PMU Placement Optimization for Smart Grid Observability and State Estimation

Y. Shi, H. D. Tuan, A. A. Nasir, T. Q. Duong, and H. V. Poor

Abstract—In this paper, phasor measurement unit (PMU) placement for power grid state estimation under different degrees of observability is studied. Observability degree is the depth of the buses’ reachability by the placed PMUs and thus constitutes an important characteristic for PMU placement. However, the sole observability as addressed in many works still does not guarantee a good estimate for the grid state. Some existing works also considered the PMU placement for minimizing the mean squared error or maximizing the mutual information between the measurement output and grid state. However, they ignore the observability requirements for computational tractability and thus potentially lead to artificial results such as acceptance of the estimate for an unobserved state component as its unconditional mean. In this work, the PMU placement optimization problem is considered by minimizing the mean squared error or maximizing the mutual information between the measurement output and grid state, under grid observability constraints. The provided solution is free from the mentioned fundamental drawbacks in the existing PMU placement designs. The problems are posed as binary nonlinear optimization problems, for which this paper develops efficient algorithms for computational solutions. The performance of the proposed algorithms is analyzed in detail through numerical examples on large scale IEEE power networks.

Index Terms—Phasor measurement unit (PMU), observability, power system state estimation, nonlinear binary programming

I. INTRODUCTION

Phasor measurement unit (PMU) is an advanced digital meter, which is used in smart power grids for real-time monitoring of grid operations [1]. By installing it at a bus, the state-of-the-art PMU can measure not only the phasor of the bus voltage but also the current phasors of incident power branches with high accuracy [2]. These measurements are explored by the modern energy management systems (EMSs) for critical applications such as optimal power flow, contingency analysis, and cyber security, etc. [3]–[5].

As phasor measurement units (PMUs) are costly, there is a vast amount of literature on PMU placement optimization to target the minimal number of PMUs. Under different degrees of observability, the mission is accomplished by binary linear programming (BLP) [6], [7]. Here, the complete observability means that there is no bus left unobserved by the placed PMUs, while depth-of-$n$ unobservability means that there are at most $n$ connecting buses left unobserved by the placed PMUs [8], making as many states as possible observed by restricted number of PMUs. An exhaustive binary search was proposed in [9] to deal with this objective under the complete observability condition and additional operating conditions such as the single branch outage and the presence of zero power injections. A binary particle swarm optimization algorithm was proposed in [10] to deal with it while maintaining the complete observability conditions under the contingencies of PMU loss or branch outage. Binary quadratic programming and BLP were respectively used in [11] and [12] to study the effect of conventional measurements and zero bus injections to the complete observability.

Apparently, observability alone does not necessarily lead to a meaningful state estimate or an informative PMU configuration. In fact, PMU configurations, which use the same number of PMUs to make the grid completely observable, can result in quite different estimation accuracies [13]. Intuitively, a better estimator can be obtained by appropriately employing more PMUs. PMU placement optimization to minimize the mean squared error (of grid state estimation) or to maximize the mutual information between the measurement output and grid state under a fixed allowable number of PMUs was considered in [14] and [15], respectively. Obviously, these placement tasks are mathematically modelled by optimization of nonlinear objective functions of binary variables subject to a simple linear constraint for fixing the number of PMUs. A convex relaxation with the binary constraint $\{0, 1\}$ for binary variables relaxed to the box constraint $[0, 1]$ was proposed in [14], which not only fails to provide even a local optimal solution in general but also is not scalable in the grid dimension as it involves an additional large-size semi-definite matrix variable. A greedy algorithm proposed in [15] does not provide a local optimal solution either. More importantly, both [14] and [15] ignored observability constraints for computational tractability. It was argued in [15] that its proposed mutual information criterion includes the grid complete observability, which is obviously not right simply because as shown later in the paper, the latter differentiates the state estimate from its unconditional mean, which is the trivial estimate, while the former does not.

To fill the gap due to disconnected considerations for the grid state observability and state estimation in the existing approaches, this paper considers PMU placement to optimize the estimation performance under different degrees of observability and with a fixed number of PMUs. These problems...
are posed as binary nonlinear optimization problems, which are computationally much challenging. To the authors’ best knowledge, such optimization problems are still quite open for research.

The rest of the paper is structured as follows. Section II is devoted to the problem statement, which also particularly shows the importance of imposing observability constraints in optimization formulations. Section III develops two scalable algorithms for PMU placement optimization to minimize the mean squared error (of grid state estimation) or maximize the mutual information between the measurement outputs and phasor states under a fixed number of PMUs and different degrees of observability. Section IV presents tailored path-following discrete optimization algorithms for the problems without observability constraint. Simulations are provided in Section V, which demonstrates the efficiency of our algorithms. Section VI concludes the paper. The fundamental inequalities used in Section III are given in the Appendix.

