Submodularity of Storage Placement Optimization in Power Networks

Junjie Qin, Insoon Yang, and Ram Rajagopal

***Abstract*—In this paper, we consider the problem of placing energy storage resources in a power network when all storage devices are optimally controlled to minimize system-wide costs. We propose a discrete optimization framework to accurately model heterogeneous storage capital and installation costs as these fixed costs account for the largest cost component in most grid-scale storage projects. Identifying an optimal placement strategy is challenging due to the combinatorial nature of such placement problems, and the spatial and temporal transfers of energy via transmission lines and distributed storage devices. To develop a scalable near-optimal placement strategy with a performance guarantee, we characterize a tight condition under which the placement value function is submodular by exploiting our duality-based analytical characterization of the optimal cost and prices. The proposed polyhedral analysis of a parametric economic dispatch problem with optimal storage control also leads to a simple but rigorous verification method for submodularity, and the novel insight that the spatiotemporal congestion pattern of a power network is critical to submodularity. A modified greedy algorithm provides a $(1-1/e)$-optimal placement solution and can be extended to obtain risk-aware placement strategies when submodularity is verified.**

***Index Terms*—Energy storage, greedy algorithms, optimization, power grids, power system economics, submodularity.**

Manuscript received February 17, 2018; revised July 23, 2018 and September 28, 2018; accepted October 18, 2018. Date of publication November 20, 2018; date of current version July 26, 2019. This work was presented at 55th IEEE Conference on Decision and Control, December 12–14, 2016, Las Vegas, NV, USA. The work of J. Qin was supported in part by the Satre family fellowship. The work of I. Yang was supported in part by National Science Foundation (CNS1657100), in part by Creative-Pioneering Researchers Program through Seoul National University (SNU) (0668-20180252), and in part by SNU Big Data Institute (0660-20180009). The work of R. Rajagopal was supported in part by NSF CAREER Award (ECCS1554178). Recommended by Associate Editor Roland P. Malhame. *(Corresponding author: Insoon Yang.)*

J. Qin is with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720-1776 USA (e-mail: qinj@berkeley.edu).

I. Yang is with the Department of Electrical and Computer Engineering/Automation and Systems Research Institute, Seoul National University, Seoul 08826, South Korea (e-mail: insoonyang@snu.ac.kr).

R. Rajagopal is with the Civil and Environmental Engineering Department, Stanford University, Stanford, CA 94305-4020 USA (e-mail: ramr@stanford.edu).

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

Digital Object Identifier 10.1109/TAC.2018.2882489

I. Introduction

ENERGY storage devices, ranging from batteries to hydropower plants, are considered to play a key role in improving the reliability, efficiency, and resilience of power systems. A strong growth of energy storage installation has occurred around the world in recent years. For example, the total storage deployment in the United States increased by 243% in power capacity and 188% in energy capacity during 2014–2015 [2]. This was driven, in part, by an increasing need for energy storage in modern power systems to compensate for the variability of wind and solar energy sources. The value of storage in the power grid under a large penetration of renewable energy sources has been quantified in several studies (e.g., [3]–[5]). It has also been claimed that energy storage can be used to shift load and support frequency regulation to enhance the system efficiency and reliability [6], [7]. Another primary driving force has been the rapidly decreasing cost of storage devices, especially batteries, as a consequence of growing public and commercial interests in electric vehicles [8].

The bulk of newly deployed storage devices has been front-of-meter deployment. In 2015, 85% of storage deployment in the United States was front-of-meter utility- or grid-scale storage [9]. 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 congestion [10], [11]. Therefore, there is a strong need for efficient strategies to place storage devices in power networks.

A. Related Work

A majority of prior studies has treated the energy storage placement problem together with the storage sizing problem. This line of research has led to *continuous optimization* formulations. For example, Thrampoulidis *et al.* [12] studied the allocation of a fixed total storage capacity over a power network to minimize the generation cost. By optimizing the capacity of each storage device together with the decision variables in economic dispatch, they obtained a structural characterization of the optimal allocation. This characterization eliminated the need to place storage at certain generation-only buses. Pandzic *et al.* [13] and Wogrin and Gayme [14] emphasized the multilevel nature of the placement problem. Their analyses also provided useful insights on the effect of congestion, wind penetration, and storage service types. Sjödin *et al.* [15] employed chance constraints to limit the system operation risk generated by variable renewable energy sources and jointly optimized generator dispatch and storage control and sizing. Kraining *et al.* [16]

extended their model predictive control-based storage operation optimization to address the problem of allocating storage capacities over the network. Qin and Rajagopal [17] derived a constrained linear-quadratic-Gaussian controller for distributed storage devices under uncertainty and formulated a storage-sizing problem as a convex program. These studies all used linearized dc approximation of ac power flow to avoid nonlinearity and complexity issues. Castillo and Gayme [18] studied the storage allocation problem in the presence of line losses. This led to a nonconvex quadratic-constrained quadratic program (QP) for which exact convex relaxations based on semidefinite programs and second-order cone programs were developed. Bose *et al.* [19] proposed a semidefinite relaxation approach to the storage placement problem using the ac power flow model and demonstrated its effectiveness through numerical simulations. With an ac power flow model, Castillo and Gayme [20] considered the setup in which storage is operated to maximize the total profit based on locational marginal prices (LMPs). Structural results between the storage decisions and the LMPs were identified. Tang and Low [21] focused on distribution networks by employing a branch flow model and derived the monotonicity properties of the optimal placement solution under the assumption that all load profiles have the same shape.

Another line of research considers the energy storage placement problem as a form of facility location problems (e.g., [22] and the references therein). An example is Qi *et al.* [23], which considers a planning problem for energy storage in the presence of wind power generation. Utilizing a simplified model for power flow, the authors formulate a mixed-integer second-order conic program for uncapacitated storage and propose an approximation scheme for the capacitated storage.

B. Proposed Work and Its Contributions

Departing from the aforementioned continuous optimization approaches, we propose a *discrete optimization* formulation for energy storage placement when all of the storage devices are optimally controlled to minimize the total system-wide cost. This formulation 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 installation and capital costs. Depending on the storage technology used, the installation costs can be as high as the capital costs. 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 differences in installation costs. Furthermore, additional fixed cost components, such as site acquisition costs, could be sensitive to the installation location. 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 an accurate model of considering these cost factors. Additionally, a discrete optimization framework is useful to handle practical scenarios in which fixed-capacity storage devices are to be placed. These advantages are elaborated in Section II-D.

We formulate the placement problem as maximizing the *placement value function*, a set function that represents the value

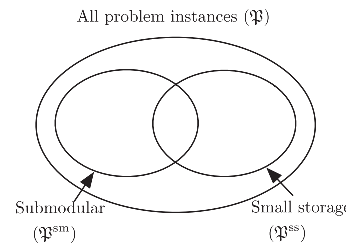


Fig. 1. Venn diagram for the set of all possible storage placement problem instances.

of a storage placement decision, subject to a knapsack constraint that models the budget limit on the aforementioned fixed costs. Unfortunately, this class of problems is NP-hard in general. To overcome this challenge, we identify rich structures of the placement value function. In particular, we characterize conditions under which the placement value function is *submodular*, suggesting that the marginal benefit of adding a storage device decreases as more devices are installed.[1] This submodular structure allows us to employ a greedy algorithm that provides a near-optimal solution with a performance guarantee [26], [27]. However, characterizing conditions under which the placement value function is submodular has remained as an unanswered question [28].

We summarize the contributions and main results of the proposed work 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 an associated multiperiod economic dispatch problem with optimal storage control and its dual problem, we analytically characterize several structural properties of the optimal system-wide cost, energy prices, and storage controls. In particular, we derive LMPs and network congestion prices as piecewise affine functions of the installed storage capacity vector and a closed-form expression of the Hessian of the optimal objective function. Third, exploiting these structural properties, we show that the submodularity of the placement value function is not guaranteed ($\mathfrak{P} \neq \mathfrak{P}^{\rm sm}$ in Fig. 1). This examination provides a unique insight into the effect of spatiotemporal (or network-storage) congestion patterns on submodularity, whereas such an effect is not observed in other applications such as sensor placement [29], [30] and actuator placement [31]–[34]. Fourth, by connecting the sign of Hessian entries to the submodularity of the storage placement value function, we identify a tight condition under which the value function is submodular through a polyhedral characterization of critical regions. Based on this polyhedral analysis, we develop an efficient and rigorous computational procedure to verify the submodularity (i.e., testing whether a particular problem instance belongs to set $\mathfrak{P}^{\rm sm}$ in Fig. 1). Fifth, motivated by the fact that the total storage deployment over the network is still small compared to the total hourly average load or generation, we define a small storage condition (set $\mathfrak{P}^{\rm ss}$ in Fig. 1) under which the verification procedure terminates in one step. The small storage condition can be tested with the problem data by solving a simple linear program. An

[1]Outside of energy storage placement, the concept of submodularity and optimization techniques exploiting submodularity have been used in a number of power system applications. See [24], [25] and references therein.

TABLE I
SUMMARY OF THE ALGORITHMS AND RESULTS

Problem	**Algorithm**	**Performance and efficiency**
Submodularity verification	Checking Hessian: multi-parametric programming (Algorithm 1)	Yes/no answer (exact up to machine precision); Terminates in a finite number of steps; Terminates in one step if the small storage condition holds.
Small storage verification	Polytope containment: linear program (19)	Yes/no answer (exact up to machine precision) via LP
Storage placement	Modified greedy algorithm (Algorithm 2)	$(1-1/e)$-optimal solution if submodularity holds; Polynomial time

extension to risk-aware placement strategies is also discussed for cases with a large penetration of wind and solar energy sources. Our algorithms and results are summarized in Table I.

The remainder of this paper is organized as follows. In Section II, we introduce the problem of jointly optimizing storage placement and control in a power network, and propose a discrete optimization formulation. Section III contains several structural properties of the optimal cost function, prices, and controls. In Section IV, we provide a computational tool to verify the submodularity of the placement value function based on the identified structural properties and a polyhedral analysis. We demonstrate the proposed approach using IEEE test cases in Section V. All the mathematical proofs of our theoretical results are contained in Appendixes A and B.

C. Notation

For a transmission network with n buses and m lines, we use $i \in \mathcal{N} := \{1,\ldots,n\}$ to index the buses, and $j = 1,\ldots,m$ to index the lines. We also use $t \in \mathcal{T} := \{1,\ldots,T\}$ to index the time periods. For a matrix $x \in \mathbb{R}^{n\times T}$ with any given positive integers n and T, we use $x_{i,t}$ to denote its (i,t)th entry, $x_t := (x_{1,t},\ldots,x_{n,t})^\top \in \mathbb{R}^{n\times 1}$ to denote its tth column, and $x_i^\top := (x_{i,1},\ldots,x_{i,T}) \in \mathbb{R}^{1\times T}$ to denote its ith row. For a sequence of matrices A_k's, $k = 1,\ldots,K$, we use $\mathrm{diag}(A_1,\ldots,A_K)$ to denote the block-diagonal matrix, with the diagonal blocks being $A_1,\ldots,A_K$. 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. We use $\mathbf{1} \in \mathbb{R}^{\mathbf{n}}$ to denote the all-one vector and $\mathbf{1_k} \in \mathbb{R}^{\mathbf{n}}$ 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 Model

We begin by considering 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.[2] Under the classical dc approximation of the steady-state ac power flow model[3] [35], the lines are characterized by their reactance and real power flow capacity $\widehat{\ell} \in \mathbb{R}^m$. Let $Y \in \mathbb{R}^{n\times n}$ be the 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 line reactance. Note that $\mathrm{rank}(Y) = n-1$. Taking bus 1 to be the reference bus, we let $\bar{Y} \in \mathbb{R}^{(n-1)\times(n-1)}$ be the submatrix of Y which contains all the entries of Y except its first row and first column. Let $Y^\dagger := \left[\begin{smallmatrix}0 & 0\\ 0 & \bar{Y}^{-1}\end{smallmatrix}\right]$ be the constrained generalized inverse of Y. For each time period $t = 1,\ldots,T$, we can then relate the line flows $f_t \in \mathbb{R}^m$ with the nodal power injection $p_t \in \mathbb{R}^n$ using a linear map $\widehat{H} := \Delta_y M^\top Y^\dagger \in \mathbb{R}^{m\times n}$, i.e., $f_t = \widehat{H} p_t$. Let

$$H := \begin{bmatrix} I \\ -I \end{bmatrix} \widehat{H} \in \mathbb{R}^{2\mathrm{m}\times n} \quad \text{and} \quad \ell := \begin{bmatrix} \widehat{\ell} \\ \widehat{\ell} \end{bmatrix} \in \mathbb{R}^{2\mathrm{m}}$$

with the real power flow capacity vector denoted by $\widehat{\ell}$. The power flow constraints can be compactly expressed as

$$\mathbf{1}^\top \mathbf{p_t} = 0 \tag{1a}$$

$$Hp_t \le \ell \tag{1b}$$

for all time periods $t \in \mathcal{T}$. Equation (1a) enforces net power balance in the network, while (1b) limits the line flows induced by the power injection vector p_t within the line capacities. The matrix H, which models the linear mapping from the nodal injections to the line flows, is commonly referred to as the *shift-factor matrix*.

