ISE

Industrial and Systems Engineering

An Energy Storage Deployment Program under Random Discharge Permissions

Somayeh Moazeni[1] and Boris Defourny[2]

[1]School of Systems and Enterprises, Stevens Institute of Technology, USA

[2]Department of Industrial and Systems Engineering, Lehigh University, USA

ISE Technical Report 15T-012

LEHIGH UNIVERSITY

An Energy Storage Deployment Program under Random Discharge Permissions

Somayeh Moazeni
School of Systems and Enterprises, Stevens Institute of Technology, 1 Castle Point Terrace, Hoboken, NJ 07030, USA, smoazeni@stevens.edu

Boris Defourny
Department of Industrial and Systems Engineering, Lehigh University, 200 W Packer Ave, Bethlehem, PA 18015, USA, defourny@lehigh.edu

Recent developments in energy storage technology and the greater use of renewables have increased interest in energy storage. Along with the unique capabilities and characteristics of storage, this interest has created the need to design and study efficient energy storage deployment programs, with the goal of providing attractive flexibility to storage owners while still indirectly supervising their operations. In this paper, we envision a framework that defines random times at which a particular participating storage unit is allowed to discharge. In this flexible energy storage deployment program, a participating storage unit receives permissions at random to discharge during peak hours in real time. Should discharge permission be issued, the storage owner has the option to discharge and be paid at a time-dependent reward that is specified contractually, or can wait for future permissions. In the present work, the discharge permission times are modeled by a Poisson process. We study the problem of optimizing discharge operations from the perspective of the storage resource owner, whose goal is to maximize his total revenue accumulated during the peak hours of a day, in expectation over the random number and times of discharge permissions. We analyze the resulting piecewise deterministic Markov decision process and provide structural results on the value function and the optimal discharge policy. Several numerical experiments illustrate the impact of the restricted discharge opportunities.

Key words: energy storage, smart grid, Markov models, dynamic programming, stochastic control

History: This paper was first submitted on October 28, 2015.

1. Introduction

In the effort to curb fossil fuel emissions and to promote renewable energy sources, energy storage plays a key role to mitigate the intermittency and uncertainty of renewables (see Denholm et al. (2010), Diaz-Gonzalez et al. (2012), Du and Lu (2014), Quadrennial Energy Review (2015)). Additionally, reliability and cost improvements in energy storage and battery technologies are underway (see Straubel (2015), Quadrennial Energy Review (2015)). Therefore, distribution and optimal management of storage resources constitute increasingly important problems.

Managing grid-level storage or controlling hybrid renewable-energy storage systems have been the topic of several previous studies, see e.g., Thompson et al. (2009), Lai et al. (2010), Lohndorf et al. (2013), Sioshansi et al. (2014), Zhou et al. (2013), Moazeni et al. (2015b), Harsha and Dahleh (2015), Moazeni et al. (2015a), Halman et al. (2015) and the references therein. These studies differ in their settings, modeling approach, and objectives. Operations optimization of storage facilities that participate in the wholesale electricity market by placing bids and commitments with the objective of profiting from price differences are studied in Carmona and Ludkovski (2010), Byrne and Verbic (2013), Xi et al. (2015). In all these models, electricity flows in and out of the storage resources during time periods that are well specified in advance. The present work differs by introducing uncertainty in the permitted discharge times. In addition, in these models, the grid or the entity, interacting with the storage owner, commits in advance to buying electricity at specific times, for example every hour in discrete time models or anytime in continuous time models.

With the increased momentum in affordable and safe energy storage, additional opportunities for energy storage deployment may emerge (Bhatnagar et al. 2013). In particular, the unique characteristics of energy storage technologies including quick ramping capabilities and the ability to provide multiple ancillary services differentiate them from existent power system assets and create potential new market opportunities for energy storage resources (see e.g., Xiao et al. (2014)). However, a number of market and regulatory barriers persist, given the difficulty of organizing, assessing and fairly rewarding the services that can be provided by storage, especially when it comes to encouraging small participants at the level of the distribution network.

In order to facilitate the deployment of storage resources by a broader range of participants, in this paper we propose and analyze a form of agreement and remuneration scheme that seeks to increase the flexibility of operations for the storage resources and to reduce the risk inherent to contracts with commitments and financial penalties for noncompliance. In this agreement, the storage owner has the option, but not the obligation, to discharge in real time at time-varying rewards agreed upon in advance. As a result, in contrast to participation in the wholesale market, the storage owner does not have to commit to providing electricity, and does not need to get involved in a bidding process. The model is scalable to hundreds, possibly thousands of storage units, in a given area managed by a single load serving entity or utility.

However, as deployment of storage devices by individuals and their participation in this flexible program become widespread, and if the power injections from storage resources are completely unregulated and ad hoc, there will be times when a large number of storage owners discharge in arbitrary amounts. In this paper, we envision a framework that restricts the times at which a particular participating storage unit can discharge. Constraining discharge times will enable distribution companies or other entities interacting with these participating distributed storage units to indirectly supervise their activities and control the aggregated effect on the grid and the market.

The times allowed for discharge are typically driven by the state of the power system and the markets, but from the perspective of the storage operator and in the agreement, these times are random. The constraints on discharge times can be fixed in advance in the contract or they can be driven by the market conditions, such as the electricity price or other stability concerns, in which case they are adjusted over time to achieve grid or market performance needs. In the latter case, the controller of a participating storage unit monitors the permitted discharge times in real time. Random discharge permission times is one of the salient features of our model. Here, the contract underwriter (for example, a utility company) does not have to commit to purchasing electricity from storage owner in advance. The random times follow a random process controlled by the load serving entity or utility that can be well defined in the contract.

This paper is concerned by the problem of optimizing discharge operations, from the perspective of the storage resource owner, under the regulation framework described above. The feature of being allowed to discharge only at times determined exogenously and randomly in real time considerably complicates the storage discharge optimization problem, but still lend itself to a tractable solution approach.

Most existing analyses of storage operations in literature (including those cited above) have focused on a discrete time model with fixed time epochs or a continuous time operation environment over a finite or infinite horizon. To the best of our knowledge, the analysis of an energy storage discharge environment restricted to exogenous random discharge permission times is novel. The absence of commitment on the permitted discharge times in such a storage introduction program allows the contract underwriter (market operator or electric utility company) to manage the level of participation of distributed storages depending on the market or grid states, or taking into account random noncompliance to regulation orders. For example random permission time epochs can be triggered during the on-peak hours of a day as a function of the price level or zonal demand level. Discharge permission times and frequencies can differ as a function of the storage technology and geographical location.

The permission flow can be modeled by an arrival process. In this paper, discharge permissions are represented by a Poisson process. At each discharge time, the participating storage reward is modeled by a time-varying concave and an increasing function of amount discharged. The concavity of the reward function produces a nonlinear pricing scheme that encourages discharges in smaller amounts, as if the utility was imposing quantity discounts upon the sellers. We assume that the arrival rate of the discharge permission times and the reward function are known by the participating storage; for instance they can be specified in the contract.

The storage discharge problem is formulated as a continuous-time stochastic control problem, in which the optimal discharge policy depend on the stochastic discharge permission flow. The optimal discharge control problem in this work belongs to the family of piecewise deterministic Markov decision processes, a class of optimal control problems introduced by Davis (1984) and studied in Davis (1993). These processes evolve through random jumps at random points in time while the evolution between jumps is deterministic. Following Kushner and Dupuis (2001), the dynamic programming principle for the control problem of interest leads into a system of nonlinear partial differential equations, which can be solved numerically. We demonstrate the effectiveness of the computational approach using several numerical examples, where for illustration purposes, permission times are triggered by spikes in the electricity price. Capturing electricity price spikes as jumps modeled via Poisson processes has been frequently considered in the literature on electricity price models, see e.g., Deng (1999), Cartea and Figueroa (2005), Culot et al. (2006), Geman and Roncoroni (2006), Kluge (2006), and Weron et al. (2004). For a comprehensive survey on the electricity price models see Carmona and Coulon (2013).

For other studies on controlled piecewise deterministic Markov processes see Yushkevich (1980), Hordijk and Schouten (1985), Almudevar (2001), Guo and Hernandez-Lerma (2009), Bauerle and Rieder (2010). For applications of these models in finance and portfolio optimization, see e.g., Jacobsen (2006), Matsumoto (2006), Pham and Tankov (2008), Bauerle and Rieder (2009), Bayraktar and Ludkovski (2011), Gassiat et al. (2011), Fujimoto et al. (2013). For applications in insurance, see e.g., Schmidli (2008) and Kirch and Runggaldier (2005). For applications in queueing theory, see e.g., Kitaev and Rykov (1995) and Rieder and Winter (2009).

To summarize, our main contributions include developing a new approach to modeling distributed energy storage deployment and analyzing several structural properties of the value function and the optimal discharge policy under this program. The analyses in this paper can guide a potential storage owner to value its participation in this flexible energy storage deployment framework. In addition, this study provides insights for the policy makers and regulators to design efficient and attractive storage deployment programs.

This paper is organized as follows. The mathematical formulation of the optimal energy storage discharge control problem is described in §2. The structure of the value function is analyzed in §3. In §4, the derivation of the partial differential equation and the procedure to compute an optimal control are discussed. §5 summarizes structural properties of the optimal policy. Illustrative examples and computational analyses are presented in §6. A discussion of the managerial insights and of possible extensions of the framework conclude the paper in §7.

Throughout this paper, "increasing" and "decreasing" mean "nondecreasing" and "nonincreasing", respectively. We denote the set of natural numbers including zero by $\mathbb{N}$ and the set of nonnegative real numbers by $\mathbb{R}_+$.

2. The Model

Consider an energy storage unit of capacity $K>0$, participating in a flexible discharge program, which enables it to discharge its stored power at permitted times over $[0,T]$. We postulate that the discharge permissions are issued randomly by the Poisson process $\{N_s\}_{s\geq 0}$ with arrival rate λ. We denote by $\{\mathcal{F}_t\}_{t\geq 0}$ the natural filtration associated with the Poisson process: $\mathcal{F}_t$ is the sigma algebra generated by $\{N_s\}_{s\leq t}$. We refer to the time-stamp of the i^{th} discharge permission that arrives on the interval $[t,\infty)$ by $\tau_{i,t}$. This implies that $\tau_{i,t}\geq t$. Thus, $\{\tau_{i,t}\}_{i\in\mathbb{N}}$ is the sequence of jump times of the Poisson process $\{N_s\}_{s\geq 0}$ since time t. We set $\tau_{0,t}=0$.

When discharge permission is communicated at some time $t\in[0,T)$, the storage owner will receive $R_t(a)$ dollars by discharging a units of electricity at this time. At terminal time T, the value of the leftover stored electricity is given by the terminal reward function $R_T(a)$. An example of the reward function is $R_t(a)=R(p_t,a)$ for some stationary function R and time-varying reward coefficients p_t, which may represent the expected nodal electricity prices. Alternatively, the reward function can be time-independent, i.e., $R_t(a)=R(p,a)$, where p can be interpreted as the average electricity price per day.

