Channel Gain Cartography for Cognitive Radios Leveraging Low Rank and Sparsity

Donghoon Lee, *Student Member, IEEE*, Seung-Jun Kim, *Senior Member, IEEE*, and Georgios B. Giannakis, *Fellow, IEEE*

***Abstract*— Channel gain cartography aims at inferring the channel gains between two arbitrary points in space based on the measurements (samples) of the gains collected by a set of radios deployed in the area. Channel gain maps are useful for various sensing and resource allocation tasks essential for the operation of cognitive radio networks. In this paper, the channel gains are modeled as the tomographic accumulations of an underlying spatial loss field (SLF), which captures the attenuation in the signal strength due to the obstacles in the propagation path. In order to estimate the map accurately with a relatively small number of measurements, the SLF is postulated to have a low-rank structure possibly with sparse deviations. Efficient batch and online algorithms are derived for the resulting map reconstruction problem. Comprehensive tests with both synthetic and real data sets corroborate that the algorithms can accurately reveal the structure of the propagation medium, and produce the desired channel gain maps.**

***Index Terms*— Channel gain cartography, cognitive radio, low rank and sparse models, RF tomography.**

I. Introduction

RECENTLY, it has been recognized that the licensed RF spectrum is often severely under-utilized depending on the time and location of communication, in spite of the evident scarcity of the spectral resources due to the growing use of wireless devices [2]. Cognitive radios (CRs) aim to mitigate this issue by opportunistically utilizing the unused licensed spectrum through spectrum sensing and dynamic spectrum access. RF cartography is an instrumental concept for such CR tasks [3]. Based on the measurements collected by spatially distributed CR sensors, RF cartography constructs the maps over the space, time, and frequency, portraying the RF landscape in which the CR network is deployed. Notable RF maps that have been proposed include the *power spectral density (PSD) maps*, which acquire the ambient interference power distribution, revealing the crowded regions that CR transceivers need to avoid [4]; and the *channel gain (CG) maps*, which capture the channel gains between any two points in space, allowing CR networks to perform accurate spectrum sensing and aggressive spatial reuse [5].

The present work focuses on channel gain cartography. Prior works capitalized on experimentally validated notion of a spatial loss field (SLF) [6], which expresses the shadow fading over an arbitrary link as the weighted integral of the underlying attenuation that the RF propagation experiences due to the blocking objects in the path. Linear interpolation techniques such as kriging were employed to estimate the shadow fading based on spatially correlated measurements, and the spatio-temporal dynamics were tracked using Kalman filtering approaches [5], [7]. It is worth noting that SLF reconstruction is tantamount to the radio tomographic imaging (RTI), useful in a wide range of applications, from locating survivors in rescue operations to environmental monitoring [8]–[10]. The method in [8] captures the variation of the propagation medium by taking SLF differences at consecutive time slots into consideration. To cope with multipath fading in a cluttered environment, multiple channel measurements were utilized to enhance localization accuracy in [11]. However, the methods in [8] and [11] do not reveal static objects in the imaging area. In contrast, a method to track moving objects using a dynamic SLF model, as well as identifying the static ones, was reported in [10]. Exploiting the sparse occupancy of the monitored area by the target objects, sparsity-leveraging algorithms for constructing obstacle maps were developed [12]–[14]. Our work adopts a related data model, but mainly focuses on the channel gain map construction for CR applications.

Although more sophisticated methodologies for channel modeling do exist [15], [16], the computational cost and requirements on various structural/geometric prior information may hinder their use in CR applications. On the other hand, the SLF model has been experimentally validated [6], as well as in the present work through a real tomographic imaging example. Our proposed approach provides a computationally efficient solution, by capitalizing on the inherent structure of measurement data, rather than relying heavily on the physics of RF propagation.

Our work interpolates the channel gains based on the SLF reconstructed from a small number of measurements using a low-rank and sparse matrix model. The key idea is to

Manuscript received December 13, 2016; revised April 12, 2017; accepted June 9, 2017. Date of publication June 23, 2017; date of current version September 8, 2017. This work was supported in part by NSF under Grant 1247885, Grant 1343248, Grant 1442686, and Grant 1547347, and in part by ARO under Grant W911NF-15-1-0492. This paper was presented at the 48th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, November 2–5, 2014 [1]. The associate editor coordinating the review of this paper and approving it for publication was P. Mähönen. *(Corresponding author: Georgios B. Giannakis.)*

D. Lee and G. B. Giannakis are with the Department of Electrical and Computer Engineering and the Digital Technology Center, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: leex6962@umn.edu; georgios@umn.edu).

S.-J. Kim is with the Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, MD 21250 USA (e-mail: sjkim@umbc.edu).

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

Digital Object Identifier 10.1109/TWC.2017.2717822

postulate that the SLF has a low-rank structure potentially corrupted by sparse outliers. Such a model is particularly appealing for urban and indoor propagation scenarios, where regular placement of buildings and walls renders a scene inherently of low rank, while sparse outliers can pick up the artifacts that do not conform to the low-rank model. While it is true that urban and indoor environments have distinct profiles due to the different scales and density of obstacles, our data model can capture the *structural regularity* of obstacles, possibly at different scales, as validated through synthetic and real data examples in Section V. The sparse term helps robustify this model by filtering out the measurements that do not conform to the low-rank structure. This is essentially the idea behind robust principal component analysis [17], which is a powerful data model that has been used widely.

In fact, since the shadow fading samples are modeled as linear tomographic measurements of the SLF, the map recovery task reduces to an instance of compressive principal component pursuit (CPCP) [18]. In general, the CPCP problem recovers the low-rank and sparse matrices from a small set of linearly projected measurements. Our algorithms are applicable to this general problem class.

We develop efficient batch and online algorithms for the map estimation task, suitable for CR network implementation. By replacing the nuclear norm-based regularizer with a bi-factorization surrogate, a block coordinate descent (BCD) algorithm becomes available to avoid costly singular value decomposition (SVD) per iteration. Although the resulting optimization problem is non-convex, the batch solver can attain the global optimum under appropriate conditions. For the online algorithm, a stochastic successive upper-bound minimization strategy is adopted, leading to a stochastic gradient descent (SGD) update rule, which enjoys low computational complexity. The iterates generated by the online algorithm are provably convergent to the stationary point of the batch problem.

The rest of the paper is organized as follows. In Section II, the system model and problem statement are provided. The map reconstruction problem is formulated and its efficient batch solution method is derived in Section III. An online algorithm is developed through a stochastic approximation, and its convergence is established in Section IV. Results from numerical tests using both synthetic and real datasets are presented in Section V, and the conclusions are offered in Section VI.

Notations: Bold uppercase (lowercase) letters denote matrices (column vectors). Calligraphic letters are used for sets; $\mathbf{I}_n$ is the $n \times n$ identity matrix. $\mathbf{0}_n$ denotes an $n \times 1$ vector of all zeros, and $\mathbf{0}_{n\times n}$ an $n \times n$ matrix of all zeros. Operators $(\cdot)^T$, $\mathrm{tr}(\cdot)$, and $\sigma_i(\cdot)$ represent the transposition, trace, and the i-th largest singular value of a matrix, respectively; $|\cdot|$ is used for the cardinality of a set, and the magnitude of a scalar. $\mathbf{R} \succeq \mathbf{0}$ signifies that $\mathbf{R}$ is positive semidefinite. The ℓ_1-norm of $\mathbf{X} \in \mathbb{R}^{n\times n}$ is $||\mathbf{X}||_1 := \sum_{i,j=1}^n |X_{ij}|$. The ℓ_∞-norm of $\mathbf{X} \in \mathbb{R}^{n\times n}$ is represented by $||\mathbf{X}||_\infty := \max\{|X_{ij}| : i, j = 1, \ldots, n\}$. For two matrices $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{n\times n}$, the matrix inner product $\langle \mathbf{X}, \mathbf{Y}\rangle := \mathrm{tr}(\mathbf{X}^T\mathbf{Y})$. The Frobenius norm of matrix $\mathbf{Y}$ is $||\mathbf{Y}||_F := \sqrt{\mathrm{tr}(\mathbf{Y}\mathbf{Y}^T)}$. The spectral norm of $\mathbf{Y}$ is $||\mathbf{Y}|| := \max_{||\mathbf{x}||_2=1} ||\mathbf{Y}\mathbf{x}||_2$, and $||\mathbf{Y}||_* := \sum_i \sigma_i(\mathbf{Y})$ is the nuclear norm of $\mathbf{Y}$. For a function $h : \mathbb{R}^{m\times n} \to \mathbb{R}$, the directional derivative of h at $\mathbf{X} \in \mathbb{R}^{m\times n}$ along a direction $\mathbf{D} \in \mathbb{R}^{m\times n}$ is denoted as $h'(\mathbf{X}; \mathbf{D}) := \lim_{t\to 0+}[h(\mathbf{X} + t\mathbf{D}) - h(\mathbf{X})]/t$.

II. System Model and Problem Statement

Consider a set of N CRs deployed over a geographical area represented by a two-dimensional plane $\mathcal{A} \subset \mathbb{R}^2$. Let $\mathbf{x}_n^{(t)} \in \mathcal{A}$ denote the position of CR $n \in \{1, 2, \ldots, N\}$ at time t. By exchanging pilot sequences, the CR nodes can estimate the channel gains among them. A typical channel gain between nodes n and n' can be modeled as the product of pathloss, shadowing, and small-scale fading. By averaging out the effect of the small-scale fading, the (averaged) channel gain measurement over a link (n, n') at time t, denoted by $G(\mathbf{x}_n^{(t)}, \mathbf{x}_{n'}^{(t)})$, can be represented (in dB) as

$$G(\mathbf{x}_n^{(t)}, \mathbf{x}_{n'}^{(t)}) = G_0 - \gamma\, 10\log_{10}||\mathbf{x}_n^{(t)} - \mathbf{x}_{n'}^{(t)}|| + s(\mathbf{x}_n^{(t)}, \mathbf{x}_{n'}^{(t)}) \tag{1}$$

where G_0 is the path gain at unit distance; $||\mathbf{x}_n^{(t)} - \mathbf{x}_{n'}^{(t)}||$ is the distance between nodes n and n'; γ is the pathloss exponent; and $s(\mathbf{x}_n^{(t)}, \mathbf{x}_{n'}^{(t)})$ is the attenuation due to the shadow fading. By subtracting the known pathloss component in (1), the noisy shadowing measurement

$$\check{s}(\mathbf{x}_n^{(t)}, \mathbf{x}_{n'}^{(t)}) = s(\mathbf{x}_n^{(t)}, \mathbf{x}_{n'}^{(t)}) + \epsilon(\mathbf{x}_n^{(t)}, \mathbf{x}_{n'}^{(t)}) \tag{2}$$

is obtained, where $\epsilon(\mathbf{x}_n^{(t)}, \mathbf{x}_{n'}^{(t)})$ denotes the measurement noise. Let $\mathcal{M}^{(t)}$ be the set of links, for which channel gain measurements are made at time t, and collect those measurements in vector $\check{\mathbf{s}}^{(t)} \in \mathbb{R}^{|\mathcal{M}^{(t)}|}$. The goal of channel gain cartography is to predict the channel gain between arbitrary points $\mathbf{x}, \mathbf{x}' \in \mathcal{A}$ at time t, based on the known nodal positions $\{\mathbf{x}_n^{(t)}\}$ and the channel gain measurements collected up to time t, that is, $\{\check{\mathbf{s}}^{(\tau)}\}_{\tau=1}^t$ [5], [7].

In order to achieve this interpolation, the structure of shadow fading experienced by co-located radio links will be leveraged. To this end, a variety of correlation models for shadow fading have been proposed [6], [19], [20]. In particular, the models in [5], [6], [9], and [10] rely on the so-termed *spatial loss field* (SLF), which captures the attenuation due to obstacles in the line-of-sight propagation.

Let $f : \mathcal{A} \to \mathbb{R}$ denote the SLF, which captures the attenuation at location $\tilde{\mathbf{x}} \in \mathcal{A}$, and $w(\mathbf{x}, \mathbf{x}', \tilde{\mathbf{x}})$ is the weight function modeling the influence of the SLF at $\tilde{\mathbf{x}}$ to the shadowing experienced by link $\mathbf{x}$–$\mathbf{x}'$. Then, $s(\mathbf{x}, \mathbf{x}')$ is expressed as [21]

$$s(\mathbf{x}, \mathbf{x}') = \int_{\mathcal{A}} w(\mathbf{x}, \mathbf{x}', \tilde{\mathbf{x}}) f(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}}. \tag{3}$$

The *normalized ellipse model* is often used for the weight function, with w taking the form [8]

$$w(\mathbf{x}, \mathbf{x}', \tilde{\mathbf{x}}) := \begin{cases} 1/\sqrt{d(\mathbf{x}, \mathbf{x}')}, & \text{if } d(\mathbf{x}, \tilde{\mathbf{x}}) + d(\mathbf{x}', \tilde{\mathbf{x}}) \\ & \quad < d(\mathbf{x}, \mathbf{x}') + \delta \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where $d(\mathbf{x}, \mathbf{x}') := \|\mathbf{x} - \mathbf{x}'\|$ is the distance between positions $\mathbf{x}$ and $\mathbf{x}'$, and $\delta > 0$ is a tunable parameter. The value of δ is commonly set to half the wavelength to assign non-zero weights only within the first Fresnel zone. The integral in (3) can be approximated by

$$s(\mathbf{x}, \mathbf{x}') \simeq \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} w(\mathbf{x}, \mathbf{x}', \tilde{\mathbf{x}}_{i,j}) f(\tilde{\mathbf{x}}_{i,j}) \tag{5}$$

where $\{\tilde{\mathbf{x}}_{i,j}\}_{i=1,j=1}^{N_x,N_y}$ are the pre-specified grid points over $\mathcal{A}$. Let matrix $\mathbf{F} \in \mathbb{R}^{N_x \times N_y}$ denote the SLF, sampled by the N_x-by-N_y grid. Similarly, the weight matrix $\mathbf{W}_{\mathbf{x}\mathbf{x}'}$ corresponding to link $\mathbf{x}$–$\mathbf{x}'$ is constructed. The shadow fading over link $\mathbf{x}$–$\mathbf{x}'$ in (5) can then be expressed as a linear projection of the SLF given by

$$s(\mathbf{x}, \mathbf{x}') \simeq \langle \mathbf{W}_{\mathbf{x}\mathbf{x}'}, \mathbf{F} \rangle = \mathrm{tr}(\mathbf{W}_{\mathbf{x}\mathbf{x}'}^{\mathcal{T}} \mathbf{F}). \tag{6}$$

The goal is to form an estimate $\hat{\mathbf{F}}^{(t)}$ of $\mathbf{F}^{(t)}$ at time t, based on $\{\mathbf{x}_n^{(t)}\}$ and $\{\check{\mathbf{s}}^{(\tau)}\}_{\tau=1}^{t}$. Once $\hat{\mathbf{F}}^{(t)}$ is obtained, the shadowing and the overall channel gain across any link $\mathbf{x}$–$\mathbf{x}'$ at time t can be estimated via (6) and (1) as

$$\hat{s}(\mathbf{x}^{(t)}, \mathbf{x}'^{(t)}) = \langle \mathbf{W}_{\mathbf{x}\mathbf{x}'}^{(t)}, \hat{\mathbf{F}}^{(t)} \rangle \tag{7}$$

$$\hat{G}(\mathbf{x}^{(t)}, \mathbf{x}'^{(t)}) = G_0 - \gamma\, 10 \log_{10} ||\mathbf{x}^{(t)} - \mathbf{x}'^{(t)}|| + \hat{s}(\mathbf{x}^{(t)}, \mathbf{x}'^{(t)}). \tag{8}$$

The number of unknown $\mathbf{F}^{(t)}$ entries is less than $N_x N_y$, while the number of measurements is $O(tN^2)$, provided that the SLF remains invariant for t slots. If the number of entries to be estimated in $\mathbf{F}^{(t)}$ is larger than the number of measurements, the problem is underdetermined and cannot be solved uniquely. To overcome this and further improve the performance even in the overdetermined cases, a priori knowledge on the structure of $\mathbf{F}^{(t)}$ will be exploited next to regularize the problem.

