This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2016.2605141, IEEE Transactions on Smart Grid

Co-optimizing Battery Storage for the Frequency Regulation and Energy Arbitrage Using Multi-Scale Dynamic Programming

Bolong Cheng, *Student Member, IEEE,* Warren B. Powell, *Member, IEEE,*

***Abstract*—We are interested in optimizing the use of battery storage for multiple applications, in particular energy arbitrage and frequency regulation. The nature of this problem requires the battery to make charging and discharging decisions at different time scales while accounting for the stochastic information such as load demand, electricity prices, and regulation signals. Solving the problem for even a single-day operation would be computationally intractable due to the large state space and the number of time steps. We propose a dynamic programming approach that takes advantage of the nested structure of the problem by solving smaller subproblems with reduced state spaces, over different time scales.**

***Index Terms*—Energy storage, frequency regulation, energy arbitrage, dynamic programming**

I. Introduction

THE increase in renewable energy sources such as wind and solar in recent years poses challenges to the robustness and resiliency of the electricity grid. Energy storage plays a significant role in meeting these challenges by improving the operation of the electricity grid while minimizing infrastructure investments. The earliest discussion of grid level storage can be traced back to [1], which presents storage in the context of a vertically integrated utility to mitigate peak generation through load shifting. A more recent report from EPRI provides a comprehensive overview of storage technologies and usages. Among many goals, it clearly addresses that energy storage of the future "should be recognized for its value in providing multiple benefits simultaneously" [2].

There is an abundance of research on the use of storage for the application of energy arbitrage, the buying and selling of electricity by exploiting the wholesale electricity price movement. Early studies such as [3] and [4] have outlined the economic benefits for energy storage through energy arbitrage. However, the results from these studies are inflated since they both assume the electricity prices are known before making storage decisions. [5] expands upon this work by relaxing the assumption of perfect price information. It uses a "back-casting" heuristic that assumes historical price and load patterns are repeated. A separate line of this research focuses on the interaction between storage and renewables. [6] found that compressed air energy storage is a better choice as a supplemental resource to wind generation in comparison to natural gas turbines when the green house gas emissions tax is high. [7] presents a wind energy commitment problem given storage and then analytically determines the optimal policy for the infinite horizon case. [8] studies the planning and operation of a wind energy storage system in an electricity market using forecasts of the prices. [9] develops a near-optimal policy for managing wind-generation with energy storage in the presence of finite transmission capacity and negative electricity prices.

In recent years, the market has come to recognize that energy arbitrage by itself is not enough to justify the investment cost of a battery. [5] 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. Furthermore, it suggests that device owners can increase the value of storage by co-optimizing between different markets such as frequency regulation and spinning reserves. [10] observes that frequency regulation offers higher profits than energy arbitrage. [4] and [11] also have shown frequency regulation can be a substantial revenue source for energy storage.

Frequency regulation is an important ancillary service for the maintenance of electricity grid stability. It helps mitigate the constant fluctuation in the supply and demand balance, usually caused by load variation or output variation from intermittent renewable resources, such as wind and solar [12]. Battery storage is an ideal technology for frequency regulation due to its almost instantaneous response time. There are many studies on integrating charging electric vehicles to provide frequency regulation services ([13], [14], and [15]). [16] optimizes the value of battery storage for frequency regulation. It models both the state of charge and voltage, reflecting the energy conversion capabilities of the battery. This line of literature only focuses on regulation service and does not consider multiple revenue streams. [10] analyzes the economics of using storage device for both energy arbitrage and frequency regulation service. The results showed "high probability of positive NPV (net present value)" in the New York City region for both energy arbitrage and regulation. However, this work does not have an optimization model and rather relies on heuristics derived from price duration patterns. [17] co-optimizes compressed air energy storage (CAES) for both energy and reserve market, but it does not fully account for the uncertain interactions between providing energy and ancillary service. Using a single storage device for multiple revenue streams requires jointly optimizing within

Bolong Cheng is with the Department of Electrical Engineering, Princeton University, Princeton, NJ, 08544.

Warren B. Powell is with the Department of Operations Research and Financial Engineering, Princeton University.

Manuscript received [DATE] 2015.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2016.2605141, IEEE Transactions on Smart Grid

fixed capacity and power constraints. [18] co-optimizes energy storage for multiple applications such as energy, capacity, and back up services. The problem is formulated as a stochastic dynamic program that solves for an hourly optimal decision. In reality, frequency regulation requires decisions at the sub-minute level (typical every two or four seconds). In [18], the regulation capacity and signal dispatch is modeled only on an hourly aggregation using the dispatch-to-contract ratio. Co-optimizing across frequency regulation and other streams is important, because frequency regulation offers the highest revenue stream. At the same time, it is quite difficult because the time scales are so different.

Frequency regulation is typically an easy control problem at the aggregate level (if a difficult engineering challenge) because it requires simply following the frequency regulation signal from the grid operator. However, optimizing across frequency regulation and energy shifting means that the instruction to follow the frequency regulation signal has to be replaced with a decision of whether to follow the signal, and how closely. How closely the frequency regulation signal should be followed depends on the current price for following the signal (which varies over time), the penalty for not following it, the current price of electricity (the LMP), and the time of day.

Given these limitations of existing storage models, this paper describes a Markov decision process (MDP) model that co-optimizes for both frequency regulation and energy arbitrage. We model the operation of the battery down to the two-second increment. Furthermore, our model accounts for the stochastic electricity price and regulation signals, thus accurately reflecting the time dependent nature of the problem. As a result of the small time steps, we have to discretize the state variable at a fine level to capture the small changes in the level of storage. A textbook application of dynamic programming would require computing and storing 41,200 matrices (the value functions in the dynamic program), each of which requires 1.6 gigabytes of storage, for a total of 66 terabytes.

We considered using approximate dynamic programming [19], but extensive empirical research did not give us confidence that ADP could handle this demanding application [20]. Instead, we first divided the problem into three time scales (daily, hourly, and five minutes), which limited the state space to three dimensions. Even this decomposition required 60 terabytes to solve the three-dimensional matrices every 2 seconds. To overcome this hurdle, we used singular value decomposition that reduced the disk space required for the value functions by a factor of 100, with near optimal performance.

This paper is organized as follows. In Section II, we first provide an overview of the PJM frequency regulation market. We then formulate the problem as a Markov decision process. In Section III, we present a modified model that takes advantage of the nested structure of the problem. In Section IV, we discuss the challenges for computing and storing the value functions and our low rank approach for approximating the value functions. In Section V, we describe the benchmark experiments and discuss circumventing computation limitation using low rank matrix approximation. We compare our method to strategies currently employed in the industry. Section VI concludes the paper.

