Calhoun: The NPS Institutional Archive

DSpace Repository

Faculty and Researchers   Faculty and Researchers' Publications

2021-06-29

Good and Bad Optimization Models: Insights from Rockafellians

Royset, Johannes O.

ArXiv

Good and Bad Optimization Models: Insights from Rockafellians

Johannes O. Royset

Operations Research Department

Naval Postgraduate School, Monterey, California

joroyset@nps.edu

Date: June 29, 2021

Abstract. A basic requirement for a mathematical model is often that its solution (output) shouldn't change much if the model's parameters (input) are perturbed. This is important because the exact values of parameters may not be known and one would like to avoid being misled by an output obtained using incorrect values. Thus, it's rarely enough to address an application by formulating a model, solving the resulting optimization problem and presenting the solution as the answer. One would need to confirm that the model is suitable, i.e., "good," and this can, at least in part, be achieved by considering a family of optimization problems constructed by perturbing parameters as quantified by a Rockafellian function. The resulting sensitivity analysis uncovers troubling situations with unstable solutions, which we referred to as "bad" models, and indicates better model formulations. Embedding an actual problem of interest within a family of problems via Rockafellians is also a primary path to optimality conditions as well as computationally attractive, alternative problems, which under ideal circumstances, and when properly tuned, may even furnish the minimum value of the actual problem. The tuning of these alternative problems turns out to be intimately tied to finding multipliers in optimality conditions and thus emerges as a main component of several optimization algorithms. In fact, the tuning amounts to solving certain dual optimization problems. In this tutorial, we'll discuss the opportunities and insights afforded by Rockafellians.

Keywords: optimization models, Rockafellian functions, sensitivity analysis, optimality conditions, normal cones, subgradients, Rockafellian relaxation.

1   Introduction

An optimization model for a particular real-world problem isn't unique. With numerous alternatives being available to the modeler, which one is better? There's the usual trade-off between a "large," presumably accurate, but computationally costly model, and a "small," coarse model easily solved. There's also the issue of sensitivity to changes in model parameters, which is a main subject of this tutorial. Will our decision change significantly if we perturb the parameters slightly? This is a major concern because the "true" values of the parameters may not be known and one wouldn't like to be misled by a solution obtained using incorrect values. In fact, practitioners of optimization know very well that it isn't enough to obtain a solution and present it as *the* answer to a decision maker. One would need to explore the effect of changes to model parameters. Show that a recommended course of action isn't a modeling artifact. Convince the decision maker that the model is "valid" and produces reasonable results. A systematic approach addressing these issues is to consider a *family of optimization problems* constructed by changing parameters of concern. The resulting sensitivity analysis uncovers

troubling situations with unstable solutions and helps us to distinguish between "good" and "bad" models.

Viewing a model of interest as producing a family of optimization problems also gives rise to algorithmic approaches. It enables us to assess whether the model has computationally attractive properties and thus is good in that sense. Optimality conditions, multiplier vectors and relaxations stem from this perspective as well. Under ideal circumstances and when properly tuned, the relaxations may even furnish the minimum value for the actual model. The tuning of these relaxations is in of itself a process of optimization; it amounts to solving certain *dual* problems.

We organize our thinking about families of problems using the fundamental concept of a *Rockafellian,* which can be traced back to the middle of the last century with the pioneering work of Rockafellar [27], Gale [13] and others. Under the name "bifunction," the concept was formalized by Rockafellar in the seminal text [28, Chapter 29], extended to infinite dimensions in [29] and beyond the convex case in [30]. The name "Rockafellian" appears in [38], with "perturbation function" [40] and "bivariate function" [7] also being found in the literature.

1.1 Definition (Rockafellian). *For the problem of minimizing $\varphi : \mathbb{R}^n \to [-\infty, \infty]$, we say that $f : \mathbb{R}^m \times \mathbb{R}^n \to [-\infty, \infty]$ is a Rockafellian* with *anchor $\bar{u} \in \mathbb{R}^m$ if*

$$f(\bar{u}, x) = \varphi(x) \quad \forall x \in \mathbb{R}^n.$$

In a practical setting, we think of φ in the definition as the objective function produced by a particular modeling effort. Since the function is permitted to take the value ∞, it accounts for constraints implicitly as discussed further in the next section. Of course, a minimizer of φ is useful and indicates a possible course of action, but it alone fails to indicate the effect of changing parameter values. An associated Rockafellian f specifies explicitly the dependence on m parameters and defines the *family of problems*

$$\left\{ \minimize_{x \in \mathbb{R}^n} f(u, x), \quad u \in \mathbb{R}^m \right\},$$

among which the *actual problem* of minimizing φ emerges as

$$\minimize_{x \in \mathbb{R}^n} f(\bar{u}, x).$$

In this tutorial, we show that the expanded view involving the whole family of problems reveals concerning modeling assumptions underpinning the actual problem, quantifies the effect of changes in the perturbation vector u away from the anchor $\bar{u}$, furnishes optimality conditions for the problem with associated computational possibilities and defines relaxations as well as supporting dual problems.

Our scope covers a vast number of models. The actual problem can be convex or nonconvex, smooth or nonsmooth. It can involve constraints, including restrictions to integer values. It can be the result of a complicated modeling process that accounts for uncertainty in various ways. For example, φ might be an expectation of the form $\mathbb{E}[g(\boldsymbol{\xi}, x)]$ and then the perturbation vector u could specify modeling assumptions about the probability distribution of the random vector $\boldsymbol{\xi}$. We note, however, that the perturbation vector u rarely represents directly an inherently random quantity such as future product

2

demand or environmental condition, which one can't expect to know at the time of decision making, and rather it represents modeling assumptions about that quantity and other factors.

Sensitivity analysis of linear programs is well-known and the "economic interpretation" of dual variables in that context is often the crowning accomplishment of many introductory optimization courses. Besides the restriction to linear settings, the classical treatment suffers from a limited array of perturbations. Typically, one would only deal with changes on the right-hand side of the constraints, but one could imagine other possibilities as well. There's an extensive literature on more general, local stability results for optimization and variational problems examining metric regularity and calmness [39, 16, 25], tilt-stability [26, 12, 18, 11, 14], full-stability [23], connections with iterative schemes [17] and specifics of nonlinear programming [15]; see also the monographs [5, 34, 8, 20] and the surveys [24, 6]. Global sensitivity analysis emerges via the truncated Hausdorff distance in [3, 4, 35, 36]. A review of all these concepts goes beyond this tutorial and we focus instead on an introductory analysis based on epi-convergence and the Rockafellar condition for optimality as developed in [38], which in turn relies on [34]. Even this more limited scope allows us to address a vast array of applications and gain significant insight about strengths and weaknesses of a model.

The tutorial starts in the next section with basic notation and some motivating examples. Section 3 presents the epigraphical point of view, which is the foundation for subsequent developments. Section 4 applies the concepts to Rockafellians to determine whether a model is overly sensitive to parametric changes. Section 5 reviews the fundamental concepts of normal cones and subgradients, which are subsequently utilized in Section 6 to quantify local changes in minimum values. Optimality conditions emerging from Rockafellians are the subject of Section 7. The chapter ends with an extended discussion of algorithmic possibilities stemming from a Rockafellian including those associated with dual problems.

2 Notation and Examples

We often represent an optimization problem by just an objective function $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$, where the *extended real line*

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\} = [-\infty, \infty]$$

allows us to express constraints implicitly by setting $\varphi(x) = \infty$ for infeasible $x \in \mathbb{R}^n$. In this setting, the *domain* of φ, denoted by

$$\operatorname{dom} \varphi = \big\{ x \in \mathbb{R}^n \mid \varphi(x) < \infty \big\},$$

specifies the feasible set; decisions producing an objective function value of infinity are considered intolerable. If $\varphi(x) \in \mathbb{R}$ for all $x \in \mathbb{R}^n$, then φ is *real-valued* and this is also specified by writing $\varphi : \mathbb{R}^n \to \mathbb{R}$. Generally, the *minimum value* and *set of minimizers* of $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$ become

$$\inf \varphi = \inf \big\{ \varphi(x) \mid x \in \mathbb{R}^n \big\}, \qquad \operatorname{argmin} \varphi = \big\{ x \in \operatorname{dom} \varphi \mid \varphi(x) \le \inf \varphi \big\}.$$

We note that the pathological case with $\varphi(x) = \infty$ for all $x \in \mathbb{R}^n$ has $\inf \varphi = \infty$, which in fact is attained at all $x \in \mathbb{R}^n$. Still, the definition specifies that $\operatorname{argmin} \varphi = \emptyset$ in that case because $\operatorname{dom} \varphi = \emptyset$. This is meaningful because we don't want to consider a "cost" of infinity to be optimal. In fact, the

problem of minimizing φ is infeasible. We say that $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$ is *proper* if $\varphi(x) > -\infty$ for all $x \in \mathbb{R}^n$ and $\varphi(x) < \infty$ for some $x \in \mathbb{R}^n$. Thus, a proper objective function rules out an infeasible problem, but $\operatorname{argmin} \varphi$ could still be empty as the example $\varphi(x) = \exp(-x)$ illustrates. Near-minimizers exist more broadly. For $\varepsilon \geq 0$, the set of *near-minimizers* of φ is

$$\varepsilon\text{-}\operatorname{argmin} \varphi = \{x \in \operatorname{dom} \varphi \mid \varphi(x) \leq \inf \varphi + \varepsilon\}.$$

If $\varepsilon > 0$ and $\inf \varphi > -\infty$, then this set is nonempty as illustrated in Figure 1.
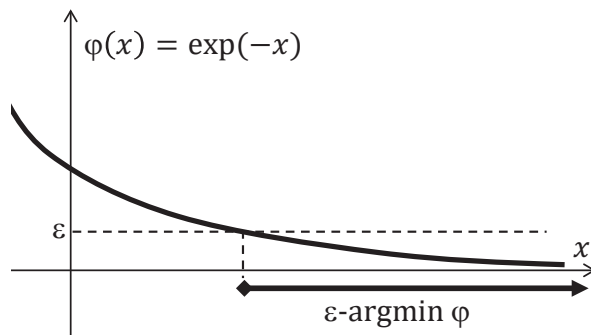

Figure 1: Example of near-minimizers when $\operatorname{argmin} \varphi$ is empty.

     While it can be useful to hide the details of a model by representing it by a single function, a more detailed analysis and computations should leverage the structural properties of the model. A traditional format is to specify an optimization problem using a real-valued objective function as well as a number of real-valued functions to define equality and inequality constraints. In addition to the glaring imbalance between the detailed specification of the feasible set in terms of *many* constraint functions and the lack of details about the objective function, the format fails to bring forth any "simple" constraints that should be treated differently than more complicated constraints. From a modeling point of view, there's also something artificial about singling out one function as objective and the rest as constraints. For example, when designing a system, we might seek a low cost and a low risk of poor performance. To insist on having a single objective function would force us to prioritize cost over risk or vice versa. Moreover, the quantity, being it cost or risk, that has been downgraded to a constraint is subject to a strict requirement of equaling or not exceeding a threshold. In reality, violations might be acceptable if appropriately penalized.

     To account for a variety of situations, we often specify optimization models using four components:

$$X \subset \mathbb{R}^n \qquad\qquad \text{(basic feasible set)}$$
$$f_0 : \mathbb{R}^n \to \mathbb{R} \qquad\qquad \text{(primary quantity of interest)}$$
$$F : \mathbb{R}^n \to \mathbb{R}^m \qquad\qquad (m \text{ secondary quantities of interest})$$
$$h : \mathbb{R}^m \to \overline{\mathbb{R}} \qquad\qquad \text{(monitoring function for secondary quantities)}$$

4

These components produce an optimization problem in *composite form*:

$$\underset{x \in X}{\text{minimize}} \ f_0(x) + h\big(F(x)\big). \tag{2.1}$$

For example, $f_0(x)$ might be the initial cost associated with decision x and $F(x) = (f_1(x), \ldots, f_m(x))$ could be subsequent costs under m different scenarios about future operating conditions each occurring with probability p_i. Then, it would be meaningful to set

$$h(z) = \sum_{i=1}^{m} p_i z_i, \quad \text{where } z = (z_1, \ldots, z_m),$$

so that the term $h(F(x))$ becomes the expected subsequent cost across the scenarios. In another setting, the secondary quantities might be subject to strict requirements expressed by

$$f_i(x) = 0, \ \ i = 1, \ldots, m$$

or, equivalently, $F(x) = 0$. These requirements are captured by setting

$$h(z) = \iota_{\{0\}^m}(z),$$

where ι_D (Greek letter iota) is the *indicator function* of the set D defined as

$$\iota_D(z) = \begin{cases} 0 & \text{for } z \in D \\ \infty & \text{otherwise.} \end{cases}$$

Thus, $F(x) = 0$ if and only if $\iota_{\{0\}^m}(F(x)) < \infty$. Since we always assume that $\alpha + \infty = \infty$ regardless of $\alpha \in \overline{\mathbb{R}}$, a decision x that fails to satisfy the equality requirements would cause $f_0(x) + h(F(x)) = \infty$, which indeed implies an infeasible decision. The composite form, especially in the context of certain monitoring functions, is also referred to as "extended nonlinear programming" [31].

Let's examine some additional examples in more detail.

2.1 Example (goal optimization). *Consider a situation with several quantities of interest (cost, risk, damage, etc.) and the goal of finding a decision that makes all these quantities low. This vague problem statement is a rather common take-away from a first meeting with a prospective client! Of course, there's little hope that one can find a decision that minimizes every quantity at the same time and the goal becomes to identify decisions that balance the various concerns. This leads to the broad area of multi-objective optimization [19]. One approach is then to identify for each quantity of interest a goal τ_i and a weight θ_i. If $f_i : \mathbb{R}^n \to \mathbb{R}$ specifies the values of the ith quantity of interest across the decision space, then we can adopt the model*

$$\underset{x \in X}{\text{minimize}} \ \sum_{i=1}^{m} \theta_i \max \big\{0, \ f_i(x) - \tau_i\big\}, \tag{2.2}$$

where X is a basic feasible set that restricts the decisions. This is referred to as goal optimization *and falls within the composite form of* (2.1).

