

PMU Placement Optimization for Efficient State Estimation in Smart Grid

Ye Shi, Hoang Duong Tuan[✉], Trung Q. Duong[✉], *Senior Member, IEEE*,
H. Vincent Poor[✉], *Fellow, IEEE*, and Andrey V. Savkin[✉]

Abstract—This paper investigates phasor measurement unit (PMU) placement for informative state estimation in smart grid by incorporating various constraints for observability. Observability constitutes an important property for PMU placement to characterize the depth of the buses' reachability by the placed PMUs, but addressing it solely by binary linear programming as in many works still does not guarantee a good estimate for the grid state. Some existing works have considered optimization of some estimation indices by ignoring the observability requirements for computational ease and thus potentially lead to trivial 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 subject to observability constraints, which incorporate operating conditions such as presence of zero injection buses, contingency of measurement loss, and limitation of communication channels per PMU. The proposed design is thus free from the fundamental shortcomings in the existing PMU placement designs. The problems are posed as large scale binary nonlinear optimization problems involving thousands of binary variables, for which this paper develops efficient algorithms for computational solutions. Their performance is analyzed in detail through numerical examples on large scale IEEE power networks. The solution method is also shown to be extendable to AC power flow models, which are formulated by nonlinear equations.

Index Terms—Phasor measurement unit (PMU), smart grid, state estimation, state observability, binary nonlinear optimization, exactly penalized method.

Manuscript received 5 June, 2019; revised 9 September, 2019; accepted 3 October, 2019. Date of publication November 6, 2019; date of current version January 31, 2020. This work was supported in part by the Institute for Computational Science and Technology, Ho Chi Minh City, Vietnam, in part by the Australian Research Council's Discovery Projects under Grant DP190102501, and in part by the U.S. National Science Foundation under Grant DMS-1736417 and Grant ECCS-1824710. (Corresponding author: Hoang Duong Tuan.)

Y. Shi and H. D. Tuan are with the School of Electrical and Data Engineering, University of Technology Sydney, Ultimo, NSW 2007, Australia (e-mail: ye.shi-1@uts.edu.au, tuan.hoang@uts.edu.au).

T. Q. Duong is with the School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Belfast BT7 1NN, U.K. (e-mail: trung.q.duong@qub.ac.uk).

H. V. Poor is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: poor@princeton.edu).

A. V. Savkin is with the School of Electrical Engineering and Telecommunications, The University of New South Wales, Sydney, NSW 2052, Australia (e-mail: a.savkin@unsw.edu.au).

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

Digital Object Identifier 10.1109/JSAC.2019.2951969

I. INTRODUCTION

A. Motivation and Literature Survey

A 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, a 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 used by modern energy management systems (EMSs) for critical applications such as optimal power flow, contingency analysis, and cyber security, etc. [3]–[5].

There is a considerable amount of literature on PMU placement optimization. From the information-theoretic perspective, complete observability constitutes an important characteristic as it means that no bus is left unobserved by the placed PMUs [6]. Under complete observability and its generalizations, PMU placement design was addressed by binary linear programming (BLP) in [6]–[9]. An exhaustive binary search was proposed in [10] under the complete observability condition and additional operating conditions such as single branch outage and the presence of zero injection buses (ZIBs). A binary particle swarm optimization algorithm was proposed in [11] to maintain the complete observability conditions under the contingencies of PMU loss or branch outage. Binary quadratic programming and BLP were respectively used in [12] and [13] to study the impact of ZIBs and power flow measurements (PFMs) to the complete observability.

PMU placement to optimize the so called gain matrix in the maximum likelihood estimate of the grid state [14] subject to a fixed allowable number of PMUs was considered in [15], which formulated it as an optimization problem of a convex objective function subject to a simple linear constraint on binary variables. A convex relaxation with the binary constraint $\{0, 1\}$ for binary variables relaxed to the box constraint $[0, 1]$, which is used in [15], not only fails to provide even a locally optimal solution in general but also is not scalable in the grid dimension as it involves an additional large-size semi-definite matrix variable. Furthermore, PMU placement to maximize the mutual information (MI) between the measurement output and grid state was solved very efficiently in [16] using a very low computational complexity greedy algorithm for submodular function optimization [17]. Neither of the computational methods used in [15] and [16] is

not capable of treating observability constraints. It was argued in [16] that its proposed mutual information criterion includes complete observability, which is an incorrect assertion 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.

B. Research Gap and Contribution

Apparently, observability alone does not necessarily lead to an acceptable state estimate or an informative PMU configuration. In fact, PMU configurations that use the same number of PMUs to make the grid completely observable can result in quite different state estimation accuracies [18]. On the other hand, ignoring the state observability requirements in PMU placement optimization as in [15] and [16] can potentially force one to accept the estimate for an unobserved state component as its unconditional mean, which is its trivial estimate. Certainly, the quality of state estimation is very critical for securing system operations and reducing outages [19].

Another challenge in PMU placement optimization is to integrate the impacts of ZIBs and PFMs to deduce the required number of placed PMUs [13], which not only helps to save cost but more importantly, to improve the smart grid cyber-security. Note that cyber-security requirements are the second most significant factor affecting PMU acquisition and installation costs [20, page iii]. Placing more PMUs makes the power system more open and thus more vulnerable to cyber/terrorist attacks with unpredictable consequences [3], [21], [22]. Furthermore, contingencies such as PMU outage may lead to unobservability and thus should be treated under PMU placement optimization.

Against the above background, the present paper aims to lay down the design foundation for PMU placement to optimize information-theoretic indices subject to various observability constraints. One should distinguish PMU placement from sensor selection for spectrum sensing (see e.g. [23], [24]) or for Kalman filtering in sensor networks (see e.g. [25] and [26]). The latter aims to select sensors from the placed ones and as such it is implemented online, while the former is implemented off-line to provide the optimal conditions for the latter online state estimation. As such PMU placement design is based on off-line optimization, which is still very computationally challenging. Exhaustive search is intractable due to massive numbers of binary decision variables involved. The paper's contributions are three-fold:

- To provide analytical models for observability-constraint aware (OCA) PMU placement optimization, which aims to minimize the MSE or maximize the MI between the measurement output and grid state in replying various observability constraints, which incorporate operating conditions such as presence of ZIBs and PFMs, contingency of single PMU outage and limitations of PMU communication channels.
- These models are large scale binary nonlinear optimization problems with no known solutions, for which a novel and scalable solver is developed. The solver relies on the development of a new class of exactly penalized optimization for binary optimization and new function

approximations for scalable computation. The provided comprehensive simulation results show that the solver works well even for large-scale networks.

- Quick solvers of global optimization are proposed to address the PMU placement optimization without observability constraints, which provides baselines for performance by OCA PMU placement optimization. These solvers outperform all the existing solvers, especially for large-scale networks. Their application to sensor selection is also immediate thanks to the extremely low computational complexity.

C. Organization and Notation

The rest of the paper is structured as follows. Section II is devoted to the development of analytical models for OCA PMU placement optimization, which also particularly shows the importance of imposing state observability constraints in state estimation. Discussion of their extensions to AC power flows is also provided. Section III develops a scalable solver for computation. Section IV presents a tailored path-following discrete optimization solver for state estimation 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, $A \succ 0$ ($A \succeq 0$, resp.) for a Hermitian symmetric matrix A means that it is positive definite (semi-definite, resp.). $\text{Trace}(\cdot)$ and $|\cdot|$ are the trace and determinant, respectively. $\mathbf{1}_N$ is an N -dimensional vector of ones. I_N is the identity matrix of size N . The cardinality of a set $\mathcal{C}$ is denoted by $|\mathcal{C}|$. $\mathbb{E}(\cdot)$ denotes expectation, so the mean $\bar{u}$ of a random vector (RV) u is $\bar{u} = \mathbb{E}(u)$. For two random vectors u and v , their cross-covariance matrix R_{uv} is $\mathbb{E}((u - \bar{u})(v - \bar{v})^T)$. Accordingly, the autocovariance $\mathcal{R}_u$ of u is $\mathbb{E}((u - \bar{u})(u - \bar{u})^T)$. $u \sim \mathcal{N}(\bar{u}, \mathcal{R}_u)$ means u is a Gaussian random vector with mean $\bar{u}$ and autocovariance $\mathcal{R}_u$. The entropy of u is $\mathcal{H}(u) = \frac{1}{2} \log_2 |\mathcal{R}_u| = \frac{1}{2 \ln 2} \ln |\mathcal{R}_u|$. Finally, denote by $u|v$ a RV u conditioned on the RV v . Lastly, $\mathbb{R}_+^N := \{(x_1, \dots, x_N)^T : x_k \geq 0, k = 1, \dots, N\}$ so $\text{int}(\mathbb{R}_+^N) = \{(x_1, \dots, x_N)^T : x_k > 0, k = 1, \dots, N\}$.

II. ANALYTICAL MODELS FOR OCA PMU PLACEMENT

A. Explicit Formulas for MSE and MI in PMU Placement

Consider a power grid with a set of buses indexed by $\mathcal{N} := \{1, 2, \dots, N\}$, where the buses are connected through a set of transmission lines $\mathcal{L} \subseteq \mathcal{N} \times \mathcal{N}$, i.e. bus k is connected to bus m if and only if $(k, m) \in \mathcal{L}$. Accordingly, $\mathcal{N}(k)$ is the set of other buses connected to bus k . For illustrative purposes only, Fig.1 depicts such a power grid with 30 buses, 41 transmission lines, 6 generators and 10 PMUs.

In a DC power model, the power injection at bus k is approximated by

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

where P_k is the power injection at bus k and θ_m is the voltage phasor angle at bus m , while B_{km} is the imaginary

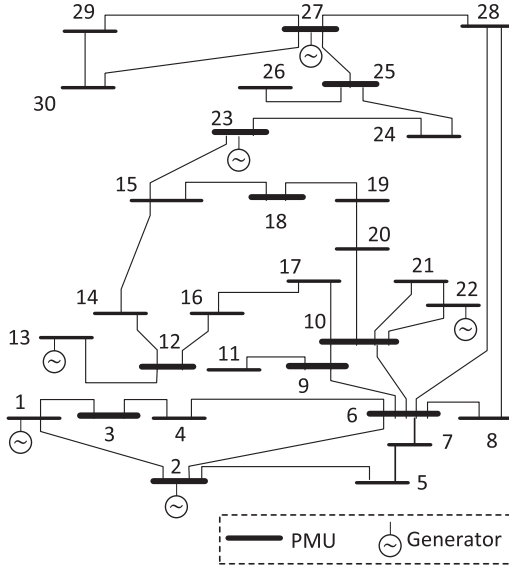


Fig. 1. IEEE 30-bus power network with PMUs.

part of the (k, m) -entry of the grid's admittance matrix Y . Let $P := (P_1, \dots, P_N)^T \in \mathbb{R}^N$ be the power injection vector and $\theta := (\theta_1, \dots, \theta_N)^T \in \mathbb{R}^N$ be the voltage phasor vector. Then the equation (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 $m \in \mathcal{N}(k)$, while $B(k, m) = 0$, otherwise. The susceptance matrix B is invertible under the assumption that the grid is fully connected [27]. Since P can be assumed to be $\mathcal{N}(u_p, \Sigma_P)$ [28], it is obvious that

$$\theta \sim \mathcal{N}(B^{-1}u_p, B^{-1}\Sigma_P(B^{-1})^T). \quad (2)$$

On the other hand, the measurement equation of a PMU installed at bus k in the linear DC power flow model is [29]

$$\begin{aligned} \zeta_k &= \theta_k + \vartheta_k, \quad k \in \mathcal{N}, \\ \zeta_{km} &= \theta_k - \theta_m + \vartheta_{km}, \quad k \in \mathcal{N}, \quad m \in \mathcal{N}(k), \end{aligned} \quad (3)$$

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 $|\mathcal{N}(k)|$. Accordingly, the measurement vector $z_k := (\zeta_k, \zeta_{k1}, \dots, \zeta_{k|\mathcal{N}(k)|})^T$ is of dimension $M_k = |\mathcal{N}(k)| + 1$. For simplicity, the equation (3) is rewritten in regression form as:

$$z_k = H_k \theta + w_k, \quad (4)$$

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

To describe the presence or absence of PMU at bus k , we introduce a selection vector $\mathbf{x} = (x_1, \dots, x_N)^T \in \{0, 1\}^N$, 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 $\sum_{k=1}^N x_k = S$. Define

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

and $\mathbf{X} = \text{diag}[x_k \mathcal{I}_{k}]_{k=1, \dots, N}$, $\mathcal{R}_w = \text{diag}[R_{w_k}]_{k=1, \dots, N}$, where $\mathcal{I}_k$ is the identity matrix of size $M_k \times M_k$.