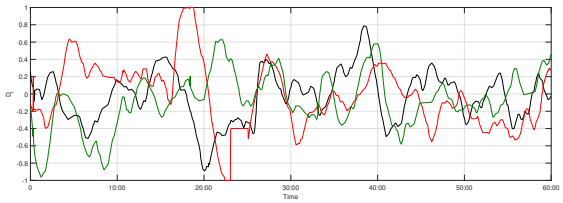


Fig. 1: Sample paths of the RegD signals over an hour

II. Battery Storage Model

In this section, we first outline the mechanism of the PJM regulation market. We explain the rules and operation of the market from which we derive our mathematical model. We then focus our attention on the control of the battery for co-optimizing frequency regulation and energy arbitrage, where we formulate the problem naturally as a Markov decision process (MDP).

A. Overview of the PJM Regulation Market

The PJM frequency regulation market provides participants, e.g. generators and various types of energy storage, with a market-based system to provide grid ancillary service in exchange for regulation credits. The PJM regulation market is a day-ahead market: resource owners wishing to sell regulation service must submit the offer by 6:00 p.m. the day prior to operation. An offer includes regulation capability, signal type, regulating hours, and various parameters of the resource. PJM clears the regulation market throughout the operating day and posts the results 30 minutes prior to the start of every operating hour. During the operating hour, PJM sends the regulation signal to the cleared resources. The RegD signal is the high pass filtered output of Area Control Error (ACE), which is a measure of the imbalance between sources and uses of power in the grid. It is specifically developed for energy storage with fast-response but limited energy capacities. The RegD signal is sent to the resource every two seconds. Fig. 1 shows three sample paths of the RegD signal over the period of an hour. PJM tracks the response of the resource and computes a performance score at the end of each service hour. The amount of regulation credit settlement received by the resource is based on this performance score. We discuss PJM's formulation of the performance score in detail in section II-F. PJM Manual 11 provides a complete overview of the market rules and operations [21].

B. Static Parameters

We take the perspective of the battery operator and model the storage problem over a finite-time horizon of one day. For our problem, we only focus on the control of the battery throughout the operating day; we assume that it has already cleared the bidding process. We also assume that the battery

is a price-taker in the energy market. The following is a list of parameters used to characterize the device:

- $R^{\max}$: The energy capacity of the battery in MWh.
- η^c, η^d: The charging and discharging efficiency of the device, respectively.
- β: The charging and discharging power capacity of the device.
- K: The assigned regulation (AReg) capacity in MW during period $[0, T]$. This is the maximum capacity assigned by PJM.

In the following subsections, we outline the five fundamental components of our model.

C. State Variables

Let $\mathcal{H} = \{0, 1, \ldots, 23\}$ be the set denoting the hours within a day and let $\mathcal{T} = \{0, \Delta t, 2\Delta t, \ldots, 1800\Delta t\}$ index the increments of two seconds (Δt) within an hour. We can index the entire day with the cross product of $(h, t) \in \mathcal{H} \times \mathcal{T}$. Naturally, the inter-hour transition follows $(h, 1800\Delta t + \Delta t) = (h+1, 0)$. Since the model is the same for all hours, we fix the operating hour h and describe the model for one hour. We temporarily drop h from the subscript for brevity.

We define $S_t = (R_t, D_t, P_t^E, G_t)$ as the state of the system at time t. R_t is the amount of energy in the battery at time t in MWh. D_t represents the RegD signal at time t in MW, which changes every two seconds. A positive D_t signal requires the battery to discharge and a negative value asks the battery to charge. P_t^E is the spot energy market price (LMP) at time t in \$/MWh, which is updated every five minutes. G_t models the performance score at time t. This is a measurement of how well the battery follows the regulation signal. PJM evaluates the score at the end of each hour.

We also define P^D as the hourly regulation market clearing price for 1MW regulation capacity. The PJM regulation market clearing price has two components: capacity and performance; we use the sum of the two components as P^D here to simplify the notation. Since P^D is constant for the entire hour and is only used at the end of the hour, we remove it from the state variable. It can be considered as a latent variable of the problem. In the following sections, variables with the superscript D are related to the frequency regulation market; those with the superscript E, the energy market.

D. The Decisions

At every time step t, our decision is given by the vector $x_t = (x_t^D, x_t^E)$. The x_t^D component is our response (in MW) to the regulation signal at time t, which changes every two seconds. The x_t^E component is the economic basepoint of the device at time t. It is the base charge/discharge rate of the battery. The regulation signal is modulated around the basepoint. PJM allows the resources to adjust its economic basepoint every five minutes. A positive value of $(x_t^D + x_t^E)$ means the battery is charging and a negative value means it is discharging. When a battery is perfectly following the RegD signal, $|D_t + x_t^D| = 0$. The economic basepoint is adjusted every five minutes to respond to the change in LMP. Alternatively, it can also be adjusted when the resource level is too high or too low from following the regulation signal.

At time t, we require that the total amount of energy stored in the device does not exceed its energy capacity:

$$0 \leq R_t + (x_t^D + x_t^E) \cdot \Delta t \leq R^{\max}. \tag{1}$$

The total amount of energy charged to or discharged from the battery is bounded by the maximum charging and discharging power capacity:

$$|x_t^D + x_t^E| \leq \beta. \tag{2}$$

Lastly, we require the economic basepoint to stay constant for the duration of the five minute interval.

$$x_t^E = x_{t-\Delta t}^E, \text{ if } (t \bmod 150\Delta t) \neq 0. \tag{3}$$

E. The Exogenous Information Process

The variable $W_t = (\hat{D}_t, \hat{P}_t^E)$ is a vector that contains exogenous information processes. $\hat{D}_t$ is the change in the RegD signal between times $t - \Delta t$ and t, which changes every two seconds. $\hat{P}_t^E$ is the change in the LMP between times $t - \Delta t$ and t. Note that the real-time LMP is updated every five minutes.