Detail. Let $f_0(x) = 0$, $F(x) = (f_1(x), \ldots, f_m(x))$ and

$$h(z) = \sum_{i=1}^{m} \theta_i \max\{0, z_i - \tau_i\}$$

in (2.1) and we recover (2.2). The resulting model leads to decisions that are as "close" as possible to satisfying all the goals. We note that lowering $f_i(x)$ below the goal τ_i has no benefit, but any value above incurs a per-unit penalty of θ_i. ◻
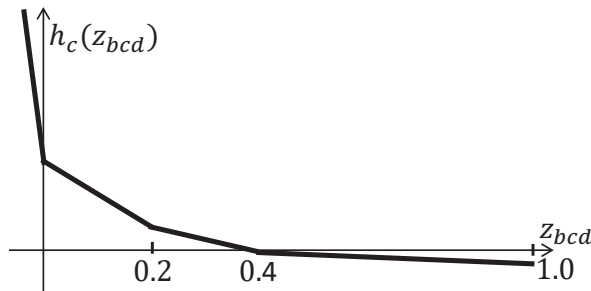

Figure 2: Monitoring function for inventory level.

2.2 Example (naval resupply at sea). *The US Navy operates worldwide without much need to enter ports for the purpose of resupply. This is achieved by a remarkably low number of transport ships that, as needed, rendezvous with the battle groups at sea. With their numerous moving "customers," the transport ships are faced with a major logistical challenge. Naval planners use an optimization model to coordinate this effort [9]. The model tracks the inventory of four commodities (dry stores and food, ship fuel, aircraft fuel and ordnance) for each battle group and one might have expected to find constraints of the kind "inventory of commodity c on day d for battle group b must be nonnegative." However, this isn't the case; transport ships are just encouraged to resupply through a system of penalties and rewards. US Navy ships are allowed to have negative inventory according to the model! This isn't a result of poor modeling. In fact, as we see in this tutorial, there's much merit to the approach. Penalties turn out to produce good models, while (hard) constraints are problematic and easily lead to bad models.*

Detail. The model in [9] specifies the actions (where to sail, when to rendezvous, what to offload, etc.) for all the transport ships across a planning horizon of up to 180 days. Suppose that the vector x specifies all these actions as well as auxiliary variables and

$f_{bcd}(x)$ is the inventory of commodity c on day d for battle group b under decision x,

which can be determined based on (assumed) known consumption rates. The decision x is constrained by limitations of the transport ships and many other concerns, which we here simply specify by the feasible set X. The feasible set, however, doesn't prevent negative values of $f_{bcd}(x)$. The objective

function only encourages high values of $f_{bcd}(x)$. Omitting some details, the model takes the composite form (2.1) with $f_0(x) = 0$ and

$$F(x) = \Big(f_{bcd}(x), \ \forall b, c, d\Big) \qquad h(z) = \sum\nolimits_{b,d,c} h_c(z_{bcd}),$$

where z is the vector with components z_{bcd} and $h_c : \mathbb{R} \to \mathbb{R}$ is a monitoring function, which might depend on the commodity c; see Figure 2. An increasingly large penalty is invoked as the inventory level decreases, while a reward is associated with high levels (above 40% capacity in Figure 2); an inventory above 100% capacity of the battle group is prevented by constraints in X. ∎

Our assessment of optimization models is carried out in the context of certain perturbations as defined by a Rockafellian. For a problem of the composite form (2.1), one might consider a Rockafellian $f : \mathbb{R}^m \times \mathbb{R}^n \to \overline{\mathbb{R}}$ given by

$$f(u, x) = \iota_X(x) + f_0(x) + h\big(F(x) + u\big),$$

which then examines the effect of changes to the secondary quantities of interest. Here, the anchor is $\bar{u} = 0$ so the actual problem is recovered by minimizing $f(0, x)$ over $x \in \mathbb{R}^n$. For example, if $h(z) = \iota_{(-\infty,0]^m}(z)$, then the Rockafellian specifies a change from the constraint $F(x) \leq 0$ to $F(x) + u \leq 0$. This corresponds to the typical "right-hand side" perturbation of linear programming sensitivity analysis and captures changes to resource budgets. For goal optimization, with h as in Example 2.1, this Rockafellian represents a change in goal for the ith quantity of interest from τ_i to $\tau_i - u_i$.

Any given model can be associated with many different Rockafellians, reflecting concerns about various parameters. The model would be considered good relative to a Rockafellian $f : \mathbb{R}^m \times \mathbb{R}^n \to \overline{\mathbb{R}}$ with anchor $\bar{u}$ if the Rockafellian exhibits certain desirable properties such as

minimizers and minimum value of $f(u, \cdot)$ tend to those of $f(\bar{u}, \cdot)$ as $u \to \bar{u}$.

We omit a formal definition and instead give a simple example illustrating that these convergence properties are far from automatically satisfied.
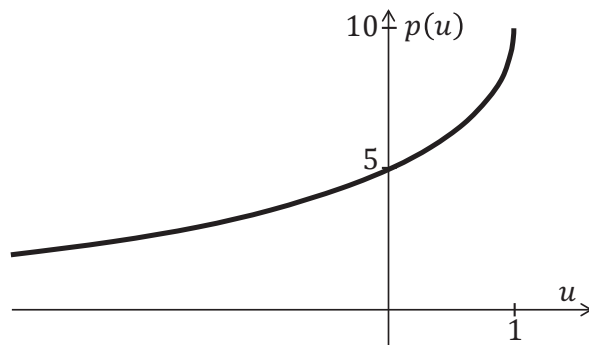

Figure 3: The minimum value $p(u) = \inf f(u, \cdot)$ as function of u in Example 2.3.

2.3 Example (constraint perturbation). *The problem of minimizing $x^2 + 1$ subject to $(x-2)(x-4)+1$ ≤ 0 can be associated with a Rockafellian defined by*

$$f(u, x) = x^2 + 1 + \iota_{(-\infty, 0]}(g(x) + u) \quad \text{and} \quad g(x) = (x-2)(x-4),$$

with $\bar{u} = 1$ as its anchor. Let $p(u) = \inf f(u, \cdot)$ so that $p(1)$ becomes the minimum value of the actual problem. It turns out that $p(u)$ doesn't tend to $p(1)$ if u approaches 1 from above. In fact, $p(u) = \infty$ for $u > 1$, while $p(1) = 10$. Thus, the Rockafellian highlights the sensitivity to changes on the right-hand side of the constraint in the actual problem. This would have raised concerns about the validity of the model had it represented a real-world situation with constraint parameters being subject to modeling assumptions.

Detail. The feasible set $\operatorname{dom} f(u, \cdot)$ is given by the constraint $(x-2)(x-4) + u \leq 0$. A simple application of the quadratic equation produces

$$\operatorname{dom} f(u, \cdot) = \begin{cases} \left[\, 3 - \sqrt{1-u}, \;\; 3 + \sqrt{1-u} \,\right] & \text{if } u \leq 1 \\ \emptyset & \text{otherwise.} \end{cases}$$

As functions of u, the minimizers and minimum values become

$$\operatorname{argmin} f(u, \cdot) = \begin{cases} \{0\} & \text{if } u < -8 \\ \{3 - \sqrt{1-u}\} & \text{if } -8 \leq u \leq 1 \\ \emptyset & \text{otherwise} \end{cases}$$

$$p(u) = \inf f(u, \cdot) = \begin{cases} 1 & \text{if } u < -8 \\ 11 - u - 6\sqrt{1-u} & \text{if } -8 \leq u \leq 1 \\ \infty & \text{otherwise.} \end{cases}$$

At every $u < 1$, p is continuous. However, at the anchor $\bar{u} = 1$, the minimum value makes a jump from 10 to ∞; see Figure 3. $\qquad \blacksquare$

3 Epigraphs

The key concept for understanding changes to minimizers and minimum values as well as several other aspects is that of an epigraph. For a function $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$, its *epigraph* is defined as

$$\operatorname{epi} \varphi = \big\{ (x, \alpha) \in \mathbb{R}^n \times \mathbb{R} \,\big|\, \varphi(x) \leq \alpha \big\}.$$

In contrast to the graph of φ, which consists of the points (x, α) with $\varphi(x) = \alpha$, the epigraph captures the crucial distinction between $\varphi(x) = \infty$ and $\varphi(x) = -\infty$ in a manner that's tailored to minimization problems. The former value designates x as infeasible, while the latter value specifies that x must be a minimizer of φ. Figure 4 shows the epigraphs of $f(0, \cdot)$, $f(3/4, \cdot)$ and $f(1, \cdot)$ as defined in Example 2.3. These epigraphs appear to converge in the sense that the "wedge" $\operatorname{epi} f(u, \cdot)$ approaches the vertical
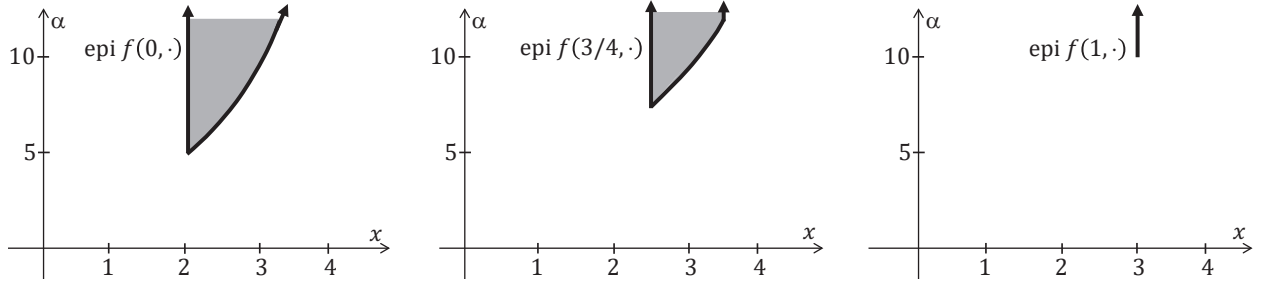
Figure 4: Epigraphs of $f(0,\cdot)$, $f(3/4,\cdot)$ and $f(1,\cdot)$ in Example 2.3. Arrows indicate that the epigraphs extend upward indefinitely.

line segment epi $f(1,\cdot)$ as $u \to 1$ from below. Moreover, the corresponding minimizers and minimum values converge as already established in Example 2.3. This isn't a coincidence. It turns out that convergence of epigraphs essentially guarantees convergence of minimizers and minimum values.

To make precise the meaning of epigraphs converging, we define the *point-to-set distance* between $\bar{x} \in \mathbb{R}^n$ and $C \subset \mathbb{R}^n$ as the distance between $\bar{x}$ and the closest point in C. Specifically,

$$\text{dist}(\bar{x}, C) = \inf_{x \in C} \|x - \bar{x}\|_2 \quad \text{when} \quad C \neq \emptyset \quad \text{and} \quad \text{dist}(\bar{x}, \emptyset) = \infty.$$

In the following, we use superscript ν (Greek nu) to index elements in a sequence, which then runs over the natural numbers $\mathbb{N} = \{1, 2, \dots\}$.

3.1 Definition (epi-convergence). *For functions $\{\varphi, \varphi^\nu : \mathbb{R}^n \to \overline{\mathbb{R}}, \ \nu \in \mathbb{N}\}$, we say that φ^ν epi-converges to φ, written $\varphi^\nu \xrightarrow{e} \varphi$, when*

$$\text{epi}\,\varphi \ \text{is a closed set} \quad \text{and} \quad \text{dist}(z, \text{epi}\,\varphi^\nu) \to \text{dist}(z, \text{epi}\,\varphi) \ \ \forall z \in \mathbb{R}^{n+1}.$$

Figure 5 illustrates epi-convergence in the context of Example 2.3: For any point $z \in \mathbb{R}^2$,

$$\text{dist}\big(z, \text{epi}(f(u,\cdot))\big) \to \text{dist}\big(z, \text{epi}(f(1,\cdot))\big)$$

when $u \to 1$ from below. The figure visualizes the situation with $z = (1, 5)$, in which case one obtains

$$\text{dist}\big(z, \text{epi}\big(f(u,\cdot)\big)\big) = \sqrt{\big(2 - \sqrt{1-u}\big)^2 + \big(6 - u - 6\sqrt{1-u}\big)^2} \quad \text{for } u \in [-8, 1].$$

Certainly, the distance tends to $\sqrt{29}$ as $u \to 1$ from below and this is indeed the distance from z to epi $f(1,\cdot)$. The situation is similar for other z.

The next theorem summarizes the main consequences of epi-convergence; see [38, Theorems 4.14, 5.5] for a proof and further details.

3.2 Theorem (epigraphical approximations). *Suppose that the functions $\varphi^\nu : \mathbb{R}^n \to \overline{\mathbb{R}}$ epi-converge to a proper function $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$. Then, the following hold:*
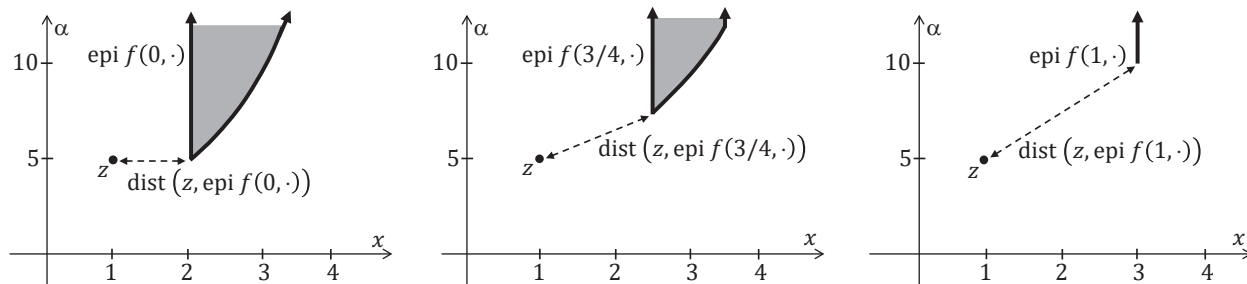
9

Figure 5: Epi-convergence in Example 2.3 as indicated by the convergence of the distance from z to the epigraphs.

