Scheduling Distributed Energy Resources under Limited Observability of Distribution Grids

Vassilis Kekatos, *Senior Member, IEEE*, Ridley Annin, *Student Member, IEEE*, Manish K. Singh, *Senior Member, IEEE*, and Junjie Qin, *Member, IEEE*

***Abstract*—Distributed energy resources (DERs) should be scheduled in a coordinated manner to postpone or avoid costly capacity upgrades. Nonetheless, the pervasive lack of data at the distribution grid edge impacts the effectiveness of real-time scheduling of DERs using the optimal power flow (OPF). Before meters are widely deployed, distribution system operators (DSOs) may have to coordinate DERs using limited data. The goals of this work are to identify which data are most important for the OPF, and solve the OPF using only this subset of selected data. The two goals can be accomplished simultaneously upon formulating a feature selection task that incorporates the OPF module, and gives rise to a nonconvex optimization problem. Stationary points of this problem can be reached using an accelerated proximal gradient algorithm. The proposed methodology is applicable to a broad range of OPF formulations. Nonetheless, linearized OPF variants posed as linear or quadratic programs feature expedited gradient calculations thanks to the neat structure of multiparametric programming. Numerical tests on a benchmark distribution feeder demonstrate that the degradation in the system's performance when running the OPF using less than half of the data is insignificant.**

***Index Terms*—Multiparametric programming, proximal gradient, optimal meter placement, smart meter data compression, group lasso, compressed sensing.**

I. Introduction

Unlike power transmission systems, distribution grids are currently challenged by limited observability. Meter deployment has been relatively limited, and smart meter readings collected at the grid edge are communicated back to the DSO with time delays or during nighttime to reduce the communication burden. Customer privacy considerations could confine the granularity of smart meter data, and concerns about cyber data attacks call for enhancing the security of customer-to-DSO communication links. At the same time, there is a pressing need for grid scheduling at finer temporal scales to reduce the magnitude and frequency of peak power injections. The predicament of real-time data scarcity can thereby hinder the effective integration of DERs. In this context, this work explores options for optimally scheduling DERs using partial grid data.

Reference [1] studies the observability of a distribution grid using smart meter data. Under limited observability, one may forego estimating the grid state and aim at scheduling DERs directly from partial data. Suppose DERs are scheduled centrally by the DSO by solving the OPF every few minutes. The OPF can be seen as a mapping between OPF input data (such as load demands) $\mathbf{x}$ and OPF solutions (DER schedules) $\mathbf{y}(\mathbf{x})$. The mapping has been studied in [2], [3], and its sensitivities (partial derivatives) have been computed in [4, Ch. 6]. When $\mathbf{x}$ is uncertain, a DSO can resort to stochastic, robust, and chance-constrained renditions of the OPF to de-risk the grid's safe operation at the expense of increased operational costs [5]. Ways to secure customers' private load demand information and their effect on OPF solutions are discussed in [6], [7].

Previous works have tried to compress the OPF model to expedite grid planning and strategic investment operations [8], [9], [10], [11]. Rather than reducing the OPF model, the goal here is to reduce the OPF input data and leave the OPF solver intact. This ensures that the OPF solver is backward-compatible and independent of data throughout specifications. Recent works propose learning OPF solutions directly from primitive data [12], [13]. For example, rather than feeding solar power data at all buses to the OPF, the DSO can feed solar irradiance data for a neighborhood to directly predict DER setpoints. Although such approaches are reasonable for solar data, they may be ineffective when controlling electric vehicles. Moreover, our proposed methodology can be conflated with the previous ideas.

Selecting which data are important for the OPF is related to feature selection in machine learning [14], wherein one tries to infer a target variable using only a subset of features. Here, the target variable is the OPF solution, and the OPF as an optimization problem comes as an integral part of the overall learning model. Data selection has appeared before in power system operations, such as in optimal meter placement for state estimation [1], or feedback controller design [15].

This work takes a fresh look at data selection having the OPF performance in mind. The technical contributions are on three fronts: *i)* Put forth a general framework for data selection and reconstruction for solving the OPF using partial data (Section II); *ii)* Devise a proximal gradient descent algorithm for solving the associated optimization problem (Section III); and *iii)* Expedite gradient computations using advances from multiparametric programming when the OPF is posed as a linear or quadratic program (Section IV). Numerical tests on the IEEE 37-bus distribution feeder corroborate the efficacy of the proposed data selection scheme. Section VI summarizes our findings and sketches open research directions.

Notation: Column vectors (matrices) are denoted by lower- (upper-) case letters. Symbol $(\cdot)^\top$ stands for transposition; $\mathbf{I}_N$ is the $N \times N$ identity matrix; $\|\mathbf{x}\|_2$ is the ℓ_2-norm of vector $\mathbf{x}$; and $\|\mathbf{X}\|_F$ is the Frobenius matrix norm.

