Power Systems Topology and State Estimation by Graph Blind Source Separation

Sivan Grotas, Student Member, IEEE, Yair Yakoby, Idan Gera, and Tirza Routtenberg, Senior Member, IEEE

Abstract—In this paper, we consider the problem of blind estimation of states and topology (BEST) in power systems. We use the linearized dc model of real power measurements with unknown voltage phases (i.e., states) and an unknown admittance matrix (i.e., topology) and show that the BEST problem can be formulated as a blind source separation (BSS) problem with a weighted Laplacian mixing matrix. We develop the constrained maximum likelihood (ML) estimator of the Laplacian matrix for this graph BSS problem with Gaussian-distributed states. The ML-BEST is shown to be only a function of the states’ second-order statistics. Since the topology recovery stage of the ML-BEST approach results in a high-complexity optimization problem, we propose two low-complexity methods to implement it: First, two-phase topology recovery, which is based on solving the relaxed convex optimization and then finding the closest Laplacian matrix, and second, augmented Lagrangian topology recovery. We derive a closed-form expression for the associated Cramér–Rao bound (CRB) on the topology matrix estimation. The performance of the proposed methods is evaluated for three case studies: the IEEE-14 bus system, the IEEE 118-bus system, and a random network, and compared with the oracle minimum mean-squared-error state estimator and with the proposed CRB.

Index Terms—Graph blind source separation (GBSS), constrained maximum likelihood, Laplacian mixing matrix, Topology identification, power system state estimation.

I. INTRODUCTION

STATE estimation is a critical component of modern energy management systems (EMSs) for multiple monitoring purposes, including analysis, security, control, situational awareness, stability assessment, power market design, and optimization of electricity dispatchment [1], [2]. In the DC model, the states are the bus voltage angles, while the grid topology includes the arrangement of loads or generators, transmission lines, transformers, and the statuses of system devices. It should be noted that this definition generalizes the computer science graph theory definition, which refers to the connectivity of the graph, since here the topology also includes the weights. In current systems it is assumed that the EMS has precise knowledge of the grid topology [1], which is used for obtaining accurate state estimation. However, knowledge of grid topology may not be available and it may change over time due to failure, opening and closing of switches on power lines, and the presence of new loads and generators. The topology data may also be incorrect due to malicious topology attacks [3]–[6]. Thus, methods for state estimation that are not based on a known topology are crucial for obtaining a reliable system model and high power quality. An additional use for topology identification is event detection, such as identifying faults, line outages, and system imbalances [7]–[9]. Moreover, it can be used to secure the system from potential cyberattacks on the topology information and to identify the potential vulnerabilities of a power system.

Several approaches to topology identification have been proposed in the literature. Detecting topological changes has been studied in [10], [11] and the conditions for the detectability of topology errors are studied in [12]. Different approaches have been suggested for topology reconstruction, such as data-driven approaches using the historical smart meter measurements [13], [14], an active sensing paradigm of grid probing of voltage data [15], and graphical models using voltage or phase measurements [16]–[18]. In particular, the voltage (i.e. state) statistic has been used in various works. For example, a data-based voltage correlation is used in [19]. The works in [20], [21] use properties of the second moments of voltage magnitude measurements to identify the radial topology through iterative algorithms. The statistical properties of voltage measurements are investigated in [20]. Recently, a few papers have addressed blind estimation of the grid topology by observing multiple power injection supervisory control and data acquisition (SCADA) measurements [22], [23], voltage and power data obtained by phasor measurement units (PMUs) [20], [24], voltage measurements and their associated correlations [16], [19], and electricity price based market data [25]. In [26], an unobservable attack is designed based on incomplete knowledge of the system matrix, which is learned via a blind identification approach. The methods proposed in [22], [23], [25], [26] can reveal part of the grid topology, such as the grid connectivity and the eigenvectors of the topology matrix, but they cannot reconstruct the full topology matrix with exact scaling and true eigenvalues. Thus, incorporating blind source separation (BSS) techniques with the specific characteristics of a graph seems promising.

BSS methods aim to restoring a set of unknown source signals from a set of observed linear mixtures of these source signals (see, e.g., [27]–[36]), without prior knowledge of the sources and the mixing system. The problem of BSS has been extensively investigated in the literature in the recent two decades. Prior works on maximum likelihood (ML) separation in BSS deal with general stationary sources [28], [37], autoregressive (AR) sources, and AR Gaussian mixture model distributed sources.
[35], [36]. The ML BSS for nonstationary structures with varying variance-profiles was considered in [38]. However, classical BSS solutions are ambiguous in the sense that the order, signs and scales of the original signals cannot be retrieved. These ambiguities cannot be tolerated in the considered power system problem. In addition, usually the distributions of the states are assumed to be Gaussian due to the central limit theorem, while most BSS methods cannot handle Gaussian sources. Therefore, new methods for BSS are required for the semiblind scenario of a Laplacian mixing matrix with Gaussian sources, without permutation and scaling problems.

In addition to state estimation in power systems, the recent field of graph signal processing (GSP) [39] has many applications [40]–[43]. Two major challenges in GSP are learning the graph structure from data under Laplacian matrix constraints (see, e.g., [44]–[46]) and blind deconvolution of signals on graphs [47], which aim to jointly identify the filter coefficients and the input signal. In future work the approach developed in this paper could be extended to general GSP applications.

In this paper, we consider the problem of state estimation and topology identification in power systems based on active power measurements. First, we show that this problem is equivalent to the problem of BSS with a weighted Laplacian mixing matrix, where the weights are determined by the branch susceptances. Then, we derive the ML blind estimation of states and topology (ML-BEST) method for Gaussian-distributed states, that incorporates the constraints of a Laplacian mixing matrix and is shown to be a second-order statistics (SOS) method. Since the topology recovery stage of the ML-BEST estimator is shown to be a NP-hard optimization problem, we suggest two practical low-complexity methods to implement this stage: (1) Two-phase topology recovery, which is based on solving the relaxed convex optimization and then finding the closest Laplacian matrix, and (2) Augmented Lagrangian topology recovery. Preliminary results can be found in [48]. We also derive a closed-form expression for the Cramér-Rao bound (CRB), which is a useful tool in evaluating the asymptotic performance of low-complexity estimators. Finally, simulations demonstrate that the proposed ML-BEST methods are applicable for different network topologies, and asymptotically achieve the CRB.

The remainder of the paper is organized as follows. In Section II we introduce the system model and the graph BSS (GBSS) problem for state and topology estimation in power systems. The ML-BEST solution is defined and two different practical methods for its topology recovery stage are suggested in Sections III and IV, respectively. Section V offers some remarks, including a parameter identifiability analysis, a complexity discussion, and possible extensions of the proposed model and methods. A closed-form expression for the CRB of the topology matrix is derived in Section VI. The proposed methods are evaluated via simulations in Section VII. Conclusions appear in Section VIII.

II. PROBLEM FORMULATION

In this section, we formulate the problem of estimating the state and topology/admittance matrix in power systems under the linear DC power model. We show that this problem is equivalent to BSS with a Laplacian mixing matrix.

A. Notation

In the rest of this paper vectors are denoted by boldface lowercase letters and matrices by boldface uppercase letters. The $K \times K$ identity matrix is denoted by $I_K$, and $1_K$ denotes the constant $K$-length one vector. The vectors $0$ and $e_m$ are zero vectors (with appropriate dimension), except for the $m$th element of $e_m$, which is 1. Based on these vectors, we define the matrices $E_{i,m} = (e_i e_m)^T$. Additionally, $\delta_{m,k}$ denotes Kronecker’s delta, which equals 1 if $m = k$ and 0 otherwise. The notations $\| \cdot \|$, $\text{Tr}(\cdot)$, and $\otimes$ denote the determinant operator, the trace operator, and the Kronecker product, respectively. For a full-rank matrix $C$, $C^\dagger = (C^TC)^{-1}C^T$ is the Moore-Penrose pseudo-inverse. The $m$th element of the vector $a$, $(m,q)$th element of the matrix $C$, and the $(m_1:m_2 \times q_1 : q_2)$ submatrix of $C$ are denoted by $a_m$, $C_{m,q}$, and $C_{m_1:m_2,q_1:q_2}$, respectively. If $C$ is a positive semidefinite matrix we denote it by $C \succeq 0$ and its square root, $C^{\frac{1}{2}}$, satisfies $C^{\frac{1}{2}}C^{\frac{1}{2}} = C$, where $C^{\frac{1}{2}}$ denotes the inverse of this square root. For any matrix $C$, $||| C |||_F$ and $||| C |||_0$ denote its Frobenius and $\ell_0$-(pseudo)norm (counting its nonzero entries), respectively. Similarly, for any symmetric matrix $S$, $\text{vech}(S)$ is a vector obtained by stacking its columns. Similarly, for any symmetric matrix $S$, $\text{vech}(S)$ is a vector obtained by stacking its columns.

B. Graph Representation of Power Systems

A power system can be represented as an undirected connected weighted graph, $G(V, \xi)$, where the set of vertices, $V = \{1, \ldots, M\}$, is the set of buses (that represent interconnections, generators or loads) and the edge set, $\xi$, is the set of connected transmission lines between the buses. An arbitrary orientation is assigned to each edge $e_i = (m, k) \in \xi$, $m, k = 1, \ldots, M$, $k < m$. $1 \leq i \leq M(M-1)$, that are ordered in a lexicographical order, which connects the vertices $m$ and $k$. The cardinality of the edge set, $|\xi| = \frac{M(M-1)}{2}$, represents all possible connections in the graph. According to the $\pi$-model of transmission lines [1], each line is characterized by the line admittance $Y_{m,k}$, $\forall (m, k) \in \xi$. The incidence matrix of a graph is $A \in \mathbb{R}^{M \times \frac{M(M-1)}{2}}$ [42], where the $(m, i)$ element of $A$ is given by

$$A_{m,i} = \begin{cases} 1 & e_i = (m, k) \text{ is connected, } m \text{ is the source} \\ -1 & e_i = (k, m) \text{ is connected, } k \text{ is the source} \\ 0 & \text{otherwise} \end{cases}$$

$\forall m = 1, \ldots, M$ and $i = 1, \ldots, \frac{M(M-1)}{2}$.

C. DC Model and Problem Formulation

We consider the DC power flow model [1], which is commonly used to represent transmission system power flow, and is based on the following assumptions on the network:

A.1 Branches are considered lossless, which results in $g_{m,k} \approx 0$, where $g_{m,k}$ and $b_{m,k}$ are the $(m,k)$th branch line conductance and susceptance, respectively.

A.2 The bus voltage magnitudes, $V_m$, $m = 1, \ldots, M$, are approximated by 1 per unit (p.u.).
A.3 Voltage angle differences across branches are small, such that \( \sin(\theta_n - \theta_k) \approx \theta_n - \theta_k \), where \( \theta_m, m = 1, \ldots, M \), are the bus voltage angles.