III. Channel Gain Prediction Using Low Rank and Sparsity

A. Problem Formulation

The low-rank plus sparse structure has been advocated in various problems in machine learning and signal processing [17], [22], [23]. Low-rank matrices are effective in capturing slow variation or regular patterns, and sparsity is instrumental for incorporating robustness against outliers. Inspired by these, we postulate that $\mathbf{F}$ has a low-rank-plus-sparse structure as

$$\mathbf{F} = \mathbf{L} + \mathbf{E} \tag{9}$$

where matrix $\mathbf{L}$ is low-rank, and $\mathbf{E}$ is sparse. This model is particularly attractive in urban or indoor scenarios where the obstacles often possess regular patterns, while the sparse term can capture irregularities that do not conform to the low-rank model.

Redefine $\mathbf{W}_{nn'}^{(t)} := \mathbf{W}_{\mathbf{x}_n \mathbf{x}_{n'}}^{(t)}$ and $\check{s}_{nn'}^{(t)} := \check{s}\ (\mathbf{x}_n^{(t)}, \mathbf{x}_{n'}^{(t)})$ for brevity.[1] Toward estimating $\mathbf{F}^{(t)}$ that obeys (9), consider the cost

$$c^{(t)}(\mathbf{L}, \mathbf{E}) := \frac{1}{2} \sum_{(n,n') \in \mathcal{M}^{(t)}} \left(\langle \mathbf{W}_{nn'}^{(t)}, \mathbf{L} + \mathbf{E} \rangle - \check{s}_{nn'}^{(t)} \right)^2 \tag{10}$$

which fits the shadowing measurements to the model. Then, with T denoting the total number of time slots taking measurements, we adopt the following optimization criterion

$$\text{(P1)} \quad \min_{\mathbf{L},\mathbf{E} \in \mathbb{R}^{N_x \times N_y}} \sum_{\tau=1}^{T} \beta^{T-\tau} \left[c^{(\tau)}(\mathbf{L}, \mathbf{E}) + \lambda ||\mathbf{L}||_* + \mu ||\mathbf{E}||_1 \right] \tag{11}$$

where $\beta \in (0, 1]$ is the forgetting factor that can be optionally put in to weigh the recent observations more heavily. The nuclear norm regularization term promotes a low-rank $\mathbf{L}$, while the ℓ_1-norm encourages sparsity in $\mathbf{E}$. Parameters λ and μ are appropriately chosen to control the effect of these regularizers. Conditions for exact recovery through a related convex formulation in the absence of measurement noise can be found in [18].

Problem (11) is convex, and can be tackled using existing efficient solvers, such as the interior-point method. Once the optimal $\hat{\mathbf{L}}$ and $\hat{\mathbf{E}}$ are found, the desired $\hat{\mathbf{F}}$ is obtained as $\hat{\mathbf{F}} = \hat{\mathbf{L}} + \hat{\mathbf{E}}$. However, the general-purpose optimization packages tend to scale poorly as the problem size grows. Specialized algorithms developed for related problems often employ costly SVD operations iteratively [18]. Furthermore, such an algorithm might not be amenable for an online implementation. Building on [24] and [25], an efficient solution is proposed next with reduced complexity.

B. Efficient Batch Solution

Without loss of generality, consider replacing $\mathbf{L}$ with the low-rank product $\mathbf{P}\mathbf{Q}^{\mathcal{T}}$, where $\mathbf{P} \in \mathbb{R}^{N_x \times \rho}$ and $\mathbf{Q} \in \mathbb{R}^{N_y \times \rho}$, and ρ is a pre-specified overestimate of the rank of $\mathbf{L}$. It is known that (e.g., [25])

$$||\mathbf{L}||_* = \min_{\mathbf{P},\mathbf{Q}} \frac{1}{2} \left(||\mathbf{P}||_F^2 + ||\mathbf{Q}||_F^2 \right) \quad \text{subject to } \mathbf{L} = \mathbf{P}\mathbf{Q}^{\mathcal{T}}. \tag{12}$$

[1] Prompted by [11], the benefit of multi-channel diversity for RTI may be incorporated in the present framework. Suppose K channels $\mathcal{K}_{nn'}^{(t)}$ are available to sensors n and n' at time t, and let $\acute{s}_{nn',k}^{(t)}$ denote the noisy measurement including fading over link $\mathbf{x}_n$–$\mathbf{x}_{n'}$ at t in channel $k \in \mathcal{K}_{nn'}^{(t)}$. Construct a new measurement as $\bar{s}_{nn'}^{(t)} = \phi(\acute{s}_{nn',1}^{(t)}, \acute{s}_{nn',2}^{(t)}, \ldots, \acute{s}_{nn',K}^{(t)})$, where $\phi(\cdot)$ is a channel selection function [11]. By replacing $\check{s}_{nn'}^{(t)}$ in (10) with $\bar{s}_{nn'}^{(t)}$, the multiple channel measurements can be incorporated without altering the method. However, the dynamic channel availability and multi-channel measurements will increase algorithm complexity. On the other hand, it is not clear whether such a multi-channel approach can be adopted for estimating *any channel gain* over multiple frequency bands, and constitutes a future research direction.

Thus, a natural re-formulation of (11) is (see also [24])

$$\text{(P2)} \min_{\mathbf{P},\mathbf{Q},\mathbf{E}} f(\mathbf{P},\mathbf{Q},\mathbf{E})4 := \sum_{\tau=1}^{T}\beta^{T-\tau}\left[c^{(\tau)}(\mathbf{P}\mathbf{Q}^{\mathcal{T}},\mathbf{E})+\frac{\lambda}{2}\left(||\mathbf{P}||_F^2+||\mathbf{Q}||_F^2\right)+\mu||\mathbf{E}||_1\right]. \tag{13}$$

Instead of seeking the N_xN_y entries of $\mathbf{L}$, the factorization approach (13) entails only $(N_x+N_y)\rho$ unknowns, thus reducing complexity and memory requirements significantly when $\rho \ll \min\{N_x,N_y\}$. Furthermore, adoption of the separable Frobenius norm regularizer in (P2) comes with no loss of optimality as asserted in the following lemma.

Lemma 1: If $\{\hat{\mathbf{L}},\hat{\mathbf{E}}\}$ *minimize (P1) and we choose* $\rho \geq$ *rank*$(\hat{\mathbf{L}})$*, then, (P2) is equivalent to (P1) at the minimum.*

Proof: It is clear that the minimum of (P1) is no larger than that of

$$\min_{\mathbf{P},\mathbf{Q},\mathbf{E}} \sum_{\tau=1}^{T}\beta^{T-\tau}\left[c^{(\tau)}(\mathbf{P}\mathbf{Q}^{\mathcal{T}},\mathbf{E})+\lambda||\mathbf{P}\mathbf{Q}^{\mathcal{T}}||_*+\mu||\mathbf{E}||_1\right] \tag{14}$$

since the search space is reduced by the reparameterization $\mathbf{L}=\mathbf{P}\mathbf{Q}^{\mathcal{T}}$ with $\rho \leq \min\{N_x,N_y\}$. Now (12) implies that the minimum of (14) is no larger than that of (P2). However, the inequality is tight since the objectives of (P1) and (P2) are identical for $\mathbf{E}:=\hat{\mathbf{E}}$, $\mathbf{P}:=\hat{\mathbf{U}}\hat{\mathbf{\Sigma}}^{1/2}$, and $\mathbf{Q}:=\hat{\mathbf{V}}\hat{\mathbf{\Sigma}}^{1/2}$, where $\hat{\mathbf{L}}=\hat{\mathbf{U}}\hat{\mathbf{\Sigma}}\hat{\mathbf{V}}^{\mathcal{T}}$ is the SVD. Consequently, (P1) and (P2) have identical costs at the minimum. ■

Although (P1) is a convex optimization problem, (P2) is not. Thus, in general, one can obtain only a *locally* optimal solution of (P2), which may not be the globally optimal solution of (P1). Interestingly, under appropriate conditions, global optimality can be guaranteed for the local optima of (P2), as claimed in the following proposition.

Proposition 1: If $\{\bar{\mathbf{P}},\bar{\mathbf{Q}},\bar{\mathbf{E}}\}$ *is a stationary point of (P2),* $\bar{\beta}:=\sum_{\tau=1}^{T}\beta^{T-\tau}$*, and* $||\tilde{f}(\bar{\mathbf{P}}\bar{\mathbf{Q}}^{\mathcal{T}},\bar{\mathbf{E}})|| \leq \lambda\bar{\beta}$ *with*

$$\tilde{f}(\hat{\mathbf{L}},\hat{\mathbf{E}}) := \sum_{\tau=1}^{T}\beta^{T-\tau}\left[\sum_{(n,n')\in\mathcal{M}^{(\tau)}}\left(\langle\mathbf{W}_{nn'}^{(\tau)},\hat{\mathbf{L}}+\hat{\mathbf{E}}\rangle-\check{s}_{nn'}^{(\tau)}\right)\mathbf{W}_{nn'}^{(\tau)}\right] \tag{15}$$

then $\{\hat{\mathbf{L}}:=\bar{\mathbf{P}}\bar{\mathbf{Q}}^{\mathcal{T}},\hat{\mathbf{E}}:=\bar{\mathbf{E}}\}$ *is a globally optimal solution to (P1).*

Proof: See Appendix A.

A stationary point of (P2) can be obtained through a block coordinate-descent (BCD) algorithm, where the optimization is performed in a cyclic fashion over one of $\{\mathbf{E},\mathbf{P},\mathbf{Q}\}$ with the remaining two variables fixed. In fact, since the term $\mu||\mathbf{E}||_1$ is separable in the individual entries as well, the cyclic update can be stretched all the way up to the individual entries of $\mathbf{E}$ without affecting convergence [26]. The proposed solver entails an iterative procedure comprising three steps per iteration $k=1,2,\ldots$

[S1] Update E:

$$\mathbf{E}[k+1]=\arg\min_{\mathbf{E}}\sum_{\tau=1}^{T}\beta^{T-\tau}\left[c^{(\tau)}(\mathbf{P}[k]\mathbf{Q}^{\mathcal{T}}[k],\mathbf{E})+\mu||\mathbf{E}||_1\right]$$

[S2] Update P:

$$\mathbf{P}[k+1]=\arg\min_{\mathbf{P}}\sum_{\tau=1}^{T}\beta^{T-\tau}\left[c^{(\tau)}(\mathbf{P}\mathbf{Q}^{\mathcal{T}}[k],\mathbf{E}[k+1])+\frac{\lambda}{2}||\mathbf{P}||_F^2\right]$$

[S3] Update Q:

$$\mathbf{Q}[k+1]=\arg\min_{\mathbf{Q}}\sum_{\tau=1}^{T}\beta^{T-\tau}\left[c^{(\tau)}(\mathbf{P}[k+1]\mathbf{Q}^{\mathcal{T}},\mathbf{E}[k+1])+\frac{\lambda}{2}||\mathbf{Q}||_F^2\right].$$

To update each block variable, the cost in (P2) is minimized while fixing the other block variables to their up-to-date iterates.

To detail the update rules, let $\boldsymbol{\mathcal{W}}^{(t)}\in\mathbb{R}^{N_xN_y\times|\mathcal{M}(t)|}$ be a matrix with columns equal to $\text{vec}\left(\mathbf{W}_{nn'}^{(t)}\right)$ for $(n,n')\in\mathcal{M}(t)$, where vec $(\cdot)$ produces a column vector by stacking the columns of a matrix one below the other (unvec$(\cdot)$ denotes the reverse process). Define $\boldsymbol{\mathcal{W}}:=[\sqrt{\beta^{T-1}}\boldsymbol{\mathcal{W}}^{(1)}\ldots\sqrt{\beta^0}\boldsymbol{\mathcal{W}}^{(T)}]$, $\check{\mathbf{s}}:=[\sqrt{\beta^{T-1}}\check{\mathbf{s}}^{(1)\mathcal{T}}\ldots\sqrt{\beta^0}\check{\mathbf{s}}^{(T)\mathcal{T}}]^{\mathcal{T}}$, and $\mathbf{e}:=\text{vec}(\mathbf{E})$. Then, one can write $\sum_{\tau=1}^{T}\beta^{T-\tau}c^{(\tau)}(\mathbf{P}\mathbf{Q}^{\mathcal{T}},\mathbf{E})=\|\boldsymbol{\mathcal{W}}^{\mathcal{T}}\text{vec}(\mathbf{P}\mathbf{Q}^{\mathcal{T}}+\mathbf{E})-\check{\mathbf{s}}\|_2^2$. Let e_l denote the l-th entry of $\mathbf{e}$, and $\mathbf{e}_{-l}$ represent the replica of $\mathbf{e}$ without its l-th entry. Similarly, let $\boldsymbol{\omega}_l^{\mathcal{T}}$ denote the l-th row of the matrix $\boldsymbol{\mathcal{W}}$, and $\boldsymbol{\mathcal{W}}_{-l}$ denote the matrix $\boldsymbol{\mathcal{W}}$ with its l-th row removed. The soft-thresholding function soft_th$(\cdot;\mu)$ is defined as

$$\text{soft\_th}(x;\mu):=\text{sgn}(x)\max\{0,|x|-\mu\}. \tag{16}$$

Minimization in **[S1]** proceeds sequentially over the individual entries of $\mathbf{e}$. At iteration k, each entry is updated via

$$e_l[k+1]=\arg\min_{e_l}\frac{1}{2}||e_l\boldsymbol{\omega}_l-\check{\mathbf{s}}||_2^2+\mu\bar{\beta}|e_l|,\ l=1,\ldots,N_xN_y \tag{17}$$

where $\check{\mathbf{s}}_l[k]:=\check{\mathbf{s}}-\boldsymbol{\mathcal{W}}^{\mathcal{T}}\text{vec}\,(\mathbf{P}[k]\mathbf{Q}^{\mathcal{T}}[k])-\boldsymbol{\mathcal{W}}_{-l}^{\mathcal{T}}\mathbf{e}_{-l}$. The closed-form solution for e_l is obtained as

$$e_l[k+1]=\frac{\text{soft\_th}(\boldsymbol{\omega}_l^{\mathcal{T}}\check{\mathbf{s}}_l[k];\mu\bar{\beta})}{||\boldsymbol{\omega}_l||_2^2}. \tag{18}$$

Matrices $\mathbf{P}$ and $\mathbf{Q}$ are similarly updated over their rows through **[S2]** and **[S3]**. Let $\mathbf{p}_i$ be the i-th row of $\mathbf{P}$, transposed to a column vector; i.e., $\mathbf{P}:=[\mathbf{p}_1,\mathbf{p}_2,\ldots,\mathbf{p}_{N_x}]^{\mathcal{T}}$. Define $\tilde{\boldsymbol{\mathcal{W}}}_i^{(t)}\in\mathbb{R}^{|\mathcal{M}(t)|\times N_y}$ to be the matrix whose rows are the i-th rows of $\{\mathbf{W}_{nn'}^{(t)}\}_{(n,n')\in\mathcal{M}(t)}$ denoted as $\tilde{\mathbf{w}}_{nn',i}^{(t)\mathcal{T}}$, and $\tilde{\mathbf{s}}_i^{(t)}\in\mathbb{R}^{|\mathcal{M}(t)|}$ a vector with entries equal to

$$\tilde{s}_{nn',i}^{(t)}:=\check{s}_{nn'}^{(t)}-\langle\mathbf{W}_{nn'}^{(t)},\mathbf{E}[k+1]\rangle-\sum_{j\neq i}^{N_x}\tilde{\mathbf{w}}_{nn',j}^{(t)\mathcal{T}}\mathbf{Q}[k]\mathbf{p}_j \tag{19}$$

for $(n,n')\in\mathcal{M}(t)$. Define also $\tilde{\boldsymbol{\mathcal{W}}}_i:=[\sqrt{\beta^{T-1}}\tilde{\boldsymbol{\mathcal{W}}}_i^{(1)\mathcal{T}}\ldots\sqrt{\beta^0}\tilde{\boldsymbol{\mathcal{W}}}_i^{(T)\mathcal{T}}]^{\mathcal{T}}$ and $\tilde{\mathbf{s}}_i:=[\sqrt{\beta^{T-1}}\tilde{\mathbf{s}}_i^{(1)\mathcal{T}}\ldots\sqrt{\beta^0}\tilde{\mathbf{s}}_i^{(T)\mathcal{T}}]^{\mathcal{T}}$. Then, $\mathbf{p}_i$ is updated by solving