In this paper, we assume that the reward function $R_t(a)$ is nonnegative and null at $a=0$, increasing in a, concave in a, and continuous in t everywhere. The nonnegativity and concavity assumptions imply that R_t is subadditive, that is, $R_t(a_1+a_2)\leq R_t(a_1)+R_t(a_2)$. We define a terminal reward function $R_T(k)$ where k is the inventory remaining at time T. We assume R_T is nonnegative, null at $k=0$, increasing in k and concave in k.The subadditivity of R_t incentivizes participating storage units to split their inventories into smaller amounts and not to discharge the entire inventory at once. This is an attractive property of the program regarding the usage of the distribution network. However, by discharging the inventory in small amounts, the storage operator bears the risk of receiving no more discharge permissions by the terminal time, in which case a leftover inventory remains at time T that is valued according to the terminal reward R_T. In our numerical work reported in §6, we use the log-utility function $R_t(a)=\log(1+p_t a)$ as the reward function for $t\in[0,T)$ and the constant function $R_T(a)=0$ for the terminal time. Therefore, the storage operator faces the problem of deciding between using current versus non-guaranteed future discharge opportunities. The selected log-utility function satisfies the property that if an inventory k is split equally among n discharge times, the cumulated reward remains below an equivalent linear reward $R^\infty(k)=p$, corresponding to the limit case where n is large: $\lim_{n\to\infty} nR(k/n)=\lim_{n\to\infty} n\log(1+pk/n)=pk$. Therefore, if p is interpreted as a contractual price, the utility company is guaranteed not to spend more than the contractual price times the total quantity available in storage.

When the storage operator is rational and risk-neutral, an optimal discharge strategy π to discharge $k \leq K$ units of power can be determined by maximizing total expected revenue over the time horizon $[0,T]$. This results in the following optimization problem,

$$V_0(k) \stackrel{\text{def}}{=} \max_{x^\pi \in \mathcal{X}_0} \mathbb{E}\left[\sum_{i=1}^{N_{T-}} R_{\tau_{i,0}}\left(x^\pi_{\tau_{i-1,0}} - x^\pi_{\tau_{i,0}}\right) + R_T\left(x^\pi_T\right) \;\middle|\; x^\pi_0 = k\right], \tag{1}$$

where $\mathcal{X}_0$ is the set of all nonnegative real-valued, right-continuous with left limits, decreasing process $x^\pi = \{x^\pi_t\}_{t\in[0,T]}$ adapted to the filtration $\{\mathcal{F}_t\}_{t>0}$. The process x^π represents the inventory level under the discharge policy π. The filtration condition imply that its values can only change at the time of jumps of the Poisson process. The random variable N_{T^-} is the number of jumps of the Poisson process $\{N_s\}_{s\geq 0}$ over the time interval $[0,T)$, and $R_T(x^\pi_T)$ captures the terminal reward. The expectation in (1) is over the Poisson process $\{N_s\}_{s\geq 0}$. Note that from the assumptions on the reward function, $V_0(0) = 0$.

The aforementioned problem constitutes a piecewise deterministic Markov decision process (Davis 1993). In parallel with the formulation in (1) in terms of the controlled inventory level x^π, one may also describe the control strategy A^π for the discharge amount over $(0,T)$. The corresponding controlled inventory process x^π_t at time t when the discharge strategy A^π is being employed satisfies

$$\begin{aligned} &x^\pi_0 = k, \\ &dx^\pi_t = -A^\pi_t\left(x^\pi_{t-}\right) dN_t, \quad \forall t \in (0,T), \\ &x^\pi_T = x^\pi_{T-}, \end{aligned} \tag{2}$$

where $\{x^\pi_{t-}\}_{t\in[0,T]}$ is the left limit process. We denote the class of policies π where $x^\pi \in \mathcal{X}_0$ by Π.

For a fixed $T < \infty$, the expected performance of a policy π from time t onwards, starting from an inventory level k at time t, is written

$$V^\pi_t(k) \stackrel{\text{def}}{=} \mathbb{E}\left[\sum_{i=1}^{N_{T-}-N_{t-}} R_{\tau_{i,t}}\left(x^\pi_{\tau_{i-1,t}} - x^\pi_{\tau_{i,t}}\right) + R_T\left(x^\pi_T\right) \;\middle|\; x^\pi_t = k\right]. \tag{3}$$

Here, $N_{T^-} - N_{t^-}$ equals the number of jumps of the Poisson process $\{N_s\}_{s\geq 0}$ over the time interval $[t,T)$. The expected performance from time t onwards with an optimal strategy is written

$$V_t(k) \stackrel{\text{def}}{=} \max_{\pi\in\Pi_t} V^\pi_t(k), \qquad \forall t \in [0,T),\ \forall k \in [0,K], \tag{4}$$

and $V_T(k) = R_T(k)$ for all k in $[0,K]$. Here, Π_t is the set of all truncated policies defined over $[t,T]$. Note that $V_t(0) = 0$ for all $t \in [0,T]$, and $V_T(k) = R_T(k)$ for all $k \in [0,K]$.

The function $V_t(k)$ in (4) satisfies the following dynamic programming equation:

$$V_t(k) = \mathbb{E}\left[\max_{a\in\mathcal{A}_k}\left\{R_{\tau_{1,t}}(a)+V_{\tau_{1,t}}(k-a)\right\}\cdot 1_{\tau_{1,t}<T} + R_T(k)\cdot 1_{\tau_{1,t}\geq T}\right], \tag{5}$$

where the expectation is over the time $\tau_{1,t}$ of the next permission. With a slight abuse of language, we call this function the value function. It represents the expectation of the cumulated reward-to-go at the upcoming decision stage. Here, $\mathcal{A}_k \subseteq \mathbb{R}_+$ is the set of all discharge amounts that the storage unit can discharge, when the inventory level is k. Examples of the set of admissible discharges include $\mathcal{A}_k = [0,k]$, or $\mathcal{A}_k = \{0\}\cup[\underline{c},\min(k,\bar{c})]$ for some constants $\underline{c}\geq 0$ and $\bar{c}\leq K$. In this paper, we assume that the admissible $\mathcal{A}_k$ is nonempty for each $k\in[0,K]$ and $\arg\max_{a\in\mathcal{A}_k}\left\{R_{\tau_{1,t}}(a)+V_{\tau_{1,t}}(k-a)\right\}\neq\emptyset$.

Let $a_t(k)$ denote the optimal discharge amount at time t, when a discharge permission arrives ($t=\tau_{0,i}$ for some i) and the inventory level is k. It follows from (5) that the optimal discharge amount is given by

$$a_t(k)\in\arg\max_{a\in\mathcal{A}_k}\ \{R_t(a)+V_t(k-a)\}. \tag{6}$$

To avoid ambiguity, we assume that if the maximizer in (6) is not unique, then $a_t(k)$ is the smallest maximizer. We set $a_t(0)=0$, for all $t\leq T$.

When the value functions $V_t(k)$ are determined, the surface $\{a_t(k)\}_{k\in[0,K],t\in[0,T]}$, computed from (6), is used in conjunction with (2) to react to the arrivals of discharge permissions in an optimal way. Thus, it is enough to fully specify the value function $V_t(k)$ for each $0\leq k\leq K$ and $0\leq t\leq T$. In the subsequent section, we analyze several structural properties of the value functions which we later use to characterize the optimal discharges under the proposed storage deployment program.

3. Structure of the Value Function

A simple observation is that the value function $V_t(k)$ is nonincreasing over t, and nondecreasing over k. This is formalized in the following two propositions.

Proposition 3.1 *For any inventory level k, $V_t(k)$ is decreasing in t.*

Proof. For times t_1 and t_2 with $t_1<t_2$, let k be the inventory level. Consider an optimal discharge policy $\pi_2\in\Pi_{t_2}$ over $[t_2,T]$ starting at the state $x_{t_2}^{\pi_2}=k$, i.e.,

$$x^{\pi_2}\in\arg\max_{\pi\in\Pi_{t_2}} V_{t_2}^{\pi}(k).$$

Then the policy π_1 resulting in the inventory process $x^{\pi_1}=\{x_t^{\pi_1}\}_{t\in[t_1,T]}$ in which $x_t^{\pi_1}=k$ for $t\in[t_1,t_2]$ and $x_t^{\pi_1}=x_t^{\pi_2}$ for $t\in(t_2,T]$ is an admissible control for discharging k units over $[t_1,T]$, which yields $V_{t_1}(k)\geq V_{t_1}^{\pi_1}(k)=V_{t_2}^{\pi_2}(k)=V_{t_2}(k)$. $\square$

Proposition 3.2 *For any time $t \in [0,T]$, $V_t(k)$ is increasing in k.*

Proof. Fix some inventory levels k_1, k_2 at time t, such that $0 < k_1 \leq k_2$. Let $\pi_1 \in \Pi_t$ be an optimal policy over $[t,T]$ from the state k_1, i.e., $x_t^{\pi_1} = k_1$ and $V_t(k_1) = V_t^{\pi_1}(k_1)$. We have

$$\begin{aligned} V_t(k_1) &= \mathbb{E}\left[\sum_{i=1}^{N_{T-}-N_{t-}} R_{\tau_{i,t}}\left(x^{\pi_1}_{\tau_{i-1,t}} - x^{\pi_1}_{\tau_{i,t}}\right) + R_T(x_T^{\pi_1}) \;\middle|\; x_t^{\pi_1} = k_1\right] \\ &= \mathbb{E}\left[R_{\tau_{1,t}}\left(k_1 - x^{\pi_1}_{\tau_{1,t}}\right) + \sum_{i=2}^{N_{T-}-N_{t-}} R_{\tau_{i,t}}\left(x^{\pi_1}_{\tau_{i-1,t}} - x^{\pi_1}_{\tau_{i,t}}\right) + R_T(x_T^{\pi_1})\right] \\ &\leq \mathbb{E}\left[R_{\tau_{1,t}}\left(k_2 - x^{\pi_1}_{\tau_{1,t}}\right) + \sum_{i=2}^{N_{T-}-N_{t-}} R_{\tau_{i,t}}\left(x^{\pi_1}_{\tau_{i-1,t}} - x^{\pi_1}_{\tau_{i,t}}\right) + R_T(x_T^{\pi_1})\right] \qquad (7) \\ &= V_t^{\pi_2}(k_2), \end{aligned}$$

where the policy π_2 starting from the state k_2 is an admissible policy over $[t,T]$, with the inventory process $x_s^{\pi_2} = k_2$ for $s \in [t, \tau_{1,t})$ and $x_s^{\pi_2} = x_s^{\pi_1}$ for $s \in [\tau_{1,t}, T]$. The inequality (7) comes from the assumption that the reward function R_t is increasing in the amount discharged and $0 < k_1 \leq k_2$. Hence, $R_{\tau_{1,t}}(k_1 - x_{\tau_{1,t}}) \leq R_{\tau_{1,t}}(k_2 - x_{\tau_{1,t}})$. Therefore,

$$V_t(k_2) = \max_{\pi \in \Pi_t} V_t^{\pi}(k_2) \geq V_t^{\pi_2}(k_2) \geq V_t(k_1),$$

which completes the proof. □

Next, we show that the value function is monotone in the discharge permission rate λ.

Proposition 3.3 *For any time $t \in [0,T]$ and inventory level k, the value function $V_t(k)$ is increasing in the arrival rate λ.*

Proof. Let λ_1 and λ_2 be arrival rates with $\lambda_1 < \lambda_2$. Let V_t^1 and V_t^2 be the corresponding value functions, i.e., V_t^1 measures cumulated rewards in expectation over a Poisson input process $\{N_s^1\}_{s \geq t}$ of arrival rate λ_1, while V_t^2 measures the expected cumulated rewards over a Poisson input process $\{N_s^2\}_{s \geq t}$ of rate λ_2.

