Article

Variables Reduction in Sequential Resource Allocation Problems

Juri Hinz [1],* and Tiziano Vargiolu [2]

[1] School of Mathematical and Physical Sciences, University of Technology Sydney, P.O. Box 123, Broadway, NSW 2007, Australia

[2] Department of Mathematics "Tullio Levi-Civita", University of Padova, via Trieste 63, 35131 Padova, Italy; vargiolu@math.unipd.it

* Correspondence: juri.hinz@uts.edu.au

check for updates

Abstract: This paper presents a general framework to address diverse notoriously difficult problems arising in the area of optimal resource management, exploitation of natural reserves, pension fund valuation, environmental protection, and storage operation. Using some common abstract features of this problem class, we present a technique which provides a significant reduction of decision variables. As an application, we discuss a battery storage control to show how a decision problem, which is practically unsolvable in the original formulation, can be treated by our method.

Keywords: dynamic programming; battery control; storage valuation

1. Introduction

Optimization of sequential decision-making under uncertainty arises in different fields, such as finance, economics, robotics, manufacturing, and telecommunication. These problems are frequently discussed under the framework of real options, as applications of optimal stochastic control. Therefore, the real options viewpoint (see [1,2]) highlights a freedom in the choice of decisions and consistently uses the operational flexibility for strategy optimization. The classical contributions [3,4] were among the first, placing sequential decision optimization into *real option* context, with applications in manufacturing [5], investment planning [6], mining [7], and commodities [8]. Therefore, a formulation of decision problems in continuous time was preferred: using a jump-diffusion setting and its well-established stochastic control toolbox, the so-called Hamilton-Jacobi Bellman (HJB) equations, (see [9]) and the Backward Stochastic Differential Equations (BSDE) (see [10]) became applicable [11–13]. However, also using discrete-time modeling, sequential decision problems have been routinely treated in terms of the *Markov Decision Theory* [14,15] whose approximate techniques are known as *Approximate Dynamic Programming*, (see [16], which comprises diverse heuristic numerical approaches).

In this paper we discuss a specific class of discrete-time stochastic control problems from the viewpoint of real options, and use a certain managerial flexibility which can be expressed in terms of a virtual reserve price. We show that in the optimal regime at any time, the actions are connected to each other (in a certain sense) via virtual reserve price. This observation can be used to significantly reduce the range of actions available for decision choice by excluding at each step those which cannot be optimal. More specifically, we consider an abstract but generic situation which frequently arises in sequential decision-making, when a certain activity (production plan) must be optimally chosen to meet the right balance between the current costs of the activity and the consumption of some resource. While all activity costs have an immediate effect, the impact of resource consumption is uncertain and becomes relevant in the future. In practice, one is frequently confronted with a huge range of possibilities to choose activities, and their combinations form a complex space whose structure is

usually determined by numerous inter-relations. Although a solution to such sequential decision problem can be theoretically obtained in terms of the standard backward induction, its practical implementation is virtually infeasible due to the high complexity of decisions.

It turns out that a certain transformation can be of great help in such situations. In particular, under specific but natural assumptions, we show how the decision variables space can be reduced to a simple one-parameter family. Such reduction is achieved by a solution to a separate deterministic optimization problem which is usually easily obtainable. Using such re-formulation, diverse numerical techniques for backward induction can be applied. In this context, in our particular situations, the Bellman's optimality principle turns out to be equivalent to certain fixed-point equations, which might lead to alternative ways to further simplify computational efforts in the backward induction.

The paper is organized as follows. Section 2 reviews dynamical Bellman principle underlying the classical and stochastic dynamic programing and explains how our method is placed within this context and its use in practice. Section 3 presents the problem and our approach to variables reduction within a general framework. Section 4 introduces an application to battery control which is reviewed in Section 5 to obtain a variables reduction for battery storages. Section 6 is devoted to formal justification of our techniques, while Section 7 concludes.

2. Dynamic Principle for Optimal Switching and Reduction of Decision Variables

In applications, sequential decision-making is usually addressed under the framework of discrete-time *Stochastic Control*. The theory of *Markov Decision Processes/Dynamic Programming* provides a variety of methods to deal with such questions. In generic situations, approaching analytical solutions for even some simplest decision processes may be a cumbersome process ([10,14,16]) Furthermore, since closed-form solutions to practically important control problems are usually unavailable, numerical approximations became popular among practitioners to obtain approximately optimal control policies. Although a huge variety of computational methods have been developed therefore, typical real-world problems are usually too complex for the existing solution techniques, in particular if the state dimension of the underlying controlled evolution is high. Let us review the finite-horizon Markov decision theory following [17].

Controlled Markov processes: Consider a random dynamics on a finite time horizon $0, \dots, T$ whose state x evolves in E and is controlled by actions a from an action set A. For each $a \in A$, we assume that $K_t^a(x, dx')$ is a stochastic transition kernel on E. A mapping $\pi_t : E \mapsto A$ which describes the action that the controller takes at time t is called a *decision rule*. A sequence of decision rules $\pi = (\pi_t)_{t=0}^{T-1}$ is called a *policy*. For each initial point $x_0 \in E$ and each policy $\pi = (\pi_t)_{t=0}^{T-1}$, there exists a probability measure $\mathbb{P}^{x_0, \pi}$ and a stochastic process $(X_t)_{t=0}^T$ satisfying the initial condition $\mathbb{P}^{x_0, \pi}(X_0 = x_0) = 1$ such that

$$\mathbb{P}^{x_0, \pi}(X_{t+1} \in B \mid X_0, \dots, X_t) = K_t^{\pi_t(X_t)}(X_t, B) \tag{1}$$

holds for each $B \subset E$ at all times $t = 0, \dots, T - 1$, i.e., given that system is in state X_t at time t, the action $a = \pi_t(X_t)$ is used to pick the transition probability $K_t^{a = \pi_t(X_t)}(X_t, \cdot)$ which randomly drives the system from X_t to X_{t+1} with the distribution $K_t^{\pi_t(X_t)}(X_t, \cdot)$. Let us use $\mathcal{K}_t^a$ to denote the one-step transition operator associated with the transition kernel K_t^a when the action $a \in A$ is chosen. In other words, for each action $a \in A$ the operator $\mathcal{K}_t^a$ acts on functions v by

$$(\mathcal{K}_t^a v)(x) = \int_E v(x') K_t^a(x, dx') \qquad x \in E, \tag{2}$$

whenever the above integrals are well-defined.

Costs of control: For each time t, we are given the *t-step reward function* $r_t : E \times A \mapsto \mathbb{R}$, where $r_t(x, a)$ represents the reward for applying an action $a \in A$ when the state of the system is $x \in E$ at time t. At

the end of the time horizon, at time T, it is assumed that no action can be taken. Here, if the system is in a state x, a *scrap value* $r_T(x)$, which is described by a pre-specified *scrap function* $r_T : E \to \mathbb{R}$, is collected. Given an initial point x_0, the goal is to maximize the expected finite-horizon total reward, in other words to find the argument $\pi^* = (\pi_t^*)_{t=0}^{T-1}$ such that

$$\pi^* = \operatorname{argmax}_\pi \mathbb{E}^{x_0, \pi} \left(\sum_{t=0}^{T-1} r_t(X_t, \pi_t(X_t)) + r_T(X_T) \right), \tag{3}$$

where A is the set of all policies, and $\mathbb{E}^{x_0, \pi}$ denotes the expectation over the controlled Markov chain defined by (1).

Decision optimization: The maximization (3) is well-defined under diverse additional assumptions (see [14], p. 199). The calculation of the optimal policy is addressed in the following setting. For $t = 0, \ldots, T - 1$, introduce the *Bellman operator*

$$\mathcal{T}_t v(x) = \sup_{a \in A} \left(r_t(x, a) + \mathcal{K}_t^a v(x) \right), \quad x \in E \tag{4}$$

which acts on each measurable function $v : E \to \mathbb{R}$ where the integrals $\mathcal{K}_t^a v$ for all $a \in A$ exist. Furthermore, consider the *Bellman recursion*

$$v_T^* = r_T, \quad v_t^* = \mathcal{T}_t v_{t+1}^* \qquad \text{for } t = T - 1, \ldots, 0. \tag{5}$$

Under appropriate assumptions, there exists a recursive solution $(v_t^*)_{t=0}^T$ to the Bellman recursion, which gives the so-called *value functions* and determines an optimal policy π^* via

$$\pi_t^*(x) = \operatorname{argmax}_{a \in A} \left(r_t(x, a) + \mathcal{K}_t^a v_{t+1}^*(x) \right), \quad x \in E, \quad t = 0, \ldots, T - 1$$

Stochastic switching: Consider now a Markov decision model whose state evolution consists of one controllable and one uncontrollable component. To be more specific, we assume that the state space $E = P \times \mathcal{Z}$ is the product of a space P (operation modes) and a set $\mathcal{Z}$ (states of environment) being a subset $\mathcal{Z} \subset \mathbb{R}^d$ of the Euclidean space. We suppose that at each time $t = 0, \ldots T - 1$ the mode component $p \in P$ is driven by actions $a \in A$ in terms of a deterministic function

$$\alpha_t : P \times \mathcal{Z} \times A \to P, \quad (p, z, a) \to \alpha_t^{p,z}(a)$$

where $\alpha_t^{p,z}(a) \in P$ stands for the new mode from if the action $a \in A$ was taken at time $t = 0, \ldots, T - 1$ in the state $(p, z) \in E$. In this setting, the transition operators are given by

$$\mathcal{K}_t^a v(p, z) = \mathbb{E}(v(\alpha_t^{p,z}(a), Z_{t+1}) \mid Z_t = z), \quad z \in \mathcal{Z} \tag{6}$$

for $t = 0, \ldots, T - 1$, and $a \in A$.