B. Energy Storage Dynamics

We consider a stylized model of energy storage:[4] for each bus i, the storage's state of charge (SOC) $s_{i,t}$ evolves as

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

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. The initial SOC is assumed to be $s_{i,1} = 0$. Given the storage capacity $x_i \ge 0$, the following constraints model the energy limit of the storage device:

$$0 \le s_{i,t} \le x_i, \quad t \in \mathcal{T}. \tag{3}$$

Note that $x_i = 0$, if there is no storage connected to bus i. Applying (2) recursively, we can express constraint (3) as $0 \le \sum_{\tau=1}^{t} -u_{i,\tau} \le x_i$, $t \in \mathcal{T}$, which can be compactly expressed

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

[3] We focus on the dc power flow model as it has been extensively used in the prior studies on transmission system planning. Nevertheless, extending our analysis to the ac power flow model is an important future research problem.

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

in the following vector form:

$$0 \le L u_i \le x_i \mathbf{1}$$

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 $i \ge j$. The information about the storage dynamics is embedded in the matrix L.

C. Economic Dispatch With Storage Control

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

$$J(x) := \min_{g,u} \quad \sum_{t=1}^{T} C_t(g_t) \tag{4a}$$

$$\text{s.t.} \quad \beta_t : H(g_t + u_t - d_t) \le \ell \qquad \forall t \in \mathcal{T} \tag{4b}$$

$$\gamma_t : \mathbf{1}^\top (g_t + u_t - d_t) = 0 \quad \forall t \in \mathcal{T} \tag{4c}$$

$$\mu_i : L u_i \le x_i \mathbf{1} \qquad \forall i \in \mathcal{N} \tag{4d}$$

$$\nu_i : L u_i \ge 0 \qquad \forall i \in \mathcal{N}. \tag{4e}$$

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-wide generation cost for time period t and is taken to be quadratic as is common in the literature [36], i.e.,

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

where Q_t is a diagonal matrix whose diagonal entries are positive, which models the increasing incremental (marginal) heat rate, and $r_t \in \mathbb{R}^n$ 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) enforce 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 connection, we set $x_i = 0$, and (4d) and (4e) reduce to $u_{i,t} = 0$ for all $t \in \mathcal{T}$. Note that we can obtain an optimal storage control schedule as well as an optimal generator dispatch schedule by solving the multiperiod economic dispatch problem. The optimal value of this problem, denoted by J, is a function of storage capacities x over the network, as storage capacities affect the feasible region of the optimization problem (4) via constraint (4d). In other words, we consider (4) as a parametric program associated with x.

D. Storage Placement as Combinatorial Optimization

The optimal cost of this multiperiod economic dispatch problem depends critically on the storage capacity vector $x \in \mathbb{R}^n$. If there is no storage connected to the network (i.e., $x = 0$), the optimal cost of this multiperiod problem reduces to the sum of the optimal costs of T single-period economic dispatch problems. Conversely, if the storage and line capacities are large enough for every node, the system cost for T periods approaches a limit where, roughly speaking, the cheapest generators across the network and over T periods are used. In this ideal case, marginal generation costs for all time periods are equalized. 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 $\bar{x}_k$ and capital and installation costs c_k, $k = 1, \dots, K$, we want to place the storage devices to minimize the system operation cost with a total budget, denoted by b, for total capital and installation costs.

We proceed to formulate the placement problem as a combinatorial optimization problem. Consider the collection of all *$n \times K$ (bus, storage "type") pairs*[5]

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

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 device with each type can be placed at each bus.[6] 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 $\mathbb{I}(X)\bar{x}$ can be expressed as $(\mathbb{I}(X)\bar{x})_i = \sum_{k:(i,k)\in X} \bar{x}_k$, which is equal to the total storage capacity at bus i. We now introduce a function, $V : 2^\Omega \to \mathbb{R}$, which we call the storage *placement value function*, defined as

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

For each placement decision X, the value $V(X)$ represents the reduction in the minimum T-period total generation cost caused by the optimal control of the storage devices given 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 *discrete optimization* problem of maximizing the placement value function

$$\max_{X\subseteq\Omega} \quad V(X) \tag{5a}$$

$$\text{s.t.} \quad \sum_{k:(i,k)\in X} c_k \le b. \tag{5b}$$

[5]It can be the case that some buses should be ruled out a priori for certain systems. In this case, we can define the set of possible storage placement decisions as $\Omega = \{(i,k) : i \in \tilde{\mathcal{N}}, k = 1, \dots, K\}$, where $\tilde{\mathcal{N}} \subseteq \{1, \dots, n\}$ is the set of buses where placing a storage is possible.

[6]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 represents the rth appearance of the same type of storage devices at the same bus, $r = 1, \dots, R$. In other words, we have $\Omega := \{(i,k') : i = 1, \dots, n,\ k' = (k-1)R + r,\ k = 1, \dots, K,\ r = 1, \dots, R\}$, where the type $k' = (k-1)R + r$ corresponds to the rth appearance of the original type k. This is a classical technique to convert an integer program into a binary program. For 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 the different fixed costs for placing the same type of storage at different buses to take into account site acquisition costs.

We claim that our problem formulation as a discrete optimization has practical advantages over continuous optimization formulations. First, our framework can handle practical scenarios in which fixed-capacity storage devices 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 postprocessing, such as thresholding, to convert the solutions of continuous optimization problems into discrete solutions. However, such postprocessing does not provide a performance guarantee in general, whereas our method directly computes a discrete solution with a provable suboptimality bound. Second, our problem formulation naturally incorporates storage devices' capital and installation costs through the knapsack constraint (5b), which accurately captures the total sum of the capital and installation costs as $\sum_{k:(i,k)\in X} c_k$ and limits it by the budget b. In contrast, it is difficult to expect such a precise regulation in continuous optimization formulations as discussed in Section I-B. Finally, the proposed discrete optimization formulation yields a very simple placement algorithm that only requires an input–output (blackbox) model of a power system. Specifically, our greedy algorithm can use simulations that capture electricity market input–output without using detailed information about the network. This is a notable advantage over continuous optimization formulations, which often require a full network model with complete information (e.g., parameters) about markets to calculate the (sub)gradient of objective functions.

III. Structures of Optimal Cost and Prices

A. Dual Analysis

In order to obtain efficient methods to solve the storage placement problem (5), we establish structural properties of the placement value function through an analytical characterization of the optimal prices.

Consider the standard dual QP of (4):

$$\max_{\lambda,\gamma,\beta,\mu,\nu} \quad \phi(\lambda,\gamma,\beta,\mu,\nu) \tag{6a}$$

$$\text{s.t.} \quad \lambda_t = \gamma_t \mathbf{1} - H^\top \beta_t \quad \forall t \in \mathcal{T} \tag{6b}$$

$$\lambda_i = L^\top(\mu_i - \nu_i) \quad \forall i \in \mathcal{N} \tag{6c}$$

$$\beta, \mu, \nu \geq 0 \tag{6d}$$

where the Lagrange dual function is given by

$$\phi(\lambda,\gamma,\beta,\mu,\nu) := \sum_{t=1}^{T} -\frac{1}{2}(\lambda_t - r_t)^\top Q_t^{-1}(\lambda_t - r_t) + d_t^\top \lambda_t - \ell^\top \beta_t - x^\top \mu_t.$$

Note that the variable $\lambda_{i,t}$ represents the LMP at bus i in period t because, by the first-order optimality condition of (4), we have

$$\nabla_{g_t} C_t(g_t) = \lambda_t, \quad t \in \mathcal{T}.$$

A more detailed derivation can be found in [37].

The standard dual QP (6) can be further simplified. Observe that (4d) and (4e) representing the lower and upper limits of SOC, respectively, 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., $s_{i,t} = (Lu_i)_t = 0$, then it must be the case that $s_{i,t} = (Lu_i)_t < x_i$. Similarly, $(Lu_i)_t = x_i$ 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 with constraint (6c), which is equivalent to $(\mu_i - \nu_i) = L^{-\top}\lambda_i$, we have

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

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

$$\mu_t = (\lambda_{t+1} - \lambda_t)^+ \quad \text{and} \quad \nu_t = (\lambda_{t+1} - \lambda_t)^- \quad \forall t \in \mathcal{T} \tag{7}$$

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

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

$$\max_{\lambda,\gamma,\beta} \quad \hat{\phi}(\lambda,\beta) \tag{8a}$$

$$\text{s.t.} \quad \lambda_t = \gamma_t \mathbf{1} - H^\top \beta_t \quad \forall t \in \mathcal{T} \tag{8b}$$

$$\beta \geq 0 \tag{8c}$$

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

$$\hat{\phi}(\lambda,\beta) := \sum_{t=1}^{T} -\frac{1}{2}(\lambda_t - r_t)^\top Q_t^{-1}(\lambda_t - r_t) + d_t^\top \lambda_t - \ell^\top \beta_t - x^\top (\lambda_{t+1} - \lambda_t)^+.$$

By strong duality, we can characterize the function $J(x)$ via a sensitivity analysis of the primal-dual pair (4) and (8). Let $(g_t^\star(x), u_t^\star(x), \lambda^\star(x), \gamma^\star(x), \beta^\star(x))$ be a pair of primal and dual solutions to (4) and (8) for a given capacity vector x. We focus on x values which induce nondegenerate solutions of (4) by assuming the following linear independence constraint qualification (LICQ) for the rest of this paper:

Assumption 1 (Flow LICQ): For each $t \in \mathcal{T}$, let $H_t^{\rm net}$ be the collection of H's rows corresponding to the congested (oriented) lines for the flow induced by $(g_t^\star(x), u_t^\star(x))$, when at least one congested line exists in period t.[7] Then, $H_t^{\rm net}$ is of full row rank for each $t \in \mathcal{T}$.

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

Proposition 1 (Uniqueness of prices): For each fixed $x \in \mathbb{R}_+^n$, the optimal dual variables $\lambda^\star(x)$ and $\gamma^\star(x)$ are unique. Furthermore, if Assumption 1 holds, then $(\lambda^\star(x), \gamma^\star(x), \beta^\star(x))$ is the unique solution to (8).

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

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

[7] The analytical expression of matrix $H_t^{\rm net}$ is provided in (11).

by

$$\nabla_x J(x) = -\sum_{t=1}^{T} \left(\lambda^\star_{t+1}(x) - \lambda^\star_t(x)\right)^+ \tag{9}$$

where $\lambda^\star_{T+1}(x) := 0$. Consequently, the optimal cost function $J(x)$ is nonincreasing in x_i for each $i \in \mathcal{N}$.

In Bose and Bitar [11], the term *locational marginal value of storage* is used to refer to the quantity $-\nabla_x J(x)$, 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 (9) for the case where the marginal cost of generation and marginal benefit of consumption are both constants. In fact, the expression (9) holds for any smooth convex cost function under mild regularity assumptions in the standard sensitivity theorem of nonlinear programming.

When the cost function is nonlinear and the size of storage to be placed is far from infinitesimal, the first-order approximation of the value function using the gradient formula (9) may not be accurate.[8] We now proceed to obtain a finer characterization of the optimal cost $J(x)$ by investigating its higher order derivatives in the following section. An immediate observation is that $J(x)$ is convex in x.

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

B. Optimal Prices and the Second-Order Sensitivity

Given that the objective function is quadratic, we expect the curvature (or second-order) information summarized by the Hessian matrix $\nabla^2_{xx} J(x)$ together with the gradient information discussed in Lemma 1 would provide a sufficient characterization of the optimal system-wide cost function $J(x)$. This is confirmed by the following result.

