Optimal Control of Energy Storage under Random Operation Permissions

Somayeh Moazeni[a], Boris Defourny[b,*]

[a] School of Business, Stevens Institute of Technology, 1 Castle Point Terrace, Hoboken, NJ 07030, USA
[b]Department of Industrial and Systems Engineering, Lehigh University, 200 W Packer Ave, Bethlehem, PA 18015, USA

ABSTRACT

This paper studies the optimal control of energy storage when operations are permitted only at random times. At the arrival of a permission, the storage operator has the option, but not the obligation, to transact. A nonlinear pricing structure incentivizes small transactions spread out among arrivals, instead of a single unscheduled massive transaction, which could stress the energy delivery system. The problem of optimizing storage operations to maximize the expected cumulated revenue over a finite horizon is modeled as a piecewise deterministic Markov decision process. Various properties of the value function and the optimal storage operation policy are established, first when permission times follow a Poisson process, and then for permissions arriving as a self-exciting point process. The sensitivity of the value function and optimal policy to the permission arrival process parameters is studied as well. A numerical scheme to compute the optimal policy is developed and employed to illustrate the theoretical results.

Current distribution systems cannot support simultaneous and identical actions of a large number of agents reacting all to an identical signal. That motivates transactive market frameworks when their access to transactions is restricted. Therefore, the optimal policy of an agent under this restriction is important to be studied. Being able to act at random arrival of permissions and under a nonlinear pricing structure are salient characteristics differentiating this study from existing work on energy storage optimization.

Keywords: Transactive energy, Energy storage, Dynamic optimization, Optimal control

---

1. Introduction

This paper is concerned with the optimal control of energy storage, given a finite horizon over which permissions for operation arrive at random times. The permission flow is modeled by an

* Contact: Boris Defourny (E-mail address: defourny@lehigh.edu)

arrival process. The goal is to maximize the cumulative expected profit from operations given the fixed horizon. The revenue from discharge operations is described by a time-varying function which is increasing and concave in the quantity. Such a nonlinear pricing structure has the property of encouraging discharges in smaller amounts.

Two key properties of this energy storage problem are the lack of control over the operation times and the nonlinear pricing scheme for the discharge payoffs. Being permitted to act only at random times is a salient characteristic differentiating this problem from existing studies on optimal energy storage management in which the controller can act at pre-specified times or act continuously. Nonlinear pricing is pervasive in electricity markets, for instance when side payments are used to compensate specific generating units for starting up or for staying idle. However, these schemes often lack transparency, and are not applied uniformly over all participants. In contrast, in the present work, the nonlinear pricing scheme is defined upfront.

This problem is motivated by market frameworks where the access of energy resources to transactions is restricted and managed in real time by a distribution system operator. Recently, there has been growing interest in so-called transactive energy markets aimed at facilitating transactions, including unscheduled transactions. Transactive energy markets can enable bilateral transactions without passing through the wholesale markets (Rahimi and Ipakchi 2012, Olken 2016, Rahimi and Ipakchi 2016, Kristov et al. 2016, Cazalet et al. 2016). Transactive energy markets would allow distributed energy resources without direct access to wholesale markets to participate in energy transactions over the distribution grid. Energy storage's unique capabilities (Denholm et al. 2010, DOE Report 2011, Diaz-Gonzalez et al. 2012, Du and Lu 2014), combined with technological advances that have been driving costs down (Straubel 2015, Quadrennial Energy Review 2015), suggests that energy storage is an asset that can play an important enabling role in the development of transactive energy markets. At the same time, further deployment of energy storage requires developing appropriate market models to address current market and regulatory barriers (Sioshansi et al. 2012, Bhatnagar et al. 2013, Xiao et al. 2014), especially when it comes to encouraging small participants at the level of the distribution network.

However, the power injections from storage resources cannot be completely unsupervised and ad hoc. Because, otherwise, as their deployment and participation become widespread, there will be times when a large number of storage owners discharge in arbitrary quantities simultaneously. This will put stress on the distribution system. Restricting distributed energy storage transactions to only those times that are specified in real time (and not in advance) by a distribution system operator, supported by an appropriate payoff structure, can alleviate this risk while enabling transactions with no pre-commitments for distributed storage. A simple framework capturing these key properties is proposed and outlined in subsection 1.1.

From the perspective of a storage resource operator participating in such a framework, the permissions for operation communicated by the distribution system operator arrive at random. Thus, the operator needs to optimally control energy storage under random operation times. The problem in this paper fits well in this situation. In fact, this analysis is the building block for further analyzing nonbinding commitment market frameworks.

1.1. Contributions

We formulate the energy storage operation problem as a continuous-time stochastic control problem, in which the optimal policy depends on the stochastic operation permission flow. The control problem in this paper 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. The dynamic programming principle for the control problem of interest leads to a system of nonlinear partial differential equations, which can be solved numerically, for instance see Kushner and Dupuis (2001). We demonstrate the effectiveness of the computational approach using several numerical examples. Next, the extension of the framework to a self-exciting permission arrival model is discussed to show how the previous analysis based on a constant arrival rate can be generalized to this more challenging context.

Several properties of the value function and the optimal policy of this problem are established, relating to the sensitivity of the value function to its state variables and to parameters of the problem. This analysis is expected to shed light on ways to influence the optimal behavior of the controller.

Summary: To summarize, our main contributions include developing a novel nonbinding commitment market framework with a number of attractive characteristics, and studying the optimal control of a storage device participating in this framework. The analyses in this paper can guide a potential storage owner to value its participation in this nonbinding commitment market framework. In addition, this study provides insights for the policy makers and regulators to design efficient and attractive storage deployment programs.

1.2. Literature Review

Energy Storage Optimal Control: 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. (2015), Moazeni et al. (2017), Harsha and Dahleh (2015), 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 most existing analyses of energy storage operations in the literature (including those cited above), electricity flows in and out of the storage resources during time periods that are well specified in advance. These models have focused on a discrete time model with fixed time epochs or a continuous time operation environment over a finite or infinite horizon. The present work differs by introducing uncertainty in the permitted discharge times. In these models, letting an energy storage device operate at specified times imposes a pre-commitment for the grid or the entity, interacting with them, to buying electricity at those specific times, for example every hour in discrete time models or anytime in continuous time models. To the best of our knowledge, the analysis of an energy storage discharge environment restricted to exogenous random operation permission times is novel.

Piecewise Deterministic Markov Processes: For studies on similar 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).

Price Spike Modeling by Poisson Processes: 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).

1.3. Outline

This paper is organized as follows. The mathematical formulation of the optimal energy storage discharge control problem is described in Section 3. The structure of the value function is analyzed in Section 4. In Section 5, the procedure to compute an optimal control is discussed. Section 6 summarizes structural properties of the optimal policy. Illustrative examples and computational analyses are presented in Section 7. The framework with uncertain arrival rates is explained in Section 8. Storage models with inefficiencies are briefly addressed in Section 9. Insights and other possible extensions conclude the paper in Section 10.

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. Nonbinding Commitment Market Framework

In this section, we describe the salient characteristics of our setting. The key property of this framework is to restrict the times at which a particular participating storage unit can discharge.

Agents

Transactions are defined between a utility company or a load serving entity, and a flexible energy resource owner (e.g., a battery in an electrical vehicle) who is unable to participate in the wholesale market, perhaps due to its commitment requirements.

Agreement

Participants enroll in a program managed by the utility company to inform the utility company of their willingness to receive transaction permission notice from the utility company.

The agreement specifies:

- A fixed time horizon over which operation permissions will be sent at random.
- The time-varying payoff structure as a function of the quantity discharged.

The utility company will issue operation permissions to a subset of enrolled available energy storage units, at its own discretion. The utility company has an internal policy for triggering discharge requests, and distributing them among program participants. For instance, load conditions, distribution network congestions, variability in supply will affect these decisions. However, the utility company ignores the response rate for the requests it sends. Actually, the response rate follows from the optimal policy of the storage controller. Therefore, as a first logical step to study the response rate and develop the permission sending policy, this paper analyzes the optimal behavior of the storage operator.

As an energy storage unit receives a permission, it has the option, but not the obligation, to discharge in real time and receive a payment, following the payoff structure agreed upon.

