Going Beyond Linear Dependencies to Unveil Connectivity of Meshed Grids

Liang Zhang, Gang Wang, and Georgios B. Giannakis
Digital Tech. Center and ECE Dept., Univ. of Minnesota, Mpls, MN 55455, USA
E-mails: {zhan3523, gangwang, georgios}@umn.edu

Abstract—Partial correlations (PC) are well suited for revealing linearly dependent (un)mediated connections in a graph when measurements (e.g., time courses) are available per node. Unfortunately, PC-based approaches to identifying the topology of a graph are less effective if nonlinear dependencies between given nodal measurements are present. To bypass this hurdle, nonlinear PCs relying on the $\ell_2$-norm regularized multi-kernel ridge regression (MKRR) have been recently proposed for brain network connectivity analysis. However, $\ell_2$-norm regularization limits the flexibility in combining kernels, which can compromise performance. For this reason, the present paper broadens the nonlinear PC approach to account for general $\ell_p$-norm regularized MKRR, in which the user-selected parameter $p \geq 1$ is attuned to the problem at hand. Aiming at a scalable algorithm, the Frank-Wolfe iterations are invoked to solve the $\ell_p$-norm based MKRR, which not only features simple closed-form updates, but it is also fast convergent. The end result is a novel scheme that leverages nonlinear dependencies captured by the generalized PC model to identify the topology of not only radial but also meshed autonomous energy grids. Improved performance is achieved at affordable computational complexity relative to existing alternatives. Simulated tests showcase the merits of the proposed schemes.

I. INTRODUCTION

Partial correlation (PC) statistics offer a well-appreciated measure of similarity between observed time-courses of node pairs that has been widely used for inferring the topology of undirected networks [1]. In contrast to ordinary correlation, PC unveils unmediated relationship between two nodes by linearly regressing out the influence from the others. This has enabled PC-based topology identification of various graphs, including those arising from brain networks [2], and power grids [3].

The resultant linear PC-based algorithm presumes a linear regression model, while the measurements collected in many real-world networks, including the brain and power networks, are often mutually nonlinearly dependent. To better capture such nonlinear dependencies, an $\ell_2$-norm regularized multi-kernel based PC (MKPC) approach was pursued in [2] to enhance accuracy in identifying edges of brain networks. The adopted multi-kernel learning (MKL) technique has been also studied in various machine learning tasks; see e.g., [4] for the $\ell_2$-norm based MKRR as well as [5] for the $\ell_p$-norm regularized MKL. Although the $\ell_p$-norm regularized MKL can lead to improved estimation accuracy, the proposed solver tailored for classification tasks is not scalable with the network size [5].

Specifically for smart grids, identifying the connectivity is of paramount importance since numerous operating, scheduling, and protection tasks require full awareness of the network topology [6]–[8]. The linear PC method as well as its concentration matrix-based variants have been advocated for power grid topology identification [3], [9]. Nevertheless, the recovery performance is suboptimal especially for meshed (loopy) networks because the nodal voltages collected are typically correlated, and voltage angles are nonlinearly dependent in practice. Although the nonlinear relationships between voltage angles are well acknowledged, grid topology learning accounting for nonlinear dependencies remains hitherto an uncharted territory.

Instead of adopting the $\ell_2$-regularization, which is computationally easy to deal with, this contribution first formulates the more challenging task of $\ell_p$-norm based MKPC. Our generalized nonlinear model is expected to offer improved estimators by exploiting the flexibility in determining the combination of kernels. For example, when $p = 1$, only a few kernels from the user-specified dictionary of kernels can be combined to form the nonlinear estimator. Therefore, the overall performance of MKPC-based topology learning can be boosted. Furthermore, an efficient implementation based on the recently revived Frank-Wolfe algorithm [10] is developed for the $\ell_p$-regularized MKRR. Besides guaranteed convergence to a global solution, it also features simple closed-form updates. At last, the performance of our $\ell_p$-norm regularized MKPC is demonstrated in the power grid topology identification application. Numerical experiments using the IEEE 14-bus power benchmark showcase the efficiency of the proposed MKPC solver, and also the improved accuracy of the generalized nonlinear PC method.