Under Assumptions A.1–A.3, the active power injected at bus \( m \) satisfies

\[
p_m = \sum_{k=1}^{M} V_m V_k \left( g_{m,k} \cos(\theta_m - \theta_k) + b_{m,k} \sin(\theta_m - \theta_k) \right)
\]

\[
\approx \sum_{k=1}^{M} b_{m,k}(\theta_m - \theta_k), \quad \forall m = 1, \ldots, M.
\]

Now, let \( \mathbf{p}[n] \triangleq [p_1[n], \ldots, p_M[n]]^T \) be the vector of active power injected and \( \mathbf{\theta}[n] \triangleq [\theta_1[n], \ldots, \theta_M[n]]^T \) the vector of voltage phase angles at time \( n, \forall n = 0, \ldots, N - 1 \). In addition, the \((m, k)\)th element of the graph Laplacian matrix, \( \mathbf{B} \), is defined in this model as

\[
\mathbf{B}_{m,k} = \begin{cases} 
  -\sum_{i=1,i\neq m}^{M} b_{m,i} & \text{if } m = k \\
  b_{m,k} & \text{if } m \neq k, \ m, k \text{ are connected ,} \\
  0 & \text{otherwise}
\end{cases}
\]

\( \forall m, k = 1, \ldots, M \), where \( b_{m,k} = -\frac{1}{x_{m,k}} \) is the susceptance of the \((m, k)\) branch if this connection exists [2]. The matrix \( \mathbf{B} \in \mathbb{R}^{M \times M} \) is a weighted Laplacian of the graph \( \mathcal{G}(V, \xi) \) describing the power network and, thus, a real, symmetric, and positive semidefinite matrix [49], which satisfies the null space property, \( \mathbf{B}1_M = 0 \), and with nonpositive off-diagonal elements.

Then, based on the model from (2), the noisy linearized DC model of the network can be written as

\[
\mathbf{p}[n] = \mathbf{B}\mathbf{\theta}[n] + \mathbf{w}[n], \quad n = 0, \ldots, N - 1,
\]

where the topology matrix, \( \mathbf{B} \), is a deterministic unknown Laplacian matrix, which is considered static for a short-period of time and under normal operating conditions. The noise is a stationary Gaussian sequence with zero mean and a covariance matrix \( \sigma^2 \mathbf{1}_M \), i.e. \( \mathbf{w}[n] \sim \mathcal{N}(0, \sigma^2 \mathbf{1}_M) \), and it is assumed that the additive noises are independent of the state vectors. The assumption of same noise level for all the buses is a simplification of the behavior of real-world power grids.

The vectors \( \{\mathbf{\theta}[n]\}, n = 0, \ldots, N - 1 \), are assumed to be unknown random states with a joint probability density function (pdf) \( f_{\mathbf{\theta}}(\cdot) \) and marginal pdfs of \( \theta_m, f_{\theta_m}(\cdot), m = 1, \ldots, M \). By subtracting the mean from the data, we can assume that \( \mathbf{\theta} \) has zero mean. The resulting centralized measurements are given by \( \mathbf{p}[n] - \bar{\mathbf{p}} \), where \( \bar{\mathbf{p}} \triangleq \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{p}[n] \) is the sample mean. For the rest of this paper, \( \mathbf{p}[n] \) will denote the mean-centered active power data.

Now, in order to reformulate the model with a full-rank mixing matrix, we use the relation

\[
\mathbf{B} = \mathbf{UB}\mathbf{U}^T,
\]

where

\[
\mathbf{U} \triangleq \begin{bmatrix} -\mathbf{I}_{M-1}^T \mathbf{I}_{M-1} \end{bmatrix} \in \mathbb{R}^{M \times (M-1)}
\]

and \( \tilde{\mathbf{B}} \triangleq \mathbf{B}_{2 \times 2 : M} \) is a 1st-order reduced-Laplacian matrix, which is obtained by removing the first row and first column of \( \mathbf{B} \). By substituting (5) in (4), one obtains

\[
\mathbf{p}[n] = \mathbf{UB}\tilde{\mathbf{\theta}}[n] + \mathbf{w}[n], \quad n = 0, \ldots, N - 1,
\]

where \( \tilde{\mathbf{\theta}}[n] \triangleq \mathbf{U}^T \mathbf{\theta}[n] = [\theta_1[n] - \theta_M[n], \ldots, \theta_M[n] - \theta_1[n]]^T, \quad n = 0, \ldots, N - 1 \). By multiplying both sides of (7) with \( \mathbf{U} \), it can be verified that the model in (7) is equivalent to

\[
\tilde{\mathbf{p}}[n] = \tilde{\mathbf{B}}\tilde{\mathbf{\theta}}[n] + \tilde{\mathbf{w}}[n], \quad n = 0, \ldots, N - 1,
\]

where \( \tilde{\mathbf{p}}[n] \triangleq \mathbf{U}^T \mathbf{p}[n] \) and \( \tilde{\mathbf{w}}[n] \triangleq \mathbf{U}^T \mathbf{w}[n], n = 0, \ldots, N - 1 \). In addition, it can be shown (see, e.g. pp. 134–144 [50]) that the modified noise sequence satisfies \( \tilde{\mathbf{w}}[n] \sim \mathcal{N}(0, \sigma^2 \mathbf{U}^T(\mathbf{U}^T)\mathbf{U}) \), \( n = 0, \ldots, N - 1 \).

We assume here that all sources are time-independent Gaussian distributed, i.e. \( \mathbf{\theta}[n] \sim \mathcal{N}(0, \mathbf{\Sigma}_{\mathbf{\theta}}) \), \( n = 0, \ldots, N - 1 \). Thus, \( \tilde{\mathbf{\theta}}[n] \sim \mathcal{N}(0, \mathbf{\Sigma}_{\tilde{\mathbf{\theta}}}) \), \( n = 0, \ldots, N - 1 \), where \( \mathbf{\Sigma}_{\tilde{\mathbf{\theta}}} \triangleq \mathbf{U}^T \mathbf{\Sigma}_{\mathbf{\theta}} \mathbf{U} \) and we assume that \( \mathbf{\theta} \) has zero mean. The zero-mean and Gaussian assumptions are for the sake of simplicity of derivations. We discuss alternative scenarios in Section VII. The necessary assumption of time-independent states is based on the assumption that we deal with a short time scale, in which the correlation can be assumed to be limited [19].

Under the assumption that \( \mathbf{\Sigma}_{\tilde{\mathbf{\theta}}} \) is known, the observation vectors are also independent Gaussian-distributed vectors, i.e.

\[
\tilde{\mathbf{p}}[n] \sim \mathcal{N}(0, \mathbf{\Sigma}_{\tilde{\mathbf{p}}} \mathbf{B} \mathbf{\Sigma}_{\tilde{\mathbf{\theta}}} \mathbf{B} + \sigma^2 \mathbf{I}_M)
\]

and, assuming nonsingular matrices,

\[
\mathbf{\Sigma}_{\tilde{\mathbf{p}}} \mathbf{B} \sim \mathcal{N}(0, \mathbf{\Sigma}_{\tilde{\mathbf{p}}} \mathbf{B} \mathbf{\Sigma}_{\tilde{\mathbf{\theta}}} \mathbf{B} + \sigma^2 \mathbf{I}_M)
\]

The reduced topology, \( \mathbf{B} \), has the following properties [42], [45]:

P.1 Positive semidefinite - Since \( \mathbf{B} \) is a symmetric, positive semidefinite matrix, \( \mathbf{B} \) is also a symmetric, positive semidefinite matrix.

P.2 Full rank - Under the assumption of a connected graph, \( \mathbf{B} \) is a nonsingular matrix of rank \( M - 1 \) and, thus, can be identified in general. In power system terminology, we assume that there are no unobservable islands in the grid.

P.3 Nonpositive off-diagonal elements - \( \tilde{\mathbf{B}}_{k,m} \leq 0 \), \( \forall k, m = 1, \ldots, M - 1, k \neq m \).

P.4 Diagonally dominant - Since \( \mathbf{B} \) is a Laplacian matrix, \( \mathbf{B} \) is a diagonally dominant matrix, i.e.

\[
\sum_{m=1, m \neq k}^{M-1} |\tilde{\mathbf{B}}_{k,m}| \leq |\tilde{\mathbf{B}}_{k,k}|, \forall k = 1, \ldots, M - 1.
\]

P.5 Sparsity (optional) - It is shown in previous works that the power system is sparse [51], i.e. the zero pseudonorm of the off-diagonal entries of \( \mathbf{B} \), \( ||\mathbf{B}||_{0,\text{off}} \), is much smaller than \( (M - 1)(M - 2) \).

III. ML-BEST

In this section, we develop the basic ML-BEST approach that jointly reconstructs the matrix \( \mathbf{B} \) and the states \( \mathbf{\theta}[n], n = 0, \ldots, N - 1 \), for the model from Section II. This problem can be interpreted as a BSS problem with a Laplacian mixing matrix, or GBSS. First, in Subsection III-A the minimum mean-squared-error (MMSE) estimator of the random states, \( \mathbf{\theta}[n] \),
$n = 0, \ldots, N - 1$, is developed. Then, in Subsection III-C, we develop the ML estimator of the noise variance, $\sigma^2$, and formulate the optimization problem describing the ML estimator of the mixing system.

A. MMSE State Estimation

For given $\mathbf{B}$ and $\sigma^2$, the sequences $\mathbf{p}[n], n = 0, \ldots, N - 1, \theta[n], n = 0, \ldots, N - 1$, are jointly Gaussian. Thus, in this case the MMSE estimator of the state vector is a linear estimator given by (see, e.g., Chapter 20 in [52], [53])

$$\hat{\theta}[n] = \Sigma_\theta \mathbf{B} \left( (\mathbf{B}^T \Sigma_\theta \mathbf{B} + \sigma^2 \mathbf{I}_M)^{-1} \right) \mathbf{p}[n], \quad (11)$$

$n = 0, \ldots, N - 1$. We refer to the estimator in (11) as the oracle MMSE state estimator, i.e., an ideal estimator which has perfect knowledge of the noise variance and the system topology.

The practical state estimator for the considered GBSS problem is obtained by plugging in the ML estimators of the noise variance and the reduced-Laplacian matrix, $\hat{\sigma}^2$ and $\hat{\Sigma}_\theta$, respectively, that are developed in the following in Subsection III-B and III-C, into (11), which results in

$$\hat{\theta}[n] = \Sigma_\theta \hat{\Sigma}_\theta \mathbf{B} + \sigma^2 \mathbf{I}_M \right) \mathbf{p}[n], \quad (12)$$

$n = 0, \ldots, N - 1$.

For high signal-to-noise ratio (SNR) values, i.e. when $\sigma^2 \to 0$, the matrix $\mathbf{B}^T \Sigma_\theta \mathbf{B}$ is a singular matrix and, thus, the covariance matrix of the data from (9) is also a singular matrix. In this case, instead of using the pseudo inverse as in (11) and (12), the unknown parameters can also be treated by removing the linearly dependent random variable (see, e.g., Chapters 3 and 10 in [54]). In power system state estimation this is usually done by setting one bus as a reference bus and setting its angle to zero (see, e.g., [1]), and then only estimating $\theta[n]$. Here we prefer to use instead the state estimation method in (11) and (12) for estimation of $\theta[n]$.

