Comparison of bounds for optimal PMU placement for state estimation in distribution grids[*]

M. Picallo, A. Anta, and B. De Schutter

Comparison of Bounds for Optimal PMU Placement for State Estimation in Distribution Grids

Miguel Picallo, Adolfo Anta and Bart De Schutter, *Fellow, IEEE*

Abstract—The lack of measurements in distribution grids poses a severe challenge for their monitoring: since there may not be enough sensors to achieve numerical observability, load forecasts (pseudo-measurements) are typically used, and thus an accurate state estimation is not guaranteed. However, an estimation is required to control distribution grids given the increasing amount of distributed generation. Therefore, we consider the problem of optimal sensor placement to improve the state estimation accuracy in large-scale, 3-phase coupled, unbalanced distribution grids. This is a combinatorial optimization problem whose optimal solution is unpractical to obtain for large networks. We explore the properties of different metrics in the context of optimal experimental design, like convexity and modularity, to propose and compare several tight lower and upper bounds on the performance of the optimal solution. Moreover, we show how to use these bounds to choose near-optimal solutions. We test the method on two IEEE benchmark test feeders, the 123-bus and the 8500-node feeders, to show the effectiveness of the approach.

Index Terms—optimal sensor placement, phasor measurement units, distribution grid state estimation, submodular maximization, projected gradient descent, optimal design of experiments

I. INTRODUCTION

The operation of a power network requires accurate monitoring of its state: bus voltages, line currents, consumption and generation, to efficiently manage its controllable elements. State Estimation (SE) serves that purpose by estimating a minimum representation of the network state, for example the bus voltage phasors. Taking several measurements and the admittance matrix of the network as parameters, SE typically solves a weighted least-squares problem using an iterative approach like Newton-Raphson [1], [2]. In transmission networks, SE is fundamental, since the volatile and distributed generation, which injects power in different locations of the network, causes fast-changing bidirectional power flows. In contrast, SE has not been so necessary in distribution grids until recently, since these grids used to have a simple radial structure with a single source bus injecting power. But this situation is changing due to the increasing penetration of distributed generation like PV panels, batteries, etc. [3].

The introduction of Phasor Measurement Units (PMUs) improves the monitoring of electrical networks. Some work in the literature is focused on using PMUs to achieve topological observability [4], which can be solved using integer linear programming [5]. Although cheap PMUs [6] are becoming available for mass deployment, their operational and network communication costs [7] may still prevent installing the required minimum number of sensors to achieve topological observability based only on PMUs, especially in distribution grids. If topological observability is not possible, neither is numerical observability ensured, which is required to solve the SE [4]. As a result, conventional Supervisory Control And Data Acquisition (SCADA) measurements need to be combined with PMUs to solve the SE problem [8]. Therefore, authors of recent work [9], [10] formulate the problem of optimal PMU placement in terms of maximizing the SE accuracy. This accuracy is typically measured through some metric in the context of optimal design of experiments [11], applied to the covariance matrix of the SE error caused by noisy measurements and unreliable forecasts. As a result, the optimal PMU placement problem is a combinatorial optimization problem with a nonlinear objective function, the SE accuracy metric, and thus cannot be solved using linear programming. Moreover, as the size of the network increases, the optimal solution becomes unpractical to obtain given the number of possible combinations of measurements [9].

In distribution grids, one of the major limitations for SE is the lack of sufficient real-time measurements to achieve observability, even using SCADA, and thus SE algorithms need to rely on pseudo-measurements, such as load forecasts. These pseudo-measurements have typically a large relative noise associated (approximately 50% [12]), which causes a high uncertainty in the SE estimates [13]. As a result, there is a growing interest in using PMUs in distribution grids [14], and some recent work proposes algorithms to place sensors to satisfy a desired performance, like algorithms using greedy and random combinations of sensors [15], [16] or evolutionary algorithms [17], [18]. However, these approaches have no optimality guarantees. Since optimal solutions are unpractical in large networks, [19] uses the results in [9] to derive lower bounds for the values of the optimal solution.

Our contribution consists in proposing and comparing several bounds for the optimal solution of the PMU placement problem in distribution grids, and extending these bounds to large-scale grids. With these bounds, we will be able to check the gap between any given suboptimal solution and the optimal solution. First, we prove properties like convexity and supermodularity for some metrics in [11]; then, we use these properties to derive a combination of bounds based on convex optimization as in [10], and bounds based on supermodular minimization similar to [9]. These bounds allow us to obtain a close bound for the performance of the optimal solution of the problem under a cardinality constraint, as well as under a budget constraint and sensors with different costs. Additionally,

M. Picallo and B. De Schutter are with the Delft Center for Systems and Control, Delft University of Technology, The Netherlands {m.picallocruz,b.deschutter}@tudelft.nl

A. Anta is with the Austrian Institute of Technology Adolfo.Anta@ait.ac.at

we show how these bounds scale to large networks by using a projected gradient descent algorithm. Moreover, we show through two examples of test feeders, one of them large-scale, how bounds based on the supermodularity property perform better than those based on convexity if the number of sensors to deploy is small, as in distribution networks.

The rest of the paper is structured as follows. Section II contains the nomenclature for the most relevant symbols. Section III-A presents some background about power networks. Section III-B discusses the different types of measurements. Section III-C summarizes the newly proposed methodology for SE [13]. Section IV presents the metrics considered for the problem and their properties. Section V states the optimal sensor placement problem and derives the lower and higher bounds for each optimum value. Section VI shows the effectiveness of the bounds on a test case. Finally, Section VII presents the conclusions. For clarity purposes, proofs can be found in the Appendices A and B. Appendices C and D describe the methods to extend the approach to large grids.

II. NOMENCLATURE

variables:

V, I, S	Vectors of bus voltages, currents and apparent power
$V_{\text{src}}, I_{\text{src}}, S_{\text{src}}$	Vectors of bus voltages, currents and apparent power at the source bus
Y	Admittance matrix
N	Number of nodes
N_{meas}	Number of measurements
σ_{psd}	Variance of the pseudo-measurements noise
$\sigma_{\text{mag}}, \sigma_{\text{ang}}$	Variance of the magnitude and angle noise of the real-time measurements
$z_{\text{meas}}, \Sigma_{\text{meas}}$	Vector and covariance of measurements
$C_{\text{meas}}, \tilde{C}_{\text{meas}}$	Matrices mapping state to measurements
$V_{\text{prior}}, V_{\text{post}}$	Prior and posterior voltage estimation
F	Subspace of the feasible voltage solutions
$\Sigma_{\text{prior}}, \Sigma_{F,\text{prior}}$	Covariance of prior estimation
$\Sigma_{\text{post}}, \Sigma_{F,\text{post}}$	Covariance of posterior estimation
$f_{\text{A}}(x), f_{\text{D}}(x)$	Covariance metrics on the vector space
$\tilde{f}_{\text{A}}(X), \tilde{f}_{\text{D}}(X)$	Covariance metrics on the set space
$x_{\text{opt}}, f_{\text{opt}}$	Optimal solution and value of the original PMU allocation problems (13), (20)
$x_{\text{convex}}, f_{\text{convex}}$	Optimal solution and value of the problems with convex relaxations (14), (21)
$x_{\text{feas}}, f_{\text{feas}}$	Feasible solution and value of (15), (22)
$x_{\text{greedy}}, f_{\text{greedy}}$	Greedy solution and value of (16) and Algorithm 1

symbols:

$\lvert \cdot \rvert$	Magnitude of a complex number
$(\bar{.}), (\cdot)^*$	Complex conjugate and conjugate transpose
$\text{diag}(\cdot)$	Diagonal operator: vector to diagonal matrix
$(\cdot)_i$	Element i of a vector
$(\cdot)_{i,j}$	Element in row i and column j of a matrix
$(\cdot)_\varepsilon$	Elements of a vector at indices in ε
$(\cdot)_{j,\bullet}, (\cdot)_{\varepsilon,\bullet}$	Row j or rows with indices ε of matrix
$\text{tr}(\cdot), \det(\cdot)$	Trace and determinant of a matrix
∇f	Gradient of f
$\{\cdot\}$	Set of elements
$(\cdot)^{(K)}$	Value at iteration K
$\Pi_{\mathcal{X}}(\cdot)$	Projection onto the set $\mathcal{X}$

