Submodularity of Energy Storage Placement in Power Networks

Junjie Qin Insoon Yang Ram Rajagopal

Abstract— **This paper studies the problem of optimally placing energy storage devices in power networks. We explicitly model capital and installation costs of storage devices because these fixed costs account for the largest cost component in most grid-scale storage projects. Finding an optimal placement strategy is a challenging task due to (i) the discrete nature of such placement problems, and (ii) the spatial and temporal transfer of energy via transmission lines and distributed energy storage resources. To develop an efficient placement framework with performance guarantees, we investigate the structural properties of the optimal value function for the multi-period economic dispatch problem with storage dynamics, and an analytical characterization of optimal storage controls and locational marginal prices. In particular, we provide a tight condition under which the optimal placement value function is submodular and an efficient computational method to certify the condition. When this condition is valid, a modified greedy algorithm for maximizing a submodular function subject to a knapsack constraint provides a $(1 - 1/e)$-optimal solution.**

I. Introduction

Recent years have witnessed a strong growth of energy storage deployment around the world. For example, the total storage deployment in the United States is increased by 243% in power capacity and 188% in energy capacity from 2014 to 2015 [1]. This is in part driven by an increasing need for energy storage in modern power systems. The value of storage in the power grid under a large penetration of renewables has been quantified in a number of prior studies (e.g., [2], [3]). It has also been shown that energy storage can be used to shift load in a way to reduce the system cost and improve system reliability. Another primary driving force is the rapidly decreasing cost of storage devices, especially batteries, partly as a consequence of increasing public and commercial interests in electric vehicles [4].

The bulk of newly deployed storage devices is front-of-meter deployment. In the U.S., 84.6% of storage deployment in the year of 2015 is utility- or grid-scale. The value of such grid-scale storage depends critically on the location at which it is installed due to the geographical heterogeneity of generation and load profiles and the possibility of network congestions [5], [6]. Therefore, developing efficient strategies for placing storage devices in power networks has attracted significant attention.

J. Qin is supported in part by the Satre family fellowship. I. Yang is supported in part by NSF under CPS:FORCES (CNS1239166). This work was partly supported by NSF CAREER Award (EECS1554178).

J. Qin and R. Rajagopal are with Stanford Sustainable Systems Laboratory, Stanford University, Stanford, CA 94305, USA {jqin, ramr}@stanford.edu

I. Yang is with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139, USA insoon@mit.edu

A majority of prior studies have treated the placement problem together with the problem of storage sizing. As a result, the placement problem is usually formulated as a form of continuous optimization. For instance, Thrampoulidis et al. [7] study the problem of allocating a fixed total storage capacity over the network to minimize the generation cost. By optimizing the capacity of each storage device together with the decision variables in economic dispatch, they obtain a structural characterization of the optimal allocation ruling out the need for placing storage at certain generation-only buses. Similar formulations are considered in [8], [9], where the multi-level nature of the placement problem is stressed and numerical approaches are provided to solve associated continuous optimization problems. With chance constraints to limit the system operation risk generated by volatile renewable energy sources, Sjödin et al. [10] study the problem of jointly optimizing generator dispatch and storage control and sizing. Qin and Rajagopal [11] derive a constrained LQG controller for (networked) distributed storage devices under uncertainty, and formulate a storage sizing problem as a convex program. Recognizing the complexity of AC power flow model, all these studies use DC approximation of AC power flow. Bose et al. [12] develop a semidefinite relaxation for storage placement problem with AC power flow model, and demonstrate its effectiveness through simulations.

Departing from continuous optimization approaches, we propose a discrete optimization formulation for energy storage placement. This is motivated by the cost structure of storage deployment: the operation and maintenance costs of storage are usually negligible compared to the fixed costs, which include the installation and capital costs. Depending on the type of storage technologies, the installation cost can be as high as the capital cost. Therefore, the cost of deploying ten units of 1 MWh battery could be dramatically different from the cost of deploying one unit of 10 MWh battery due to the differences in the installation costs. Furthermore, additional fixed cost components such as site acquisition costs could be sensitive to the location at which the storage device is to be installed. Due to the discrete nature of these heterogeneous cost factors, it is difficult to take into account all of them using a continuous optimization framework. However, discrete optimization with a budget constraint limiting the total fixed cost offers a natural and accurate modeling of these cost factors. Additionally, a discrete optimization framework also has an advantage when handling practical scenarios in which storage devices with fixed capacities are to be placed.

We formulate the placement problem as maximization of a set function, representing the *value of a storage placement* *decision*, subject to a knapsack constraint modeling the budget for the aforementioned fixed costs. Unfortunately, this class of problems is in general NP-hard. To overcome this challenge, we identify rich structures of the placement value function. In particular, we characterize conditions under which the value function is *submodular*, suggesting that the marginal benefit of adding a storage device decreases as more devices are installed. This submodular structure allows us to employ a greedy algorithm that provides a near-optimal solution with a provable suboptimality bound [13], [14]. The submodularity of energy storage placement is not intuitively unexpectable but characterizing conditions under which it holds has been recognized, e.g., in [15], as an unanswered question.

We summarize our contributions and main results as follows. First, we provide a novel discrete optimization approach to energy storage placement that allows an accurate modeling of fixed costs for storage deployment. Second, by analyzing the solution of an associated multi-period economic dispatch problem with storage dynamics, we analytically characterize the optimal system-wide cost and locational marginal prices as a function of installed storage capacities. We also show that the submodularity of the storage placement value function is not guaranteed over all problem instances through an example although such situations are unlikely to occur in practice. Our third and the most important contribution is to construct a tight condition under which the value function is submodular through a polyhedral characterization of critical regions. Furthermore, we provide an efficient and rigorous computational procedure to certify the submodularity property.

The following notations are used throughout this paper. For a transmission network with n buses and m lines, we use $i \in \mathcal{N} := \{1, \ldots, n\}$ to index the buses, and $\ell = 1, \ldots, m$ to index the lines. We also use $t \in \mathcal{T} := \{1, \ldots, T\}$ to index the time periods. For a matrix $x \in \mathbb{R}^{d \times T}$ with any given positive integers d and T, we use $x_{i,t}$ to denote its (i,t)-th entry, $x_t := (x_{1,t}, \cdots, x_{d,t})^\top \in \mathbb{R}^{d\times 1}$ to denote its tth column, and $x_i^\top := (x_{i,1}, \cdots, x_{i,T}) \in \mathbb{R}^{1\times T}$ to denote its ith row. For any real number z, we use $(z)^+ := \max(z, 0)$ to denote the positive part of z and $(z)^- := -\min(z, 0)$ to denote the negative part of z so that $z = (z)^+ - (z)^-$. For any Euclidean vector space $\mathbb{R}^d$, we use $\mathbf{1} \in \mathbb{R}^d$ to denote the all-one vector and $\mathbf{1}_k \in \mathbb{R}^d$ to denote the kth elementary vector, i.e., the vector with all zeros except for its kth element which is 1.

II. Problem Formulation

A. Power Flow and Storage Model

We first consider the operation of a connected power transmission network with n buses and m lines operated over a finite horizon of T time periods. Let $M \in \mathbb{R}^{n\times m}$ be the node-edge incidence matrix defined for the network.[1]

[1] More precisely, $M_{i,\ell} = 1$ if i is the tail of line ℓ; $M_{i,\ell} = -1$ if i is the head of line ℓ; otherwise, $M_{i,\ell} = 0$. Here, the direction of the lines are pre-determined for the purpose of defining the positive direction of flow on each line.

Under the classical DC approximation to the steady-state AC power flow [16], the lines are characterized by their susceptance $b \in \mathbb{R}^m$ and real power flow capacity $\widehat{c} \in \mathbb{R}^m$. Let $Y \in \mathbb{R}^{n\times n}$ be the DC network admittance matrix, which can be represented as $Y = M\Delta_y M^\top$, where $\Delta_y \in \mathbb{R}^{m\times m}$ is the diagonal matrix with the ℓth diagonal element being $y_\ell > 0$ which is the reciprocal of the reactance of the line. Note that $\operatorname{rank}(Y) = n-1$. Taking bus 1 to be the reference bus, we let $\bar{Y} \in \mathbb{R}^{(n-1)\times(n-1)}$ to be the sub-matrix of Y which contains the all entries of Y except its first row and first column. Define a constrained generalized inverse of Y to be $Y^\dagger := \begin{bmatrix} 0 & 0 \\ 0 & \bar{Y}^{-1} \end{bmatrix}$. For each time period $t = 1, \ldots, T$, we can then relate the line flows $f_t \in \mathbb{R}^m$ with nodal power injection $p_t \in \mathbb{R}^n$ using a linear map $\widehat{H} \in \mathbb{R}^{m\times n}$:

$$f_t = \widehat{H} p_t, \quad \text{with } \widehat{H} := \Delta_y M^\top Y^\dagger,$$

where the matrix $\widehat{H}$ is commonly referred to as the *shift-factor matrix*. Let $H := \begin{bmatrix} I \\ -I \end{bmatrix} \widehat{H} \in \mathbb{R}^{2m\times n}$ and $c := \begin{bmatrix} \widehat{c} \\ \widehat{c} \end{bmatrix} \in \mathbb{R}^{2m}$. The power flow constraints can be compactly expressed as

$$\mathbf{1}^\top p_t = 0, \quad H p_t \le c, \tag{1}$$

for all time periods $t \in \mathcal{T}$, where $p_t \in \mathbb{R}^n$ denotes the nodal power injection vector and $c := [\widehat{c}^\top, \widehat{c}^\top]^\top \in \mathbb{R}^{2m}$ with the real power flow capacity vector $\widehat{c}$. The first equation above enforces net power balance in the network, while the second inequality limits the line flows induced by the power injection vector p_t within the line capacities. The matrix H, models the linear mapping from the nodal injections to line flows, is commonly referred to as the *shift-factor matrix*.