(a) *Every cluster point of a sequence $\{x^\nu \in \varepsilon^\nu\text{-}\operatorname{argmin}\varphi^\nu, \ \nu \in \mathbb{N}\}$, with $\varepsilon^\nu \to 0$, is a minimizer of φ. Thus, in particular, if x^ν converges to $\bar{x}$ and each x^ν is a minimizer of φ^ν, then $\bar{x}$ is a minimizer of φ.*

(b) *Every cluster point $\bar{x}$ of a sequence $\{x^\nu, \ \nu \in \mathbb{N}\}$, with $\varphi^\nu(x^\nu) \leq \alpha^\nu$ and $\alpha^\nu \to \alpha$, satisfies $\varphi(\bar{x}) \leq \alpha$.*

(c) *If there's a compact set B such that $B \cap \operatorname{argmin}\varphi^\nu$ is nonempty for all ν, then $\inf \varphi^\nu \to \inf \varphi$.*

We observe that the theorem imposes essentially no assumptions on the functions besides their epi-convergence; convexity and/or smoothness aren't required. Item (b) furnishes the guarantee that a point with a low function value according to the approximation φ^ν also has a low value according to the actual function φ.

The following proposition is often helpful in establishing epi-convergence; see [38, Theorem 4.15]. Occasionally, we abbreviate "*there exist(s)*" by the symbol $\exists$.

3.3 Proposition (characterization of epi-convergence). *For $\varphi, \varphi^\nu : \mathbb{R}^n \to \overline{\mathbb{R}}$, $\varphi^\nu \xrightarrow{e} \varphi$ if and only if the following hold at each $x \in \mathbb{R}^n$:*

(a) *$\forall x^\nu \to x$, one has $\liminf \varphi^\nu(x^\nu) \geq \varphi(x)$*

(b) *$\exists x^\nu \to x$ such that $\limsup \varphi^\nu(x^\nu) \leq \varphi(x)$.*

4 Stability

Let's now bring in the concept of epi-convergence to examine whether an optimization model is good in the sense that, relative to a particular Rockafellian, minimizers and minimum values change continuously as model parameters are perturbed. The following theorem, which is an immediate consequence of Theorem 3.2, shows that the key property to check is epi-convergence of the Rockafellian.

4.1 Theorem (stability). *Suppose that $f : \mathbb{R}^m \times \mathbb{R}^n \to \overline{\mathbb{R}}$ is a Rockafellian with anchor $\bar{u} \in \mathbb{R}^m$ for the problem*

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \ \varphi(x).$$

10

Let $p(u) = \inf f(u, \cdot)$ and $P(u) = \operatorname{argmin} f(u, \cdot)$, i.e., the minimum values and minimizers of $f(u, \cdot)$ viewed as functions of u. If $f(\bar{u}, \cdot)$ is proper and, for any $u^\nu \to \bar{u}$, one has $f(u^\nu, \cdot) \xrightarrow{e} f(\bar{u}, \cdot)$ as well as a compact set $B \subset \mathbb{R}^n$ such that $B \cap \operatorname{argmin} f(u^\nu, \cdot)$ is nonempty for all ν, then the following hold:

(a) p is a continuous function at $\bar{u}$, i.e., $u^\nu \to \bar{u}$ implies that $p(u^\nu) \to \inf \varphi$

(b) P is outer semicontinuous at $\bar{u}$, i.e., $u^\nu \to \bar{u}$ implies that any cluster point $\bar{x}$ of $\{x^\nu \in P(u^\nu), \nu \in \mathbb{N}\}$ satisfies $\bar{x} \in \operatorname{argmin} \varphi$.

The stability property in the theorem is typically desirable and one might seek to develop models that satisfy the stated assumptions. Let's examine several common cases.

4.2 Example (regularization as perturbation). *In data analytics and algorithms relying on proximal point methods, an optimization problem of minimizing a continuous function $f_0 : \mathbb{R}^n \to \mathbb{R}$ over a nonempty closed set X is often modified by a regularization term involving a penalty parameter. The penalty parameter is rarely fixed. What is the effect of changing it? Specifically, the actual problem is*

$$\underset{x \in X}{\text{minimize }} f_0(x) + \theta_0 r(x),$$

where $\theta_0 \in [0, \infty)$ is the penalty parameter and $r : \mathbb{R}^n \to [0, \infty)$ is a continuous function; for instance $r(x) = \|x\|_1$ as in lasso regression. With the value of the penalty parameter in focus, we define a Rockafellian $f : \mathbb{R} \times \mathbb{R}^n \to \overline{\mathbb{R}}$ for the problem as

$$f(u, x) = \iota_X(x) + f_0(x) + \theta(u) r(x),$$

where $\theta(u)$ tends to $\theta(0) = \theta_0$ as $u \to 0$. It turns out that regardless of $\theta_0 \in [0, \infty)$, the actual problem is stable in the sense of Theorem 4.1 under changes to the penalty parameter as long as a boundedness assumption holds.

Detail. Let $u^\nu \to 0$. We use Proposition 3.3 to confirm that $f(u^\nu, \cdot) \xrightarrow{e} f(0, \cdot)$. For (a) in the proposition, let $x^\nu \to x$. We consider two cases. If $x \in X$, then $\iota_X(x) = 0$ and

$$\liminf f(u^\nu, x^\nu) = \liminf \big(\iota_X(x^\nu) + f_0(x^\nu) + \theta(u^\nu) r(x^\nu) \big)$$
$$\geq \liminf \big(f_0(x^\nu) + \theta(u^\nu) r(x^\nu) \big) = f_0(x) + \theta(0) r(x) = f(0, x).$$

If $x \notin X$, then $\iota_X(x) = \infty$ and $x^\nu \notin X$ for all ν sufficiently large because X is closed. Thus, $\liminf f(u^\nu, x^\nu) = \infty$, which matches $f(0, x)$. For (b) in the proposition, let x be arbitrary and take $x^\nu = x$ for all ν. Then,

$$\limsup f(u^\nu, x^\nu) = \limsup \big(\iota_X(x) + f_0(x) + \theta(u^\nu) r(x) \big) = \iota_X(x) + f_0(x) + \theta(0) r(x) = f(u, x)$$

and we conclude that $f(u^\nu, \cdot) \xrightarrow{e} f(0, \cdot)$. Since $f(\bar{u}, \cdot)$ is proper in view of the nonemptyness of X, Theorem 4.1 applies as long as one can determine a compact set B that "reaches" at least some minimizers in $\operatorname{argmin} f(u^\nu, \cdot)$ regardless of ν. If X is compact, then this is trivially the case but many other possibilities exist as well. $\square$

4.3 Example (risk-averse decision making). *Suppose that we would like to select a decision x from a nonempty compact set $X \subset \mathbb{R}^n$ such that a cost $g(\xi, x)$ is minimized, where $g : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$. The difficulty is that the cost depends on ξ, which is uncertain and can take values in a finite set $\Xi \subset \mathbb{R}^m$ each occurring with probability p_ξ. One modeling possibility is to select x such that the cost across the worst $(1 - \alpha)100\%$ outcomes is minimized, where $\alpha \in (0, 1)$ conveys the level of risk averseness. If α is near 1, then x is selected such that the very worst outcomes are made less costly as much as possible. If α is near 0, then the average cost governs the decision. For $\bar{\alpha} \in (0, 1)$, these modeling choices result in the following problem*

$$\underset{x \in X, \gamma \in \mathbb{R}}{\text{minimize}} \ \gamma + \frac{1}{1 - \bar{\alpha}} \sum_{\xi \in \Xi} p_\xi \max \{0, g(\xi, x) - \gamma\};$$

see [38, Section 3.C] for further details about such superquantile minimization as well as [32, 33] for applications in engineering design and in finance, where it's called CVaR minimization. In practice, it's often difficult to know the right level of risk averseness. Under the assumption that $g(\xi, \cdot)$ is continuous for all $\xi \in \Xi$, it turns out that the minimum value for the problem changes continuously as α varies in $(0, 1)$. In this sense, the underlying model is good.

Detail. Let's express $\alpha = e^u/(1 + e^u)$, with $u \in \mathbb{R}$, and adopt a Rockafellian of the form

$$f\big(u, (\gamma, x)\big) = \iota_X(x) + \gamma + (1 + e^u) \sum_{\xi \in \Xi} p_\xi \max \{0, g(\xi, x) - \gamma\}.$$

The actual problem is recovered by minimizing $f(\bar{u}, \cdot)$, where $\bar{u} = \ln(\bar{\alpha}/(1 - \bar{\alpha}))$ is the anchor. Let's check the assumptions of the Stability Theorem 4.1. Trivially, $f(\bar{u}, \cdot)$ is proper because X is nonempty. The required epi-convergence can be established using Proposition 3.3 as in the previous example and we omit the details.

Let $u^\nu \to \bar{u}$. The existence of a compact set $B \subset \mathbb{R}^{1+n}$ such that $B \cap \text{argmin} f(u^\nu, \cdot)$ is nonempty for all ν is confirmed as follows. For $x^\nu \in X$ and $\gamma^\nu \to \infty$, $f(u^\nu, (\gamma^\nu, x^\nu)) \to \infty$ because the term inside the summation is nonnegative. For $x^\nu \in X$ and $\gamma^\nu \to -\infty$, $\min_{\xi \in \Xi, \nu \in \mathbb{N}} g(\xi, x^\nu)$ is bounded from below because Ξ is a finite set, X is compact and $g(\xi, \cdot)$ is continuous. Thus,

$$\gamma^\nu + (1 + e^{u^\nu}) \sum_{\xi \in \Xi} p_\xi \max \{0, g(\xi, x^\nu) - \gamma^\nu\}$$

involves a first term tending to $-\infty$ and a second term tending to ∞, with the second term overpowering the first one due to the coefficient $(1 + e^{u^\nu})$, which is greater than one. So, in this case as well, $f(u^\nu, (\gamma^\nu, x^\nu)) \to \infty$. We've shown that the distance from the origin to $\text{argmin} f(u^\nu, \cdot)$ can't become arbitrarily large. □

4.4 Example (composite form). *Let's consider a problem of the composite form (2.1) with a nonempty closed set $X \subset \mathbb{R}^n$, continuous $f_0 : \mathbb{R}^n \to \mathbb{R}$, real-valued $h : \mathbb{R}^m \to \mathbb{R}$ and continuous $F : \mathbb{R}^n \to \mathbb{R}^m$. We're concerned about changes to F and this leads to a Rockafellian $f : \mathbb{R}^m \times \mathbb{R}^n \to \overline{\mathbb{R}}$ defined as*

$$f(u, x) = \iota_X(x) + f_0(x) + h\big(F(x) + u\big),$$

12

with anchor 0. Under a boundedness assumption, the Stability Theorem 4.1 applies and the underlying model is then good in the sense that minimizers and minimum values vary continuously as specified in the theorem. Since the naval resupply problem in Example 2.2 satisfies these assumptions, we conclude that the resulting logistical plans are actually rather stable relative to changes in the inventory levels and this can be part of the reason the model has been accepted by US Navy planners.

Detail. Since X is nonempty and h is real-valued, $f(0, \cdot)$ is proper. Let $u^\nu \to 0$. We use Proposition 3.3 to confirm that $f(u^\nu, \cdot) \xrightarrow{e} f(0, \cdot)$. For part (a) of that proposition, let $x^\nu \to x$. We consider two cases. If $x \in X$, then $\iota_X(x) = 0$ and

$$\liminf f(u^\nu, x^\nu) = \liminf \left(\iota_X(x^\nu) + f_0(x^\nu) + h\big(F(x^\nu) + u^\nu\big) \right)$$
$$\geq \liminf \left(f_0(x^\nu) + h\big(F(x^\nu) + u^\nu\big) \right) = f_0(x) + h\big(F(x)\big) = f(0, x)$$

by the continuity of f_0, h and F. If $x \notin X$, then $\iota_X(x) = \infty$ and, because X is closed, $x^\nu \notin X$ as well for all ν sufficiently large. Thus, $\liminf f(u^\nu, x^\nu) = \infty$, which matches $f(0, x)$. For (b) in the proposition, let x be arbitrary and take $x^\nu = x$ for all ν. Then,

$$\limsup f(u^\nu, x^\nu) = \limsup \left(\iota_X(x) + f_0(x) + h\big(F(x) + u^\nu\big) \right) = \iota_X(x) + f_0(x) + h\big(F(x)\big)$$

again using the continuity of h. We've established that $f(u^\nu, \cdot) \xrightarrow{e} f(0, \cdot)$. The existence of a compact set X as required by the Stability Theorem 4.1 needs to be verified separately, but this is trivial if X is compact and it holds in other cases as well.