III. STATE ESTIMATION IN DISTRIBUTION GRIDS

A. Distribution Grid Model

A distribution grid consists of buses, where power is injected or consumed, and branches, each connecting two buses. This system can be modeled as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ with nodes $\mathcal{V} = \{1, ..., N_{\text{bus}}\}$ representing the buses, edges $\mathcal{E} = \{(v_i, v_j) \mid v_i, v_j \in \mathcal{V}\}$ representing the branches, and edge weights $\mathcal{W} = \{w_{i,j} \mid (v_i, v_j) \in \mathcal{E}\}$ representing the admittance of the branches, which are determined by the length and type of the line cables. In 3-phase networks buses may have up to 3 phases, so that the voltage at bus i is $V_{\text{bus},i} \in \mathbb{C}^{n_{\phi,i}}$, where $n_{\phi,i} \leq 3$ (and the edge weights $w_{i,j} \in \mathbb{C}^{n_{\phi,i} \times n_{\phi,j}}$). The state of the network is then typically represented by the vector bus voltages $V_{\text{bus}} = [V_{\text{src}}^T, \; V^T]^T \in \mathbb{C}^{N+3}$, where $V_{\text{src}} \in \mathbb{C}^3$ denotes the known voltage of the 3 phases at the source bus, and $V \in \mathbb{C}^N$ the voltages in the non-source buses, where N depends on the number of buses and phases per bus. Then, using the Laplacian matrix $Y \in \mathbb{C}^{(N+3) \times (N+3)}$ of the weighted graph $\mathcal{G}$, called admittance matrix [1], the power flow equations to compute the currents I and the power loads S are:

$$\begin{bmatrix} I_{\text{src}} \\ I \end{bmatrix} = Y \begin{bmatrix} V_{\text{src}} \\ V \end{bmatrix}, \; S = \text{diag}(\bar{I})V \qquad (1)$$

B. Measurements

As explained in [13], several different sources of information can be available to solve the SE problem:

1) *Pseudo-measurements*, i.e., load estimations S_{psd} based on predictions and/or known installed load capacity at every bus. Since these pseudo-measurements are estimations rather than actual measurements, we model their uncertainty using a Gaussian noise with a relative large standard deviation (a typical value can be $\sigma_{\text{psd}} \approx 50\%$ [12]).

2) *Virtual measurements*, i.e., buses with zero-injections, no loads connected. They can be represented as physical constraints for the voltage states by defining the set of indices of zero-injection buses $\varepsilon = \{i, \cdots, j\}$:

$$(S)_\varepsilon = 0, \; (I)_\varepsilon = 0 \qquad (2)$$

3) *Real-time PMU measurements*, i.e., voltage and current GPS-synchronized measurements of magnitude and phase angle. According to the IEEE standard for PMU [20], these measurements may have a small error. Again, we model this uncertainty using a Gaussian noise with a low standard deviation for the magnitude and the angle, $\sigma_{\text{mag}} \approx 1\%$ and $\sigma_{\text{ang}} \approx 0.01$ rad respectively. They can be expressed using a linear approximation [13] with magnitude and angle noise due to the measurements and imperfect synchronization. For a number of N_{meas} measurements $z_{\text{meas}} \in \mathbb{C}^{N_{\text{meas}}}$ we have

$$z_{\text{meas}} \approx C_{\text{meas}}V + \text{diag}(C_{\text{meas}}V)(\omega_{\text{mag}} + j\omega_{\text{ang}}) \qquad (3)$$

where $\omega_{\text{mag}}, \omega_{\text{ang}}$ are the Gaussian noises with mean 0 and standard deviation $\sigma_{\text{mag}}, \sigma_{\text{ang}}$ respectively: $\omega_{\text{mag}} \sim \mathcal{N}(0, \sigma_{\text{mag}}I_{\text{d},N_{\text{meas}}})$, $\omega_{\text{ang}} \sim \mathcal{N}(0, \sigma_{\text{ang}}I_{\text{d},N_{\text{meas}}})$, with $I_{\text{d},n}$ denoting the identity matrix of dimension n, and where the matrix C_{meas} maps state voltages to measurements; it relates

the values of the measurements to values of state voltages. Then, for measurement j at phase l of bus i we have

$$(C_{\text{meas}}V)_j = (C_{\text{meas}})_{j,\bullet}V =$$

$$\begin{cases} V_{i_l} & \text{for a voltage measurement} \\ (Y)_{i_l,\bullet}V & \text{for a current measurement} \\ (Y)_{i_l,m_l}(V_{i_l} - V_{m_l}) & \text{for a branch-current } i \to m \\ & \text{measurement} \end{cases} \quad (4)$$

Since the measurement noises in (3) are small according to the PMUs standard [20], their covariance matrices can be approximated using the measurements:

$$\begin{aligned} \Sigma_{\text{meas}} &= (\sigma_{\text{mag}}^2 + \sigma_{\text{ang}}^2)\text{diag}(|C_{\text{meas}}V|^2) \\ &\approx (\sigma_{\text{mag}}^2 + \sigma_{\text{ang}}^2)\text{diag}(|z_{\text{meas}}|^2) \end{aligned}$$

C. State Estimation

Typically, SE consists in finding the voltages that best match the measurements by solving a weighted least-squares problem [1]. As proposed in [13], SE can be decomposed in two parts: First, using the pseudo-measurement estimations or predictions for the loads S_{psd}, we solve the power flow offline to obtain a prior estimate V_{prior}:

$$V_{\text{prior}} = \text{PowerFlow}(S_{\text{psd}}) \quad (5)$$

Then, using the real-time PMU measurements z_{meas}, a posterior solution V_{post} can be derived using a linear filter:

$$V_{\text{post}} = V_{\text{prior}} + K(z_{\text{meas}} - C_{\text{meas}}V_{\text{prior}}) \quad (6)$$

where the gain matrix K is obtained by minimizing the error covariance $\Sigma_{\text{post}} = \mathbb{E}[(V_{\text{post}} - V)^*(V_{\text{post}} - V)]$:

$$\begin{aligned} \Sigma_{\text{post}} &= \Sigma_{\text{prior}} + K(\Sigma_{\text{meas}} + C_{\text{meas}}\Sigma_{\text{prior}}C_{\text{meas}}^*)K^* \\ &\quad - K C_{\text{meas}}\Sigma_{\text{prior}} - \Sigma_{\text{prior}}C_{\text{meas}}^*K^* \\ K &= \arg\min_K \text{tr}(\Sigma_{\text{post}}) \\ &= \Sigma_{\text{prior}}C_{\text{meas}}^*(C_{\text{meas}}\Sigma_{\text{prior}}C_{\text{meas}}^* + \Sigma_{\text{meas}})^{-1} \end{aligned} \quad (7)$$

where Σ_{prior} and Σ_{meas} are the expected error covariance of the prior estimate V_{prior} and the measurements z_{meas} respectively.

Remark 1. *Extra non-synchronized real-time measurements, like magnitude measurements from SCADA, can be included in the posterior update (6) for a greater improvement of the posterior estimate V_{post} by using the first-order approximation of the measurement function [13]. If a multi-stage PMU deployment is considered [21], prior installed PMUs can be also considered as extra measurements in the same manner.*

Remark 2. *As shown in [8], splitting the problem in two steps yields the same first-order approximation as solving the problem in one step. Moreover, the posterior minimum-variance estimator using the linear update (6), is equal to the maximum-likelihood using a weighted least-squares approach [13]. Therefore, we can conclude that for an SE method that assumes Gaussian noises and performs a maximum likelihood estimation, the posterior covariance will be approximately the one in (7), and thus the method developed here for optimal sensor placement can be also extended for other SE techniques satisfying these conditions.*

Since V_{post} in (6) is an unbiased estimator, SE accuracy can be defined as a function of the posterior covariance matrix Σ_{post} in (7), which needs to be minimized to improve the SE accuracy. This motivates a deeper analysis of Σ_{post}: if there are zero-injection buses at indices ε when solving (5),

then Σ_{prior} will not be of full rank and thus will not be invertible. For convenience, we consider the restricted subspace $\{V \mid (Y)_{\varepsilon,\bullet}V = 0\}$ and the linear transformation proposed in [13] to represent the space of feasible solutions: $V = Fx + V_0$ with $x \in \mathbb{C}^{N-|\varepsilon|}$, where $|\varepsilon|$ is the cardinality of ε and F is the null space of $(Y)_{\varepsilon,\bullet}$: $F = \ker((Y)_{\varepsilon,\bullet}) \in \mathbb{C}^{N \times N-|\varepsilon|}$, so that $F^*F = I_{\text{d}}$, and V_0 denotes the voltage under zero loads. Then we have $\Sigma_{\text{prior}} = F\Sigma_{F,\text{prior}}F^*$, where $\Sigma_{F,\text{prior}}$ is the covariance of x. After some manipulations of (7), the resulting error covariance Σ_{post} ($\Sigma_{F,\text{post}}$ in this subspace) for the posterior estimation V_{post} can be expressed as

$$\begin{aligned} \Sigma_{\text{post}} &= F(\Sigma_{F,\text{prior}}^{-1} + (C_{\text{meas}}F)^*\Sigma_{\text{meas}}^{-1}C_{\text{meas}}F)^{-1}F^* \\ \Sigma_{F,\text{post}} &= (\Sigma_{F,\text{prior}}^{-1} + (C_{\text{meas}}F)^*\Sigma_{\text{meas}}^{-1}C_{\text{meas}}F)^{-1} \end{aligned} \quad (8)$$