For every $\mathbf{x} \in \mathcal{D}_S$, let $k_j \in \mathcal{N}$, $j = 1, \dots, S$ for which $x_{k_j} = 1$. Define accordingly,

$$z(\mathbf{x}) = \begin{bmatrix} z_{k_1} \\ \vdots \\ z_{k_S} \end{bmatrix}, \quad w(\mathbf{x}) = \begin{bmatrix} w_{k_1} \\ \vdots \\ w_{k_S} \end{bmatrix}, \quad \bar{H}(\mathbf{x}) = \begin{bmatrix} H_{k_1} \\ \vdots \\ H_{k_S} \end{bmatrix}. \quad (6)$$

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

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

It is obvious that $\mathcal{R}_{z(\mathbf{x})\theta} = \bar{H}(\mathbf{x})\mathcal{R}_\theta$ while $\mathcal{R}_{z(\mathbf{x})} = \bar{H}(\mathbf{x})\mathcal{R}_\theta\bar{H}(\mathbf{x})^T + \mathcal{R}_{w(\mathbf{x})}$. Let $\theta|z(\mathbf{x})$ be the RV θ conditioned on the RV $z(\mathbf{x})$. By [30, Th. 12.1]

$$\theta|z(\mathbf{x}) \sim \mathcal{N}(\hat{\theta}, \mathcal{R}_e(\mathbf{x})), \quad (7)$$

where

$$\begin{aligned} \hat{\theta} &= \bar{\theta} + \mathcal{R}_{z(\mathbf{x})\theta}^T \mathcal{R}_{z(\mathbf{x})}^{-1} (z(\mathbf{x}) - \bar{z}(\mathbf{x})) \\ &= \bar{\theta} + \mathcal{R}_\theta \bar{H}(\mathbf{x})^T (\bar{H}(\mathbf{x})\mathcal{R}_\theta\bar{H}(\mathbf{x})^T \\ &\quad + \mathcal{R}_{w(\mathbf{x})})^{-1} (z(\mathbf{x}) - \bar{H}(\mathbf{x})\bar{\theta}), \end{aligned} \quad (8)$$

which is the minimum mean squared error (MMSE) estimate of θ based on PMU output $z(\mathbf{x})$, and

$$\mathcal{R}_e(\mathbf{x}) = \mathcal{R}_\theta - \mathcal{R}_{z(\mathbf{x})\theta}^T \mathcal{R}_{z(\mathbf{x})}^{-1} \mathcal{R}_{z(\mathbf{x})\theta}. \quad (9)$$

The expression (9) only includes those $x_{k_j} = 1$ so it is not an explicit function of $\mathbf{x}$, which means that it cannot be used for systematic computation. Fortunately, its analytical form can be derived as follows:

$$\begin{aligned} \mathcal{R}_e(\mathbf{x}) &= \mathcal{R}_\theta - \mathcal{R}_{z(\mathbf{x})\theta}^T \mathcal{R}_{z(\mathbf{x})}^{-1} \mathcal{R}_{z(\mathbf{x})\theta} \\ &= \mathcal{R}_\theta - \mathcal{R}_\theta \bar{H}(\mathbf{x})^T (\bar{H}(\mathbf{x})\mathcal{R}_\theta\bar{H}(\mathbf{x})^T \\ &\quad + \mathcal{R}_{w(\mathbf{x})})^{-1} \bar{H}(\mathbf{x})\mathcal{R}_\theta \\ &= \left(\mathcal{R}_\theta^{-1} + \bar{H}(\mathbf{x})^T \mathcal{R}_{w(\mathbf{x})}^{-1} \bar{H}(\mathbf{x}) \right)^{-1} \\ &= \left(\mathcal{R}_\theta^{-1} + \sum_{j=1}^S H_{k_j}^T \mathcal{R}_{w_{k_j}}^{-1} H_{k_j} \right)^{-1} \\ &= \left(B^T \Sigma_P^{-1} B + \sum_{k=1}^N x_k H_k^T \mathcal{R}_{w_k}^{-1} H_k \right)^{-1}. \end{aligned} \quad (10)$$

Therefore, the mean squared error (MSE) $\mathbb{E}(\|\theta - \hat{\theta}\|^2) = \text{Trace}(\mathcal{R}_e(\mathbf{x}))$ is

$$f_e(\mathbf{x}) = \text{Trace} \left(\left(B^T \Sigma_P^{-1} B + \sum_{k=1}^N x_k H_k^T \mathcal{R}_{w_k}^{-1} H_k \right)^{-1} \right), \quad (11)$$

which is an analytical function of the PMU selection vector $\mathbf{x}$. This function is not only continuous but convex.

Further, the mutual information (MI) $I(\theta; z(\mathbf{x}))$ between RVs θ and $z(\mathbf{x})$ is [31, formula (6)]

$$\begin{aligned} I(\theta; z(\mathbf{x})) &= \mathcal{H}(\theta) - \mathcal{H}(\theta|z(\mathbf{x})) \\ &= \frac{1}{2 \ln 2} (\ln |\mathcal{R}_\theta| - \ln |\mathcal{R}_e(\mathbf{x})|) \\ &= \frac{1}{2 \ln 2} \ln |\mathcal{R}_\theta| + \frac{1}{2 \ln 2} \ln |B^T \Sigma_P^{-1} B + \\ &\quad + \sum_{k=1}^N x_k H_k^T \mathcal{R}_{w_k}^{-1} H_k|. \end{aligned} \quad (12)$$

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

$$\begin{aligned} f_{MI}(\mathbf{x}) &:= -\ln |\mathcal{R}_e(\mathbf{x})| \\ &= \ln |B^T \Sigma_P^{-1} B + \sum_{k=1}^N x_k H_k^T R_{w_k}^{-1} H_k|, \end{aligned} \quad (13)$$

which is not only continuous but concave.

To the authors' best knowledge, the explicit expressions (11) and (12) for the MSE and MI in PMU placement optimization are new. Incidentally, the right hand side of (10) was regarded in [14] as the gain matrix in the maximum likelihood estimate. Furthermore, [16] used the implicit expression (7) and (9) for the MI. It follows from the explicit expression (10) that

$$\mathcal{R}_e(\mathbf{x}) \succeq \mathcal{R}_e(\tilde{\mathbf{x}}) \quad \text{whenever} \quad \tilde{\mathbf{x}} - \mathbf{x} \in \mathbb{R}_+^N, \quad (14)$$

leading to

$$\text{Trace}(\mathcal{R}_e(\mathbf{x})) \geq \text{Trace}(\mathcal{R}_e(\tilde{\mathbf{x}})) \quad \& \quad \ln |\mathcal{R}_e(\mathbf{x})| \geq \ln |\mathcal{R}_e(\tilde{\mathbf{x}})| \quad (15)$$

which means that increasing the number S of PMUs improves both MMSE index $\mathbb{E}(\|\theta - \hat{\theta}\|^2)$ and MI index $I(\theta; z(\mathbf{x}))$.

B. Observability Constraints on PMU Placement

This subsection provides a machinery to maintain the information quality of MSE and MI.

1) *Complete Observability (CO)*: To assure the complete observability of system state θ , one needs the following constraint [1], [32], [33]:

$$\mathcal{A}\mathbf{x} \geq \mathbf{1}_N, \quad (16)$$

where $\mathcal{A}$ is the bus-to-bus incidence matrix defined by $\mathcal{A}_{km} = 1$ if $k = m$ or bus k is adjacent to bus m , and $\mathcal{A}_{km} = 0$ otherwise.

Let us analyse the constraint (16) from the information-theoretic view point. The constraint (16) guarantees that all state components θ_m are observable, i.e. each θ_m appears at least once in the measurement equations (3), which implies $\theta_m|z(\mathbf{x}) \neq \theta_m$, making the measurement equations (3) meaningful for estimating θ_m . When some θ_m is not observable, i.e. it does not appear in the measurement equations (3), it follows that $\theta_m|z(\mathbf{x}) = \theta_m$, so the measurement equations in (3) are useless for estimating θ_m . In this case, the estimate for θ_m is its unconditional mean θ_m with $\mathbb{E}((\theta_m - \hat{\theta}_m)^2) = \mathcal{R}_\theta(m, m)$ and $I(\theta_m; z(\mathbf{x})) = \mathcal{H}(\theta) - \mathcal{H}(\theta|z(\mathbf{x})) = 0$. In other words, the optimization problem for maximizing $I(\theta; z(\mathbf{x}))$ does not reveal a nontrivial estimate for θ_m that is a contradiction to [16, statement 1), page 448, 2nd column] which states that the mutual information metric includes the complete observability condition (16) as a special case.

2) *Zero Injection Buses (ZIBs)*: It is also known [7]–[9] that ZIBs, which are neither loads nor generators are helpful for improving observability. Let $\mathcal{Z}$ be the set of ZIBs and $\chi_{\mathcal{Z}}(\cdot)$ be its indicator function, i.e. $\chi_{\mathcal{Z}}(m) = 1$ for $m \in \mathcal{Z}$ and $\chi_{\mathcal{Z}}(m) = 0$ otherwise. For $\mathbf{y} := [y_{k,m}]_{(k,m) \in \mathcal{N} \times \mathcal{Z}}$, where

$$y_{km} \in \{0, 1\}, \quad (k, m) \in \mathcal{N} \times \mathcal{Z} \quad (17)$$

are the auxiliary binary variables to incorporate the impact of ZIBs, the observability constraints incorporating the impact of ZIBs are [8]

$$\sum_{k \in \mathcal{N}} \mathcal{A}_{km} y_{km} = \chi_{\mathcal{Z}}(m), \quad m \in \mathcal{N}, \quad (18a)$$

$$\begin{aligned} F_k(\mathbf{x}, \mathbf{y}) &:= \sum_{m \in \mathcal{N}} \mathcal{A}_{km} x_m \\ &+ \sum_{m \in \mathcal{N}} \mathcal{A}_{km} \chi_{\mathcal{Z}}(m) y_{km} \geq 1, \quad k \in \mathcal{N}. \end{aligned} \quad (18b)$$

For $m \in \mathcal{Z}$, the constraints (18) mean that all its incident buses are observable except one, which is then observable by applying the Kirchhoff's current law (KCL) at m . For $m \notin \mathcal{Z}$, the constraints (18) mean that all its incident buses are observable so m is observable by applying KCL at m .

3) *Power Flow Measurements (PFMs)*: Next, the pre-installed conventional PFMs can improve measurement redundancy, which is advantageous not only for observability but also for bad data detection [34]. All buses in a branch with a power measurement are made observable once any of them is observable. Suppose that $\mathcal{B}$ is the set of branches with power flow measurements, and $\mathcal{T}$ is the set of buses that are the terminal points of the branch in $\mathcal{B}$. For $F_k(\mathbf{x}, \mathbf{y})$ defined from (18b), the constraint (18b) can be replaced by [34]

$$F_k(\mathbf{x}, \mathbf{y}) + F_m(\mathbf{x}, \mathbf{y}) \geq 1, \quad (k, m) \in \mathcal{B}, \quad (19a)$$

$$F_k(\mathbf{x}, \mathbf{y}) \geq 1, \quad k \in \mathcal{N} \setminus \mathcal{T}, \quad (19b)$$

where $k \in \mathcal{N} \setminus \mathcal{T}$ is the bus which is not in bus set $\mathcal{T}$.