B. ML Estimation of the Noise Variance

The ML estimator of the noise variance $\sigma^2$ for Gaussian measurements with the aforementioned structure is given by

$$\hat{\sigma}^2 = \lambda_M, \quad (13)$$

where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_M$ are the eigenvalues of the sample covariance matrix,

$$\hat{\Sigma}_\theta \triangleq \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{p}[n] \mathbf{p}^T[n]. \quad (14)$$

The derivation of the estimator in (13), which appears in Appendix A, is based on results from [55]–[57]. It can be seen that the estimation performance of this estimator depends directly on the accuracy of the sample covariance matrix estimator. Thus, the sample covariance matrix is required to be a full rank matrix (and, thus, it is required that $N \geq M$) in order that the estimator in (13) will be nonzero. For the special case of some no-load and no-generator buses in the system, the power injection of these buses is identically zero. Thus, the ML estimator is the smallest eigenvalue of the reduced covariance matrix, after removing the associated power data. However, in this case the full topology cannot be recovered, since the sample covariance is not a full rank matrix.

C. System Identification: ML Estimation of the Mixing Matrix

By using the invariance property of the ML estimator [58] and the relation in (5), the ML estimator of the full Laplacian matrix can be obtained from the ML estimator of the reduced-Laplacian matrix, $\hat{\mathbf{B}}^{(\text{ML})}$, as follows:

$$\hat{\mathbf{B}}^{(\text{ML})} = \mathbf{U} \hat{\Sigma}_\theta \mathbf{B} \mathbf{U}^T. \quad (15)$$

In the following, the ML estimator of the reduced topology matrix, $\hat{\mathbf{B}}$, is formulated and is shown to be NP-hard. Practical methods to approximate the ML estimator of $\hat{\mathbf{B}}$, $\hat{\mathbf{B}}^{(\text{ML})}$, are developed in the next section. Under the model from (8) and the Gaussian-distributed sources assumptions, the normalized log-likelihood of $\mathbf{p}[n], n = 0, \ldots, N - 1$, after removing constant terms and substituting the ML estimator of the noise variance from (13), satisfies

$$\psi(\hat{\mathbf{B}}) = -\text{Tr} \left\{ \hat{\Sigma}_\theta \hat{\Sigma}_\theta^{-1} (\hat{\mathbf{B}}, \hat{\sigma}^2) \right\} - \log \left| \hat{\Sigma}_\theta (\hat{\mathbf{B}}, \hat{\sigma}^2) \right|, \quad (16)$$

where

$$\hat{\Sigma}_\theta \triangleq \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{p}[n] \mathbf{p}^T[n] = \mathbf{U} \hat{\Sigma}_\theta \mathbf{U}^T \quad (17)$$

is the modified sample covariance matrix and the last equality is obtained by substituting (14). That is, the log-likelihood from (16) depends on the data only through the sample covariance matrix, $\hat{\Sigma}_\theta$, which is the sufficient statistic for estimating $\hat{\mathbf{B}}$.

Since the reduced-Laplacian matrix satisfies Properties P1-P4, we are interested in minimizing $-\psi(\hat{\mathbf{B}})$ over the domain of symmetric matrices and under the associated constraints as follows:

$$\min_{\mathbf{B} \in \mathbb{S}^{N-1}} -\psi(\mathbf{B}) \quad \text{such that}$$

1) $\hat{\mathbf{B}} > 0$
2) $\hat{\mathbf{B}}_{m,k} \leq 0, \forall m, k = 1, \ldots, M - 1, k < m$
3) $\sum_{k=1}^{M-1} \hat{\mathbf{B}}_{m,k} \geq 0, \forall m = 1, \ldots, M - 1$

The Gaussian log-likelihood function, $\psi(\hat{\mathbf{B}})$, is a concave function of the inverse covariance matrix, $\hat{\Sigma}_\theta^{-1}(\hat{\mathbf{B}}, \hat{\sigma}^2)$. However, even without the sparsity constraint, the constraints in (18) cannot be rewritten as convex constraints on $\hat{\Sigma}_\theta^{-1}(\hat{\mathbf{B}}, \hat{\sigma}^2)$. Therefore, the resulting optimization is not a convex optimization and, in addition, a direct Karush-Kuhn-Tucker (KKT) conditions [59] solution of this constrained minimization is intractable. Two low-complexity implementation methods are described in the next section.

Imposing directly the sparsity constraint in P5 usually results in complex combinatorial searches, and, following advances in compressive sensing [60], [61], the sparsity constraint can be approximated by restricting the off-diagonal $\ell_1$-norm. We perform simulations in Subsection V-C that suggest that results obtained with simple elementwise thresholding of the estimated Laplacian matrix are competitive with those obtained with $\ell_1$-methods. Thus, at the end of the ML-BEST approach, we thresholded the off-diagonal elements of the estimator of the topology matrix, $\hat{\mathbf{B}}^{(\text{ML})}$, from (15), with a threshold, $\tau$, such
Algorithm 1: Basic ML-BEST Algorithm.

Input:
- Observations \( p[n], n = 0, \ldots, N - 1 \).
- State covariance matrix, \( \Sigma_\theta \).

Output: Estimators \( B \) and \( \theta[n], n = 0, \ldots, N - 1 \).

Algorithm Steps:
1. (Optional) Remove the sample mean, 
   \[ p = \frac{1}{N} \sum_{n=0}^{N-1} p[n], \]
   from the observations \( p[n], n = 0, \ldots, N - 1 \).
2. Obtain the sample covariance matrix, \( \Sigma_p \), by (14).
3. Perform eigendecomposition operation for the sample covariance matrix \( \Sigma_p \) to find its eigenvalues 
   \( \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_M \).
4. Estimate the noise variance by the smallest eigenvalue, 
   \( \bar{\sigma}^2 = \lambda_M \).
5. Estimate the reduced-Laplacian matrix and obtain approximation to \( \hat{\Sigma}^{(ML)} \), for example by the two-phase/augmented ML-BEST from Section IV.
6. Reconstruct the full topology matrix according to (15):
   \[ \hat{B}^{(ML)} = U \hat{\Sigma}^{(ML)} U^T. \]
7. Impose sparsity by setting the threshold according to (20):
   \[ \tau = \alpha \min_{m=1,\ldots,M} \hat{B}^{(ML)}_{m,m} \]
   and thresholding such that the \( (k,m) \)th element of the final estimation is given by (19):
   \[ \hat{B}^{(ML)}_{k,m} = \begin{cases} \hat{B}^{(ML)}_{k,m} \text{ if } |\hat{B}^{(ML)}_{k,m}| > \tau \\ 0 \text{ otherwise} \end{cases}, \]
   \( k, m = 1, \ldots, M - 1, k \neq m \).
8. Evaluate the sources according to (12):
   \[ \hat{\theta}[n] = \Sigma_\theta \hat{B}^{ML} \left( (\hat{B}^{ML})^T \Sigma_\theta \hat{B}^{ML} + \sigma^2 I_M \right)^{\dagger} p[n], \]
   \( n = 0, \ldots, N - 1 \).

That the \( (k,m) \)th element of the final estimation is given by

\[ \hat{B}^{(ML)}_{k,m} = \begin{cases} \hat{B}^{(ML)}_{k,m} \text{ if } |\hat{B}^{(ML)}_{k,m}| > \tau \\ 0 \text{ otherwise} \end{cases}, \]

\( k, m = 1, \ldots, M - 1, k \neq m \). The threshold \( \tau \) should be tuned until the desired level of sparsity is achieved, while keeping connectivity. The diagonal elements of \( B \) are known to be positive for the Laplacian matrix, which, thus, has partially known connectivity. The diagonal elements of \( B \) are known to be positive for the Laplacian matrix, which, thus, has partially known connectivity. The diagonal elements of \( B \) are known to be positive for the Laplacian matrix, which, thus, has partially known connectivity. The diagonal elements of \( B \) are known to be positive for the Laplacian matrix, which, thus, has partially known connectivity.

That the \( (k,m) \)th element of the final estimation is given by

\[ \tau = \alpha \min_{m=1,\ldots,M} \hat{B}^{(ML)}_{m,m}, \]

where \( 0 < \alpha < 1 \). The value of \( \alpha \) can be set to the inverse of the number of buses, \( \tau \), or of the average nodal degree [42].

The basic ML-BEST algorithm is summarized in Algorithm 1 for any method of estimation of the reduced-Laplacian matrix, \( B \). Two such methods are described in Section IV.

IV. Practical Implementations of the ML-BEST

In this section, two low-complexity estimation methods of the reduced topology are derived: 1) Two-phase topology recovery in Subsection IV-A; and 2) Augmented Lagrangian topology recovery in Subsection IV-B.

A. Two-Phase Topology Recovery

In this subsection, we propose a low-complexity method for solving (18) in two phases. First, we relax the original optimization problem from (18), by removing constraints 2) and 3) into

\[ \min_{\hat{B} \in \mathbb{S}^{M \times M}} -\psi(\hat{B}) \text{ such that } 1) \hat{B} \succeq 0. \] (21)

It is well known that the relaxed optimization problem from (21) is a convex optimization w.r.t. \( \Sigma_p^{-1} (B, \sigma^2) \) and the optimal solution is the sample covariance matrix inverse, \( \Sigma_p^{-1} \), under the assumption of nonsingular matrices (see, e.g. p. 466 in [62], [63]). In particular, this assumption requires that \( N \geq M - 1 \). Then, by using the invariance property of the ML estimator [58], the one-to-one mapping in (10), and the symmetry \( \hat{B}^T = \hat{B} \), one obtains that the unique minimum of (21) w.r.t. \( B \), which is the ML estimator of a symmetric positive definite mixing matrix, \( \hat{B}^{PD} \), satisfies

\[ \Sigma_p = \hat{B}^{PD} \Sigma_\theta \hat{B}^{PD} + \sigma^2 U^T (U^T)^T, \] (22)

which implies that

\[ \hat{B}^{PD} = \Sigma_\theta^{-\frac{1}{2}} \left( \Sigma_\theta^{-\frac{1}{2}} (\sigma^2 \Sigma_\theta^{-\frac{1}{2}} U^T (U^T)^T) \Sigma_\theta^{-\frac{1}{2}} \right)^{\frac{1}{2}} \Sigma_\theta^{-\frac{1}{2}}. \] (23)

For the noiseless case, i.e. when \( \sigma^2 = 0 \), the estimation in (23) is similar to the one proposed in Theorem 11 in [16] as an adhoc estimation with a diagonal sample covariance matrix of the power injection. In addition, it should be noted that, in contrast to Gaussian graphical models, in the considered model the inverse covariance matrix is not a sparse matrix and, thus, sparse inverse covariance estimation methods, such as the Graphical Lasso [16], [64], cannot be implemented to solve (21).

In the second phase, we find the closest graph Laplacian matrix to the matrix \( UB^{PD} U^T \) in the sense of Frobenius norm. Thus, we solve the following optimization problem:

\[ \min_{B \in \mathbb{S}^{M \times M}} \| UB^{PD} U^T - B \|_F \] such that
1) \( B \succeq 0 \)
2) \( B_{m,k} \leq 0, \forall m, k = 1, \ldots, M, k < m \)
3) \( \sum_{k=1}^{M} B_{m,k} = 0, \forall m = 1, \ldots, M \)

The problem in (24) is a convex optimization problem and can be efficiently computed by standard semidefinite program solvers, such as CVX [65]. This two-phase topology recovery algorithm is summarized in Algorithm 2. The ML-BEST approach with two-phase topology recovery is implemented by Algorithm 1, where Step 5 is implemented by Algorithm 2.