In the presence of zero-injection buses, $|\varepsilon| > 0$, and then $\Sigma_{\text{post}} \in \mathbb{C}^{N \times N}$ is not full rank: $\text{rank}(\Sigma_{\text{post}}) = N - |\varepsilon|$. It has the same rank as $\Sigma_{F,\text{post}} \in \mathbb{C}^{(N-|\varepsilon|) \times (N-|\varepsilon|)}$, which is full rank since $\Sigma_{F,\text{prior}}$ is full rank. Moreover, the eigenvalues of $\Sigma_{F,\text{post}}$ are all eigenvalues of Σ_{post}, because for any eigenvector v of $\Sigma_{F,\text{post}}$, Fv is an eigenvector of Σ_{post} with the same eigenvalue, since $F^*F = I_{\text{d}}$. Consequently, we can analyze $\Sigma_{F,\text{post}}$ instead of Σ_{post}. These eigenvalues represent the lengths of the axes of the confidence ellipsoid [22]. Moreover, since measurement errors are caused separately by each sensor, they are independent, i.e. Σ_{meas} is diagonal, and we can split $\Sigma_{F,\text{post}}$ by every measurement:

$$\begin{aligned} \Sigma_{F,\text{post}} &= (\Sigma_{F,\text{prior}}^{-1} + \sum_i (C_{\text{meas}}F)_{i,\bullet}^*(C_{\text{meas}}F)_{i,\bullet}(\Sigma_{\text{meas}}^{-1})_{i,i})^{-1} \\ &= (\Sigma_{F,\text{prior}}^{-1} + \sum_i x_i(\tilde{C}_{\text{meas}}F)_{i,\bullet}^*(\tilde{C}_{\text{meas}}F)_{i,\bullet}(\Sigma_{\text{meas}}^{-1})_{i,i})^{-1} \\ &= \Sigma_{F,\text{post}}(x) \end{aligned}$$

$$(9)$$

where $x_i \in \{0,1\}$, $x_i = 1$ if the physical quantity i has a sensor measuring its value, 0 otherwise; and $\tilde{C}_{\text{meas}}$ is the special case of $C_{\text{meas}}F$ with all possible measurements of all types (bus voltage, bus current, and line current) for all nodes and lines in each phase.

Definition 1. *In order to improve the accuracy of the SE, the problem of optimal sensor placement consists in minimizing $\Sigma_{post}(x)$, equivalently $\Sigma_{F,post}(x)$ in (9), according to a metric $m(\cdot)$ and under a set of constraints $h(\cdot)$ to limit the number of sensors or the total cost:*

$$\min_x m(\Sigma_{F,post}(x)) \text{ s.t. } h(x) \leq 0, \ x_i \in \{0,1\} \ \forall i \quad (10)$$

For simplicity, we define: $f(x) \equiv m(\Sigma_{F,post}(x))$.

Remark 3. *Problem (10) is equivalent to minimizing the number of sensors or the total cost of the sensors, while enforcing a given accuracy, i.e., satisfying a performance threshold λ of the metric: $\min_{\{x|f(x)\leq\lambda, \ x_i\in\{0,1\} \ \forall i\}} h(x)$, as in [15]–[18]. To achieve that, the problem (10) can be solved for different numbers of sensors or budgets until the performance value is below the threshold: $\min_{\{x|h(x)\leq0, \ x_i\in\{0,1\} \ \forall i\}} f(x) \leq \lambda$. The representation in (10), as used in [9], [10], [19] is preferred due to the availability of a computationally efficient optimization algorithm to solve the problem, see Appendix C.*

IV. METRICS FOR SENSOR PLACEMENT

There are many possible metrics $f(x)$ available in the context of optimal design of experiments [11]. Concretely, we will focus on the A-*optimal* and the D-*optimal* metrics,

because of the properties that we will show later. Nonetheless, the results of this paper could be generalized to other metrics with the same properties.

Definition 2. *The A-optimal and the D-optimal metrics are defined as follows:*

$$\text{A-}optimal : f_{\text{A}}(x) := \text{tr}(\Sigma_{\text{post}}(x)) = \text{tr}(\Sigma_{F,\text{post}}(x))$$
$$\text{D-}optimal : f_{\text{D}}(x) := \log(\det(\Sigma_{F,\text{post}}(x))) \tag{11}$$

The A-*optimal* metric represents the sum of the eigenvalues of $\Sigma_{F,\text{post}}$ and thus the sum of the lengths of axes of the confidence ellipsoid [22]. This is the metric typically used for SE methods, since standard SE maximizes the log-likelihood through a weighted least-squares minimization [1], which is equivalent to the minimum-variance estimator using the trace [13]. The D-*optimal* metric is the natural logarithm of the product of the eigenvalues of $\Sigma_{F,\text{post}}$ and it is related to the logarithm of the volume of the confidence ellipsoid [22]. These metrics have several properties relevant to the problem of optimal sensor placement:

1) *Convexity*: When relaxing x_i in (10) to be continuous, $x_i \in [0, 1]$, the metrics are convex on x [22, Section 7.5] [10], and thus a global optimum can be computed efficiently. Note that despite $\log(x)$ being concave for $x \in \mathbb{R}$, $\log(\det(X))$ is convex for $X \in \mathbb{R}^{n \times n}$ [22].

2) *Gradient computation*: The gradients ∇f for $f \in \{f_{\text{A}}, f_{\text{D}}\}$ can be derived analytically using matrix calculus [23]:

$$(\nabla f_{\text{A}}(x))_i = -\text{tr}\left(\Sigma_{F,\text{post}}^2(x)(\tilde{C}_{\text{meas}})_{i,\bullet}^*(\tilde{C}_{\text{meas}})_{i,\bullet}\right)(\Sigma_{\text{meas}}^{-1})_{i,i}$$
$$(\nabla f_{\text{D}}(x))_i = -\text{tr}\left(\Sigma_{F,\text{post}}(x)(\tilde{C}_{\text{meas}})_{i,\bullet}^*(\tilde{C}_{\text{meas}})_{i,\bullet}\right)(\Sigma_{\text{meas}}^{-1})_{i,i} \tag{12}$$

These expressions will be necessary when developing gradient methods to optimize the placement of sensors in large networks, see Appendix C.

3) *Monotonicity*: Moreover, these metrics can be seen as set functions by defining the sets $X = \{i \mid x_i = 1\}$. Note that there is a bijection between both, i.e., $i \in X \iff x_i = 1$. Therefore, we will use the notation X and x to denote the set and the vector respectively, and $\hat{f}(X) = f(x)$ to denote the metric functions applied to them. Defining

$$\Lambda_X := \sum_{i \in X}(\tilde{C}_{\text{meas}})_{i,\bullet}^*(\tilde{C}_{\text{meas}})_{i,\bullet}(\Sigma_{\text{meas}}^{-1})_{i,i} \succeq 0$$

the posterior error covariance can be expressed as $\Sigma_{F,\text{post}}(X) = (\Sigma_{F,\text{prior}}^{-1} + \Lambda_X)^{-1}$.

Proposition 1. $\hat{f}_{\text{A}}(X), \hat{f}_{\text{D}}(X)$ *are monotone decreasing.*

Proof. Consider the sets of sensors $X \subseteq Y$. Since $\Lambda_Y - \Lambda_X = \Lambda_{Y \setminus X} \succeq 0$, we have $\Lambda_Y \succeq \Lambda_X$. Then $\Sigma_{F,\text{prior}}^{-1} + \Lambda_Y \succeq \Sigma_{F,\text{prior}}^{-1} + \Lambda_X \succeq 0$ and thus $(\Sigma_{F,\text{prior}}^{-1} + \Lambda_X)^{-1} \succeq (\Sigma_{F,\text{prior}}^{-1} + \Lambda_Y)^{-1}$. Consequently, for any of the functions $\hat{f} \in \{\hat{f}_{\text{A}}, \hat{f}_{\text{D}}\}$, $\hat{f}(X) \geq \hat{f}(Y)$. ∎

4) *Modularity*: Submodularity/supermodularity are a sort of concavity/convexity properties when considering set functions. The increment of a set function when adding a new element diminishes (submodular) or increases (supermodular) as the set gets larger.