Lemma 3: The function $J(x)$ 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 $J(x)$ is quadratic, the optimal LMP vector $\lambda^\star(x)$ is affine in x.

Remark 1: The polytopes in Lemma 3 are referred to as *critical regions* in the literature of multiparametric quadratic programming (e.g., [39], [40]). In our context, each critical region is defined as a set of x 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 optimum. See Theorem 4 for a detailed characterization of the critical regions.

By 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 $z^{\rm st}_{i,t}(x) = 1$ if the constraint $(Lu_i)_t \le x_i$ is binding at the optimum, $z^{\rm st}_{i,t}(x) = -1$ if the constraint $(Lu_i)_t \ge 0$ is binding at the optimum, and $z^{\rm st}_{i,t}(x) = 0$ otherwise. In other words, $z^{\rm st}$ represents the storage congestion pattern. Under strict complementary slackness, we use (7) to obtain

$$z^{\rm st}_{i,t} := z^{\rm st}_{i,t}(x) = \begin{cases} 1, & \text{if } \lambda^\star_{i,t+1}(x) - \lambda^\star_{i,t}(x) > 0 \\ -1, & \text{if } \lambda^\star_{i,t+1}(x) - \lambda^\star_{i,t}(x) < 0 \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

We now let $\mathcal{E}^{\rm C}_t \subset \{1, \ldots, 2\,{\rm m}\}$ denote the set of transmission lines that are congested at the solution in period t and $m_t := |\mathcal{E}^{\rm C}_t|$ denote the number of congested lines. We define the selection matrix $z^{\rm net}_t \in \mathbb{R}^{m_t \times 2\,{\rm m}}$ as

$$(z^{\rm net}_t)_{i,j} := (z^{\rm net}_t(x))_{i,j} = \begin{cases} 1, & \text{if the } i\text{th element in } \mathcal{E}^{\rm C}_t \text{ is } j \\ 0, & \text{otherwise} \end{cases}$$

for $i = 1, \ldots, m_t$ and $j = 1, \ldots, 2\,{\rm m}$, and the *shift factor matrix for congested lines* as

$$H^{\rm net}_t := z^{\rm net}_t H. \tag{11}$$

Note that $z^{\rm net}_t = 0$ if all lines are uncongested in period t.

Theorem 1: In the critical region containing a given storage capacity vector x, where the storage and network congestion patterns are represented by $z^{\rm st}_t$ and $z^{\rm net}_t$, $t \in \mathcal{T}$, the optimal LMPs are affine in x. In other words, there exist matrix $W_t(z^{\rm net}, z^{\rm st}) \in \mathbb{R}^{n \times n}$ and vector $\bar{\lambda}_t(z^{\rm net}, z^{\rm st}) \in \mathbb{R}^n$ for each $t \in \mathcal{T}$ such that

$$\lambda^\star_t(x) = W_t(z^{\rm net}, z^{\rm st})x + \bar{\lambda}_t(z^{\rm net}, z^{\rm st}). \tag{12}$$

As an useful byproduct of Theorem 1, we can obtain closed-form expressions of the (reference) energy price $\gamma^\star_t = \mathbf{1}_1^\top \lambda^\star_t$ and the congestion price $\beta^\star_t$ as a function of the storage capacity vector x.

Corollary 1: Under the setting of Theorem 1, $\gamma^\star_t(x)$ and $\beta^\star_t(x)$ are affine functions of x in the critical region containing x. More explicitly, there exist $P_t(z^{\rm net}, z^{\rm st}) \in \mathbb{R}^{m_t \times n}$ and $\bar{\beta}_t(z^{\rm net}, z^{\rm st}) \in \mathbb{R}^{2\,{\rm m}}$ for each $t \in \mathcal{T}$ such that

$$\gamma^\star_t(x) = \mathbf{1}_1^\top \left(W_t(z^{\rm net}, z^{\rm st})x + \bar{\lambda}_t(z^{\rm net}, z^{\rm st})\right) \tag{13}$$

$$\beta^\star_t(x) = {z^{\rm net}_t}^\top P_t(z^{\rm net}, z^{\rm st})x + \bar{\beta}_t(z^{\rm net}, z^{\rm st}). \tag{14}$$

The Hessian of the optimal cost function plays a critical role in studying the submodularity of the storage placement value function as we will see in Section IV. Using Theorem 1 and Lemma 1, we can obtain a structural characterization (and a closed form expression) for the Hessian of $J(x)$ as follows.

Theorem 2: The optimal cost function $J(x)$ is twice differentiable almost everywhere with respect to the Lebesgue measure on $\mathbb{R}^n_+$. Furthermore, storage capacities x and x' that share the same congestion pattern, i.e., $z^{\rm net}(x) = z^{\rm net}(x')$ and $z^{\rm st}(x) = z^{\rm st}(x')$, have the same Hessian, i.e.,

$$\nabla^2_{xx} J(x) = \nabla^2_{xx} J(x')$$

provided that both $\nabla^2_{xx} J(x)$ and $\nabla^2_{xx} J(x')$ exist.

Remark 2: Although Theorem 1, Corollary 1, and Theorem 2 are stated as a structural characterization of the optimal prices $(\lambda(x), \beta(x), \gamma(x))$ and the Hessian $\nabla^2_{xx} J(x)$ with respect to x, explicit expressions for these functions are given in the proofs in Appendix A. See (21)–(23). These expressions allow us to efficiently evaluate the prices and the Hessian of the optimal cost given the storage capacity x.

[8]The first-order approximation can be used for storage placement. See [38] for an approximation algorithm based on it that gives a solution with $0.6 - 0.7$ *a posteriori* suboptimality bound.

IV. Submodularity of Placement Value Function

Equipped with the structural properties of the optimal cost function $J(x)$, we now characterize the storage placement function $V(X)$ defined in Section II-D. Recall that the set 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 contains all admissible (bus, storage type) pairs. In particular, we characterize the 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 $\mathbf{e} \in \Omega \setminus Y$,

$$F(X \cup \{\mathbf{e}\}) - F(X) \geq F(Y \cup \{\mathbf{e}\}) - F(Y). \quad (15)$$

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

$$F(X \cup \{\mathbf{e}\}) \geq F(X). \quad (16)$$

In our case, (16) implies that the marginal benefit of installing a new storage device is nonnegative and (15) states that this marginal benefit should diminish as more storage devices are connected to the system. Evidently, the nondecreasing property of $V(X)$ follows from the fact that $J(x)$ is nonincreasing (see Lemma 1). To check whether $V(X)$ is submodular, it is useful 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 well known that the following lemma provides a necessary and sufficient condition for submodularity [41].

Lemma 4: A set function $F : 2^{\Omega} \mapsto \mathbb{R}$ is submodular if and only if

$$\mathrm{D}_{\mathbf{e}} \left(\mathrm{D}_{\mathbf{e}'} F(X)\right) \leq 0 \quad \forall \mathbf{e}, \mathbf{e}' \in \Omega \text{ s.t. } \mathbf{e} \neq \mathbf{e}' \quad \forall X \subseteq \Omega. \quad (17)$$

We relate the submodularity of $V(X)$ to the sign of the Hessian entries of $J(x)$ as follows.

Theorem 3 (Sufficient condition for submodularity): The placement value function $V : 2^{\Omega} \to \mathbb{R}$ is submodular if

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

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

Theorem 3 provides a sufficient condition for the submodularity of V by just checking the sign of the Hessian entries, which can be computed using Theorem 2. This characterization is tight in the following sense.

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

This corollary is a partial converse of Theorem 3. Even when the point x that results in negative Hessian entries is contained in $\mathcal{X}$, 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) enough that its contribution to the discrete derivative is outweighed by the contribution from other critical regions with positive Hessian entries.

Theorem 3 and Corollary 2 establish a strong connection between submodularity of the storage value function $V(X)$ and the sign of the Hessian entries of the optimal cost function $J(x)$. This allows us to understand the submodularity condition through an economic interpretation of the Hessian entries.

Remark 3 (Submodularity and substitutability): Define a continuous version of the storage value function as $v(x) = J(0) - J(x)$. For any buses $i, j \in \mathcal{N}$, the Hessian entry (i.e., cross derivative) $\frac{\partial^2 v(x)}{\partial x_j \partial x_i} = \frac{\partial}{\partial x_j}(\frac{\partial v(x)}{\partial x_i})$ is the rate of change of the locational marginal value of storage at bus i when the storage capacity at bus j is changed. Thus, the condition in Theorem 3 has the following economic interpretation:

1) For $i = j$, storage at bus i has diminishing return.
2) For $i \neq j$, storage at bus j *substitutes* storage at bus i.

The convexity of the optimal cost function $J(x)$ (Lemma 2) establishes that the diagonal entries of the Hessian matrix $\nabla^2_{xx} J(x)$ are always nonnegative. The conditions for submodularity in Theorem 3 also require all off-diagonal entries of the Hessian matrix to be nonnegative, which does not follow from the properties that have already been established for the optimal value function $J(x)$.

A. Two-Bus Network

To gain insights into the sign of the off-diagonal Hessian entries, we consider a two-bus example with $T = 3$ together with synthetic load profiles. We demonstrate that, for the same network, submodularity may hold for certain load profiles but fail to hold for some other load profiles. For simplicity, we use cost functions $C_t(g_t) = \frac{1}{2} g_t^{\top} g_t$, for $t = 1, \ldots, 3$, i.e., $Q_t \equiv I \in \mathbb{R}^{2\times 2}$ and $r_t \equiv 0$. Given a time-varying demand profile over the network, if neither storage nor line capacity is constraining, then the economic dispatch solution exhibits a form of "water-filling" behavior where the optimal flows result in equalized generation from each bus for all time periods. We also notice that $g_t^{\star} = \lambda_t^{\star}$ for this cost function, by the first-order optimality condition of (4).

We now investigate the property of $J(x)$ and the optimal primal and dual solutions of the multiperiod economic dispatch problem for all storage capacities x in the region $\mathcal{X} = [0, 1] \times [0, 1]$. The line capacity is set to be 0.5. We consider the following two cases: One (Case A) is commonly observed in simulations where all of the critical regions inside of $\mathcal{X}$ have $J(x)$ with only nonnegative Hessian entries, and the other (Case B) is specially designed so that one of the critical regions has negative off-diagonal Hessian entries

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

The critical regions for these cases are depicted in Fig. 2.

For each critical region, we obtain the expression of the optimal cost function $J(x)$ (which includes its quadratic and linear coefficients); for a set of points on the mesh grid inside of each critical region, we solve the multiperiod economic

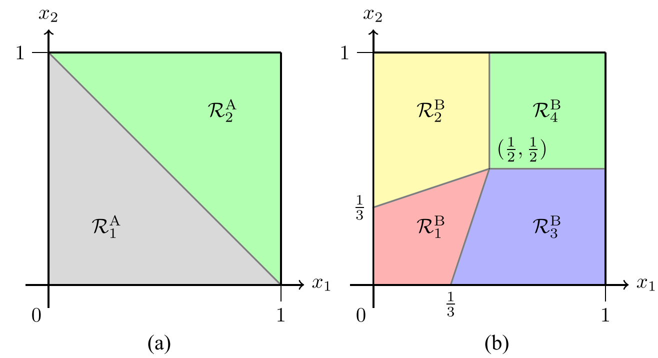


Fig. 2. Critical regions for the two-bus examples. In the figure with a slight abuse of notation, we use $\mathcal{R}_r^{\mathrm{A}}$ and $\mathcal{R}_r^{\mathrm{B}}$ to denote the rth critical region for each case. (a) Case A. (b) Case B.

