

Channel Gain Cartography via Low Rank and Sparsity

Donghoon Lee
 Dept. of ECE
 University of Minnesota
 Minneapolis, MN 55455, USA
 leex6962@umn.edu

Seung-Jun Kim
 Dept. of CSEE
 University of Maryland, Baltimore County
 Baltimore, MD 21250, USA
 sjkim@umbc.edu

Abstract—Channel gain cartography aims at inferring the channel gains between arbitrary points in space based on measurements (samples) of channel gains taken from finite pairs of transceivers. Channel gain maps are useful for various sensing and resource allocation tasks, essential for the operation of cognitive radio networks. In this work, the channel gain samples are modeled as compressive tomographic measurements of an underlying spatial loss field (SLF), postulated to have low-rank structure corrupted by sparse errors. Efficient algorithms to reconstruct the SLF are developed, from which arbitrary channel gains can be interpolated.

I. INTRODUCTION

RF cartography is an instrumental concept for spectrum sensing and dynamic spectrum access of cognitive radio (CR) networks [1]. Based on the measurements collected by spatially distributed CR sensors, RF cartography aims at constructing maps over space, time, and frequency, of useful information regarding the RF “landscape,” in which the CR network is deployed. Two kinds of maps have been proposed in the literature: i) power spectral density (PSD) maps that reconstruct the ambient interference power distribution, revealing crowded regions that CR transceivers need to avoid [2]; and ii) channel gain maps which capture shadow fading between arbitrary points in space, thus allowing CR networks to perform accurate spectrum sensing and aggressive spatial reuse [3].

The present work focuses on channel gain cartography. Prior works capitalized on the experimentally validated spatial loss field (SLF) model [4] and developed linear interpolation algorithms based on kriging techniques [3], [5]. Spatio-temporal dynamics were captured using a Kalman filtering based approach. It is worth noting that the reconstruction of the SLF itself is tantamount to RF tomographic imaging [6]. Tracking of moving objects using dynamic SLF models was investigated in [7]. However, in these works, relatively densely sampled observations were utilized for acceptable performance.

In this work, we aim to reconstruct the SLF (and subsequently interpolate channel gains) from a small number of measurements, by tapping into the recent advances in compressive sensing techniques. Specifically, it is postulated that the SLF of the two-dimensional geographical area of CR network deployment is of low rank, potentially corrupted by sparse outliers. Such a model is particularly appealing for urban and

indoor propagation scenarios, where regular placement of walls and buildings renders the scene inherently low-rank, while sparse outliers can pick up artifacts that do not conform to the low-rank model. As channel gain samples can be viewed as tomographic measurements of the SLF, the cartography problem becomes an instance of compressive principal component pursuit (CPCP) [8]. We develop computationally efficient batch and online solutions, suitable for CR network implementation.

The rest of the paper is organized as follows. In Section II, the system model is set forth. Problem formulation for channel gain prediction and its batch solution method are the subjects of Section III. An online algorithm is developed in Section IV. Results from numerical tests are provided in Section V, and the conclusions 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. Operators $(\cdot)^T$, $\text{tr}(\cdot)$, and $\sigma_i(\cdot)$ represent transposition, matrix trace, and i -th singular value. $|\cdot|$ is used for the cardinality of a set and the magnitude of a scalar. The ℓ_1 norm of $\mathbf{X} \in \mathbb{R}^{n \times n}$ is $\|\mathbf{X}\|_1 := \sum_{i,j=1}^n |X_{ij}|$. For two matrices $\mathbf{M}, \mathbf{R} \in \mathbb{R}^{n \times n}$, matrix inner product is defined as $\langle \mathbf{M}, \mathbf{R} \rangle := \text{tr}(\mathbf{M}^T \mathbf{R})$. The Frobenius norm of matrix \mathbf{R} is $\|\mathbf{R}\|_F := \sqrt{\text{tr}(\mathbf{R}\mathbf{R}^T)}$. The spectral norm of \mathbf{R} is $\|\mathbf{R}\| := \max_{\|\mathbf{x}\|_2=1} \|\mathbf{R}\mathbf{x}\|_2$, and $\|\mathbf{R}\|_* := \sum_i \sigma_i(\mathbf{R})$ is the nuclear norm of \mathbf{R} .

II. SYSTEM MODEL AND PROBLEM STATEMENT

Consider a set of N CRs deployed in a geographical area represented by a 2-dimensional plane $\mathcal{A} \subset \mathbb{R}^2$. Let $\mathbf{x}_n^{(t)}$ denote the position of CR $n \in \{1, 2, \dots, N\}$ 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 in dB scale for link (n, n') at time t , denoted by $\check{g}_{\mathbf{x}_n \mathbf{x}_{n'}}^{(t)}$, can be represented (in dBs) as

$$\check{g}_{\mathbf{x}_n \mathbf{x}_{n'}}^{(t)} = g_0 - \gamma \log_{10} \|\mathbf{x}_n^{(t)} - \mathbf{x}_{n'}^{(t)}\| + s_{\mathbf{x}_n \mathbf{x}_{n'}}^{(t)} + \epsilon_{\mathbf{x}_n \mathbf{x}_{n'}}^{(t)} \quad (1)$$