Define $p \stackrel{\text{def}}{=} \frac{\lambda_1}{\lambda_2}$. Now, let $\{Z_i\}_{i \geq 1}$ be an i.i.d. sequence of binary random variables, independent of the Poisson processes, such that $\Pr(Z_i = 1) = p$ and $\Pr(Z_i = 0) = 1 - p$.

Let the process $\{Z_i\}_{i \geq 1}$ label each arrival from the input process N_s^2. Recall that the process that counts the points labeled with ones up to time s is a Poisson process with the rate $p\lambda_2$. By our choice of p, $p\lambda_2 = \lambda_1$. This means N_s^2 compounded with Z_i defines an input process distributed as N_s^1.

Let π_1 be an optimal discharge policy that attains V_t^1 under the Poisson process $\{N_s^1\}_{s\geq t}$. Define a discharge policy π_2 adapted to N_t^2 and Z_i as follows,

$$dx_t^{\pi_2} \stackrel{\text{def}}{=} -Z_i A_t^{\pi_1}\left(x_{t^-}^{\pi_1}\right) dN_t^2, \qquad \forall t \in (0,T),$$

where $x_0^{\pi_2} = k$ and $x_T^{\pi_2} = x_{T^-}^{\pi_2}$. Here, Z_i plays the role of a coin-flipping process which, at each new arrival i from N_s^2, occurring at time $\tau_{i,t}$, permits to discharge the amount $A_{\tau_i}^{\pi_1}(x_{\tau_{i,t}^-}^{\pi_1})$ with probability p, or prevents it with probability $1-p$.

It follows that under the Poisson process N_t^2, π_2 attains the value $V_t^1(k)$ for each k. Therefore we can conclude

$$\begin{aligned} V_t^2(k) &= \max_{\pi\in\Pi_t} \mathbb{E}\left[\sum_{i=1}^{N_{T^-}^2 - N_{t^-}^2} R_{\tau_{i,t}}(x_{\tau_{i,t}}^{\pi} - x_{\tau_{i-1,t}}^{\pi}) + R_T(x_T^{\pi})\right] \\ &\geq \mathbb{E}\left[\sum_{i=1}^{N_{T^-}^2 - N_{t^-}^2} R_{\tau_{i,t}}(x_{\tau_{i,t}}^{\pi_2} - x_{\tau_{i-1,t}}^{\pi_2}) + R_T(x_T^{\pi_2})\right] \\ &= \mathbb{E}\left[\sum_{i=1}^{N_{T^-}^1 - N_{t^-}^1} R_{\tau_{i,t}}(x_{\tau_{i,t}}^{\pi_1} - x_{\tau_{i-1,t}}^{\pi_1}) + R_T(x_T^{\pi_1})\right] \\ &= V_t^1(k), \end{aligned}$$

where the first expectation is over N_t^2, the second expectation is over N_t^2 and Z_i, and the third expectation is over N_t^1. This completes the proof of $V_t^1(k) \leq V_t^2(k)$. $\square$

Monotonicity of the value function in the inventory level addressed in Proposition 3.2 implies that $V_t(k) \geq V_t(0) = 0$ for any $k \geq 0$. The following proposition indicates that the value function is also bounded above when the reward function is bounded.

Proposition 3.4 *Let $c_r := \max_{t\in[0,T]} R_t(K)$. Then, $V_t(k) \leq (1+\lambda T)\, c_r$, for all $t \in [0,T]$.*

Proof.

$$\begin{aligned} V_t(k) &= \max_{\pi\in\Pi_t} \mathbb{E}\left[\sum_{i=1}^{N_{T^-}-N_{t^-}} R_{\tau_{i,t}}\left(x_{\tau_{i-1,t}}^{\pi} - x_{\tau_{i,t}}^{\pi}\right) + R_T(x_T^{\pi}) \;\middle|\; x_t^{\pi} = k\right] \\ &\leq \mathbb{E}\left[\max_{\pi\in\Pi_t^+} \sum_{i=1}^{N_{T^-}-N_{t^-}} R_{\tau_{i,t}}\left(x_{\tau_{i-1,t}}^{\pi} - x_{\tau_{i,t}}^{\pi}\right) + R_T(x_T^{\pi}) \;\middle|\; x_t^{\pi} = k\right], \end{aligned}$$

where Π_t^+ is the extension of Π to the set of policies defined over $[t,T]$ that are $\mathcal{F}_T$ measurable. This means that the controlled process can now peak into the future until time T. Since the reward function is bounded, we have $R_{\tau_{i,t}}\left(x^{\pi}_{\tau_{i-1,t}} - x^{\pi}_{\tau_{i,t}}\right) \leq c_r$ almost surely (a.s.), for all $i = 1, \cdots, N_{T^-} - N_{t^-}$ and $R_{\tau_{i,t}}(x_T) \leq c_r$ a.s. Therefore, we have

$$V_t(k) \leq \mathbb{E}\left[\max_{\pi \in \Pi_t^+} c_r \left(N_{T^-} - N_{t^-} + 1\right) \;\middle|\; x_t = k\right] = c_r \left(\lambda(T-t) + 1\right) \leq c_r \left(\lambda T + 1\right).$$

This completes the proof. □

Next, we establish the concavity of the value function given that the reward functions R_t are concave in the inventory level. The proof of Proposition 3.5 is given in Appendix A.

Proposition 3.5 *For any time $t \in [0,T]$, the value function $V_t(k)$ is concave in k.*

The concavity of the value function $V_t(k)$ in the inventory level k implies the continuity of $V_t(k)$ in k on $[0,K]$, e.g., see Corollary 2.37 in Rockafellar and Wets (1998). In fact since $[0,K]$ is a nonempty closed and bounded subset of $\mathbb{R}$, $V_t(k)$ is uniformly continuous in k. The following proposition addresses the continuity of the value function in t. We provide a proof of Proposition 3.6 in Appendix A.

Proposition 3.6 *For any inventory level k, $V_t(k)$ is uniformly continuous in t.*

In the following section, we derive the system of partial differential equations that will be satisfied by the value function $V_t(k)$. This equation is the building block of our computational scheme to derive an optimal policy.

4. Computational Scheme

For any reward function $R_t(\cdot)$, the value function $V_t(k)$ can be computed using the Euler scheme (e.g., see Judd (1998), Kushner and Dupuis (2001)) over a time grid with the following difference equation

$$V_{t+\delta t}(k) \approx V_t(k) + \delta t \, \frac{\partial V_t(k)}{\partial t}. \tag{8}$$

The dynamic programming equation (5) for our control problem leads us to the computation of $\frac{\partial V_t(k)}{\partial t}$.

Proposition 4.1 *The derivative of the value function with respect to time equals*

$$\frac{\partial V_t(k)}{\partial t} = \lambda \left(V_t(k) - \max_{a \in \mathcal{A}_k} \left(R_t(a) + V_t(k-a)\right)\right), \tag{9}$$

where λ is the constant intensity of the discharge permissions arrival process.

Proof. Consider the time interval $[t-\delta t, t)$, where $\delta t > 0$ is a small real. Denote $A \stackrel{\text{def}}{=} \{\tau_{1,t-\delta t} > t\}$, $B \stackrel{\text{def}}{=} \{\tau_{1,t-\delta t} \leq t, \tau_{2,t-\delta t} > t\}$, and $C \stackrel{\text{def}}{=} (A \cup B)^c$. By the dynamic programming principle the value function V_t satisfies

$$V_{t-\delta t}(k) = \mathbb{E}\left[V_t(k) \cdot 1_A + X_B \cdot 1_B + X_C \cdot 1_C\right],$$

where $X_B \stackrel{\text{def}}{=} R_{\tau_{1,t-\delta t}}(a_{\tau_{1,t-\delta t}}(k)) + V_t(k - a_{\tau_{1,t-\delta t}}(k))$ and where X_C is a bounded random variable due to the fact that the rewards are bounded. Here, $a_{\tau_{1,t-\delta t}}(k)$ is defined as in (6) at time $\tau_{1,t-\delta t}$. The events A, B, C are $\mathcal{F}_t$-measurable.

The definitions of the events yield $\Pr(A) = e^{-\lambda\delta t}$, $\Pr(B) = \lambda\delta t e^{-\lambda\delta t}$, and $\Pr(C) = o(\delta t)$. Hence,

$$\begin{aligned} V_{t-\delta t}(k) &= V_t(k)\Pr(A) + \mathbb{E}\left[X_B|B\right]\Pr(B) + \mathbb{E}[X_C|C]\Pr(C) \\ &= V_t(k)e^{-\lambda\delta t} + \mathbb{E}\left[X_B|B\right]\lambda\delta t e^{-\lambda\delta t} + \mathbb{E}[X_C|C]o(\delta t). \end{aligned}$$

Using this equality, we have

$$\begin{aligned} \tfrac{\partial V_t(k)}{\partial t} &= \lim_{\delta t\to 0} \frac{V_t(k) - V_{t-\delta t}(k)}{\delta t} \\ &= \lim_{\delta t\to 0} \frac{(1-e^{-\lambda\delta t})V_t(k) - \mathbb{E}\left[X_B|B\right]\lambda\delta t e^{-\lambda\delta t} - \mathbb{E}[X_C|C]o(\delta t)}{\delta t} \\ &= \lambda V_t(k) - \lambda \lim_{\delta t\to 0} \mathbb{E}\left[X_B|B\right]. \end{aligned} \tag{10}$$

Using Proposition 3.4, for every instance $\tau_{1,t-\delta t}(\omega) \in [t-\delta t, t]$ we have

$$|X_B(\omega)| = \left|R_{\tau_{1,t-\delta t}(\omega)}(a_{\tau_{1,t-\delta t}(\omega)}(k)) + V_t(k - a_{\tau_{1,t-\delta t}(\omega)}(k))\right| \leq c_r + (1+\lambda T)c_r.$$

The bounded convergence theorem for expectations (e.g., see Cinlar (2011)) implies that the limit and the expectation in equation (10) can be interchanged. Therefore

$$\lim_{\delta t\to 0} \mathbb{E}\left[X_B \mid B\right] = \lim_{\delta t\to 0} \mathbb{E}\left[R_\nu(a_\nu + V_t(k - a_\nu(k))\right] = \mathbb{E}\left[\lim_{\delta t\to 0}\left\{R_\nu(a_\nu + V_t(k - a_\nu(k))\right\}\right] \tag{11}$$

where ν has the distribution of $\tau_{1,t-\delta t}$ given B.