II. Problem Statement and Formulation

To optimally schedule distributed energy resources (DERs), a distribution system operator (DSO) would need to solve some rendition of the optimal power flow (OPF) every few minutes. To solve the OPF, the DSO must know the feeder topology as well as reasonable predictions of load demand and DER flexibility across all buses. Although the feeder topology can be considered known and relatively time-invariant, load demand and DER flexibility may vary significantly throughout the day and within short periods. Therefore, the OPF can be interpreted as a parametric optimization model as shown in Figure 1. Load demands and DER forecasts constitute the problem parameters, which will be termed *OPF features* and abstractly denoted by a L-long vector $\mathbf{x}$. DER schedules constitute the OPF minimizers or *OPF solutions* and will be denoted by the N-long vector $\mathbf{y}$. The OPF solution associated with OPF features $\mathbf{x}$ is denoted by $\mathbf{y}(\mathbf{x})$. The parametric OPF can be abstractly expressed as:

$$\begin{aligned} \mathbf{y}(\mathbf{x}) := \arg\min_{\mathbf{y}\in\mathcal{Y}} \quad & f_0(\mathbf{y};\mathbf{x}) \\ \text{subject to (s.to)} \quad & f_m(\mathbf{y};\mathbf{x}) \leq 0, \quad m = 1:M \end{aligned} \tag{1}$$

where function f_0 models the OPF objective and functions $\{f_m\}_{m=1}^M$ the M OPF constraints. For simplicity, the feasible set of DER schedules $\mathcal{Y}$ is assumed to be independent of $\mathbf{x}$.

To find $\mathbf{y}(\mathbf{x})$, the DSO should ideally receive smart meter readings from all buses to acquire OPF features $\mathbf{x}$ in real time. This may be technically impractical due to networking and data processing limitations, and/or cyber-attack and data privacy concerns. Alternatively, the DSO may collect measurements from different grid sensors and find the power system state (typically defined as the vector of complex voltages across all buses) using state estimation. Knowing the grid state, the DSO can readily compute $\mathbf{x}$ through the power flow equations. Unfortunately, unlike power transmission systems, distribution grids are largely unobservable given the currently limited level of meter deployment.

As a practically relevant compromise, the DSO could communicate in real time with only a few buses to acquire a partial view of OPF features. Suppose that the DSO affords to collect only K out of the L OPF features. It is reasonable that this feature selection process is completed in advance: The DSO must have selected OPF features before real-time operation and established a secure communication link with them. For example, the DSO may select OPF features every few days or until a major change occurs on the distribution feeder. Under this setting, two key technical questions arise:

q1) How to select K out of L OPF parameters or features?
q2) How can the OPF be solved using only partial data?

The answers to these questions depend on the statistical properties of OPF features, the feeder topology, and the general structure of the OPF. Suppose that the DSO has a dataset $\mathcal{X} = \{\mathbf{x}_s\}_{s=1}^S$ of S OPF scenarios. Such scenarios can be sampled from historical records of OPF features using archived smart meter data. To account for future operating conditions, the DSO can synthetically generate OPF features based on anticipated levels of DER penetration and load growth. Questions related to generating $\mathcal{X}$ in a manner truthful to anticipated grid loading conditions are not discussed here.

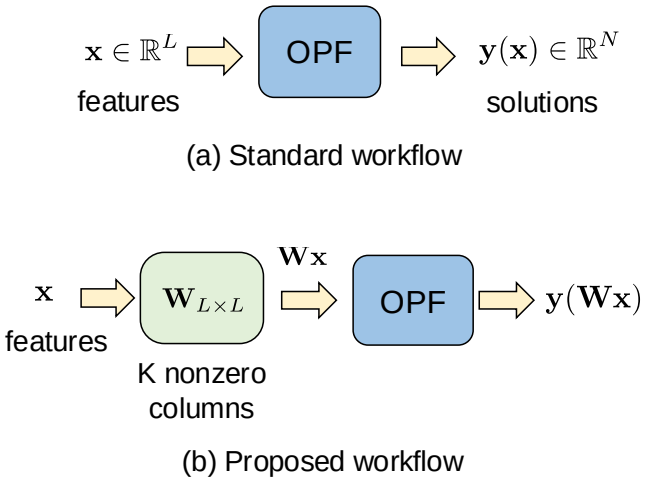

Fig. 1. *a)* According to the conventional workflow, the OPF is fed with OPF input parameters or features stored in vector $\mathbf{x} \in \mathbb{R}^L$ and returns the OPF solution $\mathbf{y}(\mathbf{x}) \in \mathbb{R}^N$. *b)* According to the proposed workflow, the OPF is fed with reconstructed data $\hat{\mathbf{x}} = \mathbf{W}\mathbf{x} \in \mathbb{R}^L$. Because matrix $\mathbf{W}$ has K nonzero columns, the reconstructed data depend linearly on only K out of the L OPF features. This effects joint data selection and linear reconstruction. The idea carries over to nonlinear reconstruction by using a neural network whose first (input) layer has a weight matrix that is also K column-sparse.