TABLE I
BATCH SOLVER OF (P2)

Initialize $\mathbf{E}[1] := \mathbf{0}_{N_x \times N_y}$, $\mathbf{P}[1]$ and $\mathbf{Q}[1]$ at random.
For $k = 1, 2, \ldots$
  [S1] *Update* $\mathbf{E}$*:*
  Set $\mathbf{e} = \mathrm{vec}(\mathbf{E}[k])$
  For $l = 1, 2, \ldots, N_x N_y$
    Set $\check{\mathbf{s}}_l[k] := \check{\mathbf{s}} - \boldsymbol{\mathcal{W}}^{\mathcal{T}} \mathrm{vec}(\mathbf{P}[k]\mathbf{Q}^{\mathcal{T}}[k]) - \boldsymbol{\mathcal{W}}^{\mathcal{T}}_{-l}\mathbf{e}_{-l}$
    $e_l[k+1] = \text{soft\_th}(\boldsymbol{\omega}_l^{\mathcal{T}} \check{\mathbf{s}}_l[k]; \mu\bar{\beta})/||\boldsymbol{\omega}_l||_2^2$
  Next l
  Set $\mathbf{E}[k+1] = \mathrm{unvec}(\mathbf{e}[k+1])$
  [S2] *Update* $\mathbf{P}$*:*
  For $i = 1, 2, \ldots, N_x$
    Set $\tilde{\boldsymbol{\mathcal{W}}}_i$ and $\tilde{\mathbf{s}}_i$
    $\mathbf{p}_i[k+1] = \left[\mathbf{Q}^{\mathcal{T}}[k]\tilde{\boldsymbol{\mathcal{W}}}_i^{\mathcal{T}}\tilde{\boldsymbol{\mathcal{W}}}_i\mathbf{Q}[k] + \lambda\bar{\beta}\mathbf{I}_\rho\right]^{-1}(\mathbf{Q}^{\mathcal{T}}[k]\tilde{\boldsymbol{\mathcal{W}}}_i^{\mathcal{T}}\tilde{\mathbf{s}}_i)$
  Next i
  $\mathbf{P}[k+1] = [\mathbf{p}_1[k+1], \mathbf{p}_2[k+1], \ldots, \mathbf{p}_{N_x}[k+1]]^{\mathcal{T}}$
  [S3] *Update* $\mathbf{Q}$*:*
  For $i = 1, 2, \ldots, N_y$
    Set $\check{\boldsymbol{\mathcal{W}}}_i$ and $\check{\mathbf{s}}_i$
    $\mathbf{q}_i[k+1] = \left[\mathbf{P}^{\mathcal{T}}[k+1]\check{\boldsymbol{\mathcal{W}}}_i^{\mathcal{T}}\check{\boldsymbol{\mathcal{W}}}_i\mathbf{P}[k+1] + \lambda\bar{\beta}\mathbf{I}_\rho\right]^{-1} \times(\mathbf{P}^{\mathcal{T}}[k+1]\check{\boldsymbol{\mathcal{W}}}_i^{\mathcal{T}}\check{\mathbf{s}}_i)$
  Next i
  $\mathbf{Q}[k+1] = [\mathbf{q}_1[k+1], \mathbf{q}_2[k+1], \ldots, \mathbf{q}_{N_y}[k+1]]^{\mathcal{T}}$
Next k

a ridge-regression problem as

$$\mathbf{p}_i[k+1] = \arg\min_{\mathbf{p}_i}\left[\frac{1}{2}||\tilde{\boldsymbol{\mathcal{W}}}_i\mathbf{Q}[k]\mathbf{p}_i - \tilde{\mathbf{s}}_i||_2^2 + \frac{\lambda\bar{\beta}}{2}||\mathbf{p}_i||_2^2\right]$$

whose solution is given in closed form by

$$\mathbf{p}_i[k+1] = \left[\mathbf{Q}^{\mathcal{T}}[k]\tilde{\boldsymbol{\mathcal{W}}}_i^{\mathcal{T}}\tilde{\boldsymbol{\mathcal{W}}}_i\mathbf{Q}[k] + \lambda\bar{\beta}\mathbf{I}_\rho\right]^{-1}\mathbf{Q}^{\mathcal{T}}[k]\tilde{\boldsymbol{\mathcal{W}}}_i^{\mathcal{T}}\tilde{\mathbf{s}}_i \tag{20}$$

which involves matrix inversion of dimension only ρ-by-ρ. Likewise, let $\mathbf{q}_i$ denote the i-th row of $\mathbf{Q}$, transposed to a column vector; i.e., $\mathbf{Q} := [\mathbf{q}_1, \ldots, \mathbf{q}_{N_y}]^{\mathcal{T}}$. Define also $\check{\boldsymbol{\mathcal{W}}}_i := [\sqrt{\beta^{T-1}}\check{\boldsymbol{\mathcal{W}}}_i^{(1)\mathcal{T}} \ldots \sqrt{\beta^0}\check{\boldsymbol{\mathcal{W}}}_i^{(T)\mathcal{T}}]^{\mathcal{T}}$ and $\check{\mathbf{s}}_i := [\sqrt{\beta^{T-1}}\check{\mathbf{s}}_i^{(1)\mathcal{T}} \ldots \sqrt{\beta^0}\check{\mathbf{s}}_i^{(T)\mathcal{T}}]^{\mathcal{T}}$, where $\check{\boldsymbol{\mathcal{W}}}_i^{(t)} \in \mathbb{R}^{|\mathcal{M}(t)| \times N_x}$ is the matrix whose rows are the transpositions of the i-th columns of $\{\mathbf{W}_{nn'}^{(t)}\}_{(n,n')\in\mathcal{M}(t)}$, denoted as $\check{\mathbf{w}}_{nn',i}^{(t)}$, and $\check{\mathbf{s}}_i^{(t)} \in \mathbb{R}^{|\mathcal{M}(t)|}$ has entries

$$\check{s}_{nn',i}^{(t)} := \check{s}_{nn'}^{(t)} - \langle\mathbf{W}_{nn'}^{(t)}, \mathbf{E}[k+1]\rangle - \sum_{j\neq i}^{N_y}\check{\mathbf{w}}_{nn',j}^{(t)\mathcal{T}}\mathbf{P}[k+1]\mathbf{q}_j \tag{21}$$

for $(n, n') \in \mathcal{M}(t)$. The update for $\mathbf{q}_i$ is then given by solving another ridge regression problem to obtain

$$\mathbf{q}_i[k+1] = \arg\min_{\mathbf{q}_i}\left[\frac{1}{2}||\check{\boldsymbol{\mathcal{W}}}_i\mathbf{P}[k+1]\mathbf{q}_i - \check{\mathbf{s}}_i||_2^2 + \frac{\lambda\bar{\beta}}{2}||\mathbf{q}_i||_2^2\right]$$

whose solution is given also in closed form by

$$\mathbf{q}_i[k+1] = \left[\mathbf{P}^{\mathcal{T}}[k+1]\check{\boldsymbol{\mathcal{W}}}_i^{\mathcal{T}}\check{\boldsymbol{\mathcal{W}}}_i\mathbf{P}[k+1] + \lambda\bar{\beta}\mathbf{I}_\rho\right]^{-1} \times\mathbf{P}^{\mathcal{T}}[k+1]\check{\boldsymbol{\mathcal{W}}}_i^{\mathcal{T}}\check{\mathbf{s}}_i \tag{22}$$

which again involves matrix inversion of dimension ρ-by-ρ. The overall algorithm is tabulated in Table I.

Although the proposed batch algorithm exhibits low computational and memory requirements, it is not suitable for online processing, since (13) must be re-solved every time a new set of measurements arrive, incurring major computational burden. Thus, the development of an online recursive algorithm is well motivated.

IV. ONLINE ALGORITHM

A. Stochastic Approximation Approach

In practice, it is often the case that a new set of data becomes available sequentially in time. Then, it is desirable to have an algorithm that can process the newly acquired data incrementally and refine the previous estimates, rather than re-computing the batch solution, which may incur prohibitively growing computational burden. Furthermore, when the channel is time-varying due to, e.g., mobile obstacles, online algorithms can readily track such variations.

Stochastic approximation (SA) is an appealing strategy for deriving online algorithms [27], [28]. Recently, techniques involving minimizing majorized surrogate functions were developed to handle nonconvex cost functions in online settings [24], [29]–[31]. An online algorithm to solve a dictionary learning problem was proposed in [30]. A stochastic gradient descent algorithm was derived for subspace tracking and anomaly detection in [24]. Here, an online algorithm for the CPCP problem is developed. The proposed approach employs quadratic surrogate functions with diagonal weighting so as to capture disparate curvatures in the directions of different block variables.

For simplicity, let the number of measurements per time slot t be constant $M := |\mathcal{M}^{(t)}|$ for all t. Define $\mathbf{X} := (\mathbf{P}, \mathbf{Q}, \mathbf{E}) \in \mathcal{X} \subset \mathcal{X}' := \mathbb{R}^{(N_x\times\rho)} \times \mathbb{R}^{(N_y\times\rho)} \times \mathbb{R}^{(N_x\times N_y)}$, where $\mathcal{X}$ is a compact convex set, and $\mathcal{X}'$ a bounded open set, and $\boldsymbol{\xi}^{(t)} := [\{\check{s}_m^{(t)}\}_{m=1}^M, \{\mathbf{W}_m^{(t)}\}_{m=1}^M] \in \Xi$, where Ξ is assumed to be bounded. Define with slight abuse of notation

$$g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)}) = g_1(\mathbf{P}, \mathbf{Q}, \mathbf{E}, \boldsymbol{\xi}^{(t)}) := \frac{1}{2}\sum_{m=1}^{M}\left(\langle\mathbf{W}_m^{(t)}, \mathbf{P}\mathbf{Q}^{\mathcal{T}} + \mathbf{E}\rangle - \check{s}_m^{(t)}\right)^2 \tag{23}$$

$$g_2(\mathbf{X}) = g_2(\mathbf{P}, \mathbf{Q}, \mathbf{E}) := \frac{\lambda}{2}\left(||\mathbf{P}||_F^2 + ||\mathbf{Q}||_F^2\right) + \mu||\mathbf{E}||_1. \tag{24}$$

A quadratic surrogate function for $g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)})$ is then constructed as

$$\begin{aligned}\check{g}_1(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) &:= g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \\ &+ \langle\mathbf{P} - \mathbf{P}^{(t-1)}, \nabla_{\mathbf{P}}g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)})\rangle + \frac{\eta_{\mathbf{P}}^{(t)}}{2}\|\mathbf{P} - \mathbf{P}^{(t-1)}\|_F^2 \\ &+ \langle\mathbf{Q} - \mathbf{Q}^{(t-1)}, \nabla_{\mathbf{Q}}g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)})\rangle + \frac{\eta_{\mathbf{Q}}^{(t)}}{2}\|\mathbf{Q} - \mathbf{Q}^{(t-1)}\|_F^2 \\ &+ \langle\mathbf{E} - \mathbf{E}^{(t-1)}, \nabla_{\mathbf{E}}g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)})\rangle + \frac{\eta_{\mathbf{E}}^{(t)}}{2}\|\mathbf{E} - \mathbf{E}^{(t-1)}\|_F^2\end{aligned} \tag{25}$$

TABLE II
ONLINE SGD SOLVER OF (P2)

Initialize $\mathbf{E}^{(0)} := \mathbf{0}_{N_x \times N_y}$, $\mathbf{P}^{(0)}$ and $\mathbf{Q}^{(0)}$ at random.
For $t = 1, 2, \ldots$
  Set $L_{\mathbf{P}} = \sum_{m=1}^{M} \left\| \mathbf{W}_m^{(t)} \mathbf{Q}^{(t-1)} \right\|_F^2$,
  $L_{\mathbf{Q}} = \sum_{m=1}^{M} \left\| \mathbf{W}_m^{(t)\mathcal{T}} \mathbf{P}^{(t-1)} \right\|_F^2$,
  $L_{\mathbf{E}} = \sum_{m=1}^{M} \left\| \mathbf{W}_m^{(t)} \right\|_F^2$, and $L_{\min} = \min\{L_{\mathbf{P}}, L_{\mathbf{Q}}, L_{\mathbf{E}}\}$.
  Set $\eta_{\mathbf{P}}^{(t)} \geq \frac{L_{\mathbf{P}}^2}{L_{\min}}, \eta_{\mathbf{Q}}^{(t)} \geq \frac{L_{\mathbf{Q}}^2}{L_{\min}}$, and $\eta_{\mathbf{E}}^{(t)} \geq \frac{L_{\mathbf{E}}^2}{L_{\min}}$.
  Set $\bar{\eta}_{\mathbf{P}}^{(t)} = \sum_{\tau=1}^{t} \eta_{\mathbf{P}}^{(\tau)}$, $\bar{\eta}_{\mathbf{Q}}^{(t)} = \sum_{\tau=1}^{t} \eta_{\mathbf{Q}}^{(\tau)}$, and $\bar{\eta}_{\mathbf{E}}^{(t)} = \sum_{\tau=1}^{t} \eta_{\mathbf{E}}^{(\tau)}$.
  $\mathbf{P}^{(t)} = \mathbf{P}^{(t-1)} - \frac{1}{\bar{\eta}_{\mathbf{P}}^{(t)} + \lambda t} \left(\nabla_{\mathbf{P}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) + \lambda \mathbf{P}^{(t-1)} \right)$
  $\mathbf{Q}^{(t)} = \mathbf{Q}^{(t-1)} - \frac{1}{\bar{\eta}_{\mathbf{Q}}^{(t)} + \lambda t} \left(\nabla_{\mathbf{Q}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) + \lambda \mathbf{Q}^{(t-1)} \right)$
  $\mathbf{Z}^{(t)} = \frac{1}{\bar{\eta}_{\mathbf{E}}^{(t)}} \left[\eta_{\mathbf{E}}^{(t)} \mathbf{E}^{(t-1)} + \bar{\eta}_{\mathbf{E}}^{(t-1)} \mathbf{Z}^{(t-1)} - \nabla_{\mathbf{E}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \right]$
  $\mathbf{E}^{(t)} = \text{soft\_th}(\mathbf{Z}^{(t)}; \mu t / \bar{\eta}_{\mathbf{E}}^{(t)})$
Next t

where $\eta_{\mathbf{P}}^{(t)}$, $\eta_{\mathbf{Q}}^{(t)}$, and $\eta_{\mathbf{E}}^{(t)}$ are positive constants, and with $\tilde{\tilde{f}}_m^{(t)}(\mathbf{P}, \mathbf{Q}, \mathbf{E}) := \langle \mathbf{W}_m^{(t)}, \mathbf{P}\mathbf{Q}^{\mathcal{T}} + \mathbf{E} \rangle - \check{s}_m^{(t)}$ it can be readily verified that

$$\nabla_{\mathbf{P}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) = \sum_{m=1}^{M} \tilde{\tilde{f}}_m^{(t)}(\mathbf{P}^{(t-1)}, \mathbf{Q}^{(t-1)}, \mathbf{E}^{(t-1)}) \mathbf{W}_m^{(t)} \mathbf{Q}^{(t-1)} \tag{26}$$

$$\nabla_{\mathbf{Q}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) = \sum_{m=1}^{M} \tilde{\tilde{f}}_m^{(t)}(\mathbf{P}^{(t-1)}, \mathbf{Q}^{(t-1)}, \mathbf{E}^{(t-1)}) \mathbf{W}_m^{(t)\mathcal{T}} \mathbf{P}^{(t-1)} \tag{27}$$

$$\nabla_{\mathbf{E}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) = \sum_{m=1}^{M} \tilde{\tilde{f}}_m^{(t)}(\mathbf{P}^{(t-1)}, \mathbf{Q}^{(t-1)}, \mathbf{E}^{(t-1)}) \mathbf{W}_m^{(t)}. \tag{28}$$

