology of [70, 24.3]. But even the continuum model does *not* make the SRC curve differentiable: it still has a kink at the peak output, and typically it has offpeak kinks, too.

the subdifferential correspondence $\mathrm{y} \mapsto \partial c_{\mathrm{SR}}(\mathrm{y})$, in the instantaneous quantity-price plane. When $k_\theta > 0$ for each θ,

$$\partial_{\mathrm{y}} c_{\mathrm{SR}}(\mathrm{y}, (k_\theta), (w_\theta)) = \begin{cases} (-\infty, w_1] & \text{if } \mathrm{y} = 0 \\ \{w_\theta\} & \text{if } \mathrm{y} \in (k_1 + \ldots + k_{\theta-1}, k_1 + \ldots + k_\theta) \\ [w_\theta, w_{\theta+1}] & \text{if } \mathrm{y} = k_1 + \ldots + k_\theta \text{ and } 1 \le \theta \le \Theta - 1 \\ [w_\Theta, +\infty) & \text{if } \mathrm{y} = \sum_{\theta=1}^{\Theta} k_\theta \\ \emptyset & \text{if } \mathrm{y} > \sum_{\theta=1}^{\Theta} k_\theta \text{ or } \mathrm{y} < 0 \end{cases} \cdot \tag{5.15.22}$$

(For the case of $\Theta = 1$, the SRMC and SRC curves have been used in Section 5.2 and are shown in Figures 5.1a and c; the supply and subdifferential correspondences, $\mathrm{p} \mapsto S(\mathrm{p})$ and $\mathrm{y} \mapsto \partial c_{\mathrm{SR}}(\mathrm{y})$, are inverse to each other.)

5.16 Derivation of the dual programmes (proofs for Section 5.5)

The dual programmes are next derived formally by using the framework of [73].

Proposition 5.16.1 (Dual to SRP programme). *The dual of the short-run profit maximisation programme (5.3.6)-(5.3.7), with k as the primal parameter ranging over the space K paired with R as the range for the dual variable r, is the fixed-input shadow-pricing programme (5.5.6), or equivalently (5.5.13)-(5.5.14) when $\mathbb{Y}$ is a cone. The dual parameter is (p, w).*

Proof. Given (p, k, w), the parametric primal constrained maximand is $\langle p \mid y \rangle - \langle w \mid v \rangle$ minus $\delta(y, -k, -v \mid \mathbb{Y})$, where y and v are the primal decision variables, and k is the primal parameter (paired with the dual decision variable r). Let d' and d'' denote the dual perturbations (paired with y and $-v$). By [73, (4.17)] with the primal problem reoriented to maximisation, the (perturbed) dual constrained minimand—a function of r and (d', d'') as well as (p, k, w)—is

$$\begin{aligned} &\sup_{y,v;\Delta k} \{\langle d', d'' \mid y, -v \rangle - \langle r \mid \Delta k \rangle + \langle p \mid y \rangle - \langle w \mid v \rangle - \delta(y, -k - \Delta k, -v \mid \mathbb{Y})\} \\ &\quad = \langle r \mid k \rangle + \sup_{y,v,\Delta k} \{\langle p + d', r, w + d'' \mid y, -k - \Delta k, -v \rangle : (y, -k - \Delta k, -v) \in \mathbb{Y}\} \\ &\quad = \langle r \mid k \rangle + \sup_{y,v,k} \{\langle p + d', r, w + d'' \mid y, -k, -v \rangle : (y, -k, -v) \in \mathbb{Y}\} \\ &\quad = \langle r \mid k \rangle + \Pi_{\mathrm{LR}}(p + d', r, w + d''). \end{aligned}$$

So, by setting $d' = 0$ and $d'' = 0$, the dual programme is (5.5.6); and when $\mathbb{Y}$ is a cone, the dual is to minimise $\langle r \mid k \rangle + \delta(p, r, w \mid \mathbb{Y}^\circ)$ over r (since $\Pi_{\mathrm{LR}} = \delta^\#(\cdot \mid \mathbb{Y}) = \delta(\cdot \mid \mathbb{Y}^\circ)$, i.e., the support function of a cone is the indicator function of the polar cone). Finally, d' and d'' perturb the dual like increments to p and w, which therefore are the dual parameters (and so d' and d'' may be renamed to Δp and Δw). □

The other duals are derived in the same way; the dual of the SRC programme is spelt out.

Proposition 5.16.2 (Dual to SRC programme). *The dual of the short-run cost minimisation programme (5.3.10)–(5.3.11), with (y, k) as the primal parameter ranging over the space $Y \times K$ paired with $P \times R$ as the range for the dual variable (p, r), is the output-and-fixed-input pricing programme (5.5.4), or equivalently (5.5.9)–(5.5.10) when $\mathbb{Y}$ is a cone. The dual parameter is w.*

5.17 Conclusions from Chapter 5

The long-run general equilibrium can be determined most efficiently through the short-run equilibrium, which itself is of central practical interest. This method uses either the producer's plant operation and valuation programmes, which form a primal-dual pair, or an optimal-value function. The choice depends on the available description of the technology but, in engineering models with multiple outputs, this is usually a production set (which favours the use of programming). The primal programme in question can be either short-run profit maximisation or short-run cost minimisation, but the profit approach is much easier. This brings to the fore the equilibrium pricing of capital goods and natural resources. Such inputs divide into those which are fixed, or nearly fixed, even in the long run (e.g., river flows for hydroelectric generation) and those which are variable in the long run but are supplied at an increasing marginal cost (like water reservoirs). Correct valuation of such inputs is essential for efficient investment decisions and operating policies, as well as to other matters (compensation payments for, e.g., land or rivers). Their values, as the key to the transition from the short-run to the long-run solution, are fundamental to the approach. Thus the use of long-run general-equilibrium analysis puts valuation on a sound basis, and the short-run programmes provide a workable method for calculating these values.

Appendix A

EXAMPLE OF DUALITY GAP BETWEEN SHORT-RUN PROFIT MAXIMISATION AND FIXED-INPUT VALUATION

Equality of the primal and dual optimal values is equivalent to semicontinuity of either value function w.r.t. its "own" parameters, i.e., Type One semicontinuity. Therefore, any sufficient condition for continuity of the one value rules out a duality gap and implies that the other value is semicontinuous. It also implies that the other programme is soluble. In this Appendix, "continuity" means Type One continuity (unless specified as Type Two).

Any result for the primal value can be transcribed for the dual value by swapping the two programmes. Below, only those sufficient conditions for continuity are considered which are put entirely and directly in terms of the primal programme. Such a criterion can be classified by the particular value whose continuity it guarantees, i.e., it is either a primal-value or a dual-value continuity criterion. In other words, it gives, in terms of the one programme, a condition that guarantees value continuity for either the same or the other programme of the pair.

There is a salient criterion in each class. A criterion of *primal*-value continuity (w.r.t. primal parameters) is Slater's Condition on the primal programme, together with its generalised forms: see [73, (8.12) and Theorem 18 (a)]. A useful criterion of *dual*-value continuity (w.r.t. dual parameters) can be based on compactness and continuity conditions on the primal constraints and the optimand: see [73, Theorem 18' (e)]. Its semicontinuity implication for the primal value, w.r.t. primal parameters, can be viewed as a version of a part of Berge's Maximum Theorem [8, VI.3: Theorem 2]; the basic semicontinuity result of [73, Example 4' after (5.13)] is simply a special case of Berge's. The semicontinuity results in [46] are closely related, being applications of Berge's Theorem.

In the context of profit or cost as the primal value function, Slater's Condition takes the form spelt out in [46]; in the case of short-run profit maximisation with conditionally fixed coefficients, it boils down to strict positivity of the fixed-input bundle k. This guarantees the continuity of $\Pi_{\mathrm{SR}}(p, \cdot, w)$ on a neighbourhood of k. The alternative upper semicontinuity result for $\Pi_{\mathrm{SR}}(p, \cdot, w)$ on K requires a price system from the predual of the commodity space, i.e., a $p \in Y'$: see [46].

Either condition (positive capacities or predual output price) rules out a duality gap between profit-maximising operation and plant valuation. Between them, the two sufficient conditions cover a lot of ground: although the alternation "$p \in Y'$ or $k \gg 0$" is not actually necessary for $\overline{\Pi}_{\mathrm{SR}}$ to equal Π_{SR} at (p, k, w), it comes close to being so with technologies such as pumped storage and hydroelectric generation. In the case of storage, if the reservoir capacity k_{St} is zero *and* the price system $p \in L^{\infty*}[0,T]$ has a singular a.k.a. purely finitely additive part $p_{\mathrm{FA}} \neq 0$ (in addition to a density a.k.a. countably additive part $p_{\mathrm{CA}} \in L^1$), then the operating profit is obviously zero, but the unit value of conversion capacity is positive. This example, spelt out next, shows also that the failure of Slater's Condition can lead to nonexistence of an exact dual solution. A similar example of a duality gap can also be given for the hydro technology.[1]

Example A.0.1 (Duality gap between operation and valuation of an incomplete plant). *Take the pumped-storage technology (5.13.5) and an output price system $p \in L^{\infty*}[0,T]$ with $p_{\mathrm{FA}} \neq 0$ and $p_{\mathrm{CA}} \in \mathrm{BV} \subset L^1$ (i.e., with a nonzero singular part and a density part of bounded variation). If additionally $k_{\mathrm{Co}} > 0$ but $k_{\mathrm{St}} = 0$ (i.e., the plant has a conversion capacity but no storage capacity), then the operating profit is zero, i.e., $\Pi_{\mathrm{SR}}^{\mathrm{PS}}(p; 0, k_{\mathrm{Co}}) = 0$. But the optimal stock price (the dual solution) is $\hat{\psi} = p_{\mathrm{CA}}$, and so the capacity value (the dual optimal value) is*

$$\overline{\Pi}_{\mathrm{SR}}^{\mathrm{PS}}(p; 0, k_{\mathrm{Co}}) = k_{\mathrm{Co}} \|p_{\mathrm{FA}}\|_{\infty}^{*} > 0 = \Pi_{\mathrm{SR}}^{\mathrm{PS}}(p; 0, k_{\mathrm{Co}}). \tag{A.0.1}$$