Variables reduction: In the above context of stochastic switching, the application of Bellman operator

$$\mathcal{T}_t v(p, z) = \sup_{a \in A} \left(r_t(p, z, a) + \mathbb{E}(v(\alpha_t^{p,z}(a), Z_{t+1}) \mid Z_t = z) \right), \quad (p, z) \in E \tag{7}$$

may cause numerical difficulties, particularly if the action space A is high-dimensional and fragmented. Unfortunately, this situation appears frequently in practice, where high-dimensional vectors of decision variables usually subject to numerous feasibility inter-dependencies. In such framework, the problem may become unsolvable due to difficulty of maximization over a complex action set A. The main contribution of our paper is to provide a significant reduction of set of actions, which are relevant

for maximization. Under specific assumptions described in this paper, we show there exists a one-parameter family (curve) $A_t^{p,z} \subset A$ which depends on the recent time $t = 0, \ldots, T = 1$ and state $(p, z) \in E$ such that the domain of maximization reduces from A to $A_t^{p,z}$ which yields instead of the infeasible maximization (7) a new problem

$$\mathcal{T}_t v(p, z) = \sup_{a \in A_t^{p,z}} \left(r_t(p, z, a) + \mathbb{E}(v(\alpha_t^{p,z}(a), Z_{t+1}) \mid Z_t = z) \right), \quad (p, z) \in E \tag{8}$$

which is significantly simpler and usually admits a (numerical) solution. To determine the curve $A_t^{p,z} \subset A$ of relevant actions, a separate deterministic optimization problem must be solved. Its solution is usually obtained explicitly and provides interesting economic insights.

Contribution of this work: Using a standard Bellman principle, we explore an abstract, but natural framework to reduce a potentially very large space of decision variables (actions) to a *single one-parameter* family. Although our technique is entirely placed within the traditional Bellman principle of stochastic/classic dynamic programming and addresses a relatively narrow problem class, a wide range of application is covered, including mining operations, pension fund management, and emission control. This approach can serve efficient numerical algorithms for notoriously difficult and important problems from practice.

3. Optimal Control via Virtual Resource Price

Let us introduce the required framework more precisely. Consider an agent who is confronted with the following problem. At the beginning $t = 0, 1 \ldots, T - 1$ of each decision epoch, an activity plan (work schedule) ξ_t is to be determined. Therefore, all costs of this work plan must be optimally balanced against their resource consumption/generation. Suppose that the limited resources are described in terms of the state variable $e \in I$ which stands for the current resource shortage and can vary within a certain interval $I \subset \mathbb{R}$. For instance, if the resource under consideration is a commodity in the storage, then $e \in I$ stands for the amount of commodity required to fill the storage to its maximal capacity. The other state variable $z \in \mathbb{R}^d$ is supposed to represent the situation in the surrounding environment. Let us agree that this environmental state variable z is relevant for decisions, but cannot be influenced by agent's actions. For instance, for commodity storage, z may comprise the driving market factors of the commodity price evolution. We furtherly assume that the environment state occurs at any time $t = 0, \ldots, T$ as a realization $z = Z_t$ of an $\mathbb{R}^d$-valued Markov process $(Z_t)_{t=0}^T$ whose dynamics carries all information, relevant for decisions.

Having observed at time $t = 0, \ldots, T - 1$ the resource level $e \in I$ and the realization of the environmental state variable $z = Z_t$, the agent selects a plan $\xi \in \Xi$ from the set Ξ of all feasible activity plans. This choice yields an immediate cost $C_t^{e,z}(\xi)$ and causes an immediate resource consumption $E_t^{e,z}(\xi)$ via pre-specified functions $C_t^{e,z}$, $E_t^{e,z}$ on Ξ which may depend on the recent state $(e, z) \in I \times \mathbb{R}^d$. While all costs are accumulated, the resource level is carried over to the next decision time $t + 1$ as $e + E_t^{e,z}(\xi)$ and will influence the decision at this time. The availability of resources becomes crucial at the end $t = T$ of the planning horizon, when a certain terminal costs $C_T(e, z)$ must be paid, which depend on the total resource level $e \in I$ and on the state $z \in \mathbb{R}^d$ of the environment. Under additional assumptions, such control problems are solved in terms of the so-called value functions $(V_t)_{t=0}^T$ which are obtained via Bellman recursions as

$$V_T(e, z) = C_T(e, z) \tag{9}$$

$$\tilde{V}_{t+1}(e, z) = \begin{cases} \mathbb{E}(V_{t+1}(e, Z_{t+1}) \mid Z_t = z) & e \in I \\ +\infty & e \notin I \end{cases}, z \in \mathbb{R}^d \tag{10}$$

$$V_t(e, z) = \inf_{\xi \in \Xi} \left(C_t^{e,z}(\xi) + \tilde{V}_{t+1}(e + E_t^{e,z}(\xi), z) \right), \tag{11}$$

for all arguments $(e,z) \in I \times \mathbb{R}^d$ representing the current situation at decision time. Please note that we implicitly agreed to penalize the violation of the admissible resource level in (10) by infinity, gaining the freedom to interpret the value function for all arguments $e \in \mathbb{R}$ with possible values $+\infty$ in order to avoid watching restriction $e + E_t^{e,z}(\xi) \in I$ in (11).

The idea underlying our variable reduction scheme is based on the realization that if the resources had a market price, then an agent would minimize all activity costs taking into account the monetary value of the consumed resources. To realize this concept, we suppose that some entity (a regulatory body) could charge a resources price $A \in \mathbb{R}$ at any time t, depending on the situation (e,z). In the presence of such virtual price $A \in \mathbb{R}$, each the decision-maker would examine the virtual charges for the resource consumption and chose an activity accordingly, by obtaining

$$\mathcal{X}_t^{e,z}(A), \text{ a minimizer of } \xi \mapsto C_t^{e,z}(\xi) + A E_t^{e,z}(\xi) \text{ over } \xi \in \Xi. \tag{12}$$

Obviously, such mapping $A \mapsto \mathcal{X}_t^{e,z}(A)$ represents an optimal activity depending on the current state $(e,z) \in I \times \mathbb{R}^d$, time $t = 0, \ldots, T-1$, given the resource price $A \in \mathbb{R}$. To some degree, this mapping can be interpreted as the willingness to save resources by following a less profitable strategy in response to an increased value of the resource. For ease of understanding, let us postpone the discussion concerning the existence of the minimizer in (12) and the properties of the relation $A \mapsto \mathcal{X}_t^{e,z}(A)$ which are crucial for the targeted results. The important assumption for now is that for each state (e,z) there exists a bounded interval $\mathcal{A}^{e,z} \subset \mathbb{R}$ such that the one-parameter family of activities

$$\xi \in \{\mathcal{X}_t^{e,z}(A) \in \Xi \mid A \in \mathcal{A}^{e,z}\} \subset \Xi \tag{13}$$

contains only the "best candidates" (for the purpose of minimization in (11)). They can be used in the Bellman recursion, replacing the minimization in (11) by (16) as follows

$$v_T(e,z) = C_T(e,z), \tag{14}$$

$$\tilde{v}_{t+1}(e,z) = \begin{cases} \mathbb{E}(v_{t+1}(e, Z_{t+1}) \mid Z_t = z) & e \in I \\ +\infty & e \notin I \end{cases} \quad z \in \mathbb{R}^d, \tag{15}$$

$$v_t(e,z) = \min_{A \in \mathcal{A}^{e,z}} \left(C_t^{e,z}(\mathcal{X}_t^{e,z}(A)) + \tilde{v}_{t+1}(e + E_t^{e,z}(\mathcal{X}_t^{e,z}(A)), z) \right). \tag{16}$$

Please note that now the minimization in (16) must be performed merely over a curve (13) instead of over the whole space Ξ as in (11). In practical applications, such reduction can provide a reasonable (numerical) approach to a virtually unsolvable problem. The main question here is whether

$$\begin{array}{c} \text{the value functions } (V_t)_{t=0}^T \text{ from (9)–(11)} \\ \text{coincide with those } (v_t)_{t=0}^T \text{ from (14)–(16).} \end{array} \tag{17}$$

In the following, we work out all conditions required therefore. Before we turn to the illustration of our technique by battery storage management in Sections 4 and 5, followed by proofs in Section 6, let us present all required assumptions which ensure the validity of the assertion (17).

