



Contents lists available at SciVerse ScienceDirect

Physical Communication

journal homepage: www.elsevier.com/locate/phycom

Full length article

Group sparse Lasso for cognitive network sensing robust to model uncertainties and outliers[☆]

Emiliano Dall'Anese, Juan Andrés Bazerque, Georgios B. Giannakis*

Department of Electrical and Computer Engineering, University of Minnesota, 200 Union Street SE, Minneapolis, MN 55455, United States

ARTICLE INFO

Article history:

Received 21 February 2011
 Received in revised form 29 April 2011
 Accepted 30 July 2011
 Available online xxxx

Keywords:

Spectrum sensing
 Spectrum cartography
 Sparse linear regression
 Total least-squares
 Outliers

ABSTRACT

To account for variations in the frequency, time, and space dimensions, dynamic re-use of licensed bands under the cognitive radio (CR) paradigm calls for innovative network-level sensing algorithms for multi-dimensional spectrum opportunity awareness. Toward this direction, the present paper develops a collaborative scheme whereby CRs cooperate to localize active primary user (PU) transmitters and reconstruct a power spectral density (PSD) map portraying the spatial distribution of power across the monitored area per frequency band and channel coherence interval. The sensing scheme is based on a parsimonious model that accounts for two forms of sparsity: one due to the narrow-band nature of transmit-PSDs compared to the large portion of spectrum that a CR can sense, and another one emerging when adopting a spatial grid of candidate PU locations. Capitalizing on this dual sparsity, an estimator of the model coefficients is obtained based on the group sparse least-absolute-shrinkage-and-selection operator (GS-Lasso). A novel reduced-complexity GS-Lasso solver is developed by resorting to the alternating direction method of multipliers (ADMoM). Robust versions of this GS-Lasso estimator are also introduced using a GS total least-squares (TLS) approach to cope with both uncertainty in the regression matrices, arising due to inaccurate channel estimation and grid-mismatch effects, and unexpected model outliers. In spite of the non-convexity of the GS-TLS criterion, the novel robust algorithm has guaranteed convergence to (at least) a local optimum. The analytical findings are corroborated by numerical tests.

Published by Elsevier B.V.

1. Introduction

To alleviate the inefficiency of the current rigid license-based spectrum assignment and make a swath of frequencies available to emerging wireless services, research efforts have focused on dynamic spectrum (re-)utilization techniques [1]. Prominent in this context is the hierarchical spectrum access model, where cognitive radios (CRs) are envisioned as autonomous devices able

to spatio-temporally re-use the licensed bands in a non-intrusive manner [2].

In lieu of coordination among primary users (PUs) and CRs, autonomous spectrum sensing is of paramount importance for the detection of ongoing PU transmissions, and thus identification of the so-called “spectrum holes”. At the expense of increasing communication overhead among CRs, cooperative sensing schemes exhibit improved performance relative to non-cooperative alternatives [3]. Conceivably, through fusion of local measurements, cooperative sensing can collect the available spatial diversity provided by multipath propagation of the underlying PU-to-CR channels. Representative past works on cooperative spectrum sensing include [4], where a bank of energy detectors is used to monitor a large portion of the spectrum, [5], where a test statistic is introduced to maximize the probability of detecting available primary bands,

[☆] This work was supported by NSF grants CCF-0830480, CCF-1016605, ECCS-0824007 and ECCS-1002180, and QNRF grant NPRP 09-341-2-128.

* Corresponding author. Tel.: +1 612 626 7781; fax: +1 612 625 4583.

E-mail addresses: emiliano@umn.edu (E. Dall'Anese), bazer002@umn.edu (J.A. Bazerque), georgios@umn.edu, georgios@ece.umn.edu (G.B. Giannakis).

and [6], where individual sensing decisions are combined using a linear-quadratic fusion rule; see also [7] for sequential alternatives.

Even if a primary band is occupied, there could be locations where the transmitted power is low enough so that these frequencies can be reused by CRs without suffering from or causing harmful interference to any PU. Thus, to enable opportunistic re-use of the licensed resources under the primary–secondary hierarchy [1], the sensing objective calls for cognition-enabling network-level algorithms that make CRs aware of PU activities across frequency, space, and time.

Initial efforts to this end have been devoted to construct power spectral density (PSD) maps (one per coherence interval), which essentially portray the PU power present at each location of the monitored area. To reconstruct the resultant PSD atlas starting from raw power measurements, a spatial interpolation technique was employed by Alaya-Feki et al. [8], and a smooth PSD map was constructed in [9] using the method of splines in order to account for shadowing. An atlas of channel gains was constructed in [10] to provide link amplitudes between any pair of points in a given geographical area; such channel gain atlas can be also used to reconstruct the PSD map provided that PU locations and transmission powers are available at the CRs.