The above analysis relies heavily on the fact that the monitoring function h is real-valued and continuous. While this is acceptable when examining goal optimization (Example 2.1) and naval resupply (Example 2.2), it would not hold in the case of (hard) constraints subject to right-hand side perturbations as in Example 2.3. In fact, that example shows one can't extend the present results without additional assumptions. This highlights the fundamental advantage of modeling requirements using penalties instead of constraints: under perturbations of the underlying functions, penalties produce good models while constraints might not. $\qquad\square$

4.5 Example (penalty approach). *Suppose that the actual problem of interest is*

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \; \varphi(x) = \iota_X(x) + f_0(x) + \iota_{(-\infty, 0]^m}\big(F(x)\big), \tag{4.1}$$

where X is nonempty closed and both $f_0 : \mathbb{R}^n \to \mathbb{R}$ and $F : \mathbb{R}^n \to \mathbb{R}^m$ are continuous. As discussed in Examples 2.3 and 4.4, the underlying model here can be bad in the sense that small changes in $F(x)$ might cause large changes in the minimizers and minimum values. It might be prudent to adjust the modeling of the secondary quantities of interest and adopt the alternative problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \; \iota_X(x) + f_0(x) + \theta \sum_{i=1}^m \max\big\{0, f_i(x)\big\}, \tag{4.2}$$

13

where $\theta \in [0, \infty)$ is a penalty parameter and $F(x) = (f_1(x), \ldots, f_m(x))$. For any fixed θ, the alternative problem is a special case of that in Example 4.4 and therefore comes with the desirable stability properties. However, we can go further and make connections with (4.1) as well.

Detail. Let's consider the Rockafellian given by

$$f(u, x) = \iota_X(x) + f_0(x) + \theta(u) \sum_{i=1}^{m} \max\{0, f_i(x) + u_i\}$$

associated with (4.2), where the penalty parameter is also made dependent on u. If $u^\nu \to 0$, $\theta(u^\nu) \to \infty$ and

$$\theta(u^\nu) \max\{0, \max_{i=1,\ldots,m} u_i^\nu\} \to 0 \quad \text{as } \nu \to \infty, \tag{4.3}$$

then $f(u^\nu, \cdot) \xrightarrow{e} \varphi$. Thus, Proposition 3.2 can be brought in to verify that minimizers and minimum values of a perturbed version of (4.2) converge not only to those of (4.2) but also to those of the *actual* problem.

To verify the claimed epi-convergence we bring in Proposition 3.3. If $x^\nu \to x$, then liminf $f(u^\nu, x^\nu) \geq \varphi(x)$ by arguments similar to those in Example 4.4. Let $x^\nu = x$ and consider three cases. (i) If $x \notin X$, then $\iota_X(x) = \infty$ and limsup $f(u^\nu, x^\nu) = \varphi(x) = \infty$. (ii) If $x \in X$ and $f_i(x) > 0$ for some i, then $\varphi(x) = \infty$ and limsup $f(u^\nu, x^\nu)$ can't exceed that value. (iii) If $x \in X$ and $f_i(x) \leq 0$ for all i, then

$$\text{limsup } f(u^\nu, x^\nu) = f_0(x) + \sum_{i=1}^{m} \text{limsup}\left(\theta(u^\nu) \max\{0, f_i(x) + u_i^\nu\}\right)$$

$$\leq f_0(x) + \sum_{i=1}^{m} \text{limsup}\left(\theta(u^\nu) \max\{0, u_i^\nu\}\right) = f_0(x) = \varphi(x)$$

and the claim about epi-convergence follows from Proposition 3.3.

We note that the requirement (4.3) has no ramification if u_i^ν approaches zero from below. However, if u_i^ν is positive, then one has to make sure that the penalty parameter $\theta(u^\nu)$ grows sufficiently slowly so that the product with u_i^ν vanishes. $\qquad\blacksquare$

5 Normal Cones and Subgradients

While continuity of minimum values as u varies provides indications of a good model, the property fails to quantify how fast the values might change. This raises the question of gradients of minimum value functions, but these might not be defined due to nonsmoothness. Moreover, to develop optimality conditions in a general setting we also need to consider nonsmooth functions; see for example the kinks in Figure 2. To address these challenges, we turn to subgradients and supporting normal cones.

5.1 Definition (normal cone). *For $\bar{x} \in C \subset \mathbb{R}^n$, a vector $v \in \mathbb{R}^n$ is normal to C at $\bar{x}$ in the regular sense, or simply a regular normal, if* [1]

$$\langle v, x - \bar{x} \rangle \leq o(\|x - \bar{x}\|_2) \quad \text{for } x \in C.$$

---

[1] The term $o(\|x - \bar{x}\|_2)$ for $x \in C$ has the property that $o(\|x - \bar{x}\|_2)/\|x - \bar{x}\|_2 \to 0$ when $x \in C \to \bar{x}$ with $x \neq \bar{x}$.

The set of all such regular normal vectors is denoted by $\widehat{N}_C(\bar{x})$.

A vector $v \in \mathbb{R}^n$ is normal to C at $\bar{x}$ in the general sense, or simply a normal, if

$$v^\nu \to v \ \text{ for some } \ v^\nu \in \widehat{N}_C(x^\nu) \ \text{ and } \ x^\nu \in C \to \bar{x}.$$

The set of all such normal vectors is $N_C(\bar{x})$, the normal cone of C at $\bar{x}$.

Figure 6 illustrates some common situations. On the left, where the boundary of the set C is smooth at $\bar{x}$, the regular normal vectors all point in the same direction; recall that $\langle v, x - \bar{x} \rangle$ is nonpositive when the angle between v and $x - \bar{x}$ is at least 90 degrees. At the corner point x', the regular normal vectors fan out as it becomes easier to form an angle of at least 90 degrees with points in C.
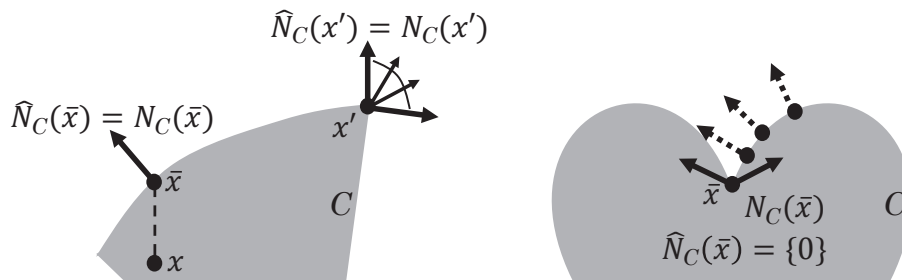

Figure 6: Normal vectors and normal cones.

On their own, regular normal vectors fail to provide a solid basis for the treatment of "complicated" sets with inward kinks and other irregularities as seen on the right in Figure 6. At the inward kink, $\widehat{N}_C(\bar{x}) = \{0\}$ since the zero vector is the only vector v that achieves nonpositive $\langle v, x - \bar{x} \rangle$ for x in C even locally. We achieve a robust notion of "normality" by considering the situation at points *near* $\bar{x}$ and this leads to the enrichment of vectors in $N_C(\bar{x})$ that aren't regular normals. Figure 6 shows as dashed arrows regular normal vectors at nearby points approaching $\bar{x}$ from the right. In the limit, these vectors give rise to the normal vectors at $\bar{x}$ pointing northwest. Likewise, points approaching $\bar{x}$ from the left result in the normal vectors at $\bar{x}$ pointing northeast. In all these illustrations, we think of normal vectors at a point in the sense of "floating arrows" and place them relative to that point in C. One could just as well visualize normal vectors as forming a cone emanating from the origin.

While normal vectors are important in expressions of optimality conditions, they also define subgradients for arbitrary functions via their epigraphs.

5.2 Definition (subgradients). *For $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $\bar{x}$ with $\varphi(\bar{x})$ finite, a vector v is a subgradient of φ at $\bar{x}$ if*

$$(v, -1) \in N_{\text{epi} \, \varphi}\big(\bar{x}, \varphi(\bar{x})\big).$$

The set of all subgradients of φ at $\bar{x}$ is denoted by $\partial \varphi(\bar{x})$.

Figure 7 illustrates the normal cones of epi φ at four different points. Near x_3, the function φ varies smoothly and the normal vectors at $(x_3, \varphi(x_3))$ all point in the same direction, which gives just one
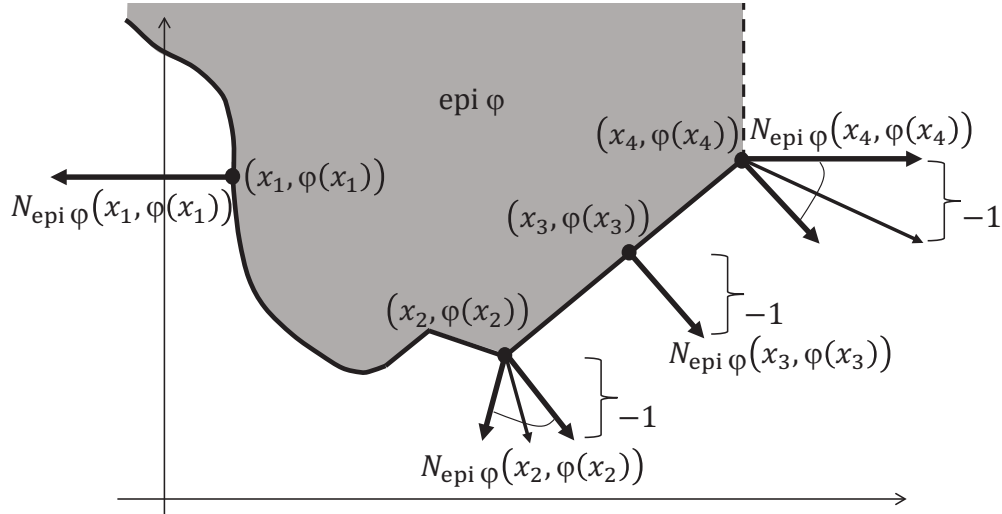
Figure 7: Normal vectors $(v, -1)$ of epigraphs produce subgradients v.

vector of the form $(v, -1)$. Thus, there's just one subgradient and it's given by this v. The value of this v appears to be roughly 1, which means that $\partial\varphi(x_3) = \{1\}$. Of course, this coincides with the slope of φ at x_3. This is confirmed by the first part of the following proposition; under smoothness, subgradients reduce to gradients. A proof is immediate from the definitions.

5.3 Proposition (subgradients and gradients). *For $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$ and a point $\bar{x}$, suppose that φ is smooth in a neighborhood of $\bar{x}$. Then,*

$$\partial\varphi(\bar{x}) = \{\nabla\varphi(\bar{x})\}.$$

At $(x_2, \varphi(x_2))$ in Figure 7, the epigraph of φ has a kink and the normal vectors fan out in many directions. The vectors of the form $(v, -1)$ in the normal cone appear to have $v \in [-1/2, 1]$, which makes this interval the set of subgradients of φ at x_2, i.e., $\partial\varphi(x_2) = [-1/2, 1]$. The values $-1/2$ and 1 correspond to the slope of φ on the left-side and the right-side of x_2, respectively.

At points to the right of x_4, the value of φ jumps to infinity. This produces a normal cone of epi φ at $(x_4, \varphi(x_4))$ that also includes "horizontal" vectors, which aren't of the form $(v, -1)$. Still, any normal vector that "tilts down" a bit can be made sufficiently long so it becomes in the form $(v, -1)$. Thus, $\partial\varphi(x_4) = [1, \infty)$.

At x_1, we have the peculiar behavior that epi φ is "vertical" and all the normal vectors are of the "horizontal" kind. Thus, there aren't any normal vectors of the form $(v, -1)$ and $\partial\varphi(x_1) = \emptyset$.

As in the case of differential calculus, there's an extensive set of rules that allow us to easily compute subgradients in many practical situations; we refer to [38, Section 4.I] for details and simply summarize two key facts here.

5.4 Proposition (sum rule). *For $\varphi_1, \varphi_2 : \mathbb{R}^n \to \overline{\mathbb{R}}$ and a point $\bar{x}$, suppose that φ_1 is smooth in a*

16

neighborhood of $\bar{x}$ and φ_2 is finite at $\bar{x}$. If φ is the function given by $\varphi(x) = \varphi_1(x) + \varphi_2(x)$, then

$$\partial\varphi(\bar{x}) = \nabla\varphi_1(\bar{x}) + \partial\varphi_2(\bar{x}).$$

5.5 Proposition (indicator function). *For $X \subset \mathbb{R}^n$ and $\bar{x} \in X$, one has*

$$\partial\iota_X(\bar{x}) = N_X(\bar{x}).$$

By combining these proposition, we see that the function given by $\varphi(x) = f_0(x) + \iota_X(x)$ has

$$\partial\varphi(x) = \nabla f_0(x) + N_X(x)$$

at any point $x \in X$ provided that f_0 is smooth. Regardless of the details of a function, we obtain the following fundamental optimality condition; see [38, Theorem 4.73].

5.6 Theorem (Fermat Rule for optimality). *For $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $x^\star$ with $\varphi(x^\star) \in \mathbb{R}$, one has*

$$x^\star \text{ local minimizer of } \varphi \implies 0 \in \partial\varphi(x^\star).$$

In Figure 7, $0 \in \partial\varphi(x_2) = [-1/2, 1]$, which means x_2 satisfies the Fermat Rule for optimality. This also illustrates the fact that while a function might be nonsmooth at just "a few" points, it tends to occur where we most care: at local minimizers. Consequently, we can't simply ignore the kinks in a function.

The Fermat Rule and our definition of subgradients apply to general functions. Convexity or smoothness isn't needed. In the case of a convex function $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$, Definition 5.2 coincides with the subgradients from convex analysis, which are often stated as those $v \in \mathbb{R}^n$ with

$$\varphi(x) \geq \varphi(\bar{x}) + \langle v, x - \bar{x}\rangle \qquad \forall x \in \mathbb{R}^n. \tag{5.1}$$

In the literature, our subgradients are sometimes referred to as "general," "limiting" or "Mordukhovich" subgradients; see the commentary of [34] for further details and explanation of why they have emerged as more central than "Clarke subgradients."

A special situation occurs when a function becomes "infinitely steep" or jumps to infinity at a point as takes place at x_1 and x_4 in Figure 7. The "horizontal" normal vectors of the epigraph in such cases are important in dealing with constraint qualifications and related issues, and are given a specific name.

We recall that a function $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$ is *lower semicontinuous (lsc)* if epi φ is a closed set.