Notation. The notation used in this paper is standard. Particularly, $A > 0$ ($A \geq 0$, resp.) for a Hermitian symmetric matrix $A$ means that it is positive definite (semi-definite, resp.). $\text{Trace}()$ and $|.|$ are the trace and determinant operator. $1_N$ is an $N$-dimensional vector of ones. $I_N$ is the identity matrix of size $N$. $a \leq b$ for two real vectors $a = (a_1, \ldots, a_n)^T$ and $b = (b_1, \ldots, b_n)^T$ is componentwise understood, i.e., $a_i \leq b_i$, $i = 1, \ldots, n$. The cardinality of a set $C$ is denoted by $|C|$. $\mathbb{E}(\cdot)$ denotes expectation, so the mean $\bar{u}$ of a random variable (RV) $U$ is $\bar{u} = \mathbb{E}(u)$. For two random variables $u$ and $v$, their cross-covariance matrix $\mathcal{R}_{uv}$ is $\mathbb{E}((u-u)(v-v)^T)$. Accordingly, the autocovariance $\mathcal{R}_u$ of $u$ is $\mathbb{E}((u-u)(u-u)^T)$. $u \sim \mathcal{N}(\bar{u}, \mathcal{R}_u)$ means $u$ is a Gaussian random variable with mean $\bar{u}$ and autocovariance $\mathcal{R}_u$, which represent the first moment of $u$. The entropy of $u$ is $H(u) = \frac{1}{2} \log_2 |\mathcal{R}_u| = \frac{1}{2} \log \det |\mathcal{R}_u|$. Finally, denote by $u|v$ a RV $u$ conditioned on the RV $v$.

II. PROBLEM STATEMENT

Consider a power grid with a set of buses indexed by $N := \{1, 2, \ldots, N\}$, where buses are connected by a set of transmission lines $L \subseteq N \times N$, i.e. bus $k$ is connected to bus $m$ if and only if $(k, m) \in L$. Accordingly, $N(k)$ is the set of other buses connected to bus $k$. In a DC power model, the power injection at bus $k$ is approximated by

$$P_k = B_{kk}\theta_k + \sum_{m \in N(k)} B_{km}\theta_m, \quad (1)$$

where $P_k$ is the power injection at bus $k$ and $\theta_m$ is the voltage phasor angle at bus $m$, while $B_{km}$ is the imaginary part of the $(k,m)$-entry of the grid’s admittance matrix $Y$. Let $P := (P_1, \ldots, P_N)^T \in \mathbb{R}^N$ be the power injection vector and $\theta := (\theta_1, \ldots, \theta_N)^T \in \mathbb{R}^N$ be the voltage phasor vector. Then [1] can be re-written as $P = B\theta$, where $B \in \mathbb{R}^{N \times N}$ is the so called susceptance matrix with the entries $B(k,k) = B_{kk}$ and $B(k,m) = B_{km}$, if $m \in N(k)$, while $B(k,m) = 0$, otherwise. The susceptance matrix $B$ is invertible under the assumption that the grid is fully connected [16]. Since $P$ can be assumed to be $N(u_p, \Sigma_p)$ [17], it is obvious that $\theta \sim N(B^{-1}u_p, B^{-1}\Sigma_p(B^{-1})^T)$.

On the other hand, the measurement equation of a PMU installed at bus $k$ in the linear DC power flow model [18] is [2], [13], [19].

$$\begin{align*}
\zeta_k &= \theta_k + \vartheta_k, \\
\zeta_{km} &= \theta_k - \theta_m + \vartheta_{km}, \quad k \in N, m \in N(k), \quad (2)
\end{align*}$$

with noises $\vartheta_k \sim \mathcal{N}(0, r_k)$ and $\vartheta_{km} \sim \mathcal{N}(0, \rho_{k})$. The number of incident lines of bus $k$ is the cardinality $|N(k)|$. Accordingly, the measurement vector $z_k := (\zeta_k, \zeta_{k1}, \ldots, \zeta_{k|N(k)|})^T$ is of dimension $M_k = |N(k)| + 1$. For simplicity, (2) is rewritten in regression form as:

$$z_k = H_k \theta + w_k, \quad (3)$$

where $H_k \in \mathbb{R}^{M_k \times N}$ is the associated regression matrix, $w_k := (\vartheta_k, \vartheta_{k1}, \ldots, \vartheta_{k|N(k)|})^T \sim \mathcal{N}(0, R_{w_k})$ with diagonal covariance $R_{w_k}$.

To describe the presence or absence of PMU at bus $k$, we introduce a selection vector $x = (x_1, \ldots, x_N)^T \in \{0, 1\}^N$, where $x_k = 1$ if a PMU is installed at bus $k$, and $x_k = 0$ otherwise. Let us assume that we have $S$ PMUs in total for installation, so

$$\sum_{k \in N} x_k = S. \quad (4)$$

Define

$$D_S := \{x \in \{0, 1\}^N : \sum_{k \in N} x_k = S\} \quad (5)$$

and $X = \text{diag}[x_k I_k]_{k=1,\ldots,N}$, $R_w = \text{diag}[R_{w_k}]_{k \in N}$, where $I_k$ is the identity matrix of size $M_k \times M_k$.

For every $x \in D_S$, let $k_j \in N$, $j = 1, \ldots, S$ for which $x_{k_j} = 1$. Define accordingly, $R_w(x) = \text{diag}[R_{w_k}]_{j=1,\ldots,S}$, and

$$z(x) = \begin{bmatrix} z_{k_1} \\ \vdots \\ z_{k_S} \end{bmatrix}, \quad w(x) = \begin{bmatrix} w_{k_1} \\ \vdots \\ w_{k_S} \end{bmatrix}, \quad H(x) = \begin{bmatrix} H_{k_1} \\ \vdots \\ H_{k_S} \end{bmatrix}.$$ 