B. Augmented Lagrangian Topology Recovery

In this subsection we develop a constrained independent component analysis (cICA) method [66] to solve (18). This
Algorithm 2: Two-phase Topology Recovery Algorithm.

Input: \( \Sigma_\theta, \Sigma_p, \sigma^2 \).

Output: Estimator \( \hat{\Theta}^{ML} \).

Algorithm Steps:

1) Evaluate the reduced sample covariance matrix from (17) by \( \Sigma_p = U^T \Sigma_p U \).

2) Evaluate the optimal solution of the optimization in (21) by (23):

\[
\hat{\Theta}^{PD} = \Sigma^{-\frac{1}{2}}_\theta \left( \Sigma^{-\frac{1}{2}}_\theta - \sigma^2 U^T (U^T)^\dagger \right) \Sigma^{-\frac{1}{2}}_\theta \dagger \Sigma^{-\frac{1}{2}}_\theta.
\]

3) Find the closest Laplacian matrix, \( \tilde{B}^{(ML)} \), to \( UB \) in (27), by solving the convex optimization problem in (24) (by solvers such as CVX [65]).

\[
\begin{align*}
Q_a(W, \mu, \Lambda, \Gamma) &= -\psi(W^{-1}) \\
&- \sum_{m=1}^{M-1} \sum_{k=1}^{m-1} \Gamma_{m,k} (W_{m,k} - W_{k,m}) \\
&+ \gamma \sum_{m=1}^{M-1} \sum_{k=1}^{m-1} (W_{m,k} - W_{k,m})^2 \\
&+ \frac{1}{2\gamma} \sum_{m=1}^{M-1} \sum_{k=1}^{m-1} \left( \gamma W^{-1}_{m,k} + \Lambda_{k,m} \right)^+ - \Lambda_{k,m}^2 \\
&+ \frac{1}{2\gamma} \sum_{m=1}^{M-1} \left( \gamma W^{-1} - \mu_m \right)^+ \right)^2 - \mu_m^2,
\end{align*}
\]

where \( \mu, \Lambda \succeq 0 \), and \( \Gamma \) are the nonnegative vector, positive semidefinite matrix, and symmetric matrix, respectively, of Lagrange multipliers, and \( \gamma > 0 \) is the penalty parameter. The first term on the r.h.s. of (25), \( -\psi(W^{-1}) \), is the objective function, defined in (16). The following two terms are related to the symmetry constraint and its corresponding penalty term, which guarantees the local convexity hypothesis in the minimization problem. Similarly, the fourth and fifth rows of (25) are related to Constraints 2) and 3) from (18), respectively, including the associated penalty terms and after algebraic manipulations that are described in detail in [66] for general inequality constraints.

The minimization of (25) w.r.t. \( W \) results in the following natural gradient descent learning rule [69] for \( W \):

\[
W(t+1) = W(t) - \eta \nu(W(t), \mu(t+1), \Lambda(t+1), \Gamma(t+1)),
\]

where \( t = 0, 1, \ldots \) is the iteration index,

\[
\nu(W, \mu, \Lambda, \Gamma) \triangleq W^T \frac{\partial Q_a(W, \mu, \Lambda, \Gamma)}{\partial W} W^T,
\]

and \( 0 < \eta \leq 1 \) is the learning rate that determines the step size. By substituting (10) and \( W = B^{-1} \) in (16) and then taking the derivative of the result w.r.t. \( W \), we obtain

\[
\frac{\partial \psi(W^{-1})}{\partial W} = -W^{-T} \left( \Sigma_p - \sigma^2 U^T (U^T)^\dagger \right) W^{-1} \Sigma^{-1}_p W^{-T} + W^{-T}.
\]

By substituting (28) and the derivative of (25) w.r.t. \( W \) in (27), we obtain

\[
\nu(W, \mu, \Lambda, \Gamma) = \left( \Sigma_p - \sigma^2 U^T (U^T)^\dagger \right) W^{-1} \Sigma^{-1}_p W^{-T} - \Lambda + I_{M-1} \mu^T.
\]

Finally, the Lagrange multipliers, \( \Gamma, \Lambda, \) and \( \mu \), according to the gradient ascent method are updated as follows:

\[
\begin{align*}
\Gamma(t+1) &= \Gamma(t) - \gamma \left( W(t) - (W(t))^T \right), \\
\Lambda(t+1) &= \left\{ \Lambda(t) + \gamma \text{off}(W(t)) \right\}^+, \\
\mu(t+1) &= \left\{ \mu(t) - \gamma \text{off}(W(t)) \right\}^+ - I_{M-1} \mu^T.
\end{align*}
\]

m, k = 1, . . . , M - 1, and \( \Lambda(t+1) \) is a symmetric matrix with nonnegative elements and zero diagonal. Then, it is updated according to (26)–(32) until convergence.

The augmented Lagrangian topology recovery is summarized in Algorithm 3. The ML-BEST approach with augmented Lagrangian topology recovery is implemented by Algorithm 1, where Step 5 is implemented by Algorithm 3. It should be noted that, in contrast to the two-phase method, the augmented Lagrangian topology recovery method does not require the non-singularity of the sample covariance matrix.

V. REMARKS

In this section, we discuss the identifiability conditions and complexity in Subsection V-A and V-B, respectively, and describe a few extensions for the proposed model and methods in Subsection V-C.

A. Identifiability Conditions

In this subsection, we discuss the GBSS identifiability conditions, under which the topology matrix and the state vectors can be recovered [56] for the model from Section II with zero-mean measurements. It is well known that Gaussian sources with i.i.d. time-structures cannot be separated [27], [28], [30]. Nevertheless, the following theorem states that when the mixing matrix is a symmetric matrix, consistent separation can rely exclusively on the SOS of the source covariance, even for Gaussian sources.

Theorem 1: Let \( \mathbf{p}[n] \), n = 0, . . . , N - 1 be measurements from the model in (4) with the topology matrix \( B \) and let \( \tilde{B} \) be defined as in (5). Further assume the following conditions:

- \( \tilde{B} \) is a symmetric positive definite matrix
- The covariance of the states, \( \Sigma_\theta \), is known and is a positive definite matrix
- The matrix \( \Sigma_p - \sigma^2 U^T (U^T)^\dagger \), where \( \Sigma_p \) and \( \sigma^2 \) are defined in (17) and (13), respectively, is a positive semidefinite matrix.

...
Algorithm 3: Augmented Lagrangian Topology Recovery Algorithm.

Input: $\Sigma_\theta$, $\Sigma_p$, $\sigma^2$.

Output: Estimator $\hat{B}^{ML}$.

Algorithm Steps:

1) Evaluate the reduced sample covariance matrix from (17) by $\Sigma_p = U^T \Sigma_p (U^T)^T$.

2) Initialize $\hat{B}^{(0)}$, for example by the estimator from (23):

$$\hat{B}^{(0)} = \hat{B}^{PD}.$$

3) Set $t = 0$, $u^{(0)} = 0$, $A^{(0)} = 0$, $W^{(0)} = \left(\hat{B}^{(t)}\right)^{-1}$, and $\gamma, \eta > 0$ to small positive scalar values.

4) Repeat

   a) Update

   $$W^{(t+1)} = W^{(t)} - \eta \nu \left(W^{(t)}, \mu^{(t+1)}, A^{(t+1)}, F^{(t+1)}\right),$$

   where $\nu(\cdot)$ is given in (29).

   b) Update the Lagrange multipliers, $A^{(t+1)}$, $F^{(t+1)}$, and $u^{(t+1)}$, according to (30), (31), and (32), respectively.

   c) $t \rightarrow t + 1$

   Until criterion $||W^{(t+1)} - W^{(t)}||_F \leq \epsilon$.

5) Evaluate the reduced topology matrix

$$\hat{B} = \left(W^{(t+1)}\right)^{-1}.$$
linear constraints on these variables that stem from Constraints 1) - 3) in (24). Thus, the computational complexity of Step 3 is around $O(M^3(M^3 - M^2 - M + 1))$, and the total complexity of the two-phase topology recovery algorithm is $O(M^3(M^3 - M^2 - M + 1))$.

3) Augmented Lagrangian topology recovery

Algorithm 3 shows the augmented Lagrangian topology recovery algorithm. The complexity of the initialization step depends on the selected initial estimator. If, for example, we initialize with $\hat{B}^{PD}$, then it costs $O(5M^3)$, as explained for the previous algorithm. For each iteration the computational complexity of Step 4.a is based on $M \times M$ matrix multiplications and inversions, which costs $O(5M^3)$. The complexity of Step 4.b of calculating the Lagrange multipliers by the thresholding operator (versus zero) is of order $O(5M^3)$. Typically, it takes $100 - 1000$ iterations to converge.

Based on the above exposition, the computational complexities of Algorithms 2 and 3 for topology recovery are of the order $O(M^3(M^3 - M^2 - M + 1))$ and $O(M^3)$, respectively. Thus, if we were to let $M$ grow while keeping $N$ fixed, the augmented Lagrangian topology recovery method would exhibit significant computational savings when compared to the two-phase topology recovery.

C. Possible Extensions

1) Extension for General States Distribution: In the case where the states are non-Gaussian, we can develop the constrained ML-BEST similarly to the derivations in Section III. That is, we assume the model from (7) and compute the reduced source pdf, $f_{\theta}(\cdot)$, by using a transformation of pdf rules (see, e.g. pp. 134–144 [50]). Under these assumptions, the normalized log likelihood of $\tilde{p}[n]$, $n = 0, \ldots, N - 1$, is given by [30]

$$\psi(\hat{B}) = \frac{1}{N} \sum_{n=0}^{N-1} \log f_\theta (\hat{B}^{-1}\tilde{p}[n]) - \log |\hat{B}|. \quad (35)$$

Then, the ML is obtained by minimizing (35) under the reduced-Laplacian matrix Properties P1-P5, similarly to in the problem formulated in (18). If direct KKT solution of this constrained minimization is intractable, we can develop associated low-complexity methods, similarly to in Subsection IV-A and IV-B.

Alternatively the proposed Gaussian ML-BEST methods can also be applied for non-Gaussian distributions with the same covariance, since the structure of the covariance matrices in (9) and (10) holds for any distribution. Although this ML-BEST approach may not be optimal for non-Gaussian distributions, it has the advantage of only requiring the SOS. In addition, SOS methods are expected to be more robust in adverse SNRs [29].

2) Shunt in Admittance Matrix: In many cases, the bus admittance matrix contains a shunt, representing the bus admittance-to-ground connection. Shunt elements are not considered here; nevertheless, the proposed model and methods can be easily extended to the case of some shunt elements by adding the shunt elements to the diagonal terms of the matrix $B$. In this case, the symmetry of the matrix $B$ is preserved, but $B$ becomes a nonsingular matrix and the assumption of a reference bus is redundant.