5.7 Definition (horizon subgradients). *For a lsc function $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$ and a point $\bar{x}$ with $\varphi(\bar{x})$ finite, a vector v is a* horizon subgradient *of φ at $\bar{x}$ if*

$$(v, 0) \in N_{\mathrm{epi}\,\varphi}\big(\bar{x}, \varphi(\bar{x})\big).$$

The set of all horizon subgradients of φ at $\bar{x}$ is denoted by $\partial^\infty\varphi(\bar{x})$.

In Figure 7, $\partial^\infty f(x_1) = (-\infty, 0]$ and $\partial^\infty f(x_4) = [0, \infty)$. At the points x_2 and x_3, there aren't any "horizontal" normal vectors and thus $\partial^\infty \varphi(x_2) = \partial^\infty \varphi(x_3) = \{0\}$. The situation at x_3 can be settled by the following general fact, which is a direct consequence of the definitions.

5.8 Proposition (horizon subgradients for smooth functions). *If $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$ is smooth in a neighborhood of $\bar{x}$, then $\partial^\infty \varphi(\bar{x}) = \{0\}$.*

Just as the case for subgradients, horizon subgradients can often be easily computed using a series of calculation rules. We recall two main facts; see [38, Section 4.I] for further details.

5.9 Proposition (expressions for horizon subgradients). *If $X \subset \mathbb{R}^n$ and $\bar{x} \in X$, then the indicator function ι_X has*

$$\partial^\infty \iota_X(\bar{x}) = N_X(\bar{x}).$$

If $h : \mathbb{R}^m \to \overline{\mathbb{R}}$ is proper, lsc and convex and $\bar{u} \in \operatorname{dom} h$, then

$$\partial^\infty h(\bar{u}) = N_{\operatorname{dom} h}(\bar{u}).$$

A function $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$ is *locally Lipschitz continuous* at $\bar{x}$ when there are $\delta \in (0, \infty)$ and $\kappa \in [0, \infty)$ such that

$$\left|\varphi(x) - \varphi(x')\right| \leq \kappa \|x - x'\|_2 \quad \text{whenever } \|x - \bar{x}\|_2 \leq \delta, \ \|x' - \bar{x}\|_2 \leq \delta.$$

If φ is locally Lipschitz continuous at every $\bar{x} \in \mathbb{R}^n$, then φ is locally Lipschitz continuous.

5.10 Proposition (subgradients and local Lipschitz continuity). *Suppose that $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$ is lsc and $\bar{x}$ is a point at which φ is finite. Then,*

$$\varphi \text{ is locally Lipschitz continuous at } \bar{x} \iff \partial^\infty \varphi(\bar{x}) = \{0\}.$$

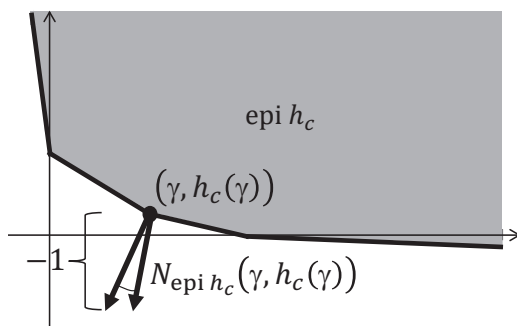*Under these circumstances, $\partial \varphi(\bar{x})$ is nonempty and compact.*

Figure 8: Normal vectors $(v, -1)$ produce subgradients v of monitoring function in naval resupply problem.

5.11 Example (naval resupply at sea; subgradients). *The monitoring function in the naval resupply problem in Example 2.2 is expressed by a piecewise affine function of the form*

$$h_c(\gamma) = \max_{i=1,\ldots,m} \alpha_i \gamma + \beta_i,$$

where α_i, β_i are given. This function is nonsmooth, but convex. Its subgradients are given by the formula

$$\partial h_c(\gamma) = \left\{ \sum_{i=1}^{m} \mu_i \alpha_i \,\middle|\, \sum_{i=1}^{m} \mu_i = 1, \;\; \mu_i = 0 \text{ if } \alpha_i \gamma + \beta_i < h_c(\gamma), \;\; \mu_i \geq 0 \text{ otherwise} \right\}.$$

Detail. Figure 8 illustrates the epigraph of h_c and a corresponding normal cone, which in turn produces the subgradients. At most points, the set of subgradients contains only a single value, the slope α_i of the corresponding line segment. However, when the graph of h_c has a kink, the normal cone at that point widens as in the figure and includes all convex combinations of the adjacent slope coefficients.☐

6 Quantitative Analysis

With the confidence that subgradients are defined for arbitrary functions, we now return to the main subject of assessing whether a particular model is good in the sense that the resulting minimum value doesn't change much as a parameter of concern varies. While Section 4 furnishes conditions under which the minimum value changes continuously, we now step further and quantify the *rate of change*. This provides a deeper understanding of the sensitivity to changes in model parameters. In particular, one might be able to identify which of the parameters are more significant and these can then receive additional scrutiny.

Certain pathological cases might occur if a problem is unbounded in some sense and these are avoided by imposing a boundedness assumption. Specifically, in the context of a function $f : \mathbb{R}^m \times \mathbb{R}^n \to \overline{\mathbb{R}}$ with values $f(u, x)$, we say that the function is *level-bounded in x locally uniformly in u* when

$$\forall \bar{u} \in \mathbb{R}^m \text{ and } \alpha \in \mathbb{R} \quad \exists \varepsilon > 0 \text{ and a bounded set } B \subset \mathbb{R}^n$$
$$\text{such that } \left\{ x \in \mathbb{R}^n \mid f(u, x) \leq \alpha \right\} \subset B \quad \text{when } \|u - \bar{u}\|_2 \leq \varepsilon.$$

Informally, the property amounts to having, for each $\bar{u}$ and α, a bounded level-set $\{x \mid f(\bar{u}, x) \leq \alpha\}$ with the bound remaining valid under perturbation around $\bar{u}$.

Under this assumption, we can, at least partially, characterize the subgradients of the minimum values of a Rockafellian. Since minimum values rarely vary smoothly with parametric changes, we indeed need to bring in subgradients as gradients may not be defined. The following theorem is given in [38, Theorem 5.13]. For refinements, we refer to [21, 22] and also [2] for the convex case.

6.1 Theorem (subgradients of min-value function). *For a proper lsc function $f : \mathbb{R}^m \times \mathbb{R}^n \to \overline{\mathbb{R}}$, with $f(u, x)$ level-bounded in x locally uniformly in u, let*

$$p(u) = \inf f(u, \cdot) \quad \text{and} \quad P(u) = \operatorname{argmin} f(u, \cdot) \;\; \forall u \in \mathbb{R}^m.$$

Then, at any $\bar{u} \in \operatorname{dom} p$, one has

$$\partial p(\bar{u}) \subset \bigcup_{\bar{x} \in P(\bar{u})} \{y \in \mathbb{R}^m \mid (y, 0) \in \partial f(\bar{u}, \bar{x})\}.$$

If f is convex, then p is convex, the inclusion holds with equality and each set in the union coincides.

Typically, f in the theorem is a Rockafellian with anchor $\bar{u}$ for an actual problem of minimizing $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$. Then, $p(\bar{u})$ is the minimum value of the actual problem and $\partial p(\bar{u})$ estimates the effect of perturbation on the minimum value. Neither smoothness nor convexity is needed in the first part of the theorem. Though, convexity allows us to sharpen the result from an inclusion to an equality.

In essence, if the subgradients of the Rockafellian are small in length, then the subgradients of the minimum value function are also small. Consequently, the effect on the minimum value of perturbing u away from $\bar{u}$ is also small. In contrast, if the Rockafellian has large subgradients, then the change in minimum value can be substantial and this calls into question the suitability of the formulation.

6.2 Example (perturbation of inequality). *Let's return to Example 2.3, but with the constraint in the actual problem now being $g(x) \leq 0$. The Rockafellian then has anchor $\bar{u} = 0$. Using Theorem 6.1, one can show that*

$$\partial p(0) = \{2\}.$$

Figure 9 visualizes the minimum value function p and its affine approximation given by this subgradient, i.e.,

$$u \mapsto p(0) + 2(u - 0) = 5 + 2u.$$

Detail. We compute the subgradients of the Rockafellian using Proposition 5.4 and this produces

$$\partial f(u, x) = (0, \ 2x) + \partial \iota_{(-\infty, 0]}\big(g(x) + u\big).$$

The last term can be calculated using a chain rule (see [38, Theorem 4.64]) so that

$$\partial f(u, x) = (0, \ 2x) + (1, \nabla g(x)) N_{(-\infty, 0]}\big(g(x) + u\big) = (0, \ 2x) + (1, \ 2x - 6) Y(u, x),$$

for $(u, x) \in \operatorname{dom} f$, where $Y(u, x) = [0, \infty)$ if $g(x) + u = 0$ and $Y(u, x) = \{0\}$ otherwise. Thus,

$$(y, 0) \in \partial f(0, x) \quad \Longleftrightarrow \quad 0 = 2x + y(2x - 6) \quad \text{for some } y \in Y(0, x).$$

Since we trivially observe that $\bar{x} = 2$ is the unique minimizer of the actual problem, and this makes $P(\bar{u}) = \{2\}$ in Theorem 6.1, the previous equivalence produces a unique $y = 2$. Consequently,

$$\partial p(0) \subset \{2\}.$$

We see directly that f is convex so the inclusion can be replaced by an equality. Moreover, in view of the convexity inequality (5.1),

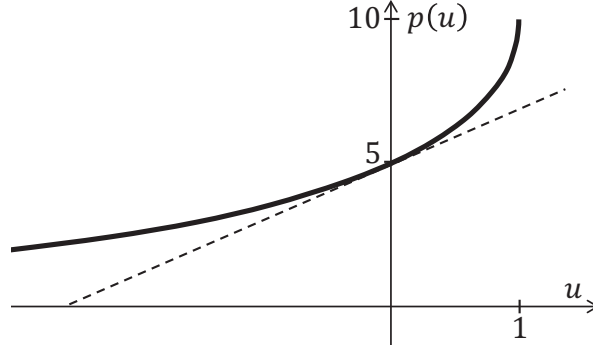$$p(u) \geq 5 + 2u \quad \forall u \in \mathbb{R}$$

Figure 9: Minimum value function p and its estimates in Example 6.2.

and this is also confirmed by Figure 9.

If we shift the focus to $\bar{u} = 1$ as in Example 2.3, then the above analysis breaks down because there aren't any subgradients of p at 1; the slope becomes vertical at 1 in Figure 9. □

For problems in the composite form, we obtain the following detailed result; see [38, Proposition 5.16].

6.3 Proposition (sensitivity in composite optimization). *For smooth $f_0 : \mathbb{R}^n \to \mathbb{R}$, smooth $F : \mathbb{R}^n \to \mathbb{R}^m$, proper, lsc and convex $h : \mathbb{R}^m \to \overline{\mathbb{R}}$ and nonempty closed $X \subset \mathbb{R}^n$, consider the problem*

$$\underset{x \in X}{\text{minimize }} f_0(x) + h\big(F(x)\big)$$

and the associated Rockafellian $f : \mathbb{R}^m \times \mathbb{R}^n \to \overline{\mathbb{R}}$ given by

$$f(u, x) = \iota_X(x) + f_0(x) + h\big(F(x) + u\big).$$

Suppose that $f(u, x)$ is level-bounded in x locally uniformly in u. Let $p(u) = \inf f(u, \cdot)$, $P(u) = \text{argmin } f(u, \cdot)$ and, for $x \in P(u)$,

$$Y(u, x) = \big\{y \in \partial h\big(F(x) + u\big) \ \big| \ -\nabla f_0(x) - \nabla F(x)^\top y \in N_X(x)\big\}.$$

Then, for $\bar{u} \in \text{dom } p$, one has

$$\partial p(\bar{u}) \subset \bigcup_{\bar{x} \in P(\bar{u})} Y(\bar{u}, \bar{x}).$$

We note that the actual problem in the proposition might be neither smooth nor convex. Still, we obtain a rather specific expression for the subgradients of the minimum value function. In particular, if all the y-vectors in $Y(\bar{u}, \bar{x})$ are small, then we conclude that the minimum value changes just a little under perturbation of u away from $\bar{u}$ and this would indicate a good model formulation.

21

6.4 Example (economic order quantity in inventory management). *A store manager needs to determine how many units to order of a given product each time the product runs out. The number of units sold per year is $\rho > 0$, which means that an order quantity of x units results in ρ/x orders per year. Suppose that each one of these orders has a fixed cost of $\beta > 0$. (The cost of the units doesn't factor in here because the total number of units ordered across the whole year is always ρ.) The manager also faces an inventory cost of $\alpha > 0$ per unit and year. With an order quantity x, the average inventory is $x/2$ so the annual inventory cost becomes $\alpha x/2$. The manager would like to determine an order quantity x such that both the ordering cost $\beta\rho/x$ and the inventory cost $\alpha x/2$ are low. Let's adopt the goal optimization model (cf. Example 2.1)*

$$\underset{x \in X \subset \mathbb{R}}{\text{minimize}} \ \max \left\{ \beta\rho/x - \tau, \ \tfrac{1}{2}\alpha x - \sigma \right\},$$

where τ, σ are the goals for the ordering cost and the inventory cost, respectively. What's the effect of changing the goals?