We consider a stylized model of energy storage:[2] for each bus i, the storage's state of charge (SOC) $x_{i,t}$ evolves as

$$x_{i,t+1} = x_{i,t} - u_{i,t}, \quad t = 1, \ldots, T-1, \tag{2}$$

where $u_{i,t}$ is the amount of energy discharged (if $u_{i,t} > 0$) or charged (if $u_{i,t} < 0$) in time period t and the initial state of charge is assumed to be $x_{i,1} = 0$. Given the storage capacity $s_i \ge 0$, the following constraints model the energy limit of the storage device

$$0 \le x_{i,t} \le s_i, \quad t \in \mathcal{T}, \tag{3}$$

where $s_i = 0$ if there is no storage connected to bus i. Equations (2) and (3) can be compactly expressed in the following vector form: $0 \le L u_i \le s_i \mathbf{1}$, where $u_i \in \mathbb{R}^T$ is the vector of storage control over T periods, and $L \in \mathbb{R}^{T\times T}$ is a lower triangular matrix with entries $L_{ij} = -1$ for $i \ge j$. In other words, the information about the storage dynamics is embedded in the matrix L.

[2] Our analysis and results can be straightforwardly extended using a more detailed storage model with charging efficiency and SOC decay. For the sake of simplicity, we use the idealized model.

B. Multi-Period Economic Dispatch

For net demand $d_t \in \mathbb{R}^n$, $t \in \mathcal{T}$, which is defined as load minus uncontrollable (renewable) generation, the *economic dispatch problem* aims to identify an efficient generator dispatch to serve the net demand. When there are storage devices connected to the network, by moving energy across time periods, a careful operation of storage could reduce the total system cost. This is achieved by linking T-single period economic dispatch problems, which results in the *multi-period economic dispatch problem with storage dynamics*:

$$\begin{aligned} J(s) := \min_{g,u} \quad & \sum_{t=1}^{T} C_t(g_t) && && \text{(4a)} \\ \text{subject to} \quad & \beta_t : H(g_t + u_t - d_t) \le c, && t \in \mathcal{T}, && \text{(4b)} \\ & \gamma_t : \mathbf{1}^\top (g_t + u_t - d_t) = 0, && t \in \mathcal{T}, && \text{(4c)} \\ & \mu_i : L u_i \le s_i \mathbf{1}, && i \in \mathcal{N}, && \text{(4d)} \\ & \nu_i : L u_i \ge 0, && i \in \mathcal{N}. && \text{(4e)} \end{aligned}$$

Here, $g_t \in \mathbb{R}^n$ is the vector of controllable power generation for each time period $t \in \mathcal{T}$, $C_t(g_t) := \sum_{i \in \mathcal{N}} C_{i,t}(g_{i,t})$ is the system-wise cost for time period t, and is taken to be quadratic as common in the literature [17], so that

$$C_t(g_t) := \frac{1}{2} g_t^\top Q_t g_t + a_t^\top g_t, \quad t \in \mathcal{T},$$

where Q_t is a diagonal matrix whose diagonal entries are positive, modeling the increasing incremental (marginal) heat rate[3], and $a_t \in \mathbb{R}^n$ is the linear cost coefficient for generators over the network. The cost function mainly models the fuel cost of generating $g_{i,t}$ MW of real power. The constraints (4b) and (4c) enforces power flow constraints (1) with the nodal power injection $p_t = g_t + u_t - d_t$ for each period t. The storage dynamics and energy limit constraints are captured by (4d) and (4e). At buses with no storage connected, we set $s_i = 0$, and (4d) and (4e) reduce to $u_{i,t} = 0$ for all t.

C. Storage Placement as Optimization of A Set Function

Note that the optimal cost of this multi-period economic dispatch problem depends critically on the storage capacity vector $s \in \mathbb{R}^n$ over the network. When only a finite budget is available for installing storage devices, the location at which a storage device is installed could have a large impact on its contribution to the cost reduction due to line congestions that could isolate the benefits of storage.

In particular, given K different types of storage devices, each with some storage capacity $\bar{s}_k$ and capital and installation cost r_k, $k = 1, \ldots, K$, we want to place the storage devices to minimize the system operation cost with a given budget R for the total capital and installation cost.

We proceed to formulate the problem as an optimization of a set function. Consider the collection of all $n \times K$ *(bus, storage type) pairs*

$$\Omega := \{(i,k) : i = 1, \ldots, n, \ k = 1, \ldots, K\}.$$

[3]Heat rate is the unit amount of heat contained in fuel needed to produce a unit MW of power output. For each generator with a fixed type of fuel supply, an increasing marginal heat rate implies an increasing marginal cost with a given fuel price.

Each subset X of Ω represents a valid placement decision, and all placement decisions can be represented by a subset of Ω if we assume that only one storage with each type can be placed at each bus.[4] For notational convenience, let $\mathbb{I} : 2^\Omega \to \{0,1\}^{n \times K}$ be a set indicator function such that $\mathbb{I}_{i,k}(X) := 0$ if $(i,k) \notin X$ and $\mathbb{I}_{i,k}(X) := 1$ if $(i,k) \in X$. Note that the ith entry of the matrix-vector product $\mathbb{I}(X)\bar{s}$ is

$$(\mathbb{I}(X)\bar{s})_i = \sum_{k:(i,k) \in X} \bar{s}_k, \tag{5}$$

which is equal to the total storage capacity at bus i. We introduce a function, $V : 2^\Omega \to \mathbb{R}$, which we call the *storage placement value function*, defined as

$$V(X) := J(\mathbb{I}(\emptyset)\bar{s}) - J(\mathbb{I}(X)\bar{s}).$$

For each fixed placement decision X, the value $V(X)$ represents the reduction in the minimum T-period total generation cost by placing and optimally operating the storage devices according to the (bus, storage type) pairs contained in X. The value function V is normalized such that $V(\emptyset) = 0$. An optimal placement solution can be obtained by solving the following combinatorial optimization problem:

$$\begin{aligned} \max_{X \subseteq \Omega} \quad & V(X) && \text{(6a)} \\ \text{s.t.} \quad & \sum_{(i,k) \in X} r_k \le R. && \text{(6b)} \end{aligned}$$

We claim that our problem formulation as discrete optimization has practical advantages over continuous optimization formulations. First, our framework can handle the practical scenarios in which storage devices with fixed capacities are to be placed. Existing continuous optimization formulations are valid under a strong assumption that the System Operator can optimize the storage capacity at each bus. One can perform a post-processing, such as thresholding, to convert the solutions of continuous optimization problems into discrete solutions. However, such post-processing does not give a performance guarantee in general, whereas our method directly computes a discrete solution with a provable suboptimality bound. Second, our problem formulation naturally incorporates investment and installation costs for storage devices through the knapsack constraint (6b). In contrast, it is difficult to expect such a precise regulation in continuous optimization formulations as discussed in Section I. Lastly, the proposed discrete optimization formulation yields a very

[4]This assumption can be easily relaxed by extending the size of Ω. Suppose that we can place R number of storage devices with the same type at each bus. For each type of storage, we can create R "sub-types", each of which presents the rth appearance of the same type of the storage at the same bus, $r = 1, \ldots, R$. In other words, we have $\Omega := \{(i,k') : i = 1, \ldots, n, \ k' = (k-1)R + r, \ k = 1, \ldots, K, \ r = 1, \ldots, R\}$, where the type $k' = (k-1)R + r$ corresponds the rth appearance of the original type k. This is a classical technique to convert a integer program into a binary program. For the sake of simplicity, we will use the problem formulation (4), but all of our results are valid when we can place multiple storage devices with the same type at each bus. A similar treatment can be used to model different fixed costs for placing the same type of storage at different buses to take into account site acquisition costs.

simple placement algorithm that only requires an input-output (blackbox) model of a power system. Specifically, our greedy algorithm utilizes simulations that capture electricity market input-ouput without using detailed information about the network. This is a notable advantage over continuous optimization formulations as they often require a full network model with complete information (e.g., parameters) about markets to calculate (sub)gradients of objective functions.

III. Structures of Optimal Cost and Prices

In order to obtain efficient methods to solve the placement problem (6), which is NP-hard, we proceed to establish structural properties of the value function through an analytical characterization of the optimal prices, i.e., the solution to the dual program of (4).

We begin by considering the standard dual QP of (4):

$$\begin{aligned}\max_{\lambda,\gamma,\beta,\mu,\nu}\quad & \phi(\lambda,\gamma,\beta,\mu,\nu)\\ \text{s.t.}\quad & \lambda_t=\gamma_t\mathbf{1}-H^\top\beta_t,\quad t\in\mathcal{T},\\ & \lambda_i=L^\top(\mu_i-\nu_i),\quad i\in\mathcal{N},\\ & \beta,\mu,\nu\geq 0,\end{aligned}$$

where the Lagrange dual function is given by

$$\begin{aligned}\phi(\lambda,\gamma,\beta,\mu,\nu):=\sum_{t=1}^{T}&-\frac{1}{2}(\lambda_t-a_t)^\top Q_t^{-1}(\lambda_t-a_t)\\ &+d_t^\top\lambda_t-c^\top\beta_t-s^\top\mu_t,\end{aligned}$$

and the new variable $\lambda_{i,t}$ represents the locational marginal price (LMP) vector at bus i in period t.