3) Extension to Power Flow Data: In Section II, we assumed that we only have bus injection sensors. If, in addition, we have line flow sensors, then, under the DC model, a line flow sensor measures the line flow from bus $k$ to bus $m$:

$$p_{m,k} = b_{m,k}(\theta_m - \theta_k), \quad m, k = 1, \ldots, M,$$

for all lines. In this case, we know the connectivity of the system, i.e. the matrix $A$ from (1), and the estimation task is reduced to estimating the susceptances of the branches and the states, based on an extended linear model. The recovery in this case can be done, for example, by developing a framework of constrained estimation or by using the BEST methods, and then enforcing nonzero (or zero) values on connected (or disconnected) transmission lines, according to the additional knowledge on the connectivity.

VI. CRAMÉR-RAO BOUND ON THE LAPLACIAN MATRIX ESTIMATION

The CRB is a commonly-used lower bound on the mean-squared error (MSE) matrix of any unbiased estimator of a deterministic parameters vector. Thus, it is a useful tool for performance analysis and system design. Moreover, in the asymptotic region, i.e. at high SNR and/or large number of observations, the performance of the ML estimator asymptotically approaches the CRB [70]. Therefore, the CRB can be used to assess the performance of the ML estimator and of its low-complexity implementations. In this section, we derive a closed-form expression for the CRB of the mixing Laplacian matrix and the noise variance, by modeling the sources as nuisance random parameters and using their marginal pdf [71]. The developed CRB is used as a benchmark for system design and for performance analysis of the proposed low-complexity techniques.

In order to develop the CRB, we use the symmetry of the matrix $B$ to define the vector of unknown parameters for the CRB as

$$\alpha \triangleq [\vech(\tilde{B}^T \sigma)^T, \sigma^T] \in \mathbb{R}^{M(M-1)/2 + 1},$$

which consists of the lower triangular elements of $B$, including the diagonal, and the noise variance, $\sigma^2$. Then, we define the matrices

$$Q \triangleq \left( \Sigma_p^{-1}(\tilde{B}, \sigma^2) \otimes \Sigma_p^{-1}(\tilde{B}, \sigma^2) \right),$$

$$K \triangleq \left( (B\Sigma_p^{-1} \otimes I_{M-1}) + (I_{M-1} \otimes (B\Sigma_p^{-1}) \right),$$

and $\Psi$, which is an $(M-1)^2 \times (M(M-1)/2 + 1)$ matrix, where the first $M(M-1)/2$ columns are the vectors

$$\left(1 - \frac{1}{2} \delta_{k,l}\right) \vech(E_{k,l}), \quad k, l = 1, \ldots, M - 1,$$

ordered with the same order as $\vech(\tilde{B})$, and the last column is $K^{-1} \vech(U^T)$. The following Theorem states the CRB for the considered GBSS problem.

Theorem 2: Consider estimating the unknown vector, $\alpha$, based on $N$ data samples obeying the model in (4), where the noise and the states are assumed to be mutually independent and to have Gaussian distributions. Then, under some mild
regularity condition, the MSE of any unbiased estimator of $\alpha$ obeys
\[ E \left[ ( \hat{\alpha} - \alpha ) ( \hat{\alpha} - \alpha )^T \right] \geq \text{LCRB}(\alpha), \quad (38) \]
where
\[ \text{LCRB}(\alpha) \triangleq \frac{2}{N} \left( \tilde{\Psi}^T K^T QK \tilde{\Psi} \right)^{\dagger} \quad (39) \]
is the CRB for this case.

**Proof:** The proof is given in Appendix B.

The bound from (38) implies, in particular, the lower bound on the MSE matrix of the lower triangular of the reduced-Laplacian matrix:
\[ E \left[ ( \text{vech}(\hat{B}) - \text{vech}(B) ) ( \text{vech}(\hat{B}) - \text{vech}(B) )^T \right] \geq \left[ \text{LCRB}(\alpha) \right]_{1, \frac{M(M-1)}{4}, 1, \frac{M(M-1)}{4}}. \quad (40) \]

In general BSS problems, the CRB cannot be calculated and the induced CRB has been proposed as an alternative [31]–[33], [72]. Here, due to the symmetry of the mixing matrix, we obtain the associated CRB from Theorem 2. Alternatively, this bound could be derived via the constrained CRB (CCRB) approach (see, e.g. [73]–[75]) under the equality parametric constraint of a symmetric mixing matrix. It should be emphasized that in the evaluation of the CRB, which is a local bound, the inequality constraints from the optimization in (18), i.e. the constraints of positive diagonal elements and nonpositive off-diagonal elements of the Laplacian matrix, do not contribute any side information [73]–[75] and the sparsity constraint also does not affect the CRB if the exact sparsity level is unknown [76]. Thus, the proposed CRB can also be used to investigate the contribution of the inequality constraints, by comparing it with the asymptotic performance of the proposed estimation methods that take into account both the equality and the inequality constraints.

**VII. SIMULATIONS**

In this section, we present simulation examples conducted in order to evaluate the performance of the proposed ML-BEST methods from Algorithm 1, combined with two-phase topology recovery and with augmented Lagrangian topology recovery from Algorithms 2 and 3, respectively. The optimization problems are solved using the CVX toolbox [65]. The sparsity threshold is set according to (20) where $\alpha$, as well as the step sizes, $\eta$ and $\gamma$ in Algorithm 3, are tuned experimentally. The simulations include three scenarios: IEEE 14-bus system, IEEE 118-bus system, and a random topology graph, with at least 100 Monte-Carlo simulations for each scenario and with different scenarios.

The MSE performance of the state estimators is compared with that of the oracle MMSE estimator from (11). In addition to the MSE of the vectorized topology estimators, vech($\hat{B}$), the topology estimation performance is also measured by the F-score metric [77]:
\[ FS(B, \hat{B}) \triangleq \frac{2tp}{2tp + fn + fp}, \]
where $tp$, $fp$, and $fn$ are the true-positive, false-positive, and false-negative detection of graph edges in $\hat{B}$ w.r.t. the ground truth edges in $B$. The F-score takes values between 0 and 1, where the value 1 means perfect classification. The F-score is a measure for the error probability in the connectivity matrix. In addition, we use the trace of the CRB from Theorem 2 as a benchmark on the trace of the MSE of the Laplacian matrix estimation.

**A. IEEE 14-Bus Power System**

In this subsection, we implement the proposed methods for the IEEE 14-bus system, representing a portion of a power system in the Midwestern U.S. The system parameters, such as branch susceptances, are taken from [78] and $M = 14$. The power flow measurements are generated using (4). The state covariance matrix is set to $\Sigma_\theta = c^2 I_M$. The SNR is defined as $\text{SNR} = 10 \log_{10} \left( \frac{1}{\text{tr} \{ B \Sigma_\theta B \}} \right)$.

We first show in Fig. 1 visual comparisons between the Laplacian matrix of the IEEE 14-bus system and the associated estimators of the Laplacian matrix, $\hat{B}$, obtained by the two-phase ML-BEST and augmented Laplacian ML-BEST for $N = 200$ and $\text{SNR} = 15 \text{ dB}$. (a) the original Laplacian matrix; (b) and (c) the estimated Laplacian by two-phase ML-BEST and augmented Lagrangian ML-BEST methods, respectively. The black circles indicate false connections.

![Fig. 1](image-url)

**Fig. 1.** Illustration of the ML-BEST topology recovery methods to estimate the Laplacian matrix of the IEEE-14 bus system with $N = 200$ samples and $\text{SNR} = 15 \text{ dB}$: (a) the original Laplacian matrix; (b) and (c) the estimated Laplacian by two-phase ML-BEST and augmented Lagrangian ML-BEST methods, respectively. The black circles indicate false connections.

In Fig. 2.a the MSE of the proposed ML-BEST methods for topology estimation and the associated CRB are presented, and in Fig. 2.b the F-score metric of the two ML-BEST methods is presented. It can be seen that the performance improves in any sense as $N$ increases, as expected. In Fig 2.a the MSE of the proposed ML-BEST methods for topology estimation and the associated CRB are presented, and in Fig. 2.b the F-score metric of the two ML-BEST methods is presented. It can be seen that while the two-phase topology recovery performs better in terms of F-score, the two ML-BEST methods have similar performance in terms of MSE. That is, the two-phase topology recovery is better in terms of estimating the connectivity matrix, i.e. it distinguishes between existing and absent links, while the performance of both topology recovery methods is close to the CRB for high SNR. It should be noted that since the CRB does not take into account the information on inequality and sparsity constraints [73]–[76], it could be higher than the true performance. The MSE of the state estimators presented in Fig. 2.c is similar for
the two methods in this case. It can be seen that for high SNRs, the state estimation performance of the ML-BEST methods with estimated topology converges to that of the oracle method, which uses the true topology. Therefore, we can conclude that for high SNRs the topology estimation converges to the true topology.

In order to demonstrate the influence of non-Gaussian distributions and of nonzero mean of the states on the performance of the proposed methods, we simulated the IEEE 14-bus system with uniformly distributed states. That is, the states, \( \theta[n] \), \( n = 0, \ldots, N - 1 \), are modeled as uniformly distributed measurements around the nominal value of the buses, \( \theta^{(\text{nom})} \), according to [79], such that \( \theta[n] \in [\theta^{(\text{nom})} - \frac{\pi}{8}, \theta^{(\text{nom})} + \frac{\pi}{8}] \), \( n = 0, \ldots, N - 1 \). Since in this case the states are not assumed to have zero mean, we replace the estimator from (12) by

\[
\hat{\theta}[n] = \frac{1}{N} \sum_{m=1}^{M} \sum_{k=1}^{M} \sin(\theta_m - \theta_k) \mathbf{p}[m | k] \theta^{(\text{nom})} + \theta^{(\text{nom})},
\]

(41)

\( n = 0, \ldots, N - 1 \), where \( \theta^{(\text{nom})} \) is the assumed mean, which can be obtained by using historical data. The results in terms of topology MSE, F-score metric, and the MSE of the state estimators are presented in Figs. 3.a–3.c for \( N = 200 \) and \( N = 1,500 \). It can be seen that the performance is similar to those that were obtained for the zero-mean Gaussian case in Fig. 2 and that the methods are robust to non-Gaussian distributions with nonzero mean.