The multi-input-multi-output PMU measurement equation is

$$z(x) = \bar{H}(x)\theta + w(x).$$

It is obvious that $\mathcal{R}_{z(x)\theta} = \bar{H}(x)R_{\theta}\bar{H}(x)^T$ while $\mathcal{R}_{z(x)} = \bar{H}(x)R_{\theta}\bar{H}(x)^T + R_{w(x)}$. Let $\theta|z(x)$ be the RV $\theta$ conditioned on the RV $z(x)$. By [20]

$$\theta|z(x) \sim N(\hat{\theta}, \mathcal{R}_{c}(x)), \quad (6)$$

where

$$\hat{\theta} = \bar{\theta} + R_{z(x)}^{-1}(\bar{z}(x) - \bar{\bar{z}}(x))$$

$$= \bar{\theta} + R_{\theta}\bar{H}(x)^T(\bar{H}(x)R_{\theta}\bar{H}(x)^T + R_{w(x)})^{-1} \times (z(x) - \bar{H}(x)\bar{\theta}),$$
which is the minimum mean squared error (MMSE) estimate of \( \theta \) based on PMU output \( z(x) \), and

\[
\mathcal{R}_e(x) = \mathcal{R}_\theta - \mathcal{R}_e^9(z(x)z(x)^\top) \mathcal{R}_e \mathcal{R}_\theta^{-1}
\]

\[
= \mathcal{R}_\theta - \mathcal{R}_\theta \mathcal{H}(x)^\top (\mathcal{H}(x) \mathcal{R}_\theta \mathcal{H}(x)^\top + \mathcal{R}_{w(x)})^{-1} \times \mathcal{H}(x) \mathcal{R}_\theta
\]

\[
= \left( \mathcal{R}_\theta^{-1} + \mathcal{H}(x)^\top \mathcal{R}_{w(x)}^{-1} \mathcal{H}(x)^\top \right)^{-1}
\]

\[
= \left( \mathcal{R}_\theta^{-1} + \sum_{j=1}^S H_j^\top \mathcal{R}_{w_j}^{-1} H_j \right)^{-1}
\]

\[
= \left( B^\top \Sigma_p^{-1} B + \sum_{k \in \mathcal{N}} x_k H_k^\top \mathcal{R}_{w_k}^{-1} H_k \right)^{-1}
\]

The mean squared error (MSE) \( \mathbb{E}([\theta - \hat{\theta}]^2) \) is

\[
f_e(x) := \text{Trace}(\mathcal{R}_e(x)),
\]

which obviously is an analytical function of the PMU selection vector \( x \).

Further, the mutual information (MI) \( I(\theta; z(x)) \) between RVs \( \theta \) and \( z(x) \) is \([21] \) formula (6)

\[
I(\theta; z(x)) = \mathcal{H}(\theta) - \mathcal{H}(\theta | z(x))
\]

\[
= \frac{1}{2 \ln 2} (\ln |\mathcal{R}_\theta| - \ln |\mathcal{R}_e(x)|).
\]

Maximizing the MI \( I(\theta; z(x)) \) is thus equivalent to maximizing \( f_{MI}(x) \) for

\[
f_{MI}(x) := -\ln |\mathcal{R}_e(x)| = \ln |B^\top \Sigma_p^{-1} B + \sum_{k \in \mathcal{N}} x_k H_k^\top \mathcal{R}_{w_k}^{-1} H_k|,
\]

It should be realized that either the MSE \( f_e(x) \) or MI \( f_{MI}(x) \) does not indicate the depth of the placed PMUs in reaching the measurement for the whole phasor state. One needs either the constraint

\[
Ax \geq 1_N,
\]

of the complete observability to assure that the phasor state \( \theta \) is completely observable \([19], [22], [23] \), where \( A \) is the bus-to-bus incidence matrix defined by \( A_{km} = 1 \) if \( k = m \) or bus \( k \) is adjacent to bus \( m \), and \( A_{km} = 0 \) otherwise, or the constraint

\[
B Ax \geq 1_{NB},
\]

of the depth-of-one unobservability to assure that there are no two connecting buses that are unobservable \([8] \). Here and after \( B \) is the branch-to-bus incident matrix and \( NB \) is the total number of branches. The general case of dept-of-n unobservability with an arbitrary \( n \) is treated similarly though its practicability is unknown.

Let us analyse the constraints \([9] \) and \([10] \) from the information-theoretic viewpoint. The constraint \([9] \) guarantees that all state components \( \theta_m \) are observable, i.e. each \( \theta_m \) appears at least once in the measurement equations \([7] \), which implies \( \theta_m | z(x) \neq \theta_m \), making the measurement equations \([3] \) meaningful for estimating \( \theta_m \). When some \( \theta_m \) is not observable, i.e. it does not appear in the measurement equations \([7] \), it follows that \( \theta_m | z(x) = \theta_m \) so the measurement equations in \([7] \) are useless for estimating \( \theta_m \). In this case, the estimate for \( \theta_m \) is its unconditional mean \( \hat{\theta}_m \) with \( \mathbb{E}((\theta_m - \hat{\theta}_m)^2) = \mathcal{R}_\theta(m,m) \) and \( I(\theta_m; z(x)) = \mathcal{H}(\theta) - \mathcal{H}(\theta | z(x)) = 0 \).

In other words, the optimization problem for maximizing \( I(\theta; z(x)) \) does not reveal a nontrivial estimate for \( \theta_m \) that is a contradiction to \([15] \) statement 1), page 448, 2nd column) which states that the mutual information metric includes the complete observability condition \([3] \) as a special case. Of course, the number of PMUs, \( S \), needs to be sufficient enough to make the constraint \([7] \) fulfilled. When \( S \) is not allowed to be sufficient, one may go for more relaxed constraint \([10] \), which forces all neighboring buses of any unobservable bus to be observable and thus essentially makes as many states as possible be observable by the PMUs.