Benefit for the Storage Operator

In contrast to participation in the wholesale market, the storage owner does not have to commit in advance to providing energy, and does not need to get involved in a bidding process. This provides the energy storage unit opportunities to participate and benefit from energy transactions without the financial risk of a binding commitment for energy injection. Thus, it is a nonbinding commitment for the storage operator.

Benefit for the Utility

The utility company gets access to installed storage capacity without having to invest itself in those assets. Although the access to individual units is intermittent, overall in aggregate capacity is obtained at a certain confidence level. Once again this relates back to the optimal behavior of the program enrollees, which is the main focus of this paper.

The utility company does not commit in advance to buy from the enrolled energy storage units. Thus, it is a nonbinding commitment for the utility company.

Through operation permission times, the utility company can indirectly supervise these participants and avoid their ad hoc inferences in the distribution grid.

Broader Benefits

This framework offers a mutually beneficial agreement: it involves nonbinding commitments with attractive flexibility and financial benefits for both parties.

It promotes further deployment of distributed storage capacities at the level of the distribution grid. The presence of energy storage units across the distribution grid can help smoothing out variability, thereby firming transactions from other agents such as buyers and sellers of wind and solar energy.

Specific details and further valuation of this market needs to be studied. However, any further analyses about this framework requires understanding the optimal behavior of storage units participating in the program. Therefore, as the first step, in this paper, we focus on this building block of the framework, namely the optimal control of an energy storage participating in this market.

3. 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 $[0, T]$, where T denotes the fixed terminal time horizon.

3.1. Permission Process

We postulate that the discharge permissions are issued randomly by the Poisson process $\{N_s\}_{s\geq 0}$ with arrival rate λ. We assume that λ is fixed (this assumption is relaxed in Section 8). We denote by $\{\mathcal{F}_t\}_{t\geq 0}$ the natural filtration associated with the Poisson process, where $\mathcal{F}_t$ is the sigma-algebra generated by $\{N_s\}_{s\leq t}$. We refer to the time-stamp of the ith 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. For convenience we set $\tau_{0,t} = t$.

The total number of permissions received between the current time and the fixed terminal time is random. Because there is no guarantee of receiving future opportunities to act, the storage operator faces the problem of deciding between using current versus uncertain future discharge opportunities.

3.2. Nonlinear Pricing Scheme

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 $R_T(a)$.

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 stored quantity 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 stored energy into smaller amounts and not to discharge the entire stored amount at once. This is an attractive property of the program regarding the usage of the distribution network. However, by discharging the stored energy in small amounts, the storage operator bears the risk of receiving no more discharge permissions by the terminal time, in which case a leftover charged level remains at time T that is valued according to the terminal reward R_T.

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 our numerical work reported in §7, 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.

The selected log-utility function is motivated by the following property.

Proposition 3.1 *The cumulated payment for a total amount of k, divided equally into n transactions, cannot be greater than the equivalent linear payment pk.*

Proof. The function $f(n) = n\log(1+pk/n)$ is increasing in n. The result then follows from $\lim_{n\to\infty} nR(k/n) = \lim_{n\to\infty} n\log(1+pk/n) = pk$.

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 discharged from these resources.

3.3. Optimal Control Problem

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

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

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 charge 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 $R_T(x_T^\pi)$ captures the terminal reward. 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 charge level x^π, one may also describe the control strategy A^π for the discharge amount over $(0,T)$. The corresponding controlled charge process x_t^π at time t when the discharge strategy A^π is being employed satisfies

$$\begin{aligned}
&x_0^\pi = k,\\
&dx_t^\pi = -A_t^\pi\,(x_{t^-}^\pi)\,dN_t, \quad \forall t\in(0,T),\\
&x_T^\pi = x_{T^-}^\pi,
\end{aligned} \tag{2}$$

where $\{x_{t^-}^\pi\}_{t\in[0,T]}$ 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 a charge level k at time t, is written

$$V_t^\pi(k) \stackrel{\text{def}}{=} \mathbb{E}\left[\sum_{i=1}^{N_{T^-}-N_{t^-}} R_{\tau_{i,t}}\left(x_{\tau_{i-1,t}}^\pi - x_{\tau_{i,t}}^\pi\right) + R_T\left(x_T^\pi\right) \;\middle|\; x_t^\pi = k\right]. \tag{3}$$

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

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

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 $k\in[0,K]$.

Let $\mathcal{A}_k\subseteq\mathbb{R}_+$ be the set of all discharge amounts that the storage unit can discharge, when the charge level is 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.

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 ($t = \tau_{i,0}$ for some i) and the charge level is k. It follows from (5) that the optimal discharge amount is given by

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

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 properties of the value functions which we later use to characterize the optimal discharges.

4. 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 proposition and proved in Appendix A.

Proposition 4.1 *It holds that*

(a) For any charge level k, $V_t(k)$ is decreasing in t.

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

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

Proposition 4.2 *For any time $t \in [0,T]$ and charge 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 charge level addressed in Proposition 4.1 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 4.3 *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. Let Π_t^+ be the extension of Π_t to the set of policies defined over $[t,T]$ that are $\mathcal{F}_T$ measurable. This means that the controlled process can now peek into the future until time T. Then we have

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

Since the reward function is bounded by c_r, we have $R_{\tau_{i,t}}\left(x_{\tau_{i-1,t}}^{\pi} - x_{\tau_{i,t}}^{\pi}\right) \leq c_r$ 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. $\square$

Next, we establish the concavity of the value function given that the reward functions R_t are concave in the charge level. The proof of Proposition 4.4 is given in Appendix A.

Proposition 4.4 *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 charge 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 4.5 in Appendix A.

Proposition 4.5 *For any charge 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.

5. Computational Scheme

For any reward function $R_t(\cdot)$, the value function $V_t(k)$ can be computed using Euler's method (e.g., see Judd (1998), Kushner and Dupuis (2001)) with the difference equation

$$V_{t+\delta}(k) = V_t(k) + \delta \, \frac{\partial V_t(k)}{\partial t}. \tag{7}$$

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

Proposition 5.1 *The derivative of the value function with respect to time equals*

$$\frac{\partial V_t(k)}{\partial t} = \lambda \left(V_t(k) - \max_{a \in \mathcal{A}_k} \left(R_t(a) + V_t(k-a) \right) \right), \tag{8}$$

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

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

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

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

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

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

Using this equality, we have

$$\begin{aligned} \tfrac{\partial V_t(k)}{\partial t} &= \lim_{\delta\to 0}\frac{V_t(k) - V_{t-\delta}(k)}{\delta} \\ &= \lim_{\delta\to 0}\frac{(1-e^{-\lambda\delta})V_t(k) - \mathbb{E}\left[X_B|B\right]\lambda\delta e^{-\lambda\delta} - \mathbb{E}[X_C|C]o(\delta)}{\delta} \\ &= \lambda V_t(k) - \lambda\lim_{\delta\to 0}\mathbb{E}\left[X_B|B\right]. \end{aligned} \tag{9}$$

Using Proposition 4.3, for every instance $\tau_{1,t-\delta}(\omega) \in [t-\delta, t]$ we have

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

The bounded convergence theorem for expectations (e.g., see Çınlar (2011)) implies that the limit and the expectation in equation (9) can be interchanged. Therefore

$$\lim_{\delta\to 0}\mathbb{E}\left[X_B \mid B\right] = \lim_{\delta\to 0}\mathbb{E}\left[R_\nu(a_\nu + V_t(k - a_\nu(k))\right] = \mathbb{E}\left[\lim_{\delta\to 0}\left\{R_\nu(a_\nu + V_t(k - a_\nu(k))\right\}\right] \tag{10}$$

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

For any instance $\nu(\omega) \in [t-\delta, t]$ where ω refers to an element of the sample space, we have

$$\begin{aligned} &\lim_{\delta\to 0}\left\{R_{\nu(\omega)}(a_{\nu(\omega)}(k)) + V_t(k - a_{\nu(\omega)}(k))\right\} \\ &= \lim_{\delta\to 0}\left\{R_{\nu(\omega)}\left(a_{\nu(\omega)}(k)\right) + V_{\nu(\omega)}\left(k - a_{\nu(\omega)}(k)\right) - V_{\nu(\omega)}\left(k - a_{\nu(\omega)}(k)\right) + V_t\left(k - a_{\nu(\omega)}(k)\right)\right\} \\ &= \lim_{\delta\to 0}\left\{R_{\nu(\omega)}\left(a_{\nu(\omega)}(k)\right) + V_{\nu(\omega)}\left(k - a_{\nu(\omega)}(k)\right)\right\} - \lim_{\delta\to 0}\left\{V_{\nu(\omega)}\left(k - a_{\nu(\omega)}(k)\right) - V_t\left(k - a_{\nu(\omega)}(k)\right)\right\} \\ &= \lim_{\delta\to 0}\left\{\max_{a\in\mathcal{A}_k}\left(R_{\nu(\omega)}(a) + V_{\nu(\omega)}(k-a)\right)\right\} - \lim_{\delta\to 0}\left\{V_{\nu(\omega)}\left(k - a_{\nu(\omega)}(k)\right) - V_t\left(k - a_{\nu(\omega)}(k)\right)\right\}. \end{aligned} \tag{11}$$