The considered DC model is an approximation of nonlinear power flow. In order to examine the effect of nonlinearity on the performance of the proposed algorithms, we investigate the robustness of our methods to deviations from the DC model assumption of small voltage angle differences across branches. To this end, we implemented the IEEE 14-bus system with the measurement model from (2), but without using Assumption A.3, i.e. we use \( \sin(\theta_m - \theta_k) \) and not \( \theta_m - \theta_k \), in (2) \( \forall m = 1, \ldots, M \). The results in terms of F-score metric and the MSE of the state estimators are presented in Fig. 4 for \( N = 1,500 \) and \( \sigma^2 = 0.01 \), versus the probability \( \sum_{m,k} \) are connected \( \Pr (|\theta_m - \theta_k| < 1) \). This probability represents the probability that the measurements satisfy Assumption A.3. It is calculated by using Monte-Carlo simulations with var-
Fig. 4. The performance of the oracle estimator and the ML-BEST methods, with two-phase and augmented Lagrangian topology recovery, for IEEE-14 bus system and nonlinear measurement model with $N = 1,500$, $\sigma^2 = 0.01$, and normally distributed states. The performance is presented versus the probability that Assumption A.3., i.e. the assumption of small voltage angle differences across branches, holds.

ious experiments on the IEEE 14-bus system, when we generate Gaussian phasor angle data with different variances, using MAT-POWER [79]. The results of the F-score metric in the left figure in Fig. 4 indicate that the recovery of the connectivity matrix is not affected by the nonlinearity for both ML-BEST methods. The right figure in Fig. 4 shows that the MSE of the state estimation increases with deviations from the linear model. It can be seen that the same trend occurs at approximately the same rate for oracle model. That is, the MSE behavior is a conventional phenomenon, which is due to the use of the approximated DC model for state estimation, and it is not unique to the blind setting, in which the topology is unknown.

B. IEEE 118-Bus Power System

In this subsection we investigate the performance of the proposed algorithm when applied to large-scale networks and present the results versus the number of time samples, $N$, in order to demonstrate the identifiability conditions. In particular, we use the IEEE 118-bus system, where the system parameters are taken from [78], $\sigma^2 = 0.01$, and $c = \sqrt{10}$. Fig. 5 shows the performance of the BEST methods versus $N$, where the topology recovery is implemented by 1) two-phase method; 2) augmented Lagrangian method; and 3) single-phase method, in which the topology recovery is obtained by applying only the first stage of the two-phase ML BEST from Step 2 in Algorithm 2. Fig. 5 shows that the performance improves in any sense as $N$ increases, as expected. The results show that the proposed methods are applicable for $N > 150$ and that the performance improves when $N$ increases, as expected.

In Fig. 5.a, we compare the performance of these methods with the proposed CRB and with the oracle CRB, which is equal to the proposed CRB limited to the support of the sparse Laplacian matrix [76]. It can be seen that the CRB is a valid lower bound on the single-phase method, since this method does not use the inequality and sparsity constraints, similarly to the CRB. The gain in the estimation that arises from the inequality constraints can be deduced from the gap between the asymptotic performance of the proposed two-phase and augmented Lagrangian methods, that take into account both the equality and the inequality constraints, and the CRB. In addition, it can be seen in this figure that the proposed oracle CRB is a valid lower bound on these sparse methods, but it is not a tight bound. This is due to the fact that the oracle CRB assumes that the exact sparsity level is known, in contrast with the proposed methods that employ sparsity with an unknown level. In Fig. 5.c we can see that the performance of all the state estimators, including the oracle estimator, degrades significantly as $N$ decreases. This is due to the fact that the state estimator from (12) is a function of the noise variance, $\hat{\sigma}^2$, which cannot be estimated for small number of measurements, as discussed in Subsection III-B.

Comparison between the results of the IEEE 14-bus system in Fig. 2 and those of the IEEE 118-bus system in Fig. 5 indicates that the inequality and sparsity constraints play a more significant role as the network size increases.

C. Random Topology

In this subsection we simulate synthetic graphs from the Watts-Strogatz ‘small world’ graph model [80] with varying numbers of buses, $M$, and an average nodal degree of 4, which is shown to be appropriate for the simulation of synthetic power grid data [51]. It should be noted that the average nodal degree of a power network is almost invariant to the size of the network and, thus, the sparsity level is usually constant around $\frac{4M}{M^2}$. The state covariance matrix is set to $\Sigma_0 = c^2 I_M$, with $c = 0.5$. In
order to achieve uniform SNR simulations, we set the Frobenius norm of the Laplacian matrix to a constant value. Typical values of the Frobenius norm of the topology matrix are shown in Table I for some IEEE test grids [78]. Based on this table, we choose to set the norm to $\|B\|_F = 3$.

In this subsection, we also perform simulations that show the performance of the proposed methods, while imposing the sparsity constraint via the $\ell_1$-norm minimization method. For these methods, instead of Step 7 in Algorithm 1, we implemented the two-phase topology recovery where the objective function in the optimization in the second stage from (24) is replaced with

$$\min_{B \in \mathbb{S}^M} \|U \tilde{B}^{PD} U^T - B\|_F + \rho \|B\|_{1-off},$$

where the constraints are the same as in (24). Similarly, for the augmented Lagrangian topology recovery we add $\rho \|B\|_{1-off}$ to the objective function from (25), $Q_S(W, \mu, \Lambda, \Gamma)$. The term $\rho \|B\|_{1-off}$ acts as a regularizer on the Laplacian matrix, $B$, to control its sparsity level. The regularization parameter, $\rho$, can be tuned until the desired level of sparsity is achieved. In the following, we refer to these topology recovery methods as “two-phase $\ell_1$” and “augmented Lagrangian $\ell_1$”.

The performance of the different methods for this random topology with different sparse methods is presented in Fig. 6 versus the number of buses in the system for $\alpha^2 = 0.01$ [1], $N = 200, 1, 500$, and $c = \sqrt{10}$. In Fig 6.a the normalized MSE (NMSE) of the ML-BEST methods for topology estimation and the associated CRB and oracle CRB are presented. The NMSE of the different methods is lower than the CRB, which does not take into account the sparsity, and higher than the oracle CRB, which assumes perfect knowledge of the connectivity. In Fig. 6.b the F-score metric of the two ML-BEST methods is presented. It can be seen that the two-phase topology recovery performs better in terms of F-score for $N = 200$ and $N = 1, 500$. The MSE of the state estimators presented in Fig. 6.c is almost identical for all the ML-BEST methods with $N = 200$ and $N = 1, 500$. The performance of the two methods, with any sparse method, becomes closer to those of the oracle performance as $N$ increases. It can be seen from Figs. 6.a–6.c that the performance, in terms of topology NMSE, F-score, and MSE of the state estimators, degraded as $M$ increases since there are more parameters to estimate. We can also conclude that incorporating sparsity by simple elementwise thresholding of the estimated Laplacian matrix is competitive with the $\ell_1$ methods in terms of performance, and, thus, is preferable for simple implementation of the ML-BEST methods.

In order to demonstrate the empirical complexity of the proposed methods for different problem dimensions, the average computation time, “runtime”, was evaluated by running the algorithm using Matlab on an Intel Core(TM) i7-7600U CPU computer, 2.80 GHz. Fig. 7 shows the runtime of the ML-BEST methods as a function of the number of buses, $M$, for a random topology and $N = 200, 1, 500$ samples. It can be seen that the runtime increases polynomially with the number of buses, $M$, and it is higher for the two-phase topology recovery than for the augmented Lagrangian topology recovery with 100 iterations, as expected from the theoretical discussion on computational complexity in Subsection V-B. The reason for this is that the two-phase topology recovery stage from Algorithm 2 requires solving an SDP problem in (24) and, therefore, has a much higher computational complexity as compared to the augmented ML-BEST estimator. Thus, the augmented ML-BEST estimator is preferable for large-scale systems. The number of measurements, $N$, has no significant effect since it is only associated with the cost of computing the sample covariance matrix and the state estimation at the beginning and the end of the basic

### Table I

| Test Case | IEEE-14 | IEEE-30 | IEEE-57 | IEEE-118 |
|-----------|---------|---------|---------|---------|
| $\|B\|_F$ | 2.79    | 4.2285  | 7.6887  | 4.7327  |

**Fig. 6.** The performance of the ML-BEST methods, with two-phase and augmented Lagrangian topology recovery, for random topology versus the number of buses with $N = 200, 1, 500$ and for $\alpha^2 = 0.01$.

**Fig. 7.** Runtime of the ML-BEST methods, with thresholding and $\ell_1$ penalty, versus number of buses, $M$, in random topology with $N = 200, 1, 500$ samples.
ML-BEST approach. Applying sparsity by using $\ell_1$-norm minimization increases the runtime of the two-phase method but does not change the runtime of the augmented Lagrangian method.

VIII. CONCLUSION

In this paper, we introduce the novel ML-BEST method for blind estimation of states and topology in power systems, by formulating the problem as a GBSS with a Laplacian mixing matrix. Since the topology recovery stage of the ML-BEST is shown to be an NP-hard optimization problem, we propose two low-complexity algorithms for the implementation of the topology recovery stage of the ML-BEST estimator: 1) a two-phase topology recovery algorithm, which finds the relaxed positive semidefinite mixing matrix solution and then finds the closest Laplacian matrix to this solution by using convex optimization; 2) an augmented Lagrangian topology recovery algorithm, which is based on classical cICA approaches. These methods rely only on the SOS of the state signals and, in contrast to classical BSS techniques, enable the separation of Gaussian sources.

We present some identifiability conditions for this GBSS problem, complexity analysis of the proposed ML-BEST methods, and the associated CRB of the demixing parameters. We show by simulations that the proposed techniques may be treated as a large-sample approximation of the ML estimator, with the same asymptotic accuracy for the considered scenarios. Moreover, for large systems the MSE of the topology estimator by the proposed methods is shown to be lower than the CRB, which does not take into account the inequality constraints. Numerical simulations show that the proposed ML-BEST methods succeed in reconstructing the topology and estimating the states, and that the topology estimators achieve the CRB asymptotically. The two-phase topology recovery method is shown to be better in terms of estimating the connectivity matrix in the sense of higher F-score. The augmented Lagrangian ML-BEST is preferable for large networks, since the two-phase ML-BEST is a computationally heavy algorithm, as described in Subsection V-B, and since it does not require the sample covariance matrix to be a nonsingular matrix. Additionally, the state estimators converge to the oracle state estimator, which assumes perfect knowledge of the topology.

State estimation is the backbone of power system monitoring and processing. The presented results indicate that even if the topology recovery is not perfect, the MSE of the state estimation is close to the MSE of the oracle performance. Thus, the proposed ML-BEST methods can be applied for practical power system operations without assuming knowledge of the topology. In future work, the proposed methods will be extended to address large-scale systems for faster and real time estimation, by efficiently exploiting the sparsity pattern of the Laplacian matrix. Finally, in order to deal with time-dependent state distributions, the proposed methods can be extended in future research along the path of the approaches from [14] or [81].