where g_0 is the path gain at unit distance; $\|\mathbf{x}_n^{(t)} - \mathbf{x}_{n'}^{(t)}\|$ the distance between nodes n and n' ; γ the pathloss exponent, $s_{\mathbf{x}_n \mathbf{x}_{n'}}^{(t)}$ the shadow fading in dB over link (n, n') at time t ,

and $\epsilon_{\mathbf{x}_n \mathbf{x}_{n'}}^{(t)}$, the measurement noise. By subtracting the known pathloss component, the shadowing measurement

$$\check{s}_{\mathbf{x}_n \mathbf{x}_{n'}}^{(t)} = s_{\mathbf{x}_n \mathbf{x}_{n'}}^{(t)} + \epsilon_{\mathbf{x}_n \mathbf{x}_{n'}}^{(t)} \quad (2)$$

is obtained. Let $\mathcal{M}^{(t)}$ be the set of links, for which channel gain measurements were made at time t , and collect the 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 $g_{\mathbf{x}\mathbf{x}'}$ 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, $\{\check{\mathbf{s}}^{(\tau)}\}_{\tau=1}^t$ [3], [5].

In order to achieve this interpolation, we will leverage the structure of shadow fading experienced by co-located radio links. A variety of correlation models for shadow fading have been proposed [3], [4], [9]. In particular, the models in [3] and [4] relied on the so-termed *spatial loss field* (SLF), which captures the attenuation in the propagating medium. Tomographic imaging applications based on such models have been pursued in [6] and [7].

Let $\mathbf{F} \in \mathbb{R}^{N_x \times N_y}$ denote the SLF, discretized to an N_x -by- N_y matrix. Then, the shadow fading of the link between points \mathbf{x} and \mathbf{x}' in \mathcal{A} is modeled as a linear combination of the field elements expressed as

$$s_{\mathbf{x}\mathbf{x}'} = \langle \mathbf{W}_{\mathbf{x}\mathbf{x}'}, \mathbf{F} \rangle \quad (3)$$

where $\mathbf{W}_{\mathbf{x}\mathbf{x}'}$ is a weight matrix corresponding to the link connecting \mathbf{x} and \mathbf{x}' . Matrix $\mathbf{W}_{\mathbf{x}\mathbf{x}'}$ is often constructed so that the shadow fading is a weighted sum of the loss field within an ellipse having the foci at \mathbf{x} and \mathbf{x}' [6]. Specifically,

$$\mathbf{W}_{\mathbf{x}\mathbf{x}'}(i, j) := \begin{cases} 1/\sqrt{d}, & \text{if } d_{ij}(\mathbf{x}) + d_{ij}(\mathbf{x}') < d + \delta \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where $d_{ij}(\mathbf{x})$ is the distance from the position represented by grid point (i, j) to \mathbf{x} , $d := \|\mathbf{x} - \mathbf{x}'\|_2$, and $d + \delta$ is the major axis of the ellipse.

Using (4), 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 $\mathbf{x}-\mathbf{x}'$ can be estimated via (3) and (1) as

$$\hat{s}_{\mathbf{x}\mathbf{x}'}^{(t)} = \langle \mathbf{W}_{\mathbf{x}\mathbf{x}'}, \hat{\mathbf{F}}^{(t)} \rangle \quad (5)$$

$$\hat{g}_{\mathbf{x}\mathbf{x}'}^{(t)} = g_0 - \gamma \log_{10} \|\mathbf{x} - \mathbf{x}'\| + \hat{s}_{\mathbf{x}\mathbf{x}'}^{(t)}. \quad (6)$$

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 also improve the performance even in the overdetermined cases, a priori knowledge on $\mathbf{F}^{(t)}$ will be exploited next to regularize the problem, using the attributes of low rank and sparsity.

III. CHANNEL GAIN PREDICTION USING LOW RANK AND SPARSITY

A. Problem Formulation

The low-rank and sparse structure has been advocated in various problems in machine learning and signal processing [10], [11]. Low-rank matrices are effective in capturing

regular patterns, and sparsity is instrumental in incorporating robustness. In this work, we postulate that \mathbf{F} has the following low-rank plus sparse structure

$$\mathbf{F}^{(t)} = \mathbf{L}^{(t)} + \mathbf{E}^{(t)} \quad (7)$$

where $\mathbf{L}^{(t)}$ has low rank and $\mathbf{E}^{(t)}$ 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 the irregularities that do not conform to the low-rank model.