F. The Transition Function

We let $S_{t+\Delta t} = S^M(S_t, x_t, W_{t+\Delta t})$ be the mapping from a state S_t to the next state $S_{t+\Delta t}$, given the decision x_t and new information $W_{t+\Delta t}$. The state variables D_t and P_t^E evolve randomly according to the following transition functions:

$$D_{t+\Delta t} = D_t + \hat{D}_{t+\Delta t}, \tag{4}$$

$$P_{t+\Delta t}^E = P_t^E + \hat{P}_{t+\Delta t}^E. \tag{5}$$

We discuss the modeling of these two processes in detail in later sections. The transition function for the energy stored in the device is given by:

$$R_{t+\Delta t} = R_t + (x_t^D + x_t^E) \Big(\mathbf{1}_{\{x_t^D + x_t^E < 0\}} + \mathbf{1}_{\{x_t^D + x_t^E > 0\}} \eta^c \Big) \cdot \Delta t, \tag{6}$$

where $\mathbf{1}_{\{\star\}}$ is the indicator function. Note that the discharge efficiency η^c reflects the loss in energy charged to the battery. Equation (6) holds for the transition from $R_{h,1800\Delta t}$ to $R_{h+1,0}$ across the boundaries at the end of each hour.

We next describe the dynamics of the PJM performance score. The performance score is a weighted sum of three components: correlation, delay, and precision. It is computed at the end of the regulation hour. The correlation and delay components measure the temporal shift of the signal and response. They are computed together using five-minute rolling correlations of the signal D_t and the response x_t^D. The precision score measures the signal/response deviation. It is computed by normalizing the hourly absolute deviation between signal and response at every measurement interval. The calculation used by PJM is unsuitable for our model since it requires storing a history of D_t and x_t.

We propose a simplified version of the performance score that can be computed recursively. A battery storage unit will

typically have correlation and delay scores close to 1 since it can ramp up or down instantaneously. Therefore we drop these two components and focus on the precision score. We denote G as the precision score of the hour and it is computed by PJM using the formula

$$G = \frac{\sum_{t=0}^{1800\Delta t} \min\left(\left(1 - \frac{|(x_t^D + D_t)\cdot \mathbf{1}_{\{t \bmod 5\Delta t = 0\}}|}{K}\right)^+, 1\right)}{360}. \tag{7}$$

This can be viewed as an average of the normalized signal/response deviation, taken at 10 second intervals ($5\Delta t$). If we want to update this formula iteratively, we first have to capture snapshots at every interval Δt. We reformulate the transition function for G_t as

$$G_{t+\Delta t} = G_t - \frac{\min\left(\frac{\left|x_t^D(\eta^d \mathbf{1}_{\{x_t^D<0\}} + \mathbf{1}_{\{x_t^D>0\}}) + D_t\right|}{K}, 1\right)}{1800}, \tag{8}$$

where $G_0 \equiv 1$. Essentially, we are tracking the degradation of G_t every time step. G_t is non-increasing in t in this formulation. Moreover, $G_{1800\Delta t} = G$ if we change the PJM sampling interval from equation (7) to 2 seconds.

G. The Revenue Function

The function $C(S_t, x_t)$ represents the from being in the state S_t and making the decision x_t at time t. We denote $T = 1800\Delta t$ as the end of the hour. The rewards earned during the hour can be characterized by

$$C(S_t, x_t) = -P_t^E(x_t^D + x_t^E)\left(\eta^d \mathbf{1}_{\{x_t^D + x_t^E < 0\}} + \mathbf{1}_{\{x_t^D + x_t^E > 0\}}\right)\cdot \Delta t, \forall t < T, \tag{9}$$

$$C_T = KP^D G_T \cdot \mathbf{1}_{\{G_T \geq 0.4\}}. \tag{10}$$

Equation (9) indicates the payment in the energy market for charging/discharging in the grid. The discharge efficiency η^d reflects the loss of energy discharged from the battery. The hourly settlement that we receive for providing frequency regulation service is described by (10). A resource with a performance score lower than 0.4 will not receive the regulation credit and can be disqualified from the regulation market.

We add back the hour-index h and let $X_{h,t}^{\pi}(S_{h,t})$ be a policy that outputs a decision $x_{h,t}$ given state $S_{h,t}$. The objective function for the horizon of 24 hours can be written as

$$F_0^* = \max_{\pi \in \Pi} \mathbf{E}\left[\sum_{h=0}^{23}\sum_{t=0}^{T} C(S_{h,t}, X_{h,t}^{\pi}(S_{h,t}))\middle| S_{0,0}\right], \tag{11}$$

where Π is the space of all admissible policies. Let $V_{h,t}^*(S_{h,t})$ be the optimal value function for a pre-decision state $S_{h,t}$. The optimal policy is characterized by Bellman's optimality equation, which is given by

$$V_{h,t}^*(S_{h,t}) = \max_{x_{h,t} \in \mathcal{X}_{h,t}} \{C(S_{h,t}, x_{h,t}) + \mathbf{E}[V_{h,t+\Delta t}^*(S_{h,t+\Delta t})|S_{h,t}]\}, \forall (h,t), \tag{12}$$

where the terminal value $V_{23,T} = 0$. Note that the hourly RegD settlement in (10) is embedded in the optimality equation. The conditional expectation is taken over the random variable $W_{h,t+\Delta t}$ and the optimal decision is given by the arg max. We can obtain the optimal policy by computing (12) traversing backward through time. For our time horizon of one day, this requires $24 \times 1800 \times |\mathcal{S}_t|$ computations. Our experiment has over 70 million states per time t; solving this problem directly becomes computationally intractable.

We notice that the two decisions happen at two different time scales: the economic basepoint is set every five minutes and the RegD response happens every two seconds. This allows us to decouple the two decisions, and formulate a nested model of the problem. We describe this new model in the following section.

III. A Nested Dynamic Programming Model

We want to formulate the problem as a nested dynamic program. In particular, we have to model the control problem on three different levels. We need to compute the value of storage for the entire day in five minute increments. Within each hour, we want to find the optimal economic basepoint every five minutes. Lastly, we need to know the best response decision every two seconds. Now we outline the mathematical model of each subproblem.

A. The Hourly Resource Model

First, we need to address the problem of representing the value of energy stored at the end of each hour. We formulate an MDP as a simplified version of the storage problem presented in [22]. We denote $\tau \triangleq 150\Delta t$, the equivalent of a five-minute interval and let $\mathcal{T}^{EB} = \{0, \tau, 2\tau, \ldots, 12\tau\}$ be the set marking the increments of five-minutes within an hour. For this problem, we let the time index be $(h,t) \in \mathcal{H} \times \mathcal{T}^{EB}$ for the horizon of 24 hours in five-minute increments.