Notation. Symbol $\top$ stands for transposition, while vectors $0$ and $1$ denote the all-zero and all-one vectors of appropriate dimensions, respectively. Operator $\langle \mathbf{x}, \mathbf{y} \rangle$ computes the inner product between $\mathbf{x}$ and $\mathbf{y}$, while $\geq$ ($\leq$) is understood entry-wise if used for vectors.

II. $\ell_p$-NORM REGULARIZED MKPC

Consider a network consisting of $N$ nodes that is modeled as a graph $G := (\mathcal{N}, \mathcal{L})$, where $\mathcal{N} := \{1, 2, \ldots, N\}$ denotes the set of all nodes, and $\mathcal{L} := \{(i, j)\} \in \mathcal{N} \times \mathcal{N}$ the set of all edges. Per node $n \in \mathcal{N}$, let $\mathbf{x}_n := [x_n[1], \ldots , x_n[T]]^\top$ be
its associated time-series measurements. For power network topology inference, \( x_n \) may consist of the nodal voltages for all time slots \( t \in \mathcal{T} := \{1, 2, \ldots, T\} \) [3]; while for brain connectivity analysis, it may correspond to the blood-oxygen-level dependent signals [2]. The task of grid connectivity identification can be described as follows: given nodal time-series \( \{x_n\}_{n \in \mathcal{N}} \), infer the network topology, namely recover the edge set \( \mathcal{L} \).

To reveal nodal connectivity, the PC method, which is effective in capturing unmediated linear influence between nodes, has been widely used [1]–[3]. Letting \( \mathcal{S} \) denote all but the \((i, j)\) nodes, consider the error vector \( \hat{x}_{ij} = x_i - \hat{x}_{ij} \), where \( \hat{x}_{ij} \) is the estimate of \( x_i \) based on \( \{x_n\}_{n \in \mathcal{S}} \). The empirical PC coefficient between \( x_i \) and \( x_j \) is given by [1]

\[
\hat{\rho}_{ij|\mathcal{S}} := \frac{(x_{ij} - \bar{x}_{ij})^T(x_{ij} - \bar{x}_{ij})}{\|x_{ij} - \bar{x}_{ij}\|_2^2} \tag{1}
\]

where \( \bar{x}_{ij} := (1/T) \sum_{t=1}^{T} x_{ij}[t] \). Having computed all PC coefficients \( \{\hat{\rho}_{ij|\mathcal{S}}\}_{(i,j) \in \mathcal{N} \times \mathcal{N}} \), determining whether node \( i \) is connected with node \( j \) entails a hypothesis test. Specifically, an edge between \( i \) and \( j \) is declared present if \( |\hat{\rho}_{ij|\mathcal{S}}| \geq \tau \), where \( \tau > 0 \) is a threshold that trades off the relative true positive for the false positive decisions.

The linear PC method assumes that \( x_{ij} \) is a linear function of \( \{x_n\}_{n \in \mathcal{S}} \). Nonetheless, the dependence of \( x_i \) on \( \{x_n\}_{n \in \mathcal{S}} \) is oftentimes nonlinear in many applications. To circumvent the limitation of linear estimators, an \( \ell_2 \)-norm based MKPC has been studied in [2], which works as follows. Let \( \{n_{11}, \ldots, n_{N-2,i}\} \) enumerate all nodes in \( \mathcal{S} \), and \( x_{\setminus i}[t] := [x_{n_{11},i}[t], \ldots, x_{n_{N-2,i},i}[t]]^T \) collect all measurements at time \( t \). By replacing \( x_{\setminus i}[t] \) with its lifted image via a feature map \( \Phi(x_{\setminus i}[t]) \), a nonlinear data generation model can be postulated as

