Low-Rank Value Function Approximation for Co-Optimization of Battery Storage

Bolong Cheng, *Student Member, IEEE*, Tsvetan Asamov, and Warren B. Powell, *Member, IEEE*

***Abstract*—We develop a near-optimal solution to the problem of co-optimizing frequency regulation and energy arbitrage with battery storage using backward approximate dynamic programming, which is shown to handle the different time scales of each revenue stream. Solution of the problem using classical backward exact dynamic programming is computationally intractable for this problem due to the large state space and long horizon. Instead, we use state sampling and low-rank approximations to estimate the entire value function, producing a high quality solution that can be computed in real time. The new algorithm is shown to reduce the computational time by one order of magnitude, and the storage requirements by two orders of magnitude, while producing near optimal policies that consistently outperform pure frequency regulation.**

***Index Terms*—Energy storage, frequency regulation, energy arbitrage, low-rank approximation.**

I. Introduction

CO-OPTIMIZING battery storage for multiple revenue streams is a very complex optimization problem for numerous reasons: it requires jointly optimizing a single device within fixed capacity and power constraints; making decisions that are on different time scales, from every two seconds for frequency regulation to every five minutes for energy arbitrage; and trading off between near term and long term profits. The co-optimization problem is particularly difficult because it requires turning an easy control problem (following the regulation signal for frequency regulation) into an optimization problem that requires balancing compliance penalties for not following the signal with the benefits of energy arbitrage.

Energy storage is an important technology that eases the integration of renewable resources in the electricity grid. There is an abundance of research dedicated to pairing storage devices with renewable generations. Reference [1] optimizes the pairing of compressed air energy storage (CAES) and wind generation, and finds it to have the lowest short-run marginal cost when the green house gas emission tax is high. Reference [2] studies the advanced energy commitment problem by pairing a wind farm with storage. An analytical optimal policy is derived for the infinite horizon case under some stationarity assumptions.

In addition, storage devices provide a wide range of economic benefits, ranging from peak shaving and backup reserves to ancillary services such as frequency and voltage regulations for the power grid [3]. In recent years, co-optimizing battery storage for multiple revenue streams has gained significant interest in both the research community and industry as the market has come to recognize that energy arbitrage by itself is not enough to justify the investment cost of a battery. Reference [4] shows that large scale energy storage can dampen the price difference between on- and off-peak hours, thus reducing the arbitrage value of a price-taking device. It suggests that device owners can increase the value of storage by co-optimizing between different markets such as frequency regulation and spinning reserves. One popular application for battery operators is frequency regulation, which is a dedicated demand response that helps stabilize the operating frequency of the grid. In the case of the Pennsylvania-New Jersey-Maryland Interconnection (PJM), the dynamic regulation (RegD) signal is specifically developed for storage units with limited capacity, where the signal is designed to be energy neutral and requires resources with a high ramp rate [5]. In the research community, [6] has surveyed the policies on the participation of storage in five major U.S. frequency regulation markets and concludes that the "pay-for-performance policies" have made frequency regulation an attractive revenue source for energy storage. Reference [7] has also shown frequency regulation can be a substantial revenue stream for energy storage.

Reference [8] first claims a high probability of positive net present value of using storage device for both energy arbitrage and frequency regulation service in the New York ISO. The economic analysis from [8] is derived from a "charge-off-peak and discharge-on-peak" heuristic policy using aggregated price duration patterns. Reference [9] formulates the co-optimization problem as a stochastic dynamic program that solves for hourly optimal decisions. In [9], the regulation signal and response are modeled using the dispatch-to-contract ratio at an hourly aggregation. However, frequency regulation requires decisions at the sub-minute level (typically every two to four seconds) in practice. A more detailed model is required in order to truly understand the intricate behavior of the storage at the sub-hourly level.

Storage optimization is a sequential decision problem, and is typically modeled as a Markov decision process (MDP).

Manuscript received October 25, 2016; revised February 9, 2017 and May 5, 2017; accepted June 1, 2017. Date of publication June 16, 2017; date of current version October 19, 2018. Paper no. TSG-01463-2016. *(Corresponding author: Bolong Cheng.)*

B. Cheng is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: bcheng@princeton.edu).

T. Asamov and W. B. Powell are with the Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSG.2017.2716382

Many realistic settings of the problem require solving MDP's with large state space (on the order of 10^6 states per time step). Lookup tables (the standard representation for states/value function) are computationally intractable in the presence of the multidimensional state variables that arise in this setting. A variety of approximation strategies have been proposed for approximate dynamic programming (ADP) for energy storage using a host of statistical learning tools (linear and locally linear models, Gaussian process regression, support vector regression), but these have not produced high quality solutions when compared against optimal benchmarks (see [10]). The reasons for the underperformance of the statistical methods are two-fold: 1) general purpose statistical learning methods do not exploit the problem structure (concavity in this case) as well as algorithms designed specifically for the occasion, and 2) the high computational cost associated with the policy improvement step (in the approximate policy iteration algorithm) limits the algorithm to optimize over a coarse subset of the state space. On the other hand, there has been success in algorithms designed to exploit different structures in the value functions. The Concave Adaptive Value Estimation (CAVE) algorithm of [11] estimates the value functions with concave, piece-wise linear approximations, speeding up the convergence rate of the approximate value iteration algorithm. Convexity has also been studied extensively in multistage linear stochastic programs (see [12]–[14]), but approximations for convex functions only handle the resource variable (i.e., the amount of energy in the battery), and offers no help with the other state variables, e.g., spot electricity price (locational marginal price or LMP), regulation market clearing price, and performance metric. Reference [15] provides a convergent algorithm for solving ADP when the value function is monotonic in all dimensions. The algorithm is shown to attain near-optimal policies while decreasing the computation requirement by nearly two orders of magnitude.

This work extends an earlier paper [16] where we first modeled the co-optimization problem as a multi-scale Markov decision process, from the hourly planning of the battery down to the two second response to the RegD signal. However, that paper fixed the regulation market clearing price (RMP) for the entire horizon, and did not consider the high correlation between the LMP and RMP processes, which could dampen the potential advantage from co-optimization. In this paper, we carefully consider this issue, and produce, for the first time, an accurate estimate of the value of co-optimization under very realistic modeling assumptions. From algorithmic and computational standpoints, [16] addresses the challenge of storing the value functions by finding low rank approximation of the value functions using singular value decomposition (SVD). However, this method still requires solving the exact value function first and then reducing it to a lower-dimensional representation; the step to compute the exact value function is computationally demanding as is the SVD operation. In this paper, we introduce for the first time the idea of computing the exact value function for only a small sample of states (1∼2% of the entire state space), and then approximating the entire value function using rank-1 submatrices which are estimated by solving sequences of linear programming problems. Unlike classical ADP techniques (forward approximate dynamic programming), our method still uses a single backward pass as is used in standard backward dynamic programming. Forward ADP methods were found to produce solutions around 60 to 80 percent of an optimal benchmark [10], which would not be sufficient to outperform a pure frequency regulation policy. Reference [17] has also explored using low rank approximation of the value function for MDP; however, this works suffers the same drawback as our previous work, where we need to compute the exact value function before performing dimension reduction.