Suppose that all information, relevant for decision-making, is carried by $\mathbb{R}^d$-valued Markov process $(Z_t)_{t=0}^T$ which is realized on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t=0}^T)$. As introduced above, we assume that the resource level can vary within a certain bounded interval $I \subset \mathbb{R}$. Having selected an activity ξ from the set Ξ of all feasible activity plans at time $t = 0, \ldots, T-1$ in the state $(e,z) \in I \times \mathbb{R}^d$, the costs $C_t^{e,z}(\xi)$ and resource consumption $E_t^{e,z}(\xi)$ are determined via pre-specified functions

$$C_t^{e,z}, E_t^{e,z} : \Xi \to \mathbb{R}, \qquad t = 0, \ldots, T-1, \quad (e,z) \in I \times \mathbb{R}^d, \tag{18}$$

while the terminal costs are determined by

$$C_T : I \times \mathbb{R}^d \to \mathbb{R}, \quad (e, z) \mapsto C_T(e, z). \tag{19}$$

All our considerations rely on additional assumptions on the functions (18) which we formulate next. To ensure that the minimization in (11) is well-defined, let us agree that there exists an *idle* activity which does not consume any resources:

$$\left. \begin{array}{l} \text{for each } (e, z) \in I \times \mathbb{R}^d \text{ there} \\ \text{exits } \xi \in \Xi \text{ such that } E_t^{e,z}(\xi) = 0. \end{array} \right\} \tag{20}$$

To ease our argumentation, we suppose that

$$\left. \begin{array}{l} \text{for each } (e, z) \in I \times \mathbb{R}^d \text{ there is an interval } \mathcal{A}^{e,z} \subset \mathbb{R} \text{ such that} \\ \text{the minimizer } \mathcal{X}_t^{e,z}(A) \in \Xi \text{ to (12) exists for each } A \in \mathcal{A}^{e,z}. \end{array} \right\} \tag{21}$$

Furthermore, let us propose a mild technical assumption

$$\left. \begin{array}{l} \text{for each } z \in \mathbb{R}^d, \xi' \in \Xi \text{ and } e, e' \in I \text{ with } e < e' \text{ there exists} \\ \xi \in \Xi \text{ reaching the same level } e + E_t^{e,z}(\xi) = e' + E_t^{e',z}(\xi') \text{ of} \\ \text{resource consumption at no greater costs } C_t^{e,z}(\xi) \leq C_t^{e',z}(\xi'). \end{array} \right\} \tag{22}$$

To determine the desired minimizer $\xi^* = \mathcal{X}_t^{e,z}(A^*)$ to (11), we rely on the following natural technical assumption

$$\left. \begin{array}{l} \text{for each } (e, z) \in I \times \mathbb{R}^d \text{ and } t = 0, \dots, T - 1, \\ \text{the function } A \mapsto E_t^{e,z}(\mathcal{X}_t^{e,z}(A)) \text{ is continuous on} \\ \mathcal{A}^{e,z}, \text{ strictly decreasing, and possesses a root.} \end{array} \right\} \tag{23}$$

Finally, we require some convexity properties in the sense that

$$\left. \begin{array}{l} \text{the set } \Xi \text{ is convex and for each } z \in \mathbb{R}^d, t = 0, \dots, T - 1 \text{ the} \\ \text{functions } (e, \xi) \mapsto C_t^{e,z}(\xi), (e, \xi) \mapsto E_t^{e,z}(\xi) \text{ are convex on } I \times \Xi, \\ \text{furthermore } e \mapsto C_T(e, z) \text{ is convex and non-decreasing on } I. \end{array} \right\} \tag{24}$$

As mentioned above, the advantage of our approach is that a simpler form (14)–(16) of Bellman recursion occurs, whose efficient (numerical) treatment may be easily possible, unlike that of the original problem (9)–(11). Indeed, it turns out that such simplification solves the original problem in the following sense:

Theorem 1. *Under the assumptions (20)–(24) consider $(V_t)_{t=0}^T$ and $(v_t)_{t=0}^T$ defined by (9)–(11) and (14)–(16) respectively. If a solution $(v_t)_{t=0}^T$ to (14)–(16) exists, then*

$$\textit{the value functions coincide } v_t = V_t \textit{ for all } t = 0, \dots, T. \tag{25}$$

In view of the above result, the practical solution now requires obtaining for each $t = 0, \dots, T - 1$ the optimal decision

$$\pi_t^* : I \times \mathbb{R}^d \to \mathbb{R} \tag{26}$$

via minimization in the simplified Bellman recursion where the virtual price for the optimal regime is obtained via $\pi_t^*(e, z) = A^*$ as a minimizer

$$A^* = \text{argmin}_{A \in \mathcal{A}^{e,z}} \left(C_t^{e,z}(\mathcal{X}_t^{e,z}(A)) + \tilde{V}_{t+1}(e + E_t^{e,z}(\mathcal{X}_t^{e,z}(A)), z) \right). \tag{27}$$

Remark 1. *We also prove that the optimal decision $\pi_t^*(e, z) = A^*$ can be obtained as a solution to the fixed-point equation*

$$A^* \in \nabla \tilde{V}_{t+1}(e + E_t^{e,z}(\mathcal{X}_t^{e,z}(A^*)), z),$$

meaning that A^* must be a sub-gradient of the expected value function, i.e., in the optimal regime, the virtual resource price must always be equal to the marginal change rate of the value function with respect to the resource level. The economic interpretation of this insight is natural: when choosing an activity plan, the agent increases the resource consumption to a level where the value loss caused by this consumption starts taking over the instant revenue of the activity.

4. Battery Storage Control

To illustrate our variables reduction technique, we introduce a model for battery storage installation operated within a deregulated electricity market. First, let us introduce some technical characteristics. The battery *capacity* stands for the maximal energy stored (measured in MWh), whereas the battery power is the amount of *electrical power* (measured in MW) the installation can provide at any moment (see [18]). The conditions under which batteries are operated affect their performance in terms of the so-called *cycle life*, which can be defined as the number of cycles completed before the effective battery capacity falls below than 60% of its nominal capacity. Therefore, the Depth of Discharge (DoD) is essential. To give the reader a quantitative understanding of this phenomena, let us consider an example of a *lithium ion* battery. Assuming that each charge/discharge cycle causes a battery deterioration, one would assume that emptying completely a battery (which corresponds to 100% DoD) 200 times is roughly equivalent (in terms of performance decline) to 400 cycles at 50% DoD and 600 cycles at 33% DoD. However, the actual behavior is different. Usually, a lithium-type battery serves longer than 400 cycles at 50% DoD and significantly longer than 600 cycles at 33% DoD, (see [19]) In our model, we suppose that visiting deep discharge states is costly since it affects battery life. For this reason, we suggest including a user-defined function which penalizes deep discharge states accordingly. Figure 1 illustrates its dependence measured in charge cycles and dependence on depth of discharge. We include such effects into our model by a cost penalization of deep discharge states.
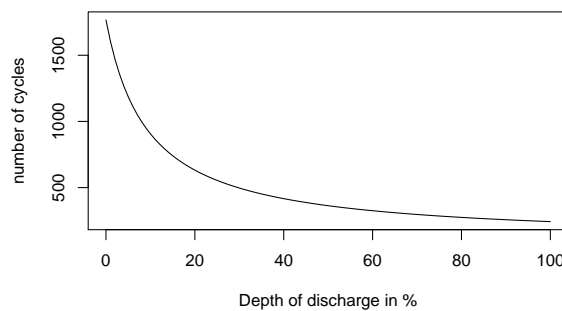
Figure 1. Typical relationship between battery life and depth of discharge.

Beyond preventing deep discharge, further operational improvements encompass avoiding that the battery is fully charged (by reducing voltage when charging) and diminishing the so-called charge/discharge *rate*, which stands for the maximal electric flow during charging/discharging.

We assume that the agent attempts to optimally manage an electricity storage by sequential decisions on the amount of energy procured/purchased through the intraday market and on the

optimal charge/discharge of batteries. Obviously, these decisions must take into account the current electricity price, market state, storage level, and all costs and technical restrictions.

Consider an energy retailer facing the obligation to satisfy an unknown energy demand of its customers at times $t = 0, \ldots, T - 1$, while renewable energy sources produce a random electricity amount given a certain battery storage. To manage a potential energy imbalance, appropriate forward positions are taken in advance and decisions are made to drive the battery storage. This is a typical dynamic control problem, since at any time $t = 0, \ldots, T$ an action must be chosen (encompassing a simultaneous energy trade and battery control) which immediately causes some costs but also determines a transition to the next system state (future battery level and market conditions). Clearly, one needs to optimally balance the recent costs against those incurred in the future, based on the current situation. The problems of this type are naturally formulated and solved in terms of dynamic programming.