The standard dual QP above can be further simplified. Observe that (4d) and (4e), representing the lower and upper limits of state of charge, cannot bind simultaneously for any storage i and time period t. In other words, if the storage device connected to bus i is empty in period t, i.e., $x_{i,t}=(Lu_i)_t=0$, then it must be the case that $x_{i,t}=(Lu_i)_t<s_i$. Similarly, $(Lu_i)_t=s_i$ signifies that $(Lu_i)_t>0$. By complementary slackness, this implies that $\mu_{i,t}\nu_{i,t}=0$ for all i and t, that is, at most one of $\mu_{i,t}$ and $\nu_{i,t}$ can be positive at the optimal solution. Combining this fact with constraint (7), which is equivalent to $(\mu_i-\nu_i)=L^{-\top}\lambda_i$, we have

$$\mu_i=(L^{-\top}\lambda_i)^+\quad\text{and}\quad\nu_i=(L^{-\top}\lambda_i)^-,\ \ i\in\mathcal{N}.$$

One can verify that, given the structure of matrix L, a more explicit display of the previous relation is

$$\mu_t=(\lambda_{t+1}-\lambda_t)^+\quad\text{and}\quad\nu_t=(\lambda_{t+1}-\lambda_t)^-,\ \ t\in\mathcal{T},\ (8)$$

where for convenience we define $\lambda_{T+1}:=0\in\mathbb{R}^n$. That is, the storage congestion price $\mu_{i,t}$ is nonzero only when the LMP λ_i ramps up in the next time period in which case its value equals the LMP increment.

Substituting the expression of μ_t into the dual QP, we get the following reduced dual program:

$$\max_{\lambda,\gamma,\beta}\quad \tilde{\phi}(\lambda,\beta) \tag{9a}$$

$$\text{s.t.}\quad \lambda_t=\gamma_t\mathbf{1}-H^\top\beta_t,\quad t\in\mathcal{T}, \tag{9b}$$

$$\beta\geq 0, \tag{9c}$$

where $\tilde{\phi}$ is a piecewise quadratic function defined as

$$\begin{aligned}\tilde{\phi}(\lambda,\beta):=\sum_{t=1}^{T}&-\frac{1}{2}(\lambda_t-a_t)^\top Q_t^{-1}(\lambda_t-a_t)\\ &+d_t^\top\lambda_t-c^\top\beta_t-s^\top(\lambda_{t+1}-\lambda_t)^+.\end{aligned}$$

By strong duality, we can characterize the function $J(s)$ via a sensitivity analysis of the primal-dual pair (4) and (9). Let $(g_t^\star(s),u_t^\star(s),\lambda^\star(s),\gamma^\star(s),\beta^\star(s))$ be a pair of primal and dual solutions to (4) and (9) for a given capacity vector s. We focus on s values which will induce nondegenerate solutions of (4). In particular, we assume the following constraint qualification for the rest of this paper.

Assumption 1 (Flow LICQ): For each $t\in\mathcal{T}$, let H_t be the collection of H's rows corresponding to the congested (oriented) lines for the flow induced by $(g_t^\star(s),u_t^\star(s))$, when there exists at least one congested line in period t.[5] Then, H_t is of full row rank for all $t\in\mathcal{T}$.

We first show that in almost all practical scenarios the prices are uniquely defined:

Proposition 1 (Uniqueness of prices): For each fixed $s\in\mathbb{R}_+^n$, the optimal dual variables $\lambda^\star(s)$ and $\gamma^\star(s)$ are unique. Furthermore, if Assumption 1 holds, then $(\lambda^\star(s),\gamma^\star(s),\beta^\star(s))$ is the unique solution to the dual problem (9).

Proof: As the objective function of (9) is strongly convex in λ, we know that $\lambda^\star(s)$ must be unique. Suppose that the first bus is the reference bus of the network, then by the definition of the shift-factor matrix, $H\mathbf{1}_1=0$. Thus constraint (9b) implies that $\gamma_t^\star(s)=\lambda_{1,t}^\star(s)$, and therefore $\gamma^\star(s)$ is also unique. Under Assumption 1, the set of primal flow constraints (4b) that are binding at the optimal solution is given by those corresponding to H_t. That is, β_t can be partitioned into $\tilde{\beta}_t$ for the binding constraints and β_t' for the slack constraints for which we know that $\beta_t'=0$ by complementary slackness. In fact, using this decomposition, the dual constraint (9b) can be written as

$$\lambda_t=\gamma_t\mathbf{1}-H_t^\top\tilde{\beta}_t,\quad t\in\mathcal{T}.$$

Now as H_t has full row rank, $\tilde{\beta}_t$ is uniquely determined by the equation above holding λ_t and γ_t fixed. This implies that β is also unique. ■

In view of Proposition 1, we assume the constraint qualification and take $(\lambda^\star(s),\gamma^\star(s),\beta^\star(s))$ as the unique dual solution for the rest of the paper. The following result characterizes the *locational marginal value of storage* via the optimal LMP:

Remark 1: Coined in Bose and Bitar [6], the term *locational marginal value of storage* is used to refer the quantity $-\nabla_s J(s)$, which characterizes the benefit of placing storage at different locations of the network when the size of storage is infinitesimal. They also obtain the expression (10) for the case where the marginal cost of generation and marginal benefit of consumption are both constants (i.e., the cost function is a piecewise linear function with two pieces). In

[5]The matrix H_t is formally defined later in (11).

fact, the expression (10) holds for any smooth convex cost function under mild regularity assumptions as in the standard sensitivity theorem of nonlinear programming.

Lemma 1 (First order sensitivity): The optimal cost function $J(s)$ is continuously differentiable and its gradient is given by

$$\nabla_s J(s) = -\sum_{t=1}^{T} \left(\lambda_{t+1}^\star(s) - \lambda_t^\star(s)\right)^+, \qquad (10)$$

where again $\lambda_{T+1}^\star(s) := 0$. Consequently, the optimal cost function $J(s)$ is nonincreasing in s_i for each $i \in \mathcal{N}$.

Proof: Consider the primal program (4), which has an infinitely differentiable objective function and linear constraints. Under the non-degeneracy condition, we can apply standard sensitivity theorem of nonlinear programming [18], which suggests the differentiability of $J(s)$ and that

$$\frac{\partial J(s)}{\partial s_i} = -\sum_{t=1}^{T} \mu_{i,t}^\star = -\sum_{t=1}^{T} \left(\lambda_{i,t+1}^\star(s) - \lambda_{i,t}^\star(s)\right)^+,$$

for any $i \in \mathcal{N}$. To show that $\partial J(s)/\partial s_i$ itself is again a continuous function, we observe that $\lambda^\star(s)$ is the unique solution of the dual QP (7). By the smoothness of the objective and constraints of (7), we know that the parameter to solution mapping $\lambda^\star(s)$ is continuous in s. Furthermore, the positive part function $(\cdot)^+$ is a continuous function from $\mathbb{R}$ to $\mathbb{R}_+$. Therefore, we conclude that $\partial J(s)/\partial s_i$ is continuous and $J(s)$ is continuously differentiable. As $\partial J(s)/\partial s_i \leq 0$, the function $J(s)$ is nonincreasing in s_i for each $i \in \mathcal{N}$. ∎

When the cost function is nonlinear and the size of the storage to be placed is far from infinitesimal, the first-order approximation of the value function using the gradient formula (10) may not be accurate.[6] We proceed to obtain a finer characterization of the optimal cost $J(s)$ by investigating its higher order derivatives. An immediate observation is that $J(s)$ is convex in s:

Lemma 2: The optimal cost function $J(s)$ is convex in s.

Proof: We write the primal problem (4) as

$$J(s) = \min_{g} \sum_{t=1}^{T} C_t(g_t) + \omega(g, s),$$

where extended real-valued function $\omega(g, s)$ is defined to be 0 if, given (g, s), there exists a control u satisfying all the constraints of (4), and $+\infty$ otherwise. Let $s^1, s^2 \in \mathbb{R}_+^n$ be two arbitrary vectors of storage capacities, and let (g^1, u^1) and (g^2, u^2) be the optimal primal solutions associated with s^1 and s^2 respectively. We claim that the function $\omega(g, s)$ is convex in (g, s). Indeed, it is easy to verify that, for $\rho \in [0, 1]$, $\omega(\rho g^1 + (1-\rho) g^2, \rho s^1 + (1-\rho) s^2) = 0$ if $\omega(g^i, s^i) = 0$ for $i = 1, 2$, as $\rho u^1 + (1-\rho) u^2$ is a feasible solution for the set of constraints given $(\rho g^1 + (1-\rho) g^2, \rho s^1 + (1-\rho) s^2)$. Therefore, $J(s)$ is convex as it is the minimum value of another convex function optimized in g over a convex set. ∎

Given that the objective function is quadratic, one would expect that the curvature information summarized by the Hessian matrix would be sufficient. This is confirmed by the following result:

Lemma 3: The optimal cost function $J(s)$ is a piecewise quadratic function with a finite number of pieces, each of which is defined on a polytope in $\mathbb{R}_+^n$. In each polytope where $J(s)$ is a quadratic function, the optimal LMPs $\lambda^\star(s)$ is affine in s.

Proof: This is a standard multi-parametric quadratic programming result. See e.g. [20]. ∎

Remark 2: The polytopes in Lemma 3 are referred to as *critical regions* in the literature of multi-parametric quadratic programming (e.g., [20], [21]). In our context, each critical region is defined as a set of s values such that the inequality constraints binding at the optimum remain unchanged. In a single period economic dispatch problem, the set of binding constraints conveys the network congestion pattern. When there are storage devices connected to the system, the definition of critical regions also depends on whether the storage constraints (4d) and (4e) bind at the solution.