How can we use the K selected features to feed into the OPF, given that the OPF solver accepts L rather than K features? One idea would be to use a machine learning (ML) model taking the K selected features as its input and returning an L-long vector as its output to be fed into the OPF. To keep the exposition uncluttered, we restrict our study to a linear model, according to which the K selected features are linearly combined to construct the L-long vector. More specifically, given the OPF feature vector $\mathbf{x} \in \mathbb{R}^L$, the vector to be fed into the OPF can be expressed as:

$$\hat{\mathbf{x}} = \mathbf{W}\mathbf{x} = \sum_{\ell=1}^{L} \mathbf{w}_\ell x_\ell \tag{2a}$$

$$\text{where} \quad \mathbf{W} := [\mathbf{w}_1 \ \mathbf{w}_2 \ \ldots \ \mathbf{w}_L] \tag{2b}$$

where $\mathbf{W}$ is an $L \times L$ matrix partitioned column-wise as shown in (2b) and x_ℓ is the ℓ-th OPF feature. To enable OPF feature selection, matrix $\mathbf{W}$ should be *column-sparse*. If $\mathbf{w}_\ell = \mathbf{0}_L$, feature x_ℓ does not contribute to $\hat{\mathbf{x}}$. In particular, we would like $\mathbf{W}$ to have only K nonzero columns, so that the constructed OPF feature vectors $\hat{\mathbf{x}}$ depend only on K of the original OPF features. Under this requirement, matrix $\mathbf{W}$ serves the joint purpose of OPF feature selection and reconstruction. Therefore, addressing questions *q1)–q2)* is tantamount to designing $\mathbf{W}$.

Given matrix $\mathbf{W}$, the OPF is provided with feature vector $\hat{\mathbf{x}} = \mathbf{W}\mathbf{x}$ instead of $\mathbf{x}$. The associated OPF solution is now $\mathbf{y}(\mathbf{W}\mathbf{x})$ instead of $\mathbf{y}(\mathbf{x})$. Of course, the OPF feature vector $\mathbf{W}\mathbf{x}$ is only a proxy of the actual grid loading conditions. The OPF solution $\mathbf{y}(\mathbf{W}\mathbf{x})$ would be applied to the grid while the grid experiences $\mathbf{x}$. Matrix $\mathbf{W}$ can be designed based

on different metrics. To showcase the main idea, the metric considered here is fidelity in terms of the OPF solution.

Given dataset $\mathcal{X}$, matrix $\mathbf{W}$ can be optimally designed as the solution to the optimization problem:

$$\min_{\mathbf{W}\in\mathbb{R}^{L\times L}} \quad f(\mathbf{W})+\lambda g(\mathbf{W}) \tag{3}$$

$$\text{where} \quad f(\mathbf{W}):=\frac{1}{2S}\sum_{s=1}^{S}\|\mathbf{y}(\mathbf{x}_s)-\mathbf{y}(\mathbf{W}\mathbf{x}_s)\|_2^2$$

$$\text{and} \quad g(\mathbf{W}):=\sum_{\ell=1}^{L}\|\mathbf{w}_\ell\|_2$$

and $\lambda>0$ is a tuning parameter. The objective function involves two terms. Function $f(\mathbf{W})$ measures the fidelity in the Euclidean sense between the OPF solutions associated with the actual feature vector $\mathbf{x}_s$ and those associated with the reconstructed feature vector $\mathbf{W}\mathbf{x}_s$. Function $g(\mathbf{W})$ has been used as a penalty in statistical learning problems to promote solutions that are column-sparse; see for example the group-Lasso approach [14]. It is easy to show that solving (3) for larger λ would yield solutions with smaller value of $g(\mathbf{W})$.

Despite being neatly expressed, solving (3) is non-trivial. For example, evaluating $f(\mathbf{W})$ entails solving $2S$ instances of the OPF: To compute $\{\mathbf{y}(\mathbf{x}_s)\}_{s=1}^S$, we need to solve the OPF for $\{\mathbf{x}_s\}_{s=1}^S$; these OPFs can be solved once as they do not depend on $\mathbf{W}$. To compute $\{\mathbf{y}(\mathbf{W}\mathbf{x}_s)\}_{s=1}^S$ however, we need to solve the OPF for $\{\mathbf{W}\mathbf{x}_s\}_{s=1}^S$ each time $\mathbf{W}$ changes. Heed that the term $g(\mathbf{W})$ is used as a surrogate to enforce column-sparse solutions. Otherwise, finding a $\mathbf{W}$ with exactly K nonzero columns would entail solving a mixed-integer program, which scales unfavorably with N and S. To avoid the computational burden, we propose handling (3) using a first-order iterative optimization algorithm.