In the resource model, the state variable consists of $S_{h,t}^R = (R_{h,t}, P_{h,t}^E)$, where the battery only responds to the LMP. The decision $x_{h,t}^R$ is the amount of energy to charge or discharge every five minutes. It must satisfy the energy capacity and power capacity constraints of the battery:

$$0 \leq R_{h,t} + x_{h,t}^R \leq R^{\max}, \tag{13}$$

$$|x_{h,t}^R| \leq \beta \cdot 150\Delta t. \tag{14}$$

The transition functions for the state variables are as follows

$$R_{h,t+\tau} = R_{h,t} + x_{h,t}^R(\eta^c \mathbf{1}_{\{x_{h,t}^R > 0\}} + \mathbf{1}_{\{x_{h,t}^R < 0\}}), \tag{15}$$

$$P_{h,t+\tau}^E = P_{h,t}^E + \hat{P}_{h,t+\tau}^E, \tag{16}$$

where $\hat{P}_{t+\tau}^E$ is the only exogenous variable of the model. The revenue function simply computes the revenue from discharging/charging, i.e.

$$C^R(S_{h,t}^R, x_{h,t}^R) = -P_{h,t}^E x_{h,t}^R\left(\mathbf{1}_{\{x_{h,t}^R > 0\}} + \eta^d \mathbf{1}_{\{x_{h,t}^R < 0\}}\right). \tag{17}$$

The objective function is defined as

$$\max_{\pi^R \in \Pi^R} \mathbf{E}\left[\sum_{h=0}^{23}\sum_{t \in \mathcal{T}^{EB}} C^R(S_{h,t}^R, X_{h,t}^{\pi^R}(S_{h,t}^R))|S_{0,0}^R\right], \tag{18}$$

where $X^{\pi^R}_{h,t}(S^R_{h,t})$ is a policy that maps the state $S^R_{h,t}$ to the decision $x^R_{h,t}$ and the space of all admissible policies for this subproblem is defined by Π^R.

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

In the EB model, we are interested in setting the economic basepoint for energy arbitrage. We assume the decision x^D_t follows a certain policy π^{FR}. Since the horizon of this subproblem is one hour, we fixed the hour-index h and temporarily drop it from the subscript. The time index $t \in \mathcal{T}^{EB}$ denotes the increments of five minutes.

We need to modify the system model according to the five-minute dynamics. Within every five minute increment, the change in the storage due to following the regulation signal can be represented as a random variable. We augment the state space with two new exogenous variables $\hat{R}^+_t$ and $\hat{R}^-_t$. We define $\hat{R}^+_{t+\tau}$ as 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 then becomes

$$R_{t+\tau} = R_t + x^E_t \cdot \tau(\mathbf{1}_{\{x^E_t<0\}} + \mathbf{1}_{\{x^E_t>0\}}\eta^c) + \eta^c \hat{R}^+_{t+\tau} + \hat{R}^-_{t+\tau}. \quad (19)$$

The LMP process P^E_t remains the same as in (16). The RegD signal D_t is omitted from the state variable for this subproblem since we only model frequency regulation on an aggregated level. The performance score G_t now becomes a random process that depends on the policy π^{FR}. We modify the transition accordingly, where

$$G_{t+\tau} = G_t + \hat{G}_{t+\tau}. \quad (20)$$

In this subproblem, the economic basepoint x^E_t is the only intrinsic decision that we have to make; however, we need to add a decision to represent the trade-off between frequency regulation and energy arbitrage. Let $x^G_t \in [0,1]$ denote the limit of the degradation in G_t from t to $t+\tau$, where $x^G_t = 0$ represents that we must strictly follow the signal from within the five minute interval and $x^G_t = 1$ allows us to disobey the signal completely. We thus derive a new set of constraints for this subproblem:

$$0 \le R_t + \Big((1-x^G_t)K(\mathbf{1}_{\{x^E_t>0\}} - \mathbf{1}_{\{x^E_t<0\}}) + x^E_t\Big)\cdot\tau \le R^{\max} \quad (21)$$

$$\left|x^E_t + (1-x^G_t)K(\mathbf{1}_{\{x^E_t>0\}} - \mathbf{1}_{\{x^E_t<0\}})\right| \le \beta \quad (22)$$

$$0 \le x^G_t \le 1 \quad (23)$$

Equation (21) enforces the battery to satisfy the energy capacity limit, even when the regulation signal requires charing/discharging at the full rate for the entire five minutes. Equation (22) guarantees the battery never exceeds the power capacity even in the worst case. In summary, the state variable for the five minute problem is $S^{EB}_t = (R_t, G_t, P^E_t)$ and the exogenous variable that becomes available at time t is $W_t = (\hat{R}^+_t, \hat{R}^-_t, \hat{G}_t, \hat{P}^E_t)$. The state transition function

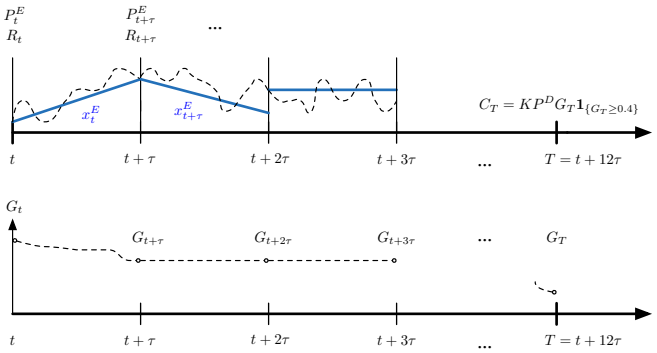


Fig. 2: Model of the economic basepoint problem. In the first figure, the slope of the blue lines can be viewed as the economic basepoint, and the dotted line represents the actual supply/demand due to following regulation signal. In the second figure, the vertical axis is the performance score G_t. It decreases over the first time period but remains constant in the next two.

$S^{EB}_{t+\tau} = S^M(S^{EB}_t, x_t, W_{t+\tau})$ is characterized by equations (19), (20), and (5) respectively. We illustrate the economic basepoint model in Fig. 2. Finally we modify the revenue function to fit into the new time scale.

$$C^{EB}(S^{EB}_t, x_t, W_{t+\tau}) = -P^E_t\left(x^E_t(\eta^d \mathbf{1}_{\{x^E_t<0\}} + \mathbf{1}_{\{x^E_t>0\}})\cdot\tau + \hat{R}^+_{t+\tau}/\eta^c + \hat{R}^-_{t+\tau}\right), \quad \forall t < T. \quad (24)$$

The contribution from hourly regulation credit (10) remains unchanged for T. Our goal is to find the optimal policy π^{EB} defined by the objective function:

$$\max_{\pi^{EB}\in\Pi^{EB}} \mathbf{E}^{\pi^{FR}}\left[\sum_{t\in\mathcal{T}^{EB}} C^{EB}(S^{EB}_t, X^{\pi^{EB}}_t(S^{EB}_t), W_{t+\tau})|S^{EB}_0\right]. \quad (25)$$

The superscript π^{FR} over the expectation implies that the frequency regulation policy influences the underlying stochastic processes of this subproblem. This subproblem needs to be computed for every hour $h \in \mathcal{H}$.