This paper is organized as follows. In Section II, we present a modified multi-scale dynamic programming model of the problem. In Section III, we discuss the challenges of computing the value functions and introduce our sparse low rank approach for approximating the value functions within the dynamic program. In Section IV, we simulate the policies obtained from the sparse low rank value function approximation on real price data from PJM and compare it with the baseline frequency regulation revenue. Section V concludes the paper.

II. The Multi-Scale Dynamic Programming Model

We follow the formulation in [16] and set up the problem as a nested MDP. Whereas [16] decomposes the problem into three levels (a daily problem in hourly increments, an hourly problem in 5-minute increments, and a 5-minute problem in 2-second increments), we only use two levels in this paper and use a stochastic model of the regulation market clearing price and LMP (along with their correlation). The coarse level now computes the optimal energy basepoint decisions for the horizon of the entire day in increments of five minutes. The finer level then computes the optimal response decision every two seconds for the horizon of the five minutes when the real-time prices are locked in. We restate the mathematical model of each sub-problem in the subsequent subsections.

A. The Parameters of the System

We model our problem from the perspective of the battery operator and optimize the operation of the battery over the finite-time horizon of one day. In order to reflect the two time-scales, we use two sets of notations to enumerate the time intervals of a day. First, let $\mathcal{T} = \{0, \Delta t, 2\Delta t, \ldots, 43200\Delta t = T\}$ index the increments of two seconds (Δt) in a day. We denote $\tau \triangleq 150\Delta t$, the equivalent of a five-minute interval and let $\mathcal{T}^{EB} = \{0, \tau, 2\tau, \ldots, 288\tau = T\}$ be the set marking the increments of five-minutes within a day. The battery is characterized by the following operational parameters:

- $R^{\max}$: The maximum capacity of the battery in MWh.
- η^c, η^d: The charging and discharging efficiency of the device, where $0 \leq \eta^c, \eta^d \leq 1$ and the roundtrip efficiency is the product of the two.
- β: The maximum power capacity of the device in MW: the battery cannot charge or discharge beyond this rate.

- K: The hourly regulation capacity in MW assigned by the grid operator. The magnitude of the regulation signal is bounded by K.

Lastly, we assume that the battery has cleared the (energy and regulation) markets for the time horizon and behaves as a price taker. We recognize that the bidding strategy is an essential part of storage optimization, but it is a complex problem that is well beyond the scope of this paper.

B. The Five-Minute Energy Basepoint (EB) Model

In the EB model, we are interested in setting the energy basepoint for participating in the energy market. We now write out the five components of this MDP: states, decisions, exogenous information, transition function, and objective function. Since the real-time LMP updates every five minutes, we use the set $\mathcal{T}^{EB}$ as the time indexing system.

We let $S_t^{EB} = (R_t, G_t, P_t^D, P_t^E)$ describe the state of the system at time t. R_t is the amount of energy in the battery at time t in MWh. G_t represents the performance score at time t. The performance score measures how closely the battery follows the regulation signal: It is bounded in [0, 1], with 1 being full compliance of the RegD signal. At the end of each hour, the performance score is used in the hourly regulation market settlement (see (11)). We define $G_t \equiv 1$ at the beginning of every hour. P_t^D is the regulation market clearing price (RMP) at time t in \$/MW. P_t^E is the spot energy market price (LMP) at time t in \$/MWh.

There are two decisions for this sub-problem: the energy basepoint x_t^E, the charge/discharge rate of the battery for the five-minute interval, and the performance degradation limit x_t^G, the auxiliary decision that limits how much the battery can deviate from the RegD signal for the five-minute interval. Usually, the battery adjusts the energy point x_t^E to react to a price signal (LMP) or state-of-charge level. Note that the frequency regulation signal will be modulated around the energy basepoint rate, which we model in Section II-C. Both decisions are made every five minutes and must satisfy the following constraints:

$$0 \le R_t + \left((1 - x_t^G)K\left(\mathbf{1}_{\{x_t^E>0\}} - \mathbf{1}_{\{x_t^E<0\}}\right) + x_t^E\right) \cdot \tau \le R^{\max}, \tag{1}$$

$$\left|x_t^E + (1 - x_t^G)K\left(\mathbf{1}_{\{x_t^E>0\}} - \mathbf{1}_{\{x_t^E<0\}}\right)\right| \le \beta, \tag{2}$$

$$0 \le x_t^G \le 1, \tag{3}$$

where $\mathbf{1}_{\{\star\}}$ is the indicator function. These constraints are viewed as a trade-off between frequency regulation and energy arbitrage. Equation (1) enforces the battery capacity to satisfy the energy capacity limit, even when the regulation signal requires charging/discharging at the full rate for the entire five minutes. Equation (2) guarantees the battery never exceeds the power capacity. In (3), the lower limit $x_t^G = 0$ represents that the battery must strictly follow the signal for the next five minutes and the upper limit $x_t^G = 1$ allows it to disobey the signal completely. We can see from these constraints that whenever x^G is close to 0, the magnitude of x^E is forced to be close to 0. Intuitively, this is equivalent of stating that the energy basepoint should be 0 (or neutral) if we want to perfectly follow the RegD signal under all circumstances.