Let us focus on the case without the forgetting factor, i.e., $\beta = 1$. A convergent SA algorithm for (P2) is obtained by considering the following surrogate problem

$$\text{(P3)} \quad \min_{\mathbf{X}} \frac{1}{t} \sum_{\tau=1}^{t} \left[\check{g}_1(\mathbf{X}, \mathbf{X}^{(\tau-1)}, \boldsymbol{\xi}^{(\tau)}) + g_2(\mathbf{X}) \right]. \tag{29}$$

In fact, solving (P3) yields a stochastic gradient descent (SGD) algorithm. In particular, since variables $\mathbf{P}$, $\mathbf{Q}$, and $\mathbf{E}$ can be separately optimized in (P3), the proposed algorithm updates the variables in parallel in each time slot t as

$$\mathbf{P}^{(t)} = \arg\min_{\mathbf{P}} \sum_{\tau=1}^{t} \Big[\langle \mathbf{P} - \mathbf{P}^{(\tau-1)}, \nabla_{\mathbf{P}} g_1(\mathbf{X}^{(\tau-1)}, \boldsymbol{\xi}^{(\tau)}) \rangle + \frac{\eta_{\mathbf{P}}^{(\tau)}}{2} ||\mathbf{P} - \mathbf{P}^{(\tau-1)}||_F^2 + \frac{\lambda}{2} ||\mathbf{P}||_F^2 \Big] \tag{30}$$

$$\mathbf{Q}^{(t)} = \arg\min_{\mathbf{Q}} \sum_{\tau=1}^{t} \Big[\langle \mathbf{Q} - \mathbf{Q}^{(\tau-1)}, \nabla_{\mathbf{Q}} g_1(\mathbf{X}^{(\tau-1)}, \boldsymbol{\xi}^{(\tau)}) \rangle + \frac{\eta_{\mathbf{Q}}^{(\tau)}}{2} ||\mathbf{Q} - \mathbf{Q}^{(\tau-1)}||_F^2 + \frac{\lambda}{2} ||\mathbf{Q}||_F^2 \Big] \tag{31}$$

$$\mathbf{E}^{(t)} = \arg\min_{\mathbf{E}} \sum_{\tau=1}^{t} \Big[\langle \mathbf{E} - \mathbf{E}^{(\tau-1)}, \nabla_{\mathbf{E}} g_1(\mathbf{X}^{(\tau-1)}, \boldsymbol{\xi}^{(\tau)}) \rangle + \frac{\eta_{\mathbf{E}}^{(\tau)}}{2} ||\mathbf{E} - \mathbf{E}^{(\tau-1)}||_F^2 + \mu ||\mathbf{E}||_1 \Big]. \tag{32}$$

By checking the first-order optimality conditions, and defining $\bar{\eta}_{\mathbf{P}}^{(t)} := \sum_{\tau=1}^{t} \eta_{\mathbf{P}}^{(\tau)}$ and $\bar{\eta}_{\mathbf{Q}}^{(t)} := \sum_{\tau=1}^{t} \eta_{\mathbf{Q}}^{(\tau)}$, the update rules for $\mathbf{P}$ and $\mathbf{Q}$ are obtained as

$$\mathbf{P}^{(t)} = \frac{1}{\bar{\eta}_{\mathbf{P}}^{(t)} + \lambda t} \sum_{\tau=1}^{t} \left[\eta_{\mathbf{P}}^{(\tau)} \mathbf{P}^{(\tau-1)} - \nabla_{\mathbf{P}} g_1(\mathbf{X}^{(\tau-1)}, \boldsymbol{\xi}^{(\tau)}) \right] \tag{33}$$

$$\mathbf{Q}^{(t)} = \frac{1}{\bar{\eta}_{\mathbf{Q}}^{(t)} + \lambda t} \sum_{\tau=1}^{t} \left[\eta_{\mathbf{Q}}^{(\tau)} \mathbf{Q}^{(\tau-1)} - \nabla_{\mathbf{Q}} g_1(\mathbf{X}^{(\tau-1)}, \boldsymbol{\xi}^{(\tau)}) \right] \tag{34}$$

which can be written in recursive forms as

$$\mathbf{P}^{(t)} = \mathbf{P}^{(t-1)} - \frac{1}{\bar{\eta}_{\mathbf{P}}^{(t)} + \lambda t} \left(\nabla_{\mathbf{P}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) + \lambda \mathbf{P}^{(t-1)} \right) \tag{35}$$

$$\mathbf{Q}^{(t)} = \mathbf{Q}^{(t-1)} - \frac{1}{\bar{\eta}_{\mathbf{Q}}^{(t)} + \lambda t} \left(\nabla_{\mathbf{Q}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) + \lambda \mathbf{Q}^{(t-1)} \right). \tag{36}$$

Due to the non-smoothness of $||\mathbf{E}||_1$, the update for $\mathbf{E}$ proceeds in two steps. First, an auxiliary variable $\mathbf{Z}^{(t)}$ is introduced, which is computed as

$$\mathbf{Z}^{(t)} = \frac{1}{\bar{\eta}_{\mathbf{E}}^{(t)}} \left[\sum_{k=1}^{t} \eta_{\mathbf{E}}^{(k)} \mathbf{E}^{(k-1)} - \nabla_{\mathbf{E}} g_1(\mathbf{X}^{(k-1)}, \boldsymbol{\xi}^{(k)}) \right]. \tag{37}$$

Again defining $\bar{\eta}_{\mathbf{E}}^{(t)} := \sum_{\tau=1}^{t} \eta_{\mathbf{E}}^{(\tau)}$, matrix $\mathbf{Z}^{(t)}$ can be obtained recursively as

$$\mathbf{Z}^{(t)} = \frac{1}{\bar{\eta}_{\mathbf{E}}^{(t)}} \left[\eta_{\mathbf{E}}^{(t)} \mathbf{E}^{(t-1)} + \bar{\eta}_{\mathbf{E}}^{(t-1)} \mathbf{Z}^{(t-1)} - \nabla_{\mathbf{E}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \right]. \tag{38}$$

Then, $\mathbf{E}^{(t)}$ is updated as

$$\mathbf{E}^{(t)} = \text{soft\_th}(\mathbf{Z}^{(t)}; \mu t / \bar{\eta}_{\mathbf{E}}^{(t)}). \tag{39}$$

The overall online algorithm is listed in Table II.

Remark 1 (Computational Complexity): In the batch algorithm of Table I, the complexity orders for computing the updates for each of $\mathbf{p}_i$ and $\mathbf{q}_i$ are $O(N_y MT)$ and $O(N_x MT)$,

TABLE III
RECONSTRUCTION ERROR AT $T = 130$ AND COMPUTATIONAL COMPLEXITY PER ITERATION

Algorithm	Proposed (BCD)	Ridge-regularized LS	Total variation (ADMM)	LASSO
$\|\mathbf{F}_0 - \hat{\mathbf{F}}\|_F / \|\mathbf{F}_0\|_F$	0.1064	0.1796	0.1196	0.1828
Complexity per iteration	$\mathcal{O}(N_x N_y MT)$	N/A	$\mathcal{O}((N_x N_y)^3 + (N_x N_y)^2 MT)$	$\mathcal{O}(N_x N_y MT)$

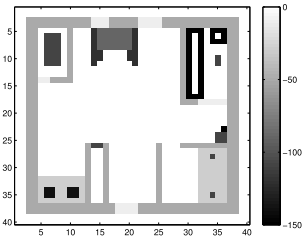
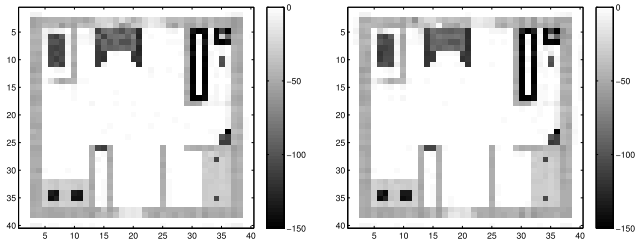
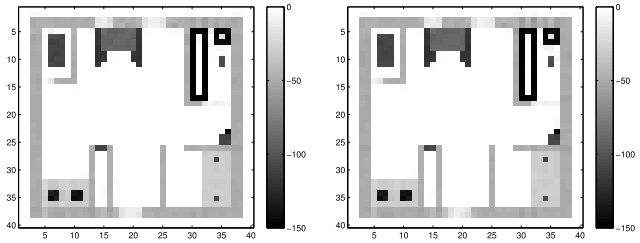


Fig. 1. True SLF.



(a) BCD ($T = 130$, $N = 52$) (b) APG ($T = 130$, $N = 52$)



(c) BCD ($T = 260$, $N = 73$) (d) APG ($T = 260$, $N = 73$)

Fig. 2. SLFs reconstructed by the batch algorithms.

respectively, due to the computation of $\tilde{\boldsymbol{\mathcal{W}}}^T \tilde{\mathbf{s}}_i$ and $\check{\boldsymbol{\mathcal{W}}}_i^T \check{\mathbf{s}}_i$. Thus, the complexity orders for updating $\mathbf{P}$ and $\mathbf{Q}$ per iteration k are both $\mathcal{O}(N_x N_y MT)$. The update of e_l incurs complexity $\mathcal{O}(MT)$ for computing $\boldsymbol{\omega}_l^T \check{\mathbf{s}}_l$. Thus, the complexity order for updating $\mathbf{E}$ per iteration k is $\mathcal{O}(N_x N_y MT)$. Accordingly, the overall per-iteration complexity of the batch algorithm becomes $\mathcal{O}(N_x N_y MT)$. On the other hand, the complexity of the online algorithm in Table II is dominated by the gradient computations, which require $\mathcal{O}(\rho N_x N_y M)$. Since ρ is smaller than N_x and N_y, and the per-iteration complexity does not grow with T, the online algorithm has a much more affordable complexity than its batch counterpart, and it is scalable for large network scenarios.

B. Convergence

The iterates $\{\mathbf{X}^{(t)}\}_{t=1}^{\infty}$ generated from the algorithm in Table II converge to a stationary point of (P2), as asserted in the following proposition. First define

$$C_t(\mathbf{X}) := \frac{1}{t}\sum_{\tau=1}^{t}\Big[g_1(\mathbf{X}, \boldsymbol{\xi}^{(\tau)}) + g_2(\mathbf{X})\Big] \tag{40}$$

$$\check{C}_t(\mathbf{X}) := \frac{1}{t}\sum_{\tau=1}^{t}\Big[\check{g}_1(\mathbf{X}, \mathbf{X}^{(\tau-1)}, \boldsymbol{\xi}^{(\tau)}) + g_2(\mathbf{X})\Big] \tag{41}$$

$$C(\mathbf{X}) := \mathbb{E}_{\boldsymbol{\xi}}\left[g_1(\mathbf{X}, \boldsymbol{\xi}) + g_2(\mathbf{X})\right]. \tag{42}$$

Note that $C_t(\mathbf{X})$ is essentially identical to the cost of (P2). Furthermore, the minimizer of $C_t(\mathbf{X})$ approaches that of $C(\mathbf{X})$ when $t \to \infty$, provided $\boldsymbol{\xi}$ obeys the law of large numbers, which is clearly the case when e.g., $\{\boldsymbol{\xi}^{(t)}\}$ is i.i.d.

Assume that $\nabla_{\mathbf{P}} g_1(\cdot, \mathbf{Q}, \mathbf{E}, \boldsymbol{\xi})$, $\nabla_{\mathbf{Q}}(\mathbf{P}, \cdot, \mathbf{E}, \boldsymbol{\xi})$ and $\nabla_{\mathbf{E}}(\mathbf{P}, \mathbf{Q}, \cdot, \boldsymbol{\xi})$ are Lipschitz with respect to $\mathbf{P}$, $\mathbf{Q}$, and $\mathbf{E}$, respectively, with constants $L_{\mathbf{P}}$, $L_{\mathbf{Q}}$, and $L_{\mathbf{E}}$, respectively (which will be shown in Appendix B). Furthermore, let $\bar{\alpha}_i^{(t)} := (\sum_{\tau=1}^{t}(\eta_i^{(\tau)} + \lambda))^{-1}$ for $i \in \{\mathbf{P}, \mathbf{Q}\}$, and $\bar{\alpha}_{\mathbf{E}}^{(t)} := (\bar{\eta}_{\mathbf{E}}^{(t)})^{-1}$ denote step sizes.

Proposition 2: If (a1) $\{\boldsymbol{\xi}^{(t)}\}_{t=1}^{\infty}$ *is an independent and identically distributed (i.i.d) random sequence;* (a2) $\{\mathbf{X}^{(t)}\}_{t=1}^{\infty}$ *are in a compact set* $\mathcal{X}$*;* (a3) Ξ *is bounded;* (a4) *For* $i \in \{\mathbf{P}, \mathbf{Q}, \mathbf{E}\}$, $\bar{\eta}_i{}^{(t)} \geq ct$ $\forall t$ *for some* $c \geq 0$*; and* (a5) $c' \geq \eta_i^{(t)} \geq L_i^2 / L_{\min}$ $\forall t$ *for some* $c' > 0$ *and* $L_{\min} := \min\{L_{\mathbf{P}}, L_{\mathbf{Q}}, L_{\mathbf{E}}\}$*, then the iterates* $\{\mathbf{X}^{(t)}\}_{t=1}^{\infty}$ *generated by the algorithm in Table II converge to the set of stationary points of* (P2) *with* $\beta = 1$*, i.e.,*

$$\lim_{t\to\infty} \inf_{\bar{\mathbf{X}} \in \bar{\mathcal{X}}} \|\mathbf{X}^{(t)} - \bar{\mathbf{X}}\|_F = 0 \quad \text{a.s.} \tag{43}$$

where $\bar{\mathcal{X}}$ *is the set of stationary points of* $C(\mathbf{X})$.

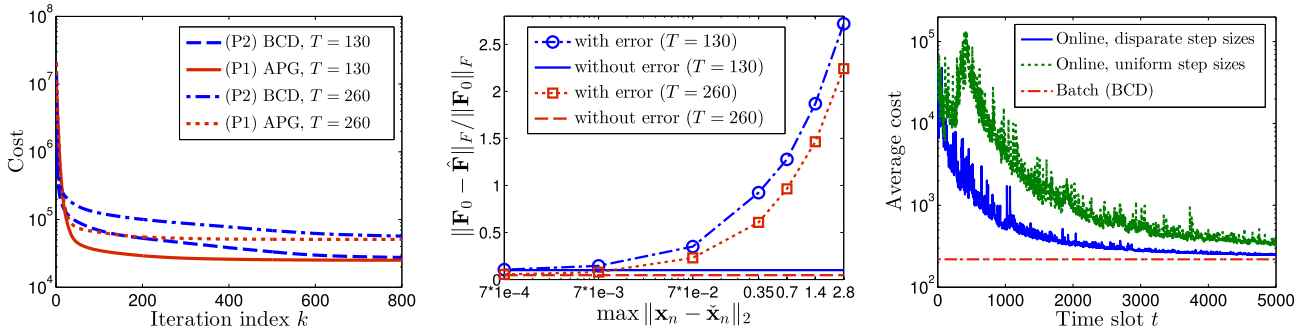
Proof: See Appendix B.

V. NUMERICAL TESTS

Performance of the proposed batch and online algorithms was assessed through numerical tests using both synthetic and real datasets. A few existing methods were also tested for comparison. The ridge-regularized least-squares (LS) scheme estimates the SLF as $\text{vec}(\hat{\mathbf{F}}) = (\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{W}}^T + \omega \mathbf{C}_{\boldsymbol{f}}^{-1})^{-1}\boldsymbol{\mathcal{W}}\check{\mathbf{s}}$, where $\mathbf{C}_{\boldsymbol{f}}$ is the spatial covariance matrix of the SLF, and ω is a regularization parameter [8], [11], [21]. The total variation (TV)-regularized LS scheme in [32] was also tested, which solves $\min_{\boldsymbol{f}} \|\check{\mathbf{s}} - \boldsymbol{\mathcal{W}}^T \boldsymbol{f}\|_2^2 + \omega\big(\sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y}|f_{i+1,j} - f_{i,j}| + \sum_{i=1}^{N_x}\sum_{j=1}^{N_y-1}|f_{i,j+1} - f_{i,j}|\big)$ where $\boldsymbol{f} := \text{vec}(\mathbf{F})$ and $f_{i,j}$ corresponds to the (i, j)-th element of $\mathbf{F}$. Finally, the LASSO estimator was obtained by solving (P1) with $\lambda = 0$.