If $p_{\mathrm{CA}} \in L^1 \setminus \mathrm{BV}$ (and still $k_{\mathrm{Co}} > 0$ but $k_{\mathrm{St}} = 0$), then the dual (stock-pricing) programme for ψ has no (exact) solution, but any sequence of ψ's in BV *that converges to p_{CA} in the L^1-norm is a sequence of approximate dual optima. The infimal capacity value is still $k_{\mathrm{Co}} \|p_{\mathrm{FA}}\|_{\infty}^{*}$ (i.e., there is the same duality gap).*

Comments (on Example A.0.1):

- It gives an example of a duality gap in *infinite* linear programming, since the SRP programme can be formulated as an LP: see (5.14.12)–(5.14.16).
- The example shows in a simple way why a duality gap must open at a point of the optimal value's discontinuity (of Type One). With the other parameters ($p \in L^{\infty*}$ and $k_{\mathrm{Co}} > 0$) kept fixed, $\overline{\Pi}_{\mathrm{SR}}$ and Π_{SR} are equal and vary continuously with k_{St} as long as it stays positive: every finite concave function on $\mathbb{R}_{++} := \mathbb{R}_+ \setminus \{0\}$ is continuous, and $\overline{\Pi}_{\mathrm{SR}} = \Pi_{\mathrm{SR}}$ when $k_{\mathrm{St}} > 0$ because this is Slater's Condition. But at $k_{\mathrm{St}} = 0$, Π_{SR} can fail to be right-continuous and then, being concave, it also fails to be u.s.c.—which means that it drops at $k_{\mathrm{St}} = 0$.[2] By contrast, Type Two semicontinuity holds automatically, i.e., $\overline{\Pi}_{\mathrm{SR}}$ is always u.s.c. and hence it is actually right-continuous at $k_{\mathrm{St}} = 0$. So the discontinuity of Π_{SR} at $k_{\mathrm{St}} = 0$ implies that $\Pi_{\mathrm{SR}}(0) < \overline{\Pi}_{\mathrm{SR}}(0)$. See Figure A.1.

[1] In the case of hydro with $p \geq 0$, $p_{\mathrm{FA}} \neq 0$ and $k_{\mathrm{St}} = 0$, if $k_{\mathrm{Tu}} > \mathrm{Sup}(e)$ then $\langle p \mid e \rangle < \langle p_{\mathrm{CA}} \mid e \rangle + k_{\mathrm{Tu}} \|p_{\mathrm{FA}}\|$; i.e., the optimal output is obviously equal to the inflow e, which yields a revenue lower than the value of hydro inputs (turbine and inflow).

[2] A finite concave function on a polyhedral set $Z \subseteq \mathbb{R}^n$ is l.s.c. on Z (so if it is u.s.c. on Z then it is continuous on Z): see [70, 10.2 and 20.5]. This applies to $Z = \mathbb{R}^n_+$ for every n (here, $n = 1$).

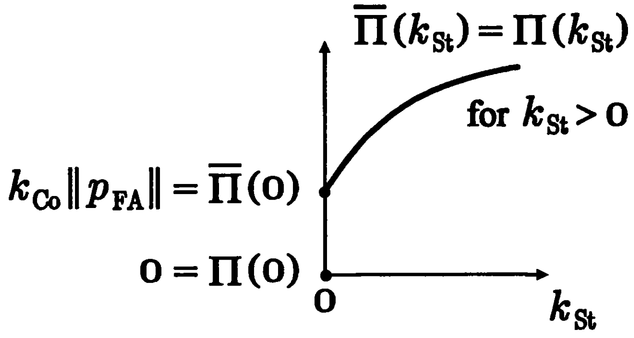

FIGURE A.1. Capacity value and operating profit for the pumped-storage technique, $\overline{\Pi}_{SR}$ and Π_{SR}, as functions of storage capacity k_{St} (for a fixed conversion capacity $k_{Co} > 0$ and a fixed good's price $p \in L^{\infty*} \setminus L^1$). When $k_{St} > 0$, Slater's Condition is met and so $\overline{\Pi} = \Pi$, but a duality gap opens at $k_{St} = 0$, where $\overline{\Pi}$ is continuous but Π drops (Example A.0.1).

- Recall from Section 5.6 that the data (here, p and k) and a pair of solutions (here, y and r) with the *same* value (i.e., without a duality gap) can be permuted to form the data and solutions to another programme pair. As the example shows, this need not be so when there *is* a duality gap. Indeed, none of the other programme pairs need have a gap. In this example, the SRP programme pair does have a gap, but the LRC and the SRC programme pairs do not, since both cost functions are semicontinuous in the quantities (which means Type One semicontinuity). That is, C_{LR} is L^1-l.s.c. (and *a fortiori* $L^{\infty*}$-l.s.c.) in $y \in L^{\infty}$.[3] The same is obviously true of C_{SR} as a function of (y, k), which is simply the 0-∞ indicator function of the closed set $\mathbb{Y}$. (There are no variable inputs with this technique, i.e., the SRC programme is merely a check of capacity sufficiency.) So permutation of p, k, y and r must fail to yield a cost-minimising solution and its dual, and it does fail: (i) the LRC programme's solution has $k_{Co} = 0$, unlike the SRP data in this example; and (ii) the OFIV (dual to SRC) programme's solution has $r_{Co} = 0$, unlike the FIV (dual to SRP) programme's solution, which has $r_{Co} = \|p_{FA}\|_{\infty}^{*} > 0$. (In detail, the SRP primal-dual solution pair—given a nonconstant $p_{CA} \in \mathrm{BV}$ and $p_{FA} \neq 0$, $k_{St} = 0$ and $k_{Co} > 0$—is $y = 0$ and $r = (r_{St}, r_{Co}) = \left(\mathrm{Var}_c^+ (p_{CA}), \|p_{FA}\|_{\infty}^{*}\right) \gg 0$. But, given $y = 0$ and $r = (r_{St}, r_{Co}) \gg 0$, the LRC solution pair is obviously $(k_{St}, k_{Co}) = (0, 0)$ with any LRMC as p—i.e., with any $p \in r_{St}\partial\check{k}_{St}(0) + r_{Co}\partial\check{k}_{Co}(0) + \text{const.}$, where $\check{k}_{St}$ and $\check{k}_{Co}$ are the input requirement functions (5.13.6)–(5.13.7). Similarly, given $y = 0$, $k_{St} = 0$ and $k_{Co} > 0$, the SRC dual solution is $r_{Co} = 0$ with any $r_{St} \geq 0$ and any $p \in r_{St}\partial\check{k}_{St}(0) + \text{const.}$)

[3] This can be shown directly from the formulae for capacity requirements (5.13.6)–(5.13.7).

Appendix B

A NONFACTORABLE JOINT SUBDIFFERENTIAL

A class is identified of jointly convex functions of two variables (which can be vector variables) such that: (i) nondifferentiability in one of the variables implies nondifferentiability in the other, and (ii) the joint subdifferentials do *not* factorise into the Cartesian product of the partial subdifferentials. This means that a partial subgradient cannot be extended to a joint one by adjoining just any partial subgradient w.r.t. the other variable. But, as is also shown, it can usually be extended by a suitable choice of the other partial subgradient.

Proposition B.0.2. *Assume that $C\colon Y \times K \to \mathbb{R} \cup \{+\infty\}$ is (jointly) positively linearly homogeneous, convex and lower semicontinuous (for the pairing of the space $Y \times K$ with $P \times R$). If additionally $(p', -r')$ and $(p'', -r'')$ are elements of $\partial_{y,k} C(y, k)$ with*[1]

$$\langle p' \mid y \rangle \neq \langle p'' \mid y \rangle \tag{B.0.1}$$

then $r' \neq r''$ (so $\partial_k C(y, k)$ is not a singleton, i.e., $C(y, \cdot)$ is not Gateaux-differentiable at k). What is more, neither $(p', -r'')$ nor $(p'', -r')$ is in $\partial_{y,k} C(y, k)$, and so

$$\partial_{y,k} C(y, k) \neq \partial_y C(y, k) \times \partial_k C(y, k) .$$

Proof. By (C.6.11), which is a variant of Euler's Theorem,

$$C(y, k) = \langle p \mid y \rangle - \langle r \mid k \rangle \tag{B.0.2}$$

for every $(p, -r) \in \partial_{y,k} C(y, k)$. So (B.0.2) holds for both $(p', -r')$ and $(p'', -r'')$, but it therefore fails for $(p', -r'')$ and $(p'', -r')$ because of (B.0.1). So neither $(p', -r'')$ nor $(p'', -r')$ is in $\partial_{y,k} C(y, k)$, which shows that this set is not a Cartesian product. □

Example B.0.3. *Take the function $c\colon \mathbb{R}^2_+ \to \mathbb{R}$ defined as in (5.2.7), i.e., $c(\mathrm{y}, k) = w\mathrm{y}$ if $0 \leq \mathrm{y} \leq k$ and $+\infty$ otherwise (given a number $w \geq 0$). With the scalar product $\langle \mathrm{p}, -r \mid \mathrm{y}, k \rangle := \mathrm{py} - rk/T$ where $T > 0$ is a given number, the joint subdifferential at a point with $\mathrm{y} = k > 0$ is*

$$\partial_{\mathrm{y},k} c(\mathrm{y}, k) = \left\{ (\mathrm{p}, -r) \in \mathbb{R}_+ \times \mathbb{R}_- : p = w + \frac{r}{T},\ r \geq 0 \right\}$$

(which, being a half-line not parallel to either axis of the plane $\mathbb{R}^2$, is not a Cartesian product).

When c serves as a convex integrand, this non-factorisation is inherited by the integral functional

$$C(y, k) := \int_0^T c(y(t), k)\, \mathrm{d}t \quad \text{for } y \in L^\infty[0, T] .$$

[1] The minus sign in $(p, -r)$ is there to make r nonnegative when $C(y, \cdot)$ is nonincreasing on K.