Definition 3. *Given a finite set Ω, a submodular/supermodular function [24] is a set function $\hat{f} : 2^\Omega \to \mathbb{R}$, with $\hat{f}(\emptyset) = 0$, where $\emptyset$ is the empty set, so that*

for any element $a \in \Omega$ and two subsets X, Y so that $X \subseteq Y \subseteq \Omega \setminus \{a\}$ we have

submodular: $\hat{f}(Y \cup \{a\}) - \hat{f}(Y) \leq \hat{f}(X \cup \{a\}) - \hat{f}(X)$

supermodular: $\hat{f}(Y \cup \{a\}) - \hat{f}(Y) \geq \hat{f}(X \cup \{a\}) - \hat{f}(X)$

Proposition 2. $\hat{f}_{\text{D}}(X) - \hat{f}_{\text{D}}(\emptyset)$ *is supermodular.*

Proof. See Appendix A ∎

This means that for f_{D}, there exist bounds between greedy solutions and optimal solutions, as we will see later in (19) and (24). These bounds will allow us to provide a limit on the values of the optimal solution, and thus to check how far the values of other suboptimal solutions can be from the value of the optimal solution.

Remark 4. *In [9] was proven that if the rows in $\tilde{C}_{meas}$ are orthogonal, $\hat{f}_{\text{A}}(X) - \hat{f}_{\text{A}}(\emptyset)$ is supermodular, and lower bounds based on the supermodularity property can also be used for the A-optimal metric. However, the rows of $\tilde{C}_{meas}$ are in general not orthogonal due to the different types of sensors in (4), the electrical connections represented in Y, and the effect of the reduced subspace F.*

V. OPTIMAL SENSOR PLACEMENT

So far we have looked at the cost function in (10), now we will focus on the constraints. In this section we state the optimal sensor placement problem under two different constraints: a cardinality constraint limiting the number of sensors, which is the most typical approach in the literature [9], [10], [12], and a budget constraint limiting the total cost of the deployed sensors, which may have different costs. For each case, we derive the respective lower and upper bounds on the optimal performance, using the metrics properties derived in Section IV.

A. Solutions with cardinality constraint

For any of the metrics $f \in \{f_{\text{A}}, f_{\text{D}}\}$, the problem of optimal placement of N_{meas} sensors is

$$x_{\text{opt}} = \arg\min_x f(x) \text{ s.t. } \sum_i x_i \leq N_{\text{meas}}, \; x_i \in \{0, 1\} \; \forall i \tag{13}$$

where N_{meas} indicated the maximum number of sensors. We denote the value of the optimal solution of (13) $f_{\text{opt}} = f(x_{\text{opt}})$.

Similar to [25], [26] for general estimation problems and in [10] for transmission power networks, we relax the constraints in (13), $x_i \in [0, 1]$, to get a continuous convex problem using

$$x_{\text{convex}} = \arg\min_x f(x) \text{ s.t. } \sum_i x_i \leq N_{\text{meas}}, \; x_i \in [0, 1] \; \forall i \tag{14}$$

We denote the value of the optimal solution of (14) as $f(x_{\text{convex}}) = f_{\text{convex}}$. However, this solution will not necessarily be feasible to (13). To create a feasible solution x_{feas} (with respective value $f(x_{\text{feas}}) = f_{\text{feas}}$), we can take the largest N_{meas} values of x_{convex}, set them to 1 and the others to 0:

$$x_{\text{feas},i} = \begin{cases} 1 \text{ if } x_{\text{convex},i} \geq x_{\text{convex},k} \\ 0 \text{ otherwise} \end{cases} \tag{15}$$

where k is such that $\left|\{i \mid x_{\text{convex},i} \geq x_{\text{convex},k}\}\right| = N_{\text{meas}}$. Ties are broken arbitrarily if there is more than one $x_{\text{convex},i}$ with value $x_{\text{convex},k}$, i.e. if $\left|\{i \mid x_{\text{convex},i} = x_{\text{convex},k}\}\right| > 1$. This also

applies for further possible ties throughout the paper. Another simple way to create a feasible solution would be using a forward greedy sensor selection: at iteration K, given the set of selected sensors $X^{(K-1)}$, add a new sensor k such that

$$k = \arg \min_{i \notin X^{(K-1)}} \hat{f}(X^{(K-1)} \cup \{i\}), \ X^{(K)} = X^{(K-1)} \cup \{k\} \quad (16)$$

We denote the solution of (16) x_{greedy} and its value $f(x_{\text{greedy}}) = f_{\text{greedy}}$. Then, since x_{opt} is a feasible suboptimal solution of (14), and x_{feas} and x_{greedy} are feasible suboptimal solutions of (13), the following holds for all metrics:

$$f_{\text{convex}} \leq f_{\text{opt}} \leq \min(f_{\text{greedy}}, f_{\text{feas}}) \quad (17)$$

For the case of $\hat{f}_{\text{D}}$, since $\hat{f}_{\text{D}}(X) - \hat{f}_{\text{D}}(\emptyset)$ is monotone non-increasing and supermodular ($-\hat{f}_{\text{D}}(X) + \hat{f}_{\text{D}}(\emptyset)$ is monotone nondecreasing submodular), we have an extra lower bound [24] for f_{opt} of (13):

$$\hat{f}_{\text{D,greedy}} - \hat{f}_{\text{D}}(\emptyset) \leq (\hat{f}_{\text{D,opt}} - \hat{f}_{\text{D}}(\emptyset))\alpha$$
$$\alpha = \left(1 - \left(1 - \frac{1}{N_{\text{meas}}}\right)^{N_{\text{meas}}}\right) \in (1 - e^{-1}, 1] \quad (18)$$

so that

$$\tilde{f}_{\text{D,greedy}} := (\hat{f}_{\text{D,greedy}} - \hat{f}_{\text{D}}(\emptyset))\alpha^{-1} + \hat{f}_{\text{D}}(\emptyset) \leq f_{\text{D,opt}} \quad (19)$$

B. Solutions with budget constraint

Although the problem of sensor placement under a cardinality constraint is simpler and has more desired properties [24], a more realistic and economic representation of the problem would be having a budget constraint on the sensors

$$x_{\text{opt}} = \arg \min_x f(x) \ \text{s.t.} \ \sum_i c_i x_i \leq b, \ x_i \in \{0, 1\} \ \forall i \quad (20)$$

where b represents the budget and c_i the cost of installing a sensor at location i. This can take into account the extra cost of installing a sensor in a remote area, or locations where specific rights might be required, or different types of sensors, etc. The relaxed convex problem is then

$$x_{\text{convex}} = \arg \min_x f(x) \ s.t. \ \sum_i c_i x_i \leq b, \ x_i \in [0, 1] \ \forall i \quad (21)$$

Similar to (15), a feasible solution x_{feas} w.r.t. (20) can be built using the convex solution x_{convex} of (21), by iteratively taking the sensor with highest $x_{\text{convex},i}$: at iteration K, and until the set of possible sensors is empty, $\mathcal{B} = \{i \mid c_i \leq b - \sum_{j \in X^{(K-1)}} c_j\} = \emptyset$, add to the set of selected sensors $X^{(K-1)}$ a new sensor $k \notin X^{(K-1)}$ such that:

$$k = \arg \max_{i \in \mathcal{B}, i \notin X^{(K-1)}} x_{\text{convex},i}, \ X^{(K)} = X^{(K-1)} \cup \{k\} \quad (22)$$

Likewise, we can adapt the forward greedy selection algorithm (16) to take costs into account. Therefore, we can use the cost-effective forward greedy selection algorithm of [27], which in every iteration adds the sensor with the lowest ratio of objective function improvement divided by sensor cost. In this version we also return intermediate values required for the bounds. For clarity we detail the whole algorithm here:

Using Algorithm 1, we denote the solutions as $X_{\text{greedy1}} = X_1$, $X_{\text{greedy2}} = X_2$ and $X_{\text{greedy1a}} = X_1 \cup \{a\}$ with corresponding values $\hat{f}(X_{\text{greedy1}}) = f_{\text{greedy1}}, \hat{f}(X_{\text{greedy2}}) = f_{\text{greedy2}}, \hat{f}(X_{\text{greedy1a}}) = f_{\text{greedy1a}}$. Since $X_{\text{greedy1}} \subseteq X_{\text{greedy2}}$, $f_{\text{greedy2}} \leq f_{\text{greedy1}}$. Note that in the cardinality constrained