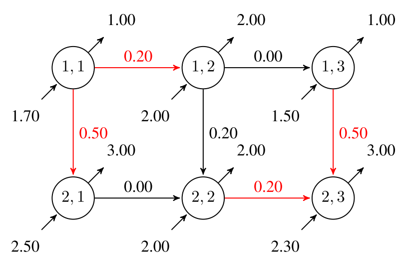


Fig. 3. Optimal flow for the case with negative Hessian entries.

dispatch problem and obtain the optimal primal dual solution. In all six critical regions across these two cases, only the red region in Case B, i.e., $\mathcal{R}_1^{\mathrm{B}}$, has negative Hessian entries. This suggests that submodularity holds for load profile d^{A}, but fails to hold for load profile d^{B}. *Therefore, submodularity does not hold in general.* Due to the page limit, we focus on region $\mathcal{R}_1^{\mathrm{B}}$ for the remainder of this section. The optimal cost function in the critical region is given by $J(x) = \frac{1}{2}x^\top \begin{bmatrix} 1.5 & -0.5 \\ -0.5 & 1.5 \end{bmatrix} x + \begin{bmatrix} -0.5 & -0.5 \end{bmatrix} x + 12.5.$

Consider a particular storage capacity vector $\tilde{x} = [0.2,\, 0.2]^\top \in \mathcal{R}_1^{\mathrm{B}}$. The solution of (4) for this given storage capacity is depicted in Fig. 3 by exploiting the following observations: First, storage can be thought of as an intertemporal link that sends energy into the future. Second, the multiperiod economic dispatch problem is a form of generalized network flow problems on a *time-extended graph* where storage edges connect the graph representations of the power network for consecutive time periods [42], [43]. In Fig. 3, each node in the graph represents a (bus, time period) pair. The vertical edges of the graph represent the transmission lines, while the horizontal edges represent the power stored for future use by each storage device. 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 vertical edge is the optimal flow sent through the line; for each horizontal storage edge, the value on it represents the amount of energy stored at the end of last time period. Red edges are congested at the optimal solution.

A key observation in this special case can be made as follows: Given the load and network congestion pattern, the use of storage links is through the following spatialtemporal flow path $(1,1) \to (1,2) \to (2,2) \to (2,3)$. In this flow path, the storage capacity at bus 1 and storage capacity at bus 2 *complements* each other, instead of having a substitution effect (cf. Theorem 3 and Remark 3). We also notice that 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. This would be unusual in practical settings, especially in planning scenarios, as the LMPs are often driven by load profiles. If such a phenomenon were to occur in practice, it would indicate that i) 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 drops to its valley, and ii) the transmission link between the two buses is weak and congested so that the optimal/equilibrium prices still follow such patterns. Given that each load bus in a transmission network often represents a collection of smaller loads, the condition in i) means that the aggregates of these small loads follow very different temporal patterns at different locations in the network. Furthermore, if we were concerned about determining which transmission lines to strengthen, conditions i) and ii) are 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.

B. Verification of Submodularity

1) General Case: In general, the storage placement value function V is not guaranteed to be submodular. More specifically, its submodularity depends on problem settings, particularly the load and network data. Thus, it is of interest to develop an efficient computation procedure which verifies the submodularity of V for any given problem instance.

This is a challenging task, as verifying the submodularity of V by definition involves checking an exponential number of inequalities. Theorem 3 reduces this problem to check the sign of Hessian entries of a continuous function $J(x)$ on $\mathcal{X}$. Theorem 2 provides an expression of the Hessian for almost every $x \in \mathcal{X}$, which is invariant in each critical region. Therefore, it is sufficient to evaluate the Hessian once per critical region in $\mathcal{X}$. We now address how we could iterate over the critical regions.

We begin by providing an explicit polyhedral characterization of the critical region that contains almost every capacity vector x. When x is on the boundary of two critical regions, strict complementary slackness fails to hold and, in general, one may obtain a degenerate solution. However, the set of boundary points has a Lebesgue measure of 0 and, therefore, does not contribute to our submodularity characterization as shown in the proof of Theorem 3. Thus, we can ignore these points for the rest of our discussion. Upon solving the multiperiod economic dispatch problem at x, we can identify the set of binding constraints and the associated z_t^{st} and z_t^{net} for $t \in \mathcal{T}$. The critical region containing x can then be expressed as the set of storage capacity vectors where the storage and network congestion states are invariant.

Theorem 4: Given $z_t^{\mathrm{st}}(x)$ and $z_t^{\mathrm{net}}(x), t \in \mathcal{T}$, evaluated at an arbitrary $x \in \mathcal{X}$ (except for a set of measure zero), the critical region containing x, denoted by $\mathcal{R}_x$, is an open convex polytope defined as the set of $\hat{x} \in \mathbb{R}_+^n$ satisfying the linear inequalities

$$\lambda^\star_{i,t+1}(\hat{x}) - \lambda^\star_{i,t}(\hat{x}) > 0 \quad \forall (i,t) \in \mathcal{N} \times \mathcal{T} \text{ s.t. } z^{\text{st}}_{i,t}(x) = 1$$
$$\lambda^\star_{i,t+1}(\hat{x}) - \lambda^\star_{i,t}(\hat{x}) < 0 \quad \forall (i,t) \in \mathcal{N} \times \mathcal{T} \text{ s.t. } z^{\text{st}}_{i,t}(x) = -1$$
$$\beta^\star_{j,t}(\hat{x}) > 0 \quad \forall (j,t) \in \mathcal{E}^{\text{C}}_t \times \mathcal{T}.$$

In other words, for each $\hat{x} \in \mathcal{R}_x$, the associated storage congestion pattern $z^{\text{st}}_t(\hat{x})$ and network congestion pattern $z^{\text{net}}_t(\hat{x})$ satisfy $z^{\text{st}}_t(\hat{x}) = z^{\text{st}}_t(x)$ and $z^{\text{net}}_t(\hat{x}) = z^{\text{net}}_t(x)$, $t \in \mathcal{T}$.

Given this polyhedral characterization of critical regions, the iterative construction of all critical regions and the complexity of this process follow from the standard practice of multiparametric quadratic programming. Geometric approaches [39] and combinatorial approaches [44] have been proposed for this task. Improvements for both types of approaches have also been developed [45]–[48]. While the basic idea of recursively partitioning the parameter space into critical regions is simple (for which a brief description is provided in Appendix C), a practical implementation of the procedure benefits from the techniques proposed in aforementioned references. Therefore, it is recommended to use a mature software tool such as MPT [49] to construct the critical regions. Both the basic implementation described in Appendix C and the implementation provided by MPT are guaranteed to terminate in a finite number of steps, though the total number of critical regions is exponential in the number of buses in the worst case. However, as we will soon see, for many practical storage placement problems where the storage size is relatively small compared to the total system load (cf., Definition 3), the procedure for constructing critical regions terminates in one step (cf., Lemma 5).

2) Small Storage: The rest of this section is devoted to a special case that garners a substantial amount of practical interest as the amount of storage to be placed is usually small *compared to the total load and generation in the network*. For example, the total power capacity of new battery storage installation in the United States was 221 MW in 2015 [2] while the average generation in the same year was 467 GW.

Definition 3: A storage placement problem is said to satisfy the *small storage condition*, if the capacity region $\mathcal{X}$ of interest is a subset of the closure of the critical region containing $x = 0$, i.e., $\mathcal{X} \subseteq \bar{\mathcal{R}}_0$, with $\mathcal{R}_0$ defined in Theorem 4.

Remark 4: The small storage condition is a condition on how much *new* storage is to be placed. In the bulk of the paper, we have assumed that the existing storage capacity is 0 for convenience. In practice, there may be a nonzero storage capacity $\tilde{x} \in \mathbb{R}_+^n$ in the system prior to the planning study. Thus, the total storage capacity after placing $x \in \mathbb{R}_+^n$ is $\tilde{x} + x$. All our technical results remain valid in this case. The small storage condition in the case of nonzero initial capacity $\tilde{x}$ is defined based on the critical region containing $\tilde{x}$ instead of 0.

The immediate consequence of this condition for the purpose of verifying submodularity is that there is only one relevant critical region. Therefore, the following result holds [39].

Lemma 5: Under the small storage condition, the multiparametric programming procedure to construct all critical regions terminates in one step.

Furthermore, 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 z^{net}_t in the base case where no storage has been installed. In other words, we can certify submodularity by simply using the solutions of the single-period economic dispatch problems for $t \in \mathcal{T}$.

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

Another consequence of the small storage condition is that it is possible to obtain much simplified expressions for the optimal prices and the Hessian as a function of the storage capacity vector x. These expressions allow us to gain insight on the role played by the spatialtemporal congestion pattern induced by any given storage capacity vector. We discuss these results in Appendix D.

To verify the small storage condition, we need to check whether the polytope $\bar{\mathcal{R}}_0$ contains the box $\mathcal{X} = [0, \bar{x}^{\max}]^n$. Instead of checking whether all 2^n vertices of the box belong to the polytope, we can simply compare the optimal value of the set containment problem $\min\{\rho \in \mathbb{R}_+ : \mathcal{X} \subseteq \rho\mathcal{R}_0\}$ with 1, where the scaled set $\rho\mathcal{R}_0$ is $\{x : (1/\rho)x \in \mathcal{R}_0\}$. Write $\bar{\mathcal{R}}_0$ as $\{x : Px \leq q\}$. This set containment problem can be formulated as the following linear program [50]:

$$\begin{aligned} \min_{\rho,\Lambda} \quad & \rho \\ \text{s.t.} \quad & \Lambda[I,\ -I]^\top = P \\ & \Lambda[\bar{x}^{\max}\mathbf{1}^\top, 0^\top]^\top \leq \rho q \\ & \rho, \Lambda \geq 0. \end{aligned} \tag{18}$$

Finally, we note that the small storage condition is equivalent to requiring that installing the storage does not affect congestion patterns for the planning scenario considered. This may not hold for some storage placement settings in which case the full-blown multiparametric programming procedure described previously should be used for verifying submodularity. The overall process of verifying submodularity for any given problem instance is summarized in Algorithm 1.

C. Modified Greedy Algorithms

If the placement value function is verified to be submodular, we can employ a (modified) greedy algorithm to obtain a near-optimal solution. In particular, Nemhauser *et al.* [26] show that a standard greedy algorithm gives a $(1 - \frac{1}{e})$-approximation of an optimal solution when maximizing monotone submodular functions subject to a cardinality or matroid constraint. This algorithm is applicable to our problem when there is only one type of storage devices, i.e., $K = 1$.

When multitype devices are used ($K > 1$), this standard greedy algorithm may not fully use the diminishing return property as the knapsack (budget) constraint (5b) can cause it become stuck at an unreasonable solution. However, a modification of the greedy algorithm is shown to achieve the same performance guarantee by considering not only the marginal benefit but also the placement cost c_k [51], [52]. Specifically, the modified algorithm optimizes the *incremental benefit-cost*

Algorithm 1: Verification of submodularity.

Verifying small storage condition by solving LP (18);
if *small storage condition holds* **then**
  Calculate $\nabla^2_{xx} J(x)$ for critical region $\mathcal{R}_0$ using (23)
  if $\left(\nabla^2_{xx} J(x)\right)_{i,j} \geq 0$ *for all* i, j **then**
    return *Yes*
  else
    return *No*
  end
else
  Construct all critical regions;
  foreach *critical region* **do**
    Calculate $\nabla^2_{xx} J(x)$ for the critical region using (24);
    if $\left(\nabla^2_{xx} J(x)\right)_{i,j} < 0$ *for some* i, j **then**
      return *No*
    end
  end
  return *Yes*
end

Algorithm 2: Modified greedy placement algorithm.

1 $X_1 \leftarrow \text{argmax}\{V(X) : |X| \leq 2, \sum_{k:(i,k)\in X} c_k \leq b\}$;
2 $X_2 \leftarrow \emptyset$;
3 **foreach** $X \subseteq \Omega$ *s.t.* $|X| = 3$, $\sum_{k:(i,k)\in X} c_k \leq b$ **do**
4   Candidates $\leftarrow \Omega \setminus X$;
5   **while** Candidates $\neq \emptyset$ **do**
6     $\mathbf{e} \leftarrow \text{argmax}_{(i,k)\in\text{Candidates}} \frac{V(X\cup\{(i,k)\})-V(X)}{c_k}$;
7   **if** $\sum_{k:(i,k)\in X\cup\{\mathbf{e}\}} c_k \leq b$ **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 $X^\star \leftarrow \text{argmax}_{X\in\{X_1,X_2\}} V(X)$;

ratio $[V(X \cup \{(i,k)\}) - V(X)]/c_k$ to reduce the possibility of terminating without finding low-cost placement configurations unlike the benefit-only standard greedy algorithm.

This algorithm also uses the partial enumeration heuristic proposed by Khuller *et al.* [27], which enumerates all subsets of up to three elements. Its details are presented in Algorithm 2. 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 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. This polynomial-time algorithm $O(|\Omega|^5)$ computational complexity can also be implemented in a distributed fashion using the parallelizable method proposed in [53].