Take a y and k with $0 \ll y \leq k$ and $\operatorname{meas}\{t \in [0,T] : y(t) = k\} > 0$. *When $L^1[0,T] \times \mathbb{R}$ is paired with $L^\infty[0,T] \times \mathbb{R}$ by the scalar product $\langle p, -r \mid y, k\rangle := \int_0^T p(t)\, y(t)\, \mathrm{d}t - rk$, one has $(p, -r) \in \partial_{y,k} C(y,k)$ if and only if both $p = w + \kappa$ and $r = \int_0^T \kappa(t)\, \mathrm{d}t$ for some $\kappa \in L^1_+[0,T]$ with $\kappa(t) = 0$ for a.e. $t \in [0,T]$ such that $y(t) < k$.*

Besides this example, Condition (B.0.1) is met by some $(p', -r')$ and $(p'', -r'')$ from $\partial_{y,k} C(y,k)$ if: (i) Y is a vector lattice, P is a sublattice of the order dual $Y^\sim$, and y is strictly positive as a linear functional on $Y^\sim$, (ii) $\partial_y C(y,k)$ contains a p' and a p'' with $p' < p''$,[2] and (iii) Every partial subgradient $p \in \partial_y C(y,k)$ can be extended to a joint subgradient $(p, -r) \in \partial_{y,k} C(y,k)$. Such extensibility can be proved in two ways. Both are based on the Hahn-Banach Extension Theorem, which can be stated as follows in terms of subgradients.

Theorem B.0.4 (Hahn-Banach). *Assume that $C\colon Y \times K \to \mathbb{R} \cup \{+\infty\}$ is a (jointly) convex function, where Y and K are topological vector spaces (with P and R as the continuous duals). If $k \in \operatorname{int}_K \operatorname{dom}(C(y,\cdot))$, i.e., $C\left(y, \widetilde{k}\right) < +\infty$ for every $\widetilde{k}$ in some neighbourhood of k, then for every $p \in \partial_y C(y,k)$ there exists an r such that $(p, -r) \in \partial_{y,k} C(y,k)$.*

Proof. See, e.g., [58, Theorem 0.28]; although that formulation applies only when (y,k) belongs to $\operatorname{int}_{Y \times K} \operatorname{dom} C$, the same proof is valid under the weaker assumption made here. □

Theorem B.0.4 does not apply to the boundary points of the function's effective domain, which is

$$\operatorname{dom} C := \{(y,k) : C(y,k) < +\infty\}.$$

And indeed, at a boundary point, a partial subgradient may have no extension (to a joint one). But it is useful to identify those cases in which such extensions do exist. This is because the boundary points can be the points of greatest interest: e.g., when C is the SRC as a function of the output bundle y and the fixed-input bundle k, all the efficient combinations of y and k lie on the boundary of $\operatorname{dom} C$. However, if C has a *finite* convex extension C^{Ex}, defined on the whole space (or at least on a neighbourhood of $\operatorname{dom} C$), and $\operatorname{dom} C$ is the sublevel set of another finite convex function C^{Do}, then Theorem B.0.4 can be applied to both functions, C^{Ex} and C^{Do}. For the original function C, this yields a result that applies also to the domain's boundary points.

Corollary B.0.5. *Let $C\colon Y \times K \to \mathbb{R} \cup \{+\infty\}$ be a (jointly) convex function. Assume that:*

1. *Its effective domain has the form*

$$\operatorname{dom} C = \left\{(y,k) : C^{\mathrm{Do}}(y,k) \leq 0 \text{ and } k \in K_0\right\} \tag{B.0.3}$$

where K_0 is a convex subset of K, and $C^{\mathrm{Do}}\colon Y \times K \to \mathbb{R}$ is a continuous convex function.

[2] Then $\langle p' \mid y\rangle < \langle p'' \mid y\rangle$, since $p' < p''$ and $y \gg 0$.

2. $k \in K_0$ *and* $C^{\mathrm{Do}}(y,k) \le 0$, *i.e.,* $(y,k) \in \operatorname{dom} C$.

3. There exists a $y^{\mathrm{S}} \in Y$ *with* $C^{\mathrm{Do}}\left(y^{\mathrm{S}},k\right) < 0$.

4. C *(or, more precisely, its restriction to* $\operatorname{dom} C$*) has a continuous convex extension* $C^{\mathrm{Ex}}\colon Y \times K \to \mathbb{R}$.

Then for every $p \in \partial_y C(y,k)$ *there exists an* r *such that* $(p,-r) \in \partial_{y,k} C(y,k)$.

Proof. This is only sketched. Every $p \in \partial_y C(y,k)$ has the form $p = p' + \alpha p''$ for some $p' \in \partial_y C^{\mathrm{Ex}}(y,k)$, $p'' \in \partial_y C^{\mathrm{Do}}(y,k)$ and a scalar $\alpha \ge 0$, with $\alpha = 0$ if $C^{\mathrm{Do}}(y,k) < 0$. By Theorem B.0.4 applied to C^{Ex} and to C^{Do}, there exist r' and r'' with

$$(p',-r') \in \partial_{y,k} C^{\mathrm{Ex}}(y,k) \text{ and } (p'',-r'') \in \partial_{y,k} C^{\mathrm{Do}}(y,k). \tag{B.0.4}$$

It now suffices to set $r := r' + \alpha r''$. For details, see [46]. □

The other way of proving Corollary B.0.5 is to establish that the relevant partial conjugate of the bivariate convex function C is superdifferentiable in the non-conjugated variable. That is, introduce the saddle (convex-concave) function on $P \times K$ defined by $\Pi := C^{\#_1}$, then show (by using Assumption 3 as Slater's Condition for maximisation of $\langle p \,|\, y\rangle - C^{\mathrm{Ex}}(y,k)$ over y subject to $C^{\mathrm{Do}}(y,k) \le 0$) that $\widehat{\partial}_k \Pi(p,k) \ne \emptyset$ for the given k and the given $p \in \partial_y C(y,k)$. Finally, apply the Subdifferential Sections Lemma (i.e., Lemma C.7.2) to conclude that any $r \in \widehat{\partial}_k \Pi(p,k)$ extends p to a $(p,-r) \in \partial_{y,k} C(y,k)$. When there is an explicit formula for Π, this can also be an effective method of calculating such an r.

Comments: Extensibility of partial subgradients means that the obvious inclusions $\partial_{y,k} C_{\mathrm{SR}} \subseteq \partial_y C_{\mathrm{SR}} \times \partial_k C_{\mathrm{SR}}$ and $\partial_y C_{\mathrm{LR}} \subseteq \partial_y C_{\mathrm{SR}}$ are "tight", each in its sense:

- $\partial_y C_{\mathrm{SR}}$ is equal to the projection of $\partial_{y,k} C_{\mathrm{SR}}$ onto Y if and only if every $p \in \partial_y C_{\mathrm{SR}}$ extends to some $(p,-r) \in \partial_{y,k} C_{\mathrm{SR}}$. A similar result applies to $\partial_k C_{\mathrm{SR}}$.
- With C_{LR} defined by (5.9.2), if every $p \in \partial_y C_{\mathrm{SR}}(y,k)$ extends to some $(p,-r) \in \partial_{y,k} C_{\mathrm{SR}}(y,k)$ then
$$\partial_y C_{\mathrm{SR}}(y,k) = \bigcup_{r \in -\partial_k C_{\mathrm{SR}}(y,k)} \partial_y C_{\mathrm{LR}}(y,r).$$
This follows from the second equivalence in (5.9.1), which is a case of the SSL (Lemma C.7.2). A similar result for C_{SR} and Π_{SR} shows that the inclusion (5.9.4) is tight.

Appendix C

Convex conjugacy and subdifferential calculus

C.1 Semicontinuous envelope

Let $C\colon Y \to \mathbb{R} \cup \{\pm\infty\}$ be a convex extended-real function on a real vector space Y that is paired with another one, P, by a bilinear form $\langle \cdot \mid \cdot \rangle : P \times Y \to \mathbb{R}$. The *effective domain* of C is the convex set

$$\operatorname{dom} C := \{y \in Y : C(y) < +\infty\}.$$

Given a locally convex topology $\mathcal{T}$ on Y that is *consistent* with P (i.e., makes P the continuous dual space), the *l.s.c. envelope* of C is the greatest lower semicontinuous (l.s.c.) minorant of C. Denoted by $\operatorname{lsc} C$, it can be determined pointwise by the formula

$$(\operatorname{lsc} C)(y) := \min\left\{C(y), \liminf_{y' \to y} C(y')\right\}$$

or globally by the formula $\operatorname{epi} \operatorname{lsc} C := \operatorname{cl} \operatorname{epi} C$, where cl means the $\mathcal{T}$-closure, and

$$\operatorname{epi} C := \{(y, \varrho) \in Y \times \mathbb{R} : C(y) \leq \varrho\}$$

is the epigraph of C. Note that $\operatorname{lsc} C$ depends on the dual space P but not on the consistent topology $\mathcal{T}$, by the Hahn-Banach Separation Theorem [32, 12A: Corollary 1]. Also, C is l.s.c. at y if and only if $C(y) = (\operatorname{lsc} C)(y)$.

A *proper convex* function is one that takes a finite value (somewhere) but does not take the value $-\infty$ (anywhere). A convex function taking the value $-\infty$ is peculiar: it may take finite values only on the algebraic boundary of its effective domain,[1] and it has no finite value at all if it is lower semicontinuous along each straight line: see, e.g., [70, 7.2 and 7.2.1], [73, Theorem 4] or [80, 5.12 with Proof].

C.2 The conjugate function

The Fenchel-Legendre *convex conjugate* of C is

$$C^{\#}(p) := \sup_{y \in Y} (\langle p \mid y \rangle - C(y)) \tag{C.2.1}$$

[1] In precise terms, $C(y) = -\infty$ for every y in the intrinsic core (a.k.a. the relative algebraic interior) of $\operatorname{dom} C$ if $C(y') = -\infty$ for some y' (and C is convex).

for $p \in P$; it is l.s.c. and either proper convex or an infinite constant ($+\infty$ or $-\infty$). Obviously

$$C^{\#}(p) \geq \langle p \,|\, y \rangle - C(y) \tag{C.2.2}$$

for every y and p; this is the Fenchel-Young Inequality.