For any instance $\nu(\omega) \in [t-\delta t, t]$, we have

$$\begin{aligned} &\lim_{\delta t\to 0}\left\{R_{\nu(\omega)}(a_{\nu(\omega)}(k)) + V_t(k - a_{\nu(\omega)}(k))\right\} \\ &= \lim_{\delta t\to 0}\left\{R_{\nu(\omega)}\left(a_{\nu(\omega)}(k)\right) + V_{\nu(\omega)}\left(k - a_{\nu(\omega)}(k)\right) - V_{\nu(\omega)}\left(k - a_{\nu(\omega)}(k)\right) + V_t\left(k - a_{\nu(\omega)}(k)\right)\right\} \\ &= \lim_{\delta t\to 0}\left\{R_{\nu(\omega)}\left(a_{\nu(\omega)}(k)\right) + V_{\nu(\omega)}\left(k - a_{\nu(\omega)}(k)\right)\right\} - \lim_{\delta t\to 0}\left\{V_{\nu(\omega)}\left(k - a_{\nu(\omega)}(k)\right) - V_t\left(k - a_{\nu(\omega)}(k)\right)\right\} \\ &= \lim_{\delta t\to 0}\left\{\max_{a\in\mathcal{A}_k}\left(R_{\nu(\omega)}(a) + V_{\nu(\omega)}(k-a)\right)\right\} - \lim_{\delta t\to 0}\left\{V_{\nu(\omega)}\left(k - a_{\nu(\omega)}(k)\right) - V_t\left(k - a_{\nu(\omega)}(k)\right)\right\}. \end{aligned} \tag{12}$$

According to Proposition 3.6, the value function $V_t(k)$ is uniformly continuous in t on $\mathcal{A}_k \cap [0, K]$. In addition, as $\delta t \to 0$, $\nu \to t$ almost surely. Therefore, for any $a \in \mathcal{A}_k \cap [0, K]$ and the instance $\nu(\omega) \in [t-\delta t, t]$ we have

$$\lim_{\delta t\to 0} V_{\nu(\omega)}(k-a) = \lim_{\nu(\omega)\to t} V_{\nu(\omega)}(k-a) = V_t(k-a).$$

In particular, for $a = a_{\nu(\omega)}(k)$, this equation yields

$$\lim_{\delta t \to 0} \left\{ V_{\nu(\omega)} \left(k - a_{\nu(\omega)}(k) \right) - V_t \left(k - a_{\nu(\omega)}(k) \right) \right\} = 0. \tag{13}$$

At the same time, since the objective function is uniformly continuous in t, $R_{\nu(\omega)}(a) + V_{\nu(\omega)}(k-a)$ epi-converges to $R_t(a) + V_t(k-a)$ as $\delta t \to 0$ by Theorem 7.15 of Rockafellar and Wets (1998). Therefore, Theorem 7.33 of Rockafellar and Wets (1998) implies that

$$\lim_{\delta t \to 0} \left\{ \max_{a \in \mathcal{A}_k} \left(R_{\nu(\omega)}(a) + V_{\nu(\omega)}(k-a) \right) \right\} = \max_{a \in \mathcal{A}_k} \left(R_t(a) + V_t(k-a) \right). \tag{14}$$

We note that equality (14) could also be derived by Theorem 2.1 of Fiacco (1974) using the uniform continuity of $V_t(k)$ in t and the continuity of the value function and the reward function in k.

Using equations (13) and (14) in equation (12) and subsequently in equation (11) follows that

$$\lim_{\delta t \to 0} \mathbb{E}\left[X_B \mid B \right] = \mathbb{E}\left[\lim_{\delta t \to 0} \left\{ R_{\nu(\omega)}(a_{\nu(\omega)}(k)) + V_t(k - a_{\nu(\omega)}(k)) \right\} \right] = \max_{a \in \mathcal{A}_k} \left(R_t(a) + V_t(k-a) \right).$$

This equation along with (10) completes the proof. □

Using Proposition 4.1 in the Euler scheme, for small $\delta t > 0$, we have

$$\begin{aligned} V_{t-\delta t}(k) &\approx V_t(k) - \delta t \; \frac{\partial V_t(k)}{\partial t} \\ &= V_t(k) - \delta t \; \lambda \left(V_t(k) - \max_{a \in \mathcal{A}_k} \left(R_t(a) + V_t(k-a) \right) \right). \end{aligned}$$

Therefore,

$$V_{t-\delta t}(k) = (1 - \lambda \delta t) V_t(k) + \lambda \delta t \max_{a \in \mathcal{A}_k} \left(R_t(a) + V_t(k-a) \right). \tag{15}$$

This difference equation along with the boundary conditions $V_t(0) = 0$ and $V_T(k) = R_T(k)$ specifies the value function $V_t(k)$ for all $0 \le k \le K$ and $0 \le t \le T$. The optimal discharge amounts $a_t(k)$ are then determined from (6) along with the boundary conditions $a_t(0) = 0$, for all $0 \le t \le T$, and $a_T(k) = k$, for all $0 \le k \le K$. Note that since both the reward functions R_t and optimal value functions V_t are concave and the set of admissible discharges $\mathcal{A}_k$ is convex, (15) involves solving convex optimization problems.

Below, we prove that the partial derivative is monotone in the inventory level.

Corollary 4.1 *Suppose that the feasible action sets $\mathcal{A}_k$ as functions of the inventory level k are such that $k_1 \le k_2$ yields $\mathcal{A}_{k_1} \subseteq \mathcal{A}_{k_2}$. Then, $\frac{\partial V_t(k)}{\partial t}$ is decreasing in k.*

Proof. Let k_1 and k_2 be two inventory levels where $k_1 \le k_2$. Therefore, the monotonicity of feasible action sets implies that $a_t(k_1) \in \mathcal{A}_{k_1} \subseteq \mathcal{A}_{k_2}$. This along with equation (9) imply that for $\lambda > 0$,

$$\begin{aligned}\tfrac{\partial V_t(k_2)}{\partial t} &= \lambda \left(V_t(k_2) - \max_{a \in \mathcal{A}_{k_2}} \left(R_t(a) + V_t(k_2 - a) \right) \right) \\ &\le \lambda \left(V_t(k_2) - \left\{ R_t\left(a_t(k_1)\right) + V_t\left(k_2 - a_t(k_1)\right) \right\} \right). \end{aligned} \tag{16}$$

In addition, the concavity of the value function V_t in the inventory level established in Proposition 3.5 implies that it has decreasing differences. Hence,

$$\begin{aligned} & V_t(k_2) - V_t(k_1) \le V_t(k_2 - a_t(k_1)) - V_t(k_1 - a_t(k_1)), \\ \Rightarrow\ & V_t(k_2) - V_t\left(k_2 - a_t(k_1)\right) \le V_t(k_1) - V_t\left(k_1 - a_t(k_1)\right). \end{aligned} \tag{17}$$

It then follows from inequality (17) in (16) that

$$\begin{aligned}\tfrac{\partial V_t(k_2)}{\partial t} &\le \lambda \left(V_t(k_2) - R_t(a_t(k_1)) - V_t(k_2 - a_t(k_1)) \right) \qquad (18) \\ &\le \lambda \left(V_t(k_1) - R_t(a_t(k_1)) - V_t(k_1 - a_t(k_1)) \right) \\ &= \lambda \left(V_t(k_1) - \max_{a \in \mathcal{A}_{k_1}} \left(R_t(a) + V_t(k_1 - a) \right) \right) = \frac{\partial V_t(k_1)}{\partial t}, \end{aligned}$$

which shows the result. □

In the next section, we provide some insights of the structure of the optimal discharge actions.

5. Structure of the Optimal Policy

This section is devoted to addressing some properties of $a_t(k)$, defined in (6), and of the inventory process x^π. Three fundamental properties are established: (i) the optimal discharged amount is nondecreasing in the inventory level, (ii) the optimal inventory trajectory over $[0,T]$ is nondecreasing in the initial inventory level, and (iii) under mild conditions, the passage of time increases the discharged amounts. Note that property (ii) implicitly bounds how the discharged amounts increase with an increase of the initial inventory level, and therefore complements property (i). We start by establishing a monotonicity result for the optimal discharge amount.

In the following, the set $\mathcal{A}_k \subseteq \mathbb{R}$ is called ascending in k, if for any $k_1 \le k_2$ and any two elements (a,b) where $a \in \mathcal{A}_{k_1}$ and $b \in \mathcal{A}_{k_2}$, we have $\min\{a,b\} \in \mathcal{A}_{k_1}$ and $\max\{a,b\} \in \mathcal{A}_{k_2}$, see e.g., Heyman and Sobel (2003) or Topkis (1998).

Proposition 5.1 *Let the set* $\mathcal{C} \overset{\text{def}}{=} \{(k,a) \in \mathbb{R}^2 : a \in \mathcal{A}_k, k \in [0,K]\}$ *be a sublattice of* $\mathbb{R}^2$ *and* $\mathcal{A}_k$ *be ascending in* k *on* $[0,K]$*. Then, for any* $t \in [0,T]$*,* $a_t(k)$ *is an increasing function of the inventory level* k*, i.e.,* $k_1 \le k_2$ *implies* $a_t(k_1) \le a_t(k_2)$*.*

Proof. The reward function R_t is a function on $\mathbb{R}$, and consequently it is supermodular in a on $\mathbb{R}$. In addition, since the value function V_t is concave in the inventory level, Lemma 2.6.2 in Topkis (1998) implies that the function $V_t(k-a)$ is supermodular in (k,a) on $\mathbb{R}^2$. Therefore, the positive linear combination of these two supermodular functions, $R_t(a)+V_t(k-a)$, is supermodular in (k,a) on $\mathbb{R}^2$. Since $\mathcal{C}$ is a sublattice of $\mathbb{R}^2$, and $\mathcal{A}_k$ is the section of $\mathcal{C}$ at k, it follows from Theorem 2.8.2 in Topkis (1998) that the optimal solution set $\arg\max_{a\in\mathcal{A}_k} \{R_t(a)+V_t(k-a)\}$ is ascending in k on $\{k:\arg\max_{a\in\mathcal{A}_k} \{R_t(a)+V_t(k-a)\}\neq\emptyset\}=[0,K]$. Therefore, it follows from Theorem 2.8.3 of Topkis (1998) that the smallest element of the optimal solution set, $a_t(k)$, is increasing in k. □

For instance, the set $\mathcal{A}_k=[0,k]$ is ascending in k on $[0,K]$ and the set $\mathcal{C}\overset{\text{def}}{=}\{(k,a)\in\mathbb{R}^2: a\in\mathcal{A}_k, k\in[0,K]\}$ is a sublattice of $\mathbb{R}^2$. In addition, $\mathcal{A}_k=[0,k]$ has the property assumed in Corollary 4.1 that is $k_1\le k_2$ implies that $\mathcal{A}_{k_1}\subseteq\mathcal{A}_{k_2}$.

The following proposition shows that an optimal policy π started at a higher initial inventory level results in an inventory process with higher levels through the entire time horizon.

Proposition 5.2 *Let $k_1\le k_2$. Denote the inventory processes corresponding to the optimal policy started at states k_1 and k_2 at time $t=0$ with x_t^1 and x_t^2, respectively. Then for all $t\in[0,T]$, $x_t^1\le x_t^2$.*

Proof. Let $\bar{t}:=\max\{t: x_s^1\le x_s^2 \text{ for } 0\le s\le t\}$. Suppose by contradiction that $\bar{t}<T$. Then, $\bar{t}$ must be the time of a discharge permission arrival, $\bar{t}=\tau_{j,0}$ for some j, such that $x_{\bar{t}-}^2\ge x_{\bar{t}-}^1$ and $x_{\bar{t}}^2<x_{\bar{t}}^1$.