An interesting feature of the proposed algorithm is that its dependence on the operation model (4) is only though the placement value function V. Therefore, we may substitute our simple operation model with one that captures more detailed characteristics of the storage technologies, power flows, and generation costs (e.g., start-up costs as considered in unit commitment problems) or even with a black-box simulator. The same greedy algorithm can be directly applied to the setting with the substituted model. However, our theoretical performance guarantees need to be extended to be useful in that setting.

D. Risk-Aware Placement

The uncertainty generated by the widespread penetration of variable renewable energy sources (including wind and solar energy) is substantial in placement decision-making. In this section, we explicitly consider the stochasticity of renewable generation or equivalently net demand d. To this end, we now view the net demand vector d as a random variable with a given density function, denoted by f_d. The optimal cost function $J(x)$ of economic dispatch depends on the realization of d. To explicitly show this dependence, we rewrite the cost function and its corresponding placement value function as $J(x; d)$ and $V(X; d)$, respectively. A standard way to account for the randomness of $V(X; d)$ is to extend the storage placement (5) as a two-stage stochastic program that maximizes $\mathbb{E}[V(X; d)]$ within the same budget constraint. In the two-stage stochastic programming formulation,[9] the first stage is for planning decisions determining the storage placement over the network; the second stage is for operation decisions determining the generator and storage dispatch given the realization of the net demand d. We consider a more general *risk-aware* formulation than this mean-performance formulation as risks generated from renewable energy sources are important factors in decision-making for power systems (e.g., [54]–[56]). Specifically, the risk-aware placement problem can be formulated as the following combinatorial stochastic program featuring a mean-risk objective

$$\begin{aligned} \max_{X\subseteq\Omega} \quad & \mathbb{E}[V(X; d)] - \kappa\rho(V(X; d)) \\ \text{s.t.} \quad & \sum_{k:(i,k)\in X} c_k \leq b \end{aligned} \tag{19}$$

where ρ is a convex risk measure and the weight $\kappa \geq 0$ represents the importance of risk relative to mean value. The following theorem suggests that the submodularity of the placement value function is preserved through the risk-aware formulation with a convex risk measure, which is monotonically nonincreasing.

Theorem 5: Suppose that $X \mapsto V(X; d)$ is nondecreasing submodular for each $d \in D$, where D is the support of the density function f_d. If ρ is a convex risk measure, then the objective function $\mathbb{E}[V(X; d)] - \kappa\rho(V(X; d))$ of the stochastic program (19) is nondecreasing submodular.

[9] In principle, it is possible to utilize a multistage stochastic programming formulation featuring a more detailed information update model for generator and storage operation. This is, however, less common in planning studies due to concerns on data availability and computational complexity.

Theorem 5, which implies that submodularity is preserved through our risk-aware formulation, is a strong result as it holds independence of the shape of the density function f_d. When we use an empirical distribution of d with a finite number of samples, we are supposed to check the submodularity of $X \mapsto V(X; d)$ for each sample of d.

We now discuss a solution method for the risk-aware placement problem through an example. *Conditional value-at-risk* (CVaR) is one of the most popular convex risk measures, which is also coherent in the sense of Artzner *et al.* [57]. It measures the expected value conditional upon being within some percentage of the worst-case scenarios. The CVaR of a random variable Z, representing a loss, is defined as[10] $\mathrm{CVaR}_\alpha(Z) := \mathbb{E}[Z \mid Z \leq \mathrm{VaR}_\alpha(Z)]$, $\alpha \in (0, 1)$, where the *value-at-risk* (VaR) of Z (with the cumulative distribution function F_Z) is given by $\mathrm{VaR}_\alpha(Z) := \inf\{z \in \mathbb{R} \mid F_Z(z) \geq \alpha\}$. In words, VaR measures $(1 - \alpha)$ worst-case quantile of a loss distribution, while CVaR is equal to the conditional expectation of the loss within that quantile. The computation of CVaR can be efficiently performed by using the following *extremal* representation [58]: $\mathrm{CVaR}_\alpha(Z) = \inf_{y\in\mathbb{R}}[y + \frac{1}{1-\alpha}\mathbb{E}[(Z - y)^+]]$. Recalling that Z represents a loss or cost, we can compute the risk term via a convex program

$$\rho(V(X; d)) = \mathrm{CVaR}_\alpha(-V(X; d)) = \inf_{y\in\mathbb{R}} \left[y + \frac{1}{1-\alpha}\mathbb{E}[(-V(X; d) - y)^+] \right].$$

Since CVaR is a convex risk measure, if $X \mapsto V(X; d)$ is nondecreasing submodular, Theorem 5 allows us to use Algorithm 2 to solve the risk-aware problem (19). An additional complexity comes from the one-dimensional (1-D) convex optimization problem to compute $\mathrm{CVaR}_\alpha(-V(X; d))$ at each greedy step. Other risk measures that result in such a bilevel combinatorial-convex optimization problem can be found in [59].

V. Numerical Experiments

A. IEEE 14 Bus Case

We first demonstrate the performance of our placement algorithms by using IEEE 14 bus test case. Hourly zonal aggregated LMP and load data are obtained from the Pennsylvania–New Jersey–Maryland (PJM) interconnection. The data correspond to 14 zones inside PJM's RTO for the year 2014. We consider the hourly operation of storage over a representative day. The input data for each hour of the representative day are obtained by averaging over all the 365 days of the year. The hourly average load in the system is 80.5 GW.

The load and price time series for these 14 zones are assigned to the 14 buses of the network, where the price data are used to specify the linear coefficient of generation costs. Quadratic cost coefficients for generators are obtained from MATPOWER [60]. The capacity of each line is set to be the average load per bus over the 24 h. We consider a simple setting in which an exhaustive search is still feasible so that the performance of the greedy placement can be compared to the exact optimal solution in the quadratic case. To this end, we let the type of storage be $K = 1$.

[10]This definition is valid when the distribution of Z has no probability atom.

We now consider placing five storage devices in the 14 bus-network, with a total energy capacity of 150 MWh. Using the optimal set containment optimization (18), we verify that this setting satisfies the small storage assumption and that in the critical region the Hessian condition in Theorem 3 holds. The greedy strategy in Algorithm 2 is implemented. We also perform an exhaustive search over all feasible storage placements 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 identifies the exact optimal placement in this case, with buses $\{13, 12, 6, 2, 5\}$ selected to place storage.

B. Other Test Cases

To further examine the performance of greedy placement, we test the algorithm by using other IEEE test cases, each of which has 14, 30, 57, and 118 buses, together with a randomized assignment of the PJM load and price data to the network buses.[11] For larger networks, it is no longer feasible to benchmark the greedy algorithm's performance against an exact optimal solution which should be identified through an exhaustive search. Therefore, we compare the greedy algorithm's performance and run-time against that of a mixed-integer quadratic programming (MIQP) solver from Gubori. As the solver implements a branch-and-bound algorithm, the Gurobi solution comes with a posterior performance bound on the optimal cost. For each of the test network topology, we follow the setup in the previous Section V-A, but vary the total storage capacity from 0.5% of the average system-wide load to 5% of the system-wide load.

For each of the 40 problem instances (four IEEE test cases and 10 total storage configurations), we have the following observations:

1) the small storage condition is valid and the placement value function is submodular;
2) the greedy algorithm achieves the same storage placement value (and the system-wide cost) as the MIQP-based method; and
3) per the error bound provided by the branch-and-bound procedure, the MIQP-based method finds the global optimal solution up to a numerical tolerance of 10^{-4} and, therefore, so does the greedy algorithm.

Fig. 4 compares the run-time of these two methods, in which the run-times are averaged across the storage capacities; we have not observed significant or systematic variation in run-time when the storage capacities are changed (the number of storage devices to be placed are fixed to be 5 as in the Section V-A). The run-time comparison demonstrates that the greedy method has superior scalability compared to the MIQP-based method, which is consistent with the fact that the greedy approach is a polynomial time algorithm, whereas branch-and-bound procedure takes exponential time in the worst case.

The observation that submodularity holds for all our tested cases and the superior performance of the greedy placement algorithm should not be generalized without caution. Our example

[11]We use a random assignment due to the lack of real demand time series for the IEEE test cases. With the random assignment, the demand (and price) time series for each node is a linear combination of the demand (and price) time series of the PJM zonal demand data with coefficients generated uniformly at random.

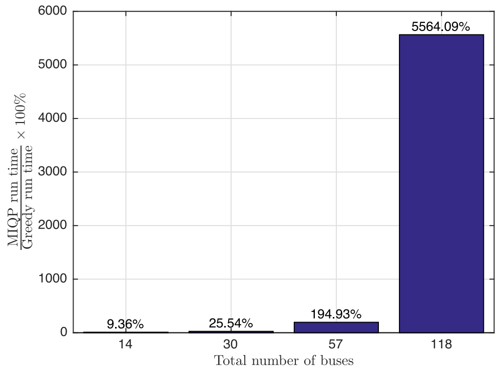


Fig. 4. Average run-time comparison between the greedy algorithm and the MIQP-based method.

in Section IV-A suggests that submodularity is not guaranteed to hold in general; our proposed verification procedure (see Algorithm 1) is a useful tool to check whether a given storage placement problem is submodular. Furthermore, the results that the greedy placement algorithm obtains an optimal solution in all test cases indicate that better performance guarantees for the greedy algorithm might be obtainable under more stringent conditions. Nonetheless, our algorithm guarantees to provide a $(1 - 1/e)$-optimal solution when the problem instance is verified to be submodular.

VI. Conclusion

In this paper, we have proposed a discrete optimization-based framework for placing energy storage devices in a power network when all storage resources are optimally controlled to minimize system-wide cost. This approach is useful when explicitly accounting for heterogeneous storage installation and capital costs. To use a scalable modified greedy algorithm to solve this NP-hard combinatorial optimization problem, we have investigated a tight condition under which the placement value function is submodular. Based on our comprehensive analytical characterization of the optimal cost, prices, and critical regions in a parametric economic dispatch problem with storage dynamics, we have also developed an efficient computational method to verify submodularity, and gained the unique insight that the spatiotemporal congestion pattern of a power network is a critical factor for submodularity.

Appendix A
Proofs for Section III

Proof of Proposition 1: As the objective function of (8) is strongly convex in λ, we know that $\lambda^\star(x)$ 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 (8b) implies that $\gamma_t^\star(x) = \lambda_{1,t}^\star(x)$, and therefore $\gamma^\star(x)$ 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 $\hat{\beta}_t$ 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 (8b) can be written as $\lambda_t = \gamma_t \mathbf{1} - H_t^{\text{net}\top}\hat{\beta}_t$, $t \in \mathcal{T}$. Now as H_t^{net} is a full row rank matrix, $\hat{\beta}_t$ is uniquely determined by the equation above given fixed λ_t and γ_t. This implies that β_t is also unique. ■

Proof of Lemma 1: Consider the primal program (4), which has an infinitely differentiable objective function and linear constraints. Under the nondegeneracy condition, we can apply standard sensitivity theorem of nonlinear programming [61], which suggests the differentiability of $J(x)$ and that

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

for any $i \in \mathcal{N}$. To show that $\partial J(x)/\partial x_i$ itself is again a continuous function, we observe that $\lambda^\star(x)$ is the unique solution of the dual QP (6). By the smoothness of the objective and constraints of (6), we know that $\lambda^\star(x)$ is continuous in x. Furthermore, the positive part function $(\cdot)^+$ is a continuous function from $\mathbb{R}$ to $\mathbb{R}_+$. Therefore, we conclude that $\partial J(x)/\partial x_i$ is continuous and $J(x)$ is continuously differentiable. As $\partial J(x)/\partial x_i \leq 0$, the function $J(x)$ is nonincreasing in x_i for each $i \in \mathcal{N}$. ■

Proof of Lemma 2: We write the primal problem (4) as

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

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

Proof of Lemma 3: This is a standard multiparametric quadratic programming result. See, e.g., [39]. ■