III. Solution Methodology

We suggest solving (3) using a proximal gradient descent (PGD) algorithm. The proximal gradient method is a generalization of the gradient descent method for solving optimization problems involving non-differentiable terms in their objective. Let PGD iterations be indexed by t. At iteration t, the PGD method updates $\mathbf{W}^t$ to $\mathbf{W}^{t+1}$ in two steps. It first performs a standard gradient descent step using the derivative of $f(\mathbf{W})$ at $\mathbf{W}^t$ to compute an intermediate variable:

$$\mathbf{Z}^t:=\mathbf{W}^t-\mu\nabla_{\mathbf{W}}f(\mathbf{W}^t). \tag{4}$$

The sought gradient can be computed as:

$$\nabla_{\mathbf{W}}f(\mathbf{W})=\frac{1}{S}\sum_{s=1}^{S}\nabla_{\mathbf{W}\mathbf{x}_s}^{\top}\mathbf{y}(\mathbf{W}\mathbf{x}_s)\left(\mathbf{y}(\mathbf{W}\mathbf{x}_s)-\mathbf{y}(\mathbf{x}_s)\right)\mathbf{x}_s^{\top}.$$

This follows from the chain rule as the gradient of $\mathbf{y}$ with respect to $\mathbf{W}$ can be expressed as the Jacobian matrix of $\mathbf{y}$ with respect to $\mathbf{W}\mathbf{x}_s$, times the Jacobian of $\mathbf{W}\mathbf{x}_s$ with respect to $\mathbf{W}$. Note that computing $\nabla_{\mathbf{W}}f(\mathbf{W}^t)$ requires not only solving the OPF to find $\mathbf{y}(\mathbf{W}\mathbf{x}_s)$, but also finding its Jacobian matrix of partial derivatives with respect to its input feature vector $\mathbf{W}\mathbf{x}_s$. We will return to this later in Section IV.

Having computed $\mathbf{Z}^t$, the PGD method proceeds by updating $\mathbf{W}^{t+1}$ as the minimizer of the convex problem:

$$\mathbf{W}^{t+1}:=\arg\min_{\mathbf{W}}\ \frac{1}{2\mu}\|\mathbf{W}-\mathbf{Z}^t\|_F^2+\lambda g(\mathbf{W}) \tag{5}$$

where $\|\cdot\|_F$ denotes the matrix Frobenius norm. Because the objective in (5) is separable across the columns of $\mathbf{W}$, the problem decouples across $\mathbf{w}_\ell$ as:

$$\mathbf{w}_\ell^{t+1}:=\arg\min_{\mathbf{w}_\ell}\ \frac{1}{2\mu}\|\mathbf{w}_\ell-\mathbf{z}_\ell^t\|_2^2+\lambda\|\mathbf{w}_\ell\|_2,\quad \ell=1:L \tag{6}$$

where $\mathbf{z}_\ell^t$ is the ℓ-th column of $\mathbf{Z}^t$. It can be shown that (6) enjoys the closed-form solution:

$$\mathbf{w}_\ell^{t+1}:=\left[1-\frac{\lambda\mu}{\|\mathbf{z}_\ell^t\|_2}\right]_+\cdot\mathbf{z}_\ell^t,\quad \ell=1:L \tag{7}$$

where the operator $[x]_+$ returns x if $x\geq 0$; and zero, otherwise. To improve the convergence rate of the method, one can try implementing accelerated variants of PGD. In particular, the accelerated PGD variant proposed in [16] bypasses stalling issues of PGD in non-convex problems. Although this method is guaranteed to converge to a stationary point of (3), it entails solving $4S$ rather than S OPF instances per PGD iteration.

Our proposed methodology is applicable regardless of the particular form of the OPF as long as (1) involves cost and constraint functions that are jointly differentiable in $(\mathbf{y},\mathbf{x})$. Following standard industry practice, to avoid scenarios where (1) becomes infeasible, network constraints can be handled as a *soft constraints*. In addition to solving the OPF, we should also be able to perform its sensitivity analysis and compute the Jacobian matrix of the OPF solution with respect to the OPF features. Fortunately, that is indeed the case for several OPF formulations. For example, one may adopt an approximate linearized grid model and formulate the OPF as a linear or quadratic program (LP/QP), whose sensitivity analysis can be found in [17], [18]. One may adopt the exact AC power flow model for a radial single-phase feeder and arrive at the second-order cone program (SOCP) formulation of the OPF, whose sensitivity is analyzed in [3]. For multi-phase feeders and meshed power transmission systems, one may pose the OPF as a semidefinite program (SDP); see [19] for the related sensitivity analysis. It is worth emphasizing that regardless of the particular formulation, computing the sensitivities of the OPF is computationally as complex as solving a system of linear equations. To put it differently, if the optimal primal and dual OPF solutions have already been computed, finding $\nabla_{\mathbf{x}}\mathbf{y}$ comes at minimal extra computational cost.

It should be clear by now that even though the proposed methodology is widely applicable, it can be computationally formidable as it entails solving multiple instances of the OPF per PGD iteration. To alleviate this burden, two options can be pursued:

- *o1)* Leverage the fact that these multiple OPF instances are not solved independently, but in *batch*.
- *o2)* Replace PGD by a *stochastic (minibatch) PGD* wherein each iteration involves only one or a few scenarios.

The two options are not mutually exclusive. We next showcase how *o1)* can be beneficial when the OPF is instantiated as a

multiparametric LP/QP. This is of practical interest considering that the OPF is oftentimes approximated by an LP/QP. Even if the DSO solves an AC-OPF during real-time operation, it is reasonable that the OPF feature selection and reconstruction process relies on an approximate OPF model during planning given the uncertainties involved.