According to Proposition 4.5, the value function $V_t(k)$ is uniformly continuous in t on $\mathcal{A}_k \cap [0, K]$. In addition, as $\delta \to 0$, $\nu \to t$ almost surely. Therefore, for any $a \in \mathcal{A}_k \cap [0, K]$ and the instance $\nu(\omega) \in [t-\delta, t]$ we have

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

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

$$\lim_{\delta\to 0}\left\{V_{\nu(\omega)}\left(k - a_{\nu(\omega)}(k)\right) - V_t\left(k - a_{\nu(\omega)}(k)\right)\right\} = 0. \tag{12}$$

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 $\delta \to 0$ by Theorem 7.15 of Rockafellar and Wets (1998). Therefore, Theorem 7.33 of Rockafellar and Wets (1998) implies that

$$\lim_{\delta\to 0}\left\{\max_{a\in\mathcal{A}_k}\left(R_{\nu(\omega)}(a) + V_{\nu(\omega)}(k-a)\right)\right\} = \max_{a\in\mathcal{A}_k}\left(R_t(a) + V_t(k-a)\right). \tag{13}$$

We note that equality (13) 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 (12) and (13) in equation (11) and subsequently in equation (10) follows that

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

This equation along with (9) completes the proof. $\square$

Using Proposition 5.1 and Euler's method, for small $\delta>0$ we obtain the difference equation

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

that is,

$$V_{t-\delta}(k)=(1-\lambda\delta)V_t(k)+\lambda\delta\max_{a\in\mathcal{A}_k}\left(R_t(a)+V_t(k-a)\right). \tag{14}$$

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, (14) involves solving convex optimization problems.

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

Corollary 5.1 *Suppose that the feasible action sets $\mathcal{A}_k$ as functions of the charge 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 charge 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 (8) imply that for $\lambda>0$,

$$\begin{aligned} \tfrac{\partial V_t(k_2)}{\partial t} &= \lambda\left(V_t(k_2)-\max_{a\in\mathcal{A}_{k_2}}\left(R_t(a)+V_t(k_2-a)\right)\right) \\ &\le \lambda\left(V_t(k_2)-\left\{R_t\left(a_t(k_1)\right)+V_t\left(k_2-a_t(k_1)\right)\right\}\right). \end{aligned} \tag{15}$$

In addition, the concavity of the value function V_t in the charge level established in Proposition 4.4 implies that it has decreasing differences. Hence, $V_t(k_2)-V_t(k_1)\le V_t(k_2-a_t(k_1))-V_t(k_1-a_t(k_1))$, and consequently

$$V_t(k_2)-V_t\left(k_2-a_t(k_1)\right)\le V_t(k_1)-V_t\left(k_1-a_t(k_1)\right). \tag{16}$$

It then follows from inequality (16) in (15) that

$$\begin{aligned}\tfrac{\partial V_t(k_2)}{\partial t} &\leq \lambda\left(V_t(k_2) - R_t(a_t(k_1)) - V_t(k_2 - a_t(k_1))\right) \\ &\leq \lambda\left(V_t(k_1) - R_t(a_t(k_1)) - V_t(k_1 - a_t(k_1))\right) \\ &= \lambda\left(V_t(k_1) - \max_{a\in\mathcal{A}_{k_1}}\left(R_t(a) + V_t(k_1 - a)\right)\right) = \frac{\partial V_t(k_1)}{\partial t},\end{aligned} \tag{17}$$

which shows the result. $\square$

6. Structure of the Optimal Policy

This section is devoted to addressing some properties of $a_t(k)$, defined in (6), and of the charge level process x^π. Three fundamental properties are established: (i) the optimal discharged amount is nondecreasing in the charge level, (ii) the optimal charge trajectory over $[0,T]$ is nondecreasing in the initial charge 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 charge level, and therefore complements property (i).

We start by establishing a monotonicity result for the optimal discharge amount. In the sequel, the set $\mathcal{A}_k \subseteq \mathbb{R}$ is called ascending in k, if for any $k_1 \leq k_2$ 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 6.1 *Let the set* $\mathcal{C} \stackrel{\text{def}}{=} \{(k,a) \in \mathbb{R}^2 : a \in \mathcal{A}_k, k \in [0,K]\}$ *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 charge level* k*, i.e.,* $k_1 \leq k_2$ *implies* $a_t(k_1) \leq a_t(k_2)$*.*

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 charge 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. $\square$

For instance, the set $\mathcal{A}_k = [0,k]$ is ascending in k on $[0,K]$ and the set $\mathcal{C} \stackrel{\text{def}}{=} \{(k,a) \in \mathbb{R}^2 : a \in \mathcal{A}_k, k \in [0,K]\}$ is a sublattice of $\mathbb{R}^2$. In addition, $\mathcal{A}_k = [0,k]$ has the property assumed in Corollary 5.1 that is $k_1 \leq k_2$ 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 charge level results in a stored quantity process with higher levels through the entire time horizon.

Proposition 6.2 *Let $k_1 \leq k_2$. Denote the charge level 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 $t \in [0,T]$, $x_t^1 \leq x_t^2$.*

Proof. Let $\bar{t} := \max\{t : x_s^1 \leq x_s^2 \text{ for } 0 \leq s \leq t\}$. 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 $x_{\bar{t}-}^2 \geq x_{\bar{t}-}^1$ 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

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

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

$$\begin{aligned} R_t\left(a^1\right) + V_t\left(x_{t-}^1 - a^1\right) &\geq R_t\left(x_{t-}^1 - x_t^2\right) + V_t\left(x_{t-}^1 - \left(x_{t-}^1 - x_t^2\right)\right) \\ &= R_t\left(x_{t-}^1 - x_t^2\right) + V_t\left(x_t^2\right). \end{aligned} \tag{18}$$

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

$$R_t\left(a^2\right) + V_t\left(x_{t-}^2 - a^2\right) \geq R_t(a) + V_t\left(x_{t-}^2 - a\right) \quad \forall a \in [0, x_{t-}^2]. \tag{19}$$

Since a^2 is the smallest maximizer, inequality (19) 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,

$$\begin{aligned} R_t\left(a^2\right) + V_t\left(x_{t-}^2 - a^2\right) &> R_t\left(x_{t-}^2 - x_t^1\right) + V_t\left(x_{t-}^2 - \left(x_{t-}^2 - x_t^1\right)\right) \\ &= R_t\left(x_{t-}^2 - x_t^1\right) + V_t\left(x_t^1\right). \end{aligned} \tag{20}$$

By combining inequalities (18) and (20), we arrive at

$$R_t\left(x_{t-}^1 - x_t^2\right) - R_t\left(a^1\right) \leq V_t\left(x_{t-}^1 - a^1\right) - V_t(x_t^2) \tag{21}$$

$$= V_t\left(x_t^1\right) - V_t\left(x_{t-}^2 - a^2\right) \tag{22}$$

$$< R_t\left(a^2\right) - R_t\left(x_{t-}^2 - x_t^1\right). \tag{23}$$

Here inequalities (21) and (23) are rearrangements of inequalities (18) and (20), and the equality (22) 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,

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

which is in contradiction with inequality (23). Thus the supposition that $\bar{t} < T$ cannot be true, i.e., for all $t \in [0,T]$, we must have $x_t^1 \leq x_t^2$. $\square$

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

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, $R_{t_1}(b) + V_{t_1}(k-b) < R_{t_1}(a_{t_1}(k)) + V_{t_1}(k - a_{t_1}(k))$, and consequently

$$V_{t_1}(k-b) - V_{t_1}(k - a_{t_1}(k)) < R_{t_1}(a_{t_1}(k)) - R_{t_1}(b). \tag{24}$$