4) *Contingency of Single PMU Outage*: As an electrical device, a PMU may be inactive in some real case. In order to guarantee that all buses are still observable when a single PMU is lost, the following contingency-aware constraint should be imposed [8]:

$$\sum_{m \in \mathcal{N}} \mathcal{A}_{km} x_m + \sum_{m \in \mathcal{N}} \mathcal{A}_{km} y_{km} \geq 2, \quad k \in \mathcal{N}. \quad (20)$$

5) *Limitation of PMU Channels*: In many scenarios, a PMU may not measure all phasors of incident buses due to the limitation of its communication channels. A binary variable c_{km} is then introduced to indicate that $c_{km} = 1$ whenever bus k is measured by a PMU installed at bus m , and $c_{km} = 0$ otherwise. Following [35, (25)–(28)], the following observability constraint is needed instead of (18b)

$$\sum_{m \in \mathcal{N}} \mathcal{A}_{km} c_{km} + \sum_{m \in \mathcal{N}} \mathcal{A}_{km} \chi_{\mathcal{Z}}(m) y_{km} \geq 1, \quad k \in \mathcal{N}, \quad (21a)$$

$$x_m \leq \sum_{k \in \mathcal{N}} \mathcal{A}_{km} c_{km} \leq C_m, \quad m \in \mathcal{N}, \quad (21b)$$

$$\mathcal{A}_{km} c_{km} \leq x_m, \quad k, m \in \mathcal{N}, \quad (21c)$$

$$c_{km} \in \{0, 1\}, \quad (k, m) \in \mathcal{N} \times \mathcal{N}, \quad (21d)$$

where C_m is the maximum number of PMU channel at bus m .

C. Analytical PMU Placement Optimization and Computational Challenges

For $\mathbf{x}, \mathbf{y}$ and $\mathbf{c} := [c_{k,m}]_{(k,m) \in \mathcal{N} \times \mathcal{N}}$ as the binary optimization variables, we are now in position to state the problem

of PMU placement optimization to minimize the MSE or to maximize the MI between the measurement output and phasor state subject to a fixed number of PMUs and observability constraints as the following binary nonlinear optimization problems

$$\min_{\mathbf{x}, \mathbf{y}, \mathbf{c}} f(\mathbf{x}) \text{ s.t. } \mathbf{x} \in \mathcal{D}_S, (17), (18a), (18b)/(19)/(20)/(21), \quad (22)$$

where $f(\mathbf{x}) \in \{f_e(\mathbf{x}), -f_{MI}(\mathbf{x})\}$, which is a convex function, and $\mathcal{D}_S$ is defined from (5). In what follows, we call (22) OCA PMU placement optimization.

Most existing works used binary linear programming in handling some but not all constraints in (22) without optimizing f . For instance, the following binary linear problem of minimizing the number of PMUs for the system observability

$$\min_{\mathbf{x}} \sum_{k=1}^N x_k : \mathbf{x} \in \{0, 1\}^N, (16) \quad (23)$$

was considered in [7], [11], [13], which can not guarantee informative PMU configuration. Particular cases of the following binary linear problems

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{y}, \mathbf{c}} \sum_{k=1}^N x_k \\ \text{s.t. } \mathbf{x} \in \{0, 1\}^N, (17), (18a), (18b)/(19)/(20)/(21), \end{aligned} \quad (24)$$

were considered in [7], [11], [13]. For instance, the constraints (20) and (21) of the contingency of PMU outage and limitation of PMU communication channels were not present in [7], the constraints (19) and (21) of the PFMs and limitation of PMU communication channels were absent in [11], while the constraint (21) of the limitation of PMU communication channels were not considered in [13].

In optimizing the so called gain matrix [14], the work [15] considered the simple binary convex problem

$$\min_{\mathbf{x}} f(\mathbf{x}) \text{ s.t. } \mathbf{x} \in \mathcal{D}_S, \quad (25)$$

by solving its convex relaxation problem, which is

$$\min_{\mathbf{x} \in \mathbb{R}_+^N, \mathbf{T} \in \mathbb{R}^{N \times N}} \text{Trace}(\mathbf{T}) \text{ s.t. } \mathbf{x} \in \text{Poly}(\mathcal{D}_S), \quad (26a)$$

$$\begin{bmatrix} B^T \Sigma_P^{-1} B + \sum_{k=1}^N x_k H_k^T R_{w_k}^{-1} H_k & I_N \\ I_N & \mathbf{T} \end{bmatrix} \succeq 0, \quad (26b)$$

when $f = f_e$, or

$$\begin{aligned} \max_{\mathbf{x} \in \mathbb{R}_+^N} \ln |B^T \Sigma_P^{-1} B + \sum_{k=1}^N x_k H_k^T R_{w_k}^{-1} H_k| \\ \text{s.t. } \mathbf{x} \in \text{Poly}(\mathcal{D}_S), \end{aligned} \quad (27)$$

when $f = f_{MI}$, for

$$\text{Poly}(\mathcal{D}_S) = \{\mathbf{x} \in [0, 1]^N : \sum_{k=1}^N x_k = S\}, \quad (28)$$

and rounding the S largest entries of their optimal solution to one and the remaining $N - S$ entries to zero. Note that there is no guarantee that the solution of (26) or (27) is sparse to

make such rounding efficient. Moreover, the size of the semi-definite optimization problem (26) is not scalable in N as it invokes the additional slack symmetric matrix variable $\mathbf{T}$ of size $N \times N$ that involves $N(N + 1)/2$ additional decision variables. For a moderate number N such as $N = 118$, the value of $N(N + 1)/2$ is 7021, which is already a huge number. There is no solver of polynomial complexity for solving (27). The reader is referred to [36] for capacity of SDR to address discrete optimization problems such as (25). Furthermore, by showing that the MI is a submodular function, the work [16] has shown that the PMU placement to maximize the MI can be very efficiently solved by the very low computational complexity greedy algorithm [17] for submodular function optimization with a really sound analytical foundation. However, both convex relaxation-based algorithm and greedy algorithm cannot be used for addressing the problem (22).

It is obvious that the problem (22) is much more computationally challenging than either the problem (23) or the problem (24). While the latter is a binary linear problem, which can be solved by very powerful binary solvers such as CPLEX [37], the former is a large scale binary nonlinear problem, which is among the most computationally difficult optimization problems with no known solution method. The next section is devoted to its computational solution.

D. Discussion on Extension to AC Power Flow Models

The optimization formulation (22) for $f(\mathbf{x}) = f_e(\mathbf{x})$ can be extended to the case of AC power flows models [38], under which the equation (1) is nonlinear. Like [25] and [26] for nonlinear sensor networks, one can use the unscented transformation and MMSE result of [39] to approximate the RV θ by a Gaussian RV as in (2) or by Gaussian mixture RV and then the conditional RV $\theta|z(\mathbf{x})$ in (7) for deriving the MMSE estimator $\hat{\theta}$ in (8) so the resultant MMSE is still an analytical function in $\mathbf{x}$ as that defined by (9).

III. SCALABLE PENALTY ALGORITHMS FOR OPTIMAL PMU PLACEMENT

To address the OCA PMU placement optimization (22) we need to handle its discrete constraint $\mathbf{x} \in \mathcal{D}_S$. An important observation is that $x^L < x$ whenever $0 < x < 1$ and $L > 1$, and $x^L = x$ whenever $x = 0$ or $x = 1$. Therefore, the binary constraint $x \in \{0, 1\}$ is expressed by two continuous constraints $x \in [0, 1]$ and $x^L = x$. This helps to express the discrete constraint $\mathbf{x} \in \mathcal{D}_S$ by continuous constraints as follows.

Lemma 1: For the polytope $\text{Poly}(\mathcal{D}_S)$ defined from (28), the discrete constraint $\mathbf{x} \in \mathcal{D}_S$ is equivalent to the continuous constraints

$$\mathbf{x} \in \text{Poly}(\mathcal{D}_S), \quad (29a)$$

$$g_1(\mathbf{x}) \geq S, \quad (29b)$$

for $g_1(\mathbf{x}) := \sum_{k=1}^N x_k^L$ with $L > 1$.

Proof. Note that $x_k^L \leq x_k \forall x_k \in [0, 1]$, so $g_1(\mathbf{x}) \leq \sum_{k=1}^N x_k = S \forall \mathbf{x} \in \text{Poly}(\mathcal{D}_S)$. Therefore constraint (29) forces $g(\mathbf{x}) = S$, which is possible if and only if $x_k^L = x_k$, $k = 1, \dots, N$, i.e. $x_k \in \{0, 1\}$, $k = 1, \dots, N$, implying $\mathbf{x} \in \mathcal{D}_S$. $\square$

Since the function $g_1(\mathbf{x})$ is convex, the constraint $g_1(\mathbf{x}) \geq S$ in (29) is a reverse convex constraint [40]. As such $\mathcal{D}_S = \text{Poly}(\mathcal{D}_S) \setminus \{\mathbf{x} : g_1(\mathbf{x}) < S\}$, i.e. $\mathcal{D}_S$ is the difference of two convex sets $\text{Poly}(\mathcal{D}_S)$ and $\{\mathbf{x} : g_1(\mathbf{x}) < S\}$. Also as L decreases, $g_1(\mathbf{x})$ tends to approach the linear function $\sum_{k=1}^N x_k$ and thus, the constraint $g_1(\mathbf{x}) \geq S$ approaches the linear constraint $\sum_{k=1}^N x_k \geq S$. However, it does not mean that choosing L closer to 1 is effective because the function $g_1(\mathbf{x}) - S$ also approaches to zero very quickly, making the constraint $g_1(\mathbf{x}) \geq S$ highly artificial. In our previous works [41], [42], $L = 2$ 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

$$\tilde{g}_1(\mathbf{x}) = 1/g_1(\mathbf{x}) - 1/S$$

can be used to measure the degree of satisfaction of the discrete constraint $\mathbf{x} \in \mathcal{D}_S$ in the sense that $\tilde{g}_1(\mathbf{x}) \geq 0 \forall \mathbf{x} \in \text{Poly}(\mathcal{D}_S)$ and $\tilde{g}_1(\mathbf{x}) = 0$ if and only if $\mathbf{x} \in \mathcal{D}_S$. $\square$

Similarly, over the box domain

$$y_{km} \in [0, 1], (k, m) \in \mathcal{N} \times \mathcal{Z}, \quad (30)$$

and

$$c_{km} \in [0, 1], (k, m) \in \mathcal{N} \times \mathcal{N}, \quad (31)$$

the following functions can be used to measure the satisfaction of the discrete constraint (17) and (21d),

$$\tilde{g}_2(\mathbf{y}) = 1/g_2(\mathbf{y}) - 1/|\mathcal{Z}|, \quad (32)$$

and

$$\tilde{g}_3(\mathbf{c}) = 1/g_3(\mathbf{c}) - 1/(N - |\mathcal{Z}|), \quad (33)$$

with

$$g_2(\mathbf{y}) := \sum_{k=1}^N \sum_{m \in \mathcal{Z}} y_{km}^L,$$

and

$$g_3(\mathbf{c}) := \sum_{k=1}^N \sum_{m=1}^N c_{km}^L.$$

Since each ZIB can exactly make one bus observable, we need at least $|\mathcal{N}| - |\mathcal{Z}|$ PMU channels for the complete observability.