Thus, we can state the problem of PMU placement optimization to minimize the MMSE or to maximize the MI between the measurement output and phasor state under a fixed number of PMUs and observability/depth-of-one unobservability as the following binary nonlinear optimization problem

\[
\min f(x) \quad \text{s.t.} \quad x \in \mathcal{D}_S, (9)/10 \tag{11}
\]

where \( f(x) \in \{ f_e(x), -f_{MI}(x) \} \), which is a convex function.

### III. Scalable Penalty Algorithms for Optimal PMU Selection

It is obvious that the main issue is regarding how to handle the discrete constraint \( x \in \mathcal{D}_S \) in \([11] \). The following result establishes the equivalence of this discrete constraint and a continuous constraint.

**Lemma 1:** For a polytope \( \text{Poly}(\mathcal{D}_S) = \{ x \in [0,1]^N : \sum_{k \in \mathcal{N}} x_k = S \} \), the discrete constraint \( x \in \mathcal{D}_S \) in \([11] \) is equivalent to the continuous constraint

\[
x \in \text{Poly}(\mathcal{D}_S), g(x) \geq S, \tag{12}
\]

for \( g(x) := \sum_{k \in \mathcal{N}} x_k^L \) with \( L > 1 \).

**Proof.** Note that \( x_k^L \leq x_k \quad \forall x_k \in [0,1] \), so \( g(x) \leq \sum_{k \in \mathcal{N}} x_k^L = S \quad \forall x \in \text{Poly}(\mathcal{D}_S) \). Therefore constraint \([12] \) forces \( g(x) = S \), which is possible if and only if \( x_k^L = x_k \), \( k \in \mathcal{N} \), i.e. \( x_k \in \{0,1\}, k \in \mathcal{N} \), implying \( x \in \mathcal{D}_S \). \( \square \)

Since \( g(x) \) is convex in \( x \), the constraint \( g(x) \geq S \) in \([12] \) is a reverse convex constraint \([24] \). As such \( \mathcal{D}_S = \text{Poly}(\mathcal{D}_S) \setminus \{ x : g(x) < S \} \), i.e. \( \mathcal{D}_S \) is difference of two convex sets \( \text{Poly}(\mathcal{D}_S) \) and \( \{ x : g(x) < S \} \). Also as \( L \) decreases, \( g(x) \) tends to approach a linear function \( \sum_{k \in \mathcal{N}} x_k \) and thus, the constraint \( g(x) \geq S \) approaches the linear constraint \( \sum_{k \in \mathcal{N}} x_k \geq S \). However, it does not mean that choosing \( L \) closer to 1 is effective because the function \( g(x) - S \) also approaches zero very quickly, making the constraint \( g(x) \geq S \) highly artificial. In our previous works \([25], [26] \), \( L = 2 \) was chosen. However, as we will see shortly, \( L = 1.5 \) is a much better choice, accelerating the convergence of the iterative computational processes. The following result is a direct consequence of Lemma 1.
Proposition 1: The function \( \hat{g}(x) = 1/g(x) - 1/S \) can be used to measure the degree of satisfaction of the discrete constraint \( x \in D_S \) in the sense that \( \hat{g}(x) \geq 0 \) if and only if \( x \in D_S \).

Following our previous developments in [25] and [26], instead of handling constraint \( \{12\} \), we incorporate the degree of its satisfaction into the objective in \( \{11\} \), leading to the following penalized optimization problem:

\[
\min_{x} F_\mu(x) := f(x) + \mu(1/g(x) - 1/S)
\text{s.t. } x \in \text{Poly}(D_S), [0]/[10], \quad (13)
\]

where \( \mu > 0 \) is a penalty parameter. This penalized optimization problem is exact with a sufficiently large \( \mu \). Note that \( \{13\} \) is a minimization of a nonconvex function over a convex set. We now develop a path-following computational procedure for its solution. For this purpose, we firstly develop an upper bounding approximation for \( \{13\} \), at some feasible point \( x^{(k)} \) (at \( k \)-th iteration). As the function \( g(x) \) is convex, it is true that [24],

\[
g(x) \geq g(x^{(k)}) := g(x^{(k)}) + (\nabla g(x^{(k)}), x - x^{(k)}) = -(L - 1) \sum_{k \in N} \kappa_k \Lambda_k^L L + \sum_{k \in N} \kappa_k \Lambda_k^{L-1} x_k.
\]

Therefore, an upper bounding approximation at \( x^{(k)} \) for \( 1/g(x) \) can be easily obtained as \( 1/g(x) \leq 1/g(x^{(k)}) \) over the trust region \( g^{(k)}(x) > 0 \).

At the \( k \)-th iteration we are supposed to solve the following convex optimization problem to generate the next iterative point \( x^{(k+1)} \):

\[
\min_{x} f(x) + \mu(1/g^{(k)}(x) - 1/S)
\text{s.t. } x \in \text{Poly}(D_S), [0]/[10], \quad (15)
\]

Although function \( f(x) \) is convex, it is not easy to optimize it. For instance, when \( f = f_e \), usually \( \{15\} \) is solved via the following semi-definite optimization problem with the introduction of slack symmetric \( N \times N \) matrix variable \( T \):

\[
\min_{x, T} \text{Trace}(T) + \mu(1/g^{(k)}(x) - 1/S)
\text{s.t. } x \in \text{Poly}(D_S), [0]/[10], \quad \text{Poly}(D_S), \quad (14)
\]

which is not scalable to \( x \). For \( f = -f_{MI} \), [15] is

\[
\max_{x \in [0,1]^n} \ln |R_e^{-1}(x)| - \mu(1/g^{(k)}(x) - 1/S)
\text{s.t. } x \in \text{Poly}(D_S), \quad (14),
\]

with no known convex solver of polynomial complexity.