IV. Accelerated Gradient Computations

This section leverages results from multiparametric programming (MPP) to expedite the gradient step in (4). Suppose that the OPF in (1) can be expressed as a multiparametric linear or quadratic program:

$$\mathbf{y}(\mathbf{x}) = \arg\min_{\mathbf{y}} \quad \mathbf{y}^\top \mathbf{H}\mathbf{y} + (\mathbf{A}\mathbf{x} + \mathbf{b})^\top \mathbf{y} \tag{8a}$$

$$\text{s.to} \quad \mathbf{C}_i\mathbf{y} \leq \mathbf{D}_i\mathbf{x} + \mathbf{e}_i \tag{8b}$$

$$\mathbf{C}_e\mathbf{y} = \mathbf{D}_e\mathbf{x} + \mathbf{e}_e \tag{8c}$$

where matrices $(\mathbf{H}, \mathbf{A}, \mathbf{C}_i, \mathbf{D}_i, \mathbf{C}_e, \mathbf{D}_e)$ and vectors $(\mathbf{b}, \mathbf{e}_i, \mathbf{e}_e)$ are known and of conformable dimensions. MPP studies how the OPF solutions $\mathbf{y}(\mathbf{x})$ depend on $\mathbf{x}$. The space of feature vectors $\mathbf{x}$ can be partitioned into *critical regions*, defined as follows. When solving (8) for all $\mathbf{x}$ falling in a critical region $\mathcal{C}_k$, the same subset of inequality constraints in (8b) becomes active or binding, that is, are satisfied with equality. Interestingly, under mild technical conditions, critical regions satisfy the ensuing properties:

p1) Each critical region is a polytope defined as

$$\mathcal{C}_k := \{\mathbf{x} : \mathbf{F}_k\mathbf{x} \leq \mathbf{g}_k\}.$$

p2) The OPF solutions for all OPF features belonging to the same critical region are affine functions of the features, i.e., there exist $(\mathbf{H}_k, \mathbf{k}_k)$ such that

$$\mathbf{y}(\mathbf{x}) = \mathbf{H}_k\mathbf{x} + \mathbf{k}_k, \quad \forall \mathbf{x} \in \mathcal{C}_k$$

and obviously $\nabla_{\mathbf{x}}\mathbf{y} = \mathbf{H}_k$ for all $\mathbf{x} \in \mathcal{C}_k$.

Matrices $(\mathbf{F}_k, \mathbf{H}_k)$ and vectors $(\mathbf{g}_k, \mathbf{k}_k)$ can be readily computed if the binding constraints are known.

Thanks to MPP results, computing the gradient $\nabla_{\mathbf{W}} f(\mathbf{W}^t)$ can be expedited through the following steps:

s0) Initialize $\mathbf{G}^t = \mathbf{0}_{N\times L}$ and $k = 0$.

s1) Given $\mathbf{W}^t$, compute OPF feature vectors $\{\mathbf{W}^t\mathbf{x}_s\}_{s=1}^S$.

s2) Solve (8) for a randomly selected scenario and increase $k := k + 1$.

s3) Define a new critical region $\mathcal{C}_k$ by computing $(\mathbf{F}_k, \mathbf{H}_k, \mathbf{g}_k, \mathbf{k}_k)$ based on the binding constraints.

s4) All scenarios satisfying $\mathbf{F}_k\mathbf{W}^t\mathbf{x}_s \leq \mathbf{g}_k$ belong to region $\mathcal{C}_k$. The corresponding OPF solutions and Jacobian matrices can be computed from *p2)*. Therefore, the sought gradient can be updated as

$$\begin{aligned}\mathbf{G}^t &:= \mathbf{G}^t + \sum_{s\in\mathcal{C}_k} \mathbf{H}_k^\top(\mathbf{H}_k\mathbf{W}^t\mathbf{x}_s - \mathbf{y}(\mathbf{x}_s))\mathbf{x}_s^\top \\ &:= \mathbf{G}^t + \mathbf{H}_k^\top\mathbf{H}_k\mathbf{W}^t\sum_{s\in\mathcal{C}_k}\mathbf{x}_s - \mathbf{H}_k^\top\sum_{s\in\mathcal{C}_k}\mathbf{y}(\mathbf{x}_s)\mathbf{x}_s^\top.\end{aligned}$$

s5) Remove the scenarios with $s \in \mathcal{C}_k$ from the dataset. Unless the dataset is now empty, return to step *s2*.

s6) Output the final gradient evaluation as

$$\nabla_{\mathbf{W}} f(\mathbf{W}^t) = \frac{1}{S}\mathbf{G}^t.$$

The advantage of the previous algorithm is that instead of solving S instances of the OPF, we can solve as many OPFs as the critical regions. The computational advantage apparently depends on the number of critical regions, which depends indirectly on the number of nonzero columns of $\mathbf{W}^t$.

V. Numerical Tests

The proposed method was evaluated using the IEEE 37-bus distribution feeder, converted to single-phase. The goal of the OPF was to optimally select the reactive power setpoints of rooftop photovoltaics (PVs). We adopted an approximate linearized power flow model, and formulated the OPF as a QP, solved in MATLAB using YALMIP and Sedumi. Reactive power setpoints were decided so they minimize the ohmic losses on lines while maintaining voltage magnitudes within the range of $\pm 3\%$ per unit.