Remark 2. *Please note that we investigate the problem of electricity storage management within an abstract context which is applicable to most of the deregulated electricity markets, possibly with minor adaptation. Namely we consider energy trading on two time scales. The longer-term trading is realized by forwards or futures or by energy delivery contracts, for instance from the day-ahead trading. In practice, these long-term positions are constantly adjusted on a short-term scale using intraday trading or balancing procedures. This two-scale structure is universal and inherent to any deregulated electricity market and we usually observe significant differences between both scales considering their prices, liquidity, and spread.*

Consider an agent who serves an energy consumer whose random demand is (partially) covered from a renewable energy generation facility whose energy output is also random. We denote by D_t the cumulative electricity demand of such facility (which stands for demand if $D_t < 0$ and surplus if $D_t > 0$) for the delivery period t. Assume that the time point $t = 0, \ldots, T - 1$ corresponds to the beginning of the period t and agree that the demand D_t is observed after t, at the end of the delivery period t. Let us agree to model $D_t = d_t + \varepsilon_t$, with a zero-mean random variable ε_t standing for the deviation of the realized demand D_t from its prediction d_t, which is observable at time t. Moreover, assume that at each time $t = 0, \ldots, T$, the producer can take a forward position which attempts to cover D_t. Let us describe such position as $d_t + f_t$ where f_t stands for the deviation of the total amount traded forward from the prediction d_t. In generic situations, the quantity f_t can be considered to be a "safety margin" which must be purchased on the top of predicted demand d_t to avoid a potential energy shortage during delivery interval. However, we do not assume that f_t or d_t must always be positive. Moreover, introduce the control variable b_t, standing for the decision to transfer the energy amount $|b_t|$ from/to the battery, where $b_t > 0$ and $b_t < 0$ represent discharging and charging actions, respectively. Here, we agree that these control actions must be decided at time t (immediately before the delivery period t starts). With these assumptions, the energy to be balanced during delivery period $t = 0, \ldots, T$ using electricity grid is

$$f_t + d_t - D_t + b_t = f_t + b_t - \varepsilon_t. \tag{28}$$

Please note that on the right-hand side of this equation, the quantities f_t and b_t must be chosen at t whereas ε_t becomes observable after time t.

Now we turn to storage control costs and introduce electricity prices

$$\Psi_t = (\Psi_t^+, \Psi_t^0, \Psi_t^-) \qquad t = 0, \ldots, T, \tag{29}$$

with the interpretation that for delivery period t, Ψ_t^0 stands for the forward price, whereas Ψ_t^+ and Ψ_t^- are the so-called upper and the lower balancing prices.

Remark 3. *Please note that Ψ_t^0 stands for the price of energy from long-term market. In the sense of the above remark, this price can represent a forward, futures, or day-ahead price of electrical energy in front of delivery, depending on modeling.*

While the forward price Ψ_t^0 is listed prior to the delivery period t and applies to energy traded in advance, the balancing prices Ψ_t^+, Ψ_t^- are determined during delivery period t and apply for purchase and procurement of the grid energy. Usually, it holds that

$$0 \leq \Psi_t^+ < \Psi_t^0 < \Psi_t^- \qquad t = 0, \ldots, T. \tag{30}$$

In practice, the price range $[\Psi_t^+, \Psi_t^-]$ can be wide, meaning that Ψ_t^+ is significantly lower than Ψ_t^0 which is lower than Ψ_t^-. This issue makes any balancing using grid energy potentially unfavorable. For this reason, the agent attempts to meet the demand as precisely as possible using a combination of the energy from the forward position and from the battery. More specifically, we suppose that the costs, associated with taking forward position, are given by

$$\Psi_t^0(d_t + f_t) + q(d_t + f_t)^2 \tag{31}$$

where $q > 0$ is a coefficient, representing the elasticity of the forward price with respect to contract volume. The total costs associated with energy balancing during the period $t = 0, \ldots, T$ is given by

$$\Psi_t^0(d_t + f_t) + q(d_t + f_t)^2 - \Psi_t^+(f_t + b_t - \varepsilon_t)^+ + \Psi_t^-(f_t + b_t - \varepsilon_t)^-. \tag{32}$$

Please note that this quantity is observable after t and is controlled by the variables f_t and b_t which must be chosen at t.

Let us precisely formulate the assumptions on random variables observables, concerning the time of their observation. Suppose that the processes $(d_t)_{t=0}^T$ $(f_t)_{t=0}^T$, $(b_t)_{t=0}^T$, $(\varepsilon_t)_{t=0}^T$ are given on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t=0}^{T+1})$ where $\mathcal{F}_t$ represents the information available at time point t, just before the start of the delivery period t. According to the above modeling, we suppose that for $t = 0, \ldots, T$

$$d_t, f_t, b_t, \Psi_t^0 \text{ are } \mathcal{F}_t\text{-measurable, and}$$
$$\varepsilon_t, \Psi_t^+, \Psi_t^- \text{ are } \mathcal{F}_{t+1}\text{-measurable.}$$

Let us suppose that

$$\begin{array}{cc} \Psi_t = (\Psi_t^-, \Psi_t^0, \Psi_t^+) \text{ and } \varepsilon_t \text{ are} & \\ \text{conditionally independent, given } \mathcal{F}_t & t = 0, \ldots, T, \end{array} \tag{33}$$

and denote by $\mathbb{E}_t(\cdot)$ the expectation, conditioned on $\mathcal{F}_t$ for $t = 0, \ldots, T$. Applying such conditional expectation $\mathbb{E}_t(\cdot)$ to (32), we use the conditional independence (4) to obtain

$$\bar{\Psi}_t^0(d_t + f_t) + q(d_t + f_t)^2 - \bar{\Psi}_t^+ \mathbb{E}_t[(f_t + b_t - \varepsilon_t)^+] + \bar{\Psi}_t^- \mathbb{E}_t[(f_t + b_t - \varepsilon_t)^-]. \tag{34}$$

In this equation, the prices expected in front of delivery, are denoted by

$$\bar{\Psi}_t^- = \mathbb{E}_t(\Psi_t^-), \quad \bar{\Psi}_t^0 = \Psi_t^0 = \mathbb{E}_t(\Psi_t^0), \quad \bar{\Psi}_t^+ = \mathbb{E}_t(\Psi_t^+), \quad t = 0, \ldots, T.$$

To simplify our energy storage management, we furtherer agree that the distribution of prediction errors does not depend on recent information in the sense that

$$\varepsilon_t \text{ and } \mathcal{F}_t \text{ are independent,} \quad t = 0, \ldots, T. \tag{35}$$

This natural assumption yields a compact form for the expected costs (34):

$$\bar{\Psi}_t^0(d_t + f_t) + q(d_t + f_t)^2 - \bar{\Psi}_t^+ h_t^+(f_t + b_t) + \bar{\Psi}_t^- h_t^-(f_t + b_t) \tag{36}$$

with functions h^+ and h^- explicitly computable from

$$h_t^+(u) = \mathbb{E}[(u - \varepsilon_t)^+], \quad h_t^-(u) = \mathbb{E}[(u - \varepsilon_t)^-], \quad u \in \mathbb{R}, \, t = 0, \dots, T. \tag{37}$$

In view of (34)–(37), the expected costs of (32) depend on the energy control variables f and b as

$$(f, b) \mapsto \Psi_t^0(d_t + f) + q(d_t + f)^2 - \Psi_t^+ h_t^+(f + b) + \Psi_t^- h_t^-(f + b) \tag{38}$$

Please note that these costs are expected at time t and can be changed by appropriate adjustment of decision variables $f, b \in \mathbb{R}$.

For an agent concerned with the minimization of all costs accumulated within the decision period ranging from $t = 0$ to $t = T$, the dynamical aspects are important. Specifically, the decision at time t to use energy b from the storage changes the storage level, which has a distinct impact on the availability of energy in the future, influencing all following decisions.

To formulate our storage control as a dynamic programming problem, we assume that a Markov dynamics $(Z_t)_{t=0}^T$ on $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t=0}^T)$ carries all relevant information. Therefore, we suppose that $(Z_t)_{t=0}^T$ is a Markovian process which takes values in $\mathbb{R}^d$. This process describes the evolution of all relevant state variables of the environment. In particular, we assume that the expected prices are represented by function $\psi : \mathbb{R}^d \to \mathbb{R}^3$ of state variables, whose components $\psi = (\psi_t^-, \psi_t^0, \psi_t^+)$ determine the prices as follows:

$$(\bar{\Psi}_t^-, \bar{\Psi}_t^0, \bar{\Psi}_t^+) = \psi_t(Z_t) = (\psi_t^-(Z_t), \psi_t^0(Z_t), \psi_t^+(Z_t)) \quad t = 0, \dots, T. \tag{39}$$

In accordance to (30), we require for $t = 0, \dots, T$ that

$$0 \le \psi_t^+(z) < \psi_t^0(z) < \psi_t^-(z), \qquad z \in \mathbb{R}^d. \tag{40}$$

Furthermore, we suppose that at any time $t = 0, \dots, T$, the conditional expectation of the next period's demand is described in terms of a deterministic function

$$d : \{0, \dots, T\} \times \mathbb{R}^d \to \mathbb{R}, \qquad (t, z) \mapsto d_t(z).$$

Besides the state of the environment $z \in \mathbb{R}^d$, the other important state variable is the current storage level $e \in I$. Having denoted the minimal and the maximal energy amounts of the battery by $\underline{e}$ and $\bar{e}$ respectively, we suppose that the storage level e represents by the amount e of energy, which is *needed to fully charge the battery*. With this interpretation, our variable is the

$$\text{resource level } e, \text{ which takes values in the interval } I =]0, \bar{e} - \underline{e}[. \tag{41}$$