Considering each critical region, we can characterize the optimal LMPs based on the network and storage congestion patterns at the optimum. For each $(i, t) \in \mathcal{N} \times \mathcal{T}$, let $\chi_{i,t}(s) = 1$ if the constraint $(Lu_i)_t \leq s_i$ is binding at the optimum and $\chi_{i,t}(s) = 0$ otherwise. In other words, χ represents the storage congestion pattern. Under strict complementary slackness, we use (8) to obtain

$$\chi_{i,t} := \chi_{i,t}(s) = \begin{cases} 1 & \text{if } \lambda_{i,t+1}^\star(s) - \lambda_{i,t}^\star(s) > 0 \\ 0 & \text{otherwise.} \end{cases}$$

We now let $\mathcal{E}_{\mathrm{C}t} \subset \{1, \ldots, 2m\}$ denote the set of oriented lines that are congested at the solution in period t and $m_t := |\mathcal{E}_{\mathrm{C}t}|$ denote the number of congested lines. Define a selection matrix $W_t \in \mathbb{R}^{m_t \times 2m}$ such that for $\ell = 1, \ldots, m_t$ and $\ell' = 1, \ldots, 2m$,

$$(W_t)_{\ell,\ell'} := (W_t(s))_{\ell,\ell'} = \begin{cases} 1 & \text{if the } \ell\text{th element in } \mathcal{E}_{\mathrm{C}t} \text{ is } \ell' \\ 0 & \text{otherwise,} \end{cases}$$

and the *shift factor matrix for congested lines* as

$$H_t := W_t H. \qquad (11)$$

Note that $W_t = 0$ if all lines are uncongested in period t.

Theorem 1: In the critical region where the storage and network congestions are represented by χ_t and W_t, $t \in \mathcal{T}$, the optimal locational marginal prices are affine in s and can be expressed as

$$\lambda_t^\star(s) = A_t(W_t)(\Delta_{\chi_t} - \Delta_{\chi_{t-1}})s + \bar{\lambda}_t(W_t), \qquad (12)$$

where Δ_{χ_t} is the diagonal matrix with vector χ_t on the

[6] The first-order approximation can be used for storage placement. We utilize it with an approximation algorithm proposed in [19] to find a solution with 0.6–0.7-*a posteriori* suboptimality bound.

diagonal with $\chi_0 := 0$, and

$$\begin{aligned} A_t(W_t) :=& \frac{1}{\mathbf{1}^\top Q_t^{-1}\mathbf{1}}\mathbf{1}\mathbf{1}^\top + Q_t M_t R_t M_t Q_t, \\ \bar{\lambda}_t(W_t) :=& A_t(W_t)\left(d_t + Q_t^{-1} a_t\right) + B_t(W_t) W_t c, \\ B_t(W_t) :=& Q_t M_t H_t^\top K_t^{-1} \end{aligned} \tag{13}$$

with $M_t := Q_t^{-1} - (Q_t^{-1}\mathbf{1}\mathbf{1}^\top Q_t^{-1})/(\mathbf{1}^\top Q_t^{-1}\mathbf{1})$, $K_t := H_t M_t H_t^\top$ and $R_t := H_t^\top K_t^{-1} H_t$. When there is no line congested in period t, all the expressions above hold with $R_t := 0$ and $B_t(W_t) := 0$.

Proof: See Appendix A. ■

To comprehend these formulas, we first set $s = 0$, in which case we obtain that locational marginal prices are the same as that for each single period economic dispatch problem without storage, i.e., $\lambda_t^\star(0) = \bar{\lambda}_t(W_t)$. Focusing on equation (13), we see that matrix $A_t(W_t)$ captures the price sensitivity of perturbing the load given the network congestion pattern W_t, while $B_t(W_t)$ represents the sensitivity of perturbing the congested lines' capacities $W_t c$. If $s \neq 0$, equation (12) suggests that when the congestion state of a storage device is changed, i.e., $\chi_{i,t} - \chi_{i,t-1} \neq 0$, perturbing storage capacity s_i has a sensitivity contribution to the prices at the period similar to that of load at the same bus. This is consistent with the intuition that the benefits of storage is achieved by modifying the effective load profile and the storage capacity could have a role in defining the prices when the storage congestion states are changed over two consecutive time periods.

As a byproduct of Theorem 1, we can obtain closed-form expressions of the (reference) energy price $\gamma_t^\star = \mathbf{1}_1^\top \lambda_t^\star$ and the congestion price $\beta_t^\star$ with respect to the capacity vector s.

Corollary 1: Under the setting of Theorem 1, $\gamma_t^\star(s)$ and $\beta_t^\star(s)$ are affine functions of $s \in \mathbb{R}_+^n$. we have

$$\gamma_t^\star(s) = \mathbf{1}_1^\top \left[A_t(W_t)(\Delta_{\chi_t} - \Delta_{\chi_{t-1}})s + \bar{\lambda}_t(W_t)\right], \tag{14}$$

$$\beta_t^\star(s) = W_t^\top B_t(W_t)^\top (\Delta_{\chi_{t-1}} - \Delta_{\chi_t})s + \bar{\beta}_t(W_t), \tag{15}$$

where

$$\bar{\beta}_t(W_t) := -W_t^\top B_t(W_t)^\top (d_t + Q_t^{-1} a_t) - W_t^\top K_t^{-1} W_t c.$$

Proof: See Appendix A. ■

Using Theorem 1 and Lemma 1, we can obtain a closed form expression for the Hessian of $J(s)$ as follows:

Theorem 2: The optimal cost function $J(s)$ is twice differentiable almost everywhere with respect to the Lebesgue measure on $\mathbb{R}_+^n$. Furthermore, for any s such that $\nabla_{ss}^2 J(s)$ exists,

$$\nabla_{ss}^2 J(s) = \mathbf{H}_\mathrm{S}(s) + \mathbf{H}_\mathrm{N}(s),$$

where $\mathbf{H}_\mathrm{S}(s)$ is the component depending on the storage congestion pattern while $\mathbf{H}_\mathrm{N}(s)$ is the component depending on the network congestion pattern, defined as

$$\mathbf{H}_\mathrm{S}(s) := \sum_{t=1}^{T} \frac{(\chi_t - \chi_{t-1})(\chi_t - \chi_{t-1})^\top}{\mathbf{1}^\top Q_t^{-1}\mathbf{1}},$$

$$\mathbf{H}_\mathrm{N}(s) := \sum_{t=1}^{T} (\Delta_{\chi_t} - \Delta_{\chi_{t-1}}) Q_t M_t R_t M_t Q_t (\Delta_{\chi_t} - \Delta_{\chi_{t-1}}).$$

Proof: See Appendix A. ■

As network congestions are only expected in a small percentage of all the operation hours, here we focus on the component $\mathbf{H}_\mathrm{S}(s)$. We show that the sign of entries of $\mathbf{H}_\mathrm{S}(s)$ depends on whether the local peaks and valleys of the LMP processes are aligned. We define the set of time indices for which each price sequence reaches a local peak (LP) and a local valley (LV) as

$$\mathcal{T}_k^{\mathrm{LP}} = \{t \in \mathcal{T} : \lambda_{k,t}^\star(s) > \lambda_{k,t-1}^\star(s), \lambda_{k,t}^\star(s) \geq \lambda_{k,t+1}^\star(s)\},$$

$$\mathcal{T}_k^{\mathrm{LV}} = \{t \in \mathcal{T} : \lambda_{k,t}^\star(s) \leq \lambda_{k,t-1}^\star(s), \lambda_{k,t}^\star(s) < \lambda_{k,t+1}^\star(s)\},$$

respectively, where $k = i, j$, $\lambda_{k,T+1}^\star(s) := 0$, and $\lambda_{k,0}^\star(s) := +\infty$ (which is consistent with the definition $\chi_0 := 0$). Furthermore, we denote the number of aligned peaks and valleys of these two price sequences as

$$T_{ij}^{\mathrm{align}} = |\mathcal{T}_i^{\mathrm{LP}} \cap \mathcal{T}_j^{\mathrm{LP}}| + |\mathcal{T}_i^{\mathrm{LV}} \cap \mathcal{T}_j^{\mathrm{LV}}|,$$

and the number of time periods when these price sequences exhibit opposite local extrema as

$$T_{ij}^{\mathrm{opp}} = |\mathcal{T}_i^{\mathrm{LP}} \cap \mathcal{T}_j^{\mathrm{LV}}| + |\mathcal{T}_i^{\mathrm{LV}} \cap \mathcal{T}_j^{\mathrm{LP}}|.$$

Then, we can express $\mathbf{H}_{\mathrm{S}ij}(s)$ simply using these two statistics of the price sequences.

Proposition 2 (LMP fluctuations and Hessian): Suppose that the quadratic cost coefficients are fixed over time, i.e., $Q_t \equiv Q$ for all $t \in \mathcal{T}$. Then, we have

$$\mathbf{H}_{\mathrm{S}ij}(s) = \frac{1}{\mathbf{1}^\top Q^{-1}\mathbf{1}}\left[T_{ij}^{\mathrm{align}} - T_{ij}^{\mathrm{opp}}\right].$$

Therefore, $\mathbf{H}_{\mathrm{S}ij}(s) \geq 0$ if and only if the fluctuations of $\lambda_i^\star(s)$ and $\lambda_j^\star(s)$ are similar in the sense that $T_{ij}^{\mathrm{align}} \geq T_{ij}^{\mathrm{opp}}$.