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

Now we turn our attention to the frequency regulation problem at the two-second time scale with a time horizon of five minutes. Since the LMP is constant over the five minute interval, P^E is treated as a *latent variable* for the subproblem, i.e. the value function is implicitly a function of P^E. We are left with a three dimensional state $S^{FR}_t = (R_t, G_t, D_t)$. Our only decision for this subproblem is the response to the regulation signal, x^D_t. The transition functions for the three state variables remain the same from equations (6), (8), and (4) respectively. We already know the optimal economic basepoint x^E and the maximal single-period degradation x^G for the entire horizon. We need to add one more constraint, where

$$\min\left(\frac{1}{K}\left|x^D_t + D_t\right|, 1\right) \le x^G. \quad (26)$$

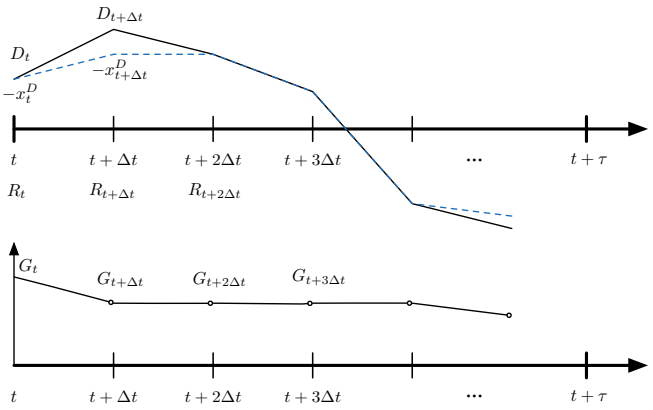


Fig. 3: Model of the frequency regulation problem. In the first figure, the solid line is D_t and the dotted line is $-x^E_t$. When we deviate from the signal, we can see a decrease in G_t as shown in the second figure.

Combining with (23), we can rewrite the above expression as

$$\frac{1}{K}\left|x^D_t + D_t\right| \leq x^G. \tag{27}$$

For this subproblem, the set of constraints is described by (1), (2), and (27). We illustrate the FR model in Fig. 3. The revenue function follows (9). The objective function for finding the optimal policy π^{FR} is written as,

$$\max_{\pi^{FR}\in\Pi^{FR}} \mathbf{E}\left[\sum_{t=t'}^{t'+\tau} C^{FR}(S^{FR}_t, X^{\pi^{FR}}_t(S^{FR}_t))|S^{FR}_{t'}, x^E, P^E\right], \tag{28}$$

given the economic basepoint x^E and the LMP P^E for the five minute interval.

IV. Computing the Value Functions

The nested model offers a more tractable solution to optimize the battery over the horizon of one day. In this section, we describe our algorithm for computing the value functions. We also discuss additional computational challenges resulting from discretization of the state space and our solution.

A. Algorithmic Approach

We have described three MDP's that operate at different time scales; we can solve them individually by computing the Bellman equations recursively backward through time. A standard algorithm can be found in all standard textbooks on the subject such as [23]. We focus our attention on how to link the three levels of MDP's together in order to find the optimal decisions. Our approach is outlined in Algorithm 1. We denote SolveR in Step 2 as the MDP model for the problem described in Section III-A. The optimal value function is defined recursively by Bellman's equation:

$$V^R_{h,t}(S^R_{h,t}) = \max_{x_{h,t}\in\mathcal{X}^R_{h,t}} \{C^R(S^R_{h,t}, x_{h,t}) + \mathbf{E}[V^R_{h,t+\tau}(S^R_{h,t+\tau})|S^R_{h,t}]\}, \forall (h,t) \in \mathcal{H}\times\mathcal{T}^{EB} \tag{29}$$

Algorithm 1 Algorithm for Solving the Value Functions

Step 1 Initialize $V_{24,12\tau} = 0$.

Step 2 Compute SolveR, obtain $V^R_{h,t}(S^R_{h,t})$, for $(h,t) \in \mathcal{H}\times\mathcal{T}^{EB}$.

Step 3 For $h = 0,\ldots,23$

- 3a Initialize $V^{EB}_{h,T}(S^{EB}_{h,T}) = V^R_{h+1,0}(S^R_{h+1,0}) + KP^E G_T \mathbf{1}_{\{G_T\geq 0.4\}}$.
- 3b Compute SolveEB$_h$, obtain $V^{EB}_{h,t}(S^{EB}_{h,t})$, for $t \in \mathcal{T}^{EB}$.

Step 4 For $h = 0,\ldots,23$, for $n = 0,1,\ldots,11$.

- 4a For all $P^E_{h,t}$ and $x^E_{h,t}$, initialize terminal value function $V^{FR}_{h,(n+1)\tau}(S^{FR}_{h,(n+1)\tau}) = V^{EB}_{h,(n+1)\tau}(S^{EB}_{h,(n+1)\tau})$.
- 4b Compute all SolveFR$_{h,t}(P^E_{h,t}, x^E_{h,t})$.

where $V_{24,12\tau} = 0$. This step gives us the value of storage for the entire day and we can move on to the next level. In step 3, we denote SolveEB$_h$ as the MDP model for the EB problem. Similarly, for each hour $h \in \mathcal{H}$ we compute the optimal five-minute value function via

$$V^{EB}_{h,t}(S^{EB}_{h,t}) = \max_{x_{h,t}\in\mathcal{X}_{h,t}} \{C^{EB}(S^{EB}_{h,t}, x_{h,t}, W_{h,t+\tau}) + \mathbf{E}[V^{EB}_{h,t+\tau}(S^{EB}_{h,t+\tau})|S^{EB}_{h,t}]\}, \forall t \in \mathcal{T}^{EB}, \forall h \in \mathcal{H}, \tag{30}$$

where $V^{EB}_{h,T}(S^{EB}_{h,T}) = V^R_{h+1,0}(S^R_{h+1,0}) + KP^E G_T \mathbf{1}_{\{G_{h,T}\geq 0.4\}}$. The expectation is taken over the exogenous variable $W_{h,t+\tau} = (\hat{R}^+_{h.t+\tau}, \hat{R}^-_{h,t+\tau}, \hat{G}_{h,t+\tau}, \hat{P}^E_{h,t+\tau})$. The terminal value functions for the EB problem is the sum of the hourly value function of the resource problem and the hourly settlement from the frequency regulation market. We have now computed the value functions down to five-minute increments for the entire day. We have also obtained the optimal economic basepoint decision. Lastly, we let SolveFR$_{h,t}(P^E, x^E)$ be the model for the MDP outlined in Section III-C. Recall that this is a function of the LMP variable P^E and the economic basepoint x^E implicitly; therefore, we have to solve $24\times 12\times|P^E|\times|x^E|$ subproblems. Now we can derive the optimal response decision x^D_t by solving