Detail. For perturbation $u \in \mathbb{R}^2$ of the goals, let's adopt the Rockafellian given by

$$f(u, x) = \iota_X(x) + h\big(F(x) + u\big),$$

where $h(z) = \max\{z_1, z_2\}$ and

$$F(x) = \big(f_1(x), f_2(x)\big) = \big(\beta\rho/x - \tau, \tfrac{1}{2}\alpha x - \sigma\big).$$

Then, the anchor of f is $\bar{u} = 0$ and the actual model corresponds to minimizing $f(0, \cdot)$. Assuming that X is a nonempty closed subset of the positive numbers, then the assumptions of Proposition 6.3 are satisfied as long as f_1 is extended in a smooth manner from being defined on the positive numbers to all of $\mathbb{R}$. (This is only a technical issue as we don't consider order quantities below 1 anyway.) We can show that in the notation of Proposition 6.3,

$$Y(u, x) = \big\{ y \in \mathbb{R}^2 \ \big| \ \beta\rho y_1 x^{-2} - \tfrac{1}{2}\alpha y_2 \in N_X(x)$$
$$y_1 + y_2 = 1; \ y_i \geq 0 \ \text{ if } f_i(x) + u_i = f(u, x), \ y_i = 0 \ \text{ otherwise}; \ i = 1, 2 \big\}$$

for $x \in P(u)$. In general, one would employ an algorithm to compute a minimizer $x^\star$ of the model at hand under $\bar{u}$, which hopefully also computes the corresponding multipliers in $Y(\bar{u}, x^\star)$. If we assume that $\sigma = \tau = 0$ and X isn't active at a minimizer, then we obtain the unique minimizer analytically by solving $f_1(x) = f_2(x)$, which produces

$$x^\star = \sqrt{2\beta\rho/\alpha}, \quad \text{with } Y(0, x^\star) = \big\{ \big(\tfrac{1}{2}, \tfrac{1}{2}\big) \big\}.$$

Then, by Proposition 6.3, the minimum value function $p(u) = \inf f(u, \cdot)$ actually has $\nabla p(0) = (1/2, 1/2)$. This provides the insight that if the current goal of zero ordering cost is changed to a small positive number τ, which corresponds to setting $u = (-\tau, 0)$, then the change in minimum value is approximately $\langle \nabla p(0), u \rangle = -\tau/2$. The negative value is reasonable as raising the goal value reduces the shortfall. $\square$

7 Optimality Conditions

The existence of a convenient optimality condition is a major advantage as we analyze and solve an optimization problem. In the nonconvex setting, computational methods might be entirely centered on satisfying the condition. A model can therefore be thought of as having good properties if the resulting optimization problem has an approachable optimality condition. Again, Rockafellians enter as key quantities. The next theorem is a major extension of the Fermat Rule in Theorem 5.6; see [38, Theorem 5.10].

7.1 Theorem (Rockafellar condition for optimality). *For the problem of minimizing $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$, suppose that $\bar{x} \in \mathbb{R}^n$ is a local minimizer, $f : \mathbb{R}^m \times \mathbb{R}^n \to \overline{\mathbb{R}}$ is a proper lsc Rockafellian with anchor $\bar{u} \in \mathbb{R}^m$ and the following qualification holds:*

$$(y, 0) \in \partial^\infty f(\bar{u}, \bar{x}) \quad \Longrightarrow \quad y = 0. \tag{7.1}$$

Then,

$$\exists \bar{y} \in \mathbb{R}^m \quad \text{such that} \quad (\bar{y}, 0) \in \partial f(\bar{u}, \bar{x}).$$

This condition is sufficient for $\bar{x}$ to be a (global) minimizer of φ when f is convex.

The auxiliary vector $y \in \mathbb{R}^m$, referred to as a *multiplier vector*, is associated with the perturbation vector u and broadens the view of such quantities beyond Lagrange multipliers from the classical Karush-Kuhn-Tucker condition. Every Rockafellian for a problem defines an optimality condition, at least as long as the qualification (7.1) holds. The Rockafellian might now be chosen more with the goal of obtaining a useful optimality condition, than based on concerns about sensitivity analysis. Still, by matching y in Theorem 6.1 with y here, we realize that solving the Rockafellar condition amounts to determining (x, y), with y furnishing sensitivity information relative to perturbations according to the chosen Rockafellian. This means that solving a problem in this broader sense not only results in a recommended course of action (x), but also a measure of stability of the underlying model (y) that can help us verify its validity.

For problems in the composite form and the Rockafellian given in Proposition 6.3, the optimality condition specializes as stated in [38, Theorem 4.75].

7.2 Theorem (optimality for composite problem). *For smooth $f_0 : \mathbb{R}^n \to \mathbb{R}$, smooth $F : \mathbb{R}^n \to \mathbb{R}^m$, closed $X \subset \mathbb{R}^n$ and proper, lsc and convex $h : \mathbb{R}^m \to \overline{\mathbb{R}}$, suppose that the following qualification holds at $x^\star$:*

$$y \in N_{\operatorname{dom} h}\big(F(x^\star)\big) \quad \text{and} \quad -\nabla F(x^\star)^\top y \in N_X(x^\star) \quad \Longrightarrow \quad y = 0. \tag{7.2}$$

If $x^\star$ is a local minimizer of the problem

$$\underset{x \in X}{\text{minimize}} \ f_0(x) + h\big(F(x)\big),$$

then

$$\exists y \in \partial h\big(F(x^\star)\big) \quad \text{such that} \quad -\nabla f_0(x^\star) - \nabla F(x^\star)^\top y \in N_X(x^\star).$$

The theorem can be specialized further in many directions. We limit the attention to the classical case of equality constraints, which goes back to Lagrange.

7.3 Example (smooth objective and equality constraint functions). *For smooth functions $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 0, 1, \ldots, m$, consider the problem*

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \ f_0(x) \ \ subject \ to \ \ f_i(x) = 0, \ i = 1, \ldots, m.$$

If $x^\star$ is a local minimizer with $\{\nabla f_i(x^\star), i = 1, \ldots, m\}$ linearly independent, then one can apply Theorem 7.2 to conclude that there is $y \in \mathbb{R}^m$ such that

$$\nabla f_0(x^\star) + \sum_{i=1}^{m} y_i \nabla f_i(x^\star) = 0, \quad f_i(x^\star) = 0, \quad y_i \in \mathbb{R}, \quad i = 1, \ldots, m.$$

Detail. The problem is of the composite form with $X = \mathbb{R}^n$, $F(x) = (f_1(x), \ldots, f_m(x))$ and $h(z) = \iota_{\{0\}^m}(z)$. This monitoring function is proper, lsc and convex, with

$$\partial h(0) = \mathbb{R}^m$$

so the multiplier vector y is unrestricted. Since $N_X(x) = \{0\}$, the gradient condition in Theorem 7.2 reduces to

$$\nabla f_0(x^\star) + \nabla F(x^\star)^\top y = 0.$$

The domain of h is $\{0\}^m$. Consequently, $N_{\text{dom } h}(z) = \mathbb{R}^m$ if $z = 0$ and $N_{\text{dom } h}(z)$ is otherwise not defined. The qualification (7.2) therefore specializes to checking whether $y \in \mathbb{R}^m$ and $\nabla F(x^\star)^\top y = 0$ imply $y = 0$. But, this is just the linear independence assumption. $\square$

8 Algorithmic Approaches

A Rockafellian associated with a particular problem also specifies algorithmic possibilities via problem relaxations. The strength of the resulting relaxations can be used to assess whether the underlying model is computationally attractive in the first place. Moreover, these relaxations define a dual problem, which highlights the many opportunities that lie beyond the classical dual problems of linear and convex programming.

This section starts with the construction of relaxations and then defines the corresponding dual problems.

8.1 Rockafellian Relaxation

A *relaxation* of a minimization problem is an alternative problem with a minimum value no higher than that of the actual one. To make this concrete, let's consider $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$, the actual problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \ \varphi(x)$$

and an associated Rockafellian $f : \mathbb{R}^m \times \mathbb{R}^n \to \overline{\mathbb{R}}$ with anchor 0. (The focus on 0 instead of a more general $\bar{u}$ as the anchor promotes symmetry below, without much loss of generality because one can always shift the perturbation vector by redefining f.) We can't do worse if permitted to *optimize* the perturbation vector u together with x. Thus, the problem

$$\underset{u\in\mathbb{R}^m, x\in\mathbb{R}^n}{\text{minimize}} \; f(u,x)$$

is a relaxation of the actual problem. Regardless of $y \in \mathbb{R}^m$, this is also the case for

$$\underset{u\in\mathbb{R}^m, x\in\mathbb{R}^n}{\text{minimize}} \; f(u,x) - \langle y, u \rangle \tag{8.1}$$

because $u = 0$ remains a possibility. In summary, for any $y \in \mathbb{R}^m$,

$$\inf_{x\in\mathbb{R}^n} \varphi(x) = \inf_{x\in\mathbb{R}^n} f(0,x) \geq \inf_{(u,x)\in\mathbb{R}^m \times \mathbb{R}^n} \big\{ f(u,x) - \langle y, u \rangle \big\}. \tag{8.2}$$

A benefit from bringing in a vector y is that we now can tune the relaxation by selecting y appropriately. We return to this subject below.

We immediately have an approach to addressing the actual problem: (i) Solve the relaxation (8.1) for some $y \in \mathbb{R}^m$ and obtain a lower bound on $\inf \varphi$. (ii) Use any (heuristic) method to obtain a candidate solution $\bar{x}$. (iii) Bound the *optimality gap* $\varphi(\bar{x}) - \inf \varphi$ of $\bar{x}$ using

$$\varphi(\bar{x}) - \inf_{(u,x)\in\mathbb{R}^m \times \mathbb{R}^n} \big\{ f(u,x) - \langle y, u \rangle \big\}.$$

If this quantity is sufficiently close to zero, then $\bar{x}$ might be deemed acceptable, making further efforts to find an even better solution superfluous. An optimality gap is measured in the units of the objective function and is thus well understood by a decision maker in most cases. We refer to this approach as *Rockafellian relaxation*.

Of course, Rockafellian relaxation is only meaningful when the Rockafellian has been selected in such a manner that (8.1) is easier to solve than the actual problem. This is often the case because "lifting" of the problem to a higher dimension, involving both x and u, allows for more flexibility. Moreover, some common choices of Rockafellians lead to explicit expressions for the minimization over u in (8.1). Let's denote by

$$l(x,y) = \inf_{u\in\mathbb{R}^m} \big\{ f(u,x) - \langle y, u \rangle \big\} \tag{8.3}$$

the resulting minimum value after such optimization over u. The function $l : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ given by this formula is the *Lagrangian* of f. If the Lagrangian corresponding to a particular Rockafellian can be expressed in a convenient form, then the solution of (8.1) might simplify.

8.1 Example (Lagrangian for equalities and inequalities). *For $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 0, 1, \ldots, m$, and $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, q$, let's consider the problem*

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \; f_0(x) + \iota_D\big(F(x)\big),$$

25

where $D = \{0\}^m \times (-\infty, 0]^q$ and

$$F(x) = \big(f_1(x), \ldots, f_m(x), g_1(x), \ldots, g_q(x)\big).$$

A Rockafellian $f : \mathbb{R}^{m+q} \times \mathbb{R}^n \to \overline{\mathbb{R}}$ for the problem is defined by

$$f(u, x) = f_0(x) + \iota_D\big(F(x) + u\big).$$

The actual problem is then to minimize $f(0, \cdot)$ and the Lagrangian has

$$l(x, y) = \begin{cases} f_0(x) + \langle F(x), y \rangle & \text{if } y_{m+1}, \ldots, y_{m+q} \geq 0 \\ -\infty & \text{otherwise.} \end{cases}$$

Detail. With $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^{m+q}$, the Rockafellian f produces the Lagrangian given by

$$l(x, y) = \inf_{u \in \mathbb{R}^{m+q}} \big\{ f_0(x) + \iota_D\big(F(x) + u\big) - \langle y, u \rangle \big\}.$$

If there's $y_i < 0$ for some $i \in \{m+1, \ldots, m+q\}$, then we can select $u_j = -f_j(x)$ for all $j \in \{1, \ldots, m\}$ and $u_j = -g_j(x)$ for all $j \in \{m+1, \ldots, m+q\} \setminus \{i\}$ so that $\iota_D(F(x) + u)$ remains zero as $u_i \to -\infty$. But, then

$$f_0(x) + \iota_D\big(F(x) + u\big) - \langle y, u \rangle \to -\infty$$

and $l(x, y) = -\infty$.

If $y_i \geq 0$ for all $i \in \{m+1, \ldots, m+q\}$, then $\bar{u}$, with components $\bar{u}_j = -f_j(x)$ for all $j \in \{1, \ldots, m\}$ and $\bar{u}_j = -g_j(x)$ for all $j \in \{m+1, \ldots, m+q\}$, solves

$$\underset{u \in \mathbb{R}^{m+q}}{\text{minimize}} \; f_0(x) + \iota_D\big(F(x) + u\big) - \langle y, u \rangle$$

and this results in

$$l(x, y) = f_0(x) - \langle y, \bar{u} \rangle = f_0(x) + \big\langle y, F(x) \big\rangle.$$

This establishes the claimed formula for the Lagrangian. $\qquad\qquad\square$

For $h : \mathbb{R}^m \to \overline{\mathbb{R}}$, we recall that the function $h^* : \mathbb{R}^m \to \overline{\mathbb{R}}$ defined by

$$h^*(v) = \sup_{u \in \mathbb{R}^m} \big\{ \langle v, u \rangle - h(u) \big\}$$

is the *conjugate* of h. This helps us to express Lagrangians for problems in the composite form as seen next; a proof is given by [38, Proposition 5.28].

8.2 Proposition (Lagrangian for composite function). *For $f_0 : \mathbb{R}^n \to \mathbb{R}$, $F : \mathbb{R}^n \to \mathbb{R}^m$ and proper, lsc and convex $h : \mathbb{R}^m \to \overline{\mathbb{R}}$, consider the problem*

$$\underset{x \in X \subset \mathbb{R}^n}{\text{minimize}} \; f_0(x) + h\big(F(x)\big).$$

The Rockafellian given by

$$f(u, x) = \iota_X(x) + f_0(x) + h\big(F(x) + u\big)$$

recovers the actual problem as minimizing $f(0, \cdot)$ and produces a Lagrangian with

$$l(x, y) = \iota_X(x) + f_0(x) + \big\langle F(x), y \big\rangle - h^*(y).$$

Let's consider a concrete application of Rockafellian relaxation.