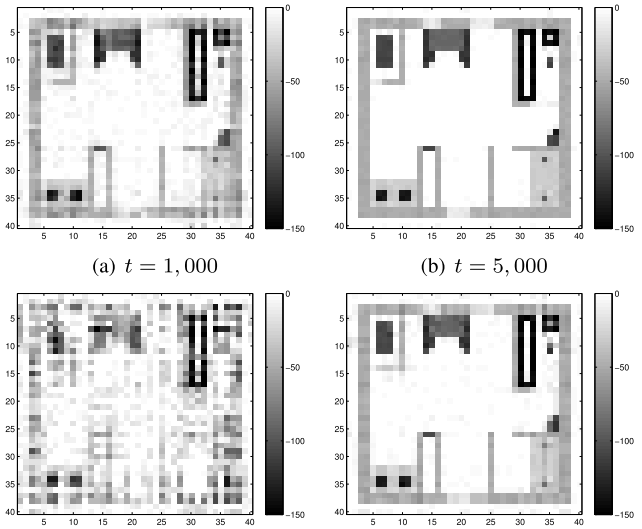
A. Test With Synthetic Data

Random tomographic measurements were taken by sensors deployed uniformly over $\mathcal{A} := [0.5, 40.5] \times [0.5, 40.5]$, from which the SLF with $N_x = N_y = 40$ was reconstructed. Per-time slot, 10 measurements were taken, corrupted by zero-mean white Gaussian noise with variance $\sigma^2 = 0.1$. The regularization parameters were set to $\lambda = 0.05$ and $\mu = 0.01$ through cross-validation by minimizing the normalized error $\|\hat{\mathbf{F}} - \mathbf{F}_0\|_F / \|\mathbf{F}_0\|_F$, where $\mathbf{F}_0$ is the ground-truth SLF depicted in Fig. 1. Other parameters were set to $\rho = 13$, $\beta = 1$, and $\delta = 0.06$; while $\mathbf{C}_{\boldsymbol{f}} = \mathbf{I}_{N_x N_y}$ and $\omega = 0.13$ were used for the ridge-regularized LS.



(a) (b) (c)

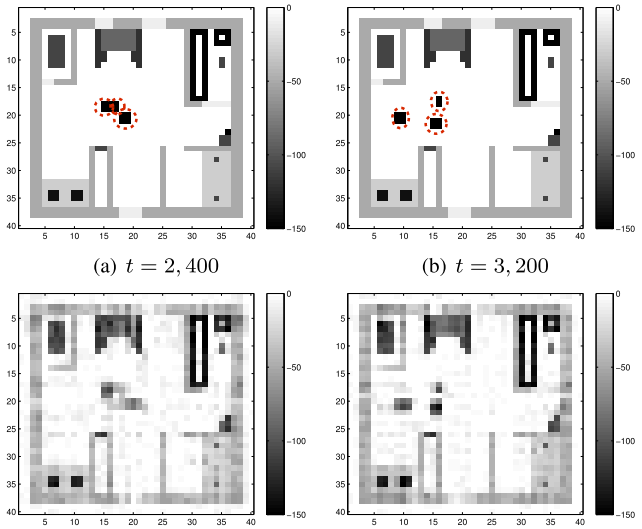
Fig. 3. SLF reconstruction using the batch and online algorithms. (a) Cost versus iterations (batch). (b) Reconstruction error versus CR location error (batch). (c) Average cost over time slots (online).



(a) $t = 1,000$ (b) $t = 5,000$

(c) $t = 1,000$ (d) $t = 5,000$

Fig. 4. SLFs reconstructed by the online algorithm. (a) and (b) correspond to using $\bar{\eta}_{\mathbf{P}}^{(t)} = \bar{\eta}_{\mathbf{Q}}^{(t)} = 300$ and $\bar{\eta}_{\mathbf{E}}^{(t)} = 10$. (c) and (d) use $\bar{\eta}_{\mathbf{P}}^{(t)} = \bar{\eta}_{\mathbf{Q}}^{(t)} = \bar{\eta}_{\mathbf{E}}^{(t)} = 300$.



(a) $t = 2,400$ (b) $t = 3,200$

(c) $t = 2,400$ (d) $t = 3,200$

Fig. 5. (a) and (b) are true SLFs; (c) and (d) show reconstructed SLFs at different time slots.

To validate the batch algorithm in Table I, two cases were tested. In the first case, the measurements were generated for $T = 130$ time slots using $N = 52$ sensors, while in the second case, $T = 260$ and $N = 73$ were used. As a comparison, the accelerated proximal gradient (APG) algorithm was also derived for (P1) [33]. Note that the APG requires the costly SVD operation of an N_x-by-N_y matrix per iteration, while only the inversion of a ρ-by-ρ matrix is necessary in the proposed BCD algorithm. Fig. 2 shows the SLFs reconstructed by APG and BCD algorithms for the two cases. Apparently, the reconstructed SLFs capture well the features of the ground-truth SLF in Fig. 1. Note that (P2) is underdetermined when $T = 130$ since the total number of unknowns in (P2) is $2,640$ while the total number of measurements is only $1,300$. This verifies that the channel gain maps can be accurately interpolated with a small number of measurements by leveraging the attributes of the low rank and sparsity. Fig. 3(a) shows the convergence of the BCD and APG algorithms. The cost of (P2) from the BCD algorithm converges to that of (P1) from APG after $k = 550$ iterations, showing that the performance of solving (P1) directly is achievable by the proposed algorithm solving (P2) instead. This can also be corroborated from the reconstructed SLFs in Fig. 2 as well.

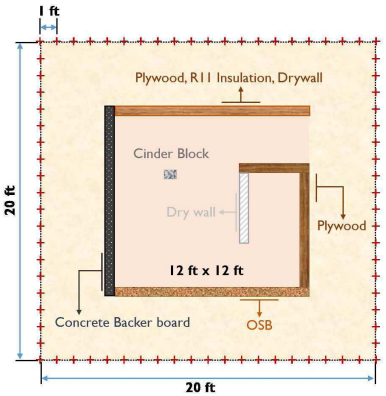
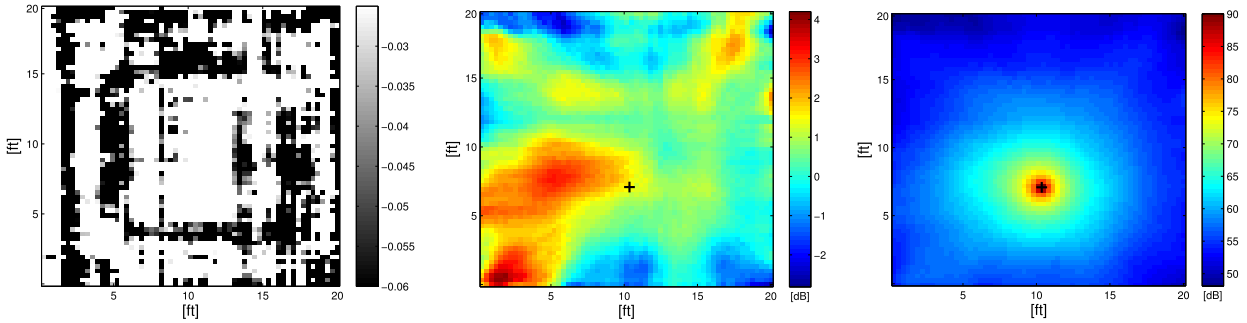


Fig. 6. Configuration of the testbed.

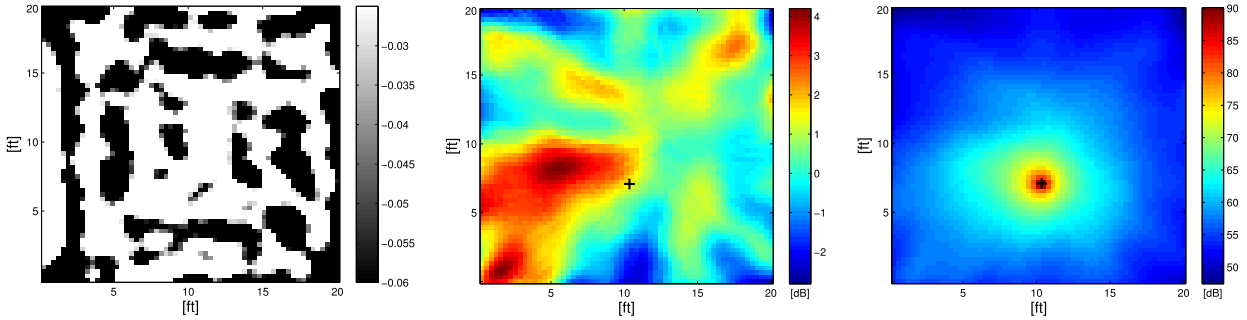
Table III lists the reconstruction error when $T = 130$ and the per-iteration complexity of the batch algorithms. It is seen that the proposed method outperforms benchmark algorithms in terms of the reconstruction error. Note that the ridge-regularized LS has a one-shot (non-iterative) complexity of $O((N_x N_y)^3)$, but its reconstruction capability is worse than the proposed algorithm as the true SLF is not smooth.

To test robustness of the proposed algorithm against imprecise CR location estimates, the reconstruction error versus the maximum sensor location error is depicted in Fig. 3(b).



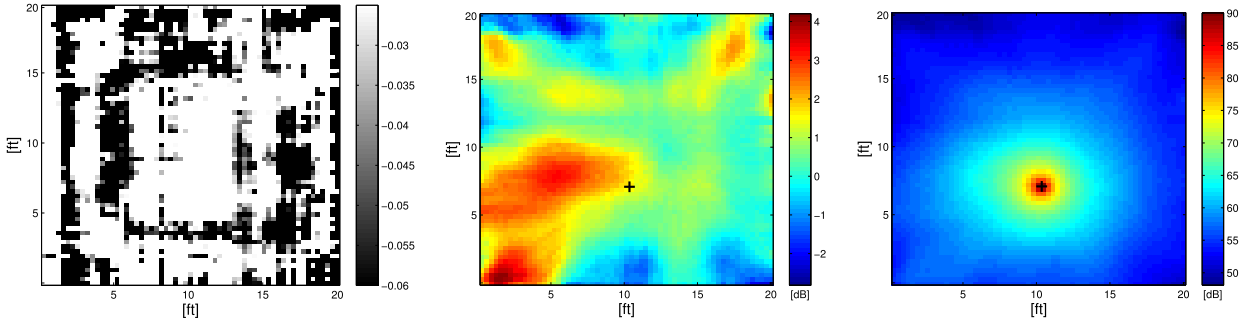
(a) SLF (b) Shadow fading map (c) Channel gain map

Fig. 7. Reconstructions by the proposed batch algorithm.



(a) SLF (b) Shadow fading map (c) Channel gain map

Fig. 8. Reconstructions by the ridge-regularized LS.



(a) SLF (b) Shadow fading map (c) Channel gain map

Fig. 9. Reconstructions by the proposed online algorithm.

To reconstruct $\mathbf{F}$ matrix, $\boldsymbol{\mathcal{W}}$ was computed via a set of erroneous sensor locations $\check{\mathbf{x}}_n^{(t)}$ obtained by adding uniformly random perturbations to true locations $\mathbf{x}_n^{(t)}$. It is seen that the SLF could be accurately reconstructed when the location error was small.

The numerical tests for the online algorithm were carried out with the same parameter setting as the batch experiments with $N = 317$. Fig. 3(c) depicts the evolution of the average cost in (40) for two sets of values for $(\bar{\eta}_{\mathbf{P}}^{(t)}, \bar{\eta}_{\mathbf{Q}}^{(t)}, \bar{\eta}_{\mathbf{E}}^{(t)})$. The green dotted curve corresponds to using $\bar{\eta}_{\mathbf{P}}^{(t)} = \bar{\eta}_{\mathbf{Q}}^{(t)} = \bar{\eta}_{\mathbf{E}}^{(t)} = 300$, while the blue solid curve is for $\bar{\eta}_{\mathbf{P}}^{(t)} = \bar{\eta}_{\mathbf{Q}}^{(t)} = 300$, and $\bar{\eta}_{\mathbf{E}}^{(t)} = 10$. It can be seen that the uniform step sizes for all variables result in convergence rate that is slower than that with the disparate step sizes. Fig. 4 shows the SLFs reconstructed via the online algorithm at $t = 1,000$ and $t = 5,000$ using the two choices of step sizes. It can be seen that for a given time slot t, flexibly choosing the step sizes yields much more accurate reconstruction. As far as reconstruction error, the online algorithm with disparate step sizes yields 6.3×10^{-2} at $t = 5,000$, while its batch counterpart has 2.4×10^{-2}. Although slightly less accurate SLF is obtained by the online algorithm, it comes with greater computational efficiency.

To assess the tracking ability of the online algorithm, the slow channel variation was simulated. The measurements were generated using the SLF in Fig. 1 with three additional objects slowly moving in the rate of unit pixel width per 70 time slots. Fig. 5 depicts instances of the true and reconstructed SLFs at $t = 2,400$ and $t = 3,200$, respectively, obtained by the online algorithm. The moving objects are marked by the red circles. It is seen that the reconstructed SLFs correctly capture the moving objects, while the stationary objects are estimated more clearly as t increases.

B. Test With Real Data

To validate the performance of the proposed framework for SLF and channel gain map estimation in realistic sce-

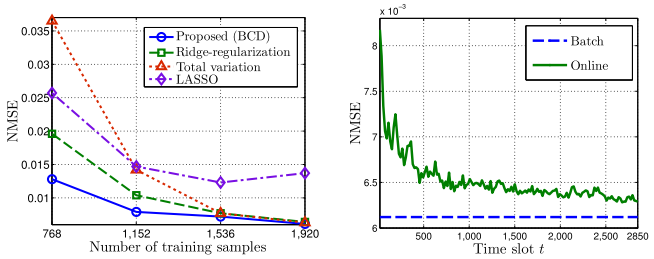
narios, real received signal strength (RSS) measurements were also processed. The data were collected by a set of $N = 20$ sensors deployed in the perimeter of a square-shaped testbed as shown in Fig. 6, where the crosses indicate the sensor positions. Data collection was performed in two steps [21]. First, free-space measurements were taken to obtain estimates of the path gain G_0 and the pathloss exponent γ via least-squares. The estimated γ was approximately 2, and G_0 was found to be 75. Then, tomographic measurements were formed with the artificial structure shown in Fig. 6. For the both measurements, 100 measurements were taken per time slot, in the 2.425 GHz frequency band, across 24 time slots. The shadowing measurements were obtained by subtracting the estimated pathloss from the RSS measurements.

The SLFs of size $N_x = N_y = 61$ were reconstructed by the proposed batch algorithm. The regularization parameters were set to $\lambda = 4.5$ and $\mu = 3.44$, which were determined by cross-validation. The parameter δ in (4) was set to 0.2 feet to capture the non-zero weights within the first Fresnel zone, and $\rho = 10$ and $\beta = 1$ were used.

For comparison, the ridge-regularized LS estimator was also tested. To construct $\mathbf{C}_f$, the exponential decay model in [6] was used, which models the covariance between points $\mathbf{x}$ and $\mathbf{x}'$ as $\mathbf{C}_f(\mathbf{x}, \mathbf{x}') = \sigma_s^2 e^{-\frac{\|\mathbf{x}-\mathbf{x}'\|_2}{\kappa}}$, where σ_s^2 and $\kappa > 0$ are model parameters. In our tests, $\sigma_s^2 = \kappa = 1$, and $\omega = 79.9$ were used.