Proof: Given that

$$\mathbf{H}_{\mathrm{S}ij}(s) = \sum_{t=1}^{T} \frac{1}{\mathbf{1}^\top Q^{-1}\mathbf{1}}(\chi_{i,t} - \chi_{i,t-1})(\chi_{j,t} - \chi_{j,t-1}),$$

we note that each term in the summation is nonzero only when both $(\chi_{i,t} - \chi_{i,t-1})$ and $(\chi_{j,t} - \chi_{j,t-1})$ are nonzero, i.e., the storage congestion states at both buses change simultaneously. By the definition of χ, for $k = i, j$ we have

$$\chi_{k,t} - \chi_{k,t-1} = \begin{cases} 1 & \text{if } t \in \mathcal{T}_k^{\mathrm{LV}}, \\ -1 & \text{if } t \in \mathcal{T}_k^{\mathrm{LP}}, \\ 0 & \text{otherwise.} \end{cases}$$

Thus

$$\begin{aligned} &(\chi_{i,t} - \chi_{i,t-1})(\chi_{j,t} - \chi_{j,t-1}) \\ &= \begin{cases} 1 & \text{if } t \in \mathcal{T}_i^{\mathrm{LP}} \cap \mathcal{T}_j^{\mathrm{LP}}, \text{or } t \in \mathcal{T}_i^{\mathrm{LV}} \cap \mathcal{T}_j^{\mathrm{LV}}, \\ -1 & \text{if } t \in \mathcal{T}_i^{\mathrm{LV}} \cap \mathcal{T}_j^{\mathrm{LP}}, \text{or } t \in \mathcal{T}_i^{\mathrm{LP}} \cap \mathcal{T}_j^{\mathrm{LV}}, \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

and therefore the claim follows. ■

Proposition 2 states that when the quadratic coefficients of generation cost functions are not time-varying, we can evaluate $\mathbf{H}_{\mathrm{S}ij}(s)$ simply by counting the number of times the local peaks and valleys of the LMP processes on bus i and j are aligned, and the number of times of the co-occurrence of opposite local extrema. When the LMPs are aligned in the sense of Proposition 2, possibly driven by the similarity of the load profiles, then this Hessian component will have nonnegative entries. We also note that this result can be easily generalized to the setting where the quadratic cost coefficients are different in each time period t, by weighting each aforementioned occurrence differently according to $\mathbf{1}^\top Q_t^{-1}\mathbf{1}$.

IV. Submodularity of Placement Value Function

Equipped with the structural properties of the optimal cost function $J(s)$, we now characterize the set function $V(X)$. Recall that the placement value function $V(X)$ models the reduction of the optimal operational cost by employing the placement decision X, which is defined as a subset of Ω that contain all admissible (bus, storage type) pairs. In particular, we provide conditions under which the value function belongs to the class of submodular functions, one of the most tractable classes in discrete optimization.

Definition 1 (Submodularity and monotonicity): For a finite set Ω, a set function $F : 2^\Omega \to \mathbb{R}$ is said to be *submodular* if for any $X \subseteq Y \subseteq \Omega$ and $\mathbf{e} \in \Omega \setminus Y$,

$$F(X \cup \{\mathbf{e}\}) - F(X) \geq F(Y \cup \{\mathbf{e}\}) - F(Y). \tag{17}$$

The function is said to be *monotonically nondecreasing* if for any $X \subseteq \Omega$ and $\mathbf{e} \in \Omega \setminus X$,

$$F(X \cup \{\mathbf{e}\}) \geq F(X). \tag{18}$$

In our case, (18) implies that the marginal benefit of installing a new storage device is nonnegative and (17) states that such marginal benefit should diminish when more storage devices are connected to the system. It is straightforward to check that any modular function is submodular. Evidently, the nondecreasing property of $V(X)$ follows from the fact that $J(s)$ is nonincreasing (Lemma 1). To check whether $V(X)$ is submodular, it is instrumental to consider an alternative characterization of submodularity, based on discrete derivatives defined for set functions.

Definition 2: For any set function $F : 2^\Omega \mapsto \mathbb{R}$, the *discrete derivative* of F in $\mathbf{e} \in \Omega$ is defined as

$$\mathrm{D}_{\mathbf{e}} F(X) := F(X \cup \{\mathbf{e}\}) - F(X \backslash \{\mathbf{e}\}).$$

It is straightforward to check that the following theorem provides a necessary and sufficient condition for submodularity [22].

Theorem 3: A set function $F : 2^\Omega \mapsto \mathbb{R}$ is submodular if and only if

$$\mathrm{D}_{\mathbf{e}} \left(\mathrm{D}_{\mathbf{e}'} F(X)\right) \leq 0, \tag{19}$$

for all $\mathbf{e}, \mathbf{e}' \in \Omega$, $\mathbf{e} \neq \mathbf{e}'$ and $X \subseteq \Omega$.

This condition allows us to relate the submodularity of $V(X)$ to the sign of the Hessian entries of $J(s)$. We relate the submodularity of $V(X)$ to the sign of the Hessian entries of $J(s)$ as follows:

Theorem 4 (Sufficient condition for submodularity): The storage placement value function $V : 2^\Omega \to \mathbb{R}$ is submodular if

$$\left(\nabla^2_{ss} J(s)\right)_{ij} \geq 0, \quad \forall i, j \in \mathcal{N},$$

for all $s \in S := [0,\ \bar{s}^{\max}]^n$, where $\bar{s}^{\max} := \sum_{k=1}^K \bar{s}_k$ is the maximum storage capacity to be achieved at each bus.

Proof: For any $X \subseteq \Omega$, and without loss of generality $\mathbf{e}_i := (i, k_i) \notin X$, $\mathbf{e}_j := (j, k_j) \notin X$, we have $\mathrm{D}_{\mathbf{e}_i} V(X) = V(X \cup \{(i, k_i)\}) - V(X)$ and

$$\begin{aligned}
&\mathrm{D}_{\mathbf{e}_j}\left(\mathrm{D}_{\mathbf{e}_i} V(X)\right) \\
&\quad = \left[V\left(X \cup \{(i,k_i),(j,k_j)\}\right) - V\left(X \cup \{(j,k_j)\}\right)\right] \\
&\qquad - \left[V(X \cup \{(i,k_i)\}) - V(X)\right].
\end{aligned}$$

Let $s^0 := \mathbb{I}(X)\bar{s}$. Using the definition of V, we have

$$\begin{aligned}
\mathrm{D}_{\mathbf{e}_j}\left(\mathrm{D}_{\mathbf{e}_i} V(X)\right) = {}& \left[J(s^0 + \bar{s}_j \mathbf{1}_j) - J(s^0 + \bar{s}_i \mathbf{1}_i + \bar{s}_j \mathbf{1}_j)\right] \\
& - \left[J(s^0) - J(s^0 + \bar{s}_i \mathbf{1}_i)\right].
\end{aligned}$$

Since $J(s)$ is continuously differentiable, the following integral expression is well-defined:

$$\begin{aligned}
&\mathrm{D}_{\mathbf{e}_j}\left(\mathrm{D}_{\mathbf{e}_i} V(X)\right) \\
&\quad = \int_0^{\bar{s}_i} \left[\frac{\partial J}{\partial s_i}(s^0 + x\mathbf{1}_i) - \frac{\partial J}{\partial s_i}(s^0 + \bar{s}_j\mathbf{1}_j + x\mathbf{1}_i)\right] \mathrm{d}x.
\end{aligned}$$

Meanwhile, given that $\partial J / \partial s_i$ is differentiable almost everywhere with respect to Lebesgue measure, we have

$$\begin{aligned}
&\mathrm{D}_{\mathbf{e}_j}\left(\mathrm{D}_{\mathbf{e}_i} V(X)\right) \\
&\quad = -\int_0^{\bar{s}_i} \int_0^{\bar{s}_j} \frac{\partial^2 J}{\partial s_j \partial s_i}(s^0 + x\mathbf{1}_i + y\mathbf{1}_j)\, \mathrm{d}y\, \mathrm{d}x.
\end{aligned}$$

As $\left(\nabla^2_{ss} J(s)\right)_{ij} \geq 0$, we have $\mathrm{D}_{\mathbf{e}_j}\left(\mathrm{D}_{\mathbf{e}_i} V(X)\right) \leq 0$ for any $i, j \in \mathcal{N}$ and any k_i and k_j. Thus, using (19), we conclude that the set function V is submodular. ■

Theorem 4 provides a sufficient condition for the submodularity of V by just checking the sign of the Hessian entries of the optimal cost function $J(s)$, which can be computed using Theorem 2. The characterization is essentially tight, in the following sense.

Corollary 2: If $\left(\nabla^2_{ss} J(s)\right)_{ij} < 0$ for some $s \in \mathbb{R}^n_+$ and $i, j \in \mathcal{N}$, then there exists a storage capacity vector $\bar{s} \in \mathbb{R}^n_+$ and the corresponding Ω such that $V(X)$ is not submodular on the subsets of Ω.

Of course, this corollary is only a partial converse of Theorem 4 as the critical regions in which $\left(\nabla^2_{ss} J(s)\right)_{ij} < 0$ may not be contained in the region S of interest given the fixed storage capacity vector $\bar{s}$. Even if the point s resulting in negative Hessian entries is contained in S, the function $V(X)$ could still be submodular if the critical region with negative Hessian entries is relatively small (or the magnitude of the negative Hessian entries is small) so that its contribution to the discrete derivative is overweighed by the contribution from other critical regions with positive Hessian entries.

By the convexity of the optimal cost function $J(s)$ (Lemma 2), it is always the case that the diagonal entries of the Hessian matrix $\nabla^2_{ss} J(s)$ are nonnegative. To build intuitions on the sign of the off-diagonal entries of the Hessian matrix, we considered a two bus example.