Following the developments in [41]–[45], instead of handling the nonconvex constraints (29b) and $g_2(\mathbf{y}) \geq |\mathcal{Z}|$ and $g_3(\mathbf{c}) \geq N - |\mathcal{Z}|$ we incorporate the degree of their satisfaction into the objective in (22), leading to the following penalized optimization problem:

$$\min_{\mathbf{x}, \mathbf{y}, \mathbf{c}} F_\mu(\mathbf{x}, \mathbf{y}, \mathbf{c}) := f(\mathbf{x}) + \mu(\tilde{g}_1(\mathbf{x}) + \tilde{g}_2(\mathbf{y}) + \tilde{g}_3(\mathbf{c})) \quad (34a)$$

$$\text{s.t. (29a), (18a), (18b)/(19)/(20)/(21), (30), (31),} \quad (34b)$$

where $\mu > 0$ is a penalty parameter. This penalized optimization problem is exact with sufficiently large μ in the sense that its optimal solution is also optimal for (22). Note that the

problem (34) 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 (34), at some its feasible point $(\mathbf{x}^{(\kappa)}, \mathbf{y}^{(\kappa)}, \mathbf{c}^{(\kappa)})$ (at κ -th iteration). As the function $g_1(\mathbf{x})$ is convex, it is true that [40],

$$\begin{aligned} g_1(\mathbf{x}) &\geq g_1^{(\kappa)}(\mathbf{x}) \\ &:= g_1(\mathbf{x}^{(\kappa)}) + \langle \nabla g_1(\mathbf{x}^{(\kappa)}), \mathbf{x} - \mathbf{x}^{(\kappa)} \rangle \\ &= -(L-1) \sum_{k=1}^N (x_k^{(\kappa)})^L + L \sum_{k=1}^N (x_k^{(\kappa)})^{L-1} x_k. \end{aligned}$$

Therefore, an upper bounding approximation at $\mathbf{x}^{(\kappa)}$ for $1/g_1(\mathbf{x})$ can be easily obtained as $1/g_1(\mathbf{x}) \leq 1/g_1^{(\kappa)}(\mathbf{x})$ over the trust region

$$g_1^{(\kappa)}(\mathbf{x}) > 0. \quad (35)$$

Analogously, $1/g_2(\mathbf{y}) \leq 1/g_2^{(\kappa)}(\mathbf{y})$ and $1/g_3(\mathbf{c}) \leq 1/g_3^{(\kappa)}(\mathbf{c})$ over the trust region

$$g_2^{(\kappa)}(\mathbf{y}) > 0, \quad (36)$$

and

$$g_3^{(\kappa)}(\mathbf{c}) > 0, \quad (37)$$

for

$$\begin{aligned} g_2^{(\kappa)}(\mathbf{y}) &:= -(L-1) \sum_{k=1}^N \sum_{m \in \mathcal{Z}} (y_{km}^{(\kappa)})^L \\ &\quad + L \sum_{k=1}^N \sum_{m \in \mathcal{Z}} (y_{km}^{(\kappa)})^{L-1} y_{km}, \end{aligned}$$

and

$$\begin{aligned} g_3^{(\kappa)}(\mathbf{c}) &:= -(L-1) \sum_{k=1}^N \sum_{m=1}^N (c_{km}^{(\kappa)})^L \\ &\quad + L \sum_{k=1}^N \sum_{m=1}^N (c_{km}^{(\kappa)})^{L-1} c_{km}. \end{aligned}$$

At the κ -th iteration we solve the following convex optimization problem to generate the next iterative point $(\mathbf{x}^{(\kappa+1)}, \mathbf{y}^{(\kappa+1)}, \mathbf{c}^{(\kappa+1)})$:

$$\min_{\mathbf{x}, \mathbf{y}, \mathbf{c}} f(\mathbf{x}) + \mu P^{(\kappa)}(\mathbf{x}, \mathbf{y}, \mathbf{c}), \quad \text{s.t. (34b), (35), (36), (37), (38)}$$

with

$$\begin{aligned} P^{(\kappa)}(\mathbf{x}, \mathbf{y}, \mathbf{c}) &:= \left(\frac{1}{g_1^{(\kappa)}(\mathbf{x})} - \frac{1}{S} \right) + \left(\frac{1}{g_2^{(\kappa)}(\mathbf{y})} - \frac{1}{|\mathcal{Z}|} \right) \\ &\quad + \left(\frac{1}{g_3^{(\kappa)}(\mathbf{c})} - \frac{1}{N - |\mathcal{Z}|} \right). \end{aligned}$$

Although the function $f(\mathbf{x})$ is already convex, it is not easy to optimize it. For instance, when $f = f_e$, usually f_e is expressed by $\text{Trace}(\mathbf{T})$, where $\mathbf{T}$ is a slack symmetric matrix variable of size $N \times N$ satisfying the semi-definite constraint (26b), which is not scalable to $\mathbf{x}$. Worse, for $f = -f_{MI}$, which is $\ln |B^T \Sigma_P^{-1} B + \sum_{k=1}^N x_k H_k^T R_{w_k}^{-1} H_k|$, there is no known convex solver of polynomial complexity.

In the following, we propose a different approach to provide for scalable iterations for (38). Obviously, there is $\epsilon > 0$ such that

$$\mathcal{A}_\epsilon := B^T \Sigma_P^{-1} B - \epsilon \sum_{k=1}^N H_k^T R_{w_k}^{-1} H_k \succ 0.$$

For $f = f_e$, applying the inequality (53) in the Appendix for

$$A_0 \rightarrow \mathcal{A}_\epsilon, \quad x_k \rightarrow x_k + \epsilon, \quad \bar{x}_k \rightarrow x_k^{(\kappa)} + \epsilon, \quad (39)$$

yields $f_e(\mathbf{x}) \leq f_e^{(\kappa)}(\mathbf{x}) := a_0^{(\kappa)} + \sum_{k=1}^N \frac{a_k^{(\kappa)}}{x_k + \epsilon}$ for

$$0 < a_0^{(\kappa)} := \text{Trace}((\mathcal{R}_e(\mathbf{x}^{(\kappa)}))^2 \mathcal{A}_\epsilon), \quad (40a)$$

$$0 < a_k^{(\kappa)} := (x_k^{(\kappa)} + \epsilon)^2 \text{Trace}((\mathcal{R}_e(\mathbf{x}^{(\kappa)}))^2 \times H_k^T R_{w_k}^{-1} H_k), k = 1, \dots, N. \quad (40b)$$

Accordingly, initialized by a feasible point $(\mathbf{x}^{(0)}, \mathbf{y}^{(0)}, \mathbf{c}^{(0)})$ for (34), at the κ -th iteration for $\kappa = 0, 1, \dots$, we solve the following convex optimization problem to generate the next iterative point $(\mathbf{x}^{(\kappa+1)}, \mathbf{y}^{(\kappa)}, \mathbf{c}^{(\kappa)})$, instead of (38):

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{y}, \mathbf{c}} F_\mu^{(\kappa)}(\mathbf{x}, \mathbf{y}, \mathbf{c}) &:= f_e^{(\kappa)}(\mathbf{x}) + \mu P^{(\kappa)}(\mathbf{x}, \mathbf{y}, \mathbf{c}) \\ \text{s.t.} \quad &(34b), (35), (36), (37). \end{aligned} \quad (41)$$

Note that $F_\mu(\mathbf{x}, \mathbf{y}, \mathbf{c}) \leq F_\mu^{(\kappa)}(\mathbf{x}, \mathbf{y}, \mathbf{c})$, $\forall (\mathbf{x}, \mathbf{y}, \mathbf{c})$, and $F_\mu(\mathbf{x}^{(\kappa)}, \mathbf{y}^{(\kappa)}, \mathbf{c}^{(\kappa)}) = F_\mu^{(\kappa)}(\mathbf{x}^{(\kappa)}, \mathbf{y}^{(\kappa)}, \mathbf{c}^{(\kappa)})$, and $F_\mu^{(\kappa)}(\mathbf{x}^{(\kappa+1)}, \mathbf{y}^{(\kappa+1)}, \mathbf{c}^{(\kappa+1)}) < F_\mu^{(\kappa)}(\mathbf{x}^{(\kappa)}, \mathbf{y}^{(\kappa)}, \mathbf{c}^{(\kappa)})$, since $(\mathbf{x}^{(\kappa+1)}, \mathbf{y}^{(\kappa+1)}, \mathbf{c}^{(\kappa+1)})$ and $(\mathbf{x}^{(\kappa)}, \mathbf{y}^{(\kappa)}, \mathbf{c}^{(\kappa)})$ are the optimal solution and a feasible point for (41). Therefore,

$$\begin{aligned} F_\mu(\mathbf{x}^{(\kappa+1)}, \mathbf{y}^{(\kappa+1)}, \mathbf{c}^{(\kappa+1)}) &\leq F_\mu^{(\kappa)}(\mathbf{x}^{(\kappa+1)}, \mathbf{y}^{(\kappa+1)}, \mathbf{c}^{(\kappa+1)}) \\ &< F_\mu^{(\kappa)}(\mathbf{x}^{(\kappa)}, \mathbf{y}^{(\kappa)}, \mathbf{c}^{(\kappa)}) \\ &= F_\mu(\mathbf{x}^{(\kappa)}, \mathbf{y}^{(\kappa)}, \mathbf{c}^{(\kappa)}), \end{aligned}$$

i.e. $(\mathbf{x}^{(\kappa+1)}, \mathbf{y}^{(\kappa+1)}, \mathbf{c}^{(\kappa+1)})$ is a better feasible point than $(\mathbf{x}^{(\kappa)}, \mathbf{y}^{(\kappa)}, \mathbf{c}^{(\kappa)})$ for (34). For a sufficient large $\mu > 0$, $\tilde{g}_1(\mathbf{x}^{(\kappa)}) + \tilde{g}_2(\mathbf{y}^{(\kappa)}) + \tilde{g}_3(\mathbf{c}^{(\kappa)}) \rightarrow 0$ as well, yielding an optimal solution of the binary nonlinear optimization problem (22) for the case $f = f_e$. Algorithm 1 provides a pseudo-code for the proposed computational procedure.

Algorithm 1 Scalable Penalized MSE Algorithm

- 1: **Initialization.** Set $\kappa = 0$. Take any feasible point $(\mathbf{x}^{(0)}, \mathbf{y}^{(0)}, \mathbf{c}^{(0)})$ for (34). Choose $\mu > 0$ such that $f_e(\mathbf{x}^{(0)})$ and $\mu(\tilde{g}_1(\mathbf{x}^{(0)}) + \tilde{g}_2(\mathbf{y}^{(0)}) + \tilde{g}_3(\mathbf{c}^{(0)}))$ have a similar magnitude.
 - 2: **Repeat until the convergence of the objective in (22):** Solve the convex optimization problem (41) to generate the next feasible point $(\mathbf{x}^{(\kappa+1)}, \mathbf{y}^{(\kappa+1)}, \mathbf{c}^{(\kappa+1)})$; Set $\kappa := \kappa + 1$.
-

Analogously, considering $f = -f_{MI}$, based on the inequality (55) in the Appendix, for A_0 , x_k , and $\bar{x}_k$ defined from (39), at the κ -th iteration we solve the following convex optimization problem to generate the next iterative point $\mathbf{x}^{(\kappa+1)}$ and $\mathbf{y}^{(\kappa+1)}$, instead of (38):

$$\begin{aligned} \max_{\mathbf{x}, \mathbf{y}, \mathbf{c}} &\left[f_{MI}^{(\kappa)}(\mathbf{x}) - \mu P^{(\kappa)}(\mathbf{x}, \mathbf{y}, \mathbf{c}) \right] \\ \text{s.t.} \quad &(34b), (35), (36), (37), \end{aligned} \quad (42)$$

for

$$f_{MI}^{(\kappa)}(\mathbf{x}) := a_0^{(\kappa)} - \sum_{k=1}^N \frac{a_k^{(\kappa)}}{x_k + \epsilon} \quad (43)$$

and

$$\begin{aligned} a_0^{(\kappa)} &:= -\ln |\mathcal{R}_e(\mathbf{x}^{(\kappa)})| + \text{Trace}(\mathcal{R}_e(\mathbf{x}^{(\kappa)}) \\ &\quad \times \left(\sum_{k=1}^N (\epsilon + x_k^{(\kappa)}) H_k^T R_{w_k}^{-1} H_k \right)), \end{aligned} \quad (44a)$$

$$\begin{aligned} a_k^{(\kappa)} &:= (x_k^{(\kappa)} + \epsilon)^2 \text{Trace}(\mathcal{R}_e(\mathbf{x}^{(\kappa)}) \\ &\quad \times H_k^T R_{w_k}^{-1} H_k), k = 1, \dots, N. \end{aligned} \quad (44b)$$

We thus adjust Algorithm 1 by solving the convex optimization problem (42) at the κ th iteration instead of (41) for computational solution of the binary nonlinear optimization problem (22) for the case $f = -f_{MI}$.