Toward estimating $\mathbf{F}^{(t)}$ that obeys (7), let $\mathbf{W}_{nn'}^{(t)} := \mathbf{W}_{\mathbf{x}_n^{(t)} \mathbf{x}_{n'}^{(t)}}$ and $\check{s}_{nn'}^{(t)} := \check{s}_{\mathbf{x}_n \mathbf{x}_{n'}}^{(t)}$, for brevity, and define the cost

$$c^{(t)}(\mathbf{L}, \mathbf{E}) := \frac{1}{2} \sum_{(n, n') \in \mathcal{M}^{(t)}} \left(\langle \mathbf{W}_{nn'}^{(t)}, \mathbf{L} + \mathbf{E} \rangle - \check{s}_{nn'}^{(t)} \right)^2 \quad (8)$$

which fits the shadowing measurements to the model. Then, the low-rank and sparse components can be recovered via

$$(P1) \quad \min_{\mathbf{L}, \mathbf{E} \in \mathbb{R}^{N_x \times N_y}} \sum_{\tau=1}^t \beta^{t-\tau} \left[c^{(\tau)}(\mathbf{L}, \mathbf{E}) + \lambda \|\mathbf{L}\|_* + \mu \|\mathbf{E}\|_1 \right] \quad (9)$$

where $\beta \in (0, 1]$ is the forgetting factor that weights the recent observations more heavily. The nuclear norm promotes low-rank of \mathbf{L} , and the ℓ_1 -norm encourages sparsity in \mathbf{E} . Parameters λ and μ are chosen appropriately to control the effect of these regularizing norms. Conditions for exact recovery through a related convex formulation in the absence of measurement noise were provided in [8].

Problem (9) is a convex optimization problem, and can be tackled using existing efficient solution techniques, such as the interior-point method and the accelerated proximal gradient (APG) method [12]. Once the optimal $\hat{\mathbf{E}}$ and $\hat{\mathbf{L}}$ are found, the desired $\hat{\mathbf{F}}^{(t)}$ is obtained as $\hat{\mathbf{F}}^{(t)} = \hat{\mathbf{L}} + \hat{\mathbf{E}}$. However, general-purpose optimization software tends to scale poorly as the problem size grows. Specialized algorithms developed for related problems often employ costly SVD operations iteratively [8]. Furthermore, such an algorithm might not be amenable for an online and distributed implementation. Next, an efficient algorithm is proposed with reduced complexity.

B. Efficient Solution

Consider replacing \mathbf{L} with the low-rank product $\mathbf{P}\mathbf{Q}^T$, 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., [11])

$$\|\mathbf{L}\|_* = \min_{\mathbf{P}, \mathbf{Q}} \frac{1}{2} (\|\mathbf{P}\|_F^2 + \|\mathbf{Q}\|_F^2) \quad (10)$$

subject to $\mathbf{L} = \mathbf{P}\mathbf{Q}^T$.

Thus, a natural re-formulation of (9) is (see also [11])

$$(P2) \quad \min_{\mathbf{P}, \mathbf{Q}, \mathbf{E}} \sum_{\tau=1}^t \beta^{t-\tau} \left[c^{(\tau)}(\mathbf{P}\mathbf{Q}^T, \mathbf{E}) + \frac{\lambda}{2} (\|\mathbf{P}\|_F^2 + \|\mathbf{Q}\|_F^2) + \mu \|\mathbf{E}\|_1 \right]. \quad (11)$$

Instead of $N_x N_y$ entries of \mathbf{L} , the factorization approach (11) entails $(N_x + N_y)\rho$ unknowns, thus reducing complexity and memory requirement for $\rho \ll N_x, N_y$.

Since (P2) is nonconvex while the original (P1) is convex, in general one expects only a *locally* optimal solution of (P2). Interestingly, under appropriate conditions, it can be shown that global optimality can be guaranteed even if (P2) is solved, as claimed next [11].

Proposition 1: *Let $\{\bar{\mathbf{P}}, \bar{\mathbf{Q}}, \bar{\mathbf{E}}\}$ be a stationary point of (P2). Define $\bar{\beta} := \sum_{\tau=1}^t \beta^{t-\tau}$ and $\bar{f}(\bar{\mathbf{L}}, \bar{\mathbf{E}}) := \sum_{x=1}^t \beta^{t-\tau} \nabla_{c^{(\tau)}}(\bar{\mathbf{L}}, \bar{\mathbf{E}})$. If $\|\bar{f}(\bar{\mathbf{P}}\bar{\mathbf{Q}}^T, \bar{\mathbf{E}})\| \leq \lambda\bar{\beta}$, then $\{\tilde{\mathbf{L}} := \bar{\mathbf{P}}\bar{\mathbf{Q}}^T, \tilde{\mathbf{E}} := \bar{\mathbf{E}}\}$ is a globally optimal solution to (P1).*

A stationary point of (11) can be obtained through a block coordinate-descent (BCD) algorithm, where the optimization is done over one of \mathbf{E} , \mathbf{P} , \mathbf{Q} with the other two fixed. In fact, since $\mu\|\mathbf{E}\|_1$ is separable over individual entries, the cyclic update can be stretched all the way up to the individual entries of \mathbf{E} without affecting convergence [13]. To update each (block) variable, the cost in (P2) is minimized while fixing other variables to their up-to-date iterates.