\[
x_i[t] = \langle \Phi(x_{\setminus i}[t]), \beta \rangle + \epsilon_i[t] \tag{2}
\]

where \( \beta \) is a parameter vector to learn, and \( \epsilon_i[t] \) captures modeling inaccuracies. Along the lines of ridge regression, an estimate of \( \beta_i \) per node \( i \) can be found as

\[
\hat{\beta}_i := \arg \min_{\beta} \frac{C}{N} \|\xi_i\|_2^2 + \frac{1}{2} \|\beta\|_2^2 \tag{3a}
\]

subject to \( \xi_i = x_i - \Phi_{\setminus i}^T \beta \) \tag{3b}

where \( \Phi_{\setminus i} := [\Phi(x_{\setminus i}[1]), \ldots, \Phi(x_{\setminus i}[T])] \), and \( C \geq 0 \) is some given constant. Since \( \Phi(x_{\setminus i}[t]) \) has high (potentially infinite) dimension, the dual of (3) which has only \( T \) variables (as many as the number of scalar equality constraints in (3b)) will be used henceforth. Specifically, the dual of (3) can be succinctly written as [11]

\[
\max_{\alpha \in \mathbb{R}^T} \frac{1}{2} \left( -\mu \alpha^T + 2\alpha^T x_i - \alpha^T K_{\setminus i} \alpha \right) \tag{4}
\]

where \( \alpha \in \mathbb{R}^T \) denotes the Lagrange multiplier associated with (3b), constant \( \mu := \frac{C}{2N} \), and \( K_{\setminus i} := \Phi_{\setminus i}^T \Phi_{\setminus i} \). Note that the maximizer of (4) can be found in closed form as \( \hat{\alpha}_i = (K_{\setminus i} + \mu I_N)^{-1} x_i \). Appealing to [2, Eq. (4)], the least-squares estimate of \( x_i[t] \) can be obtained as

\[
\hat{x}_{ij}[t] = \langle \Phi(x_{\setminus i}[t]), \hat{\beta}_i \rangle = \langle \Phi(x_{\setminus i}[t]), \Phi_{\setminus i} \hat{\alpha}_i \rangle \tag{5}
\]

which entails computing inner products between high dimensional feature vectors \( \{\langle \Phi(x_{\setminus i}[t]), \Phi(x_{\setminus i}[t']) \rangle\} \). Fortunately, such a costly computation can be significantly reduced by invoking the so-called kernel trick [12], which allows computing the wanted inner products in (5) through solely evaluating the kernel function for all pairs \((t, t') \in \mathcal{T} \times \mathcal{T} \):

\[
\kappa(x_{\setminus i}[t], x_{\setminus i}[t']) = \langle \Phi(x_{\setminus i}[t]), \Phi(x_{\setminus i}[t']) \rangle \tag{6}
\]

Clearly, the accuracy of the kernel-based estimates in (5) depends highly on the selected kernel function \( \kappa(\cdot, \cdot) \) [12]. To choose a suitable kernel, MKL is invoked here [13], which seeks \( \kappa(\cdot, \cdot) \) as a nonnegative linear combination of user-defined kernel functions; that is, \( \kappa(\cdot, \cdot) := \sum_{m=1}^{M} \theta_m \kappa_m(\cdot, \cdot) \). The coefficients \( \{\theta_m \geq 0\}_{m=1}^M \) can be deciphered from data via solving [13]

\[
\theta^* := \arg \min_{\theta \in \Theta_p} \max_{\alpha \in \mathbb{R}^N} -\mu \alpha^T + 2\alpha^T x_i - \sum_{m=1}^{M} \theta_m \alpha^T K_{m\setminus i} \alpha \tag{7}
\]

where the kernel matrix is given by

\[
K_{m\setminus i}(t, t') := \kappa_m(x_{\setminus i}[t], x_{\setminus i}[t']), \forall (t, t') \in \mathcal{T} \times \mathcal{T} \tag{8}
\]