The computational complexity of (41)/(42) is

$$\mathcal{O}(\alpha^2 \beta^{2.5} + \beta^{3.5}), \quad (45)$$

where

$$\alpha = N + \sum_{k \in \mathcal{Z}} \sum_{m \in \mathcal{N}} \mathcal{A}_{km} + \sum_{k \in \mathcal{N}} \sum_{m \in \mathcal{N}} \mathcal{A}_{km},$$

which is the number of decision variables, and

$$\beta = 4 + 4N + |\mathcal{C}|,$$

which is the number of constraints under the scenario with (18) for ZIBs and (21) for the PMU channels' limitation.

IV. TAILORED PATH-FOLLOWING DISCRETE OPTIMIZATION ALGORITHMS

In this section, we address the problem (25), which provides a lower bound for the optimal value of (22), i.e. it provides a lower bound for MSE and an upper bound for MI in OCA PMU placement optimization.

For any $\mathcal{K} \subset \mathcal{N}$ we define $\mathbf{x}_{\mathcal{K}} = (x_1, \dots, x_N)^T$ such that $x_k = 1$ for $k \in \mathcal{K}$ and $x_k = 0$ otherwise. Accordingly,

$$\mathcal{R}_e(\mathbf{x}_{\mathcal{K}}) = \left(B^T \Sigma_P^{-1} B + \sum_{k \in \mathcal{K}} H_k^T \mathcal{R}_{w_k}^{-1} H_k \right)^{-1}.$$

Thanks to the explicit expression (10), the greedy algorithm [17] for computing (25) is simply excused as to initialize from the set $\mathcal{K}$ of the selected PMUs as an empty set and process the following recursions for $\kappa = 1, \dots, S$:

$$\begin{aligned} k_\kappa &= \arg \min_{k \in \mathcal{N} \setminus \mathcal{K}} \text{Trace}((\mathcal{R}_e^{-1}(\mathbf{x}_{\mathcal{K}}) + H_k^T \mathcal{R}_{w_k}^{-1} H_k)^{-1}) \\ &\quad / \arg \max_{k \in \mathcal{N} \setminus \mathcal{K}} |\mathcal{R}_e^{-1}(\mathbf{x}_{\mathcal{K}}) + H_k^T \mathcal{R}_{w_k}^{-1} H_k| \end{aligned} \quad (46)$$

and

$$\mathcal{K} \rightarrow \mathcal{K} \cup k_\kappa. \quad (47)$$

Although such greedy algorithm is of heuristic type, the following result shows that its analytical foundation is sound thanks to the inequalities in (15).

Theorem 1: For γ_{opt} and γ_{gr} as the global optimal value of the problem (25) and that found by the greedy algorithm, the following approximation ratio is achieved

$$\frac{\gamma_{gr}}{\gamma_{opt}} \geq 1 - \frac{1}{e}. \quad (48)$$

Proof. It follows from (14) that for $\forall \mathcal{B} \subset \mathcal{A} \subset \mathcal{N}$,

$$\mathcal{R}_e(\mathbf{x}_\mathcal{A}) \preceq \mathcal{R}_e(\mathbf{x}_\mathcal{B}) \quad \& \quad \mathcal{R}_e^{-1}(\mathbf{x}_\mathcal{A}) \succeq \mathcal{R}_e^{-1}(\mathbf{x}_\mathcal{B}). \quad (49)$$

Applying the inequality (57) in the Appendix for $X = \mathcal{R}_e^{-1}(\mathbf{x}_\mathcal{A})$, $Y = \mathcal{R}_e^{-1}(\mathbf{x}_\mathcal{B})$ and $A = H_k^T \mathcal{R}_{w_k}^{-1} H_k$, $k \in \mathcal{N} \setminus \mathcal{A}$ yields

$$\begin{aligned} f_e(\mathbf{x}_\mathcal{A}) - f_e(\mathbf{x}_{\mathcal{A} \cup \{k\}}) &= \text{Trace}(\mathcal{R}_e(\mathbf{x}_\mathcal{A})) - \text{Trace}((\mathcal{R}_e^{-1}(\mathbf{x}_\mathcal{A}) \\ &\quad + H_k^T \mathcal{R}_{w_k}^{-1} H_k)^{-1}) \\ &\leq \text{Trace}(\mathcal{R}_e(\mathbf{x}_\mathcal{B})) - \text{Trace}((\mathcal{R}_e^{-1}(\mathbf{x}_\mathcal{B}) \\ &\quad + H_k^T \mathcal{R}_{w_k}^{-1} H_k)^{-1}) \\ &= f_e(\mathbf{x}_\mathcal{B}) - f_e(\mathbf{x}_{\mathcal{B} \cup \{k\}}). \end{aligned} \quad (50)$$

Furthermore, by (14),

$$\begin{aligned} f_{MI}(\mathbf{x}_{\mathcal{A} \cup \{k\}}) - f_{MI}(\mathbf{x}_\mathcal{A}) &= \ln \left| I + \mathcal{R}_{w_k}^{-1/2} H_k \mathcal{R}_e(\mathbf{x}_\mathcal{A}) H_k^T \mathcal{R}_{w_k}^{-1/2} \right| \\ &\leq \ln \left| I + \mathcal{R}_{w_k}^{-1/2} H_k \mathcal{R}_e(\mathbf{x}_\mathcal{B}) H_k^T \mathcal{R}_{w_k}^{-1/2} \right| \\ &= f_{MI}(\mathbf{x}_{\mathcal{B} \cup \{k\}}) - f_{MI}(\mathbf{x}_\mathcal{B}). \end{aligned} \quad (51)$$

The inequalities (50) and (51) show that the function f_e is supermodular and the function f_{MI} is submodular, under which the ratio (48) is valid [17]. $\square$

Note that the above result for $f = -f_{MI}$ has been shown in [16, Th. 3] by using the chain rule of MI. We have showed here that it can be proved directly based on the analytical form of the function f_{MI} . Moreover, the above result for $f = f_e$ is quite new.

We now also develop a simple but very efficient path-following discrete optimization algorithm that explores a simple structure of the discrete constraint $\mathbf{x} \in \mathcal{D}_S$ to yield the global optimal solution of (25). The following result is important for our development.

Lemma 2: The set $\mathcal{D}_S$ defined from (5) is the set of vertices of the set $\text{Poly}(\mathcal{D}_S)$ defined from (28).

Proof. For $\mathbf{x} \in \mathcal{D}_S$ define $J(\mathbf{x}) = \{k_1 < k_2 < \dots < k_S | x_{k_j} = 1, j = 1, 2, \dots, S\}$. Suppose $\bar{\mathbf{x}} \in \mathcal{D}_S$. It suffices to show that if $\bar{\mathbf{x}} = \mu \mathbf{a} + (1 - \mu) \mathbf{b}$ for $\mathbf{a}, \mathbf{b} \in \text{Poly}(\mathcal{D}_S)$ and $0 < \mu < 1$ then $\mathbf{a} = \mathbf{b} = \bar{\mathbf{x}}$. Indeed, for $i \in J(\bar{\mathbf{x}})$ 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 $i \notin J(\bar{\mathbf{x}})$ 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 $\mathbf{a} = \mathbf{b} = \bar{\mathbf{x}}$ as asserted. $\square$

Recall that point $\mathbf{x}$ is a vertex neighbouring the vertex $\bar{\mathbf{x}}$ if and only if there exists a pair (i, j) such that $x_i = 0 \neq \bar{x}_i = 1$ and $x_j = 1 \neq \bar{x}_j = 0$ and $x_\ell = \bar{x}_\ell$ whenever $\ell \neq i$ and $\ell \neq j$, i.e. $\bar{\mathbf{x}}$ and $\mathbf{x}$ are exactly different in two entries and there are $S(N - S)$ neighbouring vertices for each vertex $\bar{\mathbf{x}}$. Then $\bar{\mathbf{x}} \in \mathcal{D}_S$ is a minimizer of f over $\text{Poly}(\mathcal{D}_S)$ if and only if $f(\bar{\mathbf{x}}) \leq f(\mathbf{v})$ for every $\mathbf{v} \in \mathcal{D}_S$ neighbouring $\bar{\mathbf{x}}$. We thus

propose Algorithm 2 for solving (25), which looks like the Dantzig simplex method for linear programming.¹

Algorithm 2 Path-Following Discrete Optimization Algorithm

Initialization. Start from a $\mathbf{x}^{(0)} \in \mathcal{D}_S$. Set $\kappa = 0$.

κ -th iteration. If there is a $\bar{\mathbf{x}} \in \mathcal{D}_S$ neighbouring $\mathbf{x}^{(\kappa)}$ such that $f(\bar{\mathbf{x}}) < f(\mathbf{x}^{(\kappa)})$ then reset $\kappa + 1 \rightarrow \kappa$ and $\bar{\mathbf{x}} \rightarrow \mathbf{x}^{(\kappa)}$. Otherwise, if $f(\mathbf{x}) \geq f(\mathbf{x}^{(\kappa)})$ for all $\mathbf{x} \in \mathcal{D}_S$ neighbouring $\mathbf{x}^{(\kappa)}$ then stop: $\mathbf{x}^{(\kappa)}$ is the global optimal solution of (25).

Based on this powerful algorithm, we propose Algorithm 3 for solution of the following problem of choosing the minimum number of PMUs to satisfy given tolerance of MSE or MI:

$$\min_{\mathbf{x}} \sum_{k \in \mathcal{N}} x_k : \mathbf{x} \in \{0, 1\}^N, f(\mathbf{x}) \leq \epsilon. \quad (52)$$

Algorithm 3 Iterative Procedure

Initialization. Start from $1 < S_0 < N$ and use Algorithm 2 to find the optimal solution $\mathbf{x}^{(0)}$ of (25) for $S = S_0$.

Until $f(\mathbf{x}^{(\kappa)}) \leq \epsilon$ but $f(\mathbf{x}^{(\kappa-1)}) > \epsilon$: **Reset** $S \rightarrow S - 1$ if $f(\mathbf{x}_{opt}) < \epsilon$ and $S \rightarrow S + 1$ if $f(\mathbf{x}_{opt}) > \epsilon$; **Set** $\kappa := \kappa + 1$.

V. SIMULATION RESULTS

In the simulation, the real power injections P are normally distributed and independent across different buses [28]. Similarly to the simulation setup in [16], the mean vector of real power injection $u_p = (u_p(1), \dots, u_p(N))^T$ is obtained by properly scaling the power profiles in [47], while the diagonal entries of power injection covariance matrix are assumed to be 10% of the mean values, i.e. Σ_P is a diagonal matrix with diagonal entries $\Sigma_P(k, k) = 0.1 u_p(k)$. 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 i7-7600 processor. Sedumi [48] interfaced by CVX is used to solve the convex optimization problems (41) and (42). The commonly used benchmark power networks IEEE 30-bus, IEEE 39-bus, IEEE 57-bus, IEEE 118-bus, IEEE 300-bus, European 1354-bus and Polish 2383-bus with their structure and susceptance matrix obtained from Matpower [47] are tested. The unit of MSE is dB while that of MI is bit.