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

$$R_{t_1}(a_{t_1}(k)) - R_{t_1}(b) \leq R_{t_2}(a_{t_1}(k)) - R_{t_2}(b).$$

Combining the recent inequality in inequality (24) results in

$$V_{t_1}(k-b) - V_{t_1}(k - a_{t_1}(k)) < R_{t_2}(a_{t_1}(k)) - R_{t_2}(b). \tag{25}$$

According to Corollary 5.1, $\frac{\partial V_{t_2}(k)}{\partial t}$ 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

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

Using this inequality and inequality (25) we get

$$V_{t_2}(k-b) - V_{t_2}(k - a_{t_1}(k)) \leq V_{t_1}(k-b) - V_{t_1}(k - a_{t_1}(k)) < R_{t_2}(a_{t_1}(k)) - R_{t_2}(b).$$

A rearrangement of the recent inequality equals

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

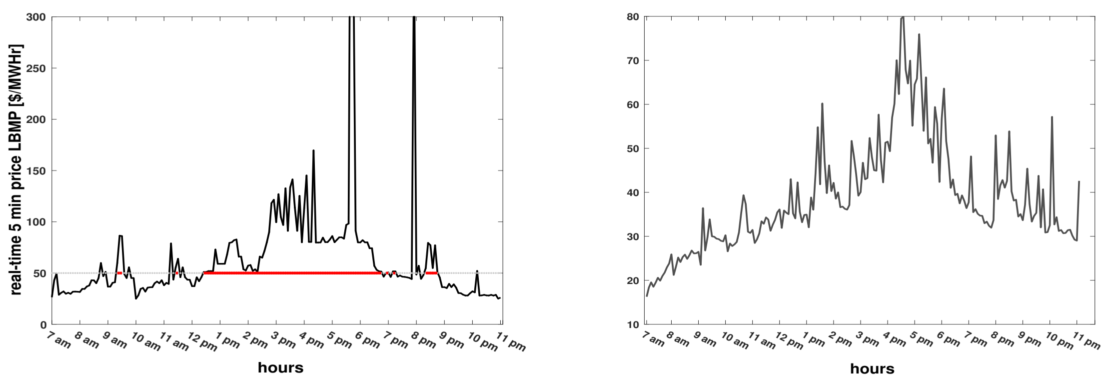
which indicates that $a_{t_1}(k)$ achieves a superior value for $R_{t_2}(a) + V_{t_2}(k-a)$ than b. Hence, b cannot be in the solution set $\operatorname{argmax}_{a\in\mathcal{A}_k} \{R_{t_2}(a) + V_{t_2}(k-a)\}$. Since this is true for any $b < a_{t_1}(k)$, we can conclude that $a_{t_2}(k) \geq a_{t_1}(k)$.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

Next we present our computational investigation of the optimal value function and optimal decisions of a storage unit.

7. Numerical Examples

We consider a storage device of capacity $K = 4$. The finite time horizon over which a participating storage unit may be issued a permission is assumed to be [7am-11pm].

We assume that the discharge permission events are triggered when the zonal electricity price is above a given price threshold. The real time 5-min prices over [7am-11pm] for August 25, 2015 for New York City (load zone J in NYISO) and the price threshold 50[$/MWHr] are illustrated in Figure 1(a). Given the price threshold, 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).



(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.

7.1. Permission Process

For a given price threshold, we estimate 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 equals to 50[$/MWHr], the average arrival rate over peak hours is 24.9355 per day. For threshold prices 100[$/MWHr] and 25[$/MWHr], the permission arrival rates become 4.7742 per day and 144.7742 per day, respectively.

7.2. Nonlinear Pricing Scheme

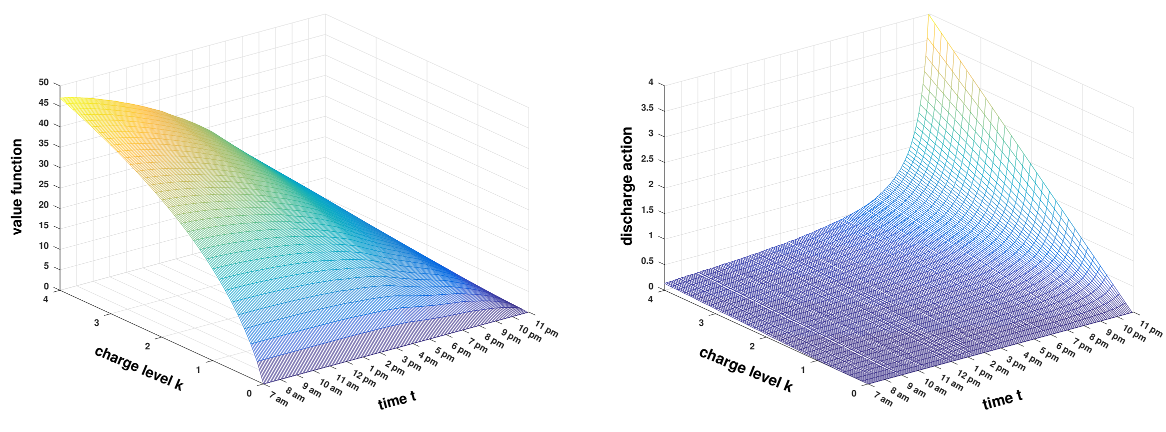
We consider the log reward function $R_t(a) = \log(1 + p_t a)$ as the reward function for $t \in [0, T)$. The average real time hourly electricity price over one month is used as a proxy of the reward coefficient p_t at every time $t \in$[7am-11pm]. The average real time hourly price curve for August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

The value of the stored power at the end of the time horizon is assumed to be zero, i.e., $R_T(a)=0$, for all a.

7.3. Results

Figure 2 summarizes the results from the computational scheme in §5 with $\delta=5$ 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 §4 on the concavity of the value function in the charge level k and its monotonicity in the charge level k and time t. The right plot also illustrates the structures analyzed in §6 that the discharge amount $a_t(k)$ is increasing in the charge level k.



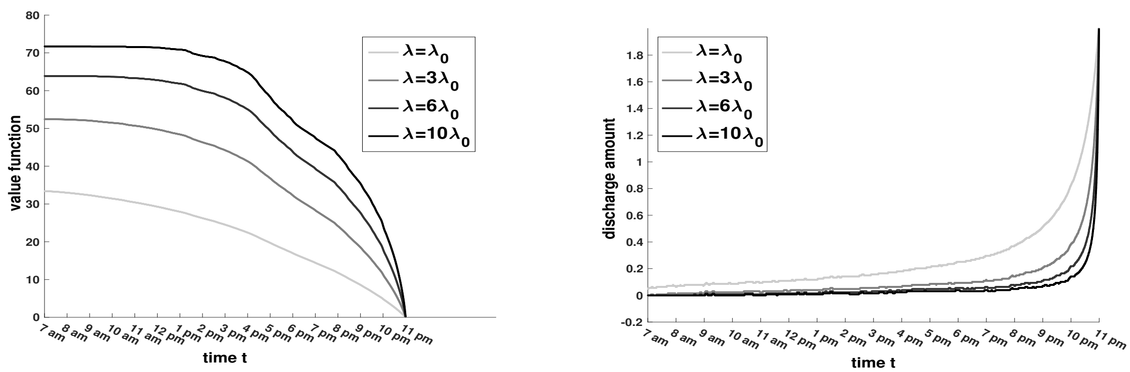
(a) optimal value function (b) optimal actions $a_t(k)$

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

7.4. Sensitivity to Permission Arrival Rate

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 charge 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 4.2.

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 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 $k=2$. Here, $\lambda_0 = 1.5585$.

8. Extension to Uncertain Arrival Rates

This section extends our analysis to the case where the Poisson arrival process is replaced by a more general point process, motivated by the need for studying the robustness of our model to the Poisson arrival process assumption. Namely, we now assume that the permissions are generated as a Markovian self-exciting point process. The generalization of the arrival model to self-exciting point processes (see e.g., Bremaud (1981), Daley and Vere-Jones (2003)) can be well-suited to the modeling of permissions arriving in clusters (Hawkes and Oakes 1974). 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. This intensity model may address the fact that balancing needs would trigger permissions from the utility company.

