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 (i) the combinatorial nature of such placement problems, and (ii) the spatial and temporal transfer 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 a novel insight that the spatio-temporal 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.

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 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 interest 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 have considered the energy storage placement problem as the problem of sizing storage. 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 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 multi-level 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. Krawing et al. [16] extended their convex 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 in recognition of the complexity of the AC power flow model. Castillo and Gayme [18] studied the storage allocation problem with line losses considered. This led to a non-convex quadratic constrained quadratic program for which exact convex relaxations based on semidefinite programs and second order cone programs were developed. Bose et al. [19] developed 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 where storage is operated to maximize profit based on locational marginal prices (LMPs) in the power network. Structural results between the storage decisions and the LMPs were derived. 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 treats the energy storage placement problem as a form of facility location problem (cf. [22] and references therein). An example is Qi et al. [23], which considers a planning problem for energy storage and transmission in the presence of wind energy 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 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 operating 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 accurate model of considering these cost factors. Additionally, a discrete optimization framework is useful when handling practical scenarios in which fixed-capacity storage devices are to be placed. These advantages are elaborated in Section II-E.

We formulate the placement problem as a maximization of the placement value function, a set function that represents the value 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 provable suboptimality bound [26], [27]. The submodularity of energy storage placement is not unexpected but characterizing conditions under which it holds has been recognized, e.g., in [28], as an unanswered question.

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 multi-period 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 locational marginal prices 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 (\(Q \neq Q_{\text{sm}}^m\) in Fig. 1), although such situations are unlikely to occur in practice. This examination provides a unique insight into the effect of spatio-temporal (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 submodularity (i.e., testing whether a particular problem instance belongs to set \(Q_{\text{sm}}^m\) 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 \(Q_{\text{sm}}^s\) in Fig. 1) under which the verification procedure terminates in one step. The small storage condition can be tested with the problem data via solving a simple linear program. An extension to risk-aware placement strategies is also discussed for cases with a large penetration of wind and solar energy sources.

\(^1\)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.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{venn_diagram.pdf}
\caption{Venn diagram for the set of all possible storage placement problem instances.}
\end{figure}

C. Organization

The remainder of this paper is organized as follows. In Section II, we introduce a 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.

II. Problem Formulation

A. 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 $t$th column, and $x_t^\top := (x_{i,1}, \ldots, x_{i,T}) \in \mathbb{R}^{1 \times T}$ to denote its $i$th row. For a sequence of matrices $A_k, k = 1, \ldots, K$, we use $\text{diag}(A_1, \ldots, A_K)$ 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 $z$ so that $z = (z)^+ - (z)^-$. For any Euclidean vector space $\mathbb{R}^n$, we use $\mathbb{I} \{\cdot\}$ to denote the all-one vector and $1_k \in \mathbb{R}^n$ to denote the $k$th elementary vector, i.e., the vector with all zeros except for its $k$th element which is 1. We use $I \{\cdot\}$ to denote the indicator function.

B. 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 $A = \mathbb{R}^{n \times m}$ be the node-edge incidence matrix defined for the network. Under the classical DC approximation to the steady-state AC power flow [35], the lines are characterized by their reactance and real power flow capacity vector denoted by $\hat{E}$. The power flow constraints can be compactly expressed as

\[
1^\top p_t = 0, \quad (1a)
\]

\[
H p_t \leq \ell, \quad (1b)
\]

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.

C. Energy Storage Dynamics

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

\[
s_{i,t+1} = s_{i,t} - u_{i,t}, \quad t = 1, \ldots, T - 1, \quad (2)
\]

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

\[
0 \leq s_{i,t} \leq x_i, \quad t \in \mathcal{T}. \quad (3)
\]

Note that $x_i = 0$ if there is no storage connected to bus $i$. Applying (2) recursively, we can express constraint (3) as

\[
0 \leq \sum_{\tau=1}^{t} -u_{i,\tau} \leq x_i, \quad t \in \mathcal{T},
\]

which can be compactly expressed in the following vector form:

\[
0 \leq L u_t \leq x_i 1, \quad (4)
\]

where $u_t \in \mathbb{R}^T$ 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 $i \geq j$. The information about the storage dynamics is embedded in the matrix $L$.

D. Economic Dispatch with Storage Control

The economic dispatch problem aims to identify an efficient generator dispatch to serve the 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

\[\text{...}\]

\[\text{...}\]
the following multi-period economic dispatch problem with storage dynamics:

\[ J(x) := \min_{g,u} \sum_{t=1}^{T} C_t(g_t) \]

subject to

\[ \beta_t : H(g_t + u_t - d_t) \leq \ell \quad \forall t \in \mathcal{T}, \]  
\[ \gamma_t : 1^T (g_t + u_t - d_t) = 0 \quad \forall t \in \mathcal{T}, \]  
\[ \mu_i : L u_i \leq x_i \mathbf{1} \quad \forall i \in \mathcal{N}, \]  
\[ \nu_i : L u_i \geq 0 \quad \forall i \in \mathcal{N}. \]

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.,

\[ C_{i,t}(g_{i,t}) := \frac{1}{2} q_{i,t}^T Q_{i,t} g_{i,t} + r_{i,t}^T g_{i,t}, \quad t \in \mathcal{T}, \]

where \( Q_{i,t} \) is a diagonal matrix whose diagonal entries are positive, which models the increasing incremental (marginal) heat rate, and \( r_{i,t} \in \mathbb{R}^n \) is the linear cost coefficient for generators over the network. The cost function mainly models the fuel costs 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 multi-period economic dispatch problem. The optimal value of the multi-period economic dispatch problem, denoted by \( J \), is a function of storage capacities over the network \( x \), as storage capacities affect the feasible region of the optimization problem (4) via constraint (4d).