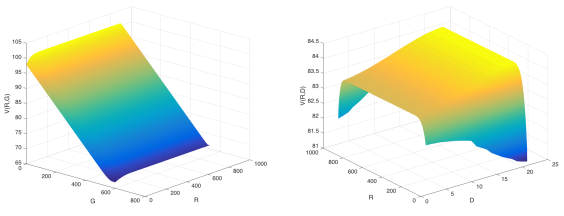
$$V^{FR}_{h,t}(S^{FR}_{h,t}) = \max_{x_{h,t}\in\mathcal{X}^{FR}_{h,t}} \{C^{FR}(S^{FR}_{h,t}, x_{h,t}) + \mathbf{E}[V^{FR}_{h,t+\Delta t}(S^{FR}_{h,t+\Delta t})|S^{FR}_{h,t}]\}, \tag{31}$$

where the terminal value functions are $V^{FR}_{h,t'} = V^{EB}_{h,t'}(S^{EB}_{h,t'})$, for $t' \in \mathcal{T}^{EB}$ and $h \in \mathcal{H}$. The expectation is taken with respect to the only exogenous variable $\hat{D}_{t+\Delta t}$.

B. Computational Challenges and Solution

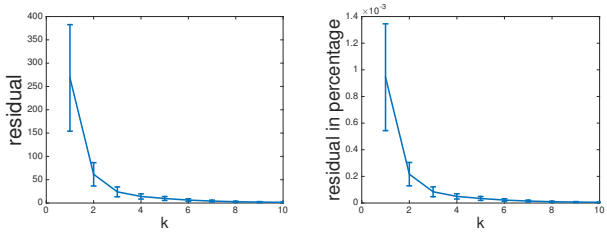
We discretize the state space and action space in order to solve the problem. In order to track the movement of resource R and the performance score G at two-second increments, we need to discretize the state very finely. We find a good discretization of the state space to be $|R|\times|G|\times|D| = 901\times 601\times 21$ for the FR problem. Assuming each element of $V(S^{FR})$ is stored as a single-precision floating point number,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2016.2605141, IEEE Transactions on Smart Grid



(a) $V(R,G)$ for a fixed D (b) $V(R,D)$ for a fixed G

Fig. 4: Sample value function plots for the FR problem



(a) Residual for the first 10 singular values (b) Residual in percentage for the first 10 singular values

Fig. 5: Residual from low rank approximation using the first 10 singular values

we need 4 bytes/state $\times$ $|S| \times 150 \approx 6$ GB to store the value functions from SolveFR for one combination of P^E and x^E. Suppose that we discretize the prices P^E to 7 levels and basepoints x^E to 5 levels, we need at least 60 TB of disk space to store the value functions for all possible states for the time horizon of one day.

Although we can compute the value functions through parallelization, we do not have enough disk space to store the value functions for even one single scenario of RegD clearing price P^D. We observe that the value functions form smooth surfaces in all three dimensions, as shown in Fig. 4. We can represent the value function V as a matrix with $|R|$ rows and the cross product $|G \times D|$ columns. Fig. 4 indicates that the difference between adjacent columns of V is close to constant for any fixed row R, hinting that V is not a full rank matrix and can be approximated with a low-rank matrix $\tilde{V}$ using singular value decomposition. SVD factors the matrix V into the following form:

$$V = U\Sigma Q^T,$$

where U is a $|R| \times |R|$ unitary matrix, Σ is a $|R|$ by $|G \times D|$ diagonal matrix, and Q^T is a $|G \times D|$ by $|G \times D|$ unitary matrix. The diagonal entries of Σ are singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{|R|} > 0$.

SVD guarantees the best rank k approximation under the Frobenius norm, i.e. it is the solution to $\min_{\tilde{V}} ||V - \tilde{V}||_F$ subject to $\text{rank}(\tilde{V}) \leq k$ [24]. Let $\tilde{V}_k$ be the rank k approximation of V obtained from SVD. We plot the residual $||V - \tilde{V}_k||_F$ in Fig. 5a, and the normalized residual $||V - \tilde{V}_k||_F / ||V||_F$ in Fig. 5b for the first 10 singular values. As we can see, the error quickly diminishes after rank 5, with the relative error lower than 0.02%. Choosing a rank of 10 will only require us to store the first 10 columns of U and the first 10 rows Q^T, taking only 73 MB of disk space comparing to 6 GB without changing discretization, a factor of almost 100 reduction in storage requirement. Now we can finally compute and store all the value functions. In the next section, we further explore the empirical trade-off between rank k and optimality in greater details.

V. Numerical Results

In this section, we present the results of our algorithm by comparing against the frequency regulation policy used in the industry, which is a greedy policy that only maximizes the FR revenue. We also evaluate the quality of the solution when limiting the SVD approximation to different ranks.

To accelerate the computational testing, we use a time step Δt in the frequency regulation problem of 10 seconds, which reduces CPU time and storage by a factor of 5. This choice does not change the policies; in addition we kept the state space the same (that is, we used the same discretization of all the state variables), which reduced rounding and truncation error.

A. Benchmark Problems

First, we consider the case of a battery of 500 KWh capacity with a power capacity of $\beta = 1$MW and a roundtrip efficiency of $0.9 \times 0.9 = 0.81$. We assume the battery has cleared for the FR market for the entire day, with $K = 1$MW regulation capacity. Over-bidding the actual battery capacity is typically done in practice to take advantage of the dynamic nature of the RegD signal.

We train our LMP data from 60 sample paths of historical prices from Jan. 13, 2013 to Mar. 12, 2013. This data is used to estimate the value functions which produces a policy that is then tested on 10 sample paths selected from the same period. We also use the historical RegD signals from the same days. To solve the EB problems, we need to model the other three exogenous variables. We model $\hat{R}^+_{t+\tau}$ and $\hat{R}^-_{t+\tau}$ using the empirical distribution of the positive and negative RegD signal aggregated at five minute increments. $\hat{G}_t$ is a discrete uniform random variable on the finite support $[G_t - x^G_t/12, G_t]$. In the FR problems, we modeled our D_t process using a bounded first-order Markov chain, where we assume $\hat{D}_{t+\tau}$ from equation (4) is a random variable with discrete pseudonormal distribution, as described in [22]. Furthermore, we assume the regulation market clearing price (P^D) is constant for the entire operating horizon. This assumption allows us to test the behavior of the policy for different values of the RegD price. For example, we expect to see closer compliance to the RegD signal as the RegD price increases.