The SLF, shadow fading map, and channel gain map reconstructed by the proposed BCD algorithm are depicted in Fig. 7. The shadow fading and channel gain maps portray the gains in dB between any point in the map and the fixed CR location at $(10.2, 7.2)$ (marked by the cross). Fig. 8 shows the results from the ridge-regularized LS estimation. It can be seen from Fig. 7(a) and Fig. 8(b) that the proposed low-rank plus sparse model produces a somewhat sharper SLF image than the ridge-regularized LS approach. Although the latter yields a smooth SLF image, it produces more artifacts near the isolated block and the boundary of the SLF. Such artifacts may lead to less accurate shadowing and channel gain maps. For instance, Fig. 7(b) and Fig. 8(b) both show that the shadow fading is stronger as more building material is crossed in the communication path. However, somewhat strong attenuations are observed near the cinder block location and the interior of the oriented strand board (OSB) wall only in Fig. 8(b), which seems anomalous.

The online algorithm was also tested with the real data. Parameters $\bar{\eta}_{\mathbf{P}}^{(t)} = \bar{\eta}_{\mathbf{Q}}^{(t)} = 620$ and $\bar{\eta}_{\mathbf{E}}^{(t)} = 200$ were selected, and 6×10^5 measurements were uniformly drawn from the original dataset with replacement to demonstrate the asymptotic performance. Fig. 9 depicts the reconstructed SLF, shadow fading and channel gain maps obtained from the online algorithm. It can be seen that the SLF shown in Fig. 9(a) is close to that depicted in Fig. 7(a). Similar observations can be made for the shadow fading and channel gain maps as well. Thus, the online algorithm is a viable alternative to the batch algorithm with reduced computational complexity, and affordable memory requirement.



(a) (b)

Fig. 10. NMSE of channel gain prediction by (a) the batch; and (b) online algorithms.

Channel gain estimation performance of the proposed algorithms was assessed via 5-fold cross-validation. Let $\check{\boldsymbol{g}}_{\text{test}}$ and $\hat{\boldsymbol{g}}_{\text{test}}$ denote RSS measurement vectors in the test set and its estimate, respectively. Prediction performance is measured by the normalized mean-square error (NMSE) $\|\check{\boldsymbol{g}}_{\text{test}} - \hat{\boldsymbol{g}}_{\text{test}}\|^2 / \|\check{\boldsymbol{g}}_{\text{test}}\|^2$. Fig. 10(a) displays the NMSE of batch algorithms with 480 test samples versus the number of training samples. It is shown that the proposed algorithm outperforms competing alternatives, particularly when a small number of training samples are available, validating the usefulness of the proposed model. The online algorithm was also tested with 2.85×10^5 measurements uniformly drawn from $1,920$ training samples with replacement. Fig. 10(b) depicts the evolution of the NMSE measured on 480 test samples at every t. It is observed that the online algorithm attains the batch performance as t increases.

VI. Conclusion

A low-rank plus sparse matrix model was proposed for channel gain cartography, which is instrumental for various CR spectrum sensing and resource allocation tasks. The channel gains were modeled as the sum of the distance-based pathloss and the tomographic accumulation of shadowing due to the underlying SLF. The SLF was postulated to have a low-rank structure corrupted by sparse outliers. Efficient batch and online algorithms were derived by leveraging a bifactor-based characterization of the matrix nuclear norm. The algorithms enjoy low computational complexity and a reduced memory requirement, without sacrificing the optimality, with provable convergence properties. Tests with both synthetic and real measurement datasets corroborated the claims and showed that the algorithms could accurately reveal the structure of the propagation medium.

Appendix

A. Proof of Proposition 1

A stationary point $\bar{\mathbf{P}}, \bar{\mathbf{Q}}$ and $\bar{\mathbf{E}}$ of (P2) must satisfy the following first-order optimality conditions [34]

$$\mathbf{0}_{N_x \times N_y} \in \partial_{\mathbf{E}} f(\bar{\mathbf{P}}, \bar{\mathbf{Q}}, \bar{\mathbf{E}}) = \Big\{ \tilde{f}(\bar{\mathbf{P}}\bar{\mathbf{Q}}^T, \bar{\mathbf{E}}) + \mu\bar{\beta}\Big[\text{sgn}(\bar{\mathbf{E}}) + \tilde{\mathbf{E}}0\Big] \Big| \bar{\mathbf{E}} \odot \tilde{\mathbf{E}} = \mathbf{0},\ \|\tilde{\mathbf{E}}\|_\infty \leq 1 \Big\} \tag{44}$$

$$\nabla_{\mathbf{P}} f(\bar{\mathbf{P}},\bar{\mathbf{Q}},\bar{\mathbf{E}}) = \tilde{f}(\bar{\mathbf{P}}\bar{\mathbf{Q}}^{\mathcal{T}},\bar{\mathbf{E}})\bar{\mathbf{Q}} + \lambda\bar{\beta}\bar{\mathbf{P}} = \mathbf{0}_{N_x\times\rho} \tag{45}$$

$$\nabla_{\mathbf{Q}^{\mathcal{T}}} f(\bar{\mathbf{P}},\bar{\mathbf{Q}},\bar{\mathbf{E}}) = \bar{\mathbf{P}}^{\mathcal{T}}\tilde{f}(\bar{\mathbf{P}}\bar{\mathbf{Q}}^{\mathcal{T}},\bar{\mathbf{E}}) + \lambda\bar{\beta}\bar{\mathbf{Q}}^{\mathcal{T}} = \mathbf{0}_{\rho\times N_y} \tag{46}$$

where $\odot$ denotes the element-wise (Hadamard) product. Through post-multiplying (45) by $\bar{\mathbf{P}}^{\mathcal{T}}$ and pre-multiplying (46) by $\bar{\mathbf{Q}}$, one can see that

$$\begin{aligned}\tilde{f}(\bar{\mathbf{P}}\bar{\mathbf{Q}}^{\mathcal{T}},\bar{\mathbf{E}}) &= -\mu\bar{\beta}(\mathrm{sgn}(\bar{\mathbf{E}}) + \tilde{\mathbf{E}})\\ \mathrm{tr}\left(\tilde{f}(\bar{\mathbf{P}}\bar{\mathbf{Q}}^{\mathcal{T}},\bar{\mathbf{E}})\bar{\mathbf{Q}}\bar{\mathbf{P}}^{\mathcal{T}}\right) &= -\lambda\bar{\beta}\mathrm{tr}(\bar{\mathbf{P}}\bar{\mathbf{P}}^{\mathcal{T}}) = -\lambda\bar{\beta}\mathrm{tr}(\bar{\mathbf{Q}}\bar{\mathbf{Q}}^{\mathcal{T}}).\end{aligned} \tag{47}$$

Define now $\kappa(\mathbf{R}_1,\mathbf{R}_2) := \frac{1}{2}(\mathrm{tr}\ (\mathbf{R}_1) + \mathrm{tr}\ (\mathbf{R}_2))$, and consider the following *convex* problem

$$\begin{aligned}\text{(P4)}\quad \min_{\substack{\mathbf{L},\mathbf{E}\in\mathbb{R}^{N_x\times N_y}\\ \mathbf{R}_1\in\mathbb{R}^{N_x\times N_x},\\ \mathbf{R}_2\in\mathbb{R}^{N_y\times N_y}}} &\sum_{\tau=1}^{T}\beta^{T-\tau}c^{(\tau)}(\mathbf{L},\mathbf{E})\\ &+\lambda\bar{\beta}\ \kappa(\mathbf{R}_1,\mathbf{R}_2) + \mu\bar{\beta}\ ||\mathbf{E}||_1\\ \text{subject to } \mathbf{R} &:= \begin{pmatrix}\mathbf{R}_1 & \mathbf{L}\\ \mathbf{L}^{\mathcal{T}} & \mathbf{R}_2\end{pmatrix} \succeq \mathbf{0}\end{aligned} \tag{48}$$

which is *equivalent* to (P1). Equivalence can be easily inferred by minimizing (P4) with respect to $\{\mathbf{R}_1,\mathbf{R}_2\}$ and noting an alternative characterization of the nuclear norm given by [25]

$$\begin{aligned}\|\mathbf{L}\|_* = \min_{\mathbf{R}_1,\mathbf{R}_2}\ &\kappa(\mathbf{R}_1,\mathbf{R}_2)\\ \text{subject to } &\mathbf{R}\succeq\mathbf{0}.\end{aligned} \tag{49}$$

In what follows, the optimality conditions of the conic program (P4) are explored. Introducing a Lagrange multiplier matrix $\mathbf{M} \in \mathbb{R}^{(N_x+N_y)\times(N_x+N_y)}$ associated with the conic constraint in (48), the Lagrangian is first formed as

$$\begin{aligned}\mathcal{L}(\mathbf{L},\mathbf{E},\mathbf{R}_1,\mathbf{R}_2;\mathbf{M}) = &\sum_{\tau=1}^{T}\beta^{T-\tau}c^{(\tau)}(\mathbf{L},\mathbf{E})\\ &+\lambda\bar{\beta}\ \kappa(\mathbf{R}_1,\mathbf{R}_2)+\mu\bar{\beta}\ ||\mathbf{E}||_1-\langle\mathbf{M},\mathbf{R}\rangle.\end{aligned} \tag{50}$$

For notational convenience, partition $\mathbf{M}$ as

$$\mathbf{M} := \begin{pmatrix}\mathbf{M}_1 & \mathbf{M}_2\\ \mathbf{M}_4 & \mathbf{M}_3\end{pmatrix} \tag{51}$$

in accordance with the block structure of $\mathbf{R}$ in (48), where $\mathbf{M}_1 \in \mathbb{R}^{N_x\times N_x}$ and $\mathbf{M}_3 \in \mathbb{R}^{N_y\times N_y}$. The optimal solution to (P4) must satisfy: (i) the stationarity conditions

$$\nabla_{\mathbf{L}}\mathcal{L}(\mathbf{L},\mathbf{E},\mathbf{R}_1,\mathbf{R}_2;\mathbf{M}) = \tilde{f}(\mathbf{L},\mathbf{E}) - \mathbf{M}_2 - \mathbf{M}_4^{\mathcal{T}} = \mathbf{0} \tag{52}$$

$$\mathbf{0}\in\partial_{\mathbf{E}}\mathcal{L}(\mathbf{L},\mathbf{E},\mathbf{R}_1,\mathbf{R}_2;\mathbf{M}) = \left\{\tilde{f}(\mathbf{L},\mathbf{E}) + \mu\bar{\beta}\left[\mathrm{sgn}(\mathbf{E})+\tilde{\mathbf{E}}\right]\ \middle|\ \mathbf{E}\odot\tilde{\mathbf{E}} = \mathbf{0},\ \|\tilde{\mathbf{E}}\|_\infty\le 1\right\} \tag{53}$$

$$\nabla_{\mathbf{R}_1}\mathcal{L}(\mathbf{L},\mathbf{E},\mathbf{R}_1,\mathbf{R}_2;\mathbf{M}) = \frac{\lambda\bar{\beta}}{2}\mathbf{I}_{N_x} - \mathbf{M}_1 = \mathbf{0} \tag{54}$$

$$\nabla_{\mathbf{R}_2}\mathcal{L}(\mathbf{L},\mathbf{E},\mathbf{R}_1,\mathbf{R}_2;\mathbf{M}) = \frac{\lambda\bar{\beta}}{2}\mathbf{I}_{N_y} - \mathbf{M}_3 = \mathbf{0} \tag{55}$$

(ii) complementary slackness condition $\langle\mathbf{M},\mathbf{R}\rangle = 0$; (iii) primal feasibility $\mathbf{R}\succeq\mathbf{0}$; and (iv) dual feasibility $\mathbf{M}\succeq\mathbf{0}$.

Using the stationary point $\bar{\mathbf{P}},\bar{\mathbf{Q}}$ and $\bar{\mathbf{E}}$ of (P2), construct a candidate solution for (P4) as $\hat{\mathbf{L}} := \bar{\mathbf{P}}\bar{\mathbf{Q}}^{\mathcal{T}}$, $\hat{\mathbf{E}} := \bar{\mathbf{E}}$, $\hat{\mathbf{R}}_1 := \bar{\mathbf{P}}\bar{\mathbf{P}}^{\mathcal{T}}$, and $\hat{\mathbf{R}}_2 := \bar{\mathbf{Q}}\bar{\mathbf{Q}}^{\mathcal{T}}$, as well as $\hat{\mathbf{M}}_1 := \frac{\lambda\bar{\beta}}{2}\mathbf{I}_{N_x}$, $\hat{\mathbf{M}}_2 := \frac{1}{2}\tilde{f}(\bar{\mathbf{P}}\bar{\mathbf{Q}}^{\mathcal{T}},\bar{\mathbf{E}})$, $\hat{\mathbf{M}}_3 := \frac{\lambda\bar{\beta}}{2}\mathbf{I}_{N_y}$, and $\hat{\mathbf{M}}_4 := \hat{\mathbf{M}}_2^{\mathcal{T}}$. After substituting these into (52)–(55), it can be readily verified that condition (i) holds. Condition (ii) also holds since

$$\begin{aligned}\langle\hat{\mathbf{M}},\hat{\mathbf{R}}\rangle &= \langle\hat{\mathbf{M}}_1,\hat{\mathbf{R}}_1\rangle + \langle\hat{\mathbf{M}}_2,\hat{\mathbf{L}}\rangle + \langle\hat{\mathbf{M}}_3,\hat{\mathbf{R}}_2\rangle + \langle\hat{\mathbf{M}}_4,\hat{\mathbf{L}}^{\mathcal{T}}\rangle\\ &= \frac{\lambda\bar{\beta}}{2}\mathrm{tr}(\bar{\mathbf{P}}\bar{\mathbf{P}}^{\mathcal{T}} + \bar{\mathbf{Q}}\bar{\mathbf{Q}}^{\mathcal{T}}) + \mathrm{tr}\left(\tilde{f}(\bar{\mathbf{P}}\bar{\mathbf{Q}}^{\mathcal{T}},\bar{\mathbf{E}})\bar{\mathbf{Q}}\bar{\mathbf{P}}^{\mathcal{T}}\right)\\ &= \mathbf{0}\end{aligned} \tag{56}$$

where the last equality follows from (47). Condition (iii) is met since $\mathbf{R}$ can be rewritten as

$$\mathbf{R} = \begin{pmatrix}\bar{\mathbf{P}}\bar{\mathbf{P}}^{\mathcal{T}} & \bar{\mathbf{P}}\bar{\mathbf{Q}}^{\mathcal{T}}\\ \bar{\mathbf{Q}}\bar{\mathbf{P}}^{\mathcal{T}} & \bar{\mathbf{Q}}\bar{\mathbf{Q}}^{\mathcal{T}}\end{pmatrix} = \begin{pmatrix}\bar{\mathbf{P}}\\ \bar{\mathbf{Q}}\end{pmatrix}\begin{pmatrix}\bar{\mathbf{P}}\\ \bar{\mathbf{Q}}\end{pmatrix}^{\mathcal{T}} \succeq \mathbf{0}. \tag{57}$$

For (iv), according to the Schur complement condition for positive semidefinite matrices, $\mathbf{M}\succeq\mathbf{0}$ holds if and only if

$$\hat{\mathbf{M}}_3 - \hat{\mathbf{M}}_4\hat{\mathbf{M}}_1^{-1}\hat{\mathbf{M}}_2 \succeq \mathbf{0} \tag{58}$$

which is equivalent to $\lambda_{\max}(\hat{\mathbf{M}}_2^{\mathcal{T}}\hat{\mathbf{M}}_2) \le (\lambda\bar{\beta}/2)^2$, or $||\tilde{f}(\bar{\mathbf{P}}\bar{\mathbf{Q}}^{\mathcal{T}},\bar{\mathbf{E}})|| \le \lambda\bar{\beta}$. ■

B. Proof of Proposition 2

The proof uses the technique similar to the one employed in [30], where the convergence of online algorithms for optimizing objectives involving non-convex bilinear terms and sparse matrices was established in the context of dictionary learning.

In order to proceed with the proof, three lemmata are first established. The first lemma concerns some properties of $g(\mathbf{X},\boldsymbol{\xi}^{(t)}) := g_1(\mathbf{X},\boldsymbol{\xi}^{(t)}) + g_2(\mathbf{X})$, and $\breve{g}(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)}) := \breve{g}_1(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)}) + g_2(\mathbf{X})$.