In the following, we propose a different approach to provide scalable iterations for [11]. Obviously, there is \( \epsilon > 0 \) such that

\[
A_e := B^T \Sigma_p^{-1} B - \epsilon \sum_{k \in N} H_k^{T} R_w^{-1} H_k > 0.
\]

For \( f = f_e \), applying inequality [23] in the Appendix for \( A_0 \to \mathcal{A}_e, x_k \to x_k + \epsilon, \bar{x}_k \to x_k^{(k)} + \epsilon, \) \( (16) \)

\[
yields f_e(x) \geq f_e(x^{(k)}) := a_0^{(k)} + \sum_{k \in N} a_k^{(k)} \frac{x_k^{(k)}}{x_k^{(k)} + \epsilon} \quad \text{for } 0 < a_0^{(k)} := \text{Trace}((R_e(x^{(k)}))^2 A_e)
\]

\[0 < a_k^{(k)} := (x_k^{(k)} + \epsilon)^2 \text{Trace}((R_e(x^{(k)}))^2 H_k^{T} R_w^{-1} H_k), \quad k \in N.\]

Accordingly, initialized by a feasible point \( x^{(0)} \) for \( \{13\} \), at the \( \kappa \)-th iteration for \( \kappa = 0, 1, \ldots \), we solve the following convex optimization problem to generate the next iterative point \( x^{(k+1)} \), instead of \( \{15\} \):

\[
\min_{x} F^{(k)}(x) := f^{(k)}(x) + \mu(1/g^{(k)}(x) - 1/S)
\text{s.t. } x \in \text{Poly}(D_S), [0]/[10], \quad (17)
\]

Note that \( F^{(k)}(x) \leq F^{(0)}(x) \forall x, \) and \( F^{(k)}(x^{(k)}) = F^{(k)}(x^{(k)}) \), and \( F^{(k)}(x^{(k+1)}) < F^{(k)}(x^{(k)}) \) (because \( x^{(k+1)} \) and \( x^{(k)} \) are the optimal solution and a feasible point for \( \{17\} \)). Therefore,

\[
F^{(k)}(x^{(k+1)}) \leq F^{(k)}(x^{(k+1)}) \leq F^{(k)}(x^{(k)}) = F^{(k)}(x^{(k)}),
\]

i.e. \( x^{(k+1)} \) is a better feasible point than \( x^{(k)} \) for \( \{13\} \). For a sufficient large \( \mu > 0 \), \( g(x^{(k)}) \to 0 \) as well, yielding an optimal solution of the binary nonlinear optimization problem \( \{11\} \) for the case \( f = f_e \). Algorithm II provides a pseudo-code for the proposed computational procedure.

**Algorithm 1 Scalable Penalized MMSE Algorithm**

1: **Initialization.** Set \( \kappa = 0 \). Take any feasible point \( x^{(0)} \in (0,1)^n \) for \( \{13\} \). Choose \( \mu \) such that \( f_e(x^{(0)}) \) and \( (1/S - 1/g(x^{(0)})) \) achieve similar magnitude.

2: **Repeat**

3: Solve the convex optimization problem \( \{17\} \) to generate the next feasible point \( x^{(k+1)} \).

4: Set \( \kappa := \kappa + 1 \).

5: Until convergence.

Analogously, based on inequality [24] in the Appendix, for \( A_0, x_k, \) and \( \bar{x}_k \) from [16], at the \( \kappa \)-th iteration we solve the following convex optimization problem to generate the next iterative point \( x^{(k+1)} \), instead of \( \{15\} \), when \( f = -f_{MI} \):

\[
\max_{x} \left[ a_0^{(k)} - \sum_{k \in N} a_k^{(k)} \frac{x_k^{(k)}}{x_k^{(k)} + \epsilon} - \mu \frac{1}{g^{(k)}(x) - 1/S} \right]
\text{s.t. } x \in \text{Poly}(D_S), [0]/[10], \quad (18)
\]

for \( a_0^{(k)} := -\ln |R_e(x^{(k)})| + \text{Trace}(R_e(x^{(k)})(\sum_{k \in N}(\epsilon + x_k^{(k)})H_k^{T} R_w^{-1} H_k)), \) \( a_k^{(k)} := (x_k^{(k)} + \epsilon)^2 \text{Trace}(R_e(x^{(k)})H_k^{T} R_w^{-1} H_k), \quad k \in N.\)

Algorithm II is a pseudo-code for solution of the binary nonlinear optimization problem \( \{11\} \) for the case \( f = -f_{MI} \).
Recall that point \(x_i \in \{0, 1\}^N\) for (13). Choose \(\mu\) such that \(f_{MI}(x_i)\) and \((1/S - 1/g(x_i))\) achieve similar magnitude.

2: **Repeat**
3: Solve the convex optimization problem (18) to generate the next feasible point \(x^{(\kappa+1)}\).
4: Set \(\kappa := \kappa + 1\).
5: Until convergence.