APPENDIX A

DERIVATION OF THE ML ESTIMATOR OF THE NOISE VARIANCE FROM (13)

Similar to (16), the normalized log likelihood of $p[n]$, $n = 0, \ldots, N - 1$, after removing constant terms, satisfies

$$
\psi(\hat{B}) = -\text{Tr} \left\{ \hat{\Sigma}_p \hat{\Sigma}_p^{-1}(\hat{B}, \sigma^2) \right\} - \log \left| \hat{\Sigma}_p(\hat{B}, \sigma^2) \right|,
$$

where $\hat{\Sigma}_p(B, \sigma^2)$ and $\hat{\Sigma}_p$ are defined in (9) and (14), respectively. Since under our assumptions (Assumption P2), rank $(\hat{B}) = M - 1$ and $\hat{\Sigma}_\theta$ is a full rank matrix, we can decompose $\hat{\Sigma}_p(B, \sigma^2)$ by using eigendecomposition

$$
B^T \hat{\Sigma}_\theta B = V \Sigma V^T,
$$

where the columns of $V \in \mathbb{R}^{M \times M}$ are the eigenvectors of $B^T \hat{\Sigma}_\theta B$ associated with the corresponding eigenvalues, $\chi_1 \geq \chi_2 \geq \ldots \geq \chi_{M-1} > 0$, where 0 is the smallest eigenvalue, $V^T V = I_M$, and $R \triangleq \text{diag}([\chi_1, \ldots, \chi_{M-1}, 0])$, i.e., it is a diagonal matrix with the elements $\chi_1, \ldots, \chi_{M-1}, 0$ on its diagonal. Then,

$$
\hat{\Sigma}_p(B, \sigma^2) = VRV^T + \sigma^2 I_M
$$

$$
= V \text{diag} \left( [\chi_1 + \sigma^2, \ldots, \chi_{M-1} + \sigma^2, \sigma^2] \right) V^T.
$$

Since the determinant of a matrix is the product of its eigenvalues, one obtains

$$
\log |\hat{\Sigma}_p(B, \sigma^2)| = \sum_{m=1}^{M-1} \log(\chi_m + \sigma^2) + \log \sigma^2
$$

$$
= \log |S| + \log \sigma^2,
$$

where $S \triangleq R_{1:1:M-1,1:M-1} + \sigma^2 I_{M-1}$ is a non-singular, diagonal matrix. In addition, based on (45) and similar to the existing derivations (see, e.g. [57]), it can be verified that

$$
\hat{\Sigma}_p^{-1}(B, \sigma^2)
$$

$$
= V \text{diag} \left( \left[ \frac{1}{\chi_1 + \sigma^2}, \ldots, \frac{1}{\chi_{M-1} + \sigma^2}, \frac{1}{\sigma^2} \right] \right) V^T
$$

$$
= \frac{1}{\sigma^2} \left( I_M - \tilde{V}V^T \right) + \tilde{V}S^{-1}\tilde{V}^T,
$$

where $\tilde{V} \overset{\Delta}{=} V_{1:1:M-1}$. Inserting (46) and (47) in (43) and using the trace operator rules yields

$$
\psi(\hat{B}) = -\frac{1}{\sigma^2} \text{Tr} \left\{ \hat{\Sigma}_p \left( I_M - \tilde{V}V^T \right) \right\}
$$

$$
- \text{Tr} \left\{ S^{-1}\hat{V}^T \hat{\Sigma}_p \hat{V} \right\} - \log |S| - \log \sigma^2.
$$

It is shown in Eq. (12) in [57] that the maximum of (48) w.r.t. $\tilde{V}$, is obtained by the ML estimator of $\tilde{V}$, which is the first $M - 1$ columns of the eigenvector matrix of the sample covariance matrix, $\hat{\Sigma}_p$. By substituting this result in (48), we obtain the following likelihood:

$$
\psi(\hat{B}) = -\frac{1}{\sigma^2} \lambda_M - \text{Tr} \left\{ S^{-1} \text{diag} ([\lambda_1, \ldots, \lambda_{M-1}] \right) \}
$$

$$
- \log |S| - \log \sigma^2,
$$

where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_M$ are the eigenvalues of the sample covariance matrix, $\hat{\Sigma}_p$. The result of the minimization of (49) w.r.t. the elements on the diagonal of $S$ and $\sigma^2$ results in

$$
\hat{S}_m = \lambda_m, \quad m = 1, \ldots, M - 1
$$

and (13), respectively, which completes this derivation.
APPENDIX B

PROOF OF THEOREM 2

Under some mild regularity condition [58], the CRB on the MSE of any unbiased estimator of \( \alpha \) is given by

\[
E \left[ (\hat{\alpha} - \alpha) (\hat{\alpha} - \alpha)^T \right] \geq \text{L}_{\text{CRB}}(\alpha) = J^{-1}(\alpha),
\]

where \( J(\alpha) \) is the associated Fisher information matrix (FIM). In order to compute the CRB, since the measurements at different times are independent, the FIM can be obtained by multiplication of the single-time measurement FIM by \( N \). Thus, due to the zero-mean Gaussian distribution of \( \overline{p}_k[\eta], \forall \eta = 0, \ldots, N-1 \), (the \( m, r \) entry of the associated \( \frac{M(M-1)}{2} + 1 \times \frac{M(M-1)}{2} + 1 \)) FIM is given by (see, e.g. p. 48 in [58])

\[
J_{m,r}(\alpha) = \frac{N}{2} \text{Tr} \left\{ \Sigma_p^{-1}(\hat{B}, \sigma^2) \frac{\partial \Sigma_p(\hat{B}, \sigma^2)}{\partial \alpha_m} \right\} \times \Sigma_p^{-1}(\hat{B}, \sigma^2) \frac{\partial \Sigma_p(\hat{B}, \sigma^2)}{\partial \alpha_r} \right\},
\]

for any \( m, r = 1, \ldots, \frac{M(M-1)}{2} + 1 \). The derivatives of \( \Sigma_p(\hat{B}, \sigma^2) \) w.r.t. the elements of \( \hat{B} \) and \( \sigma^2 \) are given by

\[
\frac{\partial \Sigma_p(\hat{B}, \sigma^2)}{\partial \alpha_r} = \frac{\partial \Sigma_p(\hat{B}, \sigma^2)}{\partial \hat{B}_{k,l}} \frac{\partial \hat{B}_{k,l}}{\partial \alpha_r} = \frac{\partial \Sigma_p(\hat{B}, \sigma^2)}{\partial \hat{B}_{k,l}} = \left( 1 - \frac{1}{2} \delta_{k,l} \right) \left( E_{k,l} \Sigma_{\theta} \hat{B} + \hat{B} \Sigma_{\theta} E_{k,l} \right),
\]

where \( E_{k,l} = e_k e_l^T + e_l e_k^T \), and for \( r = 1, \ldots, \frac{M(M-1)}{2} \), and where \( \delta \) is such that \( \alpha_r = \hat{B}_{k,l} \), and

\[
\frac{\partial \Sigma_p(\hat{B}, \sigma^2)}{\partial \alpha_r} = \frac{\partial \Sigma_p(\hat{B}, \sigma^2)}{\partial \sigma^2} = U(U^T)^T,
\]

for \( r = \frac{M(M-1)}{2} + 1 \).

By substituting (52) in (51), one obtains

\[
J_{m,r}(\alpha) = \frac{N}{2} \left( 1 - \frac{1}{2} \delta_{k,l} \right) \left( 1 - \frac{1}{2} \delta_{q,p} \right) \times \text{Tr} \left\{ \Sigma_p^{-1}(\hat{B}, \sigma^2) \left( E_{k,l} \Sigma_{\theta} \hat{B} + \hat{B} \Sigma_{\theta} E_{k,l} \right) \right\} \times \Sigma_p^{-1}(\hat{B}, \sigma^2) \left( E_{p,q} \Sigma_{\theta} \hat{B} + \hat{B} \Sigma_{\theta} E_{p,q} \right),
\]

for any set of matrices \( A_i, i = 1, 2, 3 \), of compatible dimensions. By applying (55) on (54) with the matrices \( A_1 = E_{k,l} \Sigma_{\theta} \hat{B} + \hat{B} \Sigma_{\theta} E_{k,l}, A_2 = (\Sigma_p(\hat{B}, \sigma^2))^{-1}, \) and \( A_3 = E_{p,q} \Sigma_{\theta} \hat{B} + \hat{B} \Sigma_{\theta} E_{p,q} \), and using the symmetry of these matrices, the \( (m, r) \) entry of the FIM from (54) can be rewritten as

\[
J_{m,r}(\alpha) = \frac{N}{2} \psi^T(l,k) \psi(p,q),
\]

where \( \psi(l,k) \) is defined in (36) and the last equality is obtained by using the Kronecker product rule vec \((A_1 \otimes A_2)\) vec \(A_2\). By substituting (57) in (56), we obtain

\[
J_{m,r}(\alpha) = \frac{N}{2} \left( 1 - \frac{1}{2} \delta_{k,l} \right) \left( 1 - \frac{1}{2} \delta_{p,q} \right) \times \text{vec}(E_{k,l})^T \left( \Sigma_{\theta} \right)^{-1} \psi, \]

Similarly, by substituting (52) and (53) in (51), and using the symmetry of the matrices, we obtain that the \( (m, r) \) entry of the FIM is

\[
J_{m,s}(\alpha) = \frac{N}{2} \left( 1 - \frac{1}{2} \delta_{l,k} \right) \times \text{vec}(E_{l,k})^T \left( \Sigma_{\theta} \right)^{-1} \psi, \]

for \( s = \frac{M(M-1)}{2} + 1, m = 1, \ldots, \frac{M(M-1)}{2} \), and \( m \) is such that \( \alpha_m = \hat{B}_{k,l} \). Equations (58)–(61) imply that the FIM can be formulated in a matrix form as follows:

\[
J(\alpha) = \frac{N}{2} \left( 1 - \frac{1}{2} \delta_{k,l} \right) \times \text{vec}(E_{k,l})^T \left( \Sigma_{\theta} \right)^{-1} \psi,
\]

where the matrix \( \psi \) is an \((M - 1)^2 \times \frac{M(M-1)}{2} + 1\) matrix, in which the first \((M - 1)^2 \times \frac{M(M-1)}{2} + 1\) columns are the vectors

\[
\left( 1 - \frac{1}{2} \delta_{k,l} \right) \text{vec}(E_{k,l}),
\]

ordered with the same order as vec\(B\), and the last column is \( \left( \Sigma_{\theta} \right)^{-1} \psi \). By substituting (62) in (50) we obtain the CRB in (38)–(39).

REFERENCES

[1] A. Abur and A. Gomez-Exposito, Power System State Estimation: Theory and Implementation, New York, NY, USA: Marcel Dekker, 2004.

[2] G. B. Giannakis, V. Kekatos, N. Gatsis, S. J. Kim, H. Zhu, and B. F. Wollenberg, “Monitoring and optimization for power grids: A signal processing perspective,” IEEE Signal Process. Mag., vol. 30, no. 5, pp. 107–128, Sep. 2013.
[3] J. Weimer, S. Kar, and K. H. Johansson, “Distributed detection and isolation of topology attacks in power networks,” in Proc. 1st Int. Conf. High Confidence Networked Syst. 2012, pp. 65–72.

[4] J. Kim and L. Tong, “On topology attack of a smart grid: Undetectable attack and countermeasures,” IEEE J. Sel. Areas Commun., vol. 31, no. 7, pp. 1294–1305, Jul. 2013.

[5] S. Soltan, M. Yannakakis, and G. Zussman, “Power grid state estimation following a joint cyber and physical attack,” IEEE Trans. Control Netw. Syst., vol. 5, no. 1, pp. 499–512, Mar. 2018.

[6] G. Liang, S. R. Weller, J. Zhao, F. Luo, and Z. Y. Dong, “A framework for cyber-topology attack,” 2018, Line-switching and new attack scenarios,” IEEE Trans. Smart Grid, vol. 10, no. 2, pp. 1704–1712, Mar. 2019.