E. Storage Placement as Combinatorial Optimization

The optimal cost of this multi-period economic dispatch problem depends critically on the storage capacity vector \( x \in \mathbb{R}^n \) over the network. If there is no storage connected to the network (i.e., \( x = 0 \)), the optimal cost of this multi-period 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 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 \( c_k, k = 1, \ldots, K \), we want to place the storage devices to minimize the system operation cost with a total budget of \( b \) 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\(^4\)

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

Each subset \( X \) of \( \Omega \) represents a valid placement decision, and all placement decisions can be represented by a subset of \( \Omega \) if we assume that only one storage device with each type can be placed at each bus. For notational convenience, let \( \mathbb{I} : 2^\Omega \rightarrow \{0,1\}^{\mathbb{N} \times K} \) 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 \( i \)-th entry of the matrix-vector product \( \mathbb{I}(X)x \) can be expressed as

\[ (\mathbb{I}(X)x)_i = \sum_{k: (i,k) \in X} x_k, \]

which is equal to the total storage capacity at bus \( i \). We now introduce a function, \( V : 2^\mathcal{N} \rightarrow \mathbb{R} \), which we call the storage placement value function, defined as

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

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:

\[ \max_{X \subseteq \Omega} \ V(X) \]

subject to

\[ \sum_{k: (i,k) \in X} c_k \leq b. \]

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 post-processing, such as thresholding, to convert the solutions of continuous optimization problems into discrete solutions. However, such post-processing 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 (6b), 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. Lastly, the proposed discrete

\(^4\)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 \( \Omega = \{(i,k) : i \in \mathcal{N}, k = 1, \ldots, K\} \), where \( \mathcal{N} \subseteq \{1, \ldots, n\} \) is the set of buses where placing a storage is possible.
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)gradients of objective functions.

F. Outline of Proposed Analyses

We summarize the analyses conducted in this paper as follows:

1) We characterize optimal locational marginal prices as affine functions of the storage capacity \( x \) by examining the spatio-temporal (or network-storage) congestion pattern of a power network via a dual quadratic program (Theorem 1).
2) Using the results of our dual analysis, we identify a piecewise invariant structure and a closed-form expression of the Hessian \( \nabla^2_{xx} J \) of the optimal cost function in each critical region (Theorem 2).
3) We connect the Hessian \( \nabla^2_{xx} J \) and the submodularity of the storage placement value function \( V \) (Theorem 3).
4) We investigate a polyhedral characterization of each critical region where the Hessian \( \nabla^2_{xx} J \) is invariant. We show that the spatio-temporal congestion pattern of a power network defines the critical regions (Theorem 4).
5) The polyhedral characterization is then used to develop a computational tool for verifying the submodularity of \( V \).

In Steps 1 and 2, we parametrize the multi-period economic dispatch problem with the vector \( x \) of storage capacity by relaxing its domain as \( \mathbb{R}^n \). Step 3 plays an important role in connecting our analysis in the continuous domain to the discrete notion of submodularity. In Step 4, the spatio-temporal congestion pattern of a power network is identified as an essential factor that affects the submodularity of \( V \). Verifying its submodularity via the method proposed in Step 5, we are able to find a near-optimal solution with a provable suboptimality bound via the polynomial-time modified greedy algorithm illustrated in Section IV-C.

III. Structures of Optimal Cost and Prices

A. Dual Analysis

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

Consider the standard dual quadratic program (QP) of (4):

\[
\begin{align*}
\max_{\lambda,\gamma,\beta,\mu,\nu} \quad & \phi(\lambda, \gamma, \beta, \mu, \nu) \\
\text{s.t.} \quad & \lambda_t = \gamma_t 1 - H_t^T \beta_t \quad \forall t \in T, \\
& \lambda_t = L_t^T (\mu_t - \nu_t) \quad \forall i \in N, \\
& \beta, \mu, \nu \geq 0, \\
\end{align*}
\]

where the Lagrange dual function is given by

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

Note that the variable \( \lambda_{i,t} \) represents the locational marginal price (LMP) at bus \( i \) in period \( t \) since from the first order optimality condition of (4) we have

\[
\nabla_{g_i} C_t(g_t) = \lambda_t, \quad t \in T.
\]

A more detailed derivation can be found in [37], [38].

The standard dual QP (7) can be further simplified. Observe that (4d) and (4e), representing the lower and upper limits of state of charge, 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 (7c), which is equivalent to \( (\mu_i - \nu_i) = L^{-T} \lambda_i \), we have

\[
\mu_i = (L^{-T} \lambda_i)^+ \quad \text{and} \quad \nu_i = (L^{-T} \lambda_i)^- \quad \forall i \in N.
\]

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

\[
\mu_i = (\lambda_{t+1} - \lambda_t)^+ \quad \text{and} \quad \nu_i = (\lambda_{t+1} - \lambda_t)^- \quad \forall t \in T, \quad (8)
\]

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

Substituting the expression of \( \mu_i \) into the dual QP, we obtain the following reduced dual program:

\[
\begin{align*}
\max_{\lambda,\gamma,\beta} \quad & \hat{\phi}(\lambda, \beta) \\
\text{s.t.} \quad & \lambda_t = \gamma_t 1 - H_t^T \beta_t \quad \forall t \in T, \\
& \beta \geq 0,
\end{align*}
\]

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

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

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

**Assumption 1 (Flow LICQ):** For each $t \in T$, let $H_{t}^{\text{net}}$ be the collection of $H$’s rows corresponding to the congested (oriented) lines for the flow induced by $(q_{t}^{x}(x), u_{t}^{x}(x))$, when at least one congested line exists in period $t$.\(^{5}\) Then, $H_{t}^{\text{net}}$ is of full row rank for each $t \in T$.

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 $\lambda^{*}(x)$ and $\gamma^{*}(x)$ are unique. Furthermore, if Assumption 1 holds, then $(\lambda^{*}(x), \gamma^{*}(x), \beta^{*}(x))$ is the unique solution to the dual problem (9).

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

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

$$
\nabla_{x} J(x) = - \sum_{t=1}^{T} (\lambda_{t+1}^{*}(x) - \lambda_{t}^{*}(x))^{+},
$$

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

Coined 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 (10) for the case where the marginal cost of generation and marginal benefit of consumption are both constants (i.e., the cost function is a piecewise linear function with two pieces). In fact, the expression (10) holds for any smooth convex cost function under mild regularity assumptions as in the standard sensitivity theorem of nonlinear programming.