Proof of Theorem 1 and Corollary 1: By strict complementary slackness, $\beta_{\ell,t} = 0$ if $\ell \notin \mathcal{E}_t^{\text{C}}$ and $\beta_{\ell,t} > 0$ if $\ell \in \mathcal{E}_t^{\text{C}}$. Thus, we can focus on the reduced dual variable $\hat{\beta}_t := z_t^{\text{net}}\beta_t \in \mathbb{R}^{m_t}$, $t \in \mathcal{T}$. For convenience, we denote

$$\xi_{i,t} = \left(z_{i,t}^{\text{st}}\right)^+, \quad \eta_{i,t} = \mathbb{1}\{z_{i,t}^{\text{st}} = 0\}, \quad i \in \mathcal{N}, t \in \mathcal{T}.$$

Then, we can consider the following reduced form of dual program (6) in a neighborhood of x:

$$\max_{\lambda,\gamma,\hat{\beta}} \quad \psi(\lambda, \gamma, \hat{\beta}) \tag{20a}$$

$$\text{s.t.} \quad \lambda_t = \gamma_t \mathbf{1} - H_t^{\text{net}\top}\hat{\beta}_t, \quad t \in \mathcal{T} \tag{20b}$$

$$F(\eta_t)(\lambda_{t+1} - \lambda_t) = 0, \quad t \in \mathcal{T} \tag{20c}$$

where the objective function is

$$\psi(\lambda, \gamma, \hat{\beta}) := \sum_{t=1}^{T} -\frac{1}{2}(\lambda_t - r_t)^\top Q_t^{-1}(\lambda_t - r_t) + d_t^\top \lambda_t - \hat{f}_t^\top \hat{\beta}_t - x^\top \Delta_{\xi_t}(\lambda_{t+1} - \lambda_t)$$

$\bar{f}_t := z_t^{\mathrm{net}}\bar{f}$, and the matrix $F(\eta_t) \in \mathbb{R}^{\iota_t \times n}$ with $\iota_t := \sum_{n\in\mathcal{N}} \eta_{i,t}$ is formed by removing zero rows from diagonal matrix Δ_{η_t}. We vectorize the variables and coefficients as

$$\boldsymbol{\lambda} := [\lambda_1^\top, \dots, \lambda_T^\top]^\top, \hat{\boldsymbol{\beta}} := [\hat{\beta}_1^\top, \dots, \hat{\beta}_T^\top]^\top, \boldsymbol{\gamma} := [\gamma_1, \dots, \gamma_T]^\top$$

$$\boldsymbol{r} := [r_1^\top, \dots, r_T^\top]^\top, \boldsymbol{d} := [d_1^\top, \dots, d_T^\top]^\top, \bar{\boldsymbol{f}} := [\bar{f}_1^\top, \dots, \bar{f}_T^\top]^\top$$

and let $\boldsymbol{Q} := \mathrm{diag}(Q_1, \dots, Q_T) \in \mathbb{R}^{nT\times nT}$, $\boldsymbol{G} := \mathrm{diag}(\mathbf{1}, \dots, \mathbf{1}) \in \mathbb{R}^{nT\times T}$, $\boldsymbol{H} := \mathrm{diag}(H_1^{\mathrm{net}}, \dots, H_T^{\mathrm{net}}) \in \mathbb{R}^{(\sum_{t\in\mathcal{T}} m_t)\times nT}$, and

$$\boldsymbol{F} := \begin{bmatrix} -F(\eta_1) & F(\eta_1) & & & \\ & \ddots & \ddots & & \\ & & -F(\eta_{T-1}) & F(\eta_{T-1}) \\ & & & -F(\eta_T) \end{bmatrix}$$

where $\boldsymbol{F} \in \mathbb{R}^{(\sum_{t\in\mathcal{T}} \iota_t)\times nT}$. Then, (20) can be written as

$$\begin{aligned} \max_{\boldsymbol{\lambda},\boldsymbol{\gamma},\hat{\boldsymbol{\beta}}} \quad & -\frac{1}{2}(\boldsymbol{\lambda}-\boldsymbol{r})^\top \boldsymbol{Q}^{-1}(\boldsymbol{\lambda}-\boldsymbol{r}) + \boldsymbol{\kappa}^\top\boldsymbol{\lambda} - \bar{\boldsymbol{f}}^\top\hat{\boldsymbol{\beta}} \\ \text{s.t.} \quad & \boldsymbol{\lambda} = \boldsymbol{G}\boldsymbol{\gamma} - \boldsymbol{H}^\top\hat{\boldsymbol{\beta}} \\ & \boldsymbol{F}\boldsymbol{\lambda} = 0 \end{aligned}$$

where $\boldsymbol{\kappa} := \boldsymbol{d} - \boldsymbol{Y}x$, and

$$\boldsymbol{Y} := \begin{bmatrix} -\Delta_{\xi_1} & \Delta_{\xi_1} - \Delta_{\xi_2} & \dots & \Delta_{\xi_{T-1}} - \Delta_{\xi_T} \end{bmatrix}^\top.$$

We can reduce the variable $\boldsymbol{\lambda}$ using constraint $\boldsymbol{\lambda} = \boldsymbol{G}\boldsymbol{\gamma} - \boldsymbol{H}^\top\hat{\boldsymbol{\beta}}$ and standardize the resulting optimization to obtain

$$\begin{aligned} \min_{\boldsymbol{\gamma},\hat{\boldsymbol{\beta}}} \quad & \frac{1}{2}\begin{bmatrix}\boldsymbol{\gamma}\\ \hat{\boldsymbol{\beta}}\end{bmatrix}^\top \underbrace{\begin{bmatrix} \boldsymbol{G}^\top\boldsymbol{Q}^{-1}\boldsymbol{G} & \boldsymbol{G}^\top\boldsymbol{Q}^{-1}\boldsymbol{H}^\top \\ \boldsymbol{H}\boldsymbol{Q}^{-1}\boldsymbol{G} & \boldsymbol{H}\boldsymbol{Q}^{-1}\boldsymbol{H}^\top \end{bmatrix}}_{:=\mathcal{A}} \begin{bmatrix}\boldsymbol{\gamma}\\ \hat{\boldsymbol{\beta}}\end{bmatrix} \\ & - \underbrace{\begin{bmatrix} \boldsymbol{G}^\top\boldsymbol{\kappa} + \boldsymbol{G}^\top\boldsymbol{Q}^{-1}\boldsymbol{r} \\ -\boldsymbol{H}\boldsymbol{\kappa} - \boldsymbol{H}\boldsymbol{Q}^{-1}\boldsymbol{r} - \bar{\boldsymbol{f}} \end{bmatrix}}_{:=\mathcal{B}}^\top \begin{bmatrix}\boldsymbol{\gamma}\\ \hat{\boldsymbol{\beta}}\end{bmatrix} + \frac{1}{2}\boldsymbol{r}^\top\boldsymbol{Q}^{-1}\boldsymbol{r} \\ \text{s.t.} \quad & \underbrace{\begin{bmatrix} \boldsymbol{F}\boldsymbol{G} & -\boldsymbol{F}\boldsymbol{H}^\top \end{bmatrix}}_{:=\mathcal{C}} \begin{bmatrix}\boldsymbol{\gamma}\\ \hat{\boldsymbol{\beta}}\end{bmatrix} = 0. \end{aligned}$$

This is a standard equality constrained QP. Note that the vector $\mathcal{B}$ in the objective function above is affine in $\boldsymbol{\kappa}$. Given the uniqueness of the prices (see Proposition 1), we can use the formula for the solution of standard equality constrained QP and conclude that

$$\begin{bmatrix}\boldsymbol{\gamma}\\ \hat{\boldsymbol{\beta}}\end{bmatrix} = \left(\mathcal{A}^{-1}\mathcal{C}^\top(\mathcal{C}\mathcal{A}^{-1}\mathcal{C}^\top)^{-1}\mathcal{C}\mathcal{A}^{-1} - \mathcal{A}^{-1}\right)\mathcal{B}. \tag{21}$$

Combining this expression with $\boldsymbol{\lambda} = \boldsymbol{G}\boldsymbol{\gamma} - \boldsymbol{H}^\top\hat{\boldsymbol{\beta}}$, we have

$$\boldsymbol{\lambda} = \begin{bmatrix} \boldsymbol{G} & -\boldsymbol{H}^\top \end{bmatrix} \left(\mathcal{A}^{-1}\mathcal{C}^\top(\mathcal{C}\mathcal{A}^{-1}\mathcal{C}^\top)^{-1}\mathcal{C}\mathcal{A}^{-1} - \mathcal{A}^{-1}\right)\mathcal{B}. \tag{22}$$

Notice that as $\mathcal{B}$ is affine in $\boldsymbol{\kappa}$, which is an affine function of the storage capacities x, the resulting prices $\boldsymbol{\gamma}$, $\hat{\boldsymbol{\beta}}$, and $\boldsymbol{\lambda}$ are affine functions of the storage capacities. Substituting the definition of $\mathcal{B}$ and $\boldsymbol{\kappa}$, we obtain expression for $W_t(z^{\mathrm{net}}, z^{\mathrm{st}})$, $\bar{\lambda}_t(z^{\mathrm{net}}, z^{\mathrm{st}})$, $P_t(z^{\mathrm{net}}, z^{\mathrm{st}})$, and $\bar{\beta}_t(z^{\mathrm{net}}, z^{\mathrm{st}})$, for $t \in \mathcal{T}$. ■

Proof of Theorem 2: We can compute the optimal value J, by substituting the expression of the optimal $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$, as follows:

$$J(x) = \frac{1}{2}\mathcal{B}^\top(\mathcal{A}^{-1} - \mathcal{A}^{-1}\mathcal{C}^\top(\mathcal{C}\mathcal{A}^{-1}\mathcal{C}^\top)^{-1}\mathcal{C}\mathcal{A}^{-1})\mathcal{B} - \frac{1}{2}\boldsymbol{r}^\top\boldsymbol{Q}^{-1}\boldsymbol{r}.$$

Note that $\mathcal{B} = [\begin{smallmatrix}\boldsymbol{G}^\top \\ -\boldsymbol{H}\end{smallmatrix}](\boldsymbol{d} - \boldsymbol{Y}x) + [\begin{smallmatrix}\boldsymbol{G}^\top\boldsymbol{Q}^{-1}\boldsymbol{r} \\ -\boldsymbol{H}\boldsymbol{Q}^{-1}\boldsymbol{r} - \bar{\boldsymbol{f}}\end{smallmatrix}]$. By differentiating J twice with respect to x, we obtain the Hessian for the critical region containing x as

$$\begin{aligned} &\nabla_{xx}^2 J(x) \\ &= \boldsymbol{Y}^\top \begin{bmatrix}\boldsymbol{G}^\top \\ -\boldsymbol{H}\end{bmatrix}^\top \left(\mathcal{A}^{-1} - \mathcal{A}^{-1}\mathcal{C}^\top(\mathcal{C}\mathcal{A}^{-1}\mathcal{C}^\top)^{-1}\mathcal{C}\mathcal{A}^{-1}\right) \begin{bmatrix}\boldsymbol{G}^\top \\ -\boldsymbol{H}\end{bmatrix} \boldsymbol{Y}. \end{aligned} \tag{23}$$

■

Appendix B
Proofs for Section IV

Proof of Theorem 3: 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

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

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

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

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

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

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

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

As $\left(\nabla_{xx}^2 J(x)\right)_{ij} \geq 0$, 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 (17), we conclude that the set function V is submodular. ■

Proof of Theorem 4: By the proof of Theorem 1 and Corollary 1, we know that $(\lambda^\star(x), \gamma^\star(x), \beta^\star(x))$ is a stationary point of the objective of the dual QP (8). In other words, $(\lambda^\star(x), \gamma^\star(x), \beta^\star(x))$ is the unconstrained local maximizer of (8) in an affine subspace defined by the set of equality and inequality constraints in (8) which are binding at x. Recall that, given the storage congestion state $z^{\text{st}} \in \mathbb{R}^{n\times T}$, the objective function of (8) can be written as

$$\hat{\phi}(\lambda,\beta) = \sum_{t=1}^{T} -\frac{1}{2}(\lambda_t - r_t)^\top Q_t^{-1}(\lambda_t - r_t) + d_t^\top \lambda_t - \ell^\top \beta_t - x^\top \Delta_{\xi_t}(\lambda_{t+1} - \lambda_t).$$