Lemma 2: If the assumptions (a1)–(a5) *in Proposition 2 hold, then*

(p1) $\breve{g}_1(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)})$ *majorizes* $g_1(\mathbf{X},\boldsymbol{\xi}^{(t)})$*, i.e.,* $\breve{g}_1(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)}) \ge g_1(\mathbf{X},\boldsymbol{\xi}^{(t)})\ \forall\ \mathbf{X}\in\mathcal{X}'$*;*

(p2) $\breve{g}_1$ *is locally tight, i.e.,* $\breve{g}_1(\mathbf{X}^{(t-1)},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)}) = g_1(\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)})$*;*

(p3) $\nabla\breve{g}_1(\mathbf{X}^{(t-1)},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)}) = \nabla g_1(\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)})$*;*

(p4) $\breve{g}(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)}) := \breve{g}_1(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)}) + g_2(\mathbf{X})$ *is uniformly strongly convex in* $\mathbf{X}$*, i.e.,* $\forall(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)}) \in \mathcal{X}\times\mathcal{X}\times\Xi$*, it holds that*

$$\begin{aligned}\breve{g}(\mathbf{X}+\mathbf{D},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)}) &- \breve{g}(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)})\\ &\ge \breve{g}'(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)};\mathbf{D}) + \frac{\zeta}{2}||\mathbf{D}||_F^2\end{aligned}$$

where $\zeta > 0$ *is a constant and* $\breve{g}'(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)};\mathbf{D})$ *is a directional derivative of* $\breve{g}$ *at* $\mathbf{X}$ *along the direction* $\mathbf{D}$*;*

(p5) g_1 *and* $\breve{g}_1$*, their derivatives, and their Hessians are uniformly bounded;*

(p6) g_2 *and its directional derivative* g_2' *are uniformly bounded; and*

(p7) *there exists* $\bar{g}\in\mathbb{R}$ *such that* $|\breve{g}(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)})| \le \bar{g}$.

Proof: For (p1), let us first show that $\nabla_{\mathbf{P}} g_1(\mathbf{P},\mathbf{Q},\mathbf{E},\boldsymbol{\xi}^{(t)})$, $\nabla_{\mathbf{Q}} g_1(\mathbf{P},\mathbf{Q},\mathbf{E},\boldsymbol{\xi}^{(t)})$, and $\nabla_{\mathbf{E}} g_1(\mathbf{P},\mathbf{Q},\mathbf{E},\boldsymbol{\xi}^{(t)})$ are Lipschitz continuous for $\mathbf{X} := (\mathbf{P},\mathbf{Q},\mathbf{E}) \in \mathcal{X}'$ and $\boldsymbol{\xi}^{(t)} \in \Xi$. For arbitrary $\mathbf{X}_1 := (\mathbf{P}_1,\mathbf{Q}_1,\mathbf{E}_1)$, $\mathbf{X}_2 := (\mathbf{P}_2,\mathbf{Q}_2,\mathbf{E}_2) \in \mathcal{X}'$, the variation of ∇g_1 in (26) can be bounded as

$$\begin{aligned}
&\|\nabla_{\mathbf{P}} g_1(\mathbf{P}_1,\mathbf{Q},\mathbf{E},\boldsymbol{\xi}^{(t)}) - \nabla_{\mathbf{P}} g_1(\mathbf{P}_2,\mathbf{Q},\mathbf{E},\boldsymbol{\xi}^{(t)})\|_F \\
&= \left\| \sum_{m=1}^{M} \langle \mathbf{W}_m^{(t)}, (\mathbf{P}_1-\mathbf{P}_2)\mathbf{Q}^{\mathcal{T}}\rangle \mathbf{W}_m^{(t)}\mathbf{Q} \right\|_F \\
&\overset{\text{(i1)}}{\le} \sum_{m=1}^{M} |\langle \mathbf{W}_m^{(t)}, (\mathbf{P}_1-\mathbf{P}_2)\mathbf{Q}^{\mathcal{T}}\rangle| \|\mathbf{W}_m^{(t)}\mathbf{Q}\|_F \\
&\overset{\text{(i2)}}{\le} \sum_{m=1}^{M} \|\mathbf{P}_1-\mathbf{P}_2\|_F \ \|\mathbf{W}_m^{(t)}\mathbf{Q}\|_F^2
\end{aligned}$$

where (i1) and (i2) are due to the triangle and Cauchy-Schwarz inequalities, respectively. Since $\mathcal{X}'$ and Ξ are assumed to be bounded, $\sum_{m=1}^{M} \|\mathbf{W}_m^{(t)}\mathbf{Q}\|_F^2$ is bounded. Therefore, there exists a positive constant $L_{\mathbf{P}}$ such that

$$\begin{aligned}
&\|\nabla_{\mathbf{P}} g_1(\mathbf{P}_1,\mathbf{Q},\mathbf{E},\boldsymbol{\xi}^{(t)}) - \nabla_{\mathbf{P}} g_1(\mathbf{P}_2,\mathbf{Q},\mathbf{E},\boldsymbol{\xi}^{(t)})\|_F \\
&\qquad \le L_{\mathbf{P}} \|\mathbf{P}_1-\mathbf{P}_2\|_F \quad (59)
\end{aligned}$$

meaning that $\nabla_{\mathbf{P}} g_1(\mathbf{P},\mathbf{Q},\mathbf{E},\boldsymbol{\xi}^{(t)})$ is Lipschitz continuous with constant $L_{\mathbf{P}}$. Similar arguments hold for $\nabla_{\mathbf{Q}} g_1(\mathbf{P},\mathbf{Q},\mathbf{E},\boldsymbol{\xi}^{(t)})$ and $\nabla_{\mathbf{E}} g_1(\mathbf{P},\mathbf{Q},\mathbf{E},\boldsymbol{\xi}^{(t)})$ as well, with Lipschitz constants $L_{\mathbf{Q}}$ and $L_{\mathbf{E}}$, respectively. Then, upon defining $\|\mathbf{X}\|_\Delta := \sqrt{L_{\mathbf{P}}^2\|\mathbf{P}\|_F^2 + L_{\mathbf{Q}}^2\|\mathbf{Q}\|_F^2 + L_{\mathbf{E}}^2\|\mathbf{E}\|_F^2}$, it is easy to verify

$$\|\nabla g_1(\mathbf{X}_1,\boldsymbol{\xi}^{(t)}) - \nabla g_1(\mathbf{X}_2,\boldsymbol{\xi}^{(t)})\|_F \le \|\mathbf{X}_1-\mathbf{X}_2\|_\Delta. \quad (60)$$

On the other hand, the proof of the Descent Lemma [35] can be adopted to show

$$\begin{aligned}
&g_1(\mathbf{X},\boldsymbol{\xi}^{(t)}) - g_1(\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)}) \\
&\le \langle \mathbf{X}-\mathbf{X}^{(t-1)}, \nabla g_1(\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)})\rangle + \int_0^1 \|\mathbf{X}-\mathbf{X}^{(t-1)}\|_F \\
&\quad \times \|\nabla g_1(\mathbf{X}^{(t-1)} + \alpha(\mathbf{X}-\mathbf{X}^{(t-1)}),\boldsymbol{\xi}^{(t)}) \\
&\quad - \nabla g_1(\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)})\|_F d\alpha. \quad (61)
\end{aligned}$$

Note that

$$\|\mathbf{X}\|_F \le \frac{1}{L_{\min}} \|\mathbf{X}\|_\Delta \quad (62)$$

where $L_{\min} := \min\{L_{\mathbf{P}}, L_{\mathbf{Q}}, L_{\mathbf{E}}\}$. Then, substitution of (60) into (61) with $\mathbf{X}_1 = \mathbf{X}^{(t-1)} + \alpha\ (\mathbf{X}-\mathbf{X}^{(t-1)})$ and $\mathbf{X}_2 = \mathbf{X}^{(t-1)}$ yields

$$\begin{aligned}
&g_1(\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)}) + \langle \mathbf{X}-\mathbf{X}^{(t-1)}, \nabla g_1(\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)})\rangle \\
&\qquad + \frac{1}{2L_{\min}} \|\mathbf{X}-\mathbf{X}^{(t-1)}\|_\Delta^2 \ge g_1(\mathbf{X},\boldsymbol{\xi}^{(t)}) \quad (63)
\end{aligned}$$

which completes the proof by the construction of $\check{g}_1$, provided that $\eta_i^{(t)} \ge L_i^2/L_{\min}$ for all $i \in \{\mathbf{P},\mathbf{Q},\mathbf{E}\}$.

To show (p2) and (p3), let us first denote

$$\begin{aligned}
\nabla g_1(\mathbf{X},\boldsymbol{\xi}^{(t)}) = \big(&\nabla_{\mathbf{P}} g_1(\mathbf{X},\boldsymbol{\xi}^{(t)}), \nabla_{\mathbf{Q}} g_1(\mathbf{X},\boldsymbol{\xi}^{(t)}), \\
&\nabla_{\mathbf{E}} g_1(\mathbf{X},\boldsymbol{\xi}^{(t)})\big) \quad (64)
\end{aligned}$$

$$\begin{aligned}
\nabla \check{g}_1(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)}) = \big(&\nabla_{\mathbf{P}} g_1(\mathbf{X},\boldsymbol{\xi}^{(t)}) + \eta_{\mathbf{P}}^{(t)}(\mathbf{P}-\mathbf{P}^{(t-1)}), \\
&\nabla_{\mathbf{Q}} g_1(\mathbf{X},\boldsymbol{\xi}^{(t)}) + \eta_{\mathbf{Q}}^{(t)}(\mathbf{Q}-\mathbf{Q}^{(t-1)}), \\
&\nabla_{\mathbf{E}} g_1(\mathbf{X},\boldsymbol{\xi}^{(t)}) + \eta_{\mathbf{E}}^{(t)}(\mathbf{E}-\mathbf{E}^{(t-1)})\big). \quad (65)
\end{aligned}$$

Then, it suffices to evaluate $\check{g}_1(\mathbf{X},\boldsymbol{\xi}^{(t)})$ and $\nabla \check{g}_1(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)})$ at $\mathbf{X}^{(t-1)}$ to see that (p2) and (p3) hold.

To show (p4), let us first find $\check{g}_1'$ and g_2'. Along a direction $\mathbf{D} := (\mathbf{D}_{\mathbf{P}},\mathbf{D}_{\mathbf{Q}},\mathbf{D}_{\mathbf{E}}) \in \mathcal{X}'$, it holds that $\check{g}_1'(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)};\mathbf{D}) = \langle \nabla \check{g}_1(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)}),\mathbf{D}\rangle$ since $\check{g}_1$ is differentiable. Similarly, $g_2'(\mathbf{X};\mathbf{D}) = \lambda(\langle \mathbf{P},\mathbf{D}_{\mathbf{P}}\rangle + \langle \mathbf{Q},\mathbf{D}_{\mathbf{Q}}\rangle) + \mu h'(\mathbf{E};\mathbf{D}_{\mathbf{E}})$ where $h(\mathbf{E}) := \|\mathbf{E}\|_1$, $\mathbf{d}_{\mathbf{E}} := \mathrm{vec}(\mathbf{D}_{\mathbf{E}})$ with its l-th entry being $d_{\mathbf{E},l}$, and

$$\begin{aligned}
&h'(\mathbf{E};\mathbf{D}_{\mathbf{E}}) \\
&:= \lim_{t\to 0+} \frac{h(\mathbf{E}+t\mathbf{D}_{\mathbf{E}}) - h(\mathbf{E})}{t} \\
&= \lim_{t\to 0+} \frac{\sum_{l,e_l\neq 0}(|e_l + t d_{\mathbf{E},l}| - |e_l|) + \sum_{l,e_l=0} |t d_{\mathbf{E},l}|}{t} \\
&= \sum_{l,e_l\neq 0} \mathrm{sgn}(e_l) d_{\mathbf{E},l} + \sum_{l,e_l=0} |d_{\mathbf{E},l}|. \quad (66)
\end{aligned}$$

On the other hand, the variation of $\check{g}$ can be written as

$$\begin{aligned}
&\check{g}(\mathbf{X}+\mathbf{D},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)}) - \check{g}(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)}) \\
&\quad = \check{g}_1'(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)};\mathbf{D}) + \sum_{i\in\{\mathbf{P},\mathbf{Q},\mathbf{E}\}} \frac{\eta_i^{(t)}}{2}\|\mathbf{D}_i\|_F^2 \\
&\qquad + g_2(\mathbf{X}+\mathbf{D}) - g_2(\mathbf{X}). \quad (67)
\end{aligned}$$

Note that $\sum_i \frac{\eta_i^{(t)}}{2}\|\mathbf{D}_i\|_F^2 \ge \frac{L_{\min}}{2}||\mathbf{D}||_F^2$ since $\eta_i^{(t)} \ge L_i^2/L_{\min}$ by algorithmic construction. Furthermore, $g_2(\mathbf{X}+\mathbf{D}) - g_2(\mathbf{X}) \ge g_2'(\mathbf{X};\mathbf{D})$ since g_2 is convex [36]. Then, the variation of $\check{g}$ in (67) can be lower-bounded as

$$\begin{aligned}
&\check{g}(\mathbf{X}+\mathbf{D},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)}) - \check{g}(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)}) \\
&\qquad \ge \check{g}'(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)};\mathbf{D}) + \frac{L_{\min}}{2}||\mathbf{D}||_F^2 \quad (68)
\end{aligned}$$

where $\check{g}'(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)};\mathbf{D}) = \check{g}_1'(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)};\mathbf{D}) + g_2'(\mathbf{X};\mathbf{D})$. Therefore, (p4) holds for a positive constant $\zeta \le L_{\min}$.

By the compactness of $\mathcal{X}$ and boundedness of Ξ by (a3), (p5) is automatically satisfied since g_1 and $\check{g}_1$ are continuously twice differentiable in $\mathbf{X}$ [31]. In addition, one can easily show (p6) since g_2 and g_2' are also uniformly bounded by the compactness of $\mathcal{X}$.

Let K_1 and K_2 denote constants where $|\check{g}_1| \le K_1$ and $|g_2| \le K_2$, respectively, by (p5) and (p6). Then, (p7) readily follows since

$$\begin{aligned}
|\check{g}(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)})| &= |\check{g}_1(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)}) + g_2(\mathbf{X})| \\
&\le |\check{g}_1(\mathbf{X},\mathbf{X}^{(t-1)},\boldsymbol{\xi}^{(t)})| + |g_2(\mathbf{X})| \\
&\le K_1 + K_2 =: \bar{g}. \quad (69)
\end{aligned}$$

■

The next lemma asserts that a distance between two subsequent estimates asymptotically goes to zero, which will be

used to show $\lim_{t\to\infty} \check{C}_{1,t}(\mathbf{X}^{(t)}) - C_{1,t}(\mathbf{X}^{(t)}) = 0$, almost surely.

Lemma 3: If (a2)–(a5) *hold, then* $||\mathbf{X}^{(t+1)} - \mathbf{X}^{(t)}||_F = O(1/t)$.

Proof: See [31, Lemma 2]. A proof of Lemma 3 is omitted to avoid duplication of the proof of [31, Lemma 2]. Hence, it suffices to mention that Lemma 2 guarantees the formulation of the proposed work satisfying the general assumptions on the formulation in [31]. ■

Lemma 3 does not guarantee convergence of the iterates to the stationary point of (P2). However, the final lemma asserts that the overestimated cost sequence converges to the cost of (P2), almost surely. Before proceeding to the next lemma, let us first define

$$C_{1,t}(\mathbf{X}) := \frac{1}{t}\sum_{\tau=1}^{t} g_1(\mathbf{X}, \boldsymbol{\xi}^{(\tau)}) \tag{70}$$

$$\check{C}_{1,t}(\mathbf{X}) := \frac{1}{t}\sum_{\tau=1}^{t} \check{g}_1(\mathbf{X}, \mathbf{X}^{(\tau-1)}, \boldsymbol{\xi}^{(\tau)}) \tag{71}$$

and $C_2(\mathbf{X}) := g_2(\mathbf{X})$. Note also that $\check{C}_t(\mathbf{X}) - C_t(\mathbf{X}) = \check{C}_{1,t}(\mathbf{X}) - C_{1,t}(\mathbf{X})$.