### IV. Tailored Path-Following Discrete Optimization Algorithms

In this section, we address problem (11) without the observability constraint (9)/(10):

\[
\min_{x} f(x) \quad \text{s.t.} \quad x \in D_S. \tag{19}
\]

which was considered in [14], [27] for \(f = f_x\) with the help of semi-definite relaxation (SDR). The reader is referred to [28] for capacity of SDR to address discrete optimization problems such as (19). We now develop a simple but very efficient path-following discrete optimization algorithm that explores a simple structure of the discrete constraint \(x \in D_S\) to address (19).

**Lemma 2:** \(D_S\) is the set of vertices of \(\text{Poly}(D_S)\).

**Proof.** For \(x \in D_S\) define

\[
J(x) = \{k_1 < k_2 < \ldots < k_S | x_{k_j} = 1, \ j = 1, 2, \ldots, S\}. \tag{20}
\]

Suppose \(\bar{x} \in D_S\). It suffices to show that if \(\bar{x} = \mu a + (1-\mu)b\) for \(a, b \in \text{Poly}(D_S)\) and \(0 < \mu < 1\) then \(a = b = \bar{x}\). Indeed, for \(i \in J(\bar{x})\) we have \(\bar{x}_i = 1 = \mu a_i + (1-\mu)b_i\) and since \(a_i \in [0, 1]\) and \(b_i \in [0, 1]\) it follows that \(a_i = b_i = 1\). For \(i \notin J(\bar{x})\) we have \(\bar{x}_i = 0 = \mu a_i + (1-\mu)b_i\) and since \(a_i \in [0, 1]\), and \(b_i \in [0, 1]\) it follows that \(a_i = b_i = 0\). Hence \(a = b = \bar{x}\) as asserted. □

Recall that point \(x\) is a vertex neighbouring the vertex \(\bar{x}\) if and only if there exists a pair \(i, j\) with \(i \in \{S+1, \ldots, N\}\) and \(j \in \{1, \ldots, S\}\) such that \(x_i = 1, x_j = 0\) and \(x_{k} = \bar{x}_{k} = 1\) for all \(k \in N \setminus \{j\}\) and \(x_{k} = \bar{x}_{k} = 0\) for all \(k \in \{N+1, \ldots, M\} \setminus \{i\}\).

\(\bar{x} = \bar{x}D_S\) is a minimizer of \(f\) over \(\text{Poly}(D_S)\) if and only if \(f(\bar{x}) \leq f(v)\) for every \(v \in D_S\) neighbouring \(\bar{x}\).

**Algorithm 3** Path-following discrete optimization algorithm

**Initialization.** Start from \(x^{(0)} \in D_S\). Set \(\kappa = 0\).

**\(\kappa\)-th iteration.** If there is a \(\bar{x} \in D_S\) neighbouring \(x^{(\kappa)}\) such that \(f(\bar{x}) < f(x^{(\kappa)})\) then reset \(\kappa + 1 \rightarrow \kappa\) and \(x^{(\kappa)} \rightarrow \bar{x}\). Otherwise, if \(f(x^{(\kappa)}) \geq f(x^{(\kappa)})\) for all \(x \in D_S\) neighbouring \(x^{(\kappa)}\) then stop: \(x^{(\kappa)}\) is a local optimal solution of (11).

The proposed Algorithm 3 looks like the Dantzig simplex method for linear programming, which is of the 20th century’s top ten algorithms [29] although its polynomial complexity cannot be proved (in contrast to the polynomial complexity of the interior points methods for linear programming)[1]. Based on this powerful algorithm, we propose Algorithm 4 for the following problem of choosing the minimum number of PMUs to satisfy MMSE or MI constraint:

\[
\min_{x} \sum_{k \in N} x_k : x \in \{0, 1\}^N, f(x) \leq \epsilon. \tag{21}
\]

**Algorithm 4** Iterative Procedure

**Initialization.** Start from \(1 < S_0 < N\) and use Algorithm 3 to find the optimal solution \(x^{(0)}\) of (19) for \(S = S_0\).

**\(S\)-th iteration.** Reset \(S \rightarrow S - 1\) if \(f(x_{\text{opt}}) < \epsilon\) and \(S \rightarrow S + 1\) if \(f(x_{\text{opt}}) > \epsilon\).

Set \(\kappa := \kappa + 1\).

Until \(f(x^{(\kappa)}) \leq \epsilon\) but \(f(x^{(\kappa-1)}) > \epsilon\).

### V. Simulation Results

In the simulation, the real power injections \(P\) are normally distributed and independent across different buses [17]. Similarly to the simulation setup in [15], the mean vector of real power injection \(u_p = (u_{p}(1), \ldots, u_{p}(N))^T\) is obtained by properly scaling the power profiles in [30], while the diagonal entries of power injection covariance matrix are assumed to be 10% of the mean values, i.e. \(\Sigma_P\) is a diagonal matrix with diagonal entries \(\Sigma_P(k, k) = 0.1u_{p}(k)\). The deviation of measurement noise for bus voltage and current branch are set as \(r_k = 0.01\) and \(\rho_k = 0.02\), respectively. All algorithms are solved by Matlab on a Core i7-7600U processor. Sedumi [31] interfaced by CVX is used to solve the convex optimization problems (17) and (18). The commonly used benchmark power networks IEEE 30-bus, IEEE 39-bus, IEEE 57-bus and IEEE 118-bus with their structure and susceptance matrix obtained from Matpower [30] are tested.