When the cost function is nonlinear and the size of the storage to be placed is far from infinitesimal, the first-order approximation of the value function using the gradient formula (10) may not be accurate.\(^{6}\) We now proceed to obtain a finer characterization of the optimal cost $J(x)$ by investigating its higher order derivatives. 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 Second Order Sensitivity

Given that the objective function is quadratic, we expect the curvature (or second-order) information summarized by the Hessian matrix $\nabla_{x}^{2} J(x)$ 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 optimal cost 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 a quadratic function, the optimal LMP vector $\lambda^{*}(x)$ is affine in $x$.

**Remark 1:** The polytopes in Lemma 3 are referred to as critical regions in the literature of multi-parametric quadratic programming (e.g., [40], [41]). 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 $(i, t) \in N \times T$, let $z_{i,t}^{x}(x) := 1$ if the constraint $(L_{i}u_{t})_{j} \leq x_{i}$ is binding at the optimum, $z_{i,t}^{x}(x) = -1$ if the constraint $(L_{i}u_{t})_{j} \geq 0$ is binding at the optimum, and $z_{i,t}^{x}(x) = 0$ otherwise. In other words, $z^{x}$ represents the storage congestion pattern. Under strict complementary slackness, we use (8) to obtain

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

We now let $E_{i}^{x} \subset \{1, \ldots, 2m\}$ denote the set of transmission lines that are congested at the solution in period $t$ and $m_{t} := |E_{i}^{x}|$ denote the number of congested lines. We define the selection matrix $z^{x}_{i,t} \in \mathbb{R}^{m_{t} \times 2m}$ as

$$(z^{x}_{i,t})_{i,j} := (z^{x}_{i,t}(x))_{i,j} = \begin{cases} 1 & \text{if the } i \text{th element in } E_{i}^{x} \text{ is } j \\ 0 & \text{otherwise} \end{cases}$$

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

$$H_{t}^{\text{net}} := z^{x}_{i,t}H.$$ 

Note that $z^{x}_{i,t} = 0$ 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 $z^{x}_{i,t}$ and $z^{x}_{i,t}$, $t \in T$, the optimal locational marginal prices are affine in $x$. In other words, there exist matrix $W_{t}^{x}(z^{x}, z^{x}) \in \mathbb{R}^{n \times n}$ and vector $\lambda_{t}^{x}(z^{x}, z^{x}) \in \mathbb{R}^{n}$ for each $t \in T$ such that

$$\lambda_{t}^{x}(x) = W_{t}(z^{x}, z^{x})x + \lambda_{t}(z^{x}, z^{x}).$$

As a useful byproduct of Theorem 1, we can obtain closed-form expressions of the (reference) energy price $\gamma_{t}^{x} = 1^{T} \lambda_{t}^{x}$ and the congestion price $\beta_{t}^{x}$ as a function of the storage capacity vector $x$.

**Corollary 1:** Under the setting of Theorem 1, $\gamma_{t}^{x}(x)$ and $\beta_{t}^{x}(x)$ are affine functions of $x$ in the critical region containing
functions.

It is instrumental to consider an alternative characterization of the submodularity in Theorem 3 also require all off-diagonal entries of the Hessian matrix to be nonnegative, which does not follow from the fact that $J(x)$ is submodular on $\mathcal{X}$, $e \neq e'$ and $X \subseteq \Omega$. 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 storage placement value function $V: 2^\Omega \rightarrow \mathbb{R}$ is submodular if

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

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

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) $\frac{\partial^2 v(x)}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_i} \left( \frac{\partial v(x)}{\partial x_j} \right)$ 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:

- For $i = j$, storage at bus $i$ has diminishing return;
- 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 properties that have already been established for the optimal value function $J(x)$. 

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

$$D_e F(X) := F(X \cup \{e\}) - F(X \setminus \{e\}).$$

It is straightforward to check that the following lemma provides a necessary and sufficient condition for submodularity [42].

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

$$D_e (D_e F(X)) \leq 0,$$

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

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-E. 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 $\Omega$ 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 $\Omega$, a set function $F: 2^\Omega \rightarrow \mathbb{R}$ is said to be submodular if, for any $X \subseteq Y \subseteq \Omega$ and $e \in \Omega \setminus X$,

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

(16)

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

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

(17)

In our case, (17) implies that the marginal benefit of installing a new storage device is nonnegative and (16) states that this marginal benefit should diminish as more storage devices are connected to the system. Evidently, the nonincreasing property of $V(X)$ follows from the fact that $J(x)$ is nonincreasing (Lemma 1). To check whether $V(X)$ is submodular, it is instrumental to consider an alternative characterization of submodularity based on discrete derivatives defined for set functions.
A. Two-Bus Network

To gain insights into the sign of the off-diagonal entries of the Hessian matrix, 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 certain other load profiles. For simplicity, we use cost functions $C_t(g_t) = \frac{1}{2}g_t^2g_*$, 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 $g^*_t = \lambda^*_t$ 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 for the multi-period 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.

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

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

Fig. 2: Critical regions for the two-bus examples. In the figure with a slight abuse of notation, we use $\mathcal{R}^A_i$ and $\mathcal{R}^B_j$ to denote the $r$th critical region for each case.