For the pure-FR policy, the battery maintains the economic basepoint at 0 and strictly follows the D_t signal unless it violates one of the physical constraints described in sec. II-D. In this case, the policy is defined by $\arg\min_{x^D_t} |x^D_t + D_t|$. We present our simulation results for the two policies in Table I. The co-optimization policy outperforms the greedy

TABLE I: Comparing revenues between co-optimization and pure frequency regulation

RegD price ($/MW)	co-opt ($/day)	pure-FR ($/day)	absolute improvement	relative improvement
5	180.35	100.61	79.74	79.26%
10	289.29	219.22	70.07	31.96%
20	512.18	456.42	55.76	12.22%
40	978.21	930.82	47.39	5.09%
100	2395.28	2354.03	41.25	1.75%

pure-FR policy in all price settings. The greatest proportional increase in revenue occurs when the RegD clearing price is low ranging from an 80% increase when $P^D = 5$ to a 1.75% increase when $P^D = 100$. Keep in mind that the FR revenue is much higher when the RegD price is high. In absolute terms, this translates to roughly $187,000 in yearly revenue for co-optimization when $P^D = 20$ and $357,000 when $P^D = 40$. We want to emphasize that this is not an economic assessment of co-optimization, but rather a benchmark to validate our model and algorithm. For a fair economic assessment, we need to test on separate price data and consider correlations between RegD price and the LMP (thus discarding our constant P^D assumption) in addition to other factors, which we will consider in future studies.

We note that the co-optimization policy produces increased revenue at all price levels. While this is to be expected, it is important to realize that when the RegD price is $100, the behavior almost exactly follows a pure FR policy, but using an algorithmic strategy that is dramatically different. It is our judgment that this would be very difficult to achieve with heuristic policies.

As we have assumed, when P^D is sufficiently higher than the LMP, the battery sets the economic basepoint at 0 most of the time and strictly follows the D_t signal. In contrast, when P^D is low comparing to the LMP, the battery adjusts the economic basepoint more frequently and emphasizes on energy arbitrage. In Fig. 6, the $P^D = 100$ sample path (in blue solid) mainly discharges during the price spikes and strictly follows the RegD signal with economic basepoint set at 0 the rest of the time. The $P^D = 20$ sample path (in red dotted) adjusts the economic basepoint more frequently to take advantage of every price spike. Note that the battery buys back most of the resource immediately after the first high LMP period in anticipation of later price spikes. For the $P^D = 100$ scenario, the battery is recharged following the discharge due to the price spike, but the recharging occurs more slowly

In Fig. 7, we also observe that the co-optimization policy automatically adjusts the resource level, selling when it is at full capacity (exhibited at around tick 7500). As a result, the battery can follow the regulation signal closer without hitting the boundaries, evidenced by the higher overall performance scores.

B. Low Rank Approximation and Optimality

Next, we experiment with policies computed using different assumptions for the rank in the singular value decomposition.

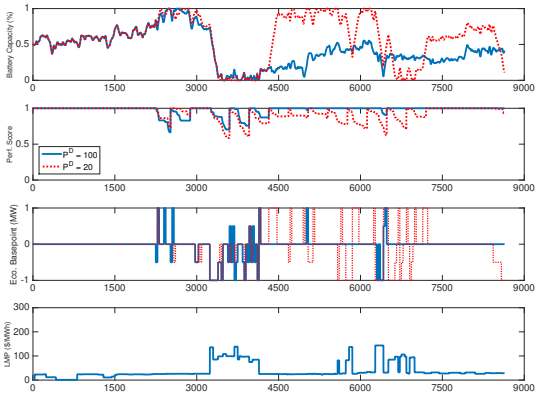


Fig. 6: A sample path for co-optimization comparing RegD clearing prices $P^D = 20$/MW (red dotted) and $P^D = 100$/MW (blue solid). The first plot shows the resource level of the battery. The second plot displays the performance scores. The economic point is shown in the third plot: a positive value means the basepoint has a charging bias; a negative value, discharging bias. The last plot shows the LMPs for the sample path.

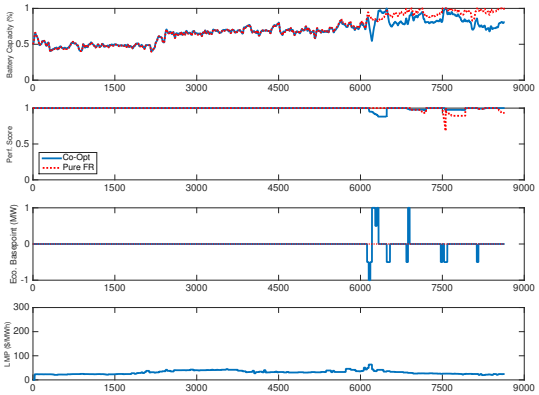


Fig. 7: A sample path comparing co-optimization (blue) and pure FR (red) policies for RegD clearing prices $P^D = 100$/MW.

,

We use the same experimental settings from the first benchmark problem but vary the rank k over the range 1, 2, 5 and 10. We present the results in Fig. 8. As the rank k decreases, the policies produce less optimal decisions. When we reduce the rank from $k = 10$ to $k = 5$, the decrease in revenue ranges from 0.2% (when $P^D = 100$) to 1.0% (when $P^D = 10$). Note that $k = 5$ decreases the disk space usage by another factor of 2, but still requires the same amount of computation time. For the policies computed using rank-1 approximation, the revenue decreases by as much as 10% for the $P^D = 5$ setting. However, the decrease in revenue is not significant (around 1%) for cases where P^D is greater than $20/MWh. Lastly, we observe that even when using rank-1 approximation, the

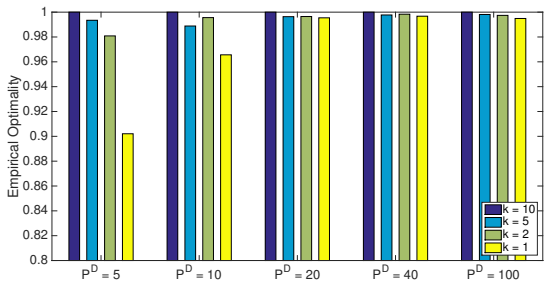


Fig. 8: Comparing the optimality of policies computed using different rank k approximation. The results are compared to the $k = 10$ revenue.