8.3 Example (constrained shortest path problem). *Let (V, E) be a directed graph with vertex set V and edge set E. Each edge $(i, j) \in E$ connects distinct vertices $i, j \in V$, and it possesses length $c_{ij} \in [0, \infty)$ and weights $D_{kij} \in [0, \infty)$ for $k = 1, \dots, q$. A directed s-t path is an ordered set of edges of the form $\{(s, i_1), (i_1, i_2), \dots, (i_{\nu-1}, t)\}$ for some $\nu \in \mathbb{N}$. Given two distinct vertices $s, t \in V$, the shortest-path problem seeks to determine a directed s-t path such that the sum of the edge lengths along the path is minimized. This is a well-studied problem that can be solved efficiently using specialized algorithms; see [1, Chapters 4 and 5].*

For nonnegative $d_k, k = 1, \dots, q$, the task becomes significantly harder if the sum of the weights D_{kij} along the path can't exceed d_k for each k. This is the constrained shortest-path problem, which can be addressed by Rockafellian relaxation—also called Lagrangian relaxation in the present context. In routing of a drone through a discretized 3-dimensional airspace, the weights D_{1ij} might represent fuel consumption along edge (i, j), which can't exceed a capacity d_1. Figure 10 illustrates a route satisfying such a fuel constraint while minimizing exposure to enemy radars expressed by c_{ij}; cf. [37].
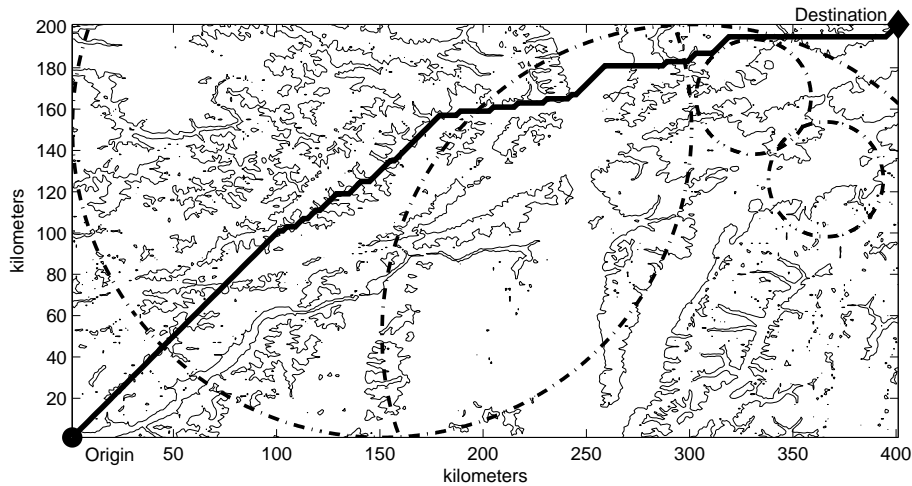

Figure 10: Route for a drone through a 3-dimensional airspace to a destination (solid line) that minimizes exposure to enemy radars (circles) while satisfying a fuel constraints [37]. Altitude changes to leverage terrain masking aren't shown.

Detail. The constrained shortest-path problem is formulated as follows. Suppose that m is the number of vertices in V and n is the number of edges in E. Let A denote the $m \times n$-vertex-edge incidence matrix such that if edge $e = (i, j)$, then $A_{ie} = 1$, $A_{je} = -1$ and $A_{i'e} = 0$ for any $i' \neq i, j$. Also, let $b_s = 1$, $b_t = -1$ and $b_i = 0$ for $i \in V \backslash \{s, t\}$ and collect them in the vector b. For each $k = 1, \dots, q$, we place the edge weights $\{D_{kij}, (i, j) \in E\}$ in the vector D_k. The $q \times n$-matrix D has D_k as its kth row. Let $d = (d_1, \dots, d_q)$. With c being the vector of $\{c_{ij}, (i, j) \in E\}$, the constrained shortest-path problem

may then be formulated as (cf. [1, p. 599])

$$\text{minimize} \langle c, x \rangle \quad \text{subject to} \quad Ax = b, \ Dx \leq d, \ x \in \{0,1\}^n.$$

A point $\bar{x} \in \{0,1\}^n$ satisfying $Ax = b$ corresponds to an s-t path with $\bar{x}_{ij} = 1$ if edge (i,j) is on the path and $\bar{x}_{ij} = 0$ otherwise. We assume there's at least one such path. Then, $\langle c, \bar{x} \rangle$ gives the length of the path and $\langle D_k, \bar{x} \rangle$ the kth weight of the path.

In the absence of the weight-constraint $Dx \leq d$, the model reduces to one for the shortest-path problem and this opens up an opportunity for applying Rockafellian relaxation via Proposition 8.2. Let

$$X = \big\{ x \in \{0,1\}^n \ \big| \ Ax = b \big\},$$

which is nonempty by assumption, and consider the Rockafellian given by

$$f(u, x) = \iota_X(x) + \langle c, x \rangle + \iota_{(-\infty,0]^q}(Dx - d + u).$$

The actual problem corresponds to minimizing $f(0, \cdot)$ over $\mathbb{R}^n$. These definitions fit the setting of Proposition 8.2 and

$$l(x, y) = \iota_X(x) + \langle c, x \rangle + \langle Dx - d, y \rangle - \iota_{[0,\infty)^q}(y),$$

where we use the fact that the conjugate of $\iota_{(-\infty,0]^q}$ is $\iota_{[0,\infty)^q}$; see [38, Example 5.29]. Thus, the chosen Rockafellian results in a Lagrangian without the constraint $Dx \leq d$. The minimization of this Lagrangian with respect to x is nothing but a shortest-path problem on the directed graph, but with edge lengths changed from c_{ij} to $c_{ij} + \sum_{k=1}^q D_{kij} y_k$, which can be solved efficiently using specialized algorithms. The resulting minimum value, modified by $\langle d, y \rangle$, yields a lower bound on the minimum value of the constrained shortest-path problem as seen from (8.2). The lower bound can be used to assess the optimality gap for any candidate path, for example obtained by a greedy search or an enumeration algorithm; see [10]. $\qquad\blacksquare$

8.2 Dual Problems

Rockafellian relaxation produces a lower bound on the minimum value of the actual problem *regardless* of the choice of auxiliary vector y. However, it would be useful to tune this vector so that the lower bound becomes as high as possible. The underlying model may also be good in the sense that the lower bound associated with the chosen Rockafellian can be brought the whole way up to the minimum value of the actual problem. This could present a major computational advantage: with a properly tuned y, minimization of the Rockafellian modified by a linear term can largely substitute for the actual problem. The process of tuning y involves optimization and gives rise to *dual problems.*

The notation y for the auxiliary vector in Rockafellian relaxation isn't coincidental. It's deeply connected with multiplier vectors emerging from the Rockafellar condition for optimality; see Theorem 7.1. In fact, solving a dual problem can be a viable approach to determining a multiplier vector and thus identify sensitivity of the minimum value to perturbations, for example via Proposition 6.3.

To make the setting concrete, let $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$ and consider the problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \ \varphi(x),$$

with an associated Rockafellian $f : \mathbb{R}^m \times \mathbb{R}^n \to \overline{\mathbb{R}}$ anchored at 0. Via the corresponding Lagrangian $l : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$, given by (8.3), this produces a *dual problem*

$$\underset{y \in \mathbb{R}^m}{\text{maximize}} \ \psi(y) = \inf_{x \in \mathbb{R}^n} l(x, y).$$

Since $\inf \varphi \geq \psi(y)$ for all y by (8.2), the dual problem indeed aims to find the best lower bound on $\inf \varphi$. Since *every* Rockafellian associated with the actual problem defines a dual problem, there are endless possibilities. One might choose a Rockafellian that produces a simple dual problem solvable by standard algorithms. This has been the traditional approach in linear programming. In the setting of Example 8.1, with all functions being affine, the Rockafellian adopted there (focusing on right-hand side perturbations) recovers the usual linear programming dual problems; see [38, Example 5.41] for details. But, this is just one possibility. Beyond the convex case, the dual problems tend to become less tractable but much depends on the structure of the actual problem as well as the choice of Rockafellian. One property is common across all dual problems: the objective function is concave. We realize this by writing

$$-\psi(y) = -\inf_{x,u} \big\{ f(u, x) - \langle y, u \rangle \big\} = \sup_{x,u} \big\{ \langle y, u \rangle - f(u, x) \big\}.$$

Thus, $-\psi$ is convex by virtue of being given by the pointwise supremum across a collection of affine functions. Interestingly, regardless of convexity of the actual primal problem and the chosen Rockafellian, the dual objective function ψ is concave, which makes subgradient, cutting plane and proximal point methods at least conceptually available.

Traditionally, the term "Lagrangian" has often been limited to the function emerging in the context of equality and inequality constraints under right-hand side perturbations; see Example 8.1. Even with the broader definition in (8.3), we view a Lagrangian as a secondary quantity stemming from a more fundamental Rockafellian. Still, Lagrangians remain important in saddle point theory, which connects the multipliers from optimality conditions with dual variables, serve as a bridge to game theory and promote an elegant symmetry with the actual problem, also expressible in terms of a Lagrangian. We refer to [38, Chapter 5] for details.

In our setting, however, we can completely bypass Lagrangians. Specifically, for any Rockafellian $f : \mathbb{R}^m \times \mathbb{R}^n \to \overline{\mathbb{R}}$ and the corresponding Lagrangian l, the dual objective function has

$$\psi(y) = \inf_x l(x, y) = \inf_{u,x} \big\{ f(u, x) - \langle u, y \rangle \big\}$$
$$= -\sup_{u,x} \big\{ \langle u, y \rangle + \langle x, 0 \rangle - f(u, x) \big\} = -f^*(y, 0).$$

Consequently, a dual objective function can just as well be defined directly in terms of the conjugate of the underlying Rockafellian.

The usefulness of a dual problem depends on the size of the resulting *duality gap*

$$\inf \varphi - \sup \psi.$$

If it's large, then the corresponding Rockafellian relaxation might not be practically helpful. We say that a Rockafellian possesses *strong duality* relative to the actual problem when the duality gap is zero. Then, the actual problem has the good property that its minimum value can be determined by solving a convex optimization problem! We indeed have strong duality in linear programming under the usual right-hand side perturbations provided that the actual problem isn't infeasible. However, strong duality is far from automatic.

8.4 Example (failure of strong duality). *For the problem of minimizing x^3 subject to $x \geq 0$ and the Rockafellian given by*

$$f(u, x) = x^3 + \iota_{(-\infty, 0]}(-x + u),$$

we obtain a Lagrangian of the form

$$l(x, y) = \begin{cases} x^3 - xy & \text{if } y \geq 0 \\ -\infty & \text{otherwise;} \end{cases}$$

see Example 8.1. The dual objective function has $\psi(y) = -\infty$ for all $y \in \mathbb{R}$, while the minimum value of the actual problem is zero so the duality gap is ∞.

Detail. In this case, the Lagrangian isn't convex in its first argument. However, strong duality may fail even under convexity. Consider the problem

$$\underset{x \in \mathbb{R}^2}{\text{minimize }} e^{-x_1} \text{ subject to } g(x) \leq 0, \text{ where } g(x) = \begin{cases} x_1^2/x_2 & \text{if } x_2 > 0 \\ \infty & \text{otherwise,} \end{cases}$$

and a Rockafellian of the form

$$f(u, x) = \begin{cases} e^{-x_1} & \text{if } g(x) + u \leq 0 \\ \infty & \text{otherwise.} \end{cases}$$

Similar to Example 8.1, this produces a Lagrangian with

$$l(x, y) = \begin{cases} e^{-x_1} + yg(x) & \text{if } x \in \text{dom } g, \ y \geq 0 \\ \infty & \text{if } x \notin \text{dom } g \\ -\infty & \text{otherwise.} \end{cases}$$

Consequently, the dual objective function has $\psi(y) = 0$ if $y \geq 0$, but $\psi(y) = -\infty$ otherwise. The maximum value of the dual problem is therefore 0. The actual problem has minimum value of 1. Thus, the duality gap is 1 even though $l(\cdot, y)$ is convex regardless of $y \in \mathbb{R}$. $\square$

Despite these discouraging examples, there are large classes of problems beyond linear optimization problems for which strong duality holds. The following result summarizes key insights from [34, Theorem 11.39] in the convex setting. For strong duality without convexity, we refer to Section 6.B in [38].

Generally, we denote by $\text{int } C$ the *interior* of a set $C \subset \mathbb{R}^n$, which informally equals C without its boundary points.

8.5 Theorem (strong duality). *For the problem of minimizing $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$ and a proper, lsc and convex Rockafellian $f : \mathbb{R}^m \times \mathbb{R}^n \to \overline{\mathbb{R}}$ with anchor 0, the corresponding dual objective function ψ satisfies strong duality provided that $0 \in \mathrm{int}(\mathrm{dom}\, p)$, where p is the minimum value function given by $p(u) = \inf f(u, \cdot)$.*

If in addition $p(0) > -\infty$, then

$$\partial p(0) = \mathrm{argmax}\, \psi,$$

which must be a nonempty and bounded set.

In Example 6.2, $\mathrm{dom}\, p = (-\infty, 1]$ so we certainly have $0 \in \mathrm{int}(\mathrm{dom}\, p)$ and strong duality holds; see Figure 9. Moreover, $p(0) = 5$ and $\partial p(0) = \{2\}$, which imply that the dual problem has 2 as its unique maximizer, with 5 as maximum value. We can determine all of this based on Theorem 8.5 without having a detailed formulation for the dual problem. The formula for $\partial p(0)$ supplements Theorem 6.1, but most significantly it highlights the profound role played by a dual problem. Under the conditions of the theorem, solving the dual problem furnishes both the minimum value of the actual problem as well as its sensitivity to perturbations as defined by a Rockafellian.