In view of (34)–(39), the expected costs of (32) are now written in terms of a function

$$\begin{aligned} C_t^{e,z}(f, b) = & \psi_t^0(z)(d_t(z) + f) + q(d_t(z) + f)^2 - \psi_t^+(z)h_t^+(f + b) \\ & + \psi_t^-(z)h_t^-(f + b) + \chi(e) \end{aligned} \tag{42}$$

of the resource level $e \in I$, the state variable $z \in \mathbb{R}^d$, and the control variables $f, b \in \mathbb{R}$ for $t = 0, \dots, T$. Please note that in (42) we include the costs $\chi(e)$ of deep discharge corresponding to the resource level $e \in I$, modeled by an

$$\text{increasing and convex function } \chi : I \to \mathbb{R}. \tag{43}$$

Remark 4. *Ideally, an understanding of the chemistry of a particular battery will suggest some generic penalization function. However, in practice, there are diverse approaches to assess (in economic terms) how the battery life is affected by visiting a deep discharge state. Here, the usual cycle life graph (as in Figure 1), does not detail all aspects required by our approach. The point is that the cycle life is determined by charge/discharge periods oscillating linearly between the minimal and the maximal capacity, rather by a certain strategy run within a random environment. Such life cycle diagrams do not provide any information whether it could be worth visiting a deep discharge state for a short period of time in order to catch an electricity price spike. Given diverse battery storage technologies, there is no simple way of determining an appropriate penalization function. For this reason, the authors suggest a pragmatic approach: For a pre-specified penalization as depicted in Figure 2, the user will calculate the corresponding optimal strategy which must be examined and assessed in simulations. If required by simulation results, the penalty function can be altered, followed by another round of optimization and simulations. Such attempts can be repeated until satisfactory results (considering battery life expectation, peak load response, and total revenue) are reached.*
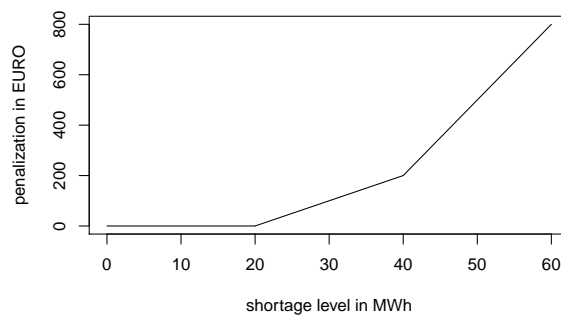
Figure 2. A function as in (43) for 60 MWh, penalizing the middle and upper range of shortage level.

Now, let us propose Bellman recursion for our battery control problem. Recall that in our modeling, the state variables at time $t = 0, \dots, T$ comprise a market situation described by the realization $z = Z_t$ of the environment state process and the current resource level $e \in I$. Having supposed that at the final date $t = T$, in the environment state $z = Z_T$ the entire battery storage content $\bar{e} - e$ can be sold at the market price $\psi_T^0(z)$, we agree that the terminal costs function is

$$C_T(e, z) = -\psi_T^0(z)(\bar{e} - \underline{e} - e) \quad (z, e) \in \mathbb{R}^d \times I. \tag{44}$$

This quantity determines the value function V_T at the final time T by

$$V_T(e, z) = C_T(e, z), \quad (z, e) \in \mathbb{R}^d \times I. \tag{45}$$

Prior to the terminal time $t = 0, \dots, T - 1$, the backward induction yields the value functions $(V_t)_{t=0}^{T-1}$ recursively by

$$\tilde{V}_{t+1}(e, z) = \mathbb{E}(V(Z_{t+1}, e) \mid Z_t = z) \quad (z, e) \in \mathbb{R}^d \times I, \tag{46}$$

$$V_t(e, z) = \min_{(f,b) \in \Xi(e,z)} \left(C_t^{e,z}(f, b) + \tilde{V}_{t+1}(e + b, z) \right) \quad (z, e) \in \mathbb{R}^d \times I, \tag{47}$$

where the minimum must be taken over the set $\Xi(e, z)$ of all admissible controls

$$\Xi(e, z) = \{(f, b) : f, b \in \mathbb{R}, \ e + b \in I\}, \quad e \in I \tag{48}$$

due to restriction that the energy transfer from/to the storage is limited by the storage capacity. However, in our approach, many arguments rely on the assumption that the set of admissible decisions

does not depend on the current situation. To replace the minimization over admissible controls $\Xi(e, z)$ in (11) by a minimization over an unrestricted set

$$\Xi = \{(f, b) : f, b \in \mathbb{R}\},$$

we introduce a penalization for violation $(f, b) \notin \Xi(e, z)$, having in mind that for a sufficiently strong penalization it will be never optimal to violate the restriction $e + b \in I$. This concept is realized by the following recursions:

$$V_T(e, z) = C_T(e, z), \quad (z, e) \in \mathbb{R}^d \times I \tag{49}$$

$$\tilde{V}_{t+1}(e, z) = \begin{cases} \mathbb{E}(V_{t+1}(e, Z_{t+1}) \mid Z_t = z) & e \in I \\ +\infty & e \notin I \end{cases}, z \in \mathbb{R}^d \tag{50}$$

$$V_t(e, z) = \inf_{(f, b) \in \Xi} \left(C_t^{e, z}(f, b) + \tilde{V}_{t+1}(e + b, z) \right), \quad (z, e) \in \mathbb{R}^d \times I. \tag{51}$$

Please note that with this definition, the functions $(V_t)_{t=0}^T$ satisfy (45)–(48) if and only if they fulfill (49)–(51).

5. Variables Reduction for Battery Control

In this section, we adopt and use our variable reduction technique to the specific situation described in Section 4. Therefore, we show that all assumptions of Theorem 1 are fulfilled. Recall from our modeling (44), (42) that the set of activities is given by

$$\Xi = \{(f, b) : f \in \mathbb{R}, \ b \in \mathbb{R}\} \tag{52}$$

with the interpretation that f and b represent the energy from trading and that from the battery, respectively. On this account, we have assumed that the costs and resource consumption functions are given by

$$C_t^{e, z}(f, b) = \psi_t^0(z)(d_t(z) + f) + q(d_t(z) + f)^2 - \psi_t^+(z)h_t^+(f + b)$$
$$+ \psi_t^-(z)h_t^-(f + b) + \chi(e) \tag{53}$$

$$E_t^{e, z}(f, b) = b \quad \text{for } z \in \mathbb{R}^d, (f, b) \in \Xi, t = 0, \ldots, T - 1. \tag{54}$$

Here the coefficient $q > 0$ represents the price elasticity whereas the functions h_t^+ and h_t^- are defined by (37) as

$$h_t^+(u) = \mathbb{E}[(u - \varepsilon_t)^+] = \int_{\mathbb{R}} (u - \epsilon)^+ \mathbb{P}_{\varepsilon_t}(d\epsilon),$$

$$h_t^-(u) = \mathbb{E}[(u - \varepsilon_t)^-] = \int_{\mathbb{R}} (u - \epsilon)^- \mathbb{P}_{\varepsilon_t}(d\epsilon),$$

in terms of the distribution $\mathbb{P}_{\varepsilon_t}$ of the demand prediction error ε_t, this distribution is non-random by assumption (35). Define the function

$$h_t(z, u) = \int_{\mathbb{R}} [-\psi_t^+(z)(u - \epsilon)^+ + \psi_t^-(z)(u - \epsilon)^-] \mathbb{P}_{\varepsilon_t}(d\epsilon) \quad z \in \mathbb{R}^d, u \in \mathbb{R} \tag{55}$$

and observe that

$$u \mapsto h_t(z, u) \text{ is convex and non-increasing on } \mathbb{R} \text{ for each } z \in \mathbb{R}^d \tag{56}$$

due to $0 \le \psi_t^+(z) < \psi_t^-(z)$ stated in (30). Using (55), we rewrite (53) as

$$
C_t^{e,z}(f, b) = \psi_t^0(z)(d_t(z) + f) + q(d_t(z) + f)^2 + h_t(z, f + b) + \chi(e) \\
\text{for } (f, b) \in \Xi, (e, z) \in I \times \mathbb{R}^d.
\tag{57}
$$

With this representation, the convexity required by (24) is obvious, since the functions $(f, b) \mapsto q(d_t(z) + f)^2$ and $(f, b) \mapsto h_t(z, f + b)$ are convex on Ξ, while $e \mapsto \chi(e)$ is convex on I.

To verify (22), consider $\xi' = (f', b') \in \Xi$ and $e, e' \in I$ with $e < e'$ to define

$$
\xi = (f, b) = (f', b' + e' - e)
$$

which yields the same resource consumption

$$
e + E_t^{e,z}(\xi) = e' + b' = e' + E_t^{e',z}(\xi')
$$

at no greater costs

$$
\begin{aligned}
C_t^{e,z}(\xi) &= \psi_t^0(z)(d_t(z) + f') + q(d_t + f')^2 + h_t(z, f' + b' + e' - e) + \chi(e) \\
&\le \psi_t^0(z)(d_t(z) + f') + q(d_t + f')^2 + h_t(z, f' + b') + \chi(e') = C_t^{e',z}(\xi'),
\end{aligned}
$$

since h_t is non-increasing by (56) in the second variable and the deep discharge penalization function χ is strictly increasing by assumption (43).