We introduce two exogenous variables $\hat{R}_t^+$ and $\hat{R}_t^-$, where $\hat{R}_{t+\tau}^+$ is the total amount of energy charged to the battery due to performing frequency regulation between t and $t+\tau$, and $\hat{R}_{t+\tau}^-$ is the total amount of energy discharged over the same interval. The transition function for the resource state R_t is expressed as

$$R_{t+\tau} = R_t + \eta^c \hat{R}_{t+\tau}^+ + \hat{R}_{t+\tau}^- + x_t^E \cdot \tau\left(\mathbf{1}_{\{x_t^E<0\}} + \mathbf{1}_{\{x_t^E>0\}}\eta^c\right), \tag{4}$$

The performance score G_t and the price processes P_t^D, P_t^E evolve randomly according to the following transition functions:

$$G_{t+\tau} = G_t + \hat{G}_{t+\tau}, \tag{5}$$

$$P_{t+\tau}^D = P_t^D + \hat{P}_{t+\tau}^D, \tag{6}$$

$$P_{t+\tau}^E = P_t^E + \hat{P}_{t+\tau}^E. \tag{7}$$

We describe the price processes in detail in Section IV. We also note that the distribution of $\hat{G}_{t+\tau}$ depends on the decision x_t^G and the policy from the frequency regulation problem. In particular, we know that $G_{t+\tau}$ is bounded above by G_t and below by $G_t - x_t^G/12$. The performance score G_t is reset to 1 at the beginning of every hour. We introduce the dummy variable $G_t^-, t \in \{12\tau, 24\tau, \ldots, T\}$ as the performance score at the end of each hour (before it is being reset) and modify (5) slightly, where

$$G_t = \begin{cases} 1, & \forall t \in \{0, 12\tau, 24\tau, \ldots, T\}, \\ G_{t-\tau} + \hat{G}_t, & \text{otherwise} \end{cases} \tag{8}$$

$$G_t^- = G_{t-\tau} + \hat{G}_t, \quad \forall t \in \{12\tau, 24\tau, \ldots, T\}, \tag{9}$$

In summary, the exogenous information that is observable at time t is $W_t = (\hat{R}_t^+, \hat{R}_t^-, \hat{G}_t, \hat{P}_t^D, \hat{P}_t^E)$. The state transition function $S_{t+\tau}^{EB} = S^M(S_t^{EB}, x_t, W_{t+\tau})$ is characterized by (4)–(9). The revenue functions for the EB problem are described by

$$C^{EB}(S_t^{EB}, x_t, W_{t+\tau}) = -P_t^E\Big(\hat{R}_{t+\tau}^+ + \eta^d \hat{R}_{t+\tau}^- + x_t^E \cdot \tau\left(\eta^d \mathbf{1}_{\{x_t^E<0\}} + \mathbf{1}_{\{x_t^E>0\}}\right)\Big), \quad \forall t \in \mathcal{T}^{EB} \setminus T. \tag{10}$$

$$C_t^{EB} = KP_{t-\tau}^D G_t^- \cdot \mathbf{1}_{\{G_t^- \ge 0.4\}}, \quad \forall t \in \{12\tau, 24\tau, \ldots, T\}. \tag{11}$$

Equation (10) is the payment associated with setting the energy basepoint every five minutes and the amount of energy used for jointly providing frequency regulation. Equation (11) represents the hourly settlement from the frequency regulation market, where a device with a performance score lower than 0.4 will forfeit the credit. Note that time index for $P_{t-\tau}^D$ is $t - \tau$ since the RMP is constant within the hour in practice. Our objective is to find the optimal policy π^{EB} from the set of all admissible policies Π^{EB} over the horizon of 24 hours defined by the objective function:

$$\max_{\pi^{EB} \in \Pi^{EB}} \mathbf{E}^{\pi^{FR}}\left[\sum_{t \in \mathcal{T}^{EB}} C^{EB}\left(S_t^{EB}, X_t^{\pi^{EB}}(S_t^{EB}), W_{t+\tau}\right) \middle| S_0^{EB}\right]. \tag{12}$$

The superscript π^{FR} over the expectation implies that the frequency regulation policy influences the underlying stochastic processes of $(\hat{R}_t^+, \hat{R}_t^-, \hat{G}_t)$ in this sub-problem.

C. The Two-Second Frequency Regulation (FR) Model

The FR problem determines the optimal response policy at the two-second (Δt) time scale for the horizon of five minutes when the LMP and RMP are fixed. The value function of this sub-problem depends on prices P^E and P^D as well as the energy basepoint x^E determined in the EB problem. We have a three dimensional state variable $S_t^{FR} = (R_t, G_t, D_t)$, where D_t is the frequency regulation signal at time t. A positive value of D_t is signaling the battery to discharge and the negative value is signaling the battery to charge.

The only decision for the FR problem is the response to the regulation signal, x_t^D. Given the optimal energy basepoint x^E and the degradation limit x^G determined from the EB problem, the response x_t^D must satisfy the following set of constraints:

$$0 \le R_t + (x_t^D + x^E) \cdot \Delta t \le R^{\max}, \tag{13}$$

$$|x_t^D + x^E| \le \beta, \tag{14}$$

$$|x_t^D + D_t| \le K x^G. \tag{15}$$

Equation (13) requires that the amount of energy stored in the device satisfies its energy capacity at any time t. Equation (14) ensures the charge/discharge rate (RegD response plus energy basepoint) does not exceed the power capacity. The D_t is modulated around the energy point of the device; therefore, the device will be limited in compliance if the energy point is non-zero due to the power capacity constraint. Equation (15) limits the deviation from the regulations signal to be within the limit x^G dictated by the EB problem.

Now we describe the evolution of the state variables in the context of two-second dynamics. The transition functions for the controllable state variables R_t and G_t are described as:

$$R_{t+\Delta t} = R_t + \left(x_t^D + x^E\right)\left(\mathbf{1}_{\{x_t^D + x^E < 0\}} + \mathbf{1}_{\{x_t^D + x^E > 0\}}\eta^c\right) \cdot \Delta t, \tag{16}$$

$$G_{t+\Delta t} = G_t - \min\left\{\left|x_t^D\left(\eta^d \mathbf{1}_{\{x_t^D < 0\}} + \mathbf{1}_{\{x_t^D > 0\}}\right) + D_t\right|K^{-1}, 1\right\} \cdot (1800)^{-1}. \tag{17}$$

The performance score G_t is deducted when x_t^D deviates from the signal D_t and is therefore non-increasing in t; 1800 is the number of two-second intervals within an hour. The RegD signal D_t evolves randomly according to the transition function:

$$D_{t+\Delta t} = D_t + \hat{D}_{t+\Delta t}, \tag{18}$$

where $\hat{D}_{t+\Delta t}$ is the only exogenous information of the system describing the change in D_t from time t to $t + \Delta t$. In practice, we find that the D_t process can be well approximated by a first-order Markov chain trained from the empirical RegD signal data. The revenue function for the FR problem is

$$C^{FR}\left(S_t^{FR}, x_t^D\right) = -P_t^E\left(x_t^D + x^E\right)\left(\eta^d \mathbf{1}_{\{x_t^D + x^E < 0\}} + \mathbf{1}_{\{x_t^D + x^E > 0\}}\right) \cdot \Delta t. \tag{19}$$

Equation (19) indicates the cost accrued at the two-second level for charging/discharging in the grid (from the FR response and the energy basepoint). The discharge efficiency η^d reflects the loss of energy discharged from the battery.