To describe the procedure, let $\mathcal{W}^{(t)} \in \mathbb{R}^{N_x N_y \times |\mathcal{M}(t)|}$ be a matrix with columns equal to $\text{vec}(\mathbf{W}_{nn'}^{(t)})$ for $(n, n') \in \mathcal{M}(t)$, where $\text{vec}(\cdot)$ produces a column vector by stacking the columns of a matrix one on top of the other. Define $\mathcal{W} := [\sqrt{\beta^{t-1}}\mathcal{W}^{(1)}, \dots, \sqrt{\beta^0}\mathcal{W}^{(t)}]$, $\tilde{\mathbf{s}} := [\sqrt{\beta^{t-1}}\tilde{\mathbf{s}}^{(1)\mathcal{T}}, \dots, \sqrt{\beta^0}\tilde{\mathbf{s}}^{(t)\mathcal{T}}]^{\mathcal{T}}$, and $\mathbf{e} := \text{vec}(\mathbf{E})$. Let e_l denote the l -th entry of \mathbf{e} and $\mathbf{e}(-l)$ represent the replica of \mathbf{e} with its l -th entry removed. Similarly, let $\mathbf{A}(l, :)$ denote the l -th row of matrix \mathbf{A} , and $\mathbf{A}(-l, :)$ the matrix \mathbf{A} with its l -th row removed. The minimization in (P2) with respect to \mathbf{E} decomposes over the individual entries of \mathbf{e} . At iteration k , each entry is updated via

$$e_l[k+1] = \arg \min_{e_l} \frac{1}{2} \|\mathcal{W}^{\mathcal{T}}(l, :)e_l - \tilde{\mathbf{s}}\|_2^2 + \mu\bar{\beta}|e_l| \quad (12)$$

where $\tilde{\mathbf{s}}[k] := \tilde{\mathbf{s}} - \mathcal{W}^{\mathcal{T}} \text{vec}(\mathbf{P}[k]\mathbf{Q}[k]^{\mathcal{T}}) - \mathcal{W}(-l, :)^{\mathcal{T}}\mathbf{e}(-l)$. Define the soft-thresholding function $\text{soft_th}_{\lambda}(\cdot)$ as $\text{soft_th}_{\mu}(x) := \text{sgn}(x) \max\{0, |x| - \mu\}$. A closed-form solution for e_l is obtained as

$$e_l[k+1] = \frac{\text{soft_th}_{\mu\bar{\beta}}(\mathcal{W}(l, :)\tilde{\mathbf{s}}[k])}{\|\mathcal{W}^{\mathcal{T}}(l, :)\|_2^2}. \quad (13)$$

Matrices \mathbf{P} and \mathbf{Q} are similarly updated over their rows. Let $\mathbf{p}_i^{\mathcal{T}}$ be the i -th row of \mathbf{P} , transposed to result in a column vector; i.e., $\mathbf{P} := [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_{N_x}]^{\mathcal{T}}$. Define $\tilde{\mathcal{W}}_i^{(t)} \in \mathbb{R}^{|\mathcal{M}(t)| \times N_y}$ to be the matrix whose rows are the i -th rows of $\{\mathbf{W}_{nn'}^{(t)}\}_{(n, n') \in \mathcal{M}(t)}$, and $\tilde{\mathbf{s}}_i^{(t)} \in \mathbb{R}^{|\mathcal{M}(t)|}$ a vector with entries equal to $\tilde{s}_{nn', i}^{(t)} = \tilde{s}_{nn', i}^{(t)} - \langle \mathbf{W}_{nn'}^{(t)}, \mathbf{E}[k+1] \rangle + \sum_{j \neq i} \mathbf{W}_{nn'}^{(t)}(j, :) \mathbf{Q}[k] \mathbf{p}_j$ for $(n, n') \in \mathcal{M}(t)$. Define also $\tilde{\mathcal{W}}_i := [\sqrt{\beta^{t-1}}\tilde{\mathcal{W}}_i^{(1)\mathcal{T}}, \dots, \sqrt{\beta^0}\tilde{\mathcal{W}}_i^{(t)\mathcal{T}}]^{\mathcal{T}}$ and $\tilde{\mathbf{s}}_i := [\sqrt{\beta^{t-1}}\tilde{\mathbf{s}}_i^{(1)\mathcal{T}}, \dots, \sqrt{\beta^0}\tilde{\mathbf{s}}_i^{(t)\mathcal{T}}]^{\mathcal{T}}$. Then, \mathbf{p}_i is updated by solving a ridge-regression problem given as

$$\mathbf{p}_i[k+1] = \arg \min_{\mathbf{p}_i} \left[\frac{1}{2} \|\tilde{\mathcal{W}}_i \mathbf{Q}[k] \mathbf{p}_i - \tilde{\mathbf{s}}_i\|_2^2 + \frac{\lambda\bar{\beta}}{2} \|\mathbf{p}_i\|_2^2 \right] \quad (14)$$

whose solution is given by

$$\mathbf{p}_i[k+1] = \left[\mathbf{Q}[k]^{\mathcal{T}} \tilde{\mathcal{W}}_i^{\mathcal{T}} \tilde{\mathcal{W}}_i \mathbf{Q}[k] + \lambda\bar{\beta} \mathbf{I}_{N_y} \right]^{-1} \mathbf{Q}[k]^{\mathcal{T}} \tilde{\mathcal{W}}_i^{\mathcal{T}} \tilde{\mathbf{s}}_i. \quad (15)$$

TABLE I. BATCH BCD UPDATE FOR (P2).