Now let us turn to the Legendre-type transform and verify (21) and (23). To obtain explicit expressions, let us suppose for the sake of concreteness that

$$
\begin{aligned}
&\text{for } t = 0, \dots, T - 1, \text{ the distribution } \mathbb{P}_{\varepsilon_t} \text{ possesses a strictly} \\
&\text{increasing and continuous distribution function } F_{\varepsilon_t}.
\end{aligned}
\tag{58}
$$

The derivative of the function h_t from (55) under the assumptions (58) is evaluated explicitly as

$$
\partial_{(0,1)} h_t(z, u) = (\psi_t^-(z) - \psi_t^+(z)) F_{\varepsilon_t}(u) - \psi_t^-(z), \qquad u \in \mathbb{R}.
\tag{59}
$$

Given an $A \in \mathbb{R}$, consider the minimization (12) of $C_t^{e,z}(\xi) + A E_t^{e,z}(\xi)$ over $\xi \in \Xi$, which yields the minimization of a convex and smooth function

$$
\mathbb{R}^2 \to \mathbb{R}, \qquad (f, b) \mapsto \psi_t^0(z)(d_t(z) + f') + q(d_t(z) + f')^2 + h_t(z, f + b) + Ab
\tag{60}
$$

whose first-order conditions are

$$
\psi_t^0(z) + 2q(d_t(z) + f) + \partial_{(0,1)} h_t(z, f + b) = 0, \qquad \partial_{(0,1)} h_t(z, f + b) + A = 0.
$$

These conditions are fulfilled by $(f^*, b^*) \in \mathbb{R}^2$ obtained from

$$
\psi_t^0(z) + 2q(d_t(z) + f^*) = A, \qquad \partial_{(0,1)} h_t(z, f^* + b^*) = -A.
\tag{61}
$$

That is, the unique minimum of (60) exists for $0 < \psi_t^+(z) < A < \psi_t^-(z)$ and is given by solution to (61) as

$$
f_t^{e,z}(A) = \frac{A - \psi_t^0(z)}{2q} - d_t(z),
\tag{62}
$$

$$
b_t^{e,z}(A) = F_{\varepsilon_t}^{-1} \left(\frac{\psi_t^-(z) - A}{\psi_t^-(z) - \psi_t^+(z)} \right) - \frac{A - \psi_t^0(z)}{2q} + d_t(z),
\tag{63}
$$

giving the minimizer

$$\mathcal{X}_t^{e,z}(A) = (f_t^{e,z}(A), b_t^{e,z}(A))$$

for (21) over the relevant interval

$$A \in \mathcal{A}^{e,z} =]\psi_t^+(z), \psi_t^-(z)[. \tag{64}$$

Furthermore, assumption (23) is satisfied, since

$$A \mapsto E_t^{e,z}(\mathcal{X}_t^{e,z}(A)) = b_t^{e,z}(A)$$

is linearly decreasing over the interval (64) ranging from a positive

$$\lim_{A \downarrow \psi_t^+(z)} b_t^{e,z}(A) = F_{\varepsilon_t}^{-1}(1_-) - \frac{\psi_t^+(z) - \psi_t^0(z)}{2q} > 0,$$

to a negative value

$$\lim_{A \uparrow \psi_t^+(z)} b_t^{e,z}(A) = F_{\varepsilon_t}^{-1}(0_+) - \frac{\psi_t^-(z) - \psi_t^0(z)}{2q} < 0.$$

All ingredients required for (14)–(16) are specified for $t = 0, \ldots, T-1$, $z \in \mathbb{R}^d$, $A \in \mathcal{A}^{e,z}$ by the following quantities

$$E_t^{e,z}(\mathcal{X}_t^{e,z}(A)) = b_t^{e,z}(A)$$
$$C_t^{e,z}(\mathcal{X}_t^{e,z}(A)) = \psi_t^0(z)(d_t(z) + f_t^{e,z})(A) + q(d_t(z) + f_t^{e,z})A^2 + h_t(z, f_t^{e,z}(A) + b_t^{e,z}(A)) + \chi(e)$$

which are relevant in (14)–(16).

6. The Variables Reduction Technique

This section is devoted to the derivation of Theorem 1. Let us prepare the proof by gradually showing auxiliary results. Our arguments rely on convexity of the functions $(V_t)_{t=0}^T$ from (9)–(11). It turns out that, based on the assumption

$$\tilde{V}_{t+1}(\cdot, z) \text{ is convex and non-decreasing on } I$$
$$\text{and } \tilde{V}_{t+1}(\cdot, z) = +\infty \text{ on } \mathbb{R} \setminus I, \text{ for all } z \in \mathbb{R}^d, \tag{65}$$

the same properties can be deduced for $\tilde{V}_t(\cdot, z)$ as formulated in the following lemma.

Lemma 1. *Given costs and resource consumption (18), let $t \in \{0, \ldots, T-1\}$, suppose that (65) holds and consider $V_t(\cdot, z)$ defined by (11). If (22) holds then $V_t(\cdot, z)$ is non-decreasing on I for all $z \in \mathbb{R}^d$. If in addition (24) is satisfied, then $V_t(\cdot, z)$ is non-decreasing and convex on I for all $z \in \mathbb{R}^d$.*

Proof. Given $z \in \mathbb{R}^d$, $e, e' \in I$ with $e < e'$, and arbitrary $\xi' \in \Xi$ there exists by (22) an action $\xi \in \Xi$ with

$$e + E_t^{e,z}(\xi) = e' + E_t^{e',z}(\xi') \text{ and } C_t^{e,z}(\xi) \le C_t^{e',z}(\xi')$$

which gives

$$C_t^{e,z}(\xi) + \tilde{V}_{t+1}(e + E_t^{e,z}(\xi), z) \le C_t^{e',z}(\xi') + \tilde{V}_{t+1}(e' + E_t^{e',z}(\xi'), z)$$

and shows $V_t(e, z) \le V_t(e', z)$ by minimization in ξ and ξ' over Ξ.

Now we turn to the convexity. Using the infimum in (11) for each $\delta > 0$, there exist $\xi, \xi' \in \Xi$ such that

$$\frac{\delta}{2} + V_t(e, z) \geq C_t^{e,z}(\xi) + \tilde{V}_{t+1}(e + E_t^{e,z}(\xi), z),$$

$$\frac{\delta}{2} + V_t(e', z) \geq C_t^{e',z}(\xi') + \tilde{V}_{t+1}(e' + E_t^{e',z}(\xi'), z).$$

Please note that these values are finite due to (20), which means that

$$e + E_t^{e,z}(\xi) \in I, \quad e' + E_t^{e',z}(\xi') \in I.$$

That is, by convex linear combinations of the activities (Ξ is convex by (24)) and resource levels we obtain

$$\xi^{(\lambda)} = \lambda \xi + (1 - \lambda) \xi' \in \Xi$$

$$e^{(\lambda)} = \lambda e + (1 - \lambda) e' \in I.$$

Now, by convexity of costs functions (24) it holds that

$$\lambda C_t^{e,z}(\xi) + (1 - \lambda) C_t^{e',z}(\xi') \geq C_t^{e^{(\lambda)},z}(\xi^{(\lambda)})$$

and using assumption (65) we conclude

$$\lambda \tilde{V}_{t+1}(e + E_t^{e,z}(\xi), z) + (1 - \lambda)\tilde{V}_{t+1}(e' + E_t^{e',z}(\xi'), z) \geq$$
$$\geq \tilde{V}_{t+1}(\lambda e + (1 - \lambda)e' + \lambda E_t^{e,z}(\xi) + (1 - \lambda)E_t^{e',z}(\xi'), z) \geq$$
$$\geq \tilde{V}_{t+1}(e^{(\lambda)} + E_t^{e^{(\lambda)},z}(\xi^{(\lambda)}), z).$$

From this, we proceed with

$$\delta + \lambda V_t(e, z) + (1 - \lambda)V_t(e', z) \geq C_t^{e^{(\lambda)},z}(\xi^{(\lambda)}) + \tilde{V}_{t+1}(e^{(\lambda)} + E_t^{e^{(\lambda)},z}(\xi^{(\lambda)}), z)$$
$$\geq \inf_{\xi \in \Xi} \left(C_t^{e^{(\lambda)},z}(\xi) + \tilde{V}_{t+1}(e^{(\lambda)} + E_t^{e,z}(\xi), z) \right)$$
$$\geq V_t(e^{(\lambda)}, z)$$

that shows the desired convexity. $\square$

In what follows, we recall the notion ∇g of the so-called *sub-gradient* of a convex function $g : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, $(n \in \mathbb{N})$ which is defined at each point $u \in \mathbb{R}^n$ as the family of linear functionals

$$\nabla g(u) = \{l : \mathbb{R}^n \to \mathbb{R} : \ l \text{ is linear with } g(u) + l(u' - u) \leq g(u'), \text{ for } u' \in \mathbb{R}^n\}.$$

For a function V on $I \times \mathbb{R}^d \to \mathbb{R}$ as in (65) which is convex in the first component on I, we agree to consider its sub-gradient in the first component, only for arguments in I. More precisely, for $(e, z) \in I \times \mathbb{R}^d$ we write $\nabla V(e, z)$ to denote the set of linear functionals $l : \mathbb{R} \to \mathbb{R}$ satisfying

$$V(e, z) + l(e' - e) \leq V(e', z) \text{ for all } e' \in \mathbb{R}.$$

Now, let us elaborate on the fixed-point property of the virtual resource price in the optimal regime in the spirit of the remark before Section 4.