Now consider a vector $\hat{x}$ that satisfies the two inequality conditions in Theorem 4. The first inequality condition ensures that, at $\hat{x}$, the storage congestion state $z^{\text{st}}(\hat{x})$ given by (10) is unchanged from $z^{\text{st}}(x)$. Therefore, $(\lambda^\star(\hat{x}), \gamma^\star(\hat{x}), \beta^\star(\hat{x}))$ is still the unconstrained local maximizer of (8) in the same affine subspace when x is replaced with $\hat{x}$ in the above expression of the objective function $\hat{\phi}$. The second inequality condition in Theorem 4 guarantees that $(\lambda^\star(\hat{x}), \gamma^\star(\hat{x}), \beta^\star(\hat{x}))$ is feasible for (8). Indeed, we observe from the expression of $\beta_t^\star(x)$ in (14) that modifying x does not affect $\beta_{j,t}^\star(x)$ for a line j that is uncongested and, hence, $\beta_{j,t}^\star(\hat{x})$ is 0 if line j is uncongested, while the second inequality condition in Theorem 4 ensures that $\beta_{j,t}^\star(\hat{x})$ remains positive if line j is congested. Therefore, $(\lambda^\star(\hat{x}), \gamma^\star(\hat{x}), \beta^\star(\hat{x}))$ must be a global maximizer of the dual QP (8) as the problem is concave. By strict complementary slackness, the conditions defining $\mathcal{R}_x$ ensure that the set of binding inequality constraints is unaltered with $\hat{x}$. Therefore, $\mathcal{R}_x$ is the critical region containing x. ■

Proof of Theorem 5: Since submodularity is preserved through addition, $\mathbb{E}[V(X;d)]$ is submodular. On the other hand, ρ is nonincreasing and convex because it is a convex risk measure. We now use the fact that the composition of a nonincreasing convex function and a nondecreasing submodular function is nonincreasing supermodular [62]. Therefore, we observe that $\rho(V(X;d))$ is nonincreasing supermodular. Finally, we again use the fact that submodularity is preserved through addition to conclude that $\mathbb{E}[V(X;d)] - \kappa\rho(V(X;d))$ is nondecreasing submodular for any $\kappa \geq 0$. ■

Appendix C
Construction of Critical Regions

Here, we provide a brief description of a basic procedure for the iterative construction of critical regions. To start, we select an initial point $x \in \mathcal{X}_0 := \mathcal{X}$ and compute the critical region $\mathcal{R}_x$ that contains it by using Theorem 4. Focusing on the part of the critical region inside of $\mathcal{X}_0$, i.e., $\mathcal{R}_x \cap \mathcal{X}_0$ while writing the inequality constraints that define this polytope as $Px \leq q$, we can partition the remaining region in $\mathcal{X}_0$ as

$$\mathcal{X}_i := \{x \in \mathcal{X}_0 : P_i^\top x \geq q_i, P_j^\top x \leq q_j \quad \forall j < i\}$$

where $P_i^\top$ is the ith row of P. Recursively applying this process to $\mathcal{X}_i$, we can obtain the collection of all critical regions in $\mathcal{X}$. This procedure is guaranteed to terminate in a finite number of steps.

Appendix D
Prices and Hessian under Small Storage

Under the small storage condition, the dual program (20) can be further simplified if the loads are heterogenous in time. In a nutshell, for any bus and any consecutive time periods, if there is a differential in the LMPs, adding *small* storage devices would not be able to fully eliminate such a price differential. In this case, we can obtain more explicit expressions for the LMPs as functions of the storage capacity vector x. Recall that $\xi_{i,t} = (z_{i,t}^{\text{st}})^+$.

Theorem 6: Suppose the small storage condition holds and the loads are heterogeneous such that $\lambda_{i,t}^\star(0) \neq \lambda_{i,t+1}^\star(0)$ for all i and t. Then, we have

$$\lambda_t^\star(x) = A_t(z_t^{\text{net}})\Delta_{(\xi_t - \xi_{t-1})}x + \tilde{\lambda}_t(z_t^{\text{net}})$$
$$\gamma_t^\star(x) = \mathbf{1}_1^\top\left(A_t(z_t^{\text{net}})\Delta_{(\xi_t - \xi_{t-1})}x + \tilde{\lambda}_t(z_t^{\text{net}})\right)$$
$$\beta_t^\star(x) = {z_t^{\text{net}}}^\top B_t(z_t^{\text{net}})^\top \Delta_{(\xi_t - \xi_{t-1})}x + \tilde{\beta}_t(z_t^{\text{net}})$$

where

$$A_t(z_t^{\text{net}}) := \frac{1}{\mathbf{1}^\top Q_t^{-1}\mathbf{1}}\mathbf{1}\mathbf{1}^\top + Q_t M_t R_t M_t Q_t$$
$$B_t(z_t^{\text{net}}) := Q_t M_t {H_t^{\text{net}}}^\top K_t^{-1}$$
$$\tilde{\lambda}_t(z_t^{\text{net}}) := A_t(z_t^{\text{net}})\left(d_t + Q_t^{-1}r_t\right) + B_t(z_t^{\text{net}})z_t^{\text{net}}\ell$$
$$\tilde{\beta}_t(z_t^{\text{net}}) := -{z_t^{\text{net}}}^\top B_t(z_t^{\text{net}})^\top(d_t + Q_t^{-1}r_t) - {z_t^{\text{net}}}^\top K_t^{-1} z_t^{\text{net}}\ell.$$

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

Proof: Under the small storage condition and the assumption that $\lambda_{i,t}^\star(0) = \lambda_{i,t+1}^\star(0)$, we no longer have the time coupling constraint (20c) in the dual program (20). In this case, given the storage congestion variable ξ and the network congestion variable z^{net}, the optimization problem above is separable across time. We can solve each of the following T optimization problems given the binding constraints:

$$J_t(x) := \max_{\lambda_t, \gamma_t, \hat{\beta}_t} \quad \psi_t(\lambda_t, \gamma_t, \hat{\beta}_t)$$
$$\text{s.t.} \quad \lambda_t = \gamma_t \mathbf{1} - {H_t^{\text{net}}}^\top \hat{\beta}_t \tag{24}$$

where

$$\psi_t(\lambda_t, \gamma_t, \hat{\beta}_t) := -\frac{1}{2}(\lambda_t - r_t)^\top Q_t^{-1}(\lambda_t - r_t) + d_t^\top \lambda_t - \ell_t^\top \hat{\beta}_t + x^\top \Delta_{(\xi_t - \xi_{t-1})}\lambda_t.$$

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

$$\lambda_t^\star = (Q_t M_t R_t M_t Q_t + \rho_t \mathbf{1}\mathbf{1}^\top)(\Delta_{(\xi_t - \xi_{t-1})}x + d_t + Q_t^{-1}r_t) + Q_t M_t {H_t^{\text{net}}}^\top K_t^{-1}\ell_t$$
$$\hat{\beta}_t^\star = -K_t^{-1}\left[H_t^{\text{net}} M_t (Q_t \Delta_{(\xi_t - \xi_{t-1})}x + Q_t d_t + r_t) + \ell_t\right]$$

where $\rho_t := 1/[\mathbf{1}^\top Q_t^{-1}\mathbf{1}]$, and M_t, K_t, and R_t are as defined above. Collecting terms and substituting the definitions for matrices $A_t(z_t^{\text{net}})$ and $B_t(z_t^{\text{net}})$, we obtain price formulas given in the theorem statement. ■

We can also obtain a simpler expression for the Hessian under the small storage condition.

Theorem 7: Under the same assumptions of Theorem 6, for any x such that $\nabla^2_{xx} J(x)$ exists, $\nabla^2_{xx} J(x) = (\nabla^2 J)^{\mathrm{st}}(x) + (\nabla^2 J)^{\mathrm{net}}(x)$, where $(\nabla^2 J)^{\mathrm{st}}(x)$ is the component that depends on the storage congestion pattern while $(\nabla^2 J)^{\mathrm{net}}(x)$ is the component that depends on the network congestion pattern, defined as

$$(\nabla^2 J)^{\mathrm{st}}(x) := \sum_{t=1}^{T} \frac{(\xi_t - \xi_{t-1})(\xi_t - \xi_{t-1})^\top}{\mathbf{1}^\top Q_t^{-1} \mathbf{1}},$$

$$(\nabla^2 J)^{\mathrm{net}}(x) := \sum_{t=1}^{T} \Delta_{(\xi_t - \xi_{t-1})} Q_t M_t R_t M_t Q_t \Delta_{(\xi_t - \xi_{t-1})}.$$

Proof: Inserting the expressions of λ_t, γ_t, and $\hat{\beta}_t$ given in Theorem 1 and Corollary 1 into (24), we obtain the optimal cost as

$$J_t(x) = \bar{J}_t + \frac{1}{2} x^\top \Delta_{(\xi_t - \xi_{t-1})} A_t(z_t^{\mathrm{net}}) \Delta_{(\xi_t - \xi_{t-1})} x + \left[A_t(z_t^{\mathrm{net}})(d_t + Q_t^{-1} r_t) + B_t(z_t^{\mathrm{net}}) \ell_t \right]^\top \Delta_{(\xi_t - \xi_{t-1})} x$$

where $\bar{J}_t$ does not depend on x. Theorem 2 then follows from computing the Hessian of $J_t(x)$ and summing it over t. ■

The simpler expression of the Hessian in Theorem 7 sheds light on how the Hessian depends on the storage and network congestion patterns. How the storage congestion pattern term of the Hessian is related to the alignment of LMP time series across the network is revealed in [1, Proposition 2]. This implies that the Hessian can be accurately approximated for some networks based on LMP alignment patterns as in practice, network congestion is only expected to occur during a small percentage of all operating hours in well-designed systems. Intuitive understanding of the network congestion term of the Hessian is left for future research.

Acknowledgment

The authors are in great debt to Prof. Yinyu Ye at Stanford University and Prof. Asu Ozdaglar at MIT for helpful discussions on various aspects of this paper.

References

[1] J. Qin, I. Yang, and R. Rajagopal, "Submodularity of energy storage placement in power networks," in *Proc. 55th IEEE Conf. Decis. Control*, 2016, pp. 686–693.

[2] GTM Research, "U.S. Energy Storage Monitor," 2015. [Online]. Available: https://www.greentechmedia.com/research/subscription/u.s.-energy-storag e-monitor

[3] J. Qin, H.-I. Su, and R. Rajagopal, "Storage in risk limiting dispatch: Control and approximation," in *Proc. Amer. Control Conf.*, 2013, pp. 4202–4208.

[4] A. ParandehGheibi, M. Roozbehani, M. A. Dahleh, and A. Ozdaglar, "The value of storage in securing reliability and mitigating risk in energy systems," *Energy Syst.*, vol. 6, no. 1, pp. 129–152, 2015.

[5] A. Castillo and D. F. Gayme, "Grid-scale energy storage applications in renewable energy integration: A survey," *Energy Convers. Manage.*, vol. 87, pp. 885–894, 2014.

[6] National Renewable Energy Laboratory, "The role of energy storage with renewable electricity generation," 2010. [Online]. Available: http://www.nrel.gov/docs/fy10osti/47187.pdf

[7] Kempton *et al.*, "A test of vehicle-to-grid (V2G) for energy storage and frequency regulation in the PJM system," 2008. [Online]. Available: https://www1.udel.edu/V2G/resources/test-v2g-in-pjm-jan09.pdf

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

[9] M. Munsell, "US energy storage market grew 243% in 2015, largest year on record," 2016. [Online]. Available: https://www.greentechmedia.com/articles/read/us-energy-storage-market-grew-243-in-2015-largest-year-on-record

[10] 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," *J. Assoc. Environ. Resour. Economists*, vol. 5, no. 1, pp. 39–75, 2018.

[11] S. Bose and E. Bitar, "Variability and the locational marginal value of energy storage," in *Proc. 53rd IEEE Conf. Decis. Control*, 2014, pp. 3259–3265.

[12] C. Thrampoulidis, S. Bose, and B. Hassibi, "Optimal placement of distributed energy storage in power networks," *IEEE Trans. Autom. Control*, vol. 61, no. 2, pp. 416–429, Feb. 2016.

[13] 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 Trans. Power Syst.*, vol. 30, no. 5, pp. 2288–2300, Sep. 2015.

[14] S. Wogrin and D. F. Gayme, "Optimizing storage siting, sizing, and technology portfolios in transmission-constrained networks," *IEEE Trans. Power Syst.*, vol. 30, no. 6, pp. 3304–3313, Nov. 2015.