The objective function for finding the optimal policy π^{FR} from the set of admissible policies Π^{FR} is written as,

$$\max_{\pi^{FR} \in \Pi^{FR}} \mathbf{E}\left[\sum_{t=t'}^{t'+\tau} C^{FR}\left(S_t^{FR}, X_t^{\pi^{FR}}\left(S_t^{FR}\right)\right)\middle| S_{t'}^{FR}, x^E, P^E, P^D\right], \tag{20}$$

for all $t \in \mathcal{T}^{EB}$, given the energy basepoint x^E, the LMP P^E and the regulation price P^D for the five-minute horizon. This sub-problem needs to be solved from all combinations of (x^E, P^D, P^E).

D. Algorithmic Approach

We first compute the optimal value function of the EB problem $V^{EB}(S_t^{EB})$ for the 24-hour horizon in five-minute increments by recursively solving the Bellman equation:

$$V_t^*\left(S_t^{EB}\right) = \max_{x_t^{EB} \in \mathcal{X}_t} \mathbf{E}\Big[C^{EB}\left(S_t^{EB}, x_t^{EB}, W_{t+\tau}\right) + V_{t+\tau}^*\left(S_{t+\tau}^{EB}\right)\Big|S_t^{EB}\Big], \forall t \in \mathcal{T}^{EB}, \tag{21}$$

with the terminal condition $V_T^{EB}(S_T^{EB}) = 0$. The expectation is taken over the exogenous variable $W_{t+\tau} = (\hat{R}_{t+\tau}^+, \hat{R}_{t+\tau}^-, \hat{G}_{t+\tau}, \hat{P}_{t+\tau}^D, \hat{P}_{t+\tau}^E)$. The value function is then used as the terminal conditions of the FR problems for every five minute interval, for all combinations of (x_t^G, P_t^D, P_t^E). For each instance of the FR problem, the optimal value function can then be computed by solving

$$V_t^*\left(S_t^{FR}\right) = \max_{x_t^D \in \mathcal{X}_t} C^{FR}\left(S_t^{FR}, x_t^D\right) + \mathbf{E}\Big[V_{t+\Delta t}^*\left(S_{t+\Delta t}^{FR}\right)\Big|S_t^{FR}\Big], \tag{22}$$

where the terminal value $V_{n\tau}^{FR} = V_{n\tau}^{EB}(S_{n\tau}^{EB})$, for $n = 1, \ldots, 288$. The expectation is taken with respect to the only exogenous variable D_t. The algorithmic approach is illustrated in Fig. 1. To optimize the battery for the horizon of 24 hours, we have to solve $24 \times 12 \times |x^E| \times |P^D| \times |P^E|$ sub-problems (50,400 instances of the FR sub-problem if using the discretization scheme from [16]). By introducing the RMP dynamics P_t^D into the state of the system, we further scale up the computational requirement for the optimization problem, which we discuss in detail next.

III. Low Rank Approximations in Dynamic Programming

Since the value functions from Section II-C are represented in the look-up table format (by discretizing the state space and the action space), the storage requirement for the entire value functions for all sub-problems becomes prohibitively high. Reference [16] estimates that the full value functions require more than 60 terabytes of disk space (when P^D is held constant for the entire horizon). This challenge is addressed by computing the low rank representation of the value functions via singular value decomposition (SVD), thus reducing

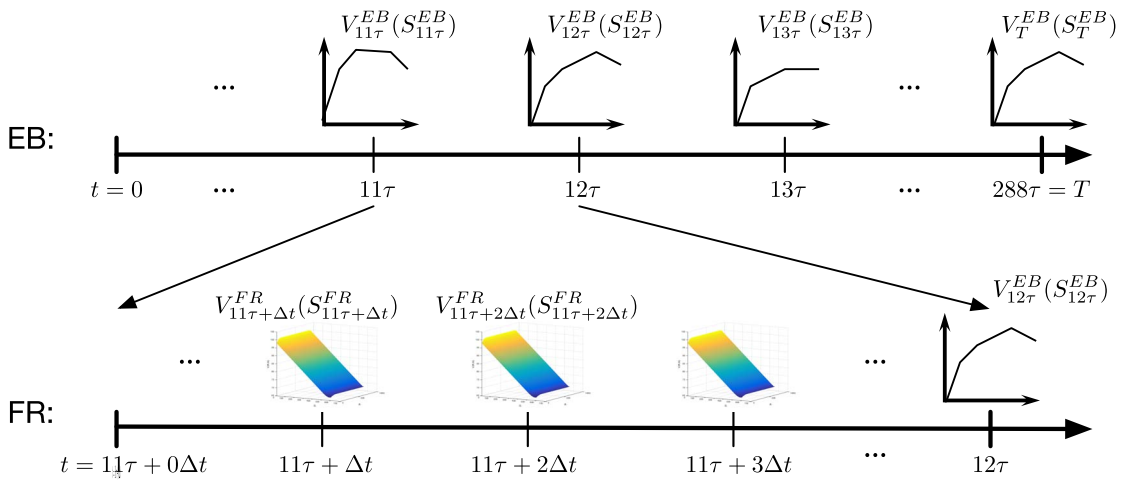


Fig. 1. Illustration of the algorithmic approach of optimizing for the entire 24-hour horizon. The first axis illustrates the value functions of the EB problem. The second axis represents the value functions of the FR problems for the five-minute interval from $t = 11\tau$ to $t = 12\tau$. Note that the terminal value function of the FR problem comes from the EB problem.

the storage requirement by a factor of 100 while still enjoying the same optimality (with a rank-10 approximation in this particular case).

As we have mentioned in the introduction, this method is inefficient since it requires solving the exact value functions first and then performing the SVD to find the low rank representation. Since we now model the RMP as a state variable, our state space is much larger than that from our previous work. Suppose that we discretize the RMP into 5 levels and use the same discretization scheme for the other state variable as before; we now need to solve equation (22) approximately 10^9 times per two-second increment, and over 3×10^{11} times for the entire 24 hour horizon. Although the FR problems can be executed in parallel, we can further improve the efficiency by avoiding repeatedly computing the value functions for similar states in the backward dynamic programs and take advantage of the low-rank structure. In the following subsections, we explore a sparse low rank approximation technique that can be embedded in the backward dynamic program and compare this new method to the SVD approximation.