,

co-optimization model still outperforms the pure FR policy.

VI. Conclusion

This paper introduces a nested MDP model that co-optimizes a storage device for energy arbitrage and frequency regulation down to two-second increments, which is the frequency of the regulation signal. This model produces an optimal policy for controlling the charging response for frequency regulation and setting the appropriate economic basepoint for energy arbitrage, taking into account the stochastic LMP prices and the RegD signal for the entire day. In order to store the value functions for all the states, we implemented a low-rank approximation using singular value decomposition.

We implemented our model and experimented on historical LMP and regulation signal data. We make the assumption that the RegD clearing price is constant during the operating day, allowing us to test the effect of the RegD price on the behavior of the policy. Our policy outperforms the pure-FR policy currently employed in the industry, especially when the RegD clearing price is low. We also experimented with different rank k in the SVD approximation. While the revenue decreases when we lower the rank k, our co-optimization method using a rank-1 approximation still outperforms the pure FR policy.

We close by emphasizing that our experiments are intended only to evaluate the performance of the algorithm. We assert, for example, that producing small but positive profits from co-optimization when the RegD price is \$100 (the level at which the optimal policy is to closely follow the RegD signal) is a significant achievement, given that the mechanics of the optimized policy are so different. However, the cost improvements in Table I should be viewed only as a demonstration that the algorithm works. A careful analysis of the economic benefits of a co-optimized policy requires testing under more realistic conditions.

Acknowledgment

This research was funded in part by a grant from the Andlinger Center for Energy and the Environment, and the National Science Foundation, grant ECCS-1127975, and the SAP Initiative for Energy Systems Research.

References

[1] EPRI, "Assessment of energy storage systems suitable for use by electric utilities." Electric Power Research Institute, Tech. Rep., 1976.
[2] DOE, "Grid energy storage," Department of Energy, Tech. Rep., 2013.
[3] P. C. Butler, J. Iannucci, and J. Eyer, "Innovative business cases for energy storage in a restructured electricity marketplace," Sandia National Laboratories, Tech. Rep., 2003.
[4] J. M. Eyer, J. J. Iannucci, and G. P. Corey, "Energy storage benefits and market analysis handbook," Sandia National Laboratories, Tech. Rep., 2004.
[5] R. Sioshansi, P. Denholm, T. Jenkin, and J. Weiss, "Estimating the value of electricity storage in PJM: Arbitrage and some welfare effects," *Energy Economics*, vol. 31, no. 2, pp. 269 – 277, 2009.
[6] 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.
[7] J. H. Kim and W. B. Powell, "Optimal energy commitments with storage and intermittent supply," *Operations Research*, vol. 59, no. 6, pp. 1347–1360, 2011.
[8] M. Dicorato, G. Forte, M. Pisani, and M. Trovato, "Planning and operating combined wind-storage system in electricity market," *Sustainable Energy, IEEE Transactions on*, vol. 3, no. 2, pp. 209–217, April 2012.
[9] Y. Zhou, A. Scheller-Wolf, N. Secomandi, and S. Smith, "Managing wind-based electricity generation in the presence of storage and transmission capacity," Tepper School of Business, Paper 1477, 2013.
[10] 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.
[11] D. Perekhodstev, "Two essays on problems of deregulated electricity markets," Ph.D. dissertation, The Tepper School of Business at Carnegie Mellon University, 2004.
[12] B. J. Kirby, "Frequency regulation basics and trends," Oak Ridge National Laboratory, Tech. Rep., 2004.
[13] J. Donadee and M. Ilic, "Stochastic co-optimization of charging and frequency regulation by electric vehicles," in *North American Power Symposium (NAPS), 2012*, Sept 2012, pp. 1–6.
[14] J. Xu and V. Wong, "An approximate dynamic programming approach for coordinated charging control at vehicle-to-grid aggregator," in *Smart Grid Communications (SmartGridComm), 2011 IEEE International Conference on*, Oct 2011, pp. 279–284.
[15] J. Tomic and W. Kempton, "Using fleets of electric-drive vehicles for grid support," *Journal of Power Sources*, vol. 168, no. 2, pp. 459 – 468, 2007.
[16] R. L. Fares, J. P. Meyers, and M. E. Webber, "A dynamic model-based estimate of the value of a vanadium redox flow battery for frequency regulation in texas," *Applied Energy*, vol. 113, pp. 189 – 198, 2014.
[17] E. Drury, P. Denholm, and R. Sioshansi, "The value of compressed air energy storage in energy and reserve markets," National Renewable Energy Laboratory, Tech. Rep., 2011.
[18] X. Xi, R. Sioshansi, and V. Marano, "A stochastic dynamic programming model for co-optimization of distributed energy storage," *Energy Systems*, vol. 5, no. 3, pp. 475–505, 2014.
[19] W. B. Powell, *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. John Wiley & Sons, Inc., 2011.
[20] D. Jiang, T. Pham, W. Powell, D. Salas, and W. Scott, "A comparison of approximate dynamic programming techniques on benchmark energy storage problems: Does anything work?" in *Adaptive Dynamic Programming and Reinforcement Learning (ADPRL), 2014 IEEE Symposium on*, Dec 2014, pp. 1–8.
[21] *PJM Manual 11: Energy & Ancillary Services Market Operations*, PJM, 2015.
[22] D. F. Salas and W. B. Powell, "Benchmarking a scalable approximate dynamic programming algorithm for stochastic control of multidimensional energy storage problems," Technical report, Working Paper, Department of Operations Research and Financial Engineering, Princeton, NJ, Tech. Rep., 2013.
[23] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New York: John Wiley & Sons, Inc., 1994.
[24] G. W. Stewart, "On the early history of the singular value decomposition," *SIAM Review*, vol. 35, no. 4, pp. 551–566, 1993.

Bolong Cheng did his Ph.D. research at Princeton University, where his

research focused on stochastic optimization, optimal learning, sequential decision making, with applications in the energy systems and electricity market.

Warren B. Powell is a Professor in the Department of Operations Research and Financial Engineering at Princeton University, and the director of CASTLE Laboratory (http://www.castlelab.princeton.edu) and the Princeton Laboratory for Energy Systems Analysis (http://energysystems.princeton.edu). He has coauthored over 200 refereed publications in stochastic optimization, stochastic resource allocation and related applications. He is the author of the book Approximate Dynamic Programming: Solving the Curses of Dimensionality and a co-author of Optimal Learning, published by John Wiley & Sons. Currently, he is involved in applications in energy, transportation, finance and healthcare.