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 multi-period economic dispatch problem and obtain the optimal primal dual solution. In all 6 critical regions across these two cases, only the red region in case B, i.e., $\mathcal{R}^B_2$, 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 space limits, we focus on this region for the remainder of this subsection. The optimal cost function in the critical region is

$$J(x) = \frac{1}{2}x^T \begin{bmatrix} 1.5 & -0.5 \\ -0.5 & 1.5 \end{bmatrix} x + \begin{bmatrix} -0.5 \\ -0.5 \end{bmatrix} x + 12.5.$$

Consider a particular storage capacity vector $\tilde{x} = [0.2, 0.2]^T \in \mathbb{R}^2$. The solution of (4) for this given storage capacity is depicted in Fig. 3 by exploiting the observation that storage can be thought of as an inter-temporal link that sends energy into the future and that the multiperiod economic dispatch problem is a form of generalized network flow problem on a time-extended graph where storage edges connect the graph representations of the power network for consecutive time periods $[43], [44]$. In Fig. 3, each node in the graph represents a (bus, time period) pair. The vertical edges of the graph represent the transmission line connecting the two buses, while the 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 $g^*_{i,t}$, 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 for this special case can be made: Given the load and network congestion pattern, the usage of storage links are through the following spatial-temporal flow path $(1, 1) \rightarrow (1, 2) \rightarrow (2, 2) \rightarrow (2, 3)$. In this flow path, the storage capacity at bus 1 and storage capacity at bus 2 complement, instead of substitutes (cf. Theorem 3 and Remark 3) each other. We also notice that optimal prices $\lambda^*$, 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 Using a Polyhedral Characterization of Critical Regions

Albeit the negative Hessian case above appears unlikely to occur in practice, there is no a priori theoretical guarantee that $V$ is submodular. In other words, 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 generally a challenging task, as verifying the submodularity of $V$ by definition involves checking an exponential number of inequalities. Theorem 3 reduces this problem to checking the sign of Hessian entries of a continuous function $J(x)$ on $X$. Theorem 2 provides an expression of the Hessian for almost every $x \in X$, which is invariant in each critical region. Therefore it is sufficient to evaluate the Hessian once per critical region in $X$. 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 the discussion. Upon solving the multi-period economic dispatch problem at $x$, we can identify the set of binding constraints and the associated $z^t_{\text{st}}$ and $z^t_{\text{net}}$ for $t \in T$. The critical region containing $x$ can then be expressed as the set of storage capacity vectors where the storage and network congestion states are not changed.

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

$$
\lambda^t_{x,t+1}(\hat{x}) - \lambda^t_{x,t}(\hat{x}) > 0 \quad \forall (i,t) \in \mathcal{N} \times \mathcal{T} \quad \text{s.t.} \quad z^t_{\text{st}}(x) = 1,
$$

$$
\lambda^t_{x,t+1}(\hat{x}) - \lambda^t_{x,t}(\hat{x}) < 0 \quad \forall (i,t) \in \mathcal{N} \times \mathcal{T} \quad \text{s.t.} \quad z^t_{\text{net}}(x) = -1,
$$

$$
\beta^t_{x,t}(\hat{x}) > 0 \quad \forall (j,t) \in \mathcal{E} \times \mathcal{T}.
$$

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

Given this polyhedral characterization of the critical regions, the iterative construction of all critical regions and the complexity of this process both follow from the standard practice of multi-parametric quadratic programming [40]. Here, we provide a brief description of this procedure. To start, we select an initial point $x \in X_0 := X$ and compute the critical region $R_x$ that contains it by using Theorem 4. Focusing on the part of the critical region inside of $X_0$, i.e., $R_x \cap X_0$ while writing the inequality constraints that define this polytope as $Px \leq q$, we can partition the remaining region in $X_0$ as

$$
X_i := \{ x \in X_0 : P^T x \geq q_i, P_j^T x \leq q_j, \forall j < i \},
$$

where $P_i^T$ is the $i$th row of $P$. Recursively applying this process to $X_i$, we obtain the collection of all critical regions in $X$. This procedure is guaranteed to terminate in a finite number of steps.

The rest of this subsection 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 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 $X$ of interest is a subset of the closure of the critical region containing $x = 0$, i.e., $X \subseteq \bar{R}_0$, with $\bar{R}_0$ defined in Theorem 4.

The immediate consequence of this condition for the purpose of verifying submodularity is as follows [40].

**Lemma 5:** The multi-parametric programming procedure terminates in one step if the small storage condition holds.

Furthermore, if the small storage condition holds, the submodularity of $V$ can be checked using merely the LMP vector $\lambda^t(0)$ and the network congestion pattern $z^t_{\text{net}}(0)$ 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 time periods $t \in T$.

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

Another consequence of small storage is that it is possible to obtain much simplified expressions for the optimal prizes 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 spatial-temporal congestion pattern induced by any given storage capacity vector. We discuss these results in Appendix C.

To verify the small storage condition, we need to check whether the polytope $\bar{R}_0$ contains the box $X = [0, \bar{x}_{\text{max}}]_n$. 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 $\min \{ \rho : \rho \in \mathbb{R}_+ : X \subseteq \rho \bar{R}_0 \}$ with 1, where the scaled set $\rho \bar{R}_0$ is $\{ x : (1/\rho) x \in \bar{R}_0 \}$. Write $\bar{R}_0$ as $\{ x : Px \leq q \}$. This set containment problem can be formulated as the following linear program [45]:

$$
\min_{\rho,\Lambda} \rho
$$

s.t.

$$
\Lambda[I - J] = P
$$

$$
\Lambda[\bar{x}_{\text{max}}1^\top, 0^\top] \leq \rho q
$$

$$
\rho, \Lambda \geq 0.
$$

(20)
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 multi-parametric programming procedure described previously should be used for verifying submodularity.

C. Modified Greedy Algorithms

If the submodularity of the value function is verified, we can employ a (modified) greedy algorithm to obtain a near-optimal solution. In particular, Nemhauser et al. [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 multi-type devices are used ($K > 1$), this standard greedy algorithm may not fully use the diminishing return property as the knapsack (budget) constraint (6b) can cause it to 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 of the greedy algorithm is shown to achieve the same performance guarantee by considering not only the marginal benefit but also the placement cost $c_k$ [46], [47]. Specifically, the modified algorithm optimizes the incremental benefit-cost ratio $\frac{V(X \cup \{(i, k)\}) - V(X)}{c_k}$ 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 1. The first candidate $X_1$ of the solution maximizes the benefit $V$ among all feasible sets of cardinality one or two as shown in Line 1. The second candidate $X_2$ is constructed in a greedy way by locally optimizing the incremental benefit-cost ratio 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 [48].