Lemma 4: If (a1)–(a5) *hold,* $\check{C}_t(\mathbf{X}^{(t)})$ *converges almost surely, and* $\lim_{t\to\infty} \check{C}_{1,t}(\mathbf{X}^{(t)}) - C_{1,t}(\mathbf{X}^{(t)}) = 0$, *almost surely.*

Proof: See [31, Lemma 1]. A proof of Lemma 4 is omitted to avoid duplication of the proof of [31, Lemma 1]. Instead, a sketch of the proof is following. It is firstly shown that the sequence $\{\check{C}_t(\mathbf{X}^{(t)})\}_{t=1}^{\infty}$ follows a quasi-martingale process and converges almost surely. Then, the lemma on positive converging sums (see [30, Lemma 8]) and Lemma 3 are used to claim that $\lim_{t\to\infty} \check{C}_{1,t}(\mathbf{X}^{(t)}) - C_{1,t}(\mathbf{X}^{(t)}) = 0$, almost surely. ■

The last step of the proof for Proposition 2 is inspired by [31]. Based on Lemma 4, it will be shown that the sequence $\{\nabla\check{C}_{1,t}(\mathbf{X}^{(t)}) - \nabla C_{1,t}(\mathbf{X}^{(t)})\}_{t=1}^{\infty}$ goes to zero, almost surely. Together with C_2', it follows that $\lim_{t\to\infty} C_t'(\mathbf{X}^{(t)}; \mathbf{D}) \geq 0\ \forall \mathbf{D}$, a.s. by algorithmic construction, implying convergence of a sequence $\{\mathbf{X}^{(t)}\}_{t=1}^{\infty}$ to the set of stationary points of $C(\mathbf{X})$.

By the compactness of $\mathcal{X}$, it is always possible to find a convergent subsequence $\{\mathbf{X}^{(t)}\}_{t=1}^{\infty}$ to a limit point $\bar{\mathbf{X}} \in \mathcal{X}$. Then, by the strong law of large numbers [37] under (a1) and equicontinuity of a family of functions $\{C_{1,t}(\cdot)\}_{t=1}^{\infty}$ due to the uniform boundedness of ∇g_1 in (p5) [38], upon restricting to the subsequence, one can have

$$\lim_{t\to\infty} C_{1,t}(\mathbf{X}^{(t)}) = \mathbb{E}_{\boldsymbol{\xi}}[g_1(\bar{\mathbf{X}}, \boldsymbol{\xi})] =: C_1(\mathbf{X}). \tag{72}$$

Similarly, a family of functions $\{\check{C}_{1,t}(\cdot)\}_{t=1}^{\infty}$ is equicontinuous due to the uniform boundedness of $\nabla\check{g}_1$ in (p5). Furthermore, $\{\check{C}_{1,t}(\cdot)\}_{t=1}^{\infty}$ is pointwisely bounded by (a1)–(a3). Thus, Arzelá-Ascoli theorem (see [38, Corollary 2.5] and [39]) implies that there exists a uniformly continuous function $\check{C}_1(\mathbf{X})$ such that $\lim_{t\to\infty} \check{C}_{1,t}(\mathbf{X}) = \check{C}_1(\mathbf{X})\ \forall\ \mathbf{X} \in \mathcal{X}$ and after restricting to the subsequence

$$\lim_{t\to\infty} \check{C}_{1,t}(\mathbf{X}^{(t)}) = \check{C}_1(\bar{\mathbf{X}}). \tag{73}$$

Furthermore, since $\check{g}_1(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \geq g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)})$ as in (p1), it follows that

$$\check{C}_{1,t}(\mathbf{X}) - C_{1,t}(\mathbf{X}) \geq 0\ \forall \mathbf{X}. \tag{74}$$

By letting $t \to \infty$ on (74) and combining Lemma 4 with (72) and (73), one deduces that

$$\check{C}_1(\bar{\mathbf{X}}) - C_1(\bar{\mathbf{X}}) = 0,\ \text{a.s.} \tag{75}$$

meaning that $\check{C}_{1,t}(\mathbf{X}) - C_{1,t}(\mathbf{X})$ takes its minimum at $\bar{\mathbf{X}}$ and

$$\nabla\check{C}_1(\bar{\mathbf{X}}) - \nabla C_1(\bar{\mathbf{X}}) = 0,\ \text{a.s.} \tag{76}$$

by the first-order optimality condition.

On the other hand, the fact that $\mathbf{X}^{(t)}$ minimizes $\check{C}_t(\mathbf{X})$ by algorithmic construction and g_2' exists (so does C_2'), yields

$$\check{C}_{1,t}(\mathbf{X}^{(t)}) + C_2(\mathbf{X}^{(t)}) \leq \check{C}_{1,t}(\mathbf{X}) + C_2(\mathbf{X}) \quad \forall \mathbf{X} \in \mathcal{X} \tag{77}$$

and $\lim_{t\to\infty} \check{C}_{1,t}(\mathbf{X}^{(t)}) + C_2(\mathbf{X}^{(t)}) \leq \lim_{t\to\infty} \check{C}_{1,t}(\mathbf{X}) + C_2(\mathbf{X})\ \forall \mathbf{X} \in \mathcal{X}$, which implies

$$\lim_{t\to\infty} \langle \nabla\check{C}_{1,t}(\mathbf{X}^{(t)}), \mathbf{D}\rangle + C_2'(\mathbf{X}^{(t)}; \mathbf{D}) \geq 0 \quad \forall \mathbf{D}. \tag{78}$$

Using the result in (76), (78) can be re-written as $\langle \nabla C_1(\bar{\mathbf{X}}), \mathbf{D}\rangle + C_2'(\bar{\mathbf{X}}; \mathbf{D}) \geq 0\ \forall \mathbf{D}$, a.s. or

$$C'(\bar{\mathbf{X}}; \mathbf{D}) \geq 0 \quad \forall \mathbf{D},\ \text{a.s.} \tag{79}$$

Thus, the subsequence $\{\mathbf{X}^{(t)}\}_{t=1}^{\infty}$ asymptotically coincides with the set of stationary points of $C(\mathbf{X})$. ■

ACKNOWLEDGEMENTS

The authors would like to thank Drs. Hamilton, Baxley, Matechik, and Ma for their helpful comments, and for kindly providing the real measurement data.

REFERENCES

[1] D. Lee and S.-J. Kim, "Channel gain cartography via low rank and sparsity," in *Proc. Asilomar Conf. Signal, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2014, pp. 1479–1483.

[2] *Spectrum Policy Task Force*, document ET Docket 02-135, FCC, Nov. 2002.

[3] S.-J. Kim, E. Dall'Anese, J. A. Bazerque, K. Rajawat, and G. B. Giannakis, "Advances in spectrum sensing and cross-layer design for cognitive radio networks," in *Academic Press Library in Signal Processing: Communications and Radar Signal Processing*, vol. 2. R. Chellapa and S. Theodoridis, Eds. San Diego, CA, USA: Academic, 2014, ch. 9, pp. 471–502.

[4] J. A. Bazerque, G. Mateos, and G. B. Giannakis, "Group-Lasso on splines for spectrum cartography," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4648–4663, Oct. 2011.

[5] S.-J. Kim, E. Dall'Anese, and G. B. Giannakis, "Cooperative spectrum sensing for cognitive radios using Kriged Kalman filtering," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 1, pp. 24–36, Feb. 2011.

[6] P. Agrawal and N. Patwari, "Correlated link shadow fading in multi-hop wireless networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4024–4036, Aug. 2009.

[7] E. Dall'Anese, S.-J. Kim, and G. B. Giannakis, "Channel gain map tracking via distributed Kriging," *IEEE Trans. Veh. Technol.*, vol. 60, no. 3, pp. 1205–1211, Mar. 2011.

[8] J. Wilson and N. Patwari, "Radio tomographic imaging with wireless networks," *IEEE Trans. Mobile Comput.*, vol. 9, no. 5, pp. 621–632, May 2010.

[9] J. Wilson and N. Patwari, "See-through walls: Motion tracking using variance-based radio tomography networks," *IEEE Trans. Mobile Comput.*, vol. 10, no. 5, pp. 612–621, May 2011.

[10] B. R. Hamilton, X. Ma, R. J. Baxley, and S. M. Matechik, "Radio frequency tomography in mobile networks," in *Proc. IEEE Statist. Signal Process. Workshop*, Ann Arbor, MI, USA, Aug. 2012, pp. 508–511.
[11] O. Kaltiokallio, M. Bocca, and N. Patwari, "Enhancing the accuracy of radio tomographic imaging using channel diversity," in *Proc. 9th IEEE Int. Conf. Mobile Ad Hoc Sensor Syst.*, Las Vegas, NV, USA, Oct. 2012, pp. 254–262.
[12] Y. Mostofi, "Compressive cooperative sensing and mapping in mobile networks," *IEEE Trans. Mobile Comput.*, vol. 10, no. 12, pp. 1769–1784, Dec. 2011.
[13] M. A. Kanso and M. G. Rabbat, "Compressed RF tomography for wireless sensor networks: Centralized and decentralized approaches," in *Proc. Int. Conf. Distrib. Comput. Sensor Syst.*, 2009, pp. 173–186.
[14] A. Moshtaghpour, M. A. Akhaee, and M. Attarifar, "Obstacle mapping in wireless sensor networks via minimum number of measurements," *IET Signal Process.*, vol. 10, no. 3, pp. 237–246, 2016.
[15] R. A. Valenzuela, "Ray tracing prediction of indoor radio propagation," in *Proc. IEEE 5th Int. Symp. Pers. Indoor Mobile Radio Commun.*, The Hague, The Netherlands, Sep. 1994, pp. 140–144.
[16] G. Wölfle, R. Wahl, P. Wertz, P. Wildbolz, and F. Landstorfer, "Dominant path prediction model for indoor scenarios," in *Proc. German Microw. Conf.*, 2005, pp. 176–179.
[17] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, May 2011, Art. no. 11.
[18] J. Wright, A. Ganesh, K. Min, and Y. Ma, "Compressive principal component pursuit," *Inf. Inference*, vol. 2, no. 1, pp. 32–68, Jun. 2013.
[19] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *Electron. Lett.*, vol. 27, no. 23, pp. 2145–2146, Nov. 1991.
[20] F. Graziosi and F. Santucci, "A general correlation model for shadow fading in mobile radio systems," *IEEE Commun. Lett.*, vol. 6, no. 3, pp. 102–104, Mar. 2002.
[21] B. R. Hamilton, X. Ma, R. J. Baxley, and S. M. Matechik, "Propagation modeling for radio frequency tomography in wireless networks," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 1, pp. 55–65, Feb. 2014.
[22] A. Eriksson and A. van den Hengel, "Efficient computation of robust low-rank matrix approximations in the presence of missing data using the L_1 norm," in *Proc. 23rd IEEE Conf. Comput. Vis. Pattern Recognit.*, San Francisco, CA, USA, Jun. 2010, pp. 771–778.
[23] M. Mardani, G. Mateos, and G. B. Giannakis, "Recovery of low-rank plus compressed sparse matrices with application to unveiling traffic anomalies," *IEEE Trans. Inf. Theory*, vol. 59, no. 8, pp. 5186–5205, Aug. 2013.
[24] M. Mardani, G. Mateos, and G. B. Giannakis, "Dynamic anomalography: Tracking network anomalies via sparsity and low rank," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 1, pp. 50–66, Feb. 2013.
[25] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, Aug. 2010.
[26] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *J. Optim. Theory Appl.*, vol. 109, pp. 475–494, Jun. 2001.
[27] K. Slavakis, S.-J. Kim, G. Mateos, and G. B. Giannakis, "Stochastic approximation vis-a-vis online learning for big data analytics," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 124–129, Nov. 2014.
[28] H. Kushner and G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*. New York, NY, USA: Springer-Verlag, 2003.
[29] M. Mardani, G. Mateos, and G. B. Giannakis, "Subspace learning and imputation for streaming big data matrices and tensors," *IEEE Trans. Signal Process.*, vol. 63, no. 10, pp. 2663–2677, May 2015.
[30] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, Mar. 2010.
[31] M. Razaviyayn, M. Sanjabi, and Z.-Q. Luo, "A stochastic successive minimization method for nonsmooth nonconvex optimization with applications to transceiver design in wireless communication networks," *Math. Program.*, vol. 157, no. 2, pp. 515–545, 2016.
[32] Z. Qin, D. Goldfarb, and S. Ma, "An alternating direction method for total variation denoising," *Optim. Method Softw.*, vol. 30, no. 3, pp. 594–615, 2015.
[33] Z. Lin, A. Ganesh, J. Right, L. Wu, M. Chen, and Y. Ma, "Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix," Dept. ECE, UIUC, Champaign, IL, USA, Tech. Rep. UILU-ENG-09-2214, 2009, pp. 1–18.
[34] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[35] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA, USA: Athena Scientific, 1999.
[36] Y. Nesterov, *Introductory Lectures on Convex optimization: A Basic Course*. Boston, MA, USA: Kluwer, 2004.
[37] B. E. Fristedt and L. F. Gray, *A Modern Approach to Probability Theory*. Boston, MA, USA: Birkhäuser, 1997.
[38] R. F. Brown, *A Topological Introduction to Nonlinear Analysis*. Boston, MA, USA: Birkhäuser, 2004.
[39] N. Dunford and J. T. Schwartz, *Linear Operators. Part 1: General Theory*. New York, NY, USA: Interscience, 1958.

Donghoon Lee (S'12) received the B.Eng. and M.Eng. degrees in electrical engineering from Korea University, Seoul, South Korea, in 2010 and 2013, respectively. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Minnesota, Twin Cities, MN, USA. From 2010 to 2011, he was with the Korea Institute of Science and Technology, Seoul.

His current research interests include statistical signal processing, big data analytics, and machine learning, with applications to wireless communication and networking.

Seung-Jun Kim (SM'12) received the B.S. and M.S. degrees from Seoul National University, Seoul, South Korea, in 1996 and 1998, respectively, and the Ph.D. degree from the University of California at Santa Barbara, CA, USA, in 2005, all in electrical engineering. From 2005 to 2008, he was with NEC Laboratories America, Princeton, NJ, USA, as a Research Staff Member. He was with the Digital Technology Center and the Department of Electrical and Computer Engineering, University of Minnesota, MN, USA, from 2008 to 2014, where his final title was Research Associate Professor. Since 2014, he has been with Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, MD, USA, as an Assistant Professor. His research interests include statistical signal processing, optimization, and machine learning, with applications to wireless communication and networking, future power systems, and (big) data analytics. He is serving as an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS.

Georgios B. Giannakis (F'97) received the Diploma degree in electrical engineering from the National Technical University of Athens, Greece, in 1981, the M.Sc. degree in electrical engineering and the M.Sc. degree in mathematics from the University of Southern California in 1983 and 1986, respectively, and the Ph.D. degree in electrical engineering from the University of Southern California in 1986. He was with the University of Virginia from 1987 to 1998. Since 1999, he has been a Professor with the University of Minnesota, where he holds an Endowed Chair in wireless telecommunications, a University of Minnesota McKnight Presidential Chair in ECE, and serves as the Director of the Digital Technology Center.

He is the co-inventor of 30 patents issued. His general interests span the areas of communications, networking and statistical signal processing, subjects on which he has authored over 400 journal papers, 700 conference papers, 25 book chapters, two edited books, and two research monographs (h-index 126). Current research focuses on learning from big data, wireless cognitive radios, and network science with applications to social, brain, and power networks with renewables. He is a fellow of the EURASIP. He has served the IEEE in a number of posts, including as a Distinguished Lecturer of the IEEE Signal Processing (SP) Society. He was a co-recipient of eight best paper awards from the IEEE-SP Society and the Communications Society, including the G. Marconi Prize Paper Award in Wireless Communications. He also received the Technical Achievement Award from the SP Society (2000), from the EURASIP (2005), the Young Faculty Teaching Award, the G. W. Taylor Award for Distinguished Research from the University of Minnesota, and the IEEE Fourier Technical Field Award (2015).