It is observed in [7] that the minimum number of PMUs for the network complete observability (CO) or depth-of-one unobservability (DoOU) can be found by solving the following binary linear program

\[
\min_{x} \sum_{k \in N} x_k : x \in \{0, 1\}^N. \tag{22}
\]

Table I provides the minimum number of PMUs needed for the network’s CO and DoOU (obtained by solving (22) by CPLEX [32]) given in the third and fourth columns.

| Network   | # Branch | # PMUs for CO | # PMUs for DoOU |
|-----------|----------|---------------|-----------------|
| 30-bus    | 41       | 10            | 4               |
| 39-bus    | 46       | 13            | 7               |
| 57-bus    | 80       | 17            | 11              |
| 118-bus   | 186      | 32            | 18              |

Fig 1 depicts the MMSE obtained by different methods versus the number of placed PMUs. The curve "Algorithm 1" is the theoretical MMSE by solving problem (17) under the

1Conceputally, Dantzig simplex method is very simple: starting from any vertex of a simplex it moves to a better neighbouring vertex until there is no better neighbouring vertex found.
constraint (9) of the complete observability, while the curve "Monte-Carlo" is obtained through Monte-Carlo simulation. The MMSEs by Algorithm 1 and Monte-Carlo simulation are seen consistent with the increase in the number of placed PMUs leading to a better MMSE. The curve "Observable" is the MMSE at feasible points for (17) that is found by CPLEX [32]. Algorithm 1 is seen to achieve much better MMSE. The last curve "Algorithm 3" is the MMSE by solving (19) by Algorithm 3. Obviously, the Algorithm 3 achieves better MMSE due to the absence of constraints (9) and (10). The curves in Fig. 2 provide normalized MI results in a similar format to Fig. 1. The capability and efficiency of Algorithm 2 and Algorithm 3 to obtain informative PMU placements are quite clear.

Table II provides numerical details of Algorithm 1, Algorithm 2 and superiority in addressing large-scale networks.

| IEEE     | Alg. 1 | Alg. 2 | CPU (s) of Alg. 2 | MMSE | MI   |
|----------|--------|--------|--------------------|------|------|
| 30-bus   | 0.1    | 65.78  | 62.94              | 4.01 | 3.17 |
| 39-bus   | 0.1    | 79.73  | 77.25              | 11.58| 7.98 |
| 57-bus   | 1      | 80.47  | 81.14              | 49.09| 46.03|
| 118-bus  | 1      | 216.31 | 193.24             | 1222.11| 2142.08|

Fig. 1: MMSE by different methods

For problem (19), Kekatos et al [14] relaxed the integer constraint \( x \in \{0, 1\}^N \) to the box constraint \( x \in [0, 1]^N \) to formulate a convex program and then round the \( S \) largest values of the solution of this convex program to 1. Obviously, their solution is hardly optimal in any sense. Fig. 3 compares the MMSE values of problem (19) founded by Algorithm 3 and by Kekatos et al [14]. The former clearly outperforms the latter, especially for large scale networks.

Due to space limitation, only IEEE 30-bus and IEEE-39 networks are selected for MMSE results solved by Algorithm 1 under the constraint (10) of depth-of-one unobservability. Fig. 4 provides MMSE performance obtained via Algorithm

![Fig. 1: MMSE by different methods](image1)

![Fig. 2: MI by different methods](image2)

![Fig. 3: MMSE found by Algorithm 3 and by [14](image3)

![Fig. 4: MMSE performance obtained via Algorithm](image4)
(under the constraint (10)) and Algorithm 3 (without any observability constraints), while Fig. 5 provides the number of bus left unobservable (for IEEE 30-bus). As expected, Algorithm 3 achieves a better MMSE but leaves more buses unobservable because it sacrifices buses to achieve the averaged performance.

For IEEE 57-bus network and IEEE 118-bus network, Fig 6 presents the number of iterations needed for the convergence of Algorithm 3 for MMSE and MI, respectively.

Fig. 4: MMSE found by Algorithm 1 under depth-of-one unobservability condition and Algorithm 3 without any observability constraints

Fig. 5: Number of buses left unobserved by Algorithm 1 under depth-of-one unobservability condition and Algorithm 3 without any observability constraints for IEEE 30-bus network

Fig. 6: Number of iterations required for the convergence of Algorithm 3

Fig. 7: Minimum number of PMUs required versus different values of tolerance level ϵ for MMSE

Given different tolerances ϵ, the required minimum number of PMUs can be obtained by Algorithm 3. For the case of f = Fε, the results are presented in Fig 7.

VI. CONCLUSIONS

In this paper, we have considered PMU placement optimization to minimize the mean squared error or maximize the mutual information between the measurement outputs and phasor states under a fixed number of PMUs and different observability conditions. These binary optimization problems are very computationally challenging due to high nonlinearity of the objective functions. Nevertheless, we have developed the scalable algorithms for their computational solution, which result at least in local optimal solutions. We also developed extremely efficient algorithms of very low computational complexity for cases of absent observability. The viability of our proposed algorithms has been confirmed through simulations with benchmark IEEE grids. The algorithmic developments for PMU placement optimization involving other practical constraints such as branch outages are under way.