An interesting feature of the proposed algorithm is that its dependence on the operation model (4) is only through the value function $V$. Therefore, we may substitute our simple operation model with one that captures more detailed characteristics of the storage technologies and power flows 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 subsection, 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 as $f_d$. The optimal cost function $J(x)$ of economic dispatch depends on the realization of $d$. To explicitly show this dependency, 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 (6) 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\footnote{In principle, it is possible to utilize a multi-stage stochastic programming formulation featuring a more detailed information update model for generator and storage operation. This is however less commonly the practice in planning studies due to data availability and computational complexity concerns.}, 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., [49–51]). Specifically, the risk-aware placement problem can be formulated as the following combinatorial stochastic program featuring a mean-risk objective:

$$\max_{X \subseteq \Omega} \quad \mathbb{E}[V(X; d)] - \kappa \rho(V(X; d))$$

s.t. \quad \sum_{k: (i, k) \in X} c_k \leq b, \quad \ldots$$

(21)

where $\rho$ 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

<table>
<thead>
<tr>
<th>Algorithm 1: Modified greedy algorithm for energy storage placement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 $X_1 \leftarrow \text{argmax}{V(X) :</td>
</tr>
<tr>
<td>2 $X_2 \leftarrow \emptyset$;</td>
</tr>
<tr>
<td>3 foreach $X \subseteq \Omega$ s.t. $</td>
</tr>
<tr>
<td>4 Candidates $\leftarrow \Omega \setminus X$;</td>
</tr>
<tr>
<td>5 while Candidates $\neq \emptyset$ do</td>
</tr>
<tr>
<td>6 $e \leftarrow \text{argmax}_{(i, k) \in \text{Candidates}} \frac{V(X \cup {(i, k)}) - V(X)}{c_k}$</td>
</tr>
<tr>
<td>7 if $\sum_{k: (i, k) \in X \cup {e}} c_k \leq b$ then</td>
</tr>
<tr>
<td>8 $X \leftarrow X \cup {e}$;</td>
</tr>
<tr>
<td>9 Candidates $\leftarrow$ Candidates $\setminus {e}$;</td>
</tr>
<tr>
<td>10 end</td>
</tr>
<tr>
<td>11 end</td>
</tr>
<tr>
<td>12 if $V(X) &gt; V(X_2)$ then</td>
</tr>
<tr>
<td>13 $X_2 \leftarrow X$;</td>
</tr>
<tr>
<td>14 end</td>
</tr>
<tr>
<td>15 end</td>
</tr>
<tr>
<td>16 $X^* \leftarrow \text{argmax}_{X \subseteq {X_1, X_2}} V(X);$</td>
</tr>
</tbody>
</table>
density function \( f_d \). If \( \rho \) is a convex risk measure, then the objective function of the stochastic program (21)

\[
E[V(X; d)] - \kappa \rho(V(X; d))
\]
is non-decreasing submodular.

Theorem 5, which implies that submodularity is preserved through our risk-aware formulation, is a strong result as it holds independent 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. [52]. It measures the expected value conditional upon being within some percentage of the worst-case scenarios. Formally, the CVaR of a random variable \( Z \), representing a loss, is defined as

\[
\text{CVaR}_\alpha(Z) := E[Z | Z \leq \text{VaR}_\alpha(Z)], \quad \alpha \in (0, 1),
\]

where the *value-at-risk* (VaR) of \( Z \) (with the cumulative distribution function \( F_Z \)) is given by

\[
\text{VaR}_\alpha(Z) := \inf \{ z \in \mathbb{R} \mid F_Z(z) \geq \alpha \}.
\]

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 [53]:

\[
\text{CVaR}_\alpha(Z) = \inf_{y \in \mathbb{R}} \left[ y + \frac{1}{1 - \alpha} E[ (Z - y)^+ ] \right].
\]

Recalling that \( Z \) represents a loss or cost, we can compute the risk term in our optimization problem as a convex program:

\[
\rho(V(X; d)) = \text{CVaR}_\alpha(-V(X; d))
\]

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

Since CVaR is a convex risk measure, if \( X \mapsto V(X; d) \) is non-decreasing submodular, Theorem 5 allows us to use Algorithm 1 to solve the risk-aware placement problem (21). This additional complexity comes from a one-dimensional convex optimization problem to compute \( \text{CVaR}_\alpha(-V(X; d)) \) at each greedy step. Other risk measures that result in such a bilevel combinatorial-convex optimization problem can be found in [54].

V. NUMERICAL EXPERIMENTS

A. IEEE 14 Bus Case

Our placement algorithms are tested using the IEEE 14 bus test case. Hourly zonal aggregated locational marginal price and load data are obtained from the 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 cost. Quadratic cost coefficients for generators are obtained from MATPOWER [55]. The capacity of each line is set to be the average load per bus over the 24 hours. 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 \( K = 1 \) and denote \( \hat{x}_1 = \hat{x} \).

We now consider placing 5 storage devices in the 14 bus-network, with a total energy capacity of 150 MWh. Using the optimal set containment optimization (20), 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 1 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 with a variety of 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.\(^8\) For larger networks, it is no longer feasible to benchmark the greedy performance against the exact optimal placement which should be identified through an exhaustive search. Therefore, we compare the greedy performance and run-time against that of a mixed-integer quadratic programming (MIQP) solver from Gurobi. 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 subsection 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 (4 IEEE test cases and 10 total storage configurations), the following observations hold:

- The small storage condition is valid and thus the placement value function is submodular;
- The greedy algorithm achieves the same storage placement value (and the system-wide cost) as the MIQP-based method.

\(^8\)We use a random assignment due to a 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.

- 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.

Figure 4 compares the run-time of these two methods, in which we have averaged the run-times across the storage capacities as 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 previous subsection). 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. Although the test results suggest it is likely for the submodularity condition to hold in practice (for fundamental reasons discussed in Section IV-A), it is not guaranteed; our proposed verification procedure is a useful tool to check whether a storage placement optimization problem is submodular. Furthermore, the rather surprising 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. Nevertheless, our algorithm guarantees to provide a $(1 - 1/e)$-optimal solution when the problem instance is verified to be submodular.

VI. CONCLUSIONS

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 at 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 spatio-temporal congestion pattern of a power network is a critical factor for submodularity.