A. Sparse Low Rank Value Function Approximation

Without loss of generality, let V_t be the value function of the state space S_t^{FR} and let V_t be represented using matrix notation with $m = |R|$ rows and the cross product $n = |G| \times |D|$ columns, so that $V_t \in \mathbb{R}^{m \times n}$. Knowing that V_t lies on some low dimensional subspace, we want to find a low rank approximation $\tilde{V}_t$ that can be used in the backward dynamic program, so that we can reduce the memory and disk space required by our method. In addition, we can also reduce the CPU time by evaluating the value function V_t only for a small sample set of the states, and using the information obtained in the sample to construct a low rank approximation to the entire value function V_t. Thus, we would be able to recover the entire value function efficiently.

Formally, we define $\mathcal{S}^S = \{(i, j) : 1 \leq i \leq m, 1 \leq j \leq n\}$ as the set of states sampled from the state space S^{FR}, where ideally we want to sample at least one entry from each row i, and each column j. Therefore the size of the sample should satisfy the following:

$$\max(m, n) \leq |\mathcal{S}^S| \leq m + n \ll mn.$$

Now we consider the standard low rank matrix completion problem which seeks to find the lowest rank approximation $\tilde{V}_t$ that matches the original matrix V_t, given the set $\mathcal{S}^S$ of observed entries. The objective can be formulated as the rank minimization problem:

$$\begin{aligned} \min_{\tilde{V}_t} \quad & Rank\big(\tilde{V}_t\big) \\ \text{s.t.} \quad & \tilde{V}_t(i, j) = V_t(i, j) \quad \forall (i, j) \in \mathcal{S}^S. \end{aligned} \tag{23}$$

However, the matrix rank function is non-convex, and the optimization problem in (23) has shown to be NP-hard. In order to overcome this challenge, [18] considers replacing the rank function with its tightest convex relaxation given by the nuclear norm. Thus it proposes solving the following problem:

$$\begin{aligned} \min_{\tilde{V}_t} \quad & \big\|\tilde{V}_t\big\|_* \\ \text{s.t.} \quad & \tilde{V}_t(i, j) = V_t(i, j) \quad \forall (i, j) \in \mathcal{S}^S, \end{aligned} \tag{24}$$

where the nuclear norm $||V_t||_*$ is the sum of its singular values. It has been shown that when the sample size is large enough, the relaxation is tight and the solution of the convex relaxation (24) is optimal for the NP-hard rank minimization problem (23). More specifically, the matrix V_t of rank r can be recovered exactly with probability $\sim$1 if the size of the observed entry sets satisfies the condition $|\mathcal{S}^S| \geq Cn^{6/5} r \log n$ for some positive numerical constant C [18]. The exact solution can be obtained by reformulating (24) as a semi-definite program. However, this approach is not practical for non-trivial

instances. In order to resolve this issue, [19] develops the singular value thresholding (SVT) algorithm to approximately solve the optimization of (24). This algorithm is a first order method and can significantly reduce the computation overhead for matrices of large size and low rank. However, the iterative algorithm requires computing the partial SVD of the incumbent solution in each shrinkage step and our empirical results indicated that it requires a large sample size in order to recover the optimal MDP value functions with high precision.

Thus, instead of finding the minimally ranked approximation, we consider a slightly different optimization problem

$$\min_{y_t, z_t} \sum_{(i,j)\in\mathcal{S}^S} (V_t(i,j) - y_t(i) z_t(j))^2, \tag{25}$$

where $y_t \in \mathbb{R}^m$, $z_t \in \mathbb{R}^n$, and $y_t(i)$ is the i-th component of the y_t. This is the equivalent to finding the best rank-1 estimation of V_t under the Frobenius norm; however, this optimization problem is non-convex (due to the mixed terms inducing indefinite Hessians) and difficult to solve. However, when the rank of the matrix V_t is one, we can closely approximate (25) with a log-linearized formulation as follows:

$$\min_{y_t, z_t} \sum_{(i,j)\in\mathcal{S}^S} \|\log V_t(i,j) - (y_t(i) + z_t(j))\|_1, \tag{26}$$

which can be solved efficiently as a linear program using any commercial optimization solver. From the optimal solution y_t^*, z_t^*, we can approximate the value function $V_t \approx \exp(y_t^*)\exp(z_t^*)^T$. In practice, the rank of V_t might be greater than one but empirically we found that we can partition V_t into sub-matrices that can be approximated very closely by rank-one matrices. In the case where computational speed is of high priority, we can also modify (26) with the L-2 norm penalty, i.e.,

$$\min_{y_t, z_t} \sum_{(i,j)\in\mathcal{S}^S} \|\log V_t(i,j) - (y_t(i) + z_t(j))\|_2^2. \tag{27}$$

This becomes an unconstrained optimization problem, and the closed form expression for the optimal solutions is precisely the ordinary least square estimator. To further reduce the computation overhead, we can use the same sample set $\mathcal{S}^S$ for all sub-matrices and compute only one Moore-Penrose pseudo-inverse matrix (the estimator). As mentioned earlier, we sample states uniformly from each row and column of the value function matrix. Algorithm 1 outlines the procedures for computing the low-rank value function approximation within a backward dynamic program.

B. Comparing Sparse Low Rank Value Function Approximation to SVD

We test the sparse low rank approximation on the test problems presented in [16] where we optimize the battery for 24 hours but use a constant regulation market price for the horizon. We discretize the state space S_t^{FR} to be $|R| \times |G| \times |D| = 900 \times 600 \times 21$ and all other parameters and settings remain unchanged. For the sparse low rank approximation, the matrix V_t is partitioned into 4×63 sub-matrices, which are computed in parallel. In practice, we keep all sub-matrices to be the same size for ease of implementation. We also find that keeping the dimension of the sub-matrix to be a unit fraction of the size of particular state variable produces a better approximation, e.g., seach sub-matrix has 200 columns which is 1/3 of discretized level of $|G|$. In general, one can exhaustively search through a small combination of sub-matrix arrangements and pick the one that yields the best value function approximation. Furthermore, we uniformly sample one state from each row and three states from each column of the sub-matrix. In Table I, we present the computation and storage comparison of the sparse low rank method to the vanilla backward dynamic program with lookup table and SVD representation for a single FR problem. Fig. 2 illustrates the three ways of representing an example value function of 40,000 states: exact value function (lookup table), rank-1 SVD representation (computed using the exact value function), and the sparse low rank representation computed using the value function of a small subset of states (993 randomly sampled states in this instance).