1: Initialize $\mathbf{E}[1] := \mathbf{0}_{N_x \times N_y}$, $\mathbf{P}[1]$ and $\mathbf{Q}[1]$ at random.
2: For $k = 1, 2, \dots$
3: For $l = 1, 2, \dots, N_x N_y$
4: Update $e_l[k+1]$ via (13)
5: Next l
6: For $i = 1, 2, \dots, N_x$
7: Update $\mathbf{p}_i[k+1]$ via (15)
8: Next i
9: For $j = 1, 2, \dots, N_y$
10: Update $\mathbf{q}_j[k+1]$ via (16)
11: Next j
12: Next k

Likewise, define $\tilde{\mathcal{W}}_i := [\sqrt{\beta^{t-1}}\tilde{\mathcal{W}}_i^{(1)\mathcal{T}}, \dots, \sqrt{\beta^0}\tilde{\mathcal{W}}_i^{(t)\mathcal{T}}]^{\mathcal{T}}$ and $\tilde{\mathbf{s}}_i := [\sqrt{\beta^{t-1}}\tilde{\mathbf{s}}_i^{(1)\mathcal{T}}, \dots, \sqrt{\beta^0}\tilde{\mathbf{s}}_i^{(t)\mathcal{T}}]^{\mathcal{T}}$, where $\tilde{\mathcal{W}}_i^{(t)} \in \mathbb{R}^{|\mathcal{M}(t)| \times N_x}$ is the matrix whose rows are the transpositions of the i -th columns of $\{\mathbf{W}_{nn'}^{(t)}\}_{(n, n') \in \mathcal{M}(t)}$, and $\tilde{\mathbf{s}}_i^{(t)} \in \mathbb{R}^{|\mathcal{M}(t)|}$ has entries equal to $\tilde{s}_{nn', i}^{(t)} = \tilde{s}_{nn', i}^{(t)} - \langle \mathbf{W}_{nn'}^{(t)}, \mathbf{E}[k+1] \rangle + \sum_{j \neq i} \mathbf{W}_{nn'}^{(t)}(j, :) \mathbf{P}[k+1] \mathbf{q}_j$ for $(n, n') \in \mathcal{M}(t)$. The update for \mathbf{q}_i is then given by

$$\mathbf{q}_i[k+1] = \left[\mathbf{P}[k+1]^{\mathcal{T}} \tilde{\mathcal{W}}_i^{\mathcal{T}} \tilde{\mathcal{W}}_i \mathbf{P}[k+1] + \lambda\bar{\beta} \mathbf{I}_{N_x} \right]^{-1} \cdot \mathbf{P}[k+1]^{\mathcal{T}} \tilde{\mathcal{W}}_i^{\mathcal{T}} \tilde{\mathbf{s}}_i. \quad (16)$$

The overall algorithm is tabulated in Table I.

IV. ONLINE ALGORITHM

Although the proposed batch algorithm exhibits low computational and memory requirements, it is not suitable for online processing, since (11) must be re-solved every time a new set of measurements arrive, thus incurring a major computational burden. Furthermore, the batch solution cannot track effectively time-varying channels. These considerations motivate the development of an efficient recursive algorithm.

Stochastic approximation (SA) is an appealing strategy for deriving online algorithms [14]. Recently, techniques involving minimizing majorized surrogate functions were developed to handle nonconvexity in online settings [15], [11], [16]. Improving upon these works, an online algorithm for the CPCP problem will be developed here. The proposed approach employs quadratic surrogate functions with diagonal weighting so as to capture disparate curvatures in the directions of different block variables.

Define $\mathbf{X} := \{\mathbf{P}, \mathbf{Q}, \mathbf{E}\} \in \mathcal{X} \subset \mathcal{X}'$, where \mathcal{X}' is an open set, and $\boldsymbol{\xi}^{(t)} := \{\{\tilde{\mathbf{s}}_m^{(t)}\}_{m=1}^M, \{\mathbf{W}_m^{(t)}\}_{m=1}^M\} \in \mathcal{E}$, where, for simplicity, the number of measurements per time slot is fixed to $M = |\mathcal{M}(t)|$ per slot t . With some abuse of notation, define $g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)}) := g_1(\mathbf{P}, \mathbf{Q}, \mathbf{E}, \boldsymbol{\xi}^{(t)}) = \frac{1}{2} \sum_{m=1}^M \left(\langle \mathbf{W}_m^{(t)}, \mathbf{P}\mathbf{Q}^{\mathcal{T}} + \mathbf{E} \rangle - \tilde{s}_m^{(t)} \right)^2$ and $g_2(\mathbf{X}) := g_2(\mathbf{P}, \mathbf{Q}, \mathbf{E}) = \frac{\lambda}{2} (\|\mathbf{P}\|_F^2 + \|\mathbf{Q}\|_F^2) + \mu\|\mathbf{E}\|_1$.