ACKNOWLEDGEMENTS

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. We also thank the anonymous reviewers and Jonathan Mather at UC Berkeley for their insightful comments and suggestions.

APPENDIX A

PROOFS FOR SECTION III

Proof of Proposition 1: As the objective function of (9) is strongly convex in $\lambda$, we know that $\lambda^*(x)$ must be unique. Suppose that the first bus is the reference bus of the network, then by the definition of the shift-factor matrix, $H_1 = 0$. Thus constraint (9b) implies that $\gamma^t(x) = \lambda^*_{1,t}(x)$, and therefore $\gamma^*(x)$ 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, $\beta_t$ can be partitioned into $\tilde{\beta}_t$ for the binding constraints and $\beta''_t$ for the slack constraints for which we know that $\beta''_t = 0$ by complementary slackness. In fact, using this decomposition, the dual constraint (9b) can be written as

$$\lambda_t = \gamma_t 1 - H_{t \text{net}}^T \tilde{\beta}_t, \quad t \in T.$$ 

Now as $H_{t \text{net}}$ is a full row rank matrix, $\tilde{\beta}_t$ is uniquely determined by the equation above given fixed $\lambda_t$ and $\gamma_t$. This implies that $\beta_t$ is also unique.

Proof of Lemma 1: Consider the primal program (4), which has an infinitely differentiable objective function and linear constraints. Under the non-degeneracy condition, we can apply standard sensitivity theorem of nonlinear programming [56], which suggests the differentiability of $J(x)$ and that

$$\frac{\partial J(x)}{\partial x_i} = -\sum_{t=1}^{T} \mu^t_{i,t} = -\sum_{t=1}^{T} (\lambda^*_t(x) - \lambda^*_t(x))^+,\quad i \in N.$$ 

for any $x \in \mathbb{N}$. To show that $\partial J(x)/\partial x_i$ itself is again a continuous function, we observe that $\lambda^*(x)$ is the unique solution of the dual QP (7). By the smoothness of the objective and constraints of (7), we know that the parameter to solution mapping $\lambda^*(x)$ 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 $x$, for each $i \in N$.

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

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

where $g$ is the vector of storage devices installed at each bus and $C_t(g_t)$ is the cost function for the $t$th bus. The objective function is continuous and differentiable in $g$. To show that $J(x)$ is concave in $g$, we need to show that the Hamiltonian $H(x, g, \lambda)$ is concave in $g$ for each $x, \lambda$. The Hamiltonian is defined as

$$H(x, g, \lambda) = \sum_{t=1}^{T} (C_t(g_t) + \lambda_t g_t).$$ 

The function $C_t(g_t)$ is concave in $g_t$ and $\lambda_t$ is a constant. Therefore, $H(x, g, \lambda)$ is concave in $g$ for each $x, \lambda$. This completes the proof.
where extended real-valued function \( b(g,x) \) 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 \( \rho \in [0,1] \), \( b((1-\rho)g^1 + \rho g^2, p x^1 + (1-\rho)p x^2) = 0 \) if \( b(g^i, x^i) = 0 \) for \( i = 1, 2 \), as \( pu^1 + (1-\rho)u^2 \) is a feasible solution for the set of constraints given \((rg^1 + (1-\rho)g^2, p x^1 + (1-\rho)p x^2)\). Therefore, \( J(x) \) is convex as it is the minimum value of another convex function minimized in \( y \) over a convex set.

**Proof of Lemma 3:** This is a standard multi-parametric quadratic programming result. See e.g. [40].

**Proof of Theorem 1 and Corollary 1:** By strict complementarity slackness,

\[
\beta_{t,i} = 0 \quad \text{if} \quad \ell \not\in \mathcal{E}_i^C \quad \text{and} \quad \beta_{t,i} > 0 \quad \text{if} \quad \ell \in \mathcal{E}_i^C.
\]

Thus, we can focus on the reduced dual variable

\[
\hat{\beta}_t := z^\text{net}_t \beta_t \in \mathbb{R}^{m_t}, \quad t \in \mathcal{T}.
\]

For convenience, we denote

\[
\xi_{t,i} = \left(z^\text{net}_{t,i}\right)^+, \quad \eta_{t,i} = \mathbb{1}\{z^\text{net}_{t,i} = 0\}, \quad i \in \mathcal{N}, t \in \mathcal{T}.
\]

Then we can consider the following reduced form of dual program (7) in a neighborhood of \( x \)

\[
\begin{align*}
\max_{\lambda, \gamma, \beta} & \quad \psi(\lambda, \gamma, \beta) \\
\text{s.t.} & \quad \lambda_t = \gamma_t 1 - H_{t}^\text{net}\beta_t, \quad t \in \mathcal{T}, \\
& \quad F(\eta_t)(\lambda_{t+1} - \lambda_t) = 0, \quad t \in \mathcal{T},
\end{align*}
\]

where the objective function is

\[
\psi(\lambda, \gamma, \beta) := \sum_{t=1}^{T} -\frac{1}{2} (\lambda_t - r_t)^\top Q^{-1}(\lambda_t - r_t) + d_t^\top \lambda_t
\]

\[
- \hat{f}_t\beta_t - x^\top \Delta \xi t(\lambda_{t+1} - \lambda_t),
\]

\[
\hat{f}_t := z^\text{net}_t \hat{f},
\]

and the matrix \( F(\eta_t) \in \mathbb{R}^{1\times n} \) with \( u_t := \sum_{\ell \in \mathcal{N}} \eta_{t,i} \) is formed by removing zero rows from diagonal matrix \( \Delta \eta_t \). We can vectorize the variables and coefficients by denoting