[15] E. Sjodin, D. F. Gayme, and U. Topcu, "Risk-mitigated optimal power flow for wind powered grids," in *Proc. Amer. Control Conf.*, 2012, pp. 4431–4437.

[16] M. Kraning, Y. Wang, E. Akuiyibo, and S. Boyd, "Operation and configuration of a storage portfolio via convex optimization," *IFAC Proc. Vols.*, vol. 44, no. 1, pp. 10 487–10 492, 2011.

[17] J. Qin and R. Rajagopal, "Dynamic programming solution to distributed storage operation and design," in *Proc. IEEE Power Energy Soc. General Meeting*, 2013.

[18] A. Castillo and D. Gayme, "Evaluating the effects of real power losses in optimal power flow based storage integration," *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 3, pp. 1132–1145, Sep. 2018.

[19] S. Bose, D. F. Gayme, U. Topcu, and K. M. Chandy, "Optimal placement of energy storage in the grid," in *Proc. 51st IEEE Conf. Decis. Control*, 2012, pp. 5605–5612.

[20] A. Castillo and D. F. Gayme, "Profit maximizing storage allocation in power grids," in *Proc. 52nd IEEE Conf. Decis. Control*, 2013, pp. 429–435.

[21] Y. Tang and S. H. Low, "Optimal placement of energy storage in distribution networks," in *Proc. 55th IEEE Conf. Decis. Control*, 2016, pp. 3258–3264.

[22] Z.-J. M. Shen, C. Coullard, and M. S. Daskin, "A joint location-inventory model," *Transp. Sci.*, vol. 37, no. 1, pp. 40–55, 2003.

[23] W. Qi, Y. Liang, and Z.-J. M. Shen, "Joint planning of energy storage and transmission for wind energy generation," *Oper. Res.*, vol. 63, no. 6, pp. 1280–1293, 2015.

[24] Q. Li, R. Negi, and M. D. Ilić, "Phasor measurement units placement for power system state estimation: A greedy approach," in *Proc. IEEE Power Energy Soc. General Meeting*, 2011.

[25] Z. Liu, A. Clark, P. Lee, L. Bushnell, D. Kirschen, and R. Poovendran, "Submodular optimization for voltage control," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 502–513, Jan. 2018.

[26] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions–I," *Math. Program.*, vol. 14, pp. 265–294, 1978.

[27] S. Khuller, A. Moss, and J. S. Naor, "The budgeted maximum coverage problem," *Inf. Process. Lett.*, vol. 70, no. 1, pp. 39–45, 1999.

[28] K. Dvijotham, M. Chertkov, and S. Backhaus, "Storage sizing and placement through operational and uncertainty-aware simulations," in *Proc. 47th Hawaii Int. Conf. Syst. Sci.*, 2014, pp. 2408–2416.

[29] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies," *J. Mach. Learn. Res.*, vol. 9, pp. 235–284, 2008.

[30] A. Krause, R. Rajagopal, A. Gupta, and C. Guestrin, "Simultaneous optimization of sensor placements and balanced schedules," *IEEE Trans. Autom. Control*, vol. 56, no. 10, pp. 2390–2405, Oct. 2011.

[31] A. Olshevsky, "Minimal controllability problems," *IEEE Trans. Control Netw. Syst.*, vol. 1, no. 3, pp. 249–258, Sep. 2014.

[32] A. Clark, L. Bushnell, and R. Poovendran, "A supermodular optimization framework for leader selection under link noise in linear multi-agent

systems," *IEEE Trans. Autom. Control*, vol. 59, no. 2, pp. 283–296, Feb. 2014.

[33] T. H. Summers, F. L. Cortesi, and J. Lygeros, "On submodularity and controllability in complex dynamical networks," *IEEE Trans. Control Netw. Syst.*, vol. 3, no. 1, pp. 91–101, Mar. 2016.

[34] V. Tzoumas, M. A. Rahimian, G. J. Pappas, and A. Jadbabaie, "Minimal actuator placement with bounds on control effort," *IEEE Trans. Control Netw. Syst.*, vol. 3, no. 1, pp. 67–78, Mar. 2016.

[35] B. Stott, J. Jardim, and O. Alsac, "DC power flow revisited," *IEEE Trans. Power Syst.*, vol. 24, no. 3, pp. 1290–1300, Aug. 2009.

[36] A. J. Wood and B. F. Wollenberg, *Power Generation, Operation, and Control*. New York, NY, USA: Wiley, 2012.

[37] D. Muñoz-Álvarez and E. Bitar, "Financial storage rights in electric power networks," *J. Regulatory Econ.*, vol. 52, no. 1, pp. 1–23, 2017.

[38] I. Yang, S. Burden, R. Rajagopal, S. S. Sastry, and C. J. Tomlin, "Approximation algorithms for optimization of combinatorial dynamical systems," *IEEE Trans. Autom. Control*, vol. 61, no. 9, pp. 2644–2649, Sep. 2016.

[39] A. Bemporad, M. Morari, V. Dua, and E. N. Pistikopoulos, "The explicit linear quadratic regulator for constrained systems," *Automatica*, vol. 38, no. 1, pp. 3–20, 2002.

[40] Y. Ji, L. Tong, and R. J. Thomas, "Probabilistic forecast of real-time LMP and network congestion," *IEEE Trans. Power Syst.*, vol. 32, no. 2, pp. 831–841, Mar. 2017.

[41] E. Boros and P. L. Hammer, "Pseudo-Boolean optimization," *Discrete Appl. Math.*, vol. 123, no. 1, pp. 155–225, 2002.

[42] Y. Xu and L. Tong, "Optimal operation and economic value of energy storage at consumer locations," *IEEE Trans. Autom. Control*, vol. 62, no. 2, pp. 792–807, Feb. 2017.

[43] R. K. Jain, J. Qin, and R. Rajagopal, "Data-driven planning of distributed energy resources amidst socio-technical complexities," *Nature Energy*, vol. 2, no. 8, 2017, Art. no. 17112.

[44] A. Gupta, S. Bhartiya, and P. Nataraj, "A novel approach to multiparametric quadratic programming," *Automatica*, vol. 47, no. 9, pp. 2112–2117, 2011.

[45] P. TøNdel, T. A. Johansen, and A. Bemporad, "An algorithm for multiparametric quadratic programming and explicit MPC solutions," *Automatica*, vol. 39, no. 3, pp. 489–497, 2003.

[46] P. Tondel, T. A. Johansen, and A. Bemporad, "Further results on multiparametric quadratic programming," in *Proc. 42nd IEEE Conf. Decis. Control (IEEE Cat. No. 03CH37475)*, 2003, pp. 3173–3178.

[47] A. Bemporad, "A multiparametric quadratic programming algorithm with polyhedral computations based on nonnegative least squares," *IEEE Trans. Autom. Control*, vol. 60, no. 11, pp. 2892–2903, Nov. 2015.

[48] P. Ahmadi-Moshkenani, T. A. Johansen, and S. Olaru, "Combinatorial approach towards multi-parametric quadratic programming based on characterizing adjacent critical regions," *IEEE Trans. Autom. Control*, vol. 63, no. 10, pp. 3221–3231, Oct. 2018.

[49] M. Herceg, M. Kvasnica, C. Jones, and M. Morari, "Multi-Parametric Toolbox 3.0," in *Proc. Eur. Control Conf.*, 2013, pp. 502–510.

[50] B. C. Eaves and R. M. Freund, "Optimal scaling of balls and polyhedra," *Math. Program.*, vol. 23, no. 1, pp. 138–147, 1982.

[51] M. Sviridenko, "A note on maximizing a submodular set function subject to a knapsack constraint," *Oper. Res. Lett.*, vol. 32, pp. 41–43, 2004.

[52] A. Krause and C. Guestrin, "A note on the budgeted maximization on submodular functions," Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. CMU-CALD-05-103, 2005.

[53] B. Mirzasoleiman, A. Karbasi, R. Sarkar, and A. Krause, "Distributed submodular maximization," *J. Mach. Learn. Res.*, vol. 17, no. 1–44, 2016.

[54] B. Zhang, R. Rajagopal, and D. Tse, "Network risk limiting dispatch: Optimal control and price of uncertainty," *IEEE Trans. Autom. Control*, vol. 59, no. 9, pp. 2442–2456, Sep. 2014.

[55] D. Bienstock, M. Chertkov, and S. Harnett, "Chance-constrained optimal power flow: Risk-aware network control under uncertainty," *SIAM Rev.*, vol. 56, no. 3, pp. 461–495, 2014.

[56] I. Yang, D. S. Callaway, and C. J. Tomlin, "Variance-constrained risk sharing in stochastic systems," *IEEE Trans. Autom. Control*, vol. 62, no. 4, pp. 1865–1879, Apr. 2017.

[57] P. Artzner, F. Delbaen, J. M. Eber, and D. Heath, "Coherent measures of risk," *Math. Finance*, vol. 9, no. 3, pp. 203–228, 1999.

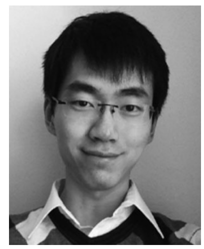
[58] R. T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," *J. Risk*, vol. 2, pp. 21–42, 2000.

[59] C. W. Miller and I. Yang, "Optimal control of conditional value-at-risk in continuous time," *SIAM J. Control Optim.*, vol. 55, no. 2, pp. 856–884, 2017.

[60] 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 Trans. Power Syst.*, vol. 26, no. 1, pp. 12–19, Feb. 2011.

[61] D. G. Luenberger and Y. Ye, *Linear and Nonlinear Programming*. vol. 2. New York, NY, USA: Springer, 1984.

[62] D. M. Topkis, "Minimizing a submodular function on a lattice," *Oper. Res.*, vol. 26, no. 2, pp. 305–321, 1978.



Junjie Qin received the B.S. degree in economics and in hydropower engineering from Tsinghua University, Beijing, China, in 2010, and the M.S. degree in civil engineering and in statistics, and the Ph.D. degree in computational and mathematical engineering from Stanford University, Stanford, CA, USA, in 2011 and 2017, respectively.

He is a Postdoctoral Researcher with UC Berkeley and Lawrence Berkeley National Laboratory, Berkeley, CA, USA, working with Prof. K. Poolla and Prof. P. Varaiya. He works on stochastic control and economic mechanism design for power systems.

Dr. Qin is a recipient of the Satre family fellowship on energy and sustainability and a finalist for the Best Student Paper Award at the 55th IEEE Conference on Decision and Control 2016.

Insoon Yang received B.S. degrees in mathematics and in mechanical engineering (*summa cum laude*) from Seoul National University, Seoul, South Korea, in 2009, and the M.S. degree in electrical engineering and computer sciences (EECS), the M.A. degree in mathematics and the Ph.D. degree in EECS from UC Berkeley, Berkeley, CA, USA, in 2012, 2013, and 2015, respectively.

He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, Seoul National University. He was an Assistant Professor with the EE Department, USC from 2016 to 2018 and a Postdoctoral Associate with the Laboratory for Information and Decision Systems, MIT from 2015 to 2016. His research interests include stochastic control and optimization with application to cyber-physical systems and safe autonomy.

Dr. Yang is a recipient of the 2015 Eli Jury Award and the 2018 SNU Creative-Pioneering Researcher Award, and a finalist for the Best Student Paper Award at the 55th IEEE Conference on Decision and Control 2016.

Ram Rajagopal received the Ph.D. degree in electrical engineering and computer sciences and the M.A. degree in statistics from the University of California Berkeley, Berkeley, CA, USA, both in 2009.

He is currently an Associate Professor of Civil and Environmental Engineering and Electrical Engineering (by Courtesy), Stanford University, where he directs the Stanford Sustainable Systems Lab (S3L), focused on large scale monitoring, data analytics, and stochastic control for infrastructure networks, in particular energy and transportation. His current research interests in power systems are in integration of renewables, smart distribution systems, and demand-side data analytics. He has also built wireless sensing systems and large-scale analytics for transportation and power networks.

Dr. Rajagopal is a recipient of the Powell Foundation Fellowship, Berkeley Regents Fellowship, the NSF CAREER Award, and the Makhoul Conjecture Challenge Award, and a finalist for the Best Student Paper Award at the 55th IEEE Conference on Decision and Control 2016. He holds more than 30 patents from his work, and has advised or founded various companies in the fields of sensor networks, power systems, and data analytics.