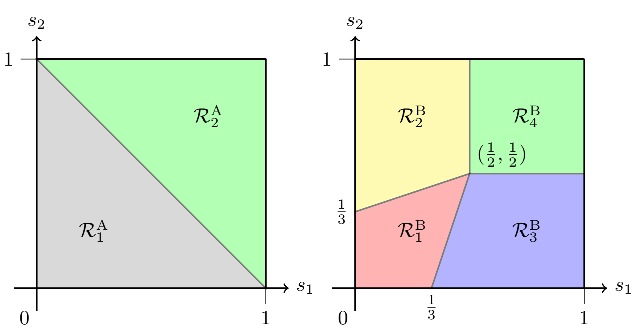
A. Two-Bus Network

To gain qualitative insight, we fix the cost function to be $C_t(g_t) = \frac{1}{2} g_t^\top g_t$, for $t = 1, \ldots, T := 3$, that is $Q_t \equiv I \in \mathbb{R}^{2\times 2}$ and $a_t \equiv 0$. With this cost function, given a time varying demand profile over the network, if neither storage nor line capacity is constraining, then the solution exhibits a form of "water-filling" behavior where the optimal flows result in equalized generation from each bus and each time period. We also notice that $g_t^\star = \lambda_t^\star$ for this cost function, by the first order optimality condition of (4).

We investigate the property of $J(s)$ and the optimal primal and dual solution of the multi-period economic dispatch problem for all storage capacities s in the region $S = [0,1] \times [0,1]$. The line capacity is fixed to be 0.5. We consider the following two cases: one commonly observed in simulation where all critical regions inside of S have $J(s)$ with only nonnegative Hessian entries, and the other specially constructed such that one of the critical regions has negative off-diagonal Hessian entries.

- *Case A:* $d^{\mathrm{A}} = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 2 & 2 \end{bmatrix}$.
- *Case B:* $d^{\mathrm{B}} = \begin{bmatrix} 1 & 2 & 1 \\ 3 & 2 & 3 \end{bmatrix}$.

The critical regions for these cases are depicted in Figure 1. For each critical region, we obtain the expression of the



(a) Case A (b) Case B

Fig. 1: Critical regions for the two-bus examples. In the figure with a slight abuse of notation, we use $\mathcal{R}_r^{\mathrm{A}}$ and $\mathcal{R}_r^{\mathrm{B}}$ to denote the rth critical region for each case.

optimal cost function $J(s)$ (which includes its quadratic and linear coefficients), and for a set of points on the a mesh grid inside of each critical region, we solve the multi-period economic dispatch problem and obtain the optimal primal dual solution. In the all 6 critical regions across these two cases, only the red region in case B, i.e., $\mathcal{R}_1^{\mathrm{B}}$, has negative Hessian entries. Due to space limit, we focus on this region for the rest of this subsection. The optimal primal variables for $\tilde{s} = [0.2, 0.2]^\top$ is shown in Figure 2.[7] The optimal cost function in the critical region is

$$J(s) = \frac{1}{2} s^\top \begin{bmatrix} 1.5 & -0.5 \\ -0.5 & 1.5 \end{bmatrix} s + \begin{bmatrix} -0.5 & -0.5 \end{bmatrix} s + 12.5,$$

with $J(\tilde{s}) = 12.34$.

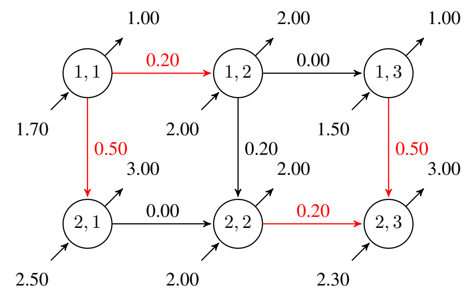


Fig. 2: Optimal flow for the case with negative Hessian entries.

A key observation for the specialty of this case can be made. The optimal prices $\lambda^\star$, as read from the generation values, follow a *low-high-low* pattern on one bus and a *high-low-high* pattern on the other bus. This is unusual in practical settings especially in planning scenarios, as the LMPs are often driven by load profiles. If occurred in practice, such a phenomenon would indicate that (a) the load profiles on these two buses complement each other in the sense that the load on bus 1 peaks when the load on bus 2 falls to its valley, and (b) the transmission link between these two buses is weak and congested so that the optimal/equilibrium prices still follow such patterns. Given that each load bus in transmission network often represents a collection of smaller loads, the condition in (a) means that the aggregates of these small loads follow very different temperol patterns at different locations in the network. Furthermore, if we were concerned with the transmission planning problem of determining which transmission lines to strengthen, conditions (a) and (b) are often strong indicators for increasing the capacity of the line connecting these two buses. In fact, for case B, doubling the line capacity eliminates the critical region with negative Hessian entries. We will further make some of these notions solid in the next subsection.

B. Certificate for Submodularity

Albeit the negative Hessian case above looks unlikely to occur in practice and in all our simulations with PJM price and load data we have not yet observed negative Hessian entries, there is no *a priori* theoretical guarantee that V is submodular. In other words, its submodularity depends on the

[7] Here, each node in the graph represents a (bus, time period) pair. The vertical edges of the graph represent the transmission line connecting the two buses, while the horizon edges represent storage device storing power for future use. Around each node (i, t), the value associated with an "inflow arrow" is the generation $g_{i,t}^\star$, and the value associated with an "outflow arrow" is the demand $d_{i,t}$. The value on each edge is the optimal flow sent through the edge; for storage edges such flow is the amount of energy stored at the end of last time period. Red edges are congested at the solution.

problem instance, in particular, the load and network data. Thus, it is of interest to develop an efficient computation procedure which certifies the submodularity of V.

This is generally a challenging task, as verifying the submodularity of V by definition involves checking an exponential number of inequalities. Theorem 4 reduces this problem to checking the sign of Hessian entries of a continuous function, $J(s)$, on the region of interest S. Theorem 2 provides a formula to compute the Hessian for almost every $s \in S$. This formula is invariant in each critical regions (Theorem 3), and hence it suffices to evaluate the Hessian once per critical region inside of S. It remains to specify how one would iterate over the critical regions, which we address here.

We start by providing an explicit characterization of the critical region containing almost every capacity vector s.[8] Upon solving the multi-period economic dispatch problem at s, we can identify the set of binding constraints and the associated χ_t and W_t for $t \in \mathcal{T}$. The critical region containing s can then be expressed as the set of storage capacity vectors where the storage and network congestion states are not changed:

Theorem 5: Given $\chi_t(s)$ and $W_t(s)$, $t \in \mathcal{T}$ evaluated at an arbitrary $s \in S$ (except a set of measure zero), the critical region containing s is a open convex polytope $\mathcal{R}_s$ defined by the set of $\tilde{s} \in \mathbb{R}_+^n$ satisfying the linear inequalities

$$\lambda^\star_{i,t+1}(\tilde{s}) - \lambda^\star_{i,t}(\tilde{s}) > 0,\, (i,t) \in \mathcal{N} \times \mathcal{T} \text{ s.t. } \chi_{i,t}(s) = 1,$$

$$\beta^\star_{\ell,t}(\tilde{s}) > 0,\, (\ell, t) \in \mathcal{E}_{\mathrm{C}t} \times \mathcal{T}.$$

In other words, for each $\tilde{s} \in \mathcal{R}_s$, the associated storage congestion pattern $\tilde{\chi}_t$ and network congestion pattern $\tilde{W}_t$ satisfy $\tilde{\chi}_t = \chi_t(s)$ and $\tilde{W}_t = W_t(s)$, $t \in \mathcal{T}$.

Proof: By the proof of Theorem 1 and Corollary 1, we know that $(\lambda^\star(s), \gamma^\star(s), \beta^\star(s))$ is a stationary point of the objective of the dual QP (9). In other words, $(\lambda^\star(s), \gamma^\star(s), \beta^\star(s))$ is the unconstrained local maximizer of (9) in an affine subspace defined by the set of equality constraints of (9) and inequality constraints of (9) which are binding at s. Recall that, given the storage congestion state $\chi \in \mathbb{R}^{n \times T}$, the objective of (9) can be written as

$$\tilde{\phi}(\lambda, \beta) = \sum_{t=1}^{T} -\frac{1}{2}(\lambda_t - a_t)^\top Q_t^{-1}(\lambda_t - a_t) + d_t^\top \lambda_t - c^\top \beta_t - s^\top \Delta_{\chi_t}(\lambda_{t+1} - \lambda_t).$$

Now consider a vector $\tilde{s}$ in a neighborhood of s. The first condition (20) in the definition of $\mathcal{R}_s$ ensures that, at $\tilde{s}$, the storage congestion state $\chi(\tilde{s})$ given by (11) is unchanged from $\chi(s)$. Therefore, $(\lambda^\star(\tilde{s}), \gamma^\star(\tilde{s}), \beta^\star(\tilde{s}))$ is still the unconstrained local maximizer of (9) in the same affine subspace when s is replaced with $\tilde{s}$ in the above expression of the objective function $\tilde{\phi}$. The second condition (20) in the definition of $\mathcal{R}_s$ guarantees that $(\lambda^\star(\tilde{s}), \gamma^\star(\tilde{s}), \beta^\star(\tilde{s}))$ is feasible for (9). Indeed, we see that by the expression of $\beta^\star_t(s)$ in (15), modifying s does not affect $\beta^\star_{\ell,t}(s)$ for a line ℓ that is not congested and hence $\beta^\star_{\ell,t}(\tilde{s})$ is 0 if line ℓ is uncongested, while (20) ensures that $\beta^\star_{\ell,t}(\tilde{s})$ remains positive if line ℓ is congested . Therefore, $(\lambda^\star(\tilde{s}), \gamma^\star(\tilde{s}), \beta^\star(\tilde{s}))$ must be a global maximizer of the dual QP (9) because the problem is concave. By strict complementary slackness, the conditions defining $\mathcal{R}_s$ ensure that the set of binding inequality constraints is unaltered. ■

Given this polyhedral characterization of the critical regions, the iterative construction of all the critical regions and the complexity of such a process follow from the standard practice of multi-parametric quadratic programming [20]. Here, we only provide a brief description. To start, pick an initial point $s \in S_0 := S$ and compute the critical region $\mathcal{R}_s$ containing it using Theorem 5. Focusing on the part of the critical region inside of S_0, i.e., $\mathcal{R}_s \cap S_0$, and writing the inequality constraints defining this polytope as $Ys \leq v$, we can partition the remaining region in S_0 as

$$S_i := \{s \in S_0 : y_i^\top s \geq v_i, y_j^\top s \leq v_j, \forall j < i\},$$

where $y_i^\top$ is the ith row of Y, and i ranges from 1 to the number of rows of Y. Recursively applying this process to S_i, we will get the collection of all critical regions in S.