Here, we investigate self-exciting shot processes. Following Chapter 6 of Çınlar (2011), let M be a standard Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}_+$. The counting process $\{N_t\}_{0\le t\le T}$ and arrival rate $\{\lambda_t\}_{0\le t\le T}$ are defined by

$$\begin{aligned} \lambda_t &= \lambda_0 e^{-\beta t} + \int_0^t \alpha e^{-\beta(t-s)} dN_s \\ N_t &= \int_{[0,t]\times\mathbb{R}_+} M(ds,dz)\mathbf{1}_{0<z\le\lambda_s}, \end{aligned} \tag{26}$$

where $\alpha \ge 0$ is the jump magnitude, $\beta \ge 0$ is the decay rate, and the indicator function $\mathbf{1}_{0<z\le\lambda_s}$ is equal to 1 if $z \in (0,\lambda_s]$ and to 0 otherwise. The process λ_t in (26) is Markov (see Section 6.27

of Çınlar (2011) for a proof). For algorithms employed in simulation studies of such self-exciting processes, see (Lewis and Shedler 1979, Ogata 1981, Sigman 2013). Figure 4 illustrates two sample realizations of this process for $\lambda_0 = 1$, $\beta = 0.8$, and $\alpha = 1$. The continuous lines represent the rate λ_t, and the dots represent the arrival times. Arrivals trigger a positive jump in the rate, while the absence of arrivals results in the exponential decay of the rate. These characteristics of the stochastic rate favor the emergence of arrivals in clusters, recognizable in Figure 4.

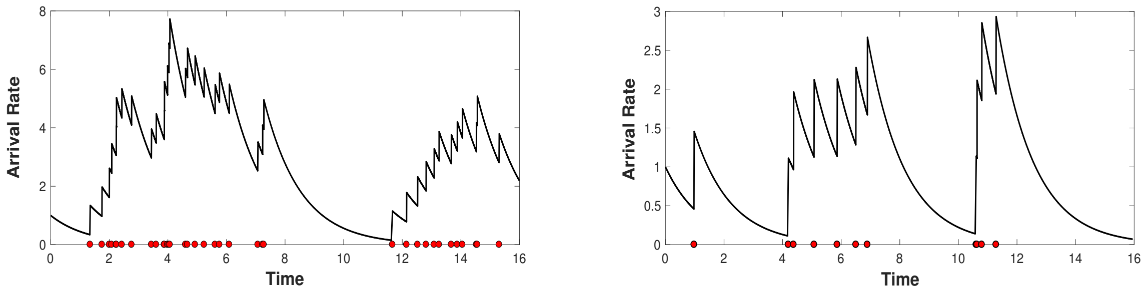


Figure 4 **Two realizations of the self-exciting shot process (curves: arrival rate; dots: arrival times).**

Given the intensity model (26), the state space at time t in the control problem of maximizing the total expected revenue is augmented to include λ_t. Equation (5) is thus extended to

$$V_t(k,\lambda_t) = \mathbb{E}\left[\max_{a\in\mathcal{A}_k}\left\{R_{\tau_{1,t}}(a) + V_{\tau_{1,t}}(k-a,\lambda_{\tau_{1,t}})\right\}\cdot \mathbf{1}_{\tau_{1,t}<T} + R_T(k)\cdot\mathbf{1}_{\tau_{1,t}\geq T}\right], \tag{27}$$

where the expectation is now also over the future stochastic rates, given the current rate λ_t. The maximization is now over the time-varying Markov discharge policies $\{A_t^\pi(k,\lambda)\}_{0\leq t\leq T}$ where k is the charge level and λ_t is the permission arrival rate at time t. An optimal discharge action at the permission time is obtained by $a_t(k,\lambda_t) = \arg\max_{a\in\mathcal{A}_k}\ \{R_t(a) + V_t(k-a,\lambda_t)\}$.

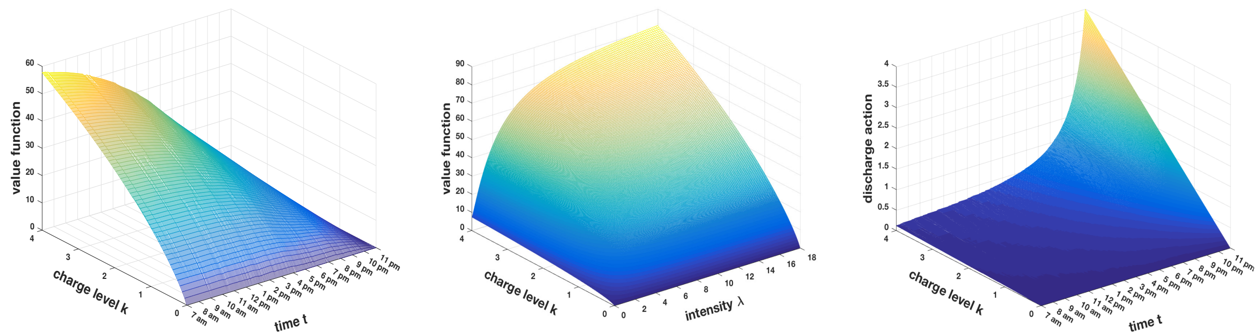
Consider the time interval $(t-\delta, t]$ for a small real $\delta > 0$. The probability of no arrival equals

$$\Pr(\tau_{1,t-\delta} > t|\lambda_{t-\delta}) = \exp\left(-\lambda_{t-\delta}\frac{(1-e^{-\beta\delta})}{\beta}\right). \tag{28}$$

See, e.g., Hawkes and Oakes (1974). This can be directly established, by the decomposition of the interval into n equal subintervals of length $\Delta = \frac{\delta}{n}$, and evaluating

$$\begin{aligned}&\Pr\left(\tau_{1,t-\delta} > t|\lambda_{t-\delta}\right) = \Pr\left(\cap_{k=1}^{n}\left\{\tau_{1,t-\delta} > t-\delta+k\Delta\right\}|\lambda_{t-\delta}\right)\\ &= \prod_{k=1}^{n}\Pr\left(\tau_{1,t-\delta} > (t-\delta)+k\Delta \mid \lambda_{t-\delta+(k-1)\Delta} = \lambda_{t-\delta}e^{-\beta(k-1)\Delta}\right) = \exp\left(-\lambda_{t-\delta}\sum_{k=1}^{n}e^{-\beta(k-1)\Delta}\Delta\right).\end{aligned}$$

As $\Delta\to 0$, this quantity can be approximated by $\exp\left(-\lambda_{t-\delta}\int_0^\delta e^{-\beta s}ds\right) = \exp\left(\frac{-\lambda_{t-\delta}(1-e^{-\beta\delta})}{\beta}\right)$, resulting in (28).



(a) value function at $\lambda_t = 2$ (b) value function at $t = 0$ (7 am) (c) optimal discharge at $\lambda_t = 2$

Figure 5 **Optimal value function and discharge actions under the self-exciting shot process for permission arrivals.**

Similarly, $\Pr(N_{t^-} - N_{(t-\delta)^-} \geq 2) = o(\delta)$ and $\Pr(N_{t^-} - N_{(t-\delta)^-} = 1) = 1 - \Pr(N_{t^-} - N_{(t-\delta)^-} = 0) + o(\delta)$ hold for small δ. Thus, it follows from the dynamic programming principle that

$$V_{t-\delta}(k, \lambda_{t-\delta}) = \left(e^{\frac{-\lambda_{t-\delta}(1-e^{-\beta\delta})}{\beta}}\right) V_t\left(k, \lambda_{t-\delta} e^{-\beta\delta}\right) + \left(1 - e^{\frac{-\lambda_{t-\delta}(1-e^{-\beta\delta})}{\beta}}\right)\left(\max_{a \in \mathcal{A}_k}\left\{R_t(a) + V_t\left(k - a, \lambda_{t-\delta} e^{-\beta\delta} + \alpha\right)\right\}\right). \tag{29}$$

This equation is employed to compute $V_t(k, \lambda_t)$ and corresponding actions $a_t(k, \lambda_t)$ for charge level k and intensity level λ_t. Figure 5 illustrates these values. Similar to §7, $K = 4$ and the log reward function $R_t(a) = \log(1 + p_t a)$ for p_t as in Figure 1(b) are considered.