---

Algorithm 1 Cost-effective forward greedy selection

Require: $b > 0, c \in \mathbb{R}^n, c \geq 0$
1: $X_1 \leftarrow \emptyset$
2: $k \leftarrow \arg \min_i \frac{\hat{f}(\{i\})}{c_i}$
3: **while** $c_k \leq b - \sum_{i \in X_1} c_i$ **do**
4: $\quad X_1 \leftarrow X_1 \cup \{k\}$
5: $\quad k \leftarrow \arg \min_i \frac{\hat{f}(X_1 \cup \{i\})}{c_i}$
6: **end while**
7: $a \leftarrow k$
8: $X_2 \leftarrow X_1$
9: **while** $\{i | c_i \leq b - \sum_{j \in X_2} c_j\} \neq \emptyset$ **do**
10: $\quad k \leftarrow \arg \min_{\{i | c_i \leq b - \sum_{j \in X_2} c_j\}} \frac{\hat{f}(X_2 \cup \{i\})}{c_i}$
11: $\quad X_2 \leftarrow X_2 \cup \{k\}$
12: **end while**
13: **return** X_1, X_2, a

---

case we would have $X_1 = X_2$, and hence $f_{\text{greedy1}} = f_{\text{greedy2}}$; therefore we use f_{greedy} in that case. Then, as in (17) we know:

$$f_{\text{convex}} \leq f_{\text{opt}} \leq \min(f_{\text{greedy2}}, f_{\text{feas}}) \quad (23)$$

Again, for the case of $\hat{f}_{\text{D}}$, since $\hat{f}_{\text{D}}(X) - \hat{f}_{\text{D}}(\emptyset)$ is nonincreasing supermodular, we have two extra lower bounds for the value f_{opt} of (20), which are derived from the proofs in [28], [29]:

$$\tilde{f}_{\text{D,greedy1}} := (f_{\text{D,greedy1}} - \hat{f}_{\text{D}}(\emptyset))\beta^{-1} + \hat{f}_{\text{D}}(\emptyset) \leq f_{\text{D,opt}}$$
$$\tilde{f}_{\text{D,greedy1a}} := (f_{\text{D,greedy1a}} - \hat{f}_{\text{D}}(\emptyset))\beta_a^{-1} + \hat{f}_{\text{D}}(\emptyset) \leq f_{\text{D,opt}} \quad (24)$$

with

$$\beta = \left(1 - \prod_{i \in X_{\text{greedy1}}} \left(1 - \frac{c_i}{b}\right)\right) \in (0, 1]$$
$$\beta_a = \left(1 - \prod_{i \in X_{\text{greedy1a}}} \left(1 - \frac{c_i}{b}\right)\right) \in (1 - e^{-1}, 1] \quad (25)$$

Remark 5. *If the costs c_i are unitary and the budget b equals the number of sensors, i.e $c_i = 1 \ \forall i$ and $b = N_{\text{meas}}$, then $\alpha = \beta$. Consequently, (19) is a particular case of (24).*

A drawback of the bound with β in (24) is that β does not have a lower bound bigger than 0, and thus it may happen that $\beta^{-1} \to \infty$ and the bound becomes trivial. Nonetheless, we can prove the following result:

Proposition 3. *Let $\gamma \in (0, 1]$ define the percentage of budget used, so that $\sum_{i \in X_{\text{greedy1}}} c_i = \gamma b \leq b$, then we have*

$$\beta \in (1 - e^{-\gamma}, 1], \beta_a \in (1 - \gamma e^{-\gamma}, 1] \quad (26)$$

Proof. See Appendix B ∎

Proposition 3 shows that if $\gamma \to 0$, then $\beta \to 0$, but $\beta_a \to 1$; so only the bound with β_a is useful. If $\gamma \to 1$, $\beta \to \beta_a$, and the bound with β may be better.

Given the supermodularity of f_{D}, see Proposition 2, we can also consider the bound proposed in [27], called online bound, generated by Algorithm 2 to obtain

$$\max_{\mathcal{A}} \tilde{f}_{\text{D,online}}(\mathcal{A}) \leq f_{\text{D,opt}} \quad (27)$$

Finally we have all bounds:

$$\max(\max_{\mathcal{A}} \tilde{f}_{\text{D,online}}(\mathcal{A}), \tilde{f}_{\text{D,greedy1}}, \tilde{f}_{\text{D,greedy1a}}, f_{\text{D,convex}})$$
$$\leq f_{\text{D,opt}} \leq \min(f_{\text{D,greedy2}}, f_{\text{feas}}) \quad (28)$$

Algorithm 2 Online bound

Require: $b > 0, c \in \mathbb{R}^n, c \geq 0, \mathcal{A}$ (any set)
1: $X \leftarrow \emptyset$
2: $\tilde{f}_{\mathrm{D,online}} = \hat{f}_\mathrm{D}(\mathcal{A})$
3: **while** $\sum_{i \in X} c_i < b$ **do**
4: $\quad k \leftarrow \arg\min_{\{i \notin \mathcal{A}\}} \frac{\hat{f}_\mathrm{D}(\mathcal{A} \cup \{i\}) - \hat{f}_\mathrm{D}(\mathcal{A})}{c_i}$
5: $\quad X \leftarrow X \cup \{k\}$
6: $\quad \tilde{f}_{\mathrm{D,online}} \leftarrow \tilde{f}_{\mathrm{D,online}} + \hat{f}_\mathrm{D}(\mathcal{A} \cup \{k\}) - \hat{f}_\mathrm{D}(\mathcal{A})$
7: **end while**
8: $\tilde{f}_{\mathrm{D,online}} \leftarrow \tilde{f}_{\mathrm{D,online}} + (\hat{f}_\mathrm{D}(\mathcal{A} \cup \{k\}) - \hat{f}_\mathrm{D}(\mathcal{A})) \frac{b - \sum_{i \in X} c_i}{c_k}$
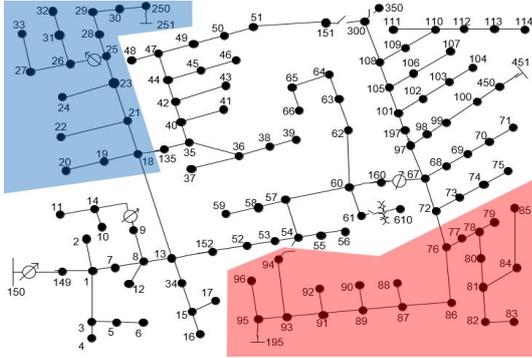

Fig. 1: 123-bus test feeder from [30] with different cost zones.

Remark 6. *Note that since $\mathcal{A}$ may be any set of sensors, $\max_\mathcal{A} \tilde{f}_{\mathrm{D,online}}(\mathcal{A})$ is computationally demanding.*

VI. TEST CASE

Now we test the bounds on the 123-bus [30] and the 8500-node [31] test feeders. For the budget-constrained case in the 123-bus feeder, we consider two more expensive zones to represent heterogeneous costs: $\tilde{c}_i = 2$ in the top left blue zone, $\tilde{c}_i = 1.5$ in the bottom right red zone, $\tilde{c}_i = 1$ elsewhere, see Fig. 1. Then, sensor costs are normalized $c_i = \frac{\tilde{c}_i}{\sum_i \tilde{c}_i}$, so that their average is 1, and thus the budget is approximately the number of sensors deployed. In the 8500-node feeder, we assign random normal distributed costs: $c_i \sim \mathcal{N}(1, 0.1)$. For any number of measurements or budget, the sensor locations can be recovered from each solution x.

The algorithms are coded in Python and run on an Intel Core i7-6700HQ CPU at 2.60GHz with 16GB of RAM. For the convex optimization problems (14) and (21), we use the function *minimize* of the *scipy.optimize* package for the 123-bus test case, and a projected gradient descent algorithm for the 8500-node, see Appendix C, where we use an efficient projection algorithm as in [32] in order to handle the computational complexity of the 8500-node feeder. Additionally, since function evaluations may take a few seconds for the 8500-node feeder, we use matrix algebra results to speed up the computation of the greedy solutions in (16) and Algorithm 1, see Appendix D. Otherwise, it could take several days to evaluate all possible nodes in every step.