The rest of this subsection is devoted to a special case that bears a substantial amount of practical interests given the relative small size of storage devices to be placed[9]:

Definition 3: A storage placement problem is said to satisfy the *small storage condition*, if the capacity region of interest is a subset of the closure of the critical region containing $s = 0$, i.e., $S \subseteq \bar{\mathcal{R}}_0$, with $\mathcal{R}_0$ as defined in Theorem 5.

Remark 3: Verifying the small storage condition involves checking whether the polytope $\bar{\mathcal{R}}_0$ contains the box $S = [0, \bar{s}^{\max}]^n$. Instead of checking whether all 2^n vertices of the box belong to the polytope, we can simply compare the optimal value of the optimal set containment problem $\min\{\rho \in \mathbb{R}_+ : S \subseteq \rho\mathcal{R}_0\}$ with 1, where the scaled set $\rho\mathcal{R}_0$ is $\{s : (1/\rho)s \in \mathcal{R}_0\}$. Write $\bar{\mathcal{R}}_0$ as $\{s : Ys \leq v\}$. This set containment problem can be formulated as the following linear program [23]:

$$\begin{aligned} \min_{\rho, \Lambda} \quad & \rho \\ \text{s.t.} \quad & \Lambda[I,\ -I]^\top = Y, \\ & \Lambda[\bar{s}^{\max}\mathbf{1}^\top, 0^\top]^\top \leq \rho v, \\ & \rho, \Lambda \geq 0. \end{aligned}$$

If the small storage condition holds, the submodularity of V can be checked using merely the LMP vector $\lambda^\star_t(0)$ and the network congestion pattern W_t in the base case where no storage has been installed. In other words, we can certify submodularity by just using the solutions of the single-period economic dispatch problems for time periods $t \in \mathcal{T}$.

[8] When s is on the boundary of two critical regions, strict complementary slackness fails to hold and in general one may get a degenerate solution. However, the set of boundary points has Lebesgue measure 0 and hence no contribution to our submodularity characterization as shown in the proof of Theorem 4.

[9] To get a sense, the total power capacity of storage installation in the U.S. in 2015 is 221 MW [1] while the average U.S. generation in the same year is 467 GW.

Corollary 3: Under the small storage condition, all $s \in S$ share the same storage and line congestion patterns as $\chi_t(0)$ and $W_t(0)$, $t \in \mathcal{T}$ which can be obtained by solving T single-period economic dispatch problems. Furthermore, if the network topology is a tree, $\chi_t(0)$ and $W_t(0)$, $t \in \mathcal{T}$ are uniquely determined using only the LMP data $\lambda^\star(0)$.

C. Placement Algorithms via Submodular Maximization

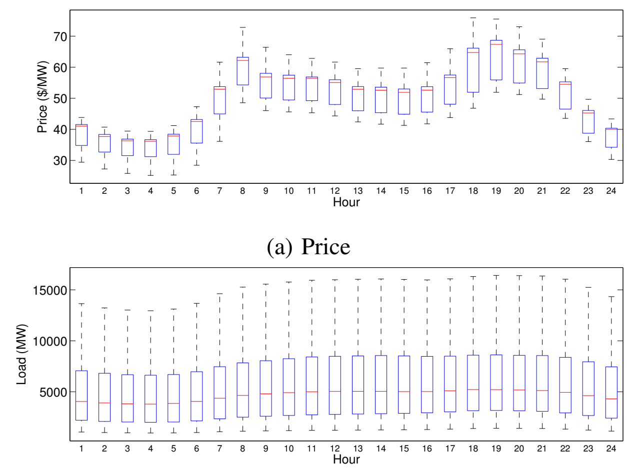
If the submodularity of the value function is verified, we can employ a (modified) greedy algorithm to obtain a near-optimal solution. In particular, Nemhauser et al. [13] show that a greedy algorithm gives a $(1 - \frac{1}{e})$-approximation of an optimal solution when maximizing monotone submodular functions subject to cardinality or matroid constraints. This algorithm is applicable to our problem when there is only one type of storage devices, i.e., $K = 1$. In the case of multi-type devices, this standard greedy algorithm may not fully utilize the diminishing return property as it can get stuck at a possibly unreasonable solution due to a knapsack (budget) constraint. However, a modification of the greedy algorithm is shown to achieve the same performance guarantee [24], [25]. This algorithm for knapsack constraints uses the partial enumeration heuristic proposed by Khuller et al. [14] which enumerates all subsets of up to three elements. Its details are presented in Algorithm 1. The first candidate X_1 of the solution maximizes the benefit V among all feasible sets of cardinality one or two as shown in Line 1. The second candidate X_2 is constructed in a greedy way by locally optimizing the *incremental benefit-cost ratio* $[V(X \cup \{(i,k)\}) - V(X)]/r_k$ starting from each set X of cardinality three as illustrated in Lines 2–15. Finally, the algorithm generates an output by comparing the two candidates X_1 and X_2.

Algorithm 1: Modified greedy algorithm for energy storage placement

1 $X_1 \leftarrow \mathrm{argmax}\{V(X) : |X| \leq 2, \sum_{(i,k)\in X} r_k \leq R\}$;
2 $X_2 \leftarrow \emptyset$;
3 **foreach** $X \subseteq \Omega$ *s.t.* $|X| = 3$, $\sum_{(i,k)\in X} r_k \leq R$ **do**
4   Candidates $\leftarrow \Omega \setminus X$;
5   **while** Candidates $\neq \emptyset$ **do**
6         $\mathbf{e} \leftarrow \mathrm{argmax}_{(i,k)\in \text{Candidates}} \frac{V(X\cup\{(i,k)\})-V(X)}{r_k}$;
7   **if** $\sum_{k:(i,k)\in X\cup\{\mathbf{e}\}} r_k \leq R$ **then**
8             $X \leftarrow X \cup \{\mathbf{e}\}$;
9   Candidates $\leftarrow$ Candidates $\setminus \{\mathbf{e}\}$;
10   **end**
11   **end**
12   **if** $V(X) > V(X_2)$ **then**
13        $X_2 \leftarrow X$;
14   **end**
15 **end**
16 $X^\star \leftarrow \mathrm{argmax}_{X\in\{X_1,X_2\}} V(X)$;

V. Numerical Experiments

The placement algorithms are tested using the network data of IEEE 14 bus test case. Hourly zonal aggregated locational marginal price and load data are obtained from PJM interconnection. The data corresponds to 14 zones inside PJM RTO for the year of 2014. We consider the hourly operation of storage over a representative day. The input data for the representative day is obtained by averaging over all the 365 days of the year for each hour of the day. The hourly price and load data distribution over 14 zones are plotted in Figure 3. The hourly average load in the system is 80.5 GW.



(a) Price

(b) Load

Fig. 3: Boxplots of price and load data.

The load and price time series for these 14 zones are assigned to the 14 buses of the network, where the price data is used to specify the linear coefficient of generation cost. We set the quadratic cost coefficients of all the generators to be 0.01, which is the median value of quadratic cost coefficients specified for IEEE 14 bus test case in MATPOWER [26]. The capacity of all lines are set to be the average load per bus over the 24 hours. We consider a simple setting in which exhaustive search is still feasible so that the performance of the greedy placement can be compared with the exact optimal solution in the quadratic case. To this end, we let the type of storage to be $K = 1$ and denote $\bar{s}_1 = \bar{s}$. Consequently, the budget constraint becomes a constraint on the number of storage to be placed.

We consider placing 5 storages over the 14 buses, with the total energy capacity being 150 MWh. Using the optimal set containment optimization (21), it is verified that this setting satisfies the small storage assumption and in the critical region the Hessian condition in Theorem 4 holds. The greedy strategy in Algorithm 1 is implemented. We also perform an exhaustive search over all the feasible storage placement to verify the actual performance of the algorithm. Instead of being $(1 - 1/e)$ suboptimal as suggested by the worst case performance bound, the greedy algorithm has in fact identified the exact optimal placement in this case, with buses $\{5, 11, 12, 1, 9\}$ selected to place storage.

VI. Conclusions and Future Work

We have developed a discrete optimization framework for energy storage placement in power networks taking into account heterogeneous storage installation and capital costs. Exploring the structural properties of a multi-period economic dispatch problem with storage dynamics and its dual program, we have derived several salient features of the placement problem including a tight characterization of conditions under which the placement value function is submodular. In particular, representing each critical region in the parametric economic dispatch problem as a convex polytope, we have proposed a rigorous and efficient method to certify the submodularity property. This work can be extended in several directions including uncertainty-aware placement strategies.