the constraint set \( \Theta_p := \{\theta \in \mathbb{R}^M | \theta \geq 0, \|\theta\|_p \leq \Lambda\} \) with \( p \geq 1 \), and \( \Lambda > 0 \) is a pre-selected constant. For notational brevity, define also

\[
F(\theta) := \max_{\alpha \in \mathbb{R}^N} -\mu \alpha^T + 2\alpha^T x_i - \sum_{m=1}^{M} \theta_m \alpha^T K_{m\setminus i} \alpha \tag{9}
\]

based on which we can rewrite (6) as follows

\[
\theta^* := \arg \min_{\theta \in \Theta_p} F(\theta). \tag{10}
\]

Upon obtaining \( \theta^* \) and \( K_{\setminus i} = \sum_{m=1}^{M} \theta_m K_{m\setminus i} \), it is easy to obtain the estimate (cf. (5))

\[
\hat{x}_{ij} = K_{ij}(\mu I + K_{\setminus i})^{-1} x_i. \tag{11}
\]

Our new formulation here generalizes the nonlinear estimator in [2] beyond \( p = 2 \). This generalization is well motivated because the estimation error \( \|x_i - \hat{x}_{ij}\|_2 \) can be reduced by choosing appropriate \( p \) values [5].

### III. EFFICIENT SOLVER FOR \( \ell_p \)-NORM BASED MKRR

Though the \( \ell_p \)-norm based MKL (9) can lead to improved estimation accuracy, solving the resulting \( \ell_p \)-norm constrained optimization problem is typically not an easy task except for a few special cases such as \( p = 1 \) or 2 [5], [14]. For this reason, we put forward an efficient and easy-to-implement solver by leveraging the projection-free Frank-Wolfe algorithm (a.k.a.,
the conditional gradient method) [10]. As a brief review, the Frank-Wolfe algorithm targets the generic convex problem
\[
y^* = \arg \min_{y \in \mathcal{Y}} f(y)
\]
where \( f \) is assumed differentiable, and the feasible set \( \mathcal{Y} \) is compact. The Frank-Wolfe solver starts with an arbitrary point \( y^0 \), and subsequently iterates between the updates for \( k \geq 0 \) as follows
\[
s^k \in \arg \min_{s \in \mathcal{Y}} s^T \nabla f(y^k)
\]
\[
y^{k+1} := y^k + \eta_k (s^k - y^k)
\]
where the diminishing step size \( \eta_k := 2/(k^p + 2) \) with \( p \) being a predefined constant in \( (0.5, 1] \) [15]. It has been shown that the successive iterates \( \{y^k\} \) are always feasible due to \( y^0 = 1, y^1 = s^0 \in \mathcal{Y} \), and \( s^k \in \mathcal{Y} \) for \( k \geq 1 \) [10].

Since \( F(\theta) \) in (9) is convex and differentiable [14], and \( \Theta_p \) is compact, the \( \ell_p \)-norm regularized MKRR (9) is in the form of (11) required by the Frank-Wolfe iterations (12). To employ (12a), the gradient of \( F(\theta) \) is given by
\[
\nabla F(\theta) = [-\alpha^T K_{1} i,j \alpha, \ldots, -\alpha^T K_{M}\alpha] \cdot \alpha
\]
with
\[
\alpha = \left( \mu I + \sum_{m=1}^{M} \theta^m K_{m} \right)^{-1} x_i.
\]
Since \( \{K_{m}\}_{m=1}^{M} \) are positive semidefinite [12], we have \( \nabla F(\theta) \leq 0, \forall \theta \in \Theta_p \). Applying (12a) to (9) yields
\[
s^k \in \arg \min_{s \in \Theta_p} s^T \nabla F(\theta^k)
\]
which we will show possible to express in closed form. Indeed, upon introducing \( g^k := -\nabla F(\theta^k) \), (15) can be equivalently posed as
\[
s^k \in \arg \max_{s \in \Theta_p} s^T g^k.
\]
By Hölder inequality and for all \( s^k \in \Theta_p \), it holds that
\[
(s^k)^T g^k \leq \|s^k\|_p \|g^k\|_q \leq \Lambda \|g^k\|_q
\]
where the norms satisfy \( 1/p + 1/q = 1 \). Using further \( g^k = -\nabla F(\theta^k) \geq 0 \), it can be readily deduced that the solution to (15) is found when equalities in (17) are attained
\[
s^k + \lambda \nabla g^k \leq \|g^k\|_q
\]
where \( q := p/(p-1) \), with \( q = \infty \), if \( p = 1 \). On the other hand, \( q = 1 \), if \( p = \infty \), and in this case, all nonzero entries of \( s^k \) are equal. Upon finding the conditional direction \( s^k \), the algorithm updates the estimate via (12b), and iterates until certain convergence criterion is met.