The second convex conjugate, $C^{\#\#}$, is the pointwise supremum of all the affine minorants of C with coefficients in P (supremum of those functions of the form $\langle p \,|\, \cdot \rangle - \varrho$, with $p \in P$ and $\varrho \in \mathbb{R}$, that nowhere exceed C), i.e.,

$$C^{\#\#}(y) = \sup_{p \in P,\, \varrho \in \mathbb{R}} \{ \langle p \,|\, y \rangle - \varrho : \langle p \,|\, y' \rangle - \varrho \leq C(y') \text{ for every } y' \in Y \}. \tag{C.2.3}$$

So $C^{\#\#}$ is l.s.c. on Y and

$$C^{\#\#} \leq \operatorname{lsc} C \leq C. \tag{C.2.4}$$

Furthermore, $C^{\#\#} = \operatorname{lsc} C$ unless $\operatorname{lsc} C$ takes the value $-\infty$ (and hence has no finite value).[2] In the latter case, $C^{\#\#} = -\infty$ (everywhere on Y) and $\operatorname{lsc} C = -\infty$ on the convex set $\operatorname{cl} \operatorname{dom} C$, but $\operatorname{lsc} C$ $= +\infty$ on the complement set. So if C is l.s.c. at y then: (i) $C^{\#\#}(y)$ and $C(y)$ can differ only by being oppositely infinite, and (ii) $C^{\#\#}(y) = C(y)$ if and only if either $C(y) < +\infty$ or both $C(y)$ $= +\infty$ and $\operatorname{lsc} C > -\infty$ everywhere on Y. Also, $C^{\#\#} = C$ (everywhere on Y) if and only if C is either l.s.c. proper convex or an infinite constant.[3] Applied to $C^{\#}$ (instead of C), this shows that

$$C^{\#\#\#} = C^{\#} \tag{C.2.5}$$

(which can also be seen directly from (C.2.1) and (C.2.4): $\left(C^{\#\#}\right)^{\#} \geq C^{\#}$ because $C^{\#\#} \leq C^{\#}$, but also $\left(C^{\#}\right)^{\#\#} \leq C^{\#}$).

For a bivariate convex function C, its *partial* second conjugate (i.e., its second conjugate taken w.r.t. just one variable y, with the other variable k kept fixed) lies always between the *total* second conjugate (i.e., the second conjugate w.r.t. both variables) and the original function itself. Formally, the partial first and second conjugates w.r.t., say, the first variable of a bivariate convex function C on $Y \times K$ (where K is another vector space) is defined by

$$C^{\#_1}(p, k) := (C(\cdot, k))^{\#}(p) := \sup_{y \in Y} (\langle p \,|\, y \rangle - C(y, k)) \tag{C.2.6}$$

for every $p \in P$ and $k \in K$. This ($C^{\#_1}$) is a saddle (convex-concave) function on $P \times K$: it is convex (like C) in the "conjugated" first variable, but (unlike C) it is concave in the non-conjugated second variable. The partial second conjugate (w.r.t. the first variable) is the bivariate convex function

$$C^{\#_1\#_1}(y, k) := (C(\cdot, k))^{\#\#}(y). \tag{C.2.7}$$

[2] When additionally Y is finite-dimensional, if $\operatorname{lsc} C$ takes the value $-\infty$, then so does C itself. This follows from [70, 7.5]; it is stated in, e.g., [73, Example 1"].

[3] In [70] and [73], C is called "closed" when $C = C^{\#\#}$, and $\operatorname{cl} C$ serves as an alternative notation for $C^{\#\#}$. This is abandoned in [74], and rightly so: $\operatorname{cl} C$ can be misinterpreted as $\operatorname{lsc} C$, especially since others—e.g., [58]—do use $\operatorname{cl} C$ instead of $\operatorname{lsc} C$ (to have $\operatorname{epi} \operatorname{cl} C := \operatorname{cl} \operatorname{epi} C$).

Remark C.2.1 (Inequality between partial and total second conjugates). *Assume that* $C: Y \times K \to \mathbb{R} \cup \{\pm\infty\}$, *where* Y *and* K *are vector spaces paired with* P *and* R. *Then*

$$C^{\#\#} \leq C^{\#_1\#_1} \leq C \tag{C.2.8}$$

on $Y \times K$. *(In other words, for each* $k \in K$, *if* C_k *means the function on* Y *defined by* $C_k(y) := C(y,k)$ *for every* y, *then* $\left(C^{\#\#}\right)_k \leq (C_k)^{\#\#} \leq C_k$ *on* Y.*)*

Proof. The second inequality of (C.2.8) is a case of (C.2.4), without the middle term. As for the first inequality of (C.2.8), this follows from a comparison, for the partial and total second conjugates, of their representations as suprema of affine minorants: by (C.2.3) applied to $C(\cdot,k)$ and to C,

$$C^{\#_1\#_1}(y,k) = \sup_{p \in P,\, \alpha \in \mathbb{R}} \{\langle p \,|\, y\rangle - \alpha : \langle p \,|\, \cdot\rangle - \alpha \leq C(\cdot,k)\} \tag{C.2.9}$$

$$C^{\#\#}(y,k) = \sup_{p \in P,\, r \in K,\, \beta \in \mathbb{R}} \{\langle p, -r \,|\, y, k\rangle - \beta : \langle p, -r \,|\, \cdot, \cdot\rangle - \beta \leq C(y')\}. \tag{C.2.10}$$

By setting α equal to $\langle r \,|\, k\rangle + \beta$, it follows that the supremum in (C.2.9) is not less than that in (C.2.10).[4] □

C.3 Subgradients

A $\mathcal{T}$-continuous *subgradient* (a.k.a. topological subgradient) of C at a $y \in Y$ is any $p \in P$ such that

$$C(y + \Delta y) \geq C(y) + \langle p \,|\, \Delta y\rangle \tag{C.3.1}$$

for every $\Delta y \in Y$. The set of all subgradients (at y) is the *subdifferential* $\partial C(y)$. In other words,

$$p \in \partial C(y) \Leftrightarrow y \text{ maximises } \langle p \,|\, \cdot\rangle - C \tag{C.3.2}$$

$$\Leftrightarrow C^{\#}(p) = \langle p \,|\, y\rangle - C(y). \tag{C.3.3}$$

So the graph of the subdifferential correspondence ($\partial C \subseteq Y \times P$) consists of those points (y,p) at which the Fenchel-Young Inequality holds as an equality.

Any linear, not necessarily $\mathcal{T}$-continuous, functional p meeting (C.3.1) is an *algebraic* subgradient of C at y, and the set of *all* such subgradients is the *algebraic subdifferential* $\partial^{\mathrm{a}} C(y)$, with $P \cap \partial^{\mathrm{a}} C(y) = \partial C(y)$ by definition. The two subdifferentials are identical, for every C, when $\mathcal{T}$ is the *strongest locally convex topology*, $\mathcal{T}_{\mathrm{SLC}}$, on Y. This is because every linear functional on Y is $\mathcal{T}_{\mathrm{SLC}}$-continuous, i.e., the $\mathcal{T}_{\mathrm{SLC}}$-continuous dual is equal to the *algebraic dual* Y^{a} (what is more, $\mathcal{T}_{\mathrm{SLC}}$ is obviously $\mathrm{m}(Y, Y^{\mathrm{a}})$, the Mackey topology for this pairing).

[4] In other words, the α in (C.2.9) is allowed to vary with k in any way (subject to the stated inequality), whilst the corresponding term in (C.2.10) is $\langle r \,|\, k\rangle + \beta$, which is additionally linear in k.

Directly from the subgradient inequality (C.3.1), if C' and C'' are convex functions with values in $\mathbb{R} \cup \{+\infty\}$, i.e., not taking the value $-\infty$, then

$$\partial (C' + C'') (y) \supseteq \partial C' (y) + \partial C'' (y) . \tag{C.3.4}$$

Equality holds for proper convex functions under a continuity assumption: if, in addition to C' and C'' being convex with values in $\mathbb{R} \cup \{+\infty\}$, there exists a point of Y at which both C' and C'' are finite and at least one (C' or C'') is continuous, then

$$\partial (C' + C'') (y) = \partial C' (y) + \partial C'' (y) \tag{C.3.5}$$

for *every* $y \in Y$. See, e.g., [73, Theorem 20 (i) under (a)] or [80, 5.38 (b)]. Applied to the case of 0-∞ indicator functions of convex subsets of Y, (C.3.5) gives the outward normal cone to the intersection of sets Z' and Z'' as the sum of their normal cones, i.e.,

$$\begin{aligned} \mathrm{N} (y \mid Z' \cap Z'') &:= \partial \delta (y \mid Z' \cap Z'') = \partial \delta (y \mid Z') + \partial \delta (y \mid Z'') \\ &=: \mathrm{N} (y \mid Z') + \mathrm{N} (y \mid Z'') \end{aligned} \tag{C.3.6}$$

for *every* $y \in Y$ if $Z' \cap \operatorname{int} Z'' \neq \emptyset$. This is stated in, e.g., [51, 4.3: Proposition 1].

Also directly from (C.3.1), for every $\alpha > 0$,

$$\partial (\alpha C) (y) = \alpha \partial C (y) \tag{C.3.7}$$

and this holds for $\alpha = 0$ as well if (and only if) $\partial C (y) \neq \emptyset$, i.e., if C is *subdifferentiable* at y.

For C to be subdifferentiable at y, it is necessary that C be l.s.c. at y and actually that $C^{\#\#} (y) = C (y)$; in this case $\partial C^{\#\#} (y) = \partial C (y)$. In other words,

$$p \in \partial C (y) \Leftrightarrow \left(p \in \partial C^{\#\#} (y) \text{ and } C^{\#\#} (y) = C (y)\right) \tag{C.3.8}$$

from (C.3.3) and (C.2.4).

Lower semicontinuity is not generally sufficient for subdifferentiability, but continuity is. In precise terms, if a proper convex function $C \colon Y \to \mathbb{R} \cup \{+\infty\}$ is continuous and finite at some point of Y, then it is subdifferentiable (and continuous) at every interior point of its effective domain, i.e., $\partial C (y)$ is nonempty and, also, $\mathrm{w} (P, Y)$-compact (weakly compact) for every $y \in \operatorname{int} \operatorname{dom} C$: see, e.g., [51, 4.2: Proposition 3], [73, Theorem 11 (a)] or [80, 5.35 (a)]. Furthermore, every algebraic subgradient is then $\mathcal{T}$-continuous, i.e., $\partial^{\mathrm{a}} C (y) = \partial C (y) \neq \emptyset$ or, equivalently,

$$\emptyset \neq \partial^{\mathrm{a}} C (y) \subseteq P \tag{C.3.9}$$

for every $y \in \operatorname{int} \operatorname{dom} C$: see, e.g., [32, 14B: Proof of Theorem] or [58, Corollary 2 to Theorem 0.27, and p. 60].