In a specific setting, strong duality is guaranteed by ensuring that the constraint functions leave some "slack," which often is easily verified.

8.6 Example (Slater constraint qualification). *For smooth convex functions $f_0, g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, q$ and the problem*

$$\underset{x \in \mathbb{R}^n}{\text{minimize}}\ f_0(x)\ \ \text{subject to}\ \ g_i(x) \leq 0,\ i = 1, \ldots, q,$$

let's consider the Rockafellian given by

$$f(u, x) = f_0(x) + \iota_{(-\infty, 0]^q}\big(G(x) + u\big), \quad \text{with } G(x) = \big(g_1(x), \ldots, g_q(x)\big).$$

The resulting dual problem satisfies strong duality provided that the following Slater condition holds:

$$\exists \bar{x}\ \ \text{such that}\ \ g_i(\bar{x}) < 0, \ \ i = 1, \ldots, q.$$

Detail. Let $p(u) = \inf f(u, \cdot)$. Under the Slater condition, there exist $\bar{x} \in \mathbb{R}^n$ and $\delta > 0$ such that $g_i(\bar{x}) + u_i \leq 0$ when $|u_i| \leq \delta$ for all i. Consequently, $p(u) \leq f_0(\bar{x}) \in \mathbb{R}$ when $\|u\|_\infty \leq \delta$, which means that $0 \in \mathrm{int}(\mathrm{dom}\, p)$ and strong duality holds by Theorem 8.5.

In the setting of Theorem 7.2, specialized to convex inequality constraints, the Slater condition ensures that the qualification (7.2) holds; see [38, Examples 4.49, 5.47] for details.

The Slater condition is by no means necessary for strong duality. For example, consider the problem of minimizing x subject to $x^2 \leq 0$ and the Rockafellian given by

$$f(u, x) = x + \iota_{(-\infty, 0]}(x^2 + u).$$

We obtain directly that

$$p(u) = \inf f(u, \cdot) = \begin{cases} -\sqrt{-u} & \text{for } u \leq 0 \\ \infty & \text{otherwise} \end{cases}$$

so the requirement $0 \in \mathrm{int}(\mathrm{dom}\, p)$ of the Strong Duality Theorem 8.5 doesn't hold. The corresponding Lagrangian has

$$l(x,y) = \begin{cases} x + yx^2 & \text{for } y \geq 0 \\ -\infty & \text{otherwise} \end{cases}$$

by Example 8.1 and the dual objective function has

$$\psi(y) = \begin{cases} -1/(4y) & \text{for } y > 0 \\ -\infty & \text{otherwise.} \end{cases}$$

Consequently, $p(0) = \sup \psi = 0$ so strong duality holds, but the Slater condition fails. $\qquad \square$

The requirement $0 \in \mathrm{int}(\mathrm{dom}\, p)$ in the Strong Duality Theorem 8.5 is ensured when $f(u, \cdot)$ approximates $f(0, \cdot)$ in a certain sense for u near zero. This is formalized in the next statements, which highlight the role of suitable approximations to ensure strong duality; see [38, Theorem 5.49, Corollary 5.50].

8.7 Theorem (strong duality from epi-convergence). *For the problem of minimizing $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$ and a proper Rockafellian $f : \mathbb{R}^m \times \mathbb{R}^n \to \overline{\mathbb{R}}$ with anchor 0, suppose that there are $u^\nu \to 0$ and $y^\nu \in \mathbb{R}^m$ such that the following hold:*

(a) *$f(u^\nu, \cdot) \xrightarrow{e} f(0, \cdot)$*

(b) *there is a compact set B such that $B \cap \mathrm{argmin}\, f(u^\nu, \cdot)$ is nonempty for all ν.*

(c) *$\liminf \langle y^\nu, u^\nu \rangle \leq 0$*

(d) *$\inf f(u^\nu, \cdot) = \sup \psi^\nu = \psi^\nu(y^\nu)$,*

where

$$\psi^\nu(y) = \inf_x l^\nu(x, y) \quad \text{and} \quad l^\nu(x,y) = \inf_u \{ f(u^\nu + u, x) - \langle y, u \rangle \}.$$

Let ψ be the dual objective function produced by f via (8.3). Then,

$$\inf \varphi = \sup \psi > -\infty.$$

The theorem shows that if we can construct perturbed functions via a Rockafellian and they epi-convergence to the actual objective function as well as possess a strong duality property, then the resulting dual problem indeed reproduces the minimum value of the actual problem provided that assumption (c) also holds.

Since $u^\nu \to 0$, assumption (c) certainly holds when $\{y^\nu, \nu \in \mathbb{N}\}$ is bounded. One can view ψ^ν as a dual objective function produced by the Rockafellian of the form $f^\nu(u, x) = f(u^\nu + u, x)$. The vector y^ν then solves the dual problem associated with f^ν. Thus, it's plausible that $\{y^\nu, \nu \in \mathbb{N}\}$ could be bounded.

32

In some cases, assumption (c) is automatic even when $\{y^\nu, \nu \in \mathbb{N}\}$ is unbounded. For example, if

$$f(u, x) = f_0(x) + \iota_{(-\infty, 0]^m}\big(F(x) + u\big)$$

for $f_0 : \mathbb{R}^n \to \mathbb{R}$ and $F : \mathbb{R}^n \to \mathbb{R}^m$, both smooth, then

$$\psi^\nu(y) = \inf_{x \in \mathbb{R}^n} f_0(x) + \big\langle F(x) + u^\nu, y \big\rangle - \iota_{[0, \infty)^m}(y)$$

by Proposition 8.2. Thus, a maximizer y^ν of ψ^ν is necessarily nonnegative. We can then choose $u^\nu \leq 0$ so that $\langle y^\nu, u^\nu \rangle \leq 0$ and $f(u^\nu, \cdot) \xrightarrow{e} f(0, \cdot)$, which can be seen by working directly from Proposition 3.3.

The theorem isn't restricted to any particular type of Rockafellian. Still, in the convex case, several aspects simplify.

8.8 Corollary *For the problem of minimizing $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$ and a proper, lsc and convex Rockafellian $f : \mathbb{R}^m \times \mathbb{R}^n \to \overline{\mathbb{R}}$ with anchor 0, suppose that there's a compact set $B \subset \mathbb{R}^n$ and $u^\nu \in \operatorname{int}(\operatorname{dom} p) \to 0$ such that $f(u^\nu, \cdot) \xrightarrow{e} f(0, \cdot)$ and $B \cap \operatorname{argmin} f(u^\nu, \cdot)$ is nonempty for all ν, where p is the minimum value function given by $p(u) = \inf f(u, \cdot)$. Let ψ be the dual objective function produced by f via (8.3).*
If $\inf \varphi < \infty$, then

$$\inf \varphi = \sup \psi$$

and this value is finite.

8.9 Example (Slater constraint qualification; cont.). *The corollary confirms the strong duality assertion towards the end of Example 8.6 even though the Slater condition fails.*

Detail. In this case, the Rockafellian, given by $f(u, x) = x + \iota_{(-\infty, 0]}(x^2 + u)$, is proper, lsc and convex. Moreover, $p(u) = -\sqrt{-u}$ for $u \leq 0$ and $p(u) = \infty$ otherwise. Thus, one can take $u^\nu = -1/\nu$ in the corollary and then $f(u^\nu, \cdot) \xrightarrow{e} f(0, \cdot)$; the compactness condition holds since $\operatorname{dom} f(u^\nu, \cdot) \subset \operatorname{dom} f(u^1, \cdot)$. $\square$

References

[1] R.K. Ahuja, T.L. Magnanti, and J.B. Orlin. *Network Flows: Theory, Algorithms, and Applications.* Pearson, 1993.

[2] A. Y. Aravkin, J. V. Burke, and M. P. Friedlander. Variational properties of value functions. *SIAM J. Optimization*, 23(3):1689–1717, 2013.

[3] H. Attouch and R. J-B Wets. Quantitative stability of variational systems: I. The epigraphical distance. *Transactions of the American Mathematical Society*, 328(2):695–729, 1991.

[4] H. Attouch and R. J-B Wets. Quantitative stability of variational systems: III. ε-approximate solutions. *Mathematical Programming*, 61:197–214, 1993.

[5] J.-P. Aubin and I. Ekeland. *Applied Nonlinear Analysis*. Issue 1237 of Pure and Applied Mathematics. Wiley, 1984.

[6] D. Aze. A survey on error bounds for lower semicontinuous functions. In *Proceedings of 2003 MODESMAI Conference, ESAIM Proc., vol. 13. EDP Sci., Les Ulis (2003)*, pages 1—-17, 2003.

[7] H.H. Bauschke and P.L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, 2011.

[8] J. F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer, 2000.

[9] G. G. Brown and M. W. Carlyle. Optimizing the US Navy's combat logistics force. *Naval Research Logistics*, 55:800–810, 2008.

[10] W. M. Carlyle, J. O. Royset, and R. K. Wood. Lagrangian relaxation and enumeration for solving constrained shortest-path problems. *Networks*, 52(4):256–270, 2008.

[11] D. Drusvyatskiy and A. S. Lewis. Tilt stability, uniform quadratic growth, and strong metric regularity of the subdifferential. *SIAM J. Optimization*, 23:256—-267, 2013.

[12] A. C. Eberhard and R. Wenczel. A study of tilt-stable optimality and sufficient conditions. *Nonlinear Analysis*, 75:1260—-1281, 2012.

[13] D. Gale. A geometric duality theorem with economic applications. *The Review of Economic Studies*, 34:19–24, 1967.

[14] H. Gfrerer and B. S. Mordukhovich. Complete characterizations of tilt stability in nonlinear programming under weakest qualification conditions. *SIAM J. Optimization*, 25:2081–2119, 2015.

[15] W. W. Hager and D. Mico-Umutesi. Error estimation in nonlinear optimization. *J. Global Optimization*, 59:327—-341, 2014.

[16] A. D. Ioffe and J. V. Outrata. On metric and calmness qualification conditions in subdifferential calculus. *Set-Valued and Variational Analysis*, 16(2-3):199—-227, 2008.

[17] D. Klatte, A. Kruger, and B. Kummer. From convergence principles to stability and optimality conditions. *J. Convex Analysis*, 19(4):1043–1072, 2012.

[18] A. S. Lewis and S. Zhang. Partial smoothness, tilt stability, and generalized Hessians. *SIAM J. Optimization*, 23:74—-94, 2013.

[19] K. Miettinen. *Nonlinear Multiobjective Optimization*. Springer, 1998.

[20] B. S. Mordukhovich. *Variational Analysis and Applications*. Springer, 2018.

[21] B. S. Mordukhovich and N.M. Nam. Variational stability and marginal functions via generalized differentiation. *Mathematics of Operations Research*, 30:800–816, 2005.

[22] B. S. Mordukhovich, N.M. Nam, and N. D. Yen. Subgradients of marginal functions in parametric mathematical programming. *Mathematical Programming*, 116:396–369, 2009.

[23] B. S. Mordukhovich, R. T. Rockafellar, and M. E. Sarabi. Characterizations of full stability in constrained optimization. *SIAM J. Optimization*, 23:1810–1849, 2013.

[24] J.-S. Pang. Error bounds in mathematical programming. *Mathematical Programming B*, 79(1-3):299—332, 1997.

[25] J. P. Penot. Error bounds, calmness and their applications in nonsmooth analysis. *Contemporary Mathematics*, 514:225–247, 2010.

[26] R. A. Poliquin and R. T. Rockafellar. Tilt stability of a local minimum. *SIAM J. Optimization*, 8:287–299, 1998.

[27] R. T. Rockafellar. *Convex Functions and Dual Extremum Problems*. PhD thesis, Harvard University, 1963.

[28] R. T. Rockafellar. *Convex Analysis*. Vol. 28 of Princeton Math. Series. Princeton University Press, 1970.

[29] R. T. Rockafellar. *Conjugate Duality and Optimization*. Conference Board on Mathematical Sciences Series, 16. SIAM Publishers, 1974.

[30] R. T. Rockafellar. Extensions of subgradient calculus with applications to optimization. *Nonlinear Analysis: Theory, Methods & Applications*, 9:665–698, 1985.

[31] R. T. Rockafellar. Extended nonlinear programming. In G. Di Pillo and F. Giannessi, editors, *Nonlinear Optimization and Related Topics*, pages 381–399. Kluwer, 1999.

[32] R. T. Rockafellar and J. O. Royset. Engineering decisions under risk-averseness. *ASCE-ASME J. Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 1(2):04015003, 2015.

[33] R. T. Rockafellar and S. Uryasev. Optimization of conditional Value-at-Risk. *Journal of Risk*, 2:493–517, 2000.

[34] R.T. Rockafellar and R. J-B Wets. *Variational Analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaft*. Springer, 3rd printing-2009 edition, 1998.

[35] J. O. Royset. Approximations and solution estimates in optimization. *Mathematical Programming*, 170(2):479–506, 2018.

[36] J. O. Royset. Stability and error analysis for optimization and generalized equations. *SIAM J. Optimization*, 30(1):752–780, 2020.

[37] J. O. Royset, W. M. Carlyle, and R. K. Wood. Routing military aircraft with a constrained shortest-path algorithm. *Military Operations Research*, 14(3):31–52, 2009.

[38] J. O. Royset and R. J-B Wets. *An Optimization Primer.* Springer, in press, 2021.

[39] J. J. Ye and X. Y. Ye. Necessary optimality conditions for optimization problems with variational inequality constraints. *Mathematics of Operations Research*, 22(4):977—-997, 1997.

[40] C. Zalinescu. *Convex Analysis in General Vector Spaces.* World Scientific, 2002.