The proposed solver for the \( \ell_p \)-regularized MKL (9) is summarized as Alg. 1. Notably, with step size \( \eta_k = 2/(k^p + 2) \), it converges to \( \theta^* \) at sublinear rate [10], namely
\[
F(\theta^k) - F(\theta^*) \leq O\left(1/k\right).
\]
It is also worth stressing that applicability of Alg. 1 goes beyond MKPC, since any \( \ell_p \)-norm based MKRR can be solved via Alg. 1.

### IV. TOPOLOGY LEARNING FOR MESHED NETWORKS

Consider a meshed power network of \( N \) buses indexed by the set \( \mathcal{N} := \{1, 2, \ldots, N\} \). Let \( \psi[t] := [\psi_1[t], \ldots, \psi_N[T]]^T \) collect the voltage angles at all buses at time \( t \). The power grid topology learning problem can be stated as follows: given voltage angle time-series measurements \( \{\psi[t]\}_{t=1}^{T} \), identify the grid topology. Expressed differently, our goal is to infer the connectivity of buses based on \( T \) voltage angle-vector measurements.

For any bus pair \( (i,j) \in \mathcal{N} \times \mathcal{N} \), define \( x_i := [\psi_1[i], \ldots, \psi_T[i]]^T \) and \( x_j := [\psi_1[j], \ldots, \psi_T[j]]^T \). Having selected candidate kernel functions \( \{\kappa_m\}_{m=1}^{M} \), one first forms the kernel matrices \( \{K_{m}\}_{m=1}^{M} \) as in (7), and learns the best kernel combination \( K_{i,j} = \sum_{m=1}^{M} \theta^m_{i,m} K_{m} \) using Alg. 1. For both \( x_i \) and \( x_j \), the next step is to obtain the \( \ell_p \)-norm MKRR based nonlinear estimates \( \hat{x}_{ij} \). The estimated PC coefficient \( \rho_{ij} \) can then be found by means of (1), and an edge connecting buses \( i \) and \( j \) is claimed to be present if \( |\hat{\rho}_{ij}| > \tau \). This pairwise hypotheses test is repeated for all pairs \( (i,j) \in \mathcal{N} \times \mathcal{N} \). The proposed \( \ell_p \)-regularized MKRR based network topology inference scheme is tabulated in Alg. 2.

### V. NUMERICAL TESTS

This section evaluates the performances of Algorithms 1 and 2 using the voltage angle data from the IEEE 14-bus benchmark [16]. To test the developed methods in a more realistic setting, the real load data from 2012 Global Energy Forecasting Competition [17] were used in all experiments. Specifically, the first ten-day loads of zones 1 – 14 were normalized to match the scale of active power demands in the
IEEE 14-bus system, which is further corrupted by random noise generated from a uniform distribution on $[0, 1]$. Voltage angle measurements $\{\psi(t)\}^{T}_{t=1}$ across $T = 240$ time slots were then obtained by solving the AC power flow equations using the publicly available MATPOWER toolbox [18]. To obtain a nonlinear voltage angle estimator, a total of 20 kernels were employed to form the dictionary, which consists of 10 polynomial kernels whose orders vary by 1 from 1 to 10, as well as 10 Gaussian kernels with variances distributed uniformly from 0.5 to 5. The regularization coefficients in (6) were set as $\mu = 1$ and $\Lambda = 3$.