Let $a^1\in\operatorname{argmax}_{a\in[0,x_{t-}^1]}\left\{R_t(a)+V_t(x_{t-}^1-a)\right\}$. Therefore, $x_t^1=x_{t-}^1-a^1$ and

$$R_t\left(a^1\right)+V_t\left(x_{t-}^1-a^1\right)\ge R_t\left(a\right)+V_t\left(x_{t-}^1-a\right)\quad\forall a\in[0,x_{t-}^1].$$

In particular, for $a=(x_{t-}^1-x_t^2)\in[0,x_{t-}^1]$ this inequality implies that

$$\begin{aligned}R_t\left(a^1\right)+V_t\left(x_{t-}^1-a^1\right) &\ge R_t\left(x_{t-}^1-x_t^2\right)+V_t\left(x_{t-}^1-\left(x_{t-}^1-x_t^2\right)\right)\\ &= R_t\left(x_{t-}^1-x_t^2\right)+V_t\left(x_t^2\right).\end{aligned}\tag{19}$$

Let a^2 be the smallest element of the solution set $\operatorname{argmax}_{a\in[0,x_{t-}^2]}\left\{R_t(a)+V_t(x_{t-}^2-a)\right\}$. We have $x_t^2=x_{t-}^2-a^2$ and

$$R_t\left(a^2\right)+V_t\left(x_{t-}^2-a^2\right)\ge R_t(a)+V_t\left(x_{t-}^2-a\right)\quad\forall a\in[0,x_{t-}^2].\tag{20}$$

Since a^2 is the smallest maximizer, inequality (20) must hold strictly for any $a<a^2$. In particular, for $(x_{t-}^2-x_t^1)\in[0,x_{t-}^2]$, it follows from the strict inequality $x_t^2<x_t^1$ that $x_{t-}^2-x_t^1<x_{t-}^2-x_t^2=a^2$. Hence,

$$\begin{aligned}R_t\left(a^2\right)+V_t\left(x_{t-}^2-a^2\right) &> R_t\left(x_{t-}^2-x_t^1\right)+V_t\left(x_{t-}^2-\left(x_{t-}^2-x_t^1\right)\right)\\ &= R_t\left(x_{t-}^2-x_t^1\right)+V_t\left(x_t^1\right).\end{aligned}\tag{21}$$

By combining inequalities (19) and (21), we arrive at

$$\begin{aligned} R_t\left(x_{t-}^1 - x_t^2\right) - R_t\left(a^1\right) &\le V_t\left(x_{t-}^1 - a^1\right) - V_t(x_t^2) && (22)\\ &= V_t\left(x_t^1\right) - V_t\left(x_{t-}^2 - a^2\right) && (23)\\ &< R_t\left(a^2\right) - R_t\left(x_{t-}^2 - x_t^1\right). && (24)\end{aligned}$$

Here inequalities (22) and (24) are rearrangements of inequalities (19) and (21), and the equality (23) comes from the facts that $x_t^1 = x_{t-}^1 - a^1$ and $x_t^2 = x_{t-}^2 - a^2$.

On the other hand, it follows from the concavity of R_t that it has decreasing differences. Thus,

$$\begin{aligned} R_t\left(x_{t-}^1 - x_t^2\right) - R_t(a^1) &= R_t\left(a^1 + (a^2 - x_{t-}^2 + x_t^1)\right) - R_t\left(a^1\right)\\ &\ge R_t\left(x_{t-}^2 - x_t^1 + (a^2 - x_{t-}^2 + x_t^1)\right) - R_t\left(x_{t-}^2 - x_t^1\right)\\ &= R_t(a^2) - R_t(x_{t-}^2 - x_t^1). \end{aligned}$$

which is in contradiction with inequality (24). Thus the supposition that $\bar{t} < T$ cannot be true, i.e., for all $t \in [0,T]$, we must have $x_t^1 \le x_t^2$. □

The following proposition discusses the monotonicity of $a_t(k)$ in t. It shows that as time approaches to the end of horizon, the participating storage unit discharges in larger amounts.

Proposition 5.3 *Suppose that $\frac{\partial R_t(a)}{\partial t}$ is increasing in a and that the set of admissible actions $\mathcal{A}_k$ is such that $k_1 \le k_2$ yields $\mathcal{A}_{k_1} \subseteq \mathcal{A}_{k_2}$. Then for any inventory level k, $a_t(k)$ is increasing in t, i.e., $t_1 < t_2$ yields $a_{t_1}(k) \le a_{t_2}(k)$.*

Proof. Fix k and let $t_2 > t_1$. For any $b < a_{t_1}(k)$, $b \notin \operatorname{argmax}_{a\in\mathcal{A}_k} \{R_{t_1}(a) + V_{t_1}(k-a)\}$. Therefore,

$$\begin{aligned} &R_{t_1}(b) + V_{t_1}(k-b) < R_{t_1}(a_{t_1}(k)) + V_{t_1}\left(k - a_{t_1}(k)\right),\\ \Rightarrow\ &V_{t_1}(k-b) - V_{t_1}\left(k - a_{t_1}(k)\right) < R_{t_1}\left(a_{t_1}(k)\right) - R_{t_1}(b). \qquad (25)\end{aligned}$$

On the other hand, the assumption that $\frac{\partial R_{t_2}(a)}{\partial t}$ is increasing in a implies that $\frac{\partial R_{t_2}(b)}{\partial t} \le \frac{\partial R_{t_2}(a_{t_1}(k))}{\partial t}$. Therefore,

$$\begin{aligned} &R_{t_2}(b) - R_{t_1}(b) \le R_{t_2}(a_{t_1}(k)) - R_{t_1}(a_{t_1}(k)),\\ \Rightarrow\ &R_{t_1}(a_{t_1}(k)) - R_{t_1}(b) \le R_{t_2}(a_{t_1}(k)) - R_{t_2}(b).\end{aligned}$$

Combining the recent inequality in inequality (25) results in

$$V_{t_1}(k-b) - V_{t_1}\left(k - a_{t_1}(k)\right) < R_{t_2}(a_{t_1}(k)) - R_{t_2}(b). \qquad (26)$$

According to Corollary 4.1, $\frac{\partial V_{t_2}(k)}{\partial t}$ is decreasing in k. In particular, $k-b > k - a_{t_1}(k)$ implies that $\frac{\partial V_{t_2}(k-b)}{\partial t} \leq \frac{\partial V_{t_2}(k-a_{t_1}(k))}{\partial t}$. Using $t_2 \geq t_1$, we have

$$V_{t_2}(k-b) - V_{t_1}(k-b) \leq V_{t_2}\left(k - a_{t_1}(k)\right) - V_{t_1}\left(k - a_{t_1}(k)\right).$$

Using this inequality and inequality (26) we get

$$\begin{aligned} V_{t_2}(k-b) - V_{t_2}\left(k - a_{t_1}(k)\right) &\leq V_{t_1}(k-b) - V_{t_1}\left(k - a_{t_1}(k)\right) \\ &< R_{t_2}(a_{t_1}(k)) - R_{t_2}(b). \end{aligned}$$

A rearrangement of the recent inequality equals

$$R_{t_2}(b) + V_{t_2}(k-b) < R_{t_2}(a_{t_1}(k)) + V_{t_2}\left(k - a_{t_1}(k)\right),$$

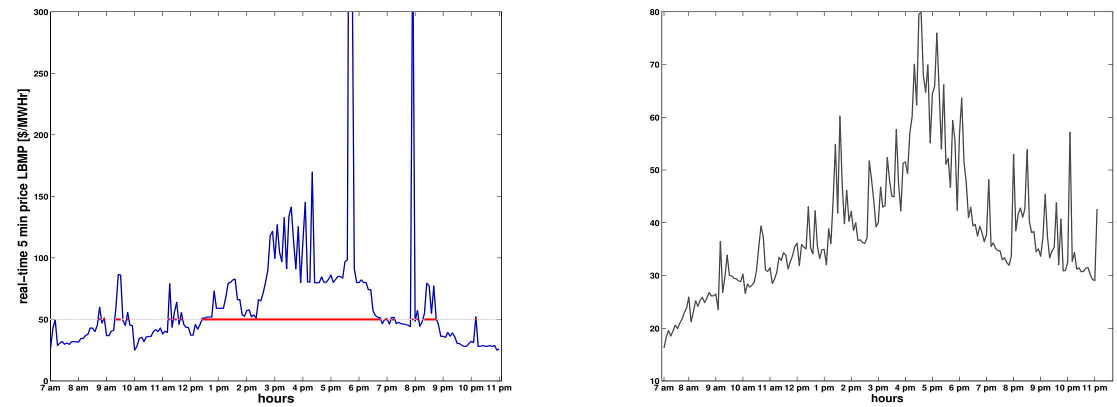
which indicates that $a_{t_1}(k)$ achieves a superior value for $R_{t_2}(a) + V_{t_2}(k-a)$ than b. Hence, b cannot be in the solution set $\text{argmax}_{a\in\mathcal{A}_k}\ \{R_{t_2}(a) + V_{t_2}(k-a)\}$. Since this is true for any $b < a_{t_1}(k)$, we can conclude that $a_{t_2}(k) \geq a_{t_1}(k)$. □

Next we present our computational investigation of the optimal value function and optimal discharge decisions of a storage unit participating in the considered flexible storage deployment program.

6. Numerical Examples

We consider a storage device of capacity $K = 4$, participating in the energy storage deployment program in New York City (load zone J in NYISO). The participating storage unit may be called to discharge power at random during peak hours of a day. More precisely, the decisions span the 7am-11pm time period. We assume that no discharge permission is issued during off-peak hours, which is a time when storage owners are likely to charge their storage devices. The value of the stored power at the end of the horizon is assumed to be zero, i.e., $R_T(a) = 0$, for all a.

We assume that the discharge permission events are triggered by the zonal electricity price levels beyond a threshold. The real time 5-min prices over peak hours [7am-11pm] for August 25, 2015 and the threshold price 50[$/MWHr] [1] is illustrated in Figure 1(a). Given the threshold price level, the discharge permission event is triggered 100 times on August 25, 2015, which is the highest number of realized discharge permission arrivals per day in August 2015. These discharge permission time slots correspond to the times specified by the solid red line in Figure 1(a). For a given price threshold, we estimated the daily arrival rate using the real time 5-min prices of peak hours from August 1, 2015 to August 31, 2015. For example, for the threshold price equal to 50[$/MWHr], the average arrival rate over peak hours is 24.9355 per day. For threshold prices



(a) August 25, 2015 (b) average real-time 5-min prices in August 2015

Figure 1 (a) Real time 5-min prices in N.Y.C. zone and the discharge permission times on August 25, 2015. Discharge permission times are indicated by the solid red line. (b) Time-varying reward coefficients, approximated by the mean real time prices in August 2015 in the N.Y.C. zone.

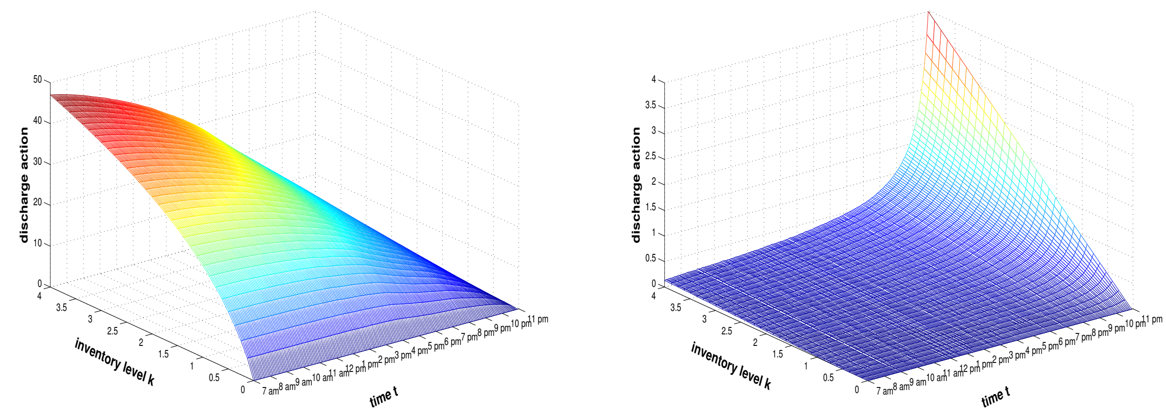
100[$/MWHr] and 25[$/MWHr], the discharge permission arrival rates become 4.7742 per day and 144.7742 per day, respectively.