Appendix

A. Proof of Theorem 1, Corollary 1 and Theorem 2

By strict complementary slackness,

$$\beta_{\ell,t} = 0 \quad \text{if} \quad \ell \notin \mathcal{E}_{\mathrm{C}t} \quad \text{and} \quad \beta_{\ell,t} > 0 \quad \text{if} \quad \ell \in \mathcal{E}_{\mathrm{C}t}.$$

Thus we can focus on the reduced dual variable

$$\widetilde{\beta}_t := W_t \beta_t \in \mathbb{R}^{m_t}, \quad t = 1, \ldots, T,$$

and the following reduced form of dual program (9) in a neighborhood of the given s vector

$$\begin{aligned} \max_{\lambda,\gamma,\widetilde{\beta}} \quad & \psi(\lambda,\gamma,\widetilde{\beta}) \\ \text{s.t.} \quad & \lambda_t = \gamma_t e - H_t^\top \widetilde{\beta}_t, \quad t = 1,\ldots,T, \end{aligned}$$

where the objective function is

$$\begin{aligned} \psi(\lambda,\gamma,\widetilde{\beta}) := \sum_{t=1}^{T} & -\frac{1}{2}(\lambda_t - \alpha_t)^\top Q_t^{-1}(\lambda_t - \alpha_t) + d_t^\top \lambda_t \\ & - c_t^\top \widetilde{\beta}_t - s^\top \Delta_{\chi_t}(\lambda_{t+1} - \lambda_t) \end{aligned}$$

and $c_t := W_t c$. Define for convenience constants $\chi_{i,0} = 0$ for $i = 1, \ldots, n$. Then optimization above is separable across time, and we can solve each of the following T optimizations given the binding constraints

$$\begin{aligned} J_t(s) := \max_{\lambda_t,\gamma_t,\widetilde{\beta}_t} \quad & \psi_t(\lambda_t,\gamma_t,\widetilde{\beta}_t) \\ \text{s.t.} \quad & \lambda_t = \gamma_t e - H_t^\top \widetilde{\beta}_t, \end{aligned}$$

where

$$\begin{aligned} \psi_t(\lambda_t,\gamma_t,\widetilde{\beta}_t) := & -\frac{1}{2}(\lambda_t - \alpha_t)^\top Q_t^{-1}(\lambda_t - \alpha_t) + d_t^\top \lambda_t \\ & - c_t^\top \widetilde{\beta}_t + s^\top(\Delta_{\chi_t} - \Delta_{\chi_{t-1}})\lambda_t. \end{aligned}$$

This equality constrained quadratic program can be solved analytically provided that solution of the original dual program is unique. In particular, we have that the optimal LMP and reduced congestion prices as follows

$$\begin{aligned} \lambda_t^\star = & (Q_t M_t R_t M_t Q_t + \rho_t \mathbf{1}\mathbf{1}^\top)\left((\Delta_{\chi_t} - \Delta_{\chi_{t-1}})s + d_t + Q_t^{-1}\alpha_t\right) \\ & + Q_t M_t H_t^\top K_t^{-1} c_t, \\ \widetilde{\beta}_t^\star = & -K_t^{-1}\left[H_t M_t\left(Q_t(\Delta_{\chi_t} - \Delta_{\chi_{t-1}})s + Q_t d_t + \alpha_t\right) + c_t\right], \end{aligned}$$

where $\rho_t := 1/[\mathbf{1}^\top Q_t^{-1}\mathbf{1}]$, and M_t, K_t and R_t are as defined in Theorem 1. Collecting terms, we have

$$\lambda_t^\star(s) = A_t(W_t)(\Delta_{\chi_t} - \Delta_{\chi_{t-1}})s + \bar{\lambda}_t(W_t),$$

with $A_t(W_t)$, $B_t(W_t)$ and $\bar{\lambda}_t(W_t)$ as defined in Theorem 1. This proves Theorem 1. Meanwhile,

$$\widetilde{\beta}_t^\star = -B_t(W_t)^\top\left[(\Delta_{\chi_t} - \Delta_{\chi_{t-1}})s + d_t + Q_t^{-1}\alpha_t\right] - K_t^{-1}c_t.$$

Corollary 1 then follows from the relationship between the (reference) energy price $\gamma_t^\star$ and the LMP $\lambda_t^\star$ and the fact that $\beta_t^\star = W_t^\top \widetilde{\beta}_t^\star$.

The optimal cost is then

$$\begin{aligned} J_t(s) = & \frac{1}{2}s^\top(\Delta_{\chi_t} - \Delta_{\chi_{t-1}})A_t(W_t)(\Delta_{\chi_t} - \Delta_{\chi_{t-1}})s \\ & + \left[A_t(W_t)(d_t + Q_t^{-1}\alpha_t) + B_t(W_t)c_t\right]^\top(\Delta_{\chi_t} - \Delta_{\chi_{t-1}})s \\ & + \bar{J}_t, \end{aligned}$$

where the last term $\bar{J}_t$ does not depend on s and has the following form

$$\begin{aligned} \bar{J}_t = & \frac{1}{2}(d_t + Q_t^{-1}\alpha_t)^\top A_t(W_t)(d_t + Q_t^{-1}\alpha_t) - \frac{1}{2}\alpha_t^\top Q_t^{-1}\alpha_t \\ & + c_t^\top B_t(W_t)^\top(d_t + Q_t^{-1}\alpha_t) + \frac{1}{2}c_t^\top K_t^{-1}c_t. \end{aligned}$$

Theorem 2 then follows from computing the Hessian of $J_t(s)$ and summing over t.

References

[1] GTM Research. U.S. Energy Storage Monitor, 2015. `https://www.greentechmedia.com/research/subscription/u.s.-energy-storage-monitor`.

[2] J. Qin, H.-I. Su, and R. Rajagopal. Storage in risk limiting dispatch: Control and approximation. In *2013 American Control Conference*, 2013.

[3] A. ParandehGheibi, M. Roozbehani, M. A. Dahleh, and A. Ozdaglar. The value of storage in securing reliability and mitigating risk in energy systems. *Energy Systems*, 6(1):129–152, 2015.

[4] B. Nykvist and M. Nilsson. Rapidly falling costs of battery packs for electric vehicles. *Nature Climate Change*, 5:329–332, 2015.

[5] D. S. Callaway, M. Fowlie, and G. McCormick. Location, location, location?: What drives variation in the marginal benefits of renewable energy and demand-side efficiency. *Manuscript: UC Berkeley*, 2015.

[6] S. Bose and E. Bitar. Variability and the locational marginal value of energy storage. In *53rd IEEE Conference on Decision and Control*, 2014.

[7] C. Thrampoulidis, S. Bose, and B. Hassibi. Optimal placement of distributed energy storage in power networks. *IEEE Transactions on Automatic Control*, 61(2):416–429, 2016.

[8] S. Wogrin and D. F. Gayme. Optimizing storage siting, sizing, and technology portfolios in transmission-constrained networks. *IEEE Transactions on Power Systems*, 30(6):3304–3313, 2015.

[9] H. Pandzic, Y. Wang, T. Qiu, Y. Dvorkin, and D. S. Kirschen. Near-optimal method for siting and sizing of distributed storage in a transmission network. *IEEE Transactions on Power Systems*, 30(5):2288–2300, 2015.

[10] E. Sjodin, D. F. Gayme, and U. Topcu. Risk-mitigated optimal power flow for wind powered grids. In *2012 American Control Conference*, 2012.

[11] J. Qin and R. Rajagopal. Dynamic programming solution to distributed storage operation and design. In *2013 IEEE Power & Energy Society General Meeting*, 2013.

[12] S. Bose, D. F. Gayme, U. Topcu, and K. M. Chandy. Optimal placement of energy storage in the grid. In *51st IEEE Conference on Decision and Control*, 2012.

[13] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher. An analysis of approximations for maximizing submodular set functions–i. *Mathematical Programming*, 14:265–294, 1978.

[14] S. Khuller, A. Moss, and J. S. Naor. The budgeted maximum coverage problem. *Information Processing Letters*, 70(1):39–45, 1999.

[15] K. Dvijotham, M. Chertkov, and S. Backhaus. Storage sizing and placement through operational and uncertainty-aware simulations. In *47th Hawaii International Conference on System Sciences*, 2014.

[16] B. Stott, J. Jardim, and O. Alsac. DC power flow revisited. *IEEE Transactions on Power Systems*, 24(3):1290–1300, 2009.

[17] A. J. Wood and B. F. Wollenberg. *Power Generation, Operation, and Control*. John Wiley & Sons, 2012.

[18] D. G Luenberger and Y. Ye. *Linear and nonlinear programming*, volume 2. Springer, 1984.

[19] I. Yang, S. Burden, R. Rajagopal, S. S. Sastry, and C. J. Tomlin. Approximation algorithms for optimization of combinatorial dynamical systems. *IEEE Transactions on Automatic Control*, to appear, 2016.

[20] A. Bemporad, M. Morari, V. Dua, and E. N. Pistikopoulos. The explicit linear quadratic regulator for constrained systems. *Automatica*, 38(1):3–20, 2002.

[21] Y. Ji, L. Tong, and R. J. Thomas. Probabilistic forecast of real-time LMP and network congestion. 2015.

[22] Endre Boros and Peter L. Hammer. Pseudo-Boolean optimization. *Discrete Applied Mathematics*, 123(1-3):155–225, nov 2002.

[23] B. C. Eaves and R. M. Freund. Optimal scaling of balls and polyhedra. *Mathematical Programming*, 23(1):138–147, 1982.

[24] M. Sviridenko. A note on maximizing a submodular set function subject to a knapsack constraint. *Operations Research Letters*, 32:41–43, 2004.

[25] A. Krause and C. Guestrin. A note on the budgeted maximization on submodular functions. Technical report, CMU-CALD-05-103, Carnegie Mellon University, 2005.

[26] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas. MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Transactions on Power Systems*, 26(1):12–19, 2011.