The feeder hosted 10 PVs, installed on buses $\{2, 4, 7, 9, 11, 14, 17, 20, 22, 25\}$. Solar profiles were extracted using solar data from the Pecan Street database, and scaled to match twice the peak loads. Loads were also extracted from the Pecan Street database as minute-based kW readings collected between 8 AM and 5 PM during year 2019. Lacking kVAR load demands, we simulated power factors as uniformly randomly sampled between 0.9 lagging and 1. Each load bus also hosted three electric vehicles, assumed to be charging in an uncoordinated manner based on synthetically generated EV data from [20].

We first explored how the fidelity in terms of OPF features $\sum_{s=1}^S \|\mathbf{x}_s - \mathbf{W}\mathbf{x}_s\|_2^2$ and in terms of OPF solutions $\sum_{s=1}^S \|\mathbf{y}(\mathbf{x}_s) - \mathbf{y}(\mathbf{W}\mathbf{x}_s)\|_2^2$ varies with K. The OPF depends on $L = 50$ features. By solving (3) for decreasing values of λ, we were able to obtain increasing values of K. The proposed method was contrasted with principal component analysis (PCA), a standard dimensionality reduction technique. PCA can design a rank-K matrix $\mathbf{W}$ to minimize the feature fidelity error $\sum_{s=1}^S \|\mathbf{x}_s - \mathbf{W}\mathbf{x}_s\|_2^2$. Nevertheless, PCA returns a generally dense matrix $\mathbf{W}$, so that reconstructed features $\mathbf{W}\mathbf{x}$ depend on all entries of $\mathbf{x}$. Although PCA cannot effect data selection, it is used here to serve as a clairvoyant benchmark.

Figure 2 shows the fidelity in terms of OPF features (top) and OPF solutions (bottom) attained by the proposed method and PCA for increasing K. As expected, the reconstructed error decreases for increasing K. Using PCA and half of the OPF features ($K = 25$) attains less than 20% reconstruction error in OPF features. Our proposed method does not achieve any drastic improvement in terms of feature reconstruction error. This is reasonable as our method has been designed to reconstruct OPF solutions, not OPF features. Indeed, as verified in the bottom panel of Figure 2, the reconstruction error in terms of OPF solutions decreases dramatically thanks to the proposed method. With only $K = 14$ OPF features, the proposed method achieves reconstruction error in OPF solutions of less than 15%.

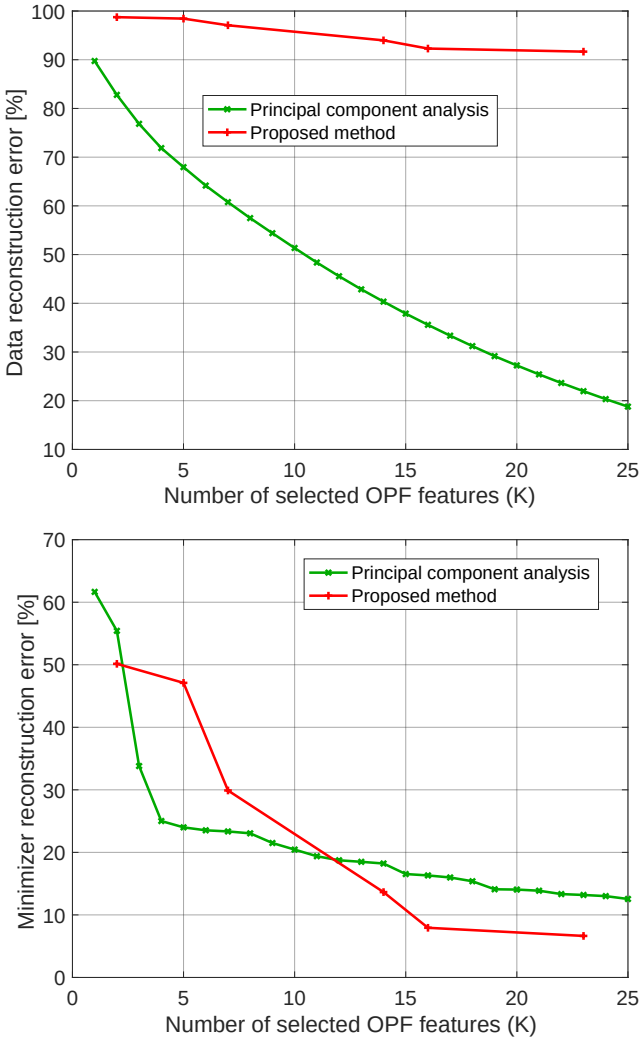

Fig. 2. Normalized reconstruction errors attained by the proposed method and PCA for increasing number of selected OPF features. *Top:* Reconstruction error in terms of OPF features. *Bottom:* Reconstruction error in terms of OPF solutions.