TABLE I
COMPARING SPARSE LOW RANK VALUE FUNCTION APPROXIMATION TO SVD FOR A SINGLE INSTANCE OF FR PROBLEM

	Look-up Table	SVD	Sparse Low Rank
Average computation time (sec)	~300	~800	~150 for (26), <30 for (27).
Sample size $\|\mathcal{S}^S\|$	N/A	N/A	185,000-190,000
Value function size $\|V_t\|$ or $\|\hat{V}_t\|$	11,340,000	135,010	107,100

Algorithm 1: Sparse Low Rank VFA in Backward DP

input : The terminal value function V_τ, the observed entry set $\mathcal{S}^S$

output: A set of low-rank approximation y_t, z_t for $t = 0, \Delta t, \ldots, \tau - \Delta t$

for $t = \tau - \Delta t, \ldots, \Delta t, 0$ **do**
  for *all sampled states* $(i,j) \in \mathcal{S}^S$ **do**
    compute (22), where
    $V_{t+\Delta t}(i', j') = \exp(y_{t+\Delta t}(i'))\exp(z_{t+\Delta t}(j'))$.
    obtain the newest rank-1 estimate y_t, z_t by solving (26) or (27).
  end
end

The sparse low rank method only requires evaluating the Bellman equation for 1-2% of the state space. The L-1 method from (26) cuts the computation time by a half compared to the look-up table method and the L-2 method from (27) reduces the computation time by one order of magnitude. In Fig. 3, we compare the new sparse low rank policies to the ones derived from the SVD approximation and the baseline pure-FR policy. We see that the policies computed using the sparse low rank approximation are about 95% optimal comparing to the policies from the SVD approximation. In addition, the sparse low rank policies outperform the pure-FR policies in all test cases, except when the regulation price is at \$100/MW (96% vs. 98%). Reference [16] observes that the SVD policy behaves

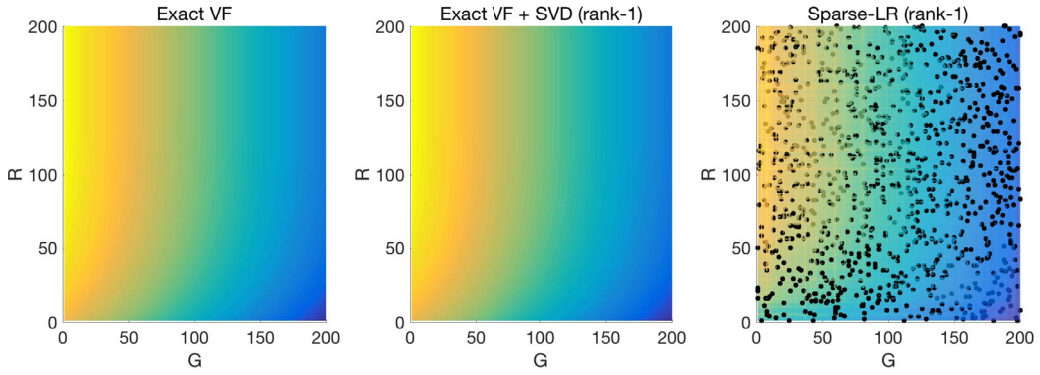


Fig. 2. Illustration of three different representations of a sample value function of 40,000 states. The first figure is the exact value function from computing the Bellman's equation for every state. The center figure is the rank-1 representation of the value function, after reducing the dimension of the exact value function using SVD. The last figure is the rank-1 approximation using the sparse low rank method. The black dots are the states where we have computed the exact value function (2.5% of the entire state space).

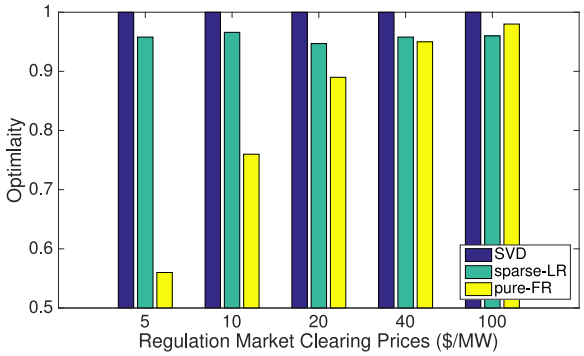
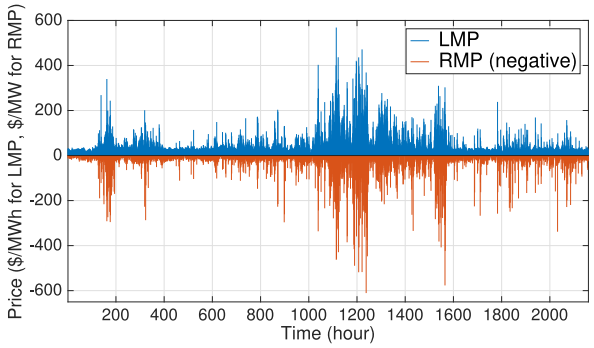
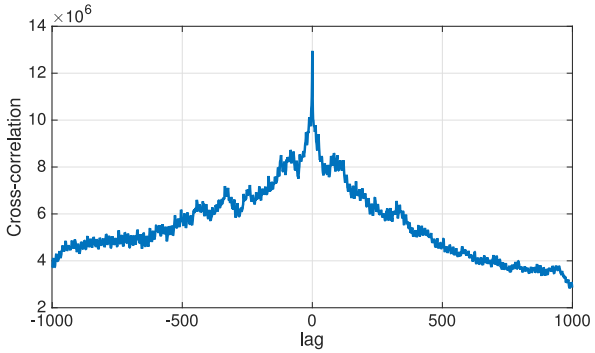


Fig. 3. Comparing the optimality of the policies computed from the sparse low rank approximation to those from the SVD approximation and the baseline pure-FR policy.



(a) The hourly LMP and RMP (negative) from January - March of 2015



(b) Cross-correlation of the LMP and RMP for +/-1000 lag terms

Fig. 4. Comparing the hourly real-time LMP and the regulation market clearing price processes.

similar to the pure-FR policy in the high regulation price scenarios, in the sense that the battery tends to closely follow the RegD signals. In this case, approximation errors from the sparse low rank method become more apparent. However, this scenario (high RMP with low LMPs) is rare in reality due to the correlation between the LMP and RMP, as we will see in the next section.