[7] T. Rountenburg and Y. C. Eldar, “Centralized identification of imbalances in power networks with synchrophasor data,” IEEE Trans. Power Syst., vol. 33, no. 2, pp. 1981–1992, Mar. 2018.

[8] T. Rountenburg, Y. Xie, R. M. Willett, and L. Tong, “PMU-based detection of imbalance in three-phase power systems,” IEEE Trans. Power Syst., vol. 30, no. 4, pp. 1966–1976, Jul. 2015.

[9] S. Cui, Z. Han, S. Kar, T. T. Kim, H. V. Poor, and A. Tajei, “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.

[10] R. Emami and A. Abur, “Tracking changes in the external network model,” in Proc. North Amer. Power Symp., Sep. 2010, pp. 1–6.

[11] Y. Sharon, A. M. Annaswamy, A. L. Motto, and A. Chakraborty, “Topology identification in distribution network with limited measurements,” in Proc. PES Innovative Smart Grid Technol., Jan. 2012, pp. 1–6.

[12] K. A. Clements and P. W. Davis, “Detection and identification of topology errors in electric power systems,” IEEE Trans. Power Syst., vol. 3, no. 4, pp. 1748–1753, Nov. 1988.

[13] Y. Liao, Y. Weng, M. Wu, and R. Rajagopal, “Distribution grid topology reconstruction: An information theoretic approach,” in Proc. North Amer. Power Syst., Oct. 2015, pp. 1–6.

[14] Y. Liao, Y. Weng, G. Liu, and R. Rajagopal, “Urban MV and LV distribution grid topology estimation via group Lasso,” IEEE Trans. Power Syst., vol. 34, no. 1, pp. 12–27, Jan. 2019.

[15] G. Cavaro and V. Kekatos, “Graph algorithms for topology identification using power grid probing,” IEEE Control Syst. Lett., vol. 2, no. 4, pp. 689–694, Oct. 2018.

[16] D. Deka, S. Talukdar, M. Chertkov, and M. V. Salapaka, “Topology estimation in bulk power grids: Guarantees on exact recovery,” in Proc. IEEE PES Int. Workshop Bulk Power Syst., 2017, pp. 1–10.

[17] D. Deka, S. Backhaus, and M. Chertkov, “Estimating distribution grid topologies: A graphical learning based approach,” in Proc. Power Syst. Comput. Conf., Jun. 2016, pp. 1–7.

[18] S. Park, D. Deka, and M. Chertkov, “Exact topology and parameter estimation in distribution grids with minimal observability,” in Proc. Power Syst. Comput. Conf., 2018, vol. 34, no. 1, 2018, pp. 9–14.

[19] S. Bolognani, N. Bof, D. Michelotti, R. Muraro, and L. Schenato, “Factorization for power grid topology tracking,” in Proc. 3rd Int. Conf. High Confidence Networked Syst., 2018, pp. 1061–1074, Sep. 2018.

[20] D. Deka, S. Backhaus, and M. Chertkov, “Learning topology of the power distribution grid with and without missing data,” in Proc. Euro. Control Conf., Jun. 2016, pp. 313–320.

[21] X. Li, H. V. Poor, and A. Scaglione, “Blind topology identification for power systems,” in Proc. IEEE Int. Conf. Smart Grid Commun., Oct. 2013, pp. 91–96.

[22] A. Anwar, A. Mahmood, and M. Pickering, “Estimation of smart grid topology using SCADA measurements,” in Proc. IEEE Int. Conf. Smart Grid Commun., Nov. 2016, pp. 539–544.

[23] Y. Yuan, O. Ardakanian, S. H. Low, and C. Tomlin, “On the inverse power flow problem,” 2016, arXiv:1610.06631. [Online]. Available: http://arxiv.org/abs/1610.06631

[24] V. Kekatos, G. B. Giannakis, and R. Baldick, “Online energy price matrix factorization for power grid topology tracking,” IEEE Trans. Smart Grid, vol. 7, no. 3, pp. 1239–1248, May 2016.

[25] S. Xie, J. Yang, K. Xie, Y. Liu, and Z. He, “Low-sparity unobservable attacks against smart grid: Attack exposure analysis and a data-driven attack scheme,” IEEE Access, vol. 5, pp. 8183–8193, 2017.

[26] P. Comon, “Independent component analysis, a new concept?” Signal Process., vol. 36, no. 3, pp. 287–314, Apr. 1994.
R. Gallager, *Stochastic Processes: Theory for Applications* (Series Stochastic Processes: Theory for Applications). Cambridge, U.K.: Cambridge Univ. Press, 2013.

T. W. Anderson and H. Rubin, “Statistical inference in factor analysis,” in *Proc. 3rd Berkeley Symp. Math. Statist. Probability*, 1956, vol. 5, pp. 111–150.

L. Tong, R. W. Liu, V. C. Soon, and Y. F. Huang, “Indeterminacy and identifiability of blind identification,” *IEEE Trans. Circuits Syst.*, vol. 38, no. 5, pp. 499–509, May 1991.

P. Stoica and M. Jansson, “On maximum likelihood estimation in factor analysis: algebraic derivation,” *Signal Process.*, vol. 89, no. 6, pp. 1260–1262, 2009.

S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993, vol. 1.

S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge Univ. Press, 2004.

D. L. Donoho and M. Elad, “Optimally sparse representation in general (non-orthogonal) dictionaries via $l_0$ minimization,” *Proc. Nat. Acad. Sci.* USA, vol. 100, no. 5, pp. 2197–2202, Mar. 2003.

M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok, “Introduction to compressed sensing,” *Compressed Sensing: Theory and Applications*, Y. C. Eldar and G. Kutyniok, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2012.

R. Horn and C. R. Johnson, *Matrix Analysis*. New York, NY, USA: Cambridge Univ. Press, 1985.

T. Zhang, A. Wiesel, and M. S. Greco, “Multivariate generalized Gaussian distribution: Convexity and graphical models,” *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 4141–4148, Aug. 2013.

J. Friedman, T. Hastie, and R. Tibshirani, “Sparse inverse covariance estimation with the graphical lasso,” *Biostatistics*, vol. 9, no. 3, pp. 432–441, 2008.

M. Grant and S. Boyd, “CVX: MATLAB software for disciplined convex programming, version 2.0,” Aug. 2012. [Online]. Available: http://cvxr.com/cvx

W. Lu and J. C. Rajapakse, “Approach and applications of constrained ICA,” *IEEE Trans. Neural Netw.*, vol. 16, no. 1, pp. 203–212, Jan. 2005.

D. P. Bertsekas, Ed., *Constrained Optimization and Lagrange Multiplier Methods*. New York, NY, USA: Academic, 1982.

W. Lu and J. C. Rajapakse, “ICA with reference,” *Neurocomputing*, vol. 69, no. 16, pp. 2244–2257, 2006.

S. Amari, “Natural gradient works efficiently in learning,” *Neural Comput.*, vol. 10, no. 2, pp. 251–276, 1998.

C. R. Rao, *Linear Statistical Inference and Its Applications*, 2nd ed. New York, NY, USA: Wiley, 1973.

Y. Noam and H. Messer, “Notes on the tightness of the hybrid Cramér–Rao lower bound,” *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2074–2084, Jun. 2009.

M. Anderson, G. S. Fu, R. Phlypo, and T. Adali, “Independent vector analysis: Identification conditions and performance bounds,” *IEEE Trans. Signal Process.*, vol. 62, no. 17, pp. 4399–4410, Sep. 2014.

J. D. Gorman and A. O. Hero, “Lower bounds for parametric estimation with constraints,” *IEEE Trans. Inf. Theory*, vol. 36, no. 6, pp. 1285–1301, Nov. 1990.

P. Stoica and B. C. Ng, “On the Cramér–Rao bound under parametric constraints,” *IEEE Signal Process. Lett.*, vol. 5, no. 7, pp. 177–179, Jul. 1998.

E. Nitzan, T. Routtenberg, and J. Tabrikian, “Cramér–Rao bound for constrained parameter estimation using lehmann-unbiasedness,” *IEEE Trans. Signal Process.*, vol. 67, no. 3, pp. 753–768, Feb. 2019.

Z. Ben-Haim and Y. C. Eldar, “The Cramér–Rao bound for estimating a sparse parameter vector,” *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3384–3389, Jun. 2010.

H. E. Egilmez, E. Pavez, and A. Ortega, “Graph learning from data under structural and Laplacian constraints,” *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 6, pp. 825–841, Sep. 2017.

“Power systems test case archive,” 1999. [Online]. Available: http://www.ee.washington.edu/research/pstca/

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.

[81] G. Cavarro and V. Kekatos, “Inverter probing for power distribution network topology processing,” 2018, arXiv:1802.06027. [Online]. Available: https://arxiv.org/abs/1801.03862

Sivan Grotas (S’18) received the B.Sc. degree (magna cum laude) in electrical engineering from the Ben-Gurion University of the Negev, Beer-Sheva, Israel, in 2018, where she is currently working toward the M.Sc. degree in electrical engineering. Her research interests are in the areas of graph signal processing, estimation theory, and optimization methods. She received the Scholarship for Outstanding Students in Engineering and Exact Sciences from the Ministry of Science and Technology, Israel and the ECE conference Travel Grant for B.Sc. students.

Yair Yakoby received the B.Sc. degrees in physics (cum laude) and in electrical engineering (summa cum laude) both from the Ben-Gurion University of the Negev, Beer-Sheva, Israel, in 2018. He is currently in Mellanox Technologies as an Algorithm Engineer with the DSP Group, Israel. His research interests include graph signal processing, optimization methods, and digital communications. He was the recipient of the award for outstanding electrical engineering project of 2017.

Idan Gera received the B.Sc. degrees in physics and (cum laude) in electrical engineering from the Ben-Gurion University of the Negev, Beer-Sheva, Israel, in 2018. He is working on a private venture to facilitate machine-learning algorithms in order to gain insights from large data sets. His research interests include graph signal processing and optimization methods. He was the recipient of the ECE Department Award for Outstanding B.Sc. Projects of 2017 and Award for unique contribution to the university’s sports teams.

Tirza Routtenberg (S’07–M’13–SM’18) received the B.Sc. degree (magna cum laude) in biomedical engineering from the Technion Israel Institute of Technology, Haifa, Israel, in 2005, and the M.Sc. (magna cum laude) and Ph.D. degrees in electrical engineering from the Ben-Gurion University of the Negev, Beer-Sheva, Israel, in 2007 and 2012, respectively. She was a Postdoctoral Fellow with the School of Electrical and Computer Engineering, Cornell University, in 2012–2014. Since October 2014, she has been a Faculty Member with the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer-Sheva, Israel. Her research interests include signal processing in smart grid, statistical signal processing, estimation and detection theory, and signal processing on graphs. She was the recipient of the Best Student Paper Award in International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2011, in 2013 the IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (coauthor), in ICASSP 2017 (coauthor), and in 2018 the IEEE Workshop on Statistical Signal Processing (coauthor). She was awarded the Negev scholarship in 2008, the Lev-Zion scholarship in 2010, and the Marc Rich foundation prize in 2011.