Fig. 2 shows the bounds for the 123-bus and the 8500-node test feeders. Both the A,D-optimal metrics are analyzed, under a cardinality and a budget constraint, for different numbers of sensors and budgets respectively. For simplicity, in the online bound $\tilde{f}_{\mathrm{D,online}}(\mathcal{A})$ in (27), we have only considered

$\mathcal{A} = \emptyset$ instead of any $\mathcal{A}$. However, we have observed that other options, like using a greedy solution for $\mathcal{A}$, produce similar bounds. The online bound has also been applied to the cardinality constrained problem. The yellow shaded area with squares shows the area between the minimum upper bound and the maximum lower bound, and thus the possible locations of the optimal solution $f_{\{\mathrm{A,D}\},\mathrm{opt}}$.

For the D-optimal metric, in Fig. 2c, 2d, 2g and 2h, it can be observed that the online bound $\tilde{f}_{\mathrm{D,online}}$ (27) outperforms the rest of the lower bounds ($\tilde{f}_{\mathrm{D,greedy}}$ (19), $\tilde{f}_{\mathrm{D,greedy1}}$ (24), $f_{\mathrm{D,convex}}$ (14), (21)) for a small number of sensors under a cardinality and under a budget constraint, especially in the 8500-node feeder. For clarity, $\tilde{f}_{\mathrm{D,greedy1a}}$ (24) has not been plotted, since the results were almost equal to $\tilde{f}_{\mathrm{D,greedy1}}$. For a small number of sensors, it is remarkable how the lower bound based on convex relaxations $f_{\mathrm{D,convex}}$ performs poorly compared to the other lower bounds, as it produces too optimistic results. This is a result of the high SE uncertainty in all nodes of the grid before the PMU placement; as a consequence, the convex optimization allocates partial unfeasible PMUs with $0 < x_i \ll 1$ in many nodes of the grid, instead of full PMUs with $x_i = 1$ in a few nodes. This is equivalent to adding more sensors with an error with larger relative standard deviation. This cannot be done, since PMUs have a fixed maximum relative error by design [20], see section III-B.

For the A-optimal metric, in Fig. 2a, 2b, 2e and 2f we do not have the lower bounds based on the supermodularity property; however, the lower bound based on convex relaxations $f_{\mathrm{A,convex}}$ (14), (21) performs better than for the D-optimal metric. It still can be observed that for a small number of sensors this convex bound is far from the greedy and feasible solutions ($f_{\mathrm{A,feas}}$ (15), $f_{\mathrm{A,greedy}}$ (16), $f_{\mathrm{A,greedy2}}$ in Algorithm (1)), but not as much as in the D-optimal metric case, and the bound approaches the values of the solutions quicker.

As expected, for both metrics simple greedy or feasible solutions ($f_{\{\mathrm{A,D}\},\mathrm{greedy}}$, $f_{\{\mathrm{A,D}\},\mathrm{greedy2}}$, $f_{\{\mathrm{A,D}\},\mathrm{feas}}$) perform better than random configurations ($f_{\{\mathrm{A,D}\},\mathrm{rand}}$), even better than the best random configuration out of the 100 samples displayed.

VII. CONCLUSIONS

We have stated the problem of optimal sensor placement for deploying PMUs to minimize the uncertainty of state estimation in distribution grids, both under a cardinality and a budget constraint. We have analyzed the properties of different metrics, concretely, convexity and supermodularity. Using these properties, we have derived a set of bounds that have enabled us to narrow the gap of the possible value of the optimal solution of this intractable problem. Since the optimal solution is unpractical to obtain in large grids, suboptimal solutions are required instead. Using these bounds, we can compute the maximum gap between any suboptimal solution and the optimal solutions. Moreover, we have observed, that the bounds produced by the supermodularity property are especially relevant when only a small number of sensor can be installed in large networks. In this case, a bound based on convex relaxation produces a too optimistic result.

Future work could include extending these results to take into account network reconfiguration due to different states of switches. Moreover, more exhaustive search algorithms could
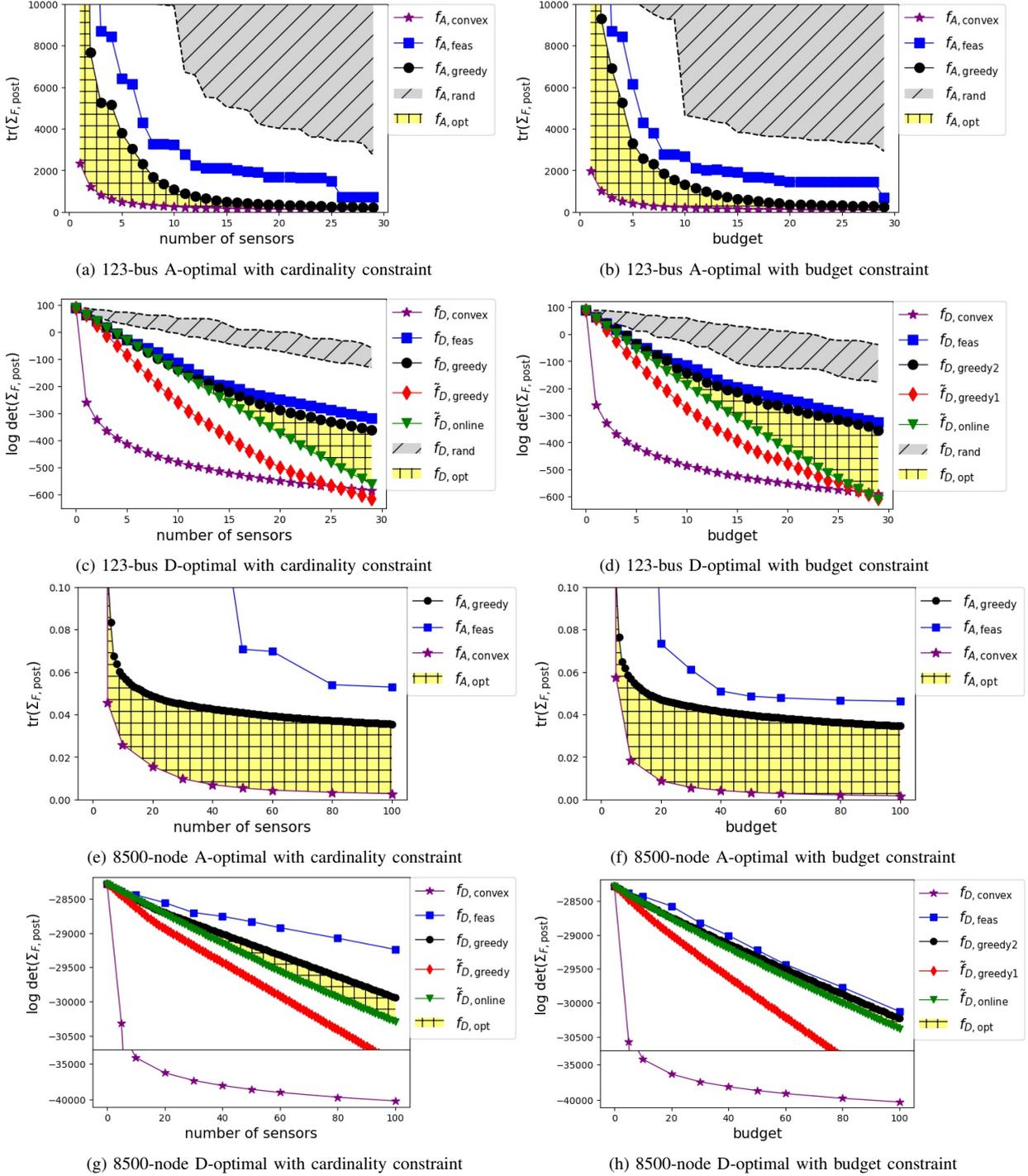
Fig. 2: Plots for the A,D-optimal metrics, under cardinality and budget constraints, the 123-bus and 8500-node cases, showing the lower bounds and the upper bounds. The grey shaded area filled with oblique lines shows the values $f_{\{A,D\},\text{rand}}$ for random configurations of sensors (100 samples). The yellow shaded area filled with a grid of horizontal and vertical lines shows the possible locations of the optimal solution $f_{\{A,D\},\text{opt}}$.

be developed to obtain solutions closer to the lower bounds. Additionally, it would be interesting to analyze how the choice of the metric affects the effectiveness of the bounds.