IV. Case Study: Numerical Simulation

In this section, we compare the co-optimization policy produced by the sparse LR algorithm to the pure-FR algorithm on the real spot (energy) prices and regulation market clearing prices from PJM, using historical data of the PSEG node.

A. Modeling the Price Processes

While our MDP model can be used for any price models, more complicated models require additional state variables to model features such as regime switching and jump terms (as seen in [20] and [21]). We would like to avoid increasing the state space, since our model already has over 50 million states per two seconds. Instead, we are interested in training our MDP with a non-parametric model using sample price paths from the historical data. Our implementation is similar to that proposed in [22], but we also need to take into consideration the correlation between the real-time LMP and RMP.

In Fig. 4a, we plot the LMP and RMP (negative, for illustration purpose) of the first three months of 2015. At a first glance, it is apparent that the two price processes are highly correlated, exhibiting matching price spikes. This matching price pattern is confirmed by examining the cross-correlation of the two processes. Recall that the cross-correlation measures the similarity of two time series as function of the lag of one series to the other; the spike at lag 0 in the Fig. 4b indicates that the two price are correlated in the time domain.

TABLE II
COMPARING REVENUES BETWEEN CO-OPTIMIZATION AND PURE FREQUENCY REGULATION FOR 2015

Months	sparse-LR revenue ($)	pure-FR revenue($)	relative improvement
January-15	22052.47	19131.03	15.27%
February-15	51282.00	46331.60	10.68%
March-15	36518.00	32329.90	12.95%
April-15	24121.20	22272.60	8.30%
May-15	31861.80	30232.44	5.39%
June-15	18975.60	17999.40	5.42%
July-15	18463.29	17152.92	7.64%
August-15	15988.56	14750.42	8.39%
September-15	22336.20	20462.40	9.16%
October-15	17714.33	16553.69	7.01%
November-15	15930.90	15033.30	5.97%
December-15	15079.33	13901.64	8.47%
Yearly Revenue	290323.68	266151.34	9.08%

We assume that the price processes may behave similarly in the same month across years; this motivates us to use the price data from the year 2014 as training data and that from 2015 as testing. We also make the assumption that prices of a particular hour are identically distributed across the days of the same month; this gives us a larger training set. The procedure to create a price model from the real price data is summarized as below:

- Cluster the hourly LMP and RMP data for a particular month. For our numerical work, we discretize the LMP into 7 levels and the RMP into 5 levels.
- Build a time-dependent Markov model for the hourly transition, where each state of the Markov chain is an order pair of LMP and RMP. Note that this is not a two-dimensional Markov chain.
- For each order pair of hourly LMP and RMP, we build an intra-hour Markov model for the real-time five-minute LMP, using sampling paths from the five-minute price data.

B. Numerical Results

For our numerical simulation, we once again set the battery capacity to be 500 KWh capacity with a power capacity of β = 1MW and a round-trip efficiency of $0.9 \times 0.9 = 0.81$. We assume the battery has cleared the regulation market with a regulation capacity of K = 1MW for the entire horizon. For each individual month, we train our model on the historical prices from the year 2014. The price data from days of the same month are aggregated together to create a single daily price model, as we have described before. For testing, we test on the actual price paths from the corresponding month of 2015. For a baseline comparison, we also test the pure-FR policy as described in [16], where the objective of the battery is to maximize the hourly regulation market settlement. In other words, the pure-FR policy minimizes the RegD signal/response deviation. We present the monthly revenues generated by these two policies in Table II.

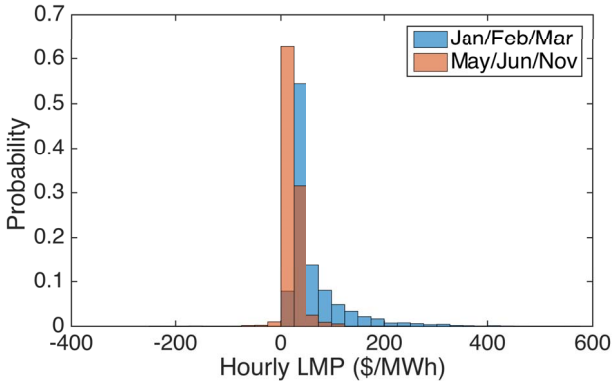


Fig. 5. Empirical distribution of the LMP for high improvement months (Jan/Feb/Mar) and low improvement months (May/Jun/Nov).

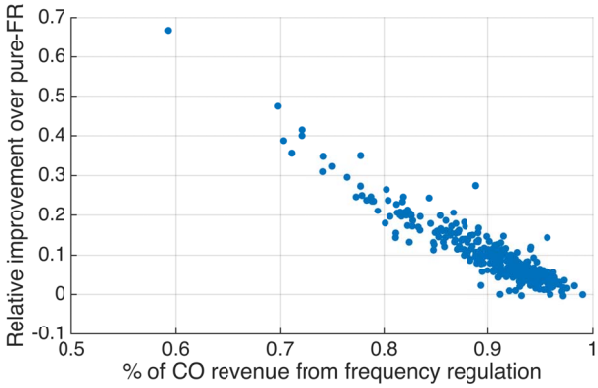


Fig. 6. Correlation between relative improvement of co-optimization over pure-FR and percentage of co-optimization revenue from frequency regulation.

The co-optimization policy consistently outperforms the pure-FR policy. The largest relative increase in revenue are from the months of January, February, and March, with over 10% improvement over frequency regulation (as a single revenue stream). The smallest increase in revenue comes from the months of May, June, and November, all lower than 6%. One explanation for this difference in improvement from co-optimization is that spot prices are higher and more volatile in the months of Jan – March than those in the months of May, Jun, and Nov. This is shown in the empirical distribution of the spot prices for these months in Fig 5. The average LMP price for the low improvement months (May/Jun/Nov) is \$24.55/MWh in comparison to the higher average of \$62.69/MWh for the high improvement months (Jan/Feb/Mar). In addition, the mean standard deviation of daily spot prices is \$40.75 for the high improvement months and \$15.64 for the low improvement months, indicating that the LMP is indeed more volatile in the month of January to March.

There does not seem to be an identifiable pattern in the EB policy as we have seen before in [16], where we observed that the battery is more likely to follow the frequency regulation signal closely when the regulation market price is high and participate in the energy market when the regulation market price is low. This is possibly due to the aforementioned high correlation between LMP and RMP. However, we notice that the relative improvement of co-optimization over pure-FR is

negatively correlated with the percentage of the total revenue, as shown in Fig. 6. In other words, frequency regulation is the more profitable revenue stream of the two, which echoes similar findings from [8] but uses a completely different logic.