Proposition 4.1 remains valid for the intensity model (26); its proof in Appendix A is directly applicable. Given c_r as in Proposition 4.3, $c_r := \max_{t \in [0,T]} R_t(K)$, and following the proof of this proposition, we arrive at $V_t(k) \leq c_r\left(\mathbb{E}[N_{T^-} - N_{t^-} \mid \lambda_t = \lambda] + 1\right)$. Equations (6.15) and (6.35) in Çınlar (2011) yield $\mathbb{E}[N_t] = \frac{\lambda_0}{(\alpha-\beta)}(e^{(\alpha-\beta)t} - 1)$. Therefore, the following upper bound on the value function associated to the arrival rate model (26) can be established,

$$V_t(k, \lambda_t) \leq \left(1 + \frac{\lambda_t}{(\alpha - \beta)}\left(e^{(\alpha-\beta)(T-t)} - 1\right)\right) c_r. \tag{30}$$

The following proposition extends the result in Proposition 4.2 and state sensitivity of the value function to the shot noise process parameters.

Proposition 8.1 *Suppose (λ_t, N_t) follows the shot noise process in (26). Consider the expected value function $V_t(k, \lambda_t)$ defined by (27). Then (i) $\lambda_1 \leq \lambda_2$ implies $V_t(k, \lambda_1) \leq V_t(k, \lambda_2)$, for all k. Furthermore, let the notation $V_t^{(\ell)}$ indicate that the arrival process follows (26) with parameters α_ℓ, β_ℓ, for $\ell = 1, 2$. Then (ii) $\beta_1 \geq \beta_2$ and $\alpha_1 \leq \alpha_2$ imply that $V_t^{(1)}(k, \lambda) \leq V_t^{(2)}(k, \lambda)$, for all k, λ.*

Proof. Let M_ω denote a fixed realization ω of a standard 2-dimensional random Poisson measure, with corresponding atoms $(t_i(\omega), z_i(\omega))$, $i \geq 0$, indexed such that $t_i < t_{i+1}$. To prove (i) and (ii) simultaneously, consider for $j = 1,2$ the processes $\{(\lambda_s^{(j)}, N_s^{(j)} - N_t^{(j)})\}_{t \leq s \leq T}$ follow (26) with $\beta = \beta_j$, $\alpha = \alpha_j$, started on $\lambda_t^{(j)} = \lambda_j$ almost surely, where $\lambda_1 \leq \lambda_2$, $\beta_1 \geq \beta_2$, and $\alpha_1 \leq \alpha_2$. Given ω, we derive the corresponding realizations $(\lambda_s^{(j)}(\omega), [N_s^{(j)} - N_t^{(j)}](\omega))_{t \leq s \leq T}$ from M_ω and (26). We have $\lambda_t^{(1)}(\omega) = \lambda_1 \leq \lambda_2 = \lambda_t^{(2)}(\omega)$. Suppose $\lambda_s^{(1)}(\omega) \leq \lambda_s^{(2)}(\omega)$ and $[N_s^{(1)} - N_t^{(1)}](\omega) \leq [N_s^{(2)} - N_t^{(2)}](\omega)$ for $s \in S$, which is true for $S = \{t\}$. Then for all times $s' \in (t_i(\omega), t_{i+1}(\omega)) \cap [s, T]$ where $i = \sup\{i \geq 0 : t_i(\omega) \leq s\}$, we have $\lambda_{s'}^{(1)}(\omega) \leq \lambda_{s'}^{(2)}(\omega)$, from the relations $\lambda_{s'}^{(j)}(\omega) = \lambda_s^{(j)}(\omega) e^{-\beta_j (s'-s)}$ with $\beta_1 \geq \beta_2$ and $\lambda_s^{(1)} \leq \lambda_s^{(2)}$, as well as $N_{s'}^{(1)} - N_s^{(1)} = N_{s'}^{(2)} - N_s^{(2)} = 0$ since there is no arrival. At the tentative jump time t_i, the inequality $\lambda^{(1)}_{t_i^-(\omega)}(\omega) \leq \lambda^{(2)}_{t_i^-(\omega)}(\omega)$ implies $\lambda^{(1)}_{t_i(\omega)}(\omega) \leq \lambda^{(2)}_{t_i(\omega)}(\omega)$, since $\lambda^{(1)}_{t_i(\omega)}(\omega) = \lambda^{(1)}_{t_i^-(\omega)}(\omega) + \alpha_1 \mathbf{1}_{(0, \lambda^{(1)}_{t_i^-(\omega)}]}(z^i(\omega)) \leq \lambda^{(2)}_{t_i^-(\omega)}(\omega) + \alpha_2 \mathbf{1}_{(0, \lambda^{(2)}_{t_i^-(\omega)}]}(z^i(\omega)) = \lambda^{(2)}_{t_i(\omega)}(\omega)$, using $\alpha_1 \leq \alpha_2$. We also have $[N^{(1)}_{t_i(\omega)} - N^{(1)}_{t_i^-(\omega)}](\omega) = \mathbf{1}_{(0, \lambda^{(1)}_{t_i^-(\omega)}]}(z^i(\omega)) \leq \mathbf{1}_{(0, \lambda^{(2)}_{t_i^-(\omega)}]}(z^i(\omega)) = [N^{(2)}_{t_i(\omega)} - N^{(2)}_{t_i^-(\omega)}](\omega)$ and thus $[N^{(1)}_{t_i(\omega)} - N_t^{(1)}](\omega) \leq [N^{(2)}_{t_i(\omega)} - N_t^{(2)}](\omega)$. Hence, the set S can be extended to $[t, T]$, that is, $\lambda_s^{(1)}(\omega) \leq \lambda_s^{(2)}(\omega)$ for all $s \in [t, T]$, and $[N_s^{(1)} - N_t^{(1)}](\omega) \leq [N_s^{(2)} - N_t^{(2)}](\omega)$ for all $s \in [t, T]$. From there, it follows, as in the proof of Proposition 4.1, that if π_1 denotes a policy over $[t, T]$ able to attain $V_t^{(1)}(k, \lambda_1)$, then we can replicate its expected value under the process $\lambda_t^{(2)}$ started on λ_2 at time t by constructing a virtual rate process λ'_t started on λ_1, decaying at rate β_1, and jumping with jumps of size α_1, at the times t_i where simultaneously, the original process $\lambda_t^{(2)}$ jumps, and the independently drawn Bernoulli random variable Z_i is equal to 1, set to happen with probability $\lambda'_{t_i}/\lambda^{(2)}_{t_i} \in [0,1]$. This proves $V_t^{(2)}(k, \lambda_2) \geq V_t^{(1)}(k, \lambda_1)$. We obtain (i) by setting $\alpha_1 = \alpha_2 = \alpha$ and $\beta_1 = \beta_2 = \beta$ to get the same shot noise stochastic process. We obtain (ii) by setting $\lambda_1 = \lambda_2 = \lambda$ to get the same start rate at time t. $\square$

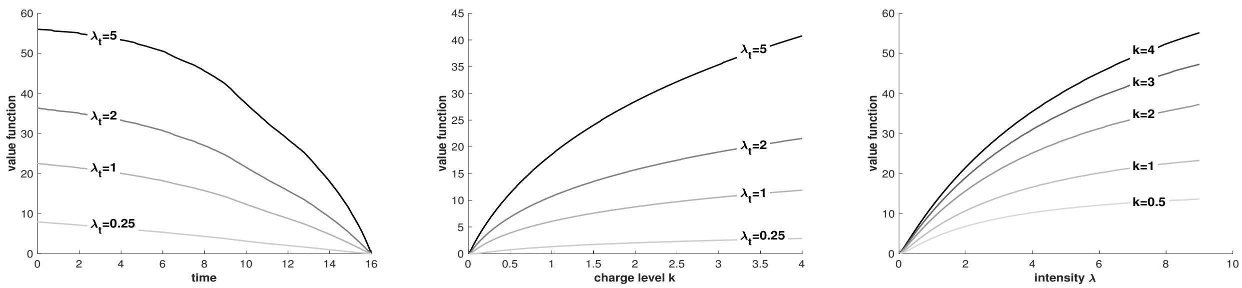
In addition, the value function $V_t(k, \lambda_t)$ is concave in k, for any time $t \in [0, T]$. The proof is similar to that of Proposition 4.4 using augmented states for $V_t^{\pi_\ell}(k_\ell, \lambda_t)$ and $A_s^{\pi_\ell}(x_{s^-}^{\pi_\ell}, \lambda_s)$ for $\ell = 1, 2$.