APPENDIX: FUNDAMENTAL INEQUALITIES

Let $\mathbb{R}_+^N := \{x \in \mathbb{R}^N : x_k \geq 0, k \in \mathcal{N}\}$ and $\text{int}(\mathbb{R}_+^N) := \{x \in \mathbb{R}^N : x_k > 0, k \in \mathcal{N}\}$. For $A_0 > 0$ and $A_k \succeq 0$, $k \in \mathcal{N}$ let $\Phi(x) := (A_0 + \sum_{k \in \mathcal{N}} x_k A_k)^{-1}$, and $\Psi(x) := (A_0 + \sum_{k=1}^N x_k A_k)^{-1}$. Recall the following result [33 Th.1]:

**Theorem 1**: Function $\varphi(x) = \text{Trace}(\Phi(x))$ is concave in the domain $\text{int}(\mathbb{R}_+^N)$, so for all $x \in \text{int}(\mathbb{R}_+^N)$ and $\bar{x} \in \text{int}(\mathbb{R}_+^N)$ one has

$$
\varphi(x) \leq \varphi(\bar{x}) + (\nabla \varphi(\bar{x}), x - \bar{x})
= \text{Trace} (\Phi(\bar{x}) A_0) + \sum_{k \in \mathcal{N}} \frac{x_k}{A_k} \text{Trace} (\Phi^2(\bar{x}) A_k).
$$

Therefore,

$$
\text{Trace}(\Psi(x)) \leq \text{Trace} \left( (\Psi(\bar{x}))^2 A_0 \right)
+ \sum_{k \in \mathcal{N}} \frac{x_k^2}{A_k} \text{Trace} \left( (\Psi(\bar{x}))^2 A_k \right). \quad (23)
$$
Next,

Theorem 2: For $A \succ 0$ function $\ln |A + HX^{-1}H^T|$ is convex in $X \succ 0$.

Proof. Since $(A + HX^{-1}H^T)^{-1} = A^{-1} - A^{-1}(H^T A^{-1} H + X)^{-1} A^{-1}$, by Appendix B, function

$$f(X) := A^{-1} - A^{-1}(H^T A^{-1} H + X)^{-1} A^{-1}$$

is concave, i.e. $f(\alpha X + \beta Y) \succeq \alpha f(X) + \beta f(Y) \quad \forall X \succ 0, Y \succ 0, \alpha \geq 0, \beta \geq 0, \alpha + \beta = 1$. Therefore $\ln |f(\alpha X + \beta Y)| \succeq \ln |\alpha f(X) + \beta f(Y)| \succeq \ln |f(X)| + \beta \ln |f(Y)|$, showing that $\ln |A + HX^{-1}H^T|^{-1} = -\ln |A + HX^{-1}H^T|$ is concave in $X$. □

The following Theorem is a direct consequence of Theorem 2.

Theorem 3: Function $\phi(x) = -\ln |\Phi(x)|$ is convex in the domain $\text{int}(\mathbb{R}^n_+)$, so for all $x \in \text{int}(\mathbb{R}^n_+)$ and $\bar{x} \in \text{int}(\mathbb{R}^n_+)$ one has

$$\phi(x) \geq \phi(\bar{x}) + \langle \nabla \phi(\bar{x}), x - \bar{x} \rangle$$

$$- \sum_{k \in \mathbb{N}} \frac{1}{x_k} \text{Trace} (\Phi(\bar{x})^{-1} A_k)$$

Therefore,

$$-\ln |\Psi(x)| \geq -\ln |\Psi(\bar{x})| + \text{Trace} (\Psi(\bar{x}) (\sum_{k \in \mathbb{N}} x_k A_k))$$

$$- \sum_{k \in \mathbb{N}} \frac{1}{x_k} \text{Trace} (\Psi(\bar{x}) A_k).$$

REFERENCES

[1] J. De La Ree, V. Centeno, J. S. Thorp, and A. G. Phadke, “Synchronized phasor measurement applications in power systems,” IEEE Trans. smart grid, vol. 1, no. 1, pp. 20–27, 2010.

[2] A. G. Phadke and J. S. Thorp, Synchronized Phasor Measurements and Their Applications. New York: Springer, 2008.

[3] J. A. Momoh, R. Adapa, and M. El-Hawary, “A review of selected optimal power flow literature to 1993. i. nonlinear and quadratic programming approaches,” IEEE Trans. Power Systems, vol. 14, no. 1, pp. 96–104, 1999.

[4] J. Zhao, G. Zhang, K. Das, G. N. Korres, N. M. Manousakis, A. K. Sinha, and Z. He, “Power system real-time monitoring by using PMU-based robust state estimation method,” IEEE Trans. Smart Grid, vol. 7, no. 1, pp. 300–309, 2016.

[5] C.-W. Ten, A. Ginter, and R. Bulbul, “Cyber-based contingency analysisis,” IEEE Trans. Power Systems, vol. 31, no. 4, pp. 3040–3050, 2016.

[6] B. Gou, “Optimal placement of PMUs by integer linear programming,” IEEE Trans. Power Systems, vol. 23, no. 3, pp. 1525–1526, 2008.

[7] B. Gou, “Generalized integer linear programming formulation for optimal PMU placement,” IEEE Trans. Power Systems, vol. 23, no. 3, pp. 1099–1104, 2008.

[8] R. F. Nguoi and A. G. Phadke, “Phasor measurement unit placement techniques for complete and incomplete observability,” IEEE Trans. Power Delivery, vol. 20, no. 4, pp. 2381–2388, 2005.

[9] S. Chakrabarti and E. Kyriakides, “Optimal placement of phasor measurement units for power system observability,” IEEE Trans. Power Systems, vol. 23, no. 3, pp. 1433–1440, 2008.

[10] M. Hajian, A. M. Ranjbar, T. Amireza, and B. Mozaffari, “Optimal placement of PMUs to maintain network observability using a modified bpso algorithm,” Int. J. Elect. Power Energy Syst., vol. 33, no. 1, pp. 28–34, 2011.