Lemma 2. *Assume that the costs and resource consumption functions are given as in (18). Let $t \in \{0, \ldots, T - 1\}$ and $\tilde{V}_{t+1}$ be a function which satisfies (65), furthermore, suppose that (21) holds. If for $(e, z) \in \mathbb{R}^d$ there exists $A^* \in \mathcal{A}^{e,z}$ satisfying the fixed-point equation*

$$A^* \in \nabla \tilde{V}_{t+1}(e + E_t^{e,z}(\mathcal{X}_t^{e,z}(A^*)), z), \tag{66}$$

then $\xi^* = \mathcal{X}_t^{e,z}(A^*)$ is a minimizer to (11).

Proof. By assumption (21) the minimizer $\xi^* = \mathcal{X}_t^{e,z}(A^*)$ to (12) satisfies

$$C_t^{e,z}(\xi^*) + A^* E_t^{e,z}(\xi^*) \le C_t^{e,z}(\xi) + A^* E_t^{e,z}(\xi) \tag{67}$$

for each $\xi \in \Xi$, which is equivalent to

$$C_t^{e,z}(\xi^*) - C_t^{e,z}(\xi) \le A^*(E_t^{e,z}(\xi) - E_t^{e,z}(\xi^*)). \tag{68}$$

On the other hand, since $A^* \in \nabla \tilde{V}_{t+1}(e + E_t^{e,z}(\xi^*), z)$ by assumption (66), we conclude that

$$A^*(E_t^{e,z}(\xi) - E_t^{e,z}(\xi^*)) \le \tilde{V}_{t+1}(e + E_t^{e,z}(\xi), z) - \tilde{V}_{t+1}(e + E_t^{e,z}(\xi^*), z). \tag{69}$$

By combining (68) and (69)

$$C_t^{e,z}(\xi^*) - C_t^{e,z}(\xi) \le \tilde{V}_{t+1}(e + E_t^{e,z}(\xi), z) - \tilde{V}_{t+1}(e + E_t^{e,z}(\xi^*), z)$$

we obtain the desired assertion

$$C_t^{e,z}(\xi^*) + \tilde{V}_{t+1}(e + E_t^{e,z}(\xi^*), z) \le C_t^{e,z}(\xi) + \tilde{V}_{t+1}(e + E_t^{e,z}(\xi), z).$$

$\square$

The above lemma suggests determining a minimizer ξ^* to (11) as $\xi^* = \mathcal{X}_t^{e,z}(A^*)$ where A^* is obtained as a solution to the fixed-point relation (66). However, according to the dynamic programing principle, the virtual resource price $A^* \in \mathcal{A}^{e,z}$ shall be determined by the minimization (27) rather than by solving a fixed-point relation. The result below shows that under natural assumptions, (27) in fact implies (66).

Lemma 3. *Given costs and resource consumption (18), let $t \in \{0, \ldots, T - 1\}$ and $(e, z) \in I \times \mathbb{R}^d$, suppose that (65) holds and assume that (21) holds and (23) is satisfied. If $A^* = A^{e,z}$ fulfills (27), then A^* also fulfills the fixed-point equation (66).*

Proof. Given $(e, z) \in I \times \mathbb{R}^d$, suppose that $A^* = A^{e,z} \in \mathcal{A}^{e,z}$ satisfies (27). Recall that by assumption (23) there exists $A_0 \in \mathcal{A}^{e,z}$ there exists a root such that $E_t^{e,z}(\mathcal{X}_t^{e,z}(A_0)) = 0$. Since this root is included in the minimization, with a finite (due to $e \in I$) value

$$V_{t+1}(e + E_t^{e,z}(\mathcal{X}_t^{e,z}(A_0)), z) = V_{t+1}(z, e) < \infty,$$

the infimum is also finite

$$C_t^{e,z}(\mathcal{X}_t^{e,z}(A^*)) + \tilde{V}_{t+1}(e + E_t^{e,z}(\mathcal{X}_t^{e,z}(A^*)), z) < \infty.$$

By the minimum property (27), for each $A \in \mathcal{A}^{e,z}$ it holds that

$$
\begin{aligned}
C_t^{e,z}(\mathcal{X}_t^{e,z}(A^*)) + \tilde{V}_{t+1}(e + E_t^{e,z}(\mathcal{X}_t^{e,z}(A^*)), z) \leq \\
\leq C_t^{e,z}(\mathcal{X}_t^{e,z}(A)) + \tilde{V}_{t+1}(e + E_t^{e,z}(\mathcal{X}_t^{e,z}(A)), z).
\end{aligned}
\tag{70}
$$

Furthermore, by definition of $\mathcal{X}_t^{e,z}(A)$, it holds on the other hand that

$$
C_t^{e,z}(\mathcal{X}_t^{e,z}(A)) + E_t^{e,z}(\mathcal{X}_t^{e,z}(A)) \leq C_t^{e,z}(\xi) + A E_t^{e,z}(\xi)
$$

for each $\xi \in \Xi$. Replacing ξ by $\mathcal{X}_t^{e,z}(A^*)$, we deduce

$$
C_t^{e,z}(\mathcal{X}_t^{e,z}(A)) \leq C_t^{e,z}(\mathcal{X}_t^{e,z}(A^*)) + A(E_t^{e,z}(\mathcal{X}_t^{e,z}(A^*)) - E_t^{e,z}(\mathcal{X}_t^{e,z}(A)))
$$

which we combine with (70) to obtain

$$
\begin{aligned}
\tilde{V}_{t+1}(e + E_t^{e,z}(\mathcal{X}_t^{e,z}(A^*)), z) \leq \\
\leq A(E_t^{e,z}(\mathcal{X}_t^{e,z}(A^*)) - E_t^{e,z}(\mathcal{X}_t^{e,z}(A))) + \tilde{V}_{t+1}(e + E_t^{e,z}(\mathcal{X}_t^{e,z}(A)), z).
\end{aligned}
\tag{71}
$$

Introducing for $(\epsilon, z) \in I \times \mathbb{R}^d$ the notations

$$
\begin{aligned}
\mathcal{E}(A) &= e + E_t^{e,z}(\mathcal{X}_t^{e,z}(A)), \quad A \in \mathcal{A}^{e,z}, \\
\tilde{V}(\epsilon) &= \tilde{V}_{t+1}(\epsilon, z),
\end{aligned}
$$

the inequality (71) is written as

$$
\tilde{V}(\mathcal{E}(A^*)) + A(\mathcal{E}(A) - \mathcal{E}(A^*)) \leq \tilde{V}(\mathcal{E}(A)), \qquad A \in \mathcal{A}^{e,z}.
\tag{72}
$$

For a strictly increasing sequence $(A_n^+)_{n \geq 1} \subset \mathcal{A}^{e,z}$ converging to A^* define a sequence

$$
\Delta_n^+ = \mathcal{E}(A_n^+) - \mathcal{E}(A^*), \quad n \geq 1
$$

which is strictly decreasing, converges to zero due to (23), and satisfies

$$
\tilde{V}(\mathcal{E}(A^*)) + A_n^+ \Delta_n^+ \leq \tilde{V}(\mathcal{E}(A^*) + \Delta_n^+),
\tag{73}
$$

because of (72). Similarly, for a strictly decreasing sequence $(A_n^-)_{n \geq 1} \subset \mathcal{A}^{e,z}$ converging to A^*, define a sequence

$$
\Delta_n^- = \mathcal{E}(A_n^-) - \mathcal{E}(A^*), \quad n \geq 1
$$

which is strictly increasing, converges to zero due to (23), and satisfies

$$
\tilde{V}(\mathcal{E}(A^*)) + A_n^- \Delta_n^- \leq \tilde{V}(\mathcal{E}(A^*) + \Delta_n^-).
\tag{74}
$$

Now apply Lemma 4 with above sequences $(\Delta_n^+)_{n \geq 1}$, $(A_n^+)_{n \geq 1}$ and $(\Delta_n^-)_{n \geq 1}$, $(A_n^-)_{n \geq 1}$ using their properties (73) and (74) to deduce the assertion

$$
A^* \in \nabla \tilde{V}(\mathcal{E}(A^*)) = \nabla \tilde{V}_{t+1}(e + E_t^{e,z}(\mathcal{X}_t^{e,z}(A^*)), z).
$$

□

Lemma 4. *Let $\tilde{V} : \mathbb{R} \mapsto \mathbb{R} \cup \{\infty\}$ be a convex function and $\varepsilon \in \mathbb{R}$ such that $\tilde{V}(\epsilon) < \infty$. Suppose that there exist sequences $(\Delta_n^+)_{n \geq 1} \subset]0, \infty[, (\Delta_n^-)_{n \geq 1} \subset] - \infty, 0[$ converging to zero and sequences $(A_n^-)_{n \geq 1}, (A_n^+)_{n \geq 1}$ converging to $A^* \in \mathbb{R}$ such that*

$$
\begin{aligned}
\tilde{V}(\epsilon) + A_n^+ \Delta_n^+ &\leq \tilde{V}(\epsilon + \Delta_n^+) \\
\tilde{V}(\epsilon) + A_n^- \Delta_n^- &\leq \tilde{V}(\epsilon + \Delta_n^-)
\end{aligned}, \quad \text{for all } n \geq 1.
\tag{75}
$$

Then it holds that $A^* \in \nabla \tilde{V}(\epsilon)$.