C.4 Continuity of convex functions

Any continuous function is bounded from above (by a finite number) on a neighbourhood of any point where its value is either finite or $-\infty$. With *convex* functions, this obvious necessary condition is also sufficient for continuity. In precise terms, if $C: Y \to \mathbb{R} \cup \{\pm\infty\}$ is convex then the following conditions are equivalent to one another:

1. C is continuous at some $y \in Y$ with $C(y) < +\infty$.

2. There exists an open set $N \subseteq Y$ and a $\varrho \in \mathbb{R}$ such that $C(y) \le \varrho$ (or, equivalently, the epigraph of C has a nonempty interior in $Y \times \mathbb{R}$).

3. C is continuous on $\operatorname{int}\operatorname{dom} C$, which is nonempty.

See, e.g., [32, 14A], [51, 3.2: Theorem 1], [73, Theorem 8] or [80, 5.20]. In particular, this shows that continuity (of a convex function) is a property that "propagates" from any single point to the whole interior of the effective domain (Part 1 $\Rightarrow$ Part 3). Also, the sufficiency of local boundedness for continuity can be combined with a Baire category argument to deduce continuity from mere lower semicontinuity for a convex function on a Banach space (or, more generally, on a barrelled space). The result has two variants (which are very similar, but not identical): see, e.g., [73, Corollary 8B] and [32, p. 84 and Exercise 3.50].

Another "automatic continuity" result, limited to finite-dimensional spaces, is that a finite convex function C on a polyhedral set $Z \subseteq \mathbb{R}^n$ is upper semicontinuous on Z (so if C is also l.s.c. on Z then it is actually continuous on Z). More generally, a convex function $C: \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ is u.s.c. on any locally simplicial (not necessarily convex or closed) subset, Z, of $\operatorname{dom} C$. See [70, 10.2 and 20.5].

C.5 Concave functions and supergradients

All of these concepts and results can be reoriented to concave functions. In particular, when $\Pi: K \to \mathbb{R} \cup \{\pm\infty\}$ is a concave function on a space K paired with another space R, its *effective domain* (in the concave sense) is the convex set

$$\widehat{\operatorname{dom}}\Pi := \{k \in K : \Pi(k) > -\infty\}$$

and the *concave conjugate* of Π is

$$\Pi_{\#}(r) := \inf_{k \in K} (\langle r \mid k \rangle - \Pi(k)) \tag{C.5.1}$$

for $r \in R$. The second concave conjugate meets the inequality

$$\Pi_{\#\#}(k) \ge \operatorname{usc}\Pi(k) \ge \Pi(k) \tag{C.5.2}$$

where $\operatorname{usc}\Pi$ is the least upper semicontinuous (u.s.c.) majorant of Π; and $\operatorname{usc}\Pi(k)$ and $\Pi_{\#\#}(k)$ differ in only one case: if $k \notin \operatorname{cl}\widehat{\operatorname{dom}}\Pi$ and $\operatorname{usc}\Pi(k'') = +\infty$ for some k'', then $\operatorname{usc}\Pi(k) = -\infty$ but $\Pi_{\#\#} = +\infty$ (on K). So if Π is u.s.c. *proper* concave (i.e., takes a finite value but does not take the value $+\infty$), then $\Pi_{\#\#} = \Pi$ (everywhere). Hence

$$\Pi_{\#\#\#} = \Pi_{\#}. \tag{C.5.3}$$

A *supergradient* of Π at a $k \in K$ is any $r \in R$ such that

$$\Pi(k + \Delta k) \leq \Pi(k) + \langle r \mid \Delta k \rangle \tag{C.5.4}$$

for every $\Delta k \in K$. The set of all supergradients (at k) is the *superdifferential*, $\widehat{\partial}\Pi(k)$, i.e.,

$$r \in \widehat{\partial}\Pi(k) \Leftrightarrow k \text{ maximises } \Pi - \langle r \mid \cdot \rangle \tag{C.5.5}$$

$$\Leftrightarrow \Pi_{\#}(r) = \langle r \mid k \rangle - \Pi(k). \tag{C.5.6}$$

Also,

$$r \in \widehat{\partial}\Pi(k) \Leftrightarrow \left(r \in \widehat{\partial}\Pi_{\#\#}(k) \text{ and } \Pi_{\#\#}(k) = \Pi(k)\right). \tag{C.5.7}$$

The concave and convex cases are linked by the rules:

$$\Pi_{\#}(r) = -(-\Pi)^{\#}(-r) \tag{C.5.8}$$

$$\Pi_{\#\#} = -(-\Pi)^{\#\#} \tag{C.5.9}$$

$$\operatorname{usc}\Pi = -\operatorname{lsc}(-\Pi) \tag{C.5.10}$$

$$\widehat{\partial}\Pi = -\partial(-\Pi). \tag{C.5.11}$$

C.6 Subgradients of conjugates

The subdifferential correspondences of mutual conjugates are inverse to each other.[5]

Theorem C.6.1 (Inversion Rule). *Assume that $C\colon Y \to \mathbb{R} \cup \{\pm\infty\}$ is convex, and Y is paired with P. Then, for every $y \in Y$ and $p \in P$*

$$p \in \partial C(y) \Leftrightarrow \left(y \in \partial C^{\#}(p) \text{ and } C^{\#\#}(y) = C(y)\right). \tag{C.6.1}$$

For a concave function Π (on a space K paired with R), this becomes

$$r \in \partial\Pi(k) \Leftrightarrow \left(k \in \partial\Pi_{\#}(r) \text{ and } \Pi_{\#\#}(k) = \Pi(k)\right). \tag{C.6.2}$$

[5] This is given in, e.g., [5, 4.4.4], [70, 23.5 (a) and (a*)] and [73, Corollary 12A].

Proof. This follows from the Fenchel-Young Inequality and from the case of equality therein as a characterisation of the subdifferential: apply (C.3.2)–(C.3.3) twice, to C and to $C^{\#}$ (in place of C), to see that the conditions $p \in \partial C(y)$ and $y \in \partial C^{\#}(p)$ are equivalent when $C^{\#\#}(y) = C(y)$. It remains to show that this equality holds when $p \in \partial C(y)$. And this is because, by (C.2.2) and by (C.3.2)–(C.3.3) applied to $C^{\#}$, $C^{\#\#}(y) \geq \langle p \mid y \rangle - C^{\#}(p) = C(y) \geq C^{\#\#}(y)$ by (C.2.4). □

The Inversion Rule and the First-Order Condition (C.3.2) are next combined in a derivative property of conjugate functions. In convex programming, this yields the derivative property of the optimal value.

Corollary C.6.2 (Derivative Property of the Conjugate). *Assume that* $C\colon Y \to \mathbb{R}\cup\{\pm\infty\}$ *is convex (and* Y *is paired with* P*). Then, for every* $y \in Y$ *and* $p \in P$,

$$y \text{ maximises } \langle p \mid \cdot \rangle - C \Leftrightarrow \left(y \in \partial C^{\#}(p) \text{ and } C^{\#\#}(y) = C(y)\right). \tag{C.6.3}$$

When C *is lower semicontinuous proper convex on* Y*, this means that*

$$\partial C^{\#}(p) = \operatorname{argmax}\left(\langle p \mid \cdot \rangle - C\right) \tag{C.6.4}$$

for every $p \in P$.[6]

Proof. The equivalence (C.6.3) follows from the FOC (C.3.2) and the Inversion Rule (C.6.1). And (C.6.4) follows from (C.6.3) because $C^{\#\#} = C$ in this case. □

The convex conjugate of the 0-∞ indicator $\delta(\cdot \mid Z)$ of a set $Z \subseteq Y$ (i.e., of the function equal to 0 on Z and $+\infty$ on $Y \setminus Z$) is the *support function* of Z, i.e.,

$$\delta^{\#}(p \mid Z) = \sup_{Z} \langle p \mid \cdot \rangle \tag{C.6.5}$$

and the Derivative Property (C.6.4) gives its subdifferential at a $p \in P$ as

$$\partial \delta^{\#}(p \mid Z) = \operatorname*{argmax}_{Z} \langle p \mid \cdot \rangle \tag{C.6.6}$$

if Z is nonempty, convex and closed. This is stated in, e.g., [70, 23.5.3] and [73, p. 36, lines 1–7]. Similarly, the *inf-support function* of a set $Z \subseteq R$ is the concave conjugate of $-\delta(\cdot \mid Z)$, i.e.,

$$\inf_{Z} \langle \cdot \mid k \rangle = (-\delta)_{\#}(k \mid Z) \tag{C.6.7}$$

for every $k \in K$ (the space paired with R). Its superdifferential at k is

$$\hat{\partial}(-\delta)_{\#}(k \mid Z) = \operatorname*{argmin}_{Z} \langle \cdot \mid k \rangle \tag{C.6.8}$$

[6] This is given in, e.g., [5, 4.4.5], [70, 23.5 (b) and (a*)] and [73, Corollary 12B]. It holds formally also when C is the constant $-\infty$ (but not when C is $+\infty$ because $\operatorname{argmax}(-\infty) := \emptyset$ by convention, whilst $\partial(-\infty)(p) := Y$).

if Z is nonempty, convex and closed.

Comment (proper and improper solutions): As in [74], $\operatorname{argmax}_Z f$ means the set of all maximum points of a function f on a set Z—provided that $\sup_Z f > -\infty$. Points of $\operatorname{argmax}_Z f$ maximise f *properly* (i.e., either to a finite value or to $+\infty$). When $f = -\infty$ on Z, any point of Z maximises f on Z, but $\operatorname{argmax}_Z f := \emptyset$. In other words, when a programme is infeasible, it is convenient to regard any point as an *improper solution*, as in [73, p. 38]. But note that in a dual pair of solutions with *equal* values both solutions are always proper (i.e., are feasible) or, equivalently, their common value is finite. To see this, let the primal programme be to maximise a concave $f\colon X \to \mathbb{R} \cup \{-\infty\}$; then the dual is to minimise a certain convex $g\colon Y \to \mathbb{R} \cup \{+\infty\}$ such that $f(x) \le g(y)$ for every x and y (where X and Y are vector spaces). If $\overline{x}$ maximises f, $\overline{y}$ minimises g and there is no duality gap, then $+\infty > f(\overline{x}) = g(\overline{y}) > -\infty$ (so $\overline{x} \in \operatorname{argmax} f$ and $\overline{y} \in \operatorname{argmin} g$).[7]

The support function of a nonempty set Z is sublinear—i.e., it is convex and positively linearly homogeneous (p.l.h.) or, equivalently, it is p.l.h. and subadditive. Conversely, every l.s.c. sublinear function $C\colon Y \to \mathbb{R} \cup \{+\infty\}$ is the support function of a nonempty, convex and closed set, viz., $\partial C(0)$—i.e.,

$$C(y) = \sup_{p \in \partial C(0)} \langle p \,|\, y \rangle \tag{C.6.9}$$

$$\text{where} \quad \partial C(0) := \{p : \langle p \,|\, y \rangle \le C(y)\}. \tag{C.6.10}$$

See, e.g., [51, 4.1: Proposition 1], [70, 13.2.1] or [80, 6.22]. By (C.6.6), it follows that

$$\partial C(y) := \left\{ p \in \partial C(0) : \langle p \,|\, y \rangle = \sup_{\partial C(0)} \langle \cdot \,|\, y \rangle = C(y) \right\} \tag{C.6.11}$$

which is stated in, e.g., [51, 4.2.1: Example 3], [70, 23.5.3] and [73, p. 36, lines 1–7]. This is a variant of Euler's Theorem on homogeneous functions.