Figure 6 illustrates the patterns for the value function in theoretical findings when permissions arrive according to the point process described in (26).

The monotonicity of the optimal action $a_t(k, \lambda_t)$ in the charge level k and time t, established in Propositions 6.1 and 6.3, is preserved for the intensity model (26); similar proofs are employed. Figure 7, obtained under the self-exiting point process model, illustrates these characteristics.

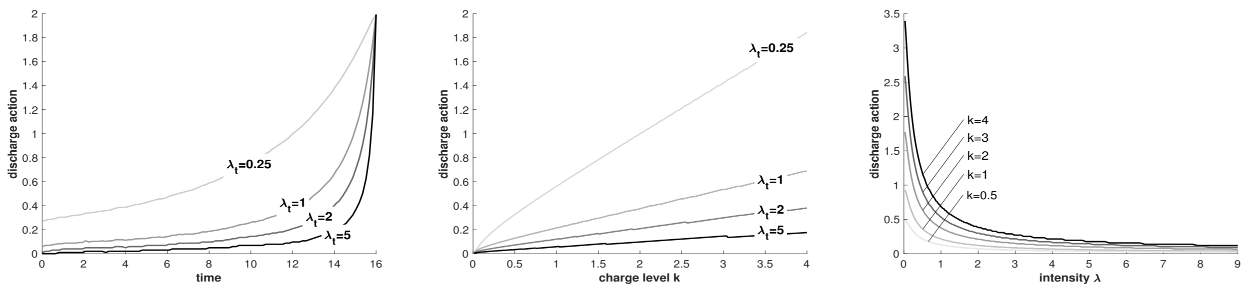
9. Extension to Storage Models with Inefficiencies

Here we clarify how one can take into account the inefficiencies of physical storage devices, thereby extending the perfectly efficient storage model assumed so far. In the presence of a storage self-discharging rate per hour, denoted by γ_{loss}, the value $V_{\tau_{1,t}}(k - a)$ in equation (5) is replaced by



(a) $k = 2$ (b) $t = 12$ (c) $t = 12$

Figure 6 Optimal value function structure under the self-exciting shot process for permission arrivals.



(a) $k = 2$ (b) $t = 12$ (c) $t = 12$

Figure 7 Optimal policy structure under the self-exciting shot process for permission arrivals.

$V_{\tau_{1,t}}([1-\gamma_{\text{loss}}(\tau_{1,t}-t)]^+(k-a))$. In the computational scheme in equation (14), both $V_t(k)$ and $V_t(k-a)$ are replaced by $V_t([1-\gamma_{\text{loss}}\delta]^+k)$ and $V_t([1-\gamma_{\text{loss}}\delta]^+(k-a))$.

When the storage discharging efficiency, denoted by η, is strictly less than one, the reward $R_t(a)$ in (5) as well as in (14) are replaced by $R_t(\eta a)$. Limitations on the discharging rate are captured by bound constraints on a when defining feasible actions. More precisely, denote the minimum and maximum discharging power of the storage unit by $a_{\min}$ and $a_{\max}$. Hence, the expression of the feasible set $\mathcal{A}_k$ should include the constraints $a_{\min} \le a \le a_{\max}$. The structural results will be preserved under these modifications.

10. Conclusion and Discussion

A novel approach to promote distributed energy storage deployment and participation in unscheduled transactions at the level of the distribution grid is proposed. This framework is promising since it enjoys several attractive features: (i) it involves nonbinding commitments, (ii) offers attractive financial benefits and flexibility for both parties, and (iii) enables the utility company to indirectly supervise operations of the energy storage units.

As the very first step to specify and value this framework, we need to study the optimal behavior of energy storage under the assumption that it is permitted to operate only at random times over a finite horizon. Thus this paper limits its scope to characterizing this component of the framework. Two salient properties of the energy storage operation model in this paper include the random operation times and nonlinear pricing scheme.

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 policy can be used by the storage operators to obtain a more precise valuation of the energy storage unit and support their investment decisions. The derivation of the optimal policy as well as the properties of the storage owner's value function enable the utility company involved in the contract to predict the response behavior of the storage controllers and the expected discharge amounts to better design the details of the framework.

Our results imply that this model is a promising framework for further research and applications in the efficient energy storage deployment and transactive energy markets. 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.

Appendix A: Proofs

Proof of Proposition 4.1 (a) For times t_1 and t_2 with $t_1 < t_2$, let k be the charge 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 stored quantity 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 $V_{t_1}(k) \geq V_{t_1}^{\pi_1}(k) = V_{t_2}^{\pi_2}(k) = V_{t_2}(k)$. This completes the proof of (a).

(b) Fix charge 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

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

where the policy π_2 starting from the state k_2 is an admissible policy over $[t, T]$, with the store quantity 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 in (A.1) comes from the assumption that the reward function R_t is increasing in the amount discharged and

$0 < k_1 \le k_2$. Hence, $R_{\tau_{1,t}}(k_1 - x_{\tau_{1,t}}) \le R_{\tau_{1,t}}(k_2 - x_{\tau_{1,t}})$. Therefore, $V_t(k_2) = \max_{\pi \in \Pi_t} V_t^{\pi}(k_2) \ge V_t^{\pi_2}(k_2) \ge V_t(k_1)$, which completes the proof of part (b). $\square$

Proof of Proposition 4.4. Fix some charge levels k_1, k_2 at time t, such that $0 < k_1 \le k_2$. Let π_1 and π_2 be optimal Markov discharge policies, respectively, starting from the charge 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 charge level $k^{\alpha} \overset{\text{def}}{=} (1-\alpha)k_1 + \alpha k_2$. Consider the controlled process x^{α} over $[t,T]$ defined as below

$$x_t^{\alpha} = k^{\alpha}, \quad dx_s^{\alpha} = -\left((1-\alpha)A_s^{\pi_1}\left(x_{s^-}^{\pi_1}\right) + \alpha A_s^{\pi_2}\left(x_{s^-}^{\pi_2}\right)\right)dN_s, \quad \forall s \in (t,T), \quad x_T^{\alpha} = x_{T^-}^{\alpha}. \tag{A.2}$$

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 charge levels k_1 and k_2. It follows from (A.2) that $x_s^{\alpha} = (1-\alpha)x_s^{\pi_1} + \alpha x_s^{\pi_2}$, for all $s \in [t,T]$. From the feasibility of the policies π_1 and π_2, we have $0 \le A_s^{\pi_\ell}\left(x_{s^-}^{\pi_\ell}\right) \le x_s^{\pi_\ell}$, for $\ell = 1,2$. Hence, $0 \le (1-\alpha)A_s^{\pi_1}\left(x_{s^-}^{\pi_1}\right) + \alpha A_s^{\pi_2}\left(x_{s^-}^{\pi_2}\right) \le (1-\alpha)x_s^{\pi_1} + \alpha x_s^{\pi_2} = x_s^{\alpha}$, which implies that x^{α} in (A.2) is an admissible charge process starting at k^{α}.

For any realization ω of the Poisson process $\{N_s\}_{s\in\mathbb{R}_+}$, equality $x_s^{\alpha} = (1-\alpha)x_s^{\pi_1} + \alpha x_s^{\pi_2}$ implies that the difference of charge 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 charge levels in the processes x^{π_1} and x^{π_2}. This along with concavity of the reward function $R_{\tau_{i,t}(\omega)}$ implies that

$$\begin{aligned}
R_{\tau_{i,t}(\omega)}\left(x_{\tau_{i-1,t}(\omega)}^{\alpha} - x_{\tau_{i,t}(\omega)}^{\alpha}\right) &= R_{\tau_{i,t}(\omega)}\left((1-\alpha)x_{\tau_{i-1,t}(\omega)}^{\pi_1} + \alpha x_{\tau_{i-1,t}(\omega)}^{\pi_2} - (1-\alpha)x_{\tau_{i,t}(\omega)}^{\pi_1} - \alpha x_{\tau_{i,t}(\omega)}^{\pi_2}\right) \\
&= R_{\tau_{i,t}(\omega)}\left((1-\alpha)\left(x_{\tau_{i-1,t}(\omega)}^{\pi_1} - x_{\tau_{i,t}(\omega)}^{\pi_1}\right) + \alpha\left(x_{\tau_{i-1,t}(\omega)}^{\pi_2} - x_{\tau_{i,t}(\omega)}^{\pi_2}\right)\right) \\
&\ge (1-\alpha)R_{\tau_{i,t}(\omega)}\left(x_{\tau_{i-1,t}(\omega)}^{\pi_1} - x_{\tau_{i,t}(\omega)}^{\pi_1}\right) + \alpha R_{\tau_{i,t}(\omega)}\left(x_{\tau_{i-1,t}(\omega)}^{\pi_2} - x_{\tau_{i,t}(\omega)}^{\pi_2}\right).
\end{aligned} \tag{A.3}$$