Proof. Suppose on the contrary that $A^* \notin \nabla \tilde{V}(\epsilon)$: then there exists $\delta \in \mathbb{R}$ with

$$
\tilde{V}(\epsilon) + A^* \delta > \tilde{V}(e + \delta).
\tag{76}
$$

First, we assume that $\delta > 0$, then there exists $A_0 < A^*$ such that

$$
\tilde{V}(\epsilon) + A_0 \delta > \tilde{V}(\epsilon + \delta)
$$

and since $\tilde{V}$ is convex, the above inequality holds for all intermediate points

$$
\tilde{V}(\epsilon) + A_0 \Delta > \tilde{V}(\epsilon + \Delta), \qquad \Delta \in]0, \delta[.
\tag{77}
$$

However, for sufficiently large $n \in \mathbb{N}$ we obtain $A_n^+ > A_0$ such that

$$
\tilde{V}(\epsilon) + A_n^+ \Delta_n^+ > \tilde{V}(\epsilon) + A_0 \Delta_n^+ > \tilde{V}(\epsilon + \Delta_n^+),
$$

which gives a contradiction to (75) for $n \in \mathbb{N}$ sufficiently large to satisfy $\Delta_n^+ \in]0, \delta[$. A similar argument in the alternative situation $\delta < 0$ also shows a contradiction and completes the proof. $\square$

Finally, let us now compose the outcomes of the Lemmas 1–4 to our technique for control variables reduction. Consider a problem defined in Section under the assumptions formulated in Section 3. Let us prove Theorem 1.

Proof. We proceed by induction to show that

$$
\left.
\begin{array}{c}
\text{the functions } (V_t(\cdot, z))_{t=0}^T \text{ and } (\tilde{V}_t(\cdot, z))_{t=0}^T \text{ are} \\
\text{convex and non-decreasing on } I \text{ for each } z \in \mathbb{R}^d.
\end{array}
\right\}
\tag{78}
$$

Since for each $z \in \mathbb{R}^d$ the function

$$
I \to \mathbb{R}, \qquad e \mapsto C_T(e, z)
$$

is convex and non-decreasing by (24), the initialization (10) ensures that $\tilde{V}_T(\cdot, z)$ is convex and non-decreasing on I. Please note that all conditions of Lemma 1 are fulfilled by assumptions, thus for $t = T - 1$ we conclude that that $V_{T-1}(\cdot, z)$ is also convex and non-decreasing on I for each $z \in \mathbb{R}^d$, and calculating the conditional expectation, we observe that $\tilde{V}_{T-1}(\cdot, z)$ is also convex and non-decreasing on I for each $z \in \mathbb{R}^d$. Proceeding for $t = T - 1, \dots, 0$ inductively, with the same argumentation, the assertion (78) follows.

Now we turn to the main claim (25). Since (9), (10) coincide with (14), (15), we obtain $\tilde{V}_T = \tilde{v}_T$. With this, we apply Lemma 3 whose assumption (65) is satisfied because of (78), ensuring that $\tilde{V}_T = \tilde{v}_T$ is convex and non-decreasing in the first argument on I. Further conditions (21), (23) of this lemma also hold. Moreover, since we have supposed in (16) that the minimum is reached, say at $A^* = A^{*,e,z}$, the fixed-point property (66)

$$
A^{*,e,z} \in \nabla \tilde{V}_T(e + E_{T-1}^{e,z}(\mathcal{X}_{T-1}^{e,z}(A^{*,e,z})), z)
$$

holds for all $(e, z) \in I \times \mathbb{R}^d$. Using Lemma 2, we conclude that

$$\xi^*_{T-1} = \mathcal{X}^{e,z}_{T-1}(A^{*,e,z})$$

is a minimizer to (10), showing that $V_{T-1} = v_{T-1}$. Repeating the argumentation for $t = T - 1, \ldots, 1$, the desired assertion $v_t = V_t$ for $t = T, \ldots, 0$ follows inductively. $\square$

7. Conclusions

New technologies have triggered a growing attention to optimization algorithms for operational management of energy storage facilities. In the generic settings, a typical electricity retailer determines an optimal strategy for purchase, procurement, generation, and storage of electrical power while taking into account a fluctuating energy price, storage costs, limited storage capacity, and uncertain production rates. Such problems are challenging, since numerous decision and control variables must be considered within a potentially high-dimensional setting. This paper addresses a variables reduction technique for such problems. Within an abstract, but natural framework, we show that a certain Legendre-type transform can be applied to equivalently reformulate the original strategy optimization into a stochastic control problem driven by a *single one-parameter* family of decision variables. It turns out that the presented technique is sufficiently general and can be applied to solve diverse dynamic resource allocation involving scarce reserves. These problems encompass a wide range of important areas including mining operations, pension fund management, and emission abatement. The authors believe that their approach can be used as a starting point for efficient numerical algorithms and will address this topic in future research.

References

1. Trigeorgis, L. *Real Options: Managerial Flexibility and Strategy in Resource Allocation*; MIT Press: Cambridge, MA, USA, 1996.
2. Trigeorgis, L. The nature of option interactions and the valuation of investments with multiple real options. *J. Financ. Quant. Anal.* **1993**, *28*, 1–20. [CrossRef]
3. Ross, S.A. A Simple Approach to the Valuation of Risky Streams. *J. Bus.* **1978**, *51*, 453–475. [CrossRef]
4. McDonald, R.L.; Siegel, D.R. Investment and the Valuation of Firms When There is an Option to Shut Down. *Int. Econ. Rev.* **1985**, *26*, 331–349. [CrossRef]
5. Dias, M. Valuation of exploration and production assets: An overview of real options models. *J. Pet. Sci. Eng.* **2004**, *44*, 93–114. [CrossRef]
6. Jaimungal, S.; de Souza, M.O.; Zubelli, J.P. Real option pricing with mean-reverting investment and project value. *Eur. J. Financ.* **2013**, *19*, 625–644. [CrossRef]
7. Cortazar, G.; Schwartz, E.S.; Casassus, J. Optimal exploration investments under price and geological-technical uncertainty: A real options model. *R&D Manag.* **2001**, *31*, 181–189. [CrossRef]
8. Devalkar, S.K.; Anupindi, R.; Sinha, A. Integrated Optimization of Procurement, Processing, and Trade of Commodities. *Oper. Res.* **2011**, *59*, 1369–1381. [CrossRef]
9. Oeksendal, B.; Sulem, A. *Applied Stochastic Control of Jump Diffusions*; Springer: Berlin/Heidelberg, Germany; New York, NY, USA, 2005.

10. Pham, H. *Continuous-Time Stochastic Control and Optimization with Financial Applications*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 2009.

11. Gobet, E.; Lemor, J.P.; Warin, X. A regression-based Monte Carlo method to solve backward stochastic differential equations. *Ann. Appl. Probab.* **2005**, *114*, 2172–2202. [CrossRef]

12. Gobet, E.; Lemor, J.P. Numerical simulation of bsdes using empirical regression methods: Theory and practice. *arXiv* **2008**, arXiv:0806.4447.

13. Bender, C.; Dokuchaev, N. A First Order BSPDE for Swing Option Pricing. *Math. Finance* **2016**, *26*, 461–491. [CrossRef]

14. Bäuerle, N.; Rieder, U. *Markov Decision Processes with Applications to Finance*; Springer: Heidelberg, Germany, 2011. [CrossRef]

15. Puterman, M. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*; Wiley: New York, NY, USA, 1994.

16. Powell, W.B. *Approximate Dynamic Programming: Solving the Curses of Dimensionality*; Wiley: Hoboken, NJ, USA, 2007.

17. Hinz, J.; Yee, J. Solving Control Problems with Linear State Dynamics—A Practical User Guide. In Proceedings of the 2016 Second International Symposium on Stochastic Models in Reliability Engineering, Life Science and Operations Management (SMRLO), Beer-Sheva, Israel, 15–18 February 2016; pp. 591–596. [CrossRef]

18. Kempener, R.; Borden, E. *Battery Storage for Renewables: Market Status and Technology Outlook*; Report; IRENA: Abu Dhabi, UAE, 2015.

19. Lawson, B. *Electropaedia*; Report; Woodbank Communications Ltd.: Chester, UK, 2019. Available online: https://www.mpoweruk.com (accessed on 14 April 2020).