APPENDIX A
PROOF OF PROPOSITION 2

Proof. The constant term $\hat{f}_\text{D}(\emptyset)$ is only necessary to ensure that the function is 0 when $X = \emptyset$. For the rest of the proof this term is not necessary since for any two sets X, Y we have:
$$(\hat{f}_\text{D}(Y) - \hat{f}_\text{D}(\emptyset)) - (\hat{f}_\text{D}(X) - \hat{f}_\text{D}(\emptyset)) = \hat{f}_\text{D}(Y) - \hat{f}_\text{D}(X).$$

This proof is very similar to the one used for controllability gramians in [33]. Consider the set of sensors $X \subseteq Y$ and a sensor a so that $\{a\} \notin Y$. Now let us define

$$g_a(X) := \hat{f}_D(X \cup \{a\}) - \hat{f}_D(X)$$

$$\tilde{f}_D(\gamma, X, Y) := \log\det\left((\Sigma_{F,\text{prior}}^{-1} + \Lambda_X + \gamma\Lambda_{Y\setminus X})^{-1}\right)$$

$$\tilde{g}_a(\gamma, X, Y) := \tilde{f}_D(\gamma, X \cup \{a\}, Y \cup \{a\}) - \tilde{f}_D(\gamma, X, Y)$$

So that we have $\tilde{f}_D(0, X, Y) = \hat{f}_D(X)$, $\tilde{f}_D(1, X, Y) = \hat{f}_D(Y)$, $\tilde{g}_a(0, X, Y) = g_a(X)$ and $\tilde{g}_a(1, X, Y) = g_a(Y)$. Now we will prove that $g_a(Y) \geq g_a(X)$ by computing the gradient of $\tilde{g}_a$:

$$\frac{\partial \tilde{f}_D}{\partial \gamma}(\gamma, X, Y) = -\text{tr}\left((\Sigma_{F,\text{prior}}^{-1} + \Lambda_X + \gamma\Lambda_{Y\setminus X})^{-1}\Lambda_{Y\setminus X}\right)$$

$$\frac{\partial \tilde{g}_a}{\partial \gamma}(\gamma, X, Y) = \frac{\partial \tilde{f}_D}{\partial \gamma}(\gamma, X \cup \{a\}, Y \cup \{a\}) - \frac{\partial \tilde{f}_D}{\partial \gamma}(\gamma, X, Y)$$

$$= -\text{tr}\Big(\big((\Sigma_{F,\text{prior}}^{-1} + \Lambda_{X\cup\{a\}} + \gamma\Lambda_{Y\setminus X})^{-1}$$
$$- (\Sigma_{F,\text{prior}}^{-1} + \Lambda_X + \gamma\Lambda_{Y\setminus X})^{-1}\big)\Lambda_{Y\setminus X}\Big)$$

Since $X \subseteq Y$, we have $\Lambda_{Y\setminus X} \succeq 0$. Since $\Lambda_{\{a\}} \succeq 0$ and $\Lambda_{X\cup\{a\}} = \Lambda_{\{a\}} + \Lambda_X$, we have

$$\Sigma_{F,\text{prior}}^{-1} + \Lambda_{X\cup\{a\}} + \gamma\Lambda_{Y\setminus X} \succeq \Sigma_{F,\text{prior}}^{-1} + \Lambda_X + \gamma\Lambda_{Y\setminus X} \succeq 0$$

and thus

$$(\Sigma_{F,\text{prior}}^{-1} + \Lambda_X + \gamma\Lambda_{Y\setminus X})^{-1} \succeq (\Sigma_{F,\text{prior}}^{-1} + \Lambda_{X\cup\{a\}} + \gamma\Lambda_{Y\setminus X})^{-1}$$

Consequently, $\frac{\partial \tilde{g}_a}{\partial \gamma}(\gamma, X, Y) \geq 0$ and

$$g_a(Y) - g_a(X) = \tilde{g}_a(1, X, Y) - \tilde{g}_a(0, X, Y)$$
$$= \int_0^1 \frac{\partial \tilde{g}_a}{\partial \gamma}(\gamma, X, Y)\partial\gamma \geq 0$$

so that

$$\hat{f}_D(Y \cup \{a\}) - \hat{f}_D(Y) \geq \hat{f}_D(X \cup \{a\}) - \hat{f}_D(X)$$

■

APPENDIX B
PROOF OF PROPOSITION 3

Proof. If $\sum_{i \in X_{\text{greedy1}}} c_i = \gamma b$, then $\prod_{i \in X_{\text{greedy1}}}\left(1 - \frac{c_i}{b}\right)$ achieves its maximum at $c_i = \frac{\gamma b}{|X_{\text{greedy1}}|}$ for all i, where $|\cdot|$ is the number of elements in a set. Then we have:

$$\beta = \left(1 - \prod_{i \in X_{\text{greedy1}}}\left(1 - \frac{c_i}{b}\right)\right)$$
$$\geq \left(1 - \left(1 - \frac{\gamma}{|X_{\text{greedy1}}|}\right)^{|X_{\text{greedy1}}|}\right) \geq (1 - e^{-\gamma})$$

We assume that $c_i \leq b$ for all i (sensors with $c_i > b$ are discarded since they cannot be installed). Then we know that $b > c_a > b - \gamma b$, where c_a is the cost of element a, and we have:

$$\beta_a = \left(1 - \prod_{i \in X_{\text{greedy1a}}}\left(1 - \frac{c_i}{b}\right)\right)$$
$$\geq \left(1 - \left(1 - \frac{\gamma}{|X_{\text{greedy1}}|}\right)^{|X_{\text{greedy1}}|}\left(1 - \frac{c_a}{b}\right)\right)$$
$$\geq \left(1 - e^{-\gamma}\left(1 - \frac{b-\gamma b}{b}\right)\right) = (1 - \gamma e^{-\gamma})$$

■

APPENDIX C
EFFICIENT PROJECTED GRADIENT DESCENT

To solve the convex problem under a cardinality constraint (14) for the 8500-node feeder, we use a projected gradient descent method using the gradient expressions in (12):

$$x^{(K+1)} = \Pi_{\mathcal{X}}(x^{(K)} - \alpha^{(K)}\nabla f(x^{(K)}))$$
$$\mathcal{X} = \{x \mid \sum_i x_i = N_{\text{meas}}, \ x_i \in [0, 1]\}$$

where the suffix $(\cdot)^{(K)}$ denotes the value at iteration K. We use $\alpha^{(K)} = \frac{\alpha}{k\|\nabla f(x^{(K)})\|_2}$ to guarantee convergence of the method [34], where α is a design parameter. For a more efficient

projection $\Pi_{\mathcal{X}}(\cdot)$, we use the scaled boxed-simplex projection algorithm proposed in [32], which converges in a finite number of iterations, as opposed to the projection algorithm proposed in [10].

When solving the problem under a budget constraint (21), we can use the change of variables $y_i = x_i c_i$, and the function $f_c(y) = f\left(\frac{y}{c}\right)$, and solve the alternative problem

$$y_{\text{convex}} = \arg\min_y f_c(y) \ s.t. \ \sum_i y_i \leq b, \ y_i \in [0, c_i] \ \forall i$$

using the projected gradient descent method:

$$y^{(K+1)} = \Pi_{\mathcal{Y}}(y^{(K)} - \alpha^{(K)}\nabla f_c(y^{(K)}))$$
$$\mathcal{Y} = \{y \mid \sum_i y_i = b, \ y_i \in [0, c_i]\}$$

where an efficient implementation of the projection $\Pi_{\mathcal{Y}}(\cdot)$ can be derived by using a modified version of the one in [32] by changing the 1 to c_i for each respective y_i. The details of the algorithm can be found in [35].

APPENDIX D
EFFICIENT GREEDY COMPUTATIONS

The posterior covariance at iteration K can be defined as

$$\Sigma_{F,\text{post}}^{(K)} = \left(\Sigma_{F,\text{prior}}^{-1} + \sum_{i \in X^{(K)}} (C_{\text{meas}}F)_{i,\bullet}^*(C_{\text{meas}}F)_{i,\bullet}(\Sigma_{\text{meas}}^{-1})_{i,i}\right)^{-1}$$

With this expression, we use Woodbury's matrix identity [23] to obtain a simplified expression that is easier to evaluate:

$$\tilde{f}_A(X^{(K)} \cup \{j\})$$
$$= \text{tr}\left(\left(((\Sigma_{F,\text{post}}^{(K)})^{-1} + (C_{\text{meas}}F)_{j,\bullet}^*(C_{\text{meas}}F)_{j,\bullet}(\Sigma_{\text{meas}}^{-1})_{j,j})^{-1}\right)\right)$$
$$= \text{tr}\left(\Sigma_{F,\text{post}}^{(K)} - \frac{\Sigma_{F,\text{post}}^{(K)}(C_{\text{meas}}F)_{j,\bullet}^*(C_{\text{meas}}F)_{j,\bullet}\Sigma_{F,\text{post}}^{(K)}}{(\Sigma_{\text{meas}}^{-1})_{j,j}^{-1} + (C_{\text{meas}}F)_{j,\bullet}\Sigma_{F,\text{post}}^{(K)}(C_{\text{meas}}F)_{j,\bullet}^*}\right)$$
$$= \text{tr}(\Sigma_{F,\text{post}}^{(K)}) - \frac{(C_{\text{meas}}F)_{j,\bullet}(\Sigma_{F,\text{post}}^{(K)})^2(C_{\text{meas}}F)_{j,\bullet}^*}{(\Sigma_{\text{meas}}^{-1})_{j,j}^{-1} + (C_{\text{meas}}F)_{j,\bullet}\Sigma_{F,\text{post}}^{(K)}(C_{\text{meas}}F)_{j,\bullet}^*}$$