Similarly, concavity of R_T and $x_s^{\alpha} = (1-\alpha)x_s^{\pi_1} + \alpha x_s^{\pi_2}$ yield

$$R_T(x^{\alpha}(\omega)) = R_T\left((1-\alpha)x_T^{\pi_1}(\omega) + \alpha x_T^{\pi_2}(\omega)\right) \ge (1-\alpha)R_T\left(x_T^{\pi_1}(\omega)\right) + \alpha R_T\left(x_T^{\pi_2}(\omega)\right). \tag{A.4}$$

Taking the sum from $i = 1$ to $N_{T^-}(\omega) - N_{t^-}(\omega)$ of inequalities (A.3) and of (A.4) results in

$$\begin{aligned}
\sum_{i=1}^{N_{T^-}(\omega)-N_{t^-}(\omega)} R\left(x_{\tau_{i-1,t}(\omega)}^{\alpha} - x_{\tau_{i,t}(\omega)}^{\alpha}\right) + R_T(x_T^{\alpha}) \ge (1-\alpha) &\sum_{i=1}^{N_{T^-}(\omega)-N_{t^-}(\omega)} R\left(x_{\tau_{i-1,t}(\omega)}^{\pi_1} - x_{\tau_{i,t}(\omega)}^{\pi_1}\right) \\
+ R_T(x_T^{\pi_1}(\omega)) + \alpha &\sum_{i=1}^{N_{T^-}(\omega)-N_{t^-}(\omega)} R\left(x_{\tau_{i-1,t}(\omega)}^{\pi_2} - x_{\tau_{i,t}(\omega)}^{\pi_2}\right) + R_T(x_T^{\pi_2}(\omega)).
\end{aligned}$$

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

$$\begin{aligned}
&\mathbb{E}\left[\sum_{i=1}^{N_{T^-}-N_{t^-}} R\left(x_{\tau_{i-1,t}}^{\alpha} - x_{\tau_{i,t}}^{\alpha}\right) + R_T(x_T^{\alpha})\right] \ge \\
&(1-\alpha)\mathbb{E}\left[\sum_{i=1}^{N_{T^-}-N_{t^-}} R\left(x_{\tau_{i-1,t}}^{\pi_1} - x_{\tau_{i,t}}^{\pi_1}\right) + R_T(x_T^{\pi_1})\right] + \alpha\mathbb{E}\left[\sum_{i=1}^{N_{T^-}-N_{t^-}} R\left(x_{\tau_{i-1,t}}^{\pi_2} - x_{\tau_{i,t}}^{\pi_2}\right) + R_T(x_T^{\pi_2})\right] \\
&= (1-\alpha)V_t^{\pi_1}(k_1) + \alpha V_t^{\pi_2}(k_2) = (1-\alpha)V_t(k_1) + \alpha V_t(k_2).
\end{aligned}$$

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 (A.2), 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. $\square$

Proof of Proposition 4.5. Fix the charge level k. For any given $\epsilon > 0$, let $\delta_\epsilon > 0$ be such that $c_r\lambda\left(\delta_\epsilon + 2T(1-e^{-\lambda\delta_\epsilon})\right) < \epsilon$. Consider any times t_1 and t_2 such that $|t_1 - t_2| < \delta_\epsilon$. 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 charge 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,

$$|V_{t_1}(k) - V_{t_2}(k)| = V_{t_1}(k) - V_{t_2}(k) = V^{\pi_1}_{t_1}(k) - V_{t_2}(k) \leq V^{\pi_1}_{t_1}(k) - V^{\pi_2}_{t_2}(k), \tag{A.5}$$

where the first equality comes from Proposition 4.1 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

$$\mathbb{E}\left[\sum_{i=1}^{N_{t_2^-}-N_{t_1^-}} R_{\tau_{i,t_1}}\left(x^{\pi_1}_{\tau_{i-1,t_1}} - x^{\pi_1}_{\tau_{i,t_1}}\right) \;\middle|\; x^{\pi_1}_{t_1} = k\right] \leq c_r\lambda(t_2 - t_1) < c_r\lambda\delta_\epsilon, \tag{A.6}$$

and consequently,

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

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 $V^{\pi_1}_{t_1}(k) - V^{\pi_2}_{t_2}(k) < c_r\lambda\delta_\epsilon + Q$, where

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

Define the events $A \stackrel{\text{def}}{=} \{\tau_{1,t_1} > t_2\}$ and $B \stackrel{\text{def}}{=} \{\tau_{1,t_1} \leq t_2\}$. When the event A occurs, the policy π_1, starting from charge level k_1 at time t_1, results in $x^{\pi_1}_{t_2} = k$. Therefore, $\mathbb{E}[Q|A] = 0$. 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 $c_r\lambda T$. Hence, $\mathbb{E}[Q|B] \leq 2c_r\lambda T$. By using $\Pr(B) = 1 - e^{-\lambda(t_2-t_1)}$, we get $\mathbb{E}[Q] = \mathbb{E}[Q|B]\Pr(B) \leq 2c_r\lambda T\left(1 - e^{-\lambda(t_2-t_1)}\right) < 2c_r\lambda T(1 - e^{-\lambda\delta_\epsilon})$. Replacing this inequality in $V^{\pi_1}_{t_1}(k) - V^{\pi_2}_{t_2}(k) < c_r\lambda\delta_\epsilon + Q$ and (A.5) yields

$$|V_{t_1}(k) - V_{t_2}(k)| < c_r\lambda\delta_\epsilon + 2c_r\lambda T\left(1 - e^{-\lambda\delta_\epsilon}\right) < \epsilon,$$

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

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grants No. 1610302 and 1610825.

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.

Bremaud, P. (1981). *Point Processes and Queues: Martingale Dynamics*. Springer, New York.

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.

Cazalet, E., P. D. Martini, J. Price, E. Woychik, and J. Caldwel (2016). Transactive energy models. *NIST Transactive Energy Challenge: Business and Regulatory Models Working Group* pp. 1–44.

Çınlar, E. (2011). *Probability and Stochastics*. Springer, New York.

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. *National Renewable Energy Laboratory Technical Report* 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.

DOE Report (2011). The importance of flexible electricity supply. *Energy Efficiency and Renewable Energy Report* DOE/GO-102011-3201, 1–4.

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.

Hawkes, A. G. and D. Oakes (1974). A cluster process representation of a self-exciting process. *Journal of Applied Probability* 11(3), 493–503.

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.

Kristov, L., P. D. Martini, and J. Taft (2016). Two visions of a transactive electric system. *IEEE Power & Energy Magazine* 14(3), 1–12.
URL `http://resnick.caltech.edu/docs/Two_Visions.pdf`

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.

Lewis, P. A. W. and G. S. Shedler (1979). Simulation of nonhomogeneous poisson processes by thinning. *Naval Research Logistics Quarterly* 26, 403–413.

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 (2017). Parallel non-stationary direct policy search for risk averse stochastic optimization. *INFORMS Journal on Computing* 29(2), 332–349.

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

Ogata, Y. (1981). On lewis' simulation method for point processes. *IEEE Transactions on Information Theory* 27(1), 23–31.

Olken, M. I. (2016). Transactive energy: Everyone gets into the act. *IEEE Power & Energy Magazine* 14(3), 4–16.

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.

Rahimi, F. and A. Ipakchi (2016). Using a transactive energy framework. *IEEE Electrification Magazine* pp. 23–29.

Rahimi, F. A. and A. Ipakchi (2012). Transactive energy techniques: Closing the gap between wholesale and retail markets. *The Electricity Journal* 25(8), 29–35.

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.

Sigman, K. (2013). *Lecture Notes on Simulation, Non-stationary Poisson processes*. Columbia University.

Sioshansi, R., P. Denholm, and T. Jenkin (2012). Market and policy barriers to deployment of energy storage. *Economics of Energy and Environmental Policy* 1(2), 1–14.

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 *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.