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

$$\begin{aligned} \check{g}_1(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) &:= g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \\ &+ \langle \mathbf{P} - \mathbf{P}^{(t-1)}, \nabla_{\mathbf{P}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \rangle + \frac{\eta_{\mathbf{P}}^{(t)}}{2} \|\mathbf{P} - \mathbf{P}^{(t-1)}\|_F^2 \\ &+ \langle \mathbf{Q} - \mathbf{Q}^{(t-1)}, \nabla_{\mathbf{Q}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \rangle + \frac{\eta_{\mathbf{Q}}^{(t)}}{2} \|\mathbf{Q} - \mathbf{Q}^{(t-1)}\|_F^2 \\ &+ \langle \mathbf{E} - \mathbf{E}^{(t-1)}, \nabla_{\mathbf{E}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \rangle + \frac{\eta_{\mathbf{E}}^{(t)}}{2} \|\mathbf{E} - \mathbf{E}^{(t-1)}\|_F^2. \end{aligned} \quad (17)$$

Upon defining $h_m(\mathbf{X}^{(t-1)}) := \langle \mathbf{W}_m^{(t)}, \mathbf{P}^{(t-1)} \mathbf{Q}^{(t-1)\top} + \mathbf{E}^{(t-1)} \rangle - \check{s}_m^{(t)}$, it can be verified that

$$\nabla_{\mathbf{P}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) = \sum_{m=1}^M h_m(\mathbf{X}^{(t-1)}) \mathbf{W}_m^{(t)} \mathbf{Q}^{(t-1)} \quad (18)$$

$$\nabla_{\mathbf{Q}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) = \sum_{m=1}^M h_m(\mathbf{X}^{(t-1)}) \mathbf{W}_m^{(t)\top} \mathbf{P}^{(t-1)} \quad (19)$$

$$\nabla_{\mathbf{E}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) = \sum_{m=1}^M h_m(\mathbf{X}^{(t-1)}) \mathbf{W}_m^{(t)}. \quad (20)$$

Suppose that $\nabla_{\mathbf{P}} g_1(\cdot, \mathbf{Q}, \mathbf{E}, \boldsymbol{\xi})$, $\nabla_{\mathbf{Q}} g_1(\cdot, \mathbf{P}, \mathbf{E}, \boldsymbol{\xi})$ and $\nabla_{\mathbf{E}} g_1(\cdot, \mathbf{P}, \mathbf{Q}, \boldsymbol{\xi})$ are Lipschitz (which can be asserted through Cauchy-Schwartz's inequality), and that the corresponding Lipschitz constants are $L_{\mathbf{P}}$, $L_{\mathbf{Q}}$, and $L_{\mathbf{E}}$, respectively. Then, upon choosing $\eta_i^{(t)} \geq L_i^2/L_{\min}$ for $i \in \{\mathbf{P}, \mathbf{Q}, \mathbf{E}\}$ and $L_{\min} := \min\{L_{\mathbf{P}}, L_{\mathbf{Q}}, L_{\mathbf{E}}\}$, it can be verified that $\check{g}_1(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)})$ satisfies the following conditions:

- (c1) $\check{g}_1(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)})$ majorizes $g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)}) \forall \mathbf{X} \in \mathcal{X}'$, i.e., $\check{g}_1(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \geq g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)})$;
- (c2) $\check{g}_1(\mathbf{X}^{(t-1)}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) = g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)})$;
- (c3) $\nabla \check{g}_1(\mathbf{X}^{(t-1)}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) = \nabla g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)})$; and
- (c4) $\check{g}_1(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) + g_2(\mathbf{X})$ is uniformly strongly convex in \mathbf{X} .

Consider the case of $\beta = 1$. Based on $\check{g}_1(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)})$ and $g_2(\mathbf{X})$, a surrogate problem for (P2) is given by

$$(P3) \quad \min_{\mathbf{X}} \frac{1}{t} \sum_{\tau=1}^t \left[\check{g}_1(\mathbf{X}, \mathbf{X}^{(\tau-1)}, \boldsymbol{\xi}^{(\tau)}) + g_2(\mathbf{X}) \right]. \quad (21)$$

Solving (P3) yields a stochastic gradient descent (SGD) update. In particular, since variables \mathbf{P} , \mathbf{Q} , and \mathbf{E} can be independently optimized in (P3), the proposed solver can perform parallel updates across variables. For \mathbf{P} and \mathbf{Q} , the update rules are obtained by checking the first-order optimality conditions, as

$$\mathbf{P}^{(t)} = \mathbf{P}^{(t-1)} - \frac{\lambda \mathbf{P}^{(t-1)} - \nabla_{\mathbf{P}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)})}{\sum_{\tau=1}^t (\eta_{\mathbf{P}}^{(\tau)} + \lambda)} \quad (22)$$

$$\mathbf{Q}^{(t)} = \mathbf{Q}^{(t-1)} - \frac{\lambda \mathbf{Q}^{(t-1)} - \nabla_{\mathbf{Q}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)})}{\sum_{\tau=1}^t (\eta_{\mathbf{Q}}^{(\tau)} + \lambda)}. \quad (23)$$

The update for \mathbf{E} proceeds in two steps. To handle the non-smoothness of $\|\mathbf{E}\|_1$, an auxiliary variable \mathbf{Z} is introduced

TABLE II. ONLINE STOCHASTIC GRADIENT ALGORITHM.