The first experiment assesses the convergence and computational performance of Alg. 1. To serve as a benchmark, (9) is first equivalently reformulated as a semidefinite program (SDP) by introducing an auxiliary variable $w \in \mathbb{R}$ such that $w \geq F(\theta)$, whose global optimum is found by SeDuMi [19]; see also [14]

$$\theta^* := \arg \min_{\theta \in \Theta, w \geq 0} w$$

subject to $$\mu I + \sum_{m=1}^{M} \theta_m K_m x_i w \geq 0. \quad (19)$$

Figure 1 depicts the evolution of the relative error $(F(\theta^k) - F(\theta^*))/\theta^*$ of (9) with $p = 1.5$ and $p = 2$, where $\{x_n\}^{21}_{n=1}$ and $\{x_m\}^{14}_{m=5}$ were used to predict $x_3$. Algorithm 1 with $\eta_k = 2/(k + 2)$ was run until $\|\hat{\theta}^k - \hat{\theta}\|_2 < 10^{-5}$, or the maximum iteration number 100 was reached. It can be observed that Alg. 1 converges to the global minimum roughly at rate $O(1/k)$.

The computation times for Alg. 1 and SeDuMi are 0.23 and 9.63 secs ($p = 1.5$); and 0.20 and 7.49 secs ($p = 2$). It is clear that Alg. 1 enjoys a clear speedup advantage over the off-the-shelf solver. This advantage is mainly due to the simple closed-form updates of (18), and it is important for the scalability of Alg. 2 which requires solving (9) $N^2 - N$ times.

The second experiment tests the topology recovery performance of Alg. 2 using the IEEE 14-bus system. Note that including the reference bus renders the covariance matrix of voltage angles non-invertible since the voltage angle at the reference bus remains always 0. Excluding the reference bus, a total of 13 buses and 18 lines are to be recovered. The performance of Alg. 2 with $p = 1.5$ and $p = 2$ was compared to the performance of linear PC- and concentration matrix-based methods [9] in terms of the empirical receiver operating characteristics (ROC) in Fig. 2. For Alg. 2 and its linear PC counterpart, the ROC curves were obtained using $||P_{ij}|(\cdot)\|_{(i,j) \in \mathbb{N}^2 \times \mathbb{N}^2}$ as test statistics, whereas for the concentration matrix-based method, entries of the negative concentration matrix were employed as test statistics. The area under the curve (AUC) for Alg. 2 with $p = 1.5$ and $p = 2$, linear PC-, and concentration matrix-based methods is 0.755, 0.743, 0.646, and 0.604, accordingly. This shows improved recovery performance of the proposed Alg. 2 relative to existing alternatives. Comparing the red and black curves in Fig. 2 further suggests that the grid topology identification can benefit from the $\ell_p$-regularized MKPC with $p$ values different from 2.

**VI. CONCLUSIONS**

Going beyond linear dependencies, the present paper advocated topology identification of meshed grids by accounting for nonlinear relationships present among time-courses of power-related quantities measured at different buses. These dependencies were captured through nonlinear partial correlations that rely on the multi-kernel based nonlinear functions of the aforementioned time-courses obtained via $\ell_p$-norm regularized regressions. For efficient implementation, the projection-free Frank-Wolfe iterations were invoked to solve the emerging $\ell_p$-norm constrained optimization problems. The novel solver features closed-form updates and guaranteed fast convergence. The performance of the generalized nonlinear model was demonstrated using the power grid topology inference application. Simulated tests on the IEEE 14-bus benchmark showcase improved computational efficiency of the projection-free solver, and increased identification accuracy of the proposed grid topology inference scheme.
REFERENCES