C.7 Subgradients of partial conjugates

In the case of *partial* conjugacy, between a bivariate convex function C and a saddle (convex-concave) function Π, the Inversion Rule not only applies to the relevant partial derivatives but also extends to the *total* derivatives (Corollaries C.7.3 and C.7.5 below). Namely, when Π and C are differentiable, their gradient maps can be obtained from each other by transposition of that pair of variables, p and y, w.r.t. which Π and C are mutual conjugates. When Π and C are nondifferentiable, the rule

[7] This argument assumes that the maximand f is nowhere $+\infty$ and that the minimand g is nowhere $-\infty$. These sensible conditions are met when the perturbed primal constrained maximand, F, is a u.s.c. proper concave function on a space $X \times A$ paired with $B \times Y$ (where A and B are the spaces of primal and dual perturbations). This is because: (i) $f(x) = F(x, 0) < +\infty$ for every x, and (ii) the perturbed dual constrained maximand, $G(b, y) := -F_{\#}(-b, y)$, is then l.s.c. proper convex, and so $g(y) := G(0, y) > -\infty$ for every y: see, e.g., [73, (4.17)].

applies to their subdifferential correspondences—i.e., to the "saddle differential" $\partial_p \Pi \times \widehat{\partial}_k \Pi$ and the joint subdifferential $\partial_{y,k} C$ (which does not usually factorise into $\partial_y C \times \partial_k C$). This rule is based on a key lemma, useful also by itself,[8] which identifies the section of the joint subdifferential $\partial_{y,k} C$ through a $p \in \partial_y C$ as $-\widehat{\partial}_k \Pi$, the partial subdifferential of $-\Pi$ w.r.t. the argument k that it shares with C (Lemma C.7.2).

These relationships between a saddle function Π and its bivariate convex "parent" C are spelt out below. First, since Π is the partial conjugate of C w.r.t. one variable, the total (bivariate) conjugate of C is the partial conjugate of $-\Pi$ w.r.t. the other variable.

Lemma C.7.1 (Total conjugacy by stages). *Assume that* $C\colon Y \times K \to \mathbb{R} \cup \{\pm\infty\}$ *and let the spaces* Y *and* K *be paired with* P *and* R*. Then, in the notation of (C.2.6),*

$$C^{\#} = \left(-C^{\#_1}\right)^{\#_2}$$

on $P \times R$*. In other words, if*

$$\Pi (p, k) = C^{\#_1} (p, k) := \sup_y \left(\langle p \,|\, y\rangle - C (y, k)\right) \tag{C.7.1}$$

for every $p \in P$ *and* $k \in K$*, then*

$$C^{\#} (p, -r) = (-\Pi)^{\#_2} (p, -r) := \sup_k \left(\Pi (p, k) - \langle r \,|\, k\rangle\right)$$

for every $p \in P$ *and* $r \in R$*.*

Proof. For every $(p, r) \in P \times R$

$$\begin{aligned} C^{\#} (p, -r) &= \sup_{y,k} \left(\langle p \,|\, y\rangle - \langle r \,|\, k\rangle - C (y, k)\right) = \sup_k \left(- \langle r \,|\, k\rangle + \sup_y \left(\langle p \,|\, y\rangle - C (y, k)\right)\right) \\ &= \sup_k \left(\Pi (p, k) - \langle r \,|\, k\rangle\right) \end{aligned}$$

as required. □

Comment ("staged" conjugacy and alternative proofs of the inequality between partial and total second conjugates): Also the second conjugate can be obtained in stages, i.e.,

$$C^{\#\#} = C^{\#_1\#_1\#_2\#_2}.$$

That is, the total second conjugate of C is equal to the partial second conjugate, w.r.t. one variable, of the partial second conjugate of C w.r.t. the other variable. This gives another proof of the first inequality in (C.2.8): $C^{\#\#} = C^{\#_1\#_1\#_2\#_2} \leq C^{\#_1\#_1}$ (by (C.2.4) applied to the function $C^{\#_1\#_1} (y, \cdot)$ on K, in place of C). Similarly, in terms of the partial second concave conjugate of $\Pi := C^{\#_1}$ w.r.t. the second variable, $C^{\#\#} = \left(\left(C^{\#_1}\right)_{\#_2\#_2}\right)^{\#_1} \leq C^{\#_1\#_1}$ (because $\Pi_{\#_2\#_2} \geq \Pi$).

[8] For example, it yields the extension (5.9.1) of the Wong-Viner Theorem.

The "staged" conjugacy is next used to "slice" the joint subdifferential of the bivariate convex function along one of the "axes" (the p-axis): the section of the set $\partial C(y,k) \subseteq P \times R$ through any $p \in \partial_y C(y,k)$ is found to be $-\widehat{\partial}_k \Pi(p,k) \subseteq \partial_k C(y,k) \subseteq R$.

Lemma C.7.2 (Subdifferential sections). *Assume that C: $Y \times K \to \mathbb{R} \cup \{+\infty\}$ is proper convex, and that Π: $P \times K \to \mathbb{R} \cup \{\pm\infty\}$ is the partial convex conjugate of C, i.e., (C.7.1) holds for each k in K (which is paired with a space R). Then the following conditions are equivalent to each other:*

1. $(p, -r) \in \partial C(y,k)$.
2. $p \in \partial_y C(y,k)$ *and* $r \in \widehat{\partial}_k \Pi(p,k)$.

Also, either condition implies that both $C(y,k)$ and $\Pi(p,k)$ are finite.

Proof. Since $C^{\#} = (-\Pi)^{\#_2}$ by Lemma C.7.1, and since $\Pi := C^{\#_1}$ by (C.7.1), one has by (C.2.1)

$$\langle p \,|\, y\rangle - C(y,k) \leq \Pi(p,k) \tag{C.7.2}$$

$$-\langle r \,|\, k\rangle + \Pi(p,k) \leq C^{\#}(p,-r) \tag{C.7.3}$$

as well as

$$\langle p \,|\, y\rangle - \langle r \,|\, k\rangle - C(y,k) \leq C^{\#}(p,-r) \tag{C.7.4}$$

for every p, y, r and k. By (C.3.3), Condition 1 is equivalent to equality in (C.7.4), which holds if and only if equalities hold in both (C.7.2) and (C.7.3). Finally, the pair of equalities is equivalent to Condition 2, again by (C.3.3).

It remains to show that the equivalent Conditions, 1 and 2, imply that $C(y,k)$ and $\Pi(p,k)$ are finite (as is also $C^{\#}(p,-r)$). For a start, note that, by assumption, C does not take the value $-\infty$, and neither does $C^{\#}$ (since C is not the constant $+\infty$). But both C and $C^{\#}$ can take the value $+\infty$. As for Π, it can take both infinite values, although for no p can the concave function $\Pi(p,\cdot)$ be the constant $-\infty$.[9]

Assume, say, Condition 1—i.e., that equality holds in (C.7.4). Since $C(y,k)$ is either finite or $+\infty$, and since so is $C^{\#}(p,-r)$, both $C(y,k)$ and $C^{\#}(p,-r)$ are actually finite (since they add up to $\langle p \,|\, y\rangle - \langle r \,|\, k\rangle$, which is finite). Given this, the inequalities (C.7.3) and (C.7.2) show that $\Pi(p,k)$ is also finite.

It is equally easy to argue from Condition 2: if equalities hold in (C.7.2) and (C.7.3), then

$$\Pi(p,k) = \langle p \,|\, y\rangle - C(y,k) < +\infty$$

$$\Pi(p,k) = C^{\#}(p,-r) + \langle r \,|\, k\rangle > -\infty$$

[9] What is more, for every $k \in K$ either (i) $\Pi(\cdot,k) = -\infty$ (everywhere on P), or (ii) $\Pi(\cdot,k)$ does not take the value $-\infty$ (anywhere on P). The latter is the case for some k (since $C(\cdot,k) \neq +\infty$ for some k); and so $\Pi(p,\cdot) \neq -\infty$ for every $p \in P$.

so $\Pi(p,k)$ is finite;[10] and hence so are $C(y,k)$ and $C^{\#}(p,-r)$. □

Finally, the Inversion Rule is applied to the partial subdifferential ($\partial_y C$) that is the range of the variable (p) indexing the sections of the joint subdifferential (∂C) in Lemma C.7.2. The result shows that, up to a sign change, the saddle-differential and the joint-subdifferential correspondences ($\partial_p \Pi \times \widehat{\partial}_k \Pi$ and $\partial_{y,k} C$) are partial inverses of each other: their graphs are identical.

Corollary C.7.3 (Partial Inversion Rule). *Under the assumptions of Lemma C.7.2, the following conditions are equivalent to each other:*[11]

1. $(p,-r) \in \partial C(y,k)$.
2. $y \in \partial_p \Pi(p,k)$ *and* $r \in \widehat{\partial}_k \Pi(p,k)$, *and* $C(\cdot,k)$ *is finite and lower semicontinuous at* y.

Also, either condition implies that both $C(y,k)$ *and* $\Pi(p,k)$ *are finite.*

Proof. By Lemma C.7.2, if $(p,-r) \in \partial C(y,k)$ then, in addition to $r \in \widehat{\partial}_k \Pi(p,k)$ and $C(y,k) < +\infty$, one has $p \in \partial_y C(y,k)$. By the Inversion Rule (C.6.1) and (C.2.4), this implies that $y \in \partial_p \Pi(p,k)$ and that $C(\cdot,k)$ is l.s.c. at y. So Condition 1 implies Condition 2.