1: Initialize $\mathbf{E}^{(0)} := \mathbf{0}_{N_x \times N_y}$, $\mathbf{P}^{(0)}$ and $\mathbf{Q}^{(0)}$ at random.
2: Set $\eta_{\mathbf{P}}^{(1)} \geq \frac{L_{\mathbf{P}}^2}{L_{\min}}$, $\eta_{\mathbf{Q}}^{(1)} \geq \frac{L_{\mathbf{Q}}^2}{L_{\min}}$, and $\eta_{\mathbf{E}}^{(1)} \geq \frac{L_{\mathbf{E}}^2}{L_{\min}}$.
3: For $t = 1, 2, \dots$
4: Update $\mathbf{P}^{(t)}$ via (22)
5: Update $\mathbf{Q}^{(t)}$ via (23)
6: Update $\mathbf{E}^{(t)}$ via (24) and (25)
7: Next t

and projected onto the desired constraint set \mathcal{E} where $\mathbf{E} \in \mathcal{E}$, leading to

$$\mathbf{Z}^{(t)} = \frac{1}{\sum_{\tau=1}^t \eta_{\mathbf{E}}^{(\tau)}} \left[\eta_{\mathbf{E}}^{(t)} \mathbf{E}^{(t-1)} - \nabla_{\mathbf{E}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) + \sum_{\tau=1}^{t-1} \eta_{\mathbf{E}}^{(\tau)} \mathbf{Z}^{(t-1)} \right] \quad (24)$$

$$\mathbf{E}^{(t)} = \text{soft_th}_{\frac{\mu t}{\sum_{\tau=1}^t \eta_{\mathbf{E}}^{(\tau)}}}(\mathbf{Z}^{(t)}). \quad (25)$$

Convergence of the proposed updates is established next.

Proposition 2: Suppose that a1) $\{\boldsymbol{\xi}^{(t)}\}_{t=1}^{\infty}$ are independent and identically distributed (i.i.d) random processes and $\beta = 1$. If a2) $\{\mathbf{X}^{(t)}\}_{t=1}^{\infty}$ are in the compact set \mathcal{X} ; a3) Ξ is bounded; and a4) for the step sizes of $i \in \{\mathbf{P}, \mathbf{Q}, \mathbf{E}\}$, $\bar{\eta}_i^{(t)} := \sum_{\tau=1}^t \eta_i^{(\tau)} \geq ct$ are used $\forall t$ for some $c \geq 0$; with a5) $c' \geq \eta_i^{(t)} \geq L_i^2/L_{\min}$, $\forall t$ for some $c' \geq 0$, the iterates $\{\mathbf{X}^{(t)}\}_{t=1}^{\infty}$ generated by the proposed algorithm asymptotically converge to the stationary points of (P2), that is

$$\lim_{t \rightarrow \infty} \left[\inf_{\bar{\mathbf{X}} \in \bar{\mathcal{X}}} \|\mathbf{X}^{(t)} - \bar{\mathbf{X}}\|_F \right] = 0 \quad \text{a.s.} \quad (26)$$

where $\bar{\mathcal{X}}$ is the set of stationary points of (P2).

The proof is inspired by [15] and [16], but it is omitted here due to space limitation.

V. NUMERICAL TESTS

To assess the proposed approach, numerical tests were performed using synthetic data. Random tomographic measurements employing 30 nodes placed along the perimeter of a square region were taken, from which a discretized SLF of size $N_x = N_y = 40$ was reconstructed. Per-time slot, 15 measurements were taken, corrupted by zero-mean white Gaussian noise with unit variance. The parameters were set as $\rho = 15$, $\beta = 1$, $\lambda = 5$, and $\mu = 0.35$.

To test the performance of the batch algorithm in Table I, synthetic data were generated for a time interval with duration 500 slots. For comparison, a benchmark algorithm derived for (P1) using the APG method was also tested [12]. As seen in Fig. 1, both algorithms attain similar estimation cost after a few tens of iterations, as asserted by Proposition 1. The reconstructed SLFs from both algorithms in Fig. 2 are visually similar to the ground truth SLF. However, the APG algorithm requires a costly SVD operation of a N_x -by- N_y matrix per iteration incurring complexity $\mathcal{O}(4N_x^2 N_y + 22N_y^3)$, while the proposed algorithm requires only inversion of ρ -by- ρ matrices with complexity $\mathcal{O}(\rho^3)$. Thus, the proposed algorithm

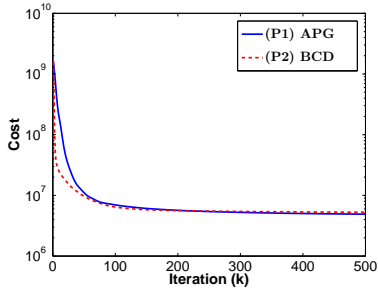


Fig. 1. Evolution of the optimization cost for batch algorithms.

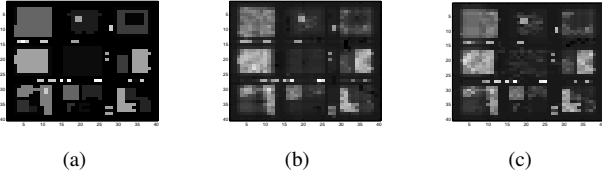


Fig. 2. True and reconstructed SLFs. (a) True SLF; (b) APG; and (c) BCD.