The following scenarios are considered for simulations:

- S-1 is with the complete observability constraint (CO) (16);
- S-2 is with the constraint (18) of ZIB's presence, which helps to reduce the number of placed PMUs for system observability;

¹Dantzig simplex method is one of the 20th century's top ten algorithms [46] although its polynomial complexity cannot be proved (in contrast to the polynomial complexity of the interior points methods for linear programming). Conceptually, it 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.

TABLE I
THE MINIMUM NUMBER OF PLACED PMUS
UNDER DIFFERENT SCENARIOS

Case	$ \mathcal{L} / \mathcal{Z} $	S-1	S-2	S-3
30	41/6	10	6	6
39	46/10	13	9	7
57	80/15	17	11	9
118	186/10	32	28	25
300	411/65	87	68	66
1354	1991/421	397	271	268
2383	2896/552	746	553	553

TABLE II
MSE (IN dB) AND MI (IN BITS) UNDER THE NUMBER S OF PMUS

Case	S-1	S-2	S-3
30	10/-13.3/7.136	6/-11.61/6.921	6/-11.61/6.921
39	13/-13.7/3.628	9/-12.9/3.407	7/-12.19/3.281
57	17/-10.7/7.558	11/-9.6/7.325	9/-9.15/7.175
118	32/-8.1/8.488	28/-7.8/8.243	25/-7.41/8.186
300	87/-2.6/42.195	68/-2.1/39.381	-
1354	397/-4.3/619.21	271/-2.8/587.43	-
2383	746/0.1/2134.45	-	-

- S-3 is with the both constraints (18a) and (19) of ZIB's presence and PFMs, which help to reduce the number of PMUs for system observability.

A. MSE and MI Performance

Table I provides some basic parameters for the networks and also the minimum of PMUs needed to guarantee different observability constraints in (22), which is obtained by using CPLEX solver [37] for solving the binary linear problem (23) or (24). In Table I, the first two columns provide the network name and its total number of branches ($|\mathcal{L}|$) and total number of ZIBs ($|\mathcal{Z}|$). The last three columns show the minimum number of PMUs obtained by solving (23) under S-1, and solving (24) under S-2 and S-3. It can be seen that, the required number of PMUs to make the system observable decreases significantly by integrating ZIBs and PFMs.

In Table II, the second column, the third column, and the fourth column provide the trio of the number of placed PMUs (S), which is the same as it is in Table I, MSE and MI that are found by Algorithm 1 under S-1, S-2, and S-3, respectively. ZIBs and PFMs in S-3 thus help in reducing the number of placed PMUs while maintaining a good MSE/MI.

In Table III, the second column and the third column of provide a similar trio that are found by Algorithm 1 for solving (22) under S-1, and S-2, respectively plus the additional contingency-constraint (20). We need to use more placed PMUs to compensate the contingency-constraint (20) and as a result they also help to improve the MSE/MI.

Fig. 2 shows the required number of placed PMUs to satisfy both (18a) to exploit ZIBs and (21) to limit the number of communication channels per PMU. Table IV provides the corresponding MSE. It should be mentioned that if the maximum number of channels per PMU is 1 then the required number of PMUs is $|\mathcal{N}| - |\mathcal{Z}|$. By allowing only one communication

TABLE III
MSE (IN dB) AND MI (IN BITS) WITH THE ADDITIONAL
CONTINGENCY-CONSTRAINT (20)

Case	S/MSE/MI under S-1	S/MSE/MI under S-2
30	21/-15.6/8.446	16/-14.6/8.073
39	28/-16.1/4.561	21/-15.4/4.315
57	33/-12.4/10.139	26/-12.1/9.732
118	68/-10.4/13.262	63/-10.3/12.973

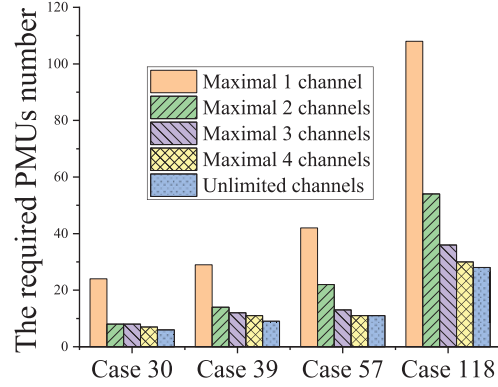


Fig. 2. The required number of placed PMUs with different limitation values of communication channels per PMU.

TABLE IV
THE REQUIRED NUMBER OF PLACED PMUS/MSE (IN dB) UNDER
DIFFERENT NUMBERS OF CHANNELS PER PMU

Case	# of channels per PMU				No channel limit
	1	2	3	4	
30	24/-15.9	8/-12.5	8/-12.5	7/-12.0	6/-11.6
39	29/-16.1	14/-14.0	12/-13.9	11/-13.2	9/-12.9
57	42/-12.8	22/-11.3	13/-10.1	11/-9.6	11/-9.6
118	108/-11.5	54/-9.8	36/-8.6	30/-8.1	28/-8.1

channel per PMU, the required number of placed PMUs is almost the same as the number of buses. By allowing two communication channels per PMU, this number is almost reduced to half. Allowing the number of communication channels per PMU more than three seems to be not so efficient as it does not help to reduce the number of placed PMUs while resulting in worse MSEs.

Fig.3 depicts the achieved MSE by using different algorithms under different scenarios versus the number of placed PMUs. The curve "S-1 Alg. 1" and "S-2 Alg. 1" are the theoretical MSE by solving problem (41) for scenarios S-1 and S-2, respectively. The MSE of "S-2 Alg. 1" is better than that of "S-1 Alg. 1" thanks to the impact of ZIBs. The curve "Monte-Carlo" is MSE obtained through Monte-Carlo simulation of scenario S-1. The curve "S-1 CPLEX" is MSE obtained by solving the feasibility problem of (23) by CPLEX. MSE obtained by the former is much better than that obtained by the latter. The MSE in the curve "S-1 CPLEX" is not monotonous, because its MSE is not optimal. The curve "LB by Alg. 2" provides a lower bound for the MSE in OCA PMU placement problem (22) by solving problem (25) by Algorithm 2. Similarly, the curves in Fig. 4 plot the MI, where the curve "UB by Alg. 2" provides an upper bound for the MI in OCA PMU placement problem (22) by solving problem (25)

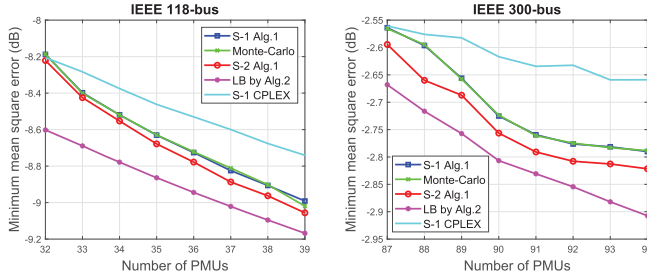


Fig. 3. MSE by different methods.

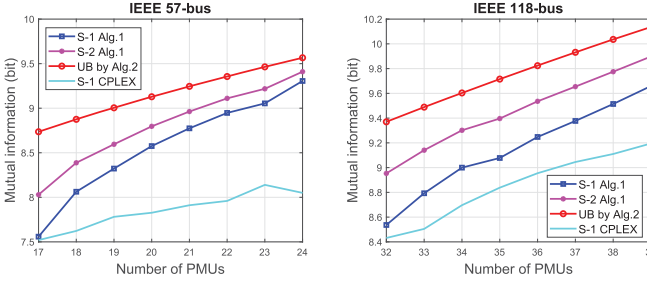


Fig. 4. MI by different methods.

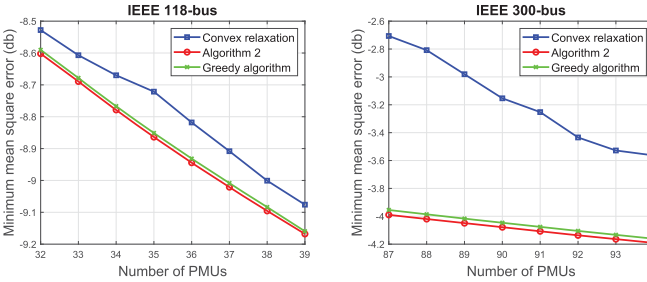


Fig. 5. MSE value of problem (25) computed by different algorithms.

by Algorithm 2. The capability and efficiency of Algorithm 1 and Algorithm 2 to obtain informative PMU placements are quite clear.

Fig. 5 and Fig. 6 provide the values of MSE and MI found by Algorithm 2, the greedy algorithm by solving (46) and convex relaxation used in [15]. Algorithm 2 and the greedy algorithm clearly outperform the convex relaxation method and the performance gaps becomes wider with the increase of network size. Observe that Algorithm 2 and the greedy algorithm perform similarly. The former provides the global optimal value for (25) while the latter guarantees a suboptimal value at least with the ratio $1 - 1/e$ according to Theorem 1. It is reasonable to expect that locating the global optimal value costs by the former more time due to the confirmation of its global optimality than that for locating a suboptimal value by the latter.

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

B. Computation Experience

The numerical details of Algorithm 1 for scenario S-1 and S-2 are summarized in Table V with the numerical value of μ

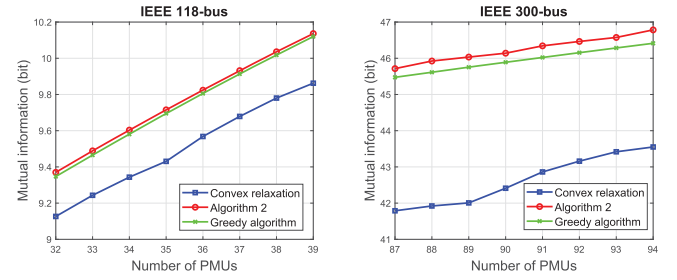


Fig. 6. MI value of problem (25) computed by different algorithms.

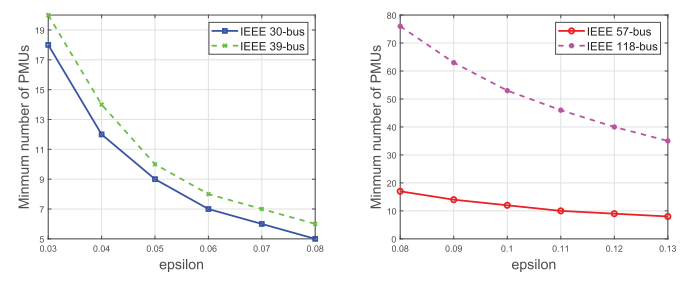
Fig. 7. Minimum number of placed PMUs required versus different values of tolerance level ϵ for MSE by Algorithm 3.

TABLE V
NUMERICAL DETAILS OF ALGORITHM 1 UNDER SCENARIO S-1 AND S-2

Case	μ		CPU (s) for S-1		CPU (s) for S-2		
	MSE	MI	MSE	MI	# y	MSE	MI
30	0.1	1	4.10	6.32	25	6.17	9.32
39	0.1	1	9.47	14.41	42	13.87	19.74
57	1	10	32.20	54.73	55	52.54	81.33
118	1	10	57.23	98.72	44	58.43	92.34
300	1	10	164.53	319.24	270	540.23	810.89
1354	10	100	473.89	715.88	1557	1643.27	2122.43
2383	10	100	1245.31	1656.74	-	-	-

in implementing Algorithm 1 given by the second and third column. The number of actual decision binary variables $y_{k,m}$ in (18) to express of the impact of ZIBs in S-2 is also provided in the sixth column. Note that this number is much smaller than $N|Z|$ because if nodes k and m are not connected, then there is no involvement of y_{km} as automatically it is zero. The CPU time of Algorithm 1 increases moderately with the increase of grid size, demonstrating its scalability. To speed up its convergence, at each iteration we check the state of variables x , y , and c to fix those, which attain binary values.