\[
\lambda = \begin{bmatrix} \lambda_1^\top, \ldots, \lambda_T^\top \end{bmatrix}^\top, \quad \hat{\beta} = \begin{bmatrix} \hat{\beta}_1^\top, \ldots, \hat{\beta}_T^\top \end{bmatrix}^\top, \quad \gamma = \begin{bmatrix} \gamma_1, \ldots, \gamma_T \end{bmatrix}^\top, \quad r = \begin{bmatrix} r_1^\top, \ldots, r_T^\top \end{bmatrix}^\top, \quad d = \begin{bmatrix} d_1^\top, \ldots, d_T^\top \end{bmatrix}^\top, \quad \hat{f} = \begin{bmatrix} \hat{f}_1^\top, \ldots, \hat{f}_T^\top \end{bmatrix}^\top,
\]

and define matrices

\[
Q = \text{diag}(Q_1, \ldots, Q_T) \in \mathbb{R}^{nT \times nT},
\]

\[
G = \text{diag}(1, \ldots, 1) \in \mathbb{R}^{nT \times n},
\]

\[
H = \text{diag}(H_{1}^\text{net}, \ldots, H_{T}^\text{net}) \in \mathbb{R}^{(\sum_{t \in \mathcal{T}} m_t) \times nT},
\]

\[
F = \begin{bmatrix}
F(\eta_1) & F(\eta_2) & \cdots & F(\eta_{T-1}) & F(\eta_T) \\
\end{bmatrix},
\]

where \( F \in \mathbb{R}^{(\sum_{t \in \mathcal{T}} m_t) \times nT} \). Then the optimization can be written as

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

where \( \kappa = d - Yx \), and

\[
Y = \begin{bmatrix}
-\Delta \xi_1 \\
\Delta \xi_1 - \Delta \xi_2 \\
\vdots \\
\Delta \xi_{T-1} - \Delta \xi_T
\end{bmatrix}.
\]

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

\[
\begin{align*}
\min_{\gamma, \beta} & \quad \frac{1}{2} \left[ \begin{array}{c}
\gamma \\
\beta
\end{array} \right]^\top \begin{bmatrix}
G^\top Q^{-1}G & G^\top Q^{-1}H^\top \\
HQ^{-1}G & HQ^{-1}H^\top
\end{bmatrix} \left[ \begin{array}{c}
\gamma \\
\beta
\end{array} \right] \\
& \quad - \left[ \begin{array}{c}
G^\top \kappa + G^\top Q^{-1}r \\
-H\kappa - HQ^{-1}r - f
\end{array} \right]^\top \left[ \begin{array}{c}
\gamma \\
\beta
\end{array} \right] + \frac{1}{2} r^\top Q^{-1}r
\end{align*}
\]

s.t. \( \begin{bmatrix} FG & -FH^\top \end{bmatrix} \left[ \begin{array}{c}
\gamma \\
\beta
\end{array} \right] = 0 \).

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

\[
\left[ \begin{array}{c}
\gamma \\
\beta
\end{array} \right] = (A^{-1}C^\top(CA^{-1}C^\top)^{-1}CA^{-1} - A^{-1}) B.
\]

Substituting this expression into \( \lambda = G\gamma - H^\top \hat{\beta} \), we have

\[
\lambda = \begin{bmatrix} G & -H^\top \end{bmatrix} (A^{-1}C^\top(CA^{-1}C^\top)^{-1}CA^{-1} - A^{-1}) B.
\]

(24)

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

**Proof of Theorem 2:** We can compute the optimal value \( J \) by substituting the expression of the optimal \( \gamma \) and \( \beta \), as a result

\[
J(x) = \frac{1}{2} B^\top (A^{-1} - A^{-1}C^\top(CA^{-1}C^\top)^{-1}CA^{-1} - A^{-1}) B - \frac{1}{2} r^\top Q^{-1}r.
\]

Recognizing that

\[
B = \begin{bmatrix} G^\top & \quad -H \\
-H^\top & \quad -HQ^{-1}r - f
\end{bmatrix}
\]

we obtain the Hessian for the critical region containing \( x \) as

\[
Y^\top (G - H^\top)(A^{-1} - A^{-1}C^\top(CA^{-1}C^\top)^{-1}CA^{-1}) \begin{bmatrix}
G^\top \\
-H
\end{bmatrix} Y
\]

by twice differentiating \( J \) with respect to \( x \).
APPENDIX B

PROOFS FOR SECTION IV

Proof of Theorem 3: For any $X \subseteq \Omega$, and without loss of generality $e_i := (i, k_i) \notin X$, $e_j := (j, k_j) \notin X$, we have $D_{e_i} V(x) = V(X \cup \{(i, k_i)\}) - V(X)$ and

$$D_{e_j} (D_{e_i} V(X)) = [V(X \cup \{(i, k_i), (j, k_j)\}) - V(X \cup \{(j, k_j)\})] - [V(X \cup \{(i, k_i)\}) - V(X)].$$

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

$$D_{e_i} (D_{e_i} V(X)) = [J(x^0 + \bar{x}_i 1_j) - J(x^0 + \bar{x}_i 1_i + \bar{x}_j 1_j)] - [J(x^0) - J(x^0 + \bar{x}_i 1_i)].$$

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

$$D_{e_i} (D_{e_i} V(X)) = \int_{x_i}^{\bar{x}_i} \left[ \frac{\partial J}{\partial x_i} (x^0 + \xi 1_i) - \frac{\partial J}{\partial x_i} (x^0 + \bar{x}_i 1_j + \xi 1_i) \right] d\xi.$$

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

$$D_{e_i} (D_{e_i} V(X)) = -\int_{\bar{x}_i}^{x_i} \int_{\bar{x}_j}^{\bar{x}_i} \frac{\partial^2 J}{\partial x_j \partial x_i} (x^0 + \xi 1_i + \xi 1_j) d\xi d\zeta.$$

As $(\nabla^2_{xx} J(x))_{ij} \geq 0$, we have $D_{e_i} (D_{e_i} V(X)) \leq 0$ for any $i, j \in N$ and any $k_i$ and $k_j$. Thus, using (18), 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 dualQP (9). In other words, $(\lambda^\star(x), \gamma^\star(x), \beta^\star(x))$ is the unconstrained local maximizer of (9) in an affine subspace defined by the set of equality and inequality constraints in (9) which are binding at $x$. Recall that, given the storage congestion state $z^{st} \in \mathbb{R}^{n \times T}$, the objective function of (9) can be written as