We consider the log reward function $R_t(a) = \log(1 + p_t a)$ as the reward function for $t \in [0, T)$. The average real time hourly electricity price over one month is used as a proxy of the reward coefficient p_t at every time $t \in$[7am-11pm]. The average real time hourly price curve for August 2015 is depicted in Figure 1(b).

Figure 2 summarizes the results from the computational scheme in §4 with $\delta t = 5$ min. Here, the permission arrival rate is set to $\lambda = 24.9355/16 = 1.5585$ per hour and the reward coefficient p_t at every time t is obtained from the curve in Figure 1(b). The left plot illustrates the value function and the right plot depicts the optimal actions $a_t(k)$. The left plot confirms the results in §3 on the concavity of the value function in the inventory level k and its monotonicity in the inventory level k and time t. The right plot also illustrates the structures analyzed in §5 that the discharge amount $a_t(k)$ is increasing in the inventory level k.

Next we investigate the structure of the value function and actions for the log reward function, as the discharge permission arrival rate λ increases. The arrival rate is an important parameter that can be controlled by the utility. The analysis for the inventory level $k = 2$ and for four values of λ, namely $\lambda = \lambda_0$, $\lambda = 3\lambda_0$, $\lambda = 6\lambda_0$, and $\lambda = 10\lambda_0$, is reported in Figure 3. The left plot depicts the value function $V_t(2)$, which increases with the arrival rate λ. This observation is consistent with Proposition 3.3.

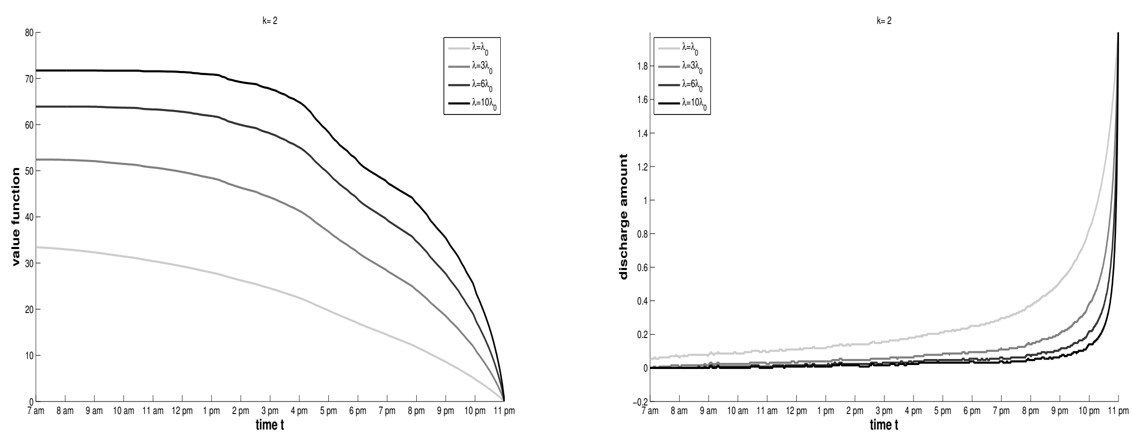
Figure 3(b) shows the optimal discharge amounts $a_t(2)$ for the four values of discharge permission rates. As the curve corresponding to $\lambda = \lambda_0$ indicates, $a_t(2)$ is nonzero even at times closer to the



(a) optimal value function (b) optimal actions $a_t(k)$

Figure 2 Results for log reward function $R_t(a) = \log(p_t a + 1)$.

beginning of the time horizon. When the expectation for having more discharge permissions is low, which corresponds to a smaller rate λ, the storage owner uses any given opportunity to discharge even if the time does not correspond to the best reward value. As the discharge permission rate increases, the optimal action is to discharge more patiently and in larger amounts when the end of time horizon is approached.



(a) optimal value function $V_t(2)$ (b) optimal actions $a_t(2)$

Figure 3 Sensitivity of the optimal value function and optimal actions to λ for inventory level $k = 2$. Here, $\lambda_0 = 1.5585$.

7. Conclusion and Discussion

In this paper, we develop a novel approach to allow individual energy storage owners and developers to inject electricity to the grid without participating in the wholesale electricity market, dealing with the bidding process, and bearing the risk of commitments. Unlike models studied in the literature on storage operation optimization, the framework in this paper has the unique feature that neither the utility nor the storage owner need to commit in advance to buying or selling electricity.

In this model, electric utilities or load serving entities communicate a need for this flexible capacity. The participating storage resource receives discharge permissions that arrive randomly at a certain rate. The storage unit has then the option to either discharge some of the stored electricity or to wait for potential future discharge permissions which however may not materialize. The concavity of the reward function in the discharged quantity makes it more profitable to release energy more often and in smaller amounts, but there is uncertainty on the number of permissions to be received within the time window. The discharge permissions could be triggered by spikes in the demand or electricity price, or other driving factors.

The results established in this paper have immediate relevance for both energy storage owners as well as electricity distribution companies, energy policy makers, and contract underwriters. The computed optimal discharge policy can be used by the storage owners to obtain a more precise valuation of the energy storage unit and support their investment decisions. In addition, the derivation of the optimal discharge policy as well as the structural properties of the storage owner's value function and optimal discharge policies allow the utilities or other entities involved in the contract to predict the response behavior of the storage controllers and better design the details of the program. Also, the results can be used to better understand the economic value of this program among developers and if needed design alternative incentives.

Our results imply that this model is an interesting framework for further research and applications in the efficient energy storage deployment. While our focus in this paper was on simple Poisson arrival model with deterministic reward curve, extending our work to more complicated models in which the arrival rate is time-varying, the reward function is also random, or each discharge permission communication includes a set of real-time constraints would also be of interest.

Finally, several possible extensions of the framework proposed in this paper are sketched below.

7.1. Disaggregation of Discharge Permissions

The expected behavior of discharge amounts observed here for different values of λ suggests that it is better for the utility company to make n contracts with smaller permission rates λ than entering in one contract where the agreed permission rate equals $n\lambda$. Since in the first case, when the utility company needs electricity and thus communicates discharge permissions to the n participating storage units, they will all discharge some nonzero amount and collectively they inject $\sum_{i=1}^{n} a_t(k_i)$ to the network, assuming that they follow the optimal policy. While in the latter case, the storage unit is likely not to discharge and to wait for the permission times with higher rewards which are likely to come.

In addition, it is noteworthy to observe that if permissions sent at a rate λ to a certain area were assigned with probability p_j to a storage resource j within that area, then the input process seen by resource j is again Poisson, with rate $p_j\lambda$. Hence, the mathematical treatment of the one-storage resource presented in this paper can be seen as a fundamental building block for the coordination of multiple storage resources.

7.2. Advance Notice on the Number of Discharge Permissions

A close model to the scheme presented in this paper is when the number of discharge permissions is declared to be $N_{T^-} = n$ in advance at time $t = 0$, before the n discharge permissions arrive during the day at random times. Assuming that the number of permissions follows a Poisson distribution with rate λT, the advance notice on the exact number of discharge permissions per day increases the expected value of the contract for the storage owner. This is formalized in the following proposition, along with some quantitative results. Recall that the joint distribution of $\tau_{1,0}, \cdots, \tau_{n,0}$ conditioned on the event $N_{T^-} = n$ coincides with the joint distribution of $u_1, \cdots, u_n$, where the growing sequence $u_1 < \cdots < u_n$ is a re-arrangement of a sequence of independent random variables which are uniformly distributed on $[0, T)$.

Proposition 7.1 *Let $V_0^+(k,n)$ denote the optimal value function at time $t = 0$ when it is known that n arrivals occur, that is,*

$$V_0^+(k,n) = \max_{x \in \mathcal{X}_0^+} \mathbb{E}\left[\sum_{i=1}^{N_{T^-}} R_{\tau_{i,0}}\left(x_{\tau_{i-1,0}} - x_{\tau_{i,0}}\right) + R_T(x_T) \middle|\, x_t = k,\ N_{T^-} = n \right],$$

where the controlled inventory process x_t is now measurable with respect to $\sigma(\{N_s : 0 \le s \le t\} \cup \{N_{T^-}\})$.

Then

1. $V_0(k) \le \mathbb{E}\left[V_0^+(k, N_{T^-})\right]$*, where the expectation is over N_{T^-} which follows a Poisson distribution with parameter λT. Recall that $V_0(k)$ is defined as in* (1)*.*

2. *The value function* $V_t^+(k,n)$ *for* $k \in [0,K]$ *and* $n = 0,1,\dots$ *is described by the recurrence relations*

$$\begin{aligned} V_t^+(k,0) &= R_T(k), \\ V_t^+(k,n) &= \mathbb{E}_{\tau_{1,t}}\left[\max_{a\in\mathcal{A}_k}\left\{R_{\tau_{1,t}}(a) + V^+_{\tau_{1,t}}(k-a,n-1)\right\} \mid N_{T^-} - N_{t^-} = n\right], \quad \textit{for } n \geq 1, \end{aligned}$$

where $\tau_{1,t}$*, given* $N_{T^-} - N_{t^-} = n$*, is* $\tau_{1,t} = t + (T-t)B$*, where the random variable* B *is distributed as*

$$\Pr(B \geq u) = (1-u)^n \quad \textit{for } 0 \leq u \leq 1.$$

3. *There exists a (time-varying) Markov policy* $\{A_s^\pi(k,n)\}_{t\leq s\leq T}$ *attaining* $V_t^+(k,n)$*, which only depends on the inventory level* k *and the count arrivals-to-come* n*.*

Proof. The first part is established by observing that

$$\begin{aligned} V_0(k) &= \max_{x\in\mathcal{X}_t^+} \mathbb{E}\left[\sum_{i=1}^{N_{T^-}} R_{\tau_{i,t}}\left(x^\pi_{\tau_{i-1,t}} - x^\pi_{\tau_{i,t}}\right) + R_T\left(x_T^\pi\right) \;\middle|\; x_t^\pi = k\right] \\ &= \max_{x\in\mathcal{X}_t^+} \mathbb{E}_{N_{T^-}}\left[\mathbb{E}_{\tau_{1,t},\cdots,\tau_{n,t}}\left[\sum_{i=1}^{n} R_{\tau_{i,t}}\left(x^\pi_{\tau_{i-1,t}} - x^\pi_{\tau_{i,t}}\right) + R_T\left(x_T^\pi\right) \;\middle|\; x_t^\pi = k, N_{T^-} = n\right]\right] \\ &\leq \mathbb{E}_{N_{T^-}}\left[\max_{x\in\mathcal{X}_t^+} \mathbb{E}_{\tau_{1,t},\cdots,\tau_{n,t}}\left[\sum_{i=1}^{n} R_{\tau_{i,t}}\left(x^\pi_{\tau_{i-1,t}} - x^\pi_{\tau_{i,t}}\right) + R_T\left(x_T^\pi\right) \;\middle|\; x_t^\pi = k, N_{T^-} = n\right]\right] \\ &= \mathbb{E}_{N_{T^-}}\left[V_0^+\left(k, N_{T^-}\right)\right]. \end{aligned}$$

The second part is established by noting that each new arrival decrements the number of future arrivals, and that the time $\tau_{1,t}$ of the upcoming future arrival given $N_{T^-} - N_{t^-} = n \geq 1$ arrivals to come follows the distribution of the minimum of n independent random variables uniformly distributed over the interval $[t,T)$.