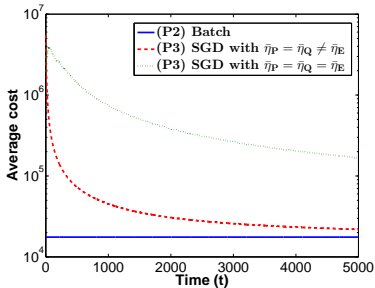


Fig. 3. Evolution of the average cost for the online algorithm.

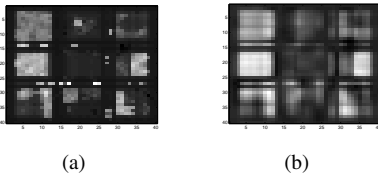


Fig. 4. Reconstructed SLFs using the online algorithm. (a) Different step sizes $\bar{\eta}_P = \bar{\eta}_Q = 4.2 \times 10^3$ and $\bar{\eta}_E = 10$; and (b) identical step sizes $\bar{\eta}_P = \bar{\eta}_Q = \bar{\eta}_E = 6000$.

achieves the performance of the benchmark algorithm with less computational complexity.

The proposed online algorithm was tested using step sizes $\bar{\eta}_P = \bar{\eta}_Q = 4.2 \times 10^3$ and $\bar{\eta}_E = 10$. The evolution of the average cost of the online algorithm is depicted in Fig. 3. Clearly, the average cost of the online estimator with different step sizes converges to the average cost solving (P2), which confirms convergence of the iterates from (P3) as asserted in Proposition 2. Next, the online algorithm with identical step sizes $\bar{\eta}_P = \bar{\eta}_Q = \bar{\eta}_E = 6000$ was tested. It resulted in slower interpolation of the SLF when compared with that of using multiple step sizes. In Fig. 3, the average cost of the online estimator with identical step sizes converges in a noticeably slow rate. The reconstructed SLFs are also shown in Fig. 4. The online algorithm with the different step sizes yields significantly better result, while the sparse components are not effectively captured with identical step sizes.

VI. CONCLUSIONS

A low-rank plus sparse matrix model was proposed for interpolation of channel gains for channel gain cartography. The low-rank component was motivated by the regular patterns of the obstacles to RF propagation in urban or indoor settings, while the sparse component could capture the artifacts that do not conform to the low-rank structure. Efficient algorithms were derived for reconstruction of SLFs using batch BCD and online SGD-type updates. The batch algorithm enjoyed low computational complexity and memory requirement. The online algorithm was shown to converge to the batch solution as claimed. Future research directions include testing with real data, and proving efficacy of the proposed techniques with severely limited number of measurements.

REFERENCES

- [1] 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 *E-Reference Signal Processing*. EURASIP, 2013, in press. [Online]. Available: <http://www.dtc.umn.edu/s/resources/eurasip2012tut.pdf>
- [2] J. A. Bazerque, G. Mateos, and G. B. Giannakis, "Group-Lasso on splines for spectrum cartography," *IEEE Trans. Sig. Proc.*, vol. 59, no. 10, pp. 4648–4663, Oct. 2011.
- [3] S.-J. Kim, E. Dall'Anese, and G. B. Giannakis, "Cooperative spectrum sensing for cognitive radios using Krighed Kalman filtering," *IEEE J. Sel. Topics Sig. Proc.*, vol. 5, no. 1, pp. 24–36, Feb. 2011.
- [4] P. Agrawal and N. Patwari, "Correlated link shadow fading in multi-hop wireless networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4024–4036, Aug. 2009.
- [5] 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.
- [6] 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.
- [7] B. R. Hamilton, X. Ma, R. J. Baxley, and S. M. Matechik, "Radio frequency tomography in mobile networks," in *Proc. of the IEEE Stat. Sig. Proc. Workshop*, Ann Arbor, MI, Aug. 2012, pp. 508–511.
- [8] J. Wright, A. Ganesh, K. Min, and Y. Ma, "Compressive principal component pursuit," *Information and Inference*, vol. 2, no. 1, pp. 32–68, Jun. 2013.
- [9] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *Elec. Lett.*, vol. 27, no. 23, pp. 2145–2146, Nov. 1991.
- [10] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. of the ACM*, vol. 58, no. 3, article no. 11, May 2011.
- [11] M. Mardani, G. Mateos, and G. B. Giannakis, "Dynamic anomalography: Tracking network anomalies via sparsity and low rank," *IEEE J. Sel. Topics Sig. Proc.*, vol. 7, no. 1, pp. 50–66, Feb. 2013.
- [12] 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," *UIUC Technical Report UILU-ENG-09-2214*, pp. 1–18, 2009.
- [13] P. Tseng, "Convergence of block coordinate descent method for non-differentiable minimization," *J. Optimiz. Theory Applicat.*, vol. 109, pp. 475–494, Jun. 2001.
- [14] K. Slavakis, S.-J. Kim, G. Mateos, and G. B. Giannakis, "Stochastic approximation vis-a-vis online learning for big data analytics," *IEEE Sig. Proc. Mag.*, vol. 31, pp. 124–129, Nov. 2014.
- [15] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Machine Learning Res.*, vol. 11, pp. 19–60, Jan. 2010.
- [16] M. Razaviyayn, M. Sanjabi, and Z. Luo, "A stochastic successive minimization method for nonsmooth nonconvex optimization." [Online]. Available: <http://arXiv:1307.4457v2>