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

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 $z^{st}(\hat{x})$ given by (11) is unchanged from $z^{st}(x)$. Therefore, $(\lambda^\star(\hat{x}), \gamma^\star(\hat{x}), \beta^\star(\hat{x}))$ is still the unconstrained local maximizer of (9) in the same affine subspace when $x$ is replaced with $\hat{x}$ in the above expression of the objective function $\phi$. The second inequality condition in Theorem 4 guarantees that $(\lambda^\star(\hat{x}), \gamma^\star(\hat{x}), \beta^\star(\hat{x}))$ is feasible for (9). Indeed, we observe from the expression of $\beta^\star(\hat{x})$ in (15) that modifying $x$ does not affect $\beta^\star_j(x)$ for a line $j$ that is uncongested and hence $\beta^\star_j(x)$ is 0 if line $j$ is uncongested, while the second inequality condition in Theorem 4 ensures that $\beta^\star_j(x)$ 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 dualQP (9) as the problem is concave. By strict complementary slackness, the conditions defining $R_x$ ensure that the set of binding inequality constraints is unaltered with $\hat{x}$. Therefore $R_x$ is the critical region containing $x$.

Proof of Theorem 5: Since submodularity is preserved through addition, $E[V(X; d)]$ is submodular. On the other hand, $\rho$ is nonincreasing and convex since 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 submodular [57]. Therefore, we observe that $\rho(V(X; d))$ is nonincreasing submodular. Finally, we again use the fact that submodularity is preserved through addition to conclude that $E[V(X; d)] - \kappa \rho(V(X; d))$ is nondecreasing submodular for any $\kappa \geq 0$.

APPENDIX C

PRICES AND HESSIAN UNDER SMALL STORAGE

Under the small storage condition, the dual program (22) 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 locational marginal prices, 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 locational marginal prices as functions of the storage capacity vector $x$. Recall that $\xi_{it} = (z_{it}^{net})^\top$.

Theorem 6: Suppose the small storage condition is in force and the loads are heterogeneous such that $\Lambda_{it}(0) \neq \Lambda_{i,t+1}(0)$ for all $i$ and $t$. Then we have

$$\lambda^t_i(x) = A_i(z^{net}_t) \Delta(\xi_{i,t-1}) x + \hat{\lambda}_i(z^{net}_t),$$

$$\gamma^t_i(x) = 1^\top (A_i(z^{net}_t) \Delta(\xi_{i,t-1}) x + \hat{\lambda}_i(z^{net}_t)),$$

$$\beta^t_i(x) = z^{net}_t^\top B_i(z^{net}_t) \Delta(\xi_{i,t-1}) x + \hat{\beta}_i(z^{net}_t),$$

where

$$A_i(z^{net}_t) := \frac{1}{1^\top Q_t^{-1}} 11^\top + Q_t M_t R_t M_t Q_t,$$

$$B_i(z^{net}_t) := Q_t M_t H_t^{net} K_t^{-1},$$

$$\lambda_i(z^{net}_t) := A_i(z^{net}_t) (d_t + Q_t^{-1} r_t) + B_i(z^{net}_t) \xi_{it},$$

$$\hat{\beta}_i(z^{net}_t) := -z^{net}_t^\top B_i(z^{net}_t) (d_t + Q_t^{-1} r_t) - z^{net}_t^\top K_t^{-1} z^{net}_t \xi_{it}.$$
where
\[ \psi_t(\lambda_t, \gamma_t, \beta_t) := \frac{1}{2}(\lambda_t - r_t)^\top Q_t^{-1}(\lambda_t - r_t) + d_t^\top \lambda_t - \ell_t \beta_t + x^\top \Delta(\xi_t - \xi_{t-1}) \lambda_t. \]

This equality constrained quadratic program 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:

\[ \lambda_t^* = (Q_t M_t R_t M_t Q_t + p_t 11^\top)(\Delta(\xi_t - \xi_{t-1}))x + d_t + Q_t^{-1}r_t \]
\[ + Q_t M_t H_t^{\text{net}} K_t^{-1} \ell_t, \]
\[ \beta_t^* = -K_t^{-1}[H_t^{\text{net}} M_t Q_t \Delta(\xi_t - \xi_{t-1})x + Q_t d_t r_t + r_t + \ell_t], \]

where \( p_t := 1/[1^\top Q_t^{-1} 1] \), 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 \( B_t(z_t^{\text{net}}, \ell_t) \), 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 \( \nabla_x^2 J(x) \) exists,

\[ \nabla_x^2 J(x) = (\nabla^2 J)^{\text{st}}(x) + (\nabla^2 J)^{\text{net}}(x), \]

where \((\nabla^2 J)^{\text{st}}(x)\) is the component that depends on the storage congestion pattern while \((\nabla^2 J)^{\text{net}}(x)\) is the component that depends on the network congestion pattern, defined as

\[ (\nabla^2 J)^{\text{st}}(x) := \sum_{t=1}^T (\xi_t - \xi_{t-1})(\xi_t - \xi_{t-1})^\top, \]
\[ 1^\top Q_t^{-1} 1, \]
\[ (\nabla^2 J)^{\text{net}}(x) := \sum_{t=1}^T \Delta(\xi_t - \xi_{t-1}) Q_t M_t R_t M_t Q_t \Delta(\xi_t - \xi_{t-1}). \]

**Proof:** Inserting the expressions of \( \lambda_t, \gamma_t, \beta_t \) given in Theorem 1 and Corollary 1 into (30), we obtain the optimal cost as

\[ J_t(x) = \frac{1}{2} x^\top \Delta(\xi_t - \xi_{t-1}) A_t(z_t^{\text{net}}) x + [A_t(z_t^{\text{net}})](d_t + Q_t^{-1} r_t) + B_t(z_t^{\text{net}}, \ell_t) \Delta(\xi_t - \xi_{t-1})x + J_t, \]

where the last term \( J_t \) 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. Proposition 2 in [1] reveals how the storage congestion pattern term of the Hessian is related to the alignment of LMP time series across the network. 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.