For the converse, since $C(y,k) < +\infty$ and $C(\cdot,k)$ is l.s.c. at y, one has $C(y,k) = C^{\#_1\#_1}(y,k)$. So if $y \in \partial_p \Pi(p,k)$ then $p \in \partial_y C(y,k)$ by the Inversion Rule (C.6.1). And if additionally $r \in \widehat{\partial}_k \Pi(p,k)$, then $(p,-r) \in \partial C(y,k)$ by Lemma C.7.2. □

Comments (on the PIR and SSL):

- Finiteness of $C(y,k)$ can be dropped from Condition 2 (and the proof of its equivalence to Condition 1 simplifies) if either (i) $C(\cdot,k)$ is assumed to be l.s.c. on the whole space Y (and not just at the particular point y), or (ii) Y is finite-dimensional. This is because, in either case, the assumption (of Lemmas C.7.2 and C.7.3) that $C(\cdot,k) > -\infty$ on Y implies that $\operatorname{lsc}(C(\cdot,k)) > -\infty$ on Y (when Y is finite-dimensional, this follows from [70, 7.5]). Therefore $\operatorname{lsc}(C(\cdot,k)) = C^{\#_1\#_1}(\cdot,k)$ on Y, and so the Inversion Rule (C.6.1) shows that $p \in \partial_y C(y,k)$ if and only if both $y \in \partial_p \Pi(p,k)$ and $C(\cdot,k)$ is l.s.c. at y. Thus Corollary C.7.3 reduces immediately to Lemma C.7.2.
- There is a structural difference between the Subdifferential Sections Lemma and the Partial Inversion Rule. The SSL turns the condition $(p,-r) \in \partial_{y,k} C$ into a pair of conditions like $p \in \partial_y C$ and $r \in \widehat{\partial}_k \Pi$—which involve two functions but use partial subdifferentials w.r.t. the *same* variables as in the joint subdifferential. The PIR turns the condition $(p,-r) \in \partial_{y,k} C$ into the pair of conditions $y \in \partial_p \Pi$ and $r \in \widehat{\partial}_k \Pi$. These use a single function Π, but only one of

[10] That $\Pi(p,k) > -\infty$ can also be deduced from $r \in \widehat{\partial}_k \Pi(p,k)$, since $\Pi(p,\cdot) \neq -\infty$.
[11] This is in, e.g., [5, 4.4.14], [69, Lemma 4], [70, 37.5] and [74, 11.48].

its arguments (k) is the same as in the original function C: the other argument (y) is replaced by its dual (p) in inverting $\partial_y C$ into $\partial_p \Pi$. This step requires the semicontinuity of C w.r.t. y—and this is why the PIR is not purely algebraic like the SSL.

Remark C.7.4. *Under the assumptions of Lemma C.7.2,*

$$\widehat{\partial}_k \Pi (p, k) \subseteq -\partial_k C (y, k) \quad \text{when } p \in \partial_y C (y, k) \tag{C.7.5}$$

i.e., when y yields the supremum defining Π in (C.7.1).

Proof. Since

$$\partial C (y, k) \subseteq \partial_y C (y, k) \times \partial_k C (y, k) \tag{C.7.6}$$

$\partial_k C (y, k)$ contains the section of $\partial C (y, k)$ through any $p \in \partial_y C (y, k)$. And this section is $-\widehat{\partial}_k \Pi (p, k)$ by Lemma C.7.2. □

Comments:

- A simpler proof of (C.7.5) comes straight from the definition (C.7.1):

  $$\Pi (p, k + \Delta k) \geq \langle p \,|\, y \rangle - C (y, k + \Delta k) \text{ for every } \Delta k$$

  with equality at $\Delta k = 0$. In other words, the graph of the convex function $-\Pi (p, \cdot)$ lies below that of $C (y, \cdot)$ + const., touching it at k. It follows that $-\widehat{\partial}_k \Pi (p, k)$ is a subset of $\partial_k C (y, k)$, although this "envelope argument" does not show it ($-\widehat{\partial}_k \Pi$) to be a section of $\partial C (y, k)$ through p.

- The inclusion (C.7.6) is usually "tight" in the sense that $\partial_y C \times \partial_k C$ is the smallest Cartesian product set encasing ∂C: by (C.7.6) itself, $\partial_y C$ and $\partial_k C$ contain the projections of ∂C (onto P and R), and the reverse inclusions can be obtained by using the Hahn-Banach Extension Theorem (Theorem B.0.4 or Corollary B.0.5).

For a saddle function S with a (bivariate) convex parent I, the following useful variant of Corollary C.7.3 transposes the saddle differential correspondence ∂S into $\partial I^{\#}$ instead of ∂I (i.e., into the subdifferential correspondence of I's total conjugate instead of I itself).

Corollary C.7.5 (Dual Partial Inversion Rule). *Assume that I: $Y \times V \rightarrow \mathbb{R} \cup \{+\infty\}$ is proper convex and (jointly) lower semicontinuous for the pairing of the space V with W (and Y with P), and that $-S$: $Y \times W \rightarrow \mathbb{R} \cup \{\pm\infty\}$ is $I^{\#_2}$ (the partial convex conjugate of I), i.e.,*

$$S (y, w) = \inf_u (I (y, u) - \langle w \,|\, u \rangle) \tag{C.7.7}$$

for every $y \in Y$ and $w \in W$. Then the following conditions are equivalent to each other:

1. $(y, u) \in \partial I^{\#}(p, w)$.

2. $p \in \partial_y S(y, w)$ *and* $-u \in \widehat{\partial}_w S(y, w)$.

Also, either condition implies that both $I(p, w)$ *and* $S(y, w)$ *are finite.*

Proof. Since $I^{\#\#} = I$ by the assumption that I is l.s.c., the Inversion Rule (C.6.1) shows that Condition 1 is equivalent to: $(p, w) \in \partial I(y, u)$. And this is equivalent to Condition 2 by the Partial Inversion Rule (Corollary C.7.3) and the first Comment thereafter. □

Comment (on another derivation of DPIR): By Lemma C.7.1, the convex function $I^{\#}$ is a partial conjugate of the saddle function S; and when this relationship can be inverted to represent S as a partial conjugate of $I^{\#}$, the equivalence of $\partial I^{\#}$ and ∂S follows from the PIR alone. But this argument requires $S(\cdot, w)$ to be l.s.c. on Y, and this is a condition that S can actually fail at some points (even when I is l.s.c.). Corollary C.7.5 obviates the need to ensure that S is l.s.c. in y.

REFERENCES

[1] Afriat, S. N. (1971): "Theory of maxima and the method of Lagrange", *SIAM Journal on Applied Mathematics*, 20, 343–357.

[2] Aliprantis, C., and O. Burkinshaw (1985): *Positive operators.* New York-London: Academic Press.

[3] Anderson, D. (1976): "Models for determining least-cost investments in electricity supply", *Bell Journal of Economics*, 7, 267–299.

[4] Anderson, E. J., and P. Nash (1987): *Linear programming in infinite-dimensional spaces.* New York-Chichester-Brisbane-Toronto-Singapore: Wiley.

[5] Aubin, J. P., and I. Ekeland (1984): *Applied nonlinear analysis.* New York: Wiley.

[6] Bair, J., and R. Fourneau (1975): *Etude géometrique des espaces vectoriels* (Lecture Notes in Mathematics, vol. 489). Berlin-Heidelberg-New York: Springer-Verlag.

[7] Bauer, W., H. Gfrerer, and H. Wacker (1984): "Optimization strategies for hydro energy storage plants", *Zeitschrift für Operations Research, Series B*, 28, 103–131.

[8] Berge, C. (1963): *Topological spaces.* Edinburgh: Oliver and Boyd.

[9] Berrie, T. (1967): "The economics of system planning in bulk electricity supply", *Electrical Review*, 181. (Also in: *Public enterprise*, ed. by R. Turvey, pp. 173–211. Harmondsworth-Baltimore, MD-Ringwood, Victoria: Penguin Books, 1968.)

[10] Bewley, T. (1972): "Existence of equilibria in economies with infinitely many commodities", *Journal of Economic Theory*, 4, 514–540.

[11] Birkhoff, G. (1967): *Lattice theory.* Providence, RI: American Mathematical Society.

[12] Boiteux, M. (1964): "Peak-load pricing", in *Marginal cost pricing in practice* (Chapter 4), ed. by J. R. Nelson. Engelwood Cliffs, NJ: Prentice Hall. (A translation of "La tarification des demandes en pointe: application de la théorie de la vente au cout marginal", *Revue Général de l'Electricité*, 58 (1949), 321–340.)

[13] Boutacoff, D. (1989): "Emerging strategies for energy storage", *Electric Power Research Institute Journal*, vol. 14, No. 5, 4–13.

[14] Chong, K. M., and N. M. Rice (1971): "Equimeasurable rearrangements of functions", *Queen's Papers in Pure and Applied Mathematics*, 28.

[15] Chvatal, V. (1983): *Linear programming.* New York: Freeman.

[16] Craven, B. D. (1978): *Mathematical programming and control theory.* London: Chapman and Hall.

[17] Crew, M. A., and P. R. Kleindorfer (1979): *Public utility economics.* London-New York: MacMillan.

[18] Day, P. W. (1972): "Rearrangement inequalities", *Canadian Journal of Mathematics*, 24, 930–943.

[19] Day, P. W. (1973): "Decreasing rearrangements and doubly stochastic operators", *Transactions of the American Mathematical Society*, 178, 383–392.

[20] Dellacherie, C., and P. A. Meyer (1978): *Probabilities and potential.* Amsterdam-New York-Oxford: North-Holland.

[21] Diewert, W. E. (1974): "Applications of duality theory", in *Frontiers of quantitative economics*, Volume II (Chapter 3), ed. by M. D. Intriligator and D. A. Kendrick. Amsterdam: North-Holland.

[22] Diewert, W. E. (1982): "Duality approaches to microeconomic theory", in *Handbook of mathematical economics*, Volume II (Chapter 12), ed. by K. J. Arrow and M. D. Intriligator. Amsterdam: North-Holland.

[23] Drèze, J. H. (1964): "Some postwar contributions of French economists to theory and public policy", *American Economic Review*, 54 (supplement, June 1964), 1–64.