Then, we use Sylvester's determinant identity [23] to get

$$\tilde{f}_D(X^{(K)} \cup \{j\})$$
$$= \log\det\left(((\Sigma_{F,\text{post}}^{(K)})^{-1} + (C_{\text{meas}}F)_{j,\bullet}^*(C_{\text{meas}}F)_{j,\bullet}(\Sigma_{\text{meas}}^{-1})_{j,j})^{-1}\right)$$
$$= \log\det(\Sigma_{F,\text{post}}^{(K)})$$
$$- \log\det\left(I_d + \Sigma_{F,\text{post}}^{(K)}(C_{\text{meas}}F)_{j,\bullet}^*(C_{\text{meas}}F)_{j,\bullet}(\Sigma_{\text{meas}}^{-1})_{j,j}\right)$$
$$= \log\det(\Sigma_{F,\text{post}}^{(K)})$$
$$- \log\left(1 + (\Sigma_{\text{meas}}^{-1})_{j,j}(C_{\text{meas}}F)_{j,\bullet}\Sigma_{F,\text{post}}^{(K)}(C_{\text{meas}}F)_{j,\bullet}^*\right)$$

REFERENCES

[1] A. Abur and A. G. Exposito, *Power System State Estimation: Theory and Implementation.* CRC Press, 2004.

[2] A. Monticelli, "Electric power system state estimation," *Proceedings of the IEEE*, vol. 88, no. 2, pp. 262–282, 2000.

[3] A. Ipakchi and F. Albuyeh, "Grid of the future," *IEEE Power and Energy Magazine*, vol. 7, no. 2, pp. 52–62, 2009.

[4] T. Baldwin, L. Mili, M. Boisen, and R. Adapa, "Power system observability with minimal phasor measurement placement," *IEEE Transactions on Power Systems*, vol. 8, no. 2, pp. 707–715, 1993.

[5] B. Gou, "Generalized integer linear programming formulation for optimal PMU placement," *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 1099–1104, 2008.

[6] B. Pinte, M. Quinlan, and K. Reinhard, "Low voltage micro-phasor measurement unit (μPMU)," in *IEEE Power and Energy Conference at Illinois (PECI)*, Feb 2015, pp. 1–4.

[7] M. B. Mohammadi, R. Hooshmand, and F. H. Fesharaki, "A new approach for optimal placement of PMUs and their required communication infrastructure in order to minimize the cost of the WAMS," *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 84–93, Jan 2016.

[8] M. Zhou, V. A. Centeno, J. S. Thorp, and A. G. Phadke, "An alternative for including phasor measurements in state estimators," *IEEE Transactions on Power Systems*, vol. 21, no. 4, pp. 1930–1937, 2006.

[9] Q. Li, R. Negi, and M. D. Ilić, "Phasor measurement units placement for power system state estimation: A greedy approach," in *IEEE Power and Energy Society General Meeting*, July 2011, pp. 1–8.

[10] V. Kekatos, G. B. Giannakis, and B. Wollenberg, "Optimal placement of phasor measurement units via convex relaxation," *IEEE Transactions on Power Systems*, vol. 27, no. 3, pp. 1521–1530, 2012.

[11] F. Pukelsheim, *Optimal Design of Experiments*. SIAM, 2006.

[12] L. Schenato, G. Barchi, D. Macii, R. Arghandeh, K. Poolla, and A. V. Meier, "Bayesian linear state estimation using smart meters and pmus measurements in distribution grids," in *IEEE International Conference on Smart Grid Communications*, Nov 2014, pp. 572–577.

[13] M. Picallo, A. Anta, A. Panosyan, and B. De Schutter, "A two-step distribution system state estimator with grid constraints and mixed measurements," in *IEEE Power Systems Computation Conference*, June 2018.

[14] A. von Meier, D. Culler, A. McEachern, and R. Arghandeh, "Micro-synchrophasors for distribution systems," in *IEEE/PES Innovative Smart Grid Technologies Conference (ISGT)*, Feb 2014, pp. 1–5.

[15] R. Singh, B. C. Pal, and R. B. Vinter, "Measurement placement in distribution system state estimation," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 668–675, 2009.

[16] R. Singh, B. C. Pal, R. A. Jabr, and R. B. Vinter, "Meter placement for distribution system state estimation: an ordinal optimization approach," *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2328–2335, 2011.

[17] J. Liu, J. Tang, F. Ponci, A. Monti, C. Muscas, and P. A. Pegoraro, "Trade-offs in pmu deployment for state estimation in active distribution grids," *IEEE Transactions on Smart Grid*, vol. 3, no. 2, pp. 915–924, June 2012.

[18] S. Prasad and D. M. V. Kumar, "Trade-offs in PMU and IED deployment for active distribution state estimation using multi-objective evolutionary algorithm," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 6, pp. 1298–1307, June 2018.

[19] F. Fusco and J. C. Villumsen, "Sensor placement for optimal estimation in power distribution grids," in *IEEE/PES Innovative Smart Grid Technologies Conference (ISGT)*, Feb 2015, pp. 1–5.

[20] K. Martin *et al.*, "Exploring the IEEE standard C37. 118–2005 synchrophasors for power systems," *IEEE Transactions on Power Delivery*, vol. 23, no. 4, pp. 1805–1811, 2008.

[21] F. Aminifar, M. Fotuhi-Firuzabad, M. Shahidehpour, and A. Khodaei, "Probabilistic multistage pmu placement in electric power systems," *IEEE Transactions on Power Delivery*, vol. 26, no. 2, pp. 841–849, April 2011.

[22] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[23] M. Brooks, "The Matrix Reference Manual," http://www.ee.ic.ac.uk/hp/staff/dmb/matrix/intro.html, 2011, [Online].

[24] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions," *Mathematical Programming*, vol. 14, no. 1, pp. 265–294, 1978.

[25] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Transactions on Signal Processing*, vol. 57, no. 2, pp. 451–462, 2009.

[26] S. P. Chepuri and G. Leus, "Sparsity-promoting sensor selection for non-linear measurement models," *IEEE Transactions on Signal Processing*, vol. 63, no. 3, pp. 684–698, 2015.

[27] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, "Cost-effective outbreak detection in networks," in *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2007, pp. 420–429.

[28] A. Krause and C. Guestrin, "A note on the budgeted maximization of submodular functions," Carnegie Mellon University, School of Computer Science, Tech. Rep., 2005.

[29] S. Khuller, A. Moss, and J. S. Naor, "The budgeted maximum coverage problem," *Information Processing Letters*, vol. 70, no. 1, pp. 39–45, 1999.

[30] W. H. Kersting, "Radial distribution test feeders," in *IEEE/PES Winter Meeting. Conference Proceedings*, vol. 2, 2001, pp. 908–912 vol.2.

[31] R. F. Arritt and R. C. Dugan, "The IEEE 8500-node test feeder," in *IEEE PES Transmission and Distribution Conference and Exposition*, April 2010, pp. 1–6.

[32] A. Aghazadeh, M. Golbabaee, A. Lan, and R. Baraniuk, "Insense: Incoherent sensor selection for sparse signals," *Signal Processing*, vol. 150, pp. 57–65, 2018.

[33] T. H. Summers, F. L. Cortesi, and J. Lygeros, "On submodularity and controllability in complex dynamical networks," *IEEE Transactions on Control of Network Systems*, vol. 3, no. 1, pp. 91–101, 2016.

[34] A. Nedic and D. P. Bertsekas, "Incremental subgradient methods for nondifferentiable optimization," *SIAM Journal on Optimization*, vol. 12, no. 1, pp. 109–138, 2001.

[35] M. Picallo, A. Anta, and B. De Schutter, "Efficient convex optimization for optimal pmu placement in large distribution grids," in *IEEE Milano PowerTech*, June 2019, in press.