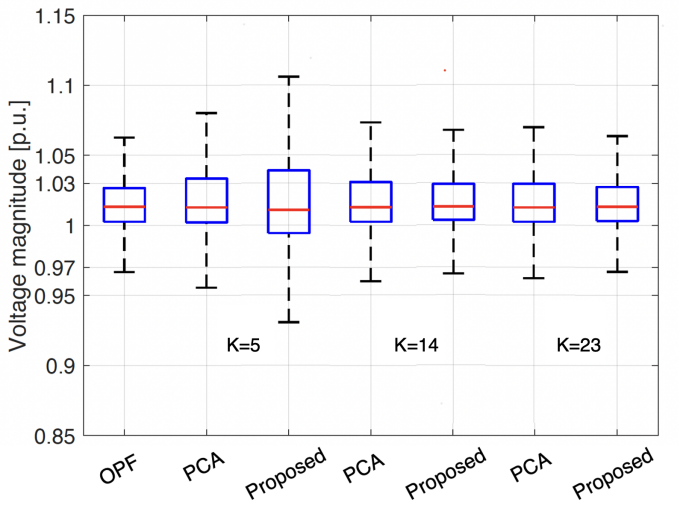

Fig. 3. Box plots of voltage magnitudes across all buses and scenarios for the OPF fed by complete data, and the OPF fed by K features decided by the proposed method.

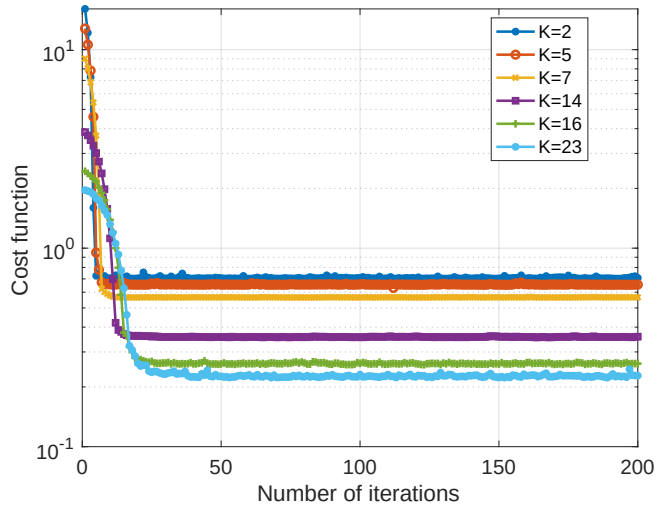

Fig. 4. Cost convergence across PGD iterations for different values of λ, and in turn, K.

How does the error in reconstructing OPF solutions translate to safe grid operation? Recall that although DER reactive power setpoints $\mathbf{y}(\mathbf{W}\mathbf{x}_s)$ were designed to regulate voltages within $\pm 3\%$ pu under loading conditions $\mathbf{W}\mathbf{x}_s$, the actual loading conditions are $\mathbf{x}_s$. What are the grid bus voltages when setpoints $\mathbf{y}(\mathbf{W}\mathbf{x}_s)$ are applied along with the actual loading conditions $\mathbf{x}_s$? Figure 3 shows the box plots of voltage magnitudes across buses and scenarios attained by the original OPF, as well as the proposed method and PCA for $K \in \{5, 14, 23\}$ OPF features. Voltages violate the $\pm 3\%$ limits occasionally even when using the original OPF, because they are enforced as soft constraints. For $K = 5$ features, the proposed method yields a slightly expanded box plot, whereas PCA results in unsafe grid operation. The box plots improve for $K = 14$ features. For $K = 23$ features, the proposed method yields box plots that are almost identical to those of the original OPF. In plain words, although the OPF is now provided with less than half of the original OPF features, it finds DER setpoints that are as good as the original OPF in terms of voltage regulation.

Finally, Figure 4 shows the convergence of PGD iterations in terms of the reconstruction error of OPF solutions. For larger K, the proposed method converges to lower reconstruction errors, as expected. Interestingly, the method seems to be converging to a stationary point within 30 iterations or less. The computational time varies depending on K, and ranges from 30 min to 4 hours on a laptop computer. This variation is reasonable because the time complexity of the proposed method depends on the number of critical regions, which in turn, depends on the value of K. Larger values of K tend to yield an increased number of critical regions, which prolongs the computational time of each PGD iteration.

VI. Conclusions

To deal with privacy, cyber-security, and low observability challenges in power distribution grids, we have proposed

a novel method to select OPF features to facilitate DER scheduling from partial data. The joint task of data selection and reconstruction has been formulated as a non-convex problem over the induced OPF mapping. To avoid mixed-integer formulations, we have developed a first-order optimization method, which is guaranteed to converge to a stationary point of the nonconvex problem. Although the method is applicable to a wide range of OPF formulations, when the OPF takes the form of a parametric LP/QP, results from multiparametric programming can expedite the calculation of the related gradients.