The third part is an immediate consequence of the recurrence relation followed by the structure of the reward functions. □

7.3. Extension to Self-Exciting Point Processes

The generalization of the arrival model to self-exciting point processes (see e.g., Daley and Vere-Jones (2003)) can be well suited to the modeling of permissions which are arriving in clusters. This may happen, for example when discharge permissions are driven by high demand levels or network perturbations, in which cases the occurrence of past discharge permission arrivals may increase the probability of occurrence of future permission arrivals.

One example leading to tractable computations is the Hawkes process with exponential decay. The arrival rate is described by the conditional intensity

$$\lambda_t = \lambda_0 + \alpha \int_0^t e^{-\beta(t-s)} dN_s$$

with $\alpha > 0$, $\beta > 0$, $\alpha/\beta < 1$. Here, λ_0 is the base permission arrival rate.

The process $y_t := \int_0^t e^{-\beta(t-s)} dN_s$ and consequently $\lambda_t = \lambda_0 + \alpha y_t$ are Markovian, with

$$dy_t = \begin{cases} 1 & \text{with probability } (\lambda_0 + \alpha y_t)dt \\ -\beta y_t dt & \text{with probability } 1 - (\lambda_0 + \alpha y_t)dt \end{cases}$$

In this case, it is sufficient to augment the state space to also include y_t, or equivalently λ_t. The maximization is now over the time-varying Markov discharge policies $\{A_t^\pi(k, \lambda)\}_{0 \le t \le T}$ where k is the inventory and λ is the permission rate at time t.

Appendix A

In this appendix, we present the proofs of Proposition 3.5 and Proposition 3.6.

Proof of Proposition 3.5. Fix some inventory levels k_1, k_2 at time t, such that $0 < k_1 \leq k_2$. Let π_1 and π_2 be optimal Markov discharge policies, respectively, starting from the inventory level k_1 and k_2. Hence, $V_t(k_1) = V_t^{\pi_1}(k_1)$ and $V_t(k_2) = V_t^{\pi_2}(k_2)$.

For any $\alpha \in [0,1]$, define the inventory level $k^\alpha \stackrel{\text{def}}{=} (1-\alpha)k_1 + \alpha k_2$. Consider the controlled process x^α over $[t,T]$ defined as below

$$\begin{aligned}
&x_t^\alpha = k^\alpha, \\
&dx_s^\alpha = -\left((1-\alpha)A_s^{\pi_1}\left(x_{s^-}^{\pi_1}\right) + \alpha A_s^{\pi_2}\left(x_{s^-}^{\pi_2}\right)\right) dN_s, \quad \forall s \in (t,T), \\
&x_T^\alpha = x_{T-}^\alpha.
\end{aligned} \tag{27}$$

Note that in general this is not equivalent to applying some Markov strategy to x_t^α, in particular the strategy keeps track of π_1 and π_2 started at inventory levels k_1 and k_2. It follows from (27) that

$$x_s^\alpha = (1-\alpha)x_s^{\pi_1} + \alpha x_s^{\pi_2}, \quad \forall s \in [t,T]. \tag{28}$$

From the feasibility of the policies π_1 and π_2, we have

$$\begin{aligned}
&0 \leq A_s^{\pi_1}\left(x_{s^-}^{\pi_1}\right) \leq x_s^{\pi_1}, \\
&0 \leq A_s^{\pi_2}\left(x_{s^-}^{\pi_2}\right) \leq x_s^{\pi_2}.
\end{aligned}$$

Therefore,

$$0 \leq (1-\alpha)A_s^{\pi_1}\left(x_{s^-}^{\pi_1}\right) + \alpha A_s^{\pi_2}\left(x_{s^-}^{\pi_2}\right) \leq (1-\alpha)x_s^{\pi_1} + \alpha x_s^{\pi_2} = x_s^\alpha,$$

which implies that x^α in (27) is an admissible inventory process starting at k^α.

For any realization ω of the Poisson process $\{N_s\}_{s \in \mathbb{R}_+}$, equation (28) implies that the difference of inventory levels in the process x^α between two consecutive arrival times $\tau_{i-1,t}(\omega)$ and $\tau_{i,t}(\omega)$ is the convex combination of the differences of inventory levels in the processes x^{π_1} and π_2. This along with concavity of the reward function $R_{\tau_{i,t}(\omega)}$ implies that

$$\begin{aligned}
R_{\tau_{i,t}(\omega)}\left(x_{\tau_{i-1,t}(\omega)}^\alpha - x_{\tau_{i,t}(\omega)}^\alpha\right) &= R_{\tau_{i,t}(\omega)}\left((1-\alpha)x_{\tau_{i-1,t}(\omega)}^{\pi_1} + \alpha x_{\tau_{i-1,t}(\omega)}^{\pi_2} - (1-\alpha)x_{\tau_{i,t}(\omega)}^{\pi_1} - \alpha x_{\tau_{i,t}(\omega)}^{\pi_2}\right) \\
&= R_{\tau_{i,t}(\omega)}\left((1-\alpha)\left(x_{\tau_{i-1,t}(\omega)}^{\pi_1} - x_{\tau_{i,t}(\omega)}^{\pi_1}\right) + \alpha\left(x_{\tau_{i-1,t}(\omega)}^{\pi_2} - x_{\tau_{i,t}(\omega)}^{\pi_2}\right)\right) \\
&\geq (1-\alpha)R_{\tau_{i,t}(\omega)}\left(x_{\tau_{i-1,t}(\omega)}^{\pi_1} - x_{\tau_{i,t}(\omega)}^{\pi_1}\right) + \alpha R_{\tau_{i,t}(\omega)}\left(x_{\tau_{i-1,t}(\omega)}^{\pi_2} - x_{\tau_{i,t}(\omega)}^{\pi_2}\right).
\end{aligned} \tag{29}$$

Similarly, concavity of R_T and (28) yield

$$R_T(x^\alpha(\omega)) = R_T\left((1-\alpha)x_T^{\pi_1}(\omega) + \alpha x_T^{\pi_2}(\omega)\right) \geq (1-\alpha)R_T\left(x_T^{\pi_1}(\omega)\right) + \alpha R_T\left(x_T^{\pi_2}(\omega)\right). \tag{30}$$

Taking the sum from $i=1$ to $N_{T^-}(\omega)-N_{t^-}(\omega)$ of inequalities (29) and of (30) results in

$$\sum_{i=1}^{N_{T^-}(\omega)-N_{t^-}(\omega)} R\left(x^{\alpha}_{\tau_{i-1,t}(\omega)}-x^{\alpha}_{\tau_{i,t}(\omega)}\right)+R_T(x^{\alpha}_T)\geq$$
$$(1-\alpha)\sum_{i=1}^{N_{T^-}(\omega)-N_{t^-}(\omega)} R\left(x^{\pi_1}_{\tau_{i-1,t}(\omega)}-x^{\pi_1}_{\tau_{i,t}(\omega)}\right)+R_T(x^{\pi_1}_T(\omega))$$
$$+\alpha\sum_{i=1}^{N_{T^-}(\omega)-N_{t^-}(\omega)} R\left(x^{\pi_2}_{\tau_{i-1,t}(\omega)}-x^{\pi_2}_{\tau_{i,t}(\omega)}\right)+R_T(x^{\pi_2}_T(\omega)).$$

Since this inequality holds for any instance ω of the arrival process $\{N_s\}_{s\in\mathbb{R}_+}$, we have

$$\mathbb{E}\left[\sum_{i=1}^{N_{T^-}-N_{t^-}} R\left(x^{\alpha}_{\tau_{i-1,t}}-x^{\alpha}_{\tau_{i,t}}\right)+R_T(x^{\alpha}_T)\right]\geq$$
$$(1-\alpha)\mathbb{E}\left[\sum_{i=1}^{N_{T^-}-N_{t^-}} R\left(x^{\pi_1}_{\tau_{i-1,t}}-x^{\pi_1}_{\tau_{i,t}}\right)+R_T(x^{\pi_1}_T)\right]+\alpha\mathbb{E}\left[\sum_{i=1}^{N_{T^-}-N_{t^-}} R\left(x^{\pi_2}_{\tau_{i-1,t}}-x^{\pi_2}_{\tau_{i,t}}\right)+R_T(x^{\pi_2}_T)\right]$$
$$=(1-\alpha)V^{\pi_1}_t(k_1)+\alpha V^{\pi_2}_t(k_2)$$
$$=(1-\alpha)V_t(k_1)+\alpha V_t(k_2).$$

Since the value function at time t and at state k^α is at least equal to the value of the admissible policy defined in (27), we have

$$V_t(k^\alpha)\geq\mathbb{E}\left[\sum_{i=1}^{N_{T^-}-N_{t^-}} R(x^{\alpha}_{\tau_{i-1,t}}-x^{\alpha}_{\tau_{i,t}})+R_T(x^{\alpha}_T)\right]\geq(1-\alpha)V_t(k_1)+\alpha V_t(k_2),$$

which establishes the concavity of V_t in k. $\square$

Next, we provide the proof of Proposition 3.6.

Proof of Proposition 3.6. Fix the inventory level k. For any given $\epsilon>0$, let $\delta>0$ be such that $c_r\lambda(\delta+2T(1-e^{-\lambda\delta}))<\epsilon$. Consider any times t_1 and t_2 such that $|t_1-t_2|<\delta$. Without loss of generality, assume that $t_1\leq t_2$. Let $\pi_1\in\Pi_{t_1}$ be an optimal policy over $[t_1,T]$ starting from the inventory level $x^{\pi_1}_{t_1}=k$. Therefore, $V_{t_1}(k)=V^{\pi_1}_{t_1}(k)$. In addition, let $\pi_2\in\Pi_{t_2}$ be any admissible policy over $[t_2,T]$ starting from state k at time t_2 as $x^{\pi_2}_{t_2}=k$. Therefore, $V_{t_2}(k)\geq V^{\pi_2}_{t_2}(k)$. Hence,

$$|V_t(k)-V_{\bar{t}}(k)|=V_{t_1}(k)-V_{t_2}(k)=V^{\pi_1}_{t_1}(k)-V_{t_2}(k)\leq V^{\pi_1}_{t_1}(k)-V^{\pi_2}_{t_2}(k),\tag{31}$$

where the first equality comes from Proposition 3.1 which yields $V_{t_1}(k)\geq V_{t_2}(k)$.