The computational complexity of Algorithm 1 is determined by (45), which is the computational complexity of the convex problem (41)/(42) solved at each iteration, and the number of iterations for its convergence. Fig. 8 provides the number of iterations that Algorithm 1 needs for computing (34) for IEEE 30-bus network, 39-bus network, 57-bus network and 118-bus network under different scenarios.

Table VI provides the CPU for implementing Algorithm 2, the greedy algorithm (46) and convex relaxation using (26)-(27). The CPU time of convex relaxation for the IEEE 300-bus is about five hours and seven hours and half due to the involvement of additional $300 \times 301/2 = 45.150$ variables. It should be mentioned that, all the proposed

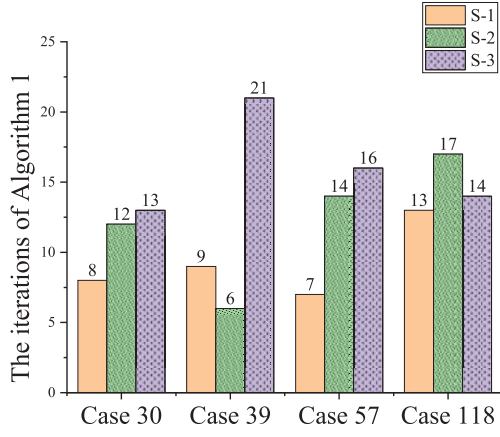


Fig. 8. Average number of iterations that Algorithm 1 needs in computing (34).

TABLE VI

COMPARISON OF CPU TIME BETWEEN ALGORITHM 2, THE GREEDY ALGORITHM (GA) (46) AND CONVEX RELAXATION (CR) (26)–(27)

Case	CPU (s) of Alg. 2		CPU (s) of GA		CPU (s) of CR	
	MSE	MI	MSE	MI	MSE	MI
30	0.30	0.57	0.03	0.10	1.85	3.46
39	0.56	1.24	0.05	0.17	3.63	5.29
57	2.88	5.23	0.12	0.38	12.34	18.52
118	45.21	95.47	0.84	1.51	1053.26	1547.65
300	537.56	941.45	62.38	87.91	17267.45	26833.46

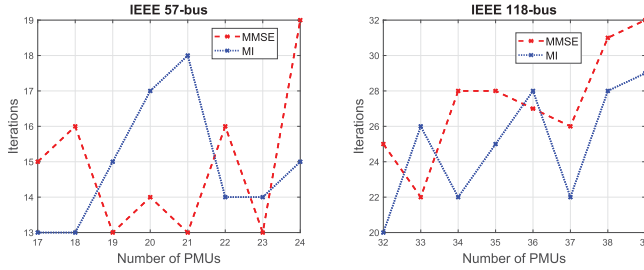


Fig. 9. Number of iterations required for the convergence of Algorithm 2.

algorithms experience difficulty for calculating matrix inverse (for $f_e(\mathbf{x})$) and determinant (for $f_{MI}(\mathbf{x})$), especially in large-scale networks.

For IEEE 57-bus network and IEEE 118-bus network, Fig. 9 presents the number of iterations needed for the convergence of Algorithm 2 for MSE and MI, respectively. The computational complexity of each iteration is $\mathcal{O}(S(N-S))$.

VI. CONCLUSION

The paper has 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 constraints to guarantee the estimation quality, which is formulated as a large scale binary nonlinear optimization problem. The paper has developed scalable algorithms for their computation, which result at least in locally optimal solutions. The paper has also developed extremely efficient algorithms of very low computational complexity for cases of

absent observability. The viability of the developed algorithms has been confirmed through comprehensive simulations with benchmark grids.

APPENDIX: FUNDAMENTAL INEQUALITIES

For $A_0 \succ 0$ and $A_k \succeq 0$, $k = 1, \dots, N$ let $\Phi(\mathbf{x}) := (A_0 + \sum_{k=1}^N \frac{1}{x_k} A_k)^{-1}$, and $\Psi(\mathbf{x}) := (A_0 + \sum_{k=1}^N x_k A_k)^{-1}$. The first result is

Theorem 2: The following inequality holds true for all $\mathbf{x} \in \text{int}(\mathbb{R}_+^N)$ and $\bar{\mathbf{x}} \in \text{int}(\mathbb{R}_+^N)$:

$$\begin{aligned} & \text{Trace}(\Psi(\mathbf{x})) \\ & \leq \text{Trace}((\Psi(\bar{\mathbf{x}}))^2 A_0) + \sum_{k=1}^N \frac{\bar{x}_k^2}{x_k} \text{Trace}((\Psi(\bar{\mathbf{x}}))^2 A_k). \end{aligned} \quad (53)$$

Proof. By [26, Th.1], the function $\varphi(\mathbf{x}) = \text{Trace}(\Phi(\mathbf{x}))$ is concave in the domain $\text{int}(\mathbb{R}_+^N)$, so for all $\mathbf{x} \in \text{int}(\mathbb{R}_+^N)$ and $\bar{\mathbf{x}} \in \text{int}(\mathbb{R}_+^N)$ one has

$$\begin{aligned} \varphi(\mathbf{x}) & \leq \varphi(\bar{\mathbf{x}}) + \langle \nabla \varphi(\bar{\mathbf{x}}), \mathbf{x} - \bar{\mathbf{x}} \rangle \\ & = \text{Trace}(\Phi^2(\bar{\mathbf{x}}) A_0) + \sum_{k=1}^N \frac{x_k}{\bar{x}_k^2} \text{Trace}(\Phi^2(\bar{\mathbf{x}}) A_k). \end{aligned} \quad (54)$$

Then the inequality (53) is obtained by replace $x_k \rightarrow 1/x_k$ and $\bar{x}_k \rightarrow 1/\bar{x}_k$, $k = 1, \dots, N$, in (54). $\square$

The next result is

Theorem 3: The following inequality holds true for all $\mathbf{x} \in \text{int}(\mathbb{R}_+^N)$ and $\bar{\mathbf{x}} \in \text{int}(\mathbb{R}_+^N)$:

$$\begin{aligned} -\ln |\Psi(\mathbf{x})| & \geq -\ln |\Psi(\bar{\mathbf{x}})| \\ & + \text{Trace} \left(\Psi(\bar{\mathbf{x}}) \left(\sum_{k=1}^N \bar{x}_k A_k \right) \right) - \sum_{k=1}^N \frac{\bar{x}_k^2}{x_k} \text{Trace}(\Psi(\bar{\mathbf{x}}) A_k). \end{aligned} \quad (55)$$

Proof. By [49, Th. 2], the function $\phi(\mathbf{x}) = -\ln |\Phi(\mathbf{x})|$ is convex in the domain $\text{int}(\mathbb{R}_+^N)$, so for all $\mathbf{x} \in \text{int}(\mathbb{R}_+^N)$ and $\bar{\mathbf{x}} \in \text{int}(\mathbb{R}_+^N)$ one has

$$\begin{aligned} \phi(\mathbf{x}) & \geq \phi(\bar{\mathbf{x}}) + \langle \nabla \phi(\bar{\mathbf{x}}), \mathbf{x} - \bar{\mathbf{x}} \rangle \\ & = -\ln |\Phi(\bar{\mathbf{x}})| + \text{Trace} \left((\Phi(\bar{\mathbf{x}}))^{-1} \left(\sum_{k=1}^N \frac{1}{\bar{x}_k} A_k \right) \right) \\ & \quad - \sum_{k=1}^N \frac{x_k}{\bar{x}_k^2} \text{Trace}((\Phi(\bar{\mathbf{x}}))^{-1} A_k). \end{aligned} \quad (56)$$

Then the inequality (55) is obtained by replace $x_k \rightarrow 1/x_k$ and $\bar{x}_k \rightarrow 1/\bar{x}_k$, $k = 1, \dots, N$, in (56). $\square$

The last result is

Theorem 4: The following inequality holds true for all $X \succeq Y \succ 0$ and $A \succeq 0$

$$\begin{aligned} & \text{Trace}(X^{-1}) - \text{Trace}((X+A)^{-1}) \\ & \leq \text{Trace}(Y^{-1}) - \text{Trace}((Y+A)^{-1}). \end{aligned} \quad (57)$$

Proof. By the mean value theorem, there is $\xi \in (0, 1)$ such that for $Z = \xi X + (1-\xi)Y \succ 0$,

$$\begin{aligned} & (\text{Trace}(X^{-1}) - \text{Trace}((X+A)^{-1})) \\ & \quad - (\text{Trace}(Y^{-1}) - \text{Trace}((Y+A)^{-1})) \end{aligned}$$

$$\begin{aligned}
&= \text{Trace}(((Z + A)^{-2} - Z^{-2})(X - Y)) \\
&= \text{Trace}(((X - Y)^{1/2}(Z + A)^{-1}(X - Y)^{1/2})^2 \\
&\quad - ((X - Y)^{1/2}Z^{-1}(X - Y)^{1/2})^2) \\
&= \text{Trace}(((X - Y)^{1/2}(Z + A)^{-1}(X - Y)^{1/2} \\
&\quad - (X - Y)^{1/2}Z^{-1}(X - Y)^{1/2}) \\
&\quad \times ((X - Y)^{1/2}(Z + A)^{-1}(X - Y)^{1/2} \\
&\quad + (X - Y)^{1/2}Z^{-1}(X - Y)^{1/2})) \leq 0,
\end{aligned}$$

because

$$\begin{aligned}
&(X - Y)^{1/2}(Z + A)^{-1}(X - Y)^{1/2} - (X - Y)^{1/2}Z^{-1}(X - Y)^{1/2} \\
&= (X - Y)^{1/2}[(Z + A)^{-1} - Z^{-1}](X - Y)^{1/2} \preceq 0,
\end{aligned}$$

due to $(Z + A)^{-1} \preceq Z^{-1}$, and

$$\begin{aligned}
&(X - Y)^{1/2}(Z + A)^{-1}(X - Y)^{1/2} + (X - Y)^{1/2}Z^{-1}(X - Y)^{1/2} \\
&= (X - Y)^{1/2}[(Z + A)^{-1} + Z^{-1}](X - Y)^{1/2} \succeq 0
\end{aligned}$$