V. Conclusion

In this paper, we introduce a sparse low-rank value function approximation technique that can be used to solve multidimensional time dependent battery co-optimization problems. This new technique reduces the computation time of the backward dynamic programming by one order of magnitude and the storage space of the value function by two orders of magnitude. We benchmark the algorithm on a set of simplified co-optimization problems where the regulation price is held constant, and our new algorithm produces near-optimal (over 95%) policies. For reference, the exact solution used in the benchmark requires 60TB of storage space and more than 24 hours to solve on a computing cluster.

We then revise the nested MDP model to take into consideration the dynamics of the regulation market prices. Using this new model, we benchmark our sparse low rank method on the five-minute real-time LMP's and hourly regulation market clearing prices from the years 2014 (as training data) and 2015 (as testing data). The new co-optimization policy consistently outperforms the pure frequency regulation policy for all twelve months. While co-optimization provides 9% increase in the yearly revenue, we recognize that frequency regulation is still the more profitable revenue stream, consisting of nearly 90% of the total revenue.

References

[1] J. B. Greenblatt, S. Succar, D. C. Denkenberger, R. H. Williams, and R. H. Socolow, "Baseload wind energy: Modeling the competition between gas turbines and compressed air energy storage for supplemental generation," *Energy Policy*, vol. 35, no. 3, pp. 1474–1492, 2007.

[2] J. H. Kim and W. B. Powell, "Optimal energy commitments with storage and intermittent supply," *Oper. Res.*, vol. 59, no. 6, pp. 1347–1360, 2011.

[3] "Grid energy storage," U.S. Dept. Energy, Washington, DC, USA, Tech. Rep., 2013.

[4] R. Sioshansi, P. Denholm, T. Jenkin, and J. Weiss, "Estimating the value of electricity storage in PJM: Arbitrage and some welfare effects," *Energy Econ.*, vol. 31, no. 2, pp. 269–277, 2009.

[5] S. Benner. (2015). *A Brief History of Regulation Signals at PJM*. [Online]. Available: http://goo.gl/9UUQts

[6] B. Xu, Y. Dvorkin, D. S. Kirschen, C. A. Silva-Monroy, and J.-P. Watson, "A comparison of policies on the participation of storage in U.S. frequency regulation markets," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, Boston, MA, USA, Jul. 2016, pp. 1–5.

[7] J. M. Eyer, J. J. Iannucci, and G. P. Corey, "Energy storage benefits and market analysis handbook," Sandia Nat. Lab., Albuquerque, NM, USA, Tech. Rep. SAND2004-6177, 2004.

[8] R. Walawalkar, J. Apt, and R. Mancini, "Economics of electric energy storage for energy arbitrage and regulation in New York," *Energy Policy*, vol. 35, no. 4, pp. 2558–2568, 2007.

[9] X. Xi, R. Sioshansi, and V. Marano, "A stochastic dynamic programming model for co-optimization of distributed energy storage," *Energy Syst.*, vol. 5, no. 3, pp. 475–505, 2014.

[10] D. R. Jiang, T. V. Pham, W. B. Powell, D. F. Salas, and W. R. Scott, "A comparison of approximate dynamic programming techniques on benchmark energy storage problems: Does anything work?" in *Proc. IEEE Symp. Adapt. Dyn. Program. Reinforcement Learn. (ADPRL)*, Orlando, FL, USA, Dec. 2014, pp. 1–8.

[11] G. A. Godfrey and W. B. Powell, "An adaptive, distribution-free algorithm for the newsvendor problem with censored demands, with applications to inventory and distribution," *Manag. Sci.*, vol. 47, no. 8, pp. 1101–1112, 2001.

[12] J. R. Birge, "Decomposition and partitioning methods for multistage stochastic linear programs," *Oper. Res.*, vol. 33, no. 5, pp. 989–1007, 1985.

[13] M. V. F. Pereira and L. M. V. G. Pinto, "Multi-stage stochastic optimization applied to energy planning," *Math. Program.*, vol. 52, no. 1, pp. 359–375, 1991.

[14] T. Asamov, D. F. Salas, and W. B. Powell. (2016). *SDDP vs. ADP: The Effect of Dimensionality in Multistage Stochastic Optimization for Grid Level Energy Storage*. [Online]. Available: https://arxiv.org/abs/1605.01521

[15] D. R. Jiang and W. B. Powell, "An approximate dynamic programming algorithm for monotone value functions," *Oper. Res.*, vol. 63, no. 6, pp. 1489–1511, 2015.

[16] B. Cheng and W. B. Powell, "Co-optimizing battery storage for the frequency regulation and energy arbitrage using multi-scale dynamic programming," *IEEE Trans. Smart Grid*, to be published.

[17] H. Y. Ong, "Value function approximation via low-rank models," *CoRR*, vol. abs/1509.00061, 2015. [Online]. Available: http://arxiv.org/abs/1509.00061

[18] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2009.

[19] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.

[20] M. Coulon, W. B. Powell, and R. Sircar, "A model for hedging load and price risk in the Texas electricity market," *Energy Econ.*, vol. 40, pp. 976–988, 2013.

[21] A. Cartea and M. G. Figueroa, "Pricing in electricity markets: A mean reverting jump diffusion model with seasonality," *Appl. Math. Finance*, vol. 12, no. 4, pp. 313–335, 2005.

[22] D. R. Jiang and W. B. Powell, "Optimal hour-ahead bidding in the real-time electricity market with battery storage using approximate dynamic programming," *INFORMS J. Comput.*, vol. 27, no. 3, pp. 525–543, 2015.

Bolong Cheng received the Ph.D. degree from Princeton University, where his research focused on stochastic optimization, optimal learning, sequential decision making, with applications in the energy systems and electricity market. He is currently a Research Engineer with SigOpt, a software start-up offering Bayesian optimization as a service.

Tsvetan Asamov was a Post-Doctoral Research Associate with the Department of Operations Research and Financial Engineering, Princeton University. His research focused on the development and implementation of stochastic optimization models for energy systems.

Warren B. Powell is a Professor with the Department of Operations Research and Financial Engineering, Princeton University, and the Director of CASTLE Laboratory and the Princeton Laboratory for Energy Systems Analysis. He has co-authored over 200 refereed publications in stochastic optimization, stochastic resource allocation, and related applications. He has authored the book entitled *Approximate Dynamic Programming: Solving the Curses of Dimensionality* and co-authored the book entitled *Optimal Learning* (Wiley). He is currently involved in applications in energy, transportation, finance, and healthcare.