Note that, since the reward function is bounded above by c_r, we have

$$\mathbb{E}\left[\sum_{i=1}^{N_{t_2^-}-N_{t_1^-}} R_{\tau_{i,t_1}}\left(x^{\pi_1}_{\tau_{i-1,t_1}}-x^{\pi_1}_{\tau_{i,t_1}}\right)\;\middle|\; x^{\pi_1}_{t_1}=k\right]\leq c_r\lambda(t_2-t_1)<c_r\lambda\delta,$$

and consequently,

$$V_{t_1}^{\pi_1}(k) < c_r\lambda\delta + \mathbb{E}\left[\sum_{i=N_{t_2^-}-N_{t_1^-}+1}^{N_{T^-}-N_{t_1^-}} R_{\tau_{i,t_1}}\left(x_{\tau_{i-1,t_1}}^{\pi_1} - x_{\tau_{i,t_1}}^{\pi_1}\right) \;\middle|\; x_{t_1}^{\pi_1} = k\right].$$

Note that for any $i \geq N_{t_2^-} - N_{t_1^-} + 1$, $\tau_{i,t_1} = \tau_{(i-N_{t_2^-}+N_{t_1^-}),t_2}$. Thus, the index in the above summation can be rewritten to start from 1 to $N_{T^-} - N_{t_2^-}$ to label arrival times τ_{i,t_2}. Therefore, we arrive at

$$V_{t_1}^{\pi_1}(k) - V_{t_2}^{\pi_2}(k) < c_r\lambda\delta + Q, \tag{32}$$

where

$$Q \stackrel{\text{def}}{=} \mathbb{E}\left[\sum_{i=1}^{N_{T^-}-N_{t_2^-}} R_{\tau_{i,t_2}}\left(x_{\tau_{i-1,t_2}}^{\pi_1} - x_{\tau_{i,t_2}}^{\pi_1}\right) \;\middle|\; x_{t_1}^{\pi_1} = k\right] - \mathbb{E}\left[\sum_{i=1}^{N_{T^-}-N_{t_2^-}} R_{\tau_{i,t_2}}\left(x_{\tau_{i-1,t_2}}^{\pi_2} - x_{\tau_{i,t_2}}^{\pi_2}\right) \;\middle|\; x_{t_2}^{\pi_2} = k\right].$$

Define the events $A \stackrel{\text{def}}{=} \{\tau_{1,t_1} > t_2\}$ and $B \stackrel{\text{def}}{=} \{\tau_{1,t_1} \leq t_2\}$. Hence, $\Pr(A) = e^{-\lambda(t_2-t_1)}$ and $\Pr(B) = 1 - e^{-\lambda(t_2-t_1)}$. When the event A occurs, the policy π_1, starting from inventory level k_1 at time t_1, results in $x_{t_2}^{\pi_1} = k$. Therefore, $\mathbb{E}[Q|A] = 0$. In addition, by invoking the upper bound on the reward function, each expectation in Q is bounded above by $c_r\lambda(T - t_2)$, which is no greater than $c_r\lambda T$. Thus, $\mathbb{E}[Q|B] \leq 2c_r\lambda T$. It then follows that

$$Q = \mathbb{E}[Q|B]\Pr(B) \leq 2c_r\lambda T\left(1 - e^{-\lambda(t_2-t_1)}\right) < 2c_r\lambda T\left(1 - e^{-\lambda\delta}\right).$$

Replacing the recent inequality in (32) and (31), we arrive that

$$|V_t(k) - V_{\bar{t}}(k)| < c_r\lambda\delta + 2c_r\lambda T\left(1 - e^{-\lambda\delta}\right) < \epsilon.$$

which completes the proof of continuity of the function V_t in t. $\square$

References

Almudevar, A. (2001). A dynamic programming algorithm for the optimal control of piecewise deterministic Markov processes. *SIAM Journal on Control and Optimization* 40(2), 525–539.

Bauerle, N. and U. Rieder (2009). MDP algorithms for portfolio optimization problems in pure jump markets. *Finance and Stochastics* 13(4), 591–611.

Bauerle, N. and U. Rieder (2010). Optimal control of piecewise deterministic Markov processes with finite time horizon. In A. B. Piunovskiy, ed., *Modern trends of controlled stochastic processes: Theory and Applications*, pp. 144–160. Luniver Press.

Bayraktar, E. and M. Ludkovski (2011). Optimal trade execution in illiquid markets. *Mathematical Finance* 21(4), 681–701.

Bhatnagar, D., A. Currier, J. Hernandez, O. Ma, and B. Kirby (2013). Market and policy barriers to energy storage deployment, a study for the energy storage systems program. *Sandia National Laboratories Report* SAND2013-7606, 1–58.

Byrne, C. and G. Verbic (2013). Feasibility of residential battery storage for energy arbitrage. *2013 Australasian Universities Power Engineering Conference (AUPEC)* pp. 1–7.

Carmona, R. and M. Coulon (2013). A survey of commodity markets and structural models for electricity prices. In F. E. Benth, V. A. Kholodnyi, and Laurence, eds., *Quantitative energy finance: Modeling, pricing, and hedging in energy and commodity markets*, pp. 41–83. Springer, New York.

Carmona, R. and M. Ludkovski (2010). Valuation of energy storage: An optimal switching approach. *Quantitative Finance* 10(4), 359–374.

Cartea, A. and M. Figueroa (2005). Pricing in electricity markets: A mean reverting jump diffusion model with seasonality. *Applied Mathematical Finance* 12(4), 313–335.

Cinlar, E. (2011). *Probability and Stochastics*. Springer.

Culot, M., V. Goffin, S. Lawford, S. de Menten, and Y. Smeers (2006). An affine jump diffusion model for electricity. *working paper, Université Catholique de Louvain* pp. 1–51.

Daley, D. J. and D. Vere-Jones (2003). *An Introduction to the Theory of Point Processes, Volume I: Elementary Theory and Methods*. Springer, second edition.

Davis, M. H. A. (1984). Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models. *Journal of the Royal Statistical Society. Series B (Methodological)* 46(3), 353–388.

Davis, M. H. A. (1993). *Markov Models and Optimization, Monographs on Statistics and Applied Probability.* Chapman & Hall, London.

Deng, S. (1999). Stochastic models of energy commodity prices and their applications: Mean-reversion with jumps and spikes. *Technical report, University of California Energy Institute* pp. 1–35.

Denholm, P., E. Ela, B. Kirby, and M. Milligan (2010). The role of energy storage with renewable electricity generation. *Technical Report National Renewable Energy Laboratory* pp. 1–61.

Diaz-Gonzalez, F., A. Sumper, O. Gomis-Bellmunt, and R. Villafafila-Robles (2012). A review of energy storage technologies for wind power applications. *Renewable and Sustainable Energy Reviews* 16(4), 2154–2171.

Du, P. and N. Lu (2014). *Energy Storage for Smart Grids: Planning and Operation for Renewable and Variable Energy Resources.* Academic Press, first edition.

Fiacco, A. V. (1974). Convergence properties of local solutions of sequences of mathematical programming problems in general spaces. *Journal of Optimization Theory and Applications* 13(1), 1–12.

Fujimoto, K., H. Nagai, and W. J. Runggaldier (2013). Expected power-utility maximization under incomplete information and with Cox-process observations. *Applied Mathematics and Optimization* 67(1), 33–72.

Gassiat, P., H. Pham, and M. Sirbu (2011). Optimal investment on finite horizon with random discrete order flow in illiquid markets. *International Journal of Theoretical and Applied Finance* 14(1), 17–40.

Geman, H. and A. Roncoroni (2006). Understanding the fine structure of electricity prices. *Journal of Business* 79(3), 1225–1261.

Guo, X. and O. Hernandez-Lerma (2009). *Continuous-time Markov Decision Processes.* Springer, New York.

Halman, N., G. Nannicini, and J. Orlin (2015). On the complexity of energy storage problems. *submitted* pp. 1–22.

Harsha, P. and M. Dahleh (2015). Optimal management and sizing of energy storage under dynamic pricing for the efficient integration of renewable energy. *IEEE Transactions on Power Systems* 30(3), 1164–1181.

Heyman, D. P. and M. J. Sobel (2003). *Stochastic Models in Operations Research, Vol. II: Stochastic Optimization.* Dover Publications.

Hordijk, A. and F. D. Schouten (1985). Markov decision drift processes: Conditions for optimality obtained by discretization. *Mathematics of Operations Research* 10(1), 160–173.

Jacobsen, M. (2006). *Point Process Theory and Applications, Marked Point and Piecewise Deterministic Processes.* Birkhauser Boston.

Judd, K. L. (1998). *Numerical Methods in Economics.* The MIT Press.

Kirch, M. and W. J. Runggaldier (2005). Efficient hedging when asset prices follow a geometric poisson process with unknown intensities. *SIAM Journal on Control and Optimization* 43(4), 1174–1195.

Kitaev, M. Y. and V. V. Rykov (1995). *Controlled Queueing Systems.* CRC Press.

Kluge, T. (2006). *Pricing Swing Options and Other Electricity Derivatives.* PhD thesis, University of Oxford.

Kushner, H. and P. G. Dupuis (2001). *Numerical Methods for Stochastic Control Problems in Continuous Time*. Springer, New York.

Lai, G., F. Margot, and N. Secomandi (2010). An approximate dynamic programming approach to benchmark practice-based heuristics for natural gas storage valuation. *Operations Research* 58(3), 564–582.

Lohndorf, N., D. Wozabal, and S. Minner (2013). Optimizing trading decisions for hydro storage systems using approximate dual dynamic programming. *Operations Research* 61(4), 810–823.

Matsumoto, K. (2006). Optimal portfolio of low liquid assets with a log-utility function. *Finance and Stochastics* 10(1), 121–145.

Moazeni, S., W. B. Powell, B. Defourny, and B. Bouzaiene-Ayari (2015a). Parallel non-stationary direct policy search for risk averse stochastic optimization. *available at http://www.optimization-online.org/DB_HTML/2015/09/5100.html* pp. 1–37.

Moazeni, S., W. B. Powell, and A. H. Hajimiragha (2015b). Mean-conditional value-at-risk optimal energy storage operation in the presence of transaction costs. *IEEE Transactions on Power Systems* 30(3), 1222–1232.

Pham, H. and P. Tankov (2008). A model of optimal consumption under liquidity risk with random trading times. *Mathematical Finance* 18(4), 613–627.

Quadrennial Energy Review (2015). Energy transmission, storage, and distribution infrastructure. *Quadrennial Energy Review* pp. 1–348.

Rieder, U. and J. Winter (2009). Optimal control of Markovian jump processes with partial information and applications to a parallel queueing model. *Mathematical Methods of Operations Research* 70(3), 567–596.

Rockafellar, R. T. and R. J. Wets (1998). *Variational Analysis*. Springer, Berlin Heidelberg.

Schmidli, H. (2008). *Stochastic control in insurance*. Springer, London.

Sioshansi, R., S. H. Madaeni, and P. Denholm (2014). A dynamic programming approach to estimate the capacity value of energy storage. *IEEE Transactions on Power Systems* 29(1), 395–403.

Straubel, J. (2015). Energy storage, EV's and the grid. *EIA Conference, Washington D.C.* pp. 1–29.

Thompson, M., H. Rasmussen, and M. Davison (2009). Natural gas storage valuation and optimization: A real options application. *Naval Research Logistics* 56(30), 226–238.

Topkis, D. M. (1998). *Supermodularity and Complementarity*. Princeton University Press, first edition.

Weron, R., M. Bierbrauer, and S. Truck (2004). Modeling electricity prices: jump diffusion and regime switching. *Physica A: Statistical Methods and its Applications* 336(1-2), 39–48.

Xi, X., R. Sioshansi, and V. Marano (2015). A stochastic dynamic programming model for co-optimization of distributed energy storage. *Energy Systems* pp. 1–31.

Xiao, Y., Q. Su, F. Bresler, R. Carroll, J. Schmitt, and M. Olaleye (2014). Performance-based regulation model in PJM wholesale markets. In *2014 IEEE PES General Meeting, Conference Exposition*, pp. 1–5.

Yushkevich, A. (1980). On reducing a jump controllable Markov model to a model with discrete time. *Theory of Probability & Its Applications* 25(1), 58–69.

Zhou, Y., A. A. Scheller-Wolf, N. Secomandi, and S. Smith (2013). Managing wind-based electricity generation in the presence of storage and transmission capacity. *Technical Report, Tepper School of Business, Carnegie Mellon University* pp. 1–40.