due to $(Z + A)^{-1} + Z^{-1} \succeq 0$. $\square$

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, Jun. 2010.
- [2] A. G. Phadke and J. S. Thorp, *Synchronized Phasor Measurements and Their Applications*. New York, NY, USA: Springer, 2008.
- [3] S. Cui, Z. Han, S. Kar, T. T. Kim, H. V. Poor, and A. Tajer, "Coordinated data-injection attack and detection in the smart grid: A detailed look at enriching detection solutions," *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 106–115, Sep. 2012.
- [4] J. Zhao *et al.*, "Power system real-time monitoring by using PMU-based robust state estimation method," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 300–309, Jan. 2016.
- [5] C.-W. Ten, A. Ginter, and R. Bulbul, "Cyber-based contingency analysis," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 3040–3050, Jul. 2016.
- [6] R. F. Nuqui and A. G. Phadke, "Phasor measurement unit placement techniques for complete and incomplete observability," *IEEE Trans. Power Del.*, vol. 20, no. 4, pp. 2381–2388, Oct. 2005.
- [7] B. Gou, "Generalized integer linear programming formulation for optimal PMU placement," *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 1099–1104, Aug. 2008.
- [8] F. Aminifar, A. Khodaei, M. Fotuhi-Firuzabad, and M. Shahidehpour, "Contingency-constrained PMU placement in power networks," *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 516–523, Feb. 2010.
- [9] J. Aghaei, A. Baharvandi, A. Rabiee, and M. A. Akbari, "Probabilistic PMU placement in electric power networks: An MILP-based multiobjective model," *IEEE Trans. Ind. Informat.*, vol. 11, no. 2, pp. 332–341, Apr. 2015.
- [10] S. Chakrabarti and E. Kyriakides, "Optimal placement of phasor measurement units for power system observability," *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 1433–1440, Aug. 2008.
- [11] M. Hajian, A. M. Ranjbar, T. Amraee, and B. Mozafari, "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.
- [12] S. Chakrabarti, E. Kyriakides, and D. G. Eliades, "Placement of synchronized measurements for power system observability," *IEEE Trans. Power Del.*, vol. 24, no. 1, pp. 12–19, Jan. 2009.
- [13] K. G. Khajeh, E. Bashar, A. M. Rad, and G. B. Ghahreghpetian, "Integrated model considering effects of zero injection buses and conventional measurements on optimal PMU placement," *IEEE Trans. Smart Grid*, vol. 8, no. 2, pp. 1006–1013, Mar. 2017.
- [14] A. Monticelli, "Electric power system state estimation," *Proc. IEEE*, vol. 88, no. 2, pp. 262–282, Feb. 2000.
- [15] V. Kekatos, G. B. Giannakis, and B. Wollenberg, "Optimal placement of phasor measurement units via convex relaxation," *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1521–1530, Aug. 2012.
- [16] Q. Li, T. Cui, Y. Weng, R. Negi, F. Franchetti, and M. D. Ilic, "An information-theoretic approach to PMU placement in electric power systems," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 446–456, Mar. 2013.
- [17] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions-I," *Math. Program.*, vol. 14, no. 1, pp. 265–294, 1978.
- [18] M. J. Rice and G. T. Heydt, "Power systems state estimation accuracy enhancement through the use of PMU measurements," in *Proc. IEEE PES Transmiss. Distrib. Conf. Expo.*, May 2006, pp. 161–165.
- [19] R. A. Sevlian, Y. Zhao, R. Rajagopal, A. Goldsmith, and H. V. Poor, "Outage detection using load and line flow measurements in power distribution systems," *IEEE Trans. Power Syst.*, vol. 33, pp. 2053–2069, Mar. 2018.
- [20] U.S. Department of Energy, Office of Electricity Delivery and Energy Reliability. (Oct. 2014). *Factors Affecting PMU Installation Costs*. [Online]. Available: https://www.smartgrid.gov/files/PMU-cost-study-final-10162014_1.pdf
- [21] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Trans. Inf. Syst. Secur.*, vol. 14, no. 1, 2011, Art. no. 13.
- [22] Y. Zhao, A. Goldsmith, and H. V. Poor, "Minimum sparsity of unobservable power network attacks," *IEEE Trans. Automat. Control*, vol. 62, no. 7, pp. 3354–3368, Jul. 2017.
- [23] E. Dahlman *et al.*, "5G wireless access: Requirements and realization," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 42–47, Dec. 2014.
- [24] Y. Selen, H. Tullberg, and J. Kronander, "Sensor selection for cooperative spectrum sensing," in *Proc. 3rd IEEE Symp. New Frontiers Dyn. Spectr. Access Netw.*, Oct. 2008, pp. 1–11.
- [25] U. Rashid, H. D. Tuan, P. Apkarian, and H. H. Kha, "Globally optimized power allocation in multiple sensor fusion for linear and nonlinear networks," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 903–915, Feb. 2012.
- [26] J. A. Bengua, H. D. Tuan, T. Q. Duong, and H. V. Poor, "Joint sensor and relay power control in tracking Gaussian mixture targets by wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 66, no. 2, pp. 492–506, Sep. 2018.
- [27] G. Krumpholz, K. Clements, and P. Davis, "Power system observability: A practical algorithm using network topology," *IEEE Trans. Power App. Syst.*, vol. PAS-99, no. 4, pp. 1534–1542, Jul. 1980.
- [28] A. Schellenberg, W. Rosehart, and J. Aguado, "Cumulant-based probabilistic optimal power flow (P-OPF) with Gaussian and gamma distributions," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 773–781, May 2005.
- [29] J. Grainger and J. W. Stevenson, *Power System Analysis*. New York, NY, USA: McGraw-Hill, 1994.
- [30] S. Kay, *Fundamentals Statistical Signal Processing: Estimation*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.
- [31] H. D. Tuan, D. H. Pham, B. Vo, and T. Q. Nguyen, "Entropy of general Gaussian distributions and MIMO channel capacity maximizing precoder and decoder," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process (ICASSP)*, May 2007, pp. III-325–III-328.
- [32] R. Kavasseri and S. K. Srinivasan, "Joint placement of phasor and power flow measurements for observability of power systems," *IEEE Trans. Power Syst.*, vol. 26, no. 4, pp. 1929–1936, Nov. 2011.
- [33] M. Göl and A. Abur, "Observability and criticality analyses for power systems measured by phasor measurements," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3319–3326, Aug. 2013.
- [34] M. Esmaili, K. Gharani, and H. A. Shayanfar, "Redundant observability PMU placement in the presence of flow measurements considering contingencies," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 3765–3773, Nov. 2013.
- [35] S. Azizi, G. B. Ghahreghpetian, and A. Salehi-Dobakhshari, "Optimal integration of phasor measurement units in power systems considering conventional measurements," *IEEE Trans. Smart Grid*, vol. 4, no. 2, pp. 1113–1121, Jun. 2013.
- [36] H. D. Tuan, T. T. Son, H. Tuy, and H. H. Nguyen, "Optimum multi-user detection by nonsmooth optimization," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, May 2011, pp. 3444–3447.
- [37] *CPLEX Optimizer*. Accessed: Feb. 14, 2018. [Online]. Available: <https://www.ibm.com/analytics/data-science/prescriptive-analytics/cplex-optimizer>
- [38] A. Gomez-Exposito and A. Abur, *Power System State Estimation: Theory and Implementation*. Boca Raton, FL, USA: CRC Press, 2004.

- [39] S. A. Pasha, H. D. Tuan, and B. N. Vo, "Nonlinear Bayesian filtering using the unscented linear fractional transformation model," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 477–489, Feb. 2010.
- [40] H. Tuy, *Convex Analysis and Global Optimization*, 2nd ed. New York, NY, USA: Springer, 2016.
- [41] E. Che, H. D. Tuan, and H. H. Nguyen, "Joint optimization of cooperative beamforming and relay assignment in multi-user wireless relay networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 10, pp. 5481–5495, Oct. 2014.
- [42] H. H. M. Tam, H. D. Tuan, D. T. Ngo, T. Q. Duong, and H. V. Poor, "Joint load balancing and interference management for small-cell heterogeneous networks with limited backhaul capacity," *IEEE Trans. Wireless Commun.*, vol. 16, no. 2, pp. 872–884, Feb. 2017.
- [43] A. H. Phan, H. D. Tuan, H. H. Kha, and D. T. Ngo, "Non-smooth optimization for efficient beamforming in cognitive radio multicast transmission," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2941–2951, Jun. 2012.
- [44] Y. Shi, H. D. Tuan, S. Su, and H. Tuy, "Global optimization for optimal power flow over transmission networks," *J. Global Optim.*, vol. 69, no. 3, pp. 745–760, 2017.
- [45] Y. Shi, H. D. Tuan, A. V. Savkin, T. Q. Duong, and H. V. Poor, "Model predictive control for smart grids with multiple electric-vehicle charging stations," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 2127–2136, Mar. 2019.
- [46] B. A. Cipra, "The best of the 20th century: Editors name top 10 algorithms," *SIAM News*, vol. 33, pp. 1–2, Dec. 2000.
- [47] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 12–19, Feb. 2011.
- [48] J. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optim. Methods Softw.*, vols. 11–12, nos. 1–4, pp. 625–653, 1999.
- [49] L. D. Nguyen, H. D. Tuan, T. Q. Duong, and H. V. Poor, "Multi-user regularized zero-forcing beamforming," *IEEE Trans. Signal Process.*, vol. 67, no. 11, pp. 2839–2853, Jun. 2019.



Ye Shi received the Ph.D. degree in electrical engineering from the University of Technology Sydney, Australia, in 2018. He is currently a Post-Doctoral Fellow with the School of Computer Science, University of Technology Sydney. His research interests include nonconvex optimization in smart grid, robust control, and fuzzy systems. He received the Best Paper Award at the 6th IEEE International Conference on Control Systems, Computing and Engineering in 2016.



Hoang Duong Tuan received the Diploma (Hons.) and Ph.D. degrees in applied mathematics from Odessa State University, Ukraine, in 1987 and 1991, respectively. He was as an Assistant Professor with the Department of Electronic-Mechanical Engineering, Nagoya University, from 1994 to 1999, and then as an Associate Professor with the Department of Electrical and Computer Engineering, Toyota Technological Institute, Nagoya, from 1999 to 2003. He was a Professor with the School of Electrical Engineering and Telecommunications, University of New South Wales, from 2003 to 2011. He is currently a Professor with the School of Electrical and Data Engineering, University of Technology Sydney. He has been involved in research with the areas of optimization, control, signal processing, wireless communication, and biomedical engineering for more than 20 years.



Trung Q. Duong (S'05–M'12–SM'13) received the Ph.D. degree in telecommunications systems from the Blekinge Institute of Technology (BTH), Sweden, in 2012. He is currently with Queen's University Belfast, U.K., where he was a Lecturer (Assistant Professor) from 2013 to 2017 and has been a Reader (Associate Professor) since 2018. He is the author or coauthor of over 340 technical articles published in scientific journals (over 200 articles) and presented at international conferences (over 140 articles). His current research interests include the Internet of Things (IoT), wireless communications, molecular communications, and signal processing.

Dr. Duong received the Best Paper Award at the IEEE Vehicular Technology Conference (VTC-Spring) in 2013, IEEE International Conference on Communications (ICC) 2014, IEEE Global Communications Conference (GLOBECOM) 2016, IEEE Digital Signal Processing Conference (DSP) 2017, and International Wireless Communications and Mobile Computing Conference (IWCMC) 2019. He was a recipient of prestigious Royal Academy of Engineering Research Fellowship (2016–2021) and a prestigious Newton Prize 2017. He is currently an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON COMMUNICATIONS, and IET COMMUNICATIONS, and a Lead Senior Editor for the IEEE COMMUNICATIONS LETTERS.



H. Vincent Poor (S'72–M'77–SM'82–F'87) received the Ph.D. degree in EECS from Princeton University, Princeton, NJ, USA, in 1977. From 1977 to 1990, he was on the faculty of the University of Illinois at Urbana-Champaign. Since 1990, he has been on the faculty at Princeton University, where he is currently the Michael Henry Strater University Professor of Electrical Engineering. From 2006 to 2016, he was the Dean of Princeton's School of Engineering and Applied Science. He has also held visiting appointments at several other universities, including most recently at Berkeley and Cambridge. His research interests are in the areas of information theory and signal processing, and their applications in wireless networks, energy systems, and related fields. Among his publications in these areas is the recent book *Multiple Access Techniques for 5G Wireless Networks and Beyond* (Springer, 2019).

Dr. Poor is a member of the National Academy of Engineering and the National Academy of Sciences, and is a Foreign Member of the Chinese Academy of Sciences, the Royal Society, and other national and international academies. He received the Marconi and Armstrong Awards of the IEEE Communications Society in 2007 and 2009, respectively. Recent recognition of his work includes the 2017 IEEE Alexander Graham Bell Medal, the 2019 ASEE Benjamin Garver Lamme Award, a D.Sc. *honoris causa* from Syracuse University in 2017, and a D.Eng. *honoris causa* from the University of Waterloo in 2019.



Andrey V. Savkin received the M.S. and Ph.D. degrees from Leningrad State University in 1987 and 1991, respectively. He has been a Full Professor with the School of Electrical Engineering and Telecommunications, The University of New South Wales, Sydney, Australia, since 2000. He has authored or coauthored seven research monograph (published by Springer, Birkhauser, IEEE Press - Wiley, and Elsevier) and about 200 journal articles. He served as an Associate and/or Guest Editor for several international journals and numerous international conferences in the field. His current research interests include the applications of control theory to robotics, power systems, sensor networks, and biomedical engineering.