The proposed methodology sets the foundations for exploring several exciting research directions, including the following: *i)* The OPF feature reconstruction scheme can be extended from linear to nonlinear using a neural network with column-sparse weights on its first (input) layer; *ii)* The ML module performing data selection and reconstruction has been designed here using a *supervised* approach requiring OPF labels $\{\mathbf{y}(\mathbf{x}_s)\}_{s=1}^{S}$. Alternatively, the ML module can be designed using an unsupervised approach similar to the one in [12]; and *iii)* To account for approximation errors and uncertainties, one may consider robust or stochastic formulations for the data selection and reconstruction process, or the OPF problem itself.

REFERENCES

[1] S. Bhela, V. Kekatos, and S. Veeramachaneni, "Power distribution system observability with smart meter data," in *Proc. IEEE Global Conf. on Signal and Inf. Process.*, Montreal, Canada, Nov. 2017.

[2] M. K. Singh, V. Kekatos, and G. B. Giannakis, "Learning to solve the AC-OPF using sensitivity-informed deep neural networks," *IEEE Trans. Power Syst.*, vol. 37, no. 4, pp. 2833–2846, Jul. 2022.

[3] M. Jalali, M. K. Singh, V. Kekatos, G. B. Giannakis, and C. C. Liu, "Fast inverter control by learning the OPF mapping using sensitivity-informed Gaussian processes," *IEEE Trans. Smart Grid*, vol. 14, no. 3, pp. 2432–2445, May 2022.

[4] A. J. Conejo, E. Castillo, R. Minguez, and R. Garcia-Bertrand, *Decomposition techniques in mathematical programming*. Springer, 2006.

[5] L. Roald, D. Pozo, A. Papavasiliou, D. Molzahn, J. Kazempour, and A. Conejo, "Power systems optimization under uncertainty: A review of methods and applications," *Electric Power Systems Research*, vol. 214, 2023.

[6] R. Mieth, J. M. Morales, and H. V. Poor, "Data valuation from data-driven optimization," *IEEE Trans. Control of Network Systems*, pp. 1–12, 2024.

[7] V. Dvorkin, F. Fioretto, P. Van Hentenryck, P. Pinson, and J. Kazempour, "Differentially private optimal power flow for distribution grids," *IEEE Trans. Power Syst.*, vol. 36, no. 3, pp. 2186–2196, May 2021.

[8] S. Chevalier and M. R. Almassalkhi, "Towards optimal Kron-based reduction of networks (Opti-KRON) for the electric power grid," in *Proc. IEEE Conf. on Decision and Control*, 2022, pp. 5713–5718.

[9] I. P. Nikolakakos, H. H. Zeineldin, M. S. El-Moursi, and J. L. Kirtley, "Reduced-order model for inter-inverter oscillations in islanded droop-controlled microgrids," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4953–4963, 2018.

[10] W. Jang, S. Mohapatra, T. J. Overbye, and H. Zhu, "Line limit preserving power system equivalent," in *2013 IEEE Power and Energy Conference at Illinois (PECI)*, 2013, pp. 206–212.

[11] J. B. Ward, "Equivalent circuits for power-flow studies," *Transactions of the American Institute of Electrical Engineers*, vol. 68, no. 1, pp. 373–382, 1949.

[12] S. Gupta, V. Kekatos, and M. Jin, "Controlling smart inverters using proxies: A chance-constrained DNN-based approach," *IEEE Trans. Smart Grid*, vol. 13, no. 2, pp. 1310–1321, Mar. 2022.

[13] J. Kotary, V. D. Vito, J. Christopher, P. V. Hentenryck, and F. Fioretto, "Predict-then-optimize by proxy: Learning joint models of prediction and optimization," 2023. [Online]. Available: https://arxiv.org/abs/2311.13087

[14] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2007.

[15] F. Dorfler, M. R. Jovanovic, M. Chertkov, and F. Bullo, "Sparsity-promoting optimal wide-area control of power networks," *IEEE Trans. Power Syst.*, vol. 29, no. 5, pp. 2281–2291, Sep. 2014.

[16] H. Li and Z. Lin, "Accelerated proximal gradient methods for nonconvex programming," in *Intl. Conf. on Neural Information Processing Systems*, vol. 1. Montreal, Canada: MIT Press, 2015, pp. 379—-387.

[17] S. Taheri, M. Jalali, V. Kekatos, and L. Tong, "Fast probabilistic hosting capacity analysis for active distribution systems," *IEEE Trans. Smart Grid*, vol. 12, no. 3, pp. 2000–2012, May 2021.

[18] M. K. Singh, S. Gupta, V. Kekatos, G. Cavraro, and A. Bernstein, "Learning to optimize power distribution grids using sensitivity-informed deep neural networks," in *Proc. IEEE Intl. Conf. on Smart Grid Commun.*, Tempe, AZ, Nov. 2020, pp. 1–6.

[19] M. K. Singh, V. Kekatos, and G. B. Giannakis, "Learning to solve the AC-OPF using sensitivity-informed deep neural networks," *IEEE Trans. Power Syst.*, vol. 37, no. 4, pp. 2833–2846, Jul. 2022.

[20] M. Muratori, "Impact of uncoordinated plug-in electric vehicle charging on residential power demand - supplementary data," Golden, CO, 2017.