[24] Dubovitskii, A. Ya., and A. A. Milutin (1965): "Extremum problems in the presence of restrictions", *USSR Computational Mathematics and Mathematical Physics*, 5, 1–80.

[25] Dunford, N., and J. T. Schwartz (1958): *Linear operators, Part I: General theory.* New York: Interscience.

[26] El-Hawary, M. E., and G. S. Christensen (1979): *Optimal economic operation of electric power systems.* New York-San Francisco-London: Academic Press.

[27] Foran, J. (1991): *Fundamentals of real analysis.* New York-Basel-Hong Kong: Dekker.

[28] Gallant, A. R., and R. W. Koenker (1984): "Costs and benefits of peak-load pricing of electricity. A continuous-time econometric approach", *Journal of Econometrics*, 26, 83–113.

[29] Gfrerer, H. (1984): "Optimization of hydro energy storage plants by variational methods", *Zeitschrift für Operations Research, Series B*, 28, 87–101.

[30] Gravelle, H. S. E. (1976): "The peak-load problem with feasible storage", *Economic Journal*, 86, 256–277.

[31] Henderson, J. M., and R. E. Quandt (1971): *Microeconomic theory*. New York-London: McGraw-Hill.

[32] Holmes, R. B. (1975): *Geometric functional analysis and its applications*. Berlin-Heidelberg-New York: Springer.

[33] Horsley, A. (1982): "Electricity pricing for large supply systems", STICERD Discussion Paper TE/82/48, LSE.

[34] Horsley, A., and A. J. Wrobel (1986): "The Mackey continuity of the monotone rearrangement", *Proceedings of the American Mathematical Society*, 97, 626–628.

[35] Horsley, A., and A. J. Wrobel (1987): "The extreme points of some convex sets in the theory of majorization", *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen, Series A*, 90, 171–176.

[36] Horsley, A., and A. J. Wrobel (1988): "Subdifferentials of convex symmetric functions: An application of the Inequality of Hardy, Littlewood and Pólya", *Journal of Mathematical Analysis and Applications*, 135, 462–475.

[37] Horsley, A., and A. J. Wrobel (1991): "The closedness of the free-disposal hull of a production set", *Economic Theory*, 1, 386–391.

[38] Horsley, A., and A. J. Wrobel (1994): "The Minkowski gauge as a subgradient constant for sensitivity analysis", manuscript.

[39] Horsley, A., and A. J. Wrobel (1996): "Uninterruptible consumption, concentrated charges, and equilibrium in the commodity space of continuous functions", STICERD Discussion Paper TE/96/300, LSE.

[40] Horsley, A., and A. J. Wrobel (1996): "Comparative statics for a partial equilibrium model of investment with Wicksell-complementary capital inputs", STICERD Discussion Paper TE/96/302, LSE.

[41] Horsley, A., and A. J. Wrobel (1999): "Efficiency rents of hydroelectric storage plants in continuous-time peak-load pricing", in *The current state of economic science* (Volume 1, pp. 453–480), ed. by S. B. Dahiya. Rohtak: Spellbound Publications.

[42] Horsley, A., and A. J. Wrobel (2000): "Localisation of continuity to bounded sets for non-metrisable vector topologies and its applications to economic equilibrium theory", *Indagationes Mathematicae* (New Series), 11, 53–61.

[43] Horsley, A., and A. J. Wrobel (2002): "Boiteux's solution to the shifting-peak problem and the equilibrium price density in continuous time", *Economic Theory*, 20, 503–537.

[44] Horsley, A., and A. J. Wrobel (2003): "The Kantorovich-Rubinshtein-Vassershtein norm for measures on the circle", a manuscript.

[45] Horsley, A., and A. J. Wrobel (2005): "Continuity of the equilibrium price density and its uses in peak-load pricing", *Economic Theory*, 26, 839–866.

[46] Horsley, A., and A. J. Wrobel (2005): "Characterizations of long-run producer optima and the short-run approach to long-run market equilibrium: a general theory with applications to peak-load pricing", STICERD Discussion Paper TE/2005/490, LSE.

http://sticerd.lse.ac.uk/publications/te.asp

[47] Horsley, A., and A. J. Wrobel (2006): "Demand continuity and equilibrium in Banach commodity spaces", *Banach Center Publications*, 71 (*Game theory and mathematical economics*, ed. by A. Wieczorek, M. Malawski and A. Wiszniewska-Matyszkiel), 163–183. (Also available as CDAM Research Report LSE-CDAM-2005-01, http://www.cdam.lse.ac.uk/Reports/reports2005.html)

[48] Horsley, A., and A. J. Wrobel (2006): "Efficiency rents of a hydroelectric storage plant with a variable head", forthcoming CDAM Research Report, LSE.

[49] Horsley, A., A. J. Wrobel, and T. Van Zandt (1998): "Berge's Maximum Theorem with two topologies on the action set", *Economics Letters*, 61, 285–291.

[50] Ioffe, A. D., and Levin, V. L. (1972): "Subdifferentials of convex functions", *Transactions of the Moscow Mathematical Society*, 26, 1–72.

[51] Ioffe, A. D., and V. M. Tihomirov (1979): *Theory of extremal problems.* Amsterdam-New York-Oxford: North-Holland.

[52] Jacoby, H. P. (1967): "Analysis of investments in electric power", Economic Development Series, Center for International Affairs, Harvard University.

[53] Kantorovich, L. W., and P. G. Akilov (1982): *Functional analysis.* Pergamon Press: Oxford.

[54] Klein, E., and A. C. Thompson (1984): *Theory of correspondences.* New York-Chichester-Brisbane-Toronto: Wiley.

[55] Koopmans, T. C. (1957): "Water storage policy in a simplified hydroelectric system", in *Proceedings of the First International Conference on Operational Research*, pp. 193–227. London-Baltimore. (Also in Koopmans, T. C. (1970): *Scientific papers of Tjalling C. Koopmans*, pp. 282–316. Berlin-Heidelberg: Springer.)

[56] Koopmans, T. C. (1977): "Concepts of optimality and their uses", *American Economic Review*, 67, 261–274 (Nobel Lecture 1975).

[57] Laurent, P.-J. (1972): *Approximation et optimisation.* Paris: Hermann.

[58] Levin, V. L. (1985): *Convex analysis in spaces of measurable functions and its applications to mathematics and economics* (in Russian). Moscow: Nauka.

[59] Luxemburg, W. A. J. (1967): "Rearrangement invariant Banach function spaces", *Queen's Papers in Pure and Applied Mathematics*, 10, 83–144.

[60] Marsh, W. D. (1980): *Economics of electric utility power generation.* Oxford-New York: Oxford University Press-Clarendon Press.

[61] Marshall, A. W., and I. Olkin (1979): *Inequalities: theory of majorization and its applications.* New York-London-Toronto-Sydney: Academic Press.

[62] Meyer-Nieberg, P. (1991): *Banach lattices.* Berlin-Heidelberg-New York: Springer.

[63] Moore, T. (1994): "Storing megawatthours with SMES", *Electric Power Research Institute Journal*, vol. 19, No. 5, 24–33.

[64] Munasinghe, M., and J. J. Warford (1982): *Electricity pricing: theory and case studies.* Baltimore-London: The Johns Hopkins University Press (for the World Bank).

[65] Nguyen, D. T. (1976): "The problems of peak loads and inventories", *Bell Journal of Economics*, 7, 232–241.

[66] Phu, H. X. (1984): "Zur Stetigkeit der Lösung der adjungierten Gleichung bei Aufgaben der optimalen Steuerung mit Zustandbeschränkungen", *Zeitschrift für Analysis und ihre Anwendungen*, 3, 527–539.

[67] Phu, H. X. (1987): "On the optimal control of a hydroelectric power plant", *Systems and Control Letters*, 8, 281–288.

[68] Pyatt, G. (1978): "Marginal costs, prices and storage", *Economic Journal*, 88, 749–762.

[69] Rockafellar, R. T. (1968): "A general correspondence between dual minimax problems and convex programs", *Pacific Journal of Mathematics*, 25, 597–611.

[70] Rockafellar, R. T. (1970): *Convex analysis.* Princeton, NJ: Princeton University Press.

[71] Rockafellar, R. T. (1970): "Conjugate convex functions in optimal control and the calculus of variations", *Journal of Mathematical Analysis and Applications*, 32, 174–222.

[72] Rockafellar, R. T. (1971): "Integrals which are convex functionals, II", *Pacific Journal of Mathematics*, 39, 439–469.

[73] Rockafellar, R. T. (1974): *Conjugate duality and optimization.* Philadelphia, PA: SIAM.

[74] Rockafellar, R. T., and R. J. B. Wets (1997): *Variational analysis.* Berlin-Heidelberg-New York: Springer.

[75] Rudin, W. (1973): *Functional analysis.* New York: McGraw-Hill.

[76] Rudin, W. (1974): *Real and complex analysis.* New York: McGraw-Hill.

[77] Ryff, J. V. (1965): "Orbits of L^1-functions under doubly stochastic transformations", *Transactions of the American Mathematical Society*, 117, 92–100.

[78] Ryff, J. V. (1967): "Extreme points of some convex subsets of $L^1(0,1)$", *Proceedings of the American Mathematical Society*, 18, 1026–1034.

[79] Takayama, A. (1985): *Mathematical economics.* Cambridge-London-New York: Cambridge University Press.

[80] Tiel, J. van (1984): *Convex analysis.* Chichester-New York-Brisbane: Wiley.

[81] Tjur, T. (1972): *On the mathematical foundations of probability* (Lecture Notes 1). Institute of Mathematical Statistics, University of Copenhagen.

[82] Valadier, M. (1987): "Une singulière forme linéaire singulière sur L^∞", *Travaux du Séminaire d'Analyse Convexe*, 17, Exposé No. 4, pp. 4.1–4.3 (Université des Sciences et Techniques du Languedoc, Montpellier).

[83] Vardi, J., and B. Avi-Itzhak (1981): *Electric energy generation: economics, reliability and rates.* Cambridge, MA-London: The MIT Press.

[84] Weir, A. (1973): *Lebesgue integration and measure.* Cambridge: Cambridge University Press.

[85] Yamamuro, S. (1974): *Differential calculus in topological vector spaces* (Lecture Notes in Mathematics, vol. 374). Berlin-Heidelberg-New York: Springer.

[86] Yosida, K., and E. Hewitt (1952): "Finitely additive measures", *Transactions of the American Mathematical Society*, 72, 46–66.