To further address the challenges encountered with this multi-dimensional sensing vision, the present paper presents a collaborative sensing scheme whereby CRs cooperate to localize the actively transmitting PUs and estimate their PSD across space in the presence of model uncertainties. This network-level sensing algorithm can be complemented by the channel gain atlas, so that the CR system can effectively estimate the PSD distribution in space and, thus, reveal areas where the CRs could re-use the primary bands in a non-destructive manner.

The novel sensing scheme here is based on a parsimonious system model accounting for the scarce presence of active PUs in the same frequency band(s), in the monitored area, due to mutual interference. Using a virtual grid-based approach for the potential PU transmitter locations, a form of *spatial-domain sparsity* emerges because actual PU transmitters are present in only few of the potential (grid) locations. A basis expansion model is then adopted to approximate the PU transmit-PSD distribution in frequency, which renders the sensing objective tantamount to estimating the PSD basis coefficients corresponding to each grid point. Since individual PU transmissions are narrow-band relative to the large swath of frequencies a CR can sense, only few of the PSD basis coefficients are nonzero – a fact giving rise to *frequency-domain sparsity*.

This parsimonious system model thus entails a form of *hierarchical dual-domain sparsity* [11] in the PSD basis coefficients that are to be estimated, in the sense that groups of coefficients corresponding to locations with no PUs will be collectively zero. In addition, some of the basis coefficients within groups corresponding to active PU locations will be zero. Capitalizing on this form of hierarchical sparsity, a group sparse regression problem is formulated, which is to be solved centrally by either a fusion center (FC), or, individual CRs using their measured

PSDs. A novel low-complexity algorithm for solving such a problem is developed using the alternating direction method of multipliers (ADMoM) [12].

A critical issue for the proposed network-level sensing problem is acquiring the grid-to-CR channel gains present in the underlying regression matrix. One way to acquire such information is through the channel gain cartography approach of [10]. However, possible inaccurate channel gains or adoption of a shadowing-agnostic path loss-only model [13,14] could deteriorate the performance of the sensing algorithm [15]. Also, a grid-based approach introduces itself possible model offsets, as the actual PU locations may not coincide with points of the grid. To account for these uncertainties, a robust version of the group sparse (GS) least-absolute-shrinkage-and-selection operator (Lasso) is developed. The main contribution in this direction consists in an extension of the sparse total least-squares (TLS) framework of [16] to incorporate the hierarchical sparsity inherent to this sensing application. Combining the merits of Lasso, group Lasso, and TLS, the proposed group sparse (GS-)TLS approach yields hierarchically-sparse PSD estimates that are also robust to model uncertainties induced by the random channel, grid offsets, and basis approximation errors. In spite of the non-convexity of the proposed GS-TLS criterion, an iterative solver with guaranteed convergence to at least a locally-optimal solution is developed.

Additional factors compromising accuracy of PSD estimates at the CRs, are abrupt changes in shadow fading that may be due to, e.g., moving obstacles or moving CRs, and, possible failures of the sensing modules themselves. A crude remedy for such effects is simple averaging of all the PSD estimates at the FC. Instead, a robust GS-TLS formulation is proposed here, that is capable of discerning and removing such so-called model outliers [17], which in turn leads to reliable PSD estimates. However, sorting out unreliable measurements not only promotes estimation accuracy, but also leads to self-healing and re-organization mechanisms for the CRs network.

The rest of the paper is organized as follows. Section 2 introduces the basis expansion model, and describes the PSD observations used for the model fitting approach. A centralized algorithm for solving GS-Lasso problems is developed in Section 3, whereas perturbations in the channel (regression) matrices are considered in Section 4. The outlier-resilient sensing algorithm is devised in Section 5, numerical results are provided in Section 6, and Section 7 draws the conclusions.¹

2. System model and problem formulation

Consider an incumbent PU system comprising N_s transmitters (sources) located in a geographical area $\mathcal{A} \subset \mathbb{R}^2$. Their activity over a frequency band B is to be

¹ Notation: upper (lower) bold face letters are used to denote matrices (column vectors); $\mathbf{1}_n$ and $\mathbf{0}_n$ denote the $n \times 1$ vectors with all ones and all zeros, respectively; $(\cdot)^T$ denotes transposition; $\|\mathbf{X}\|_F$ the Frobenius norm of matrix \mathbf{X} , and $\|\mathbf{x}\|_p := (\sum_i x_i^p)^{1/p}$, $p = 1, 2$, the ℓ_1 - and ℓ_2 -norms, respectively; finally, $\text{sgn}(\cdot)$ denotes the sign function and \mathbf{I}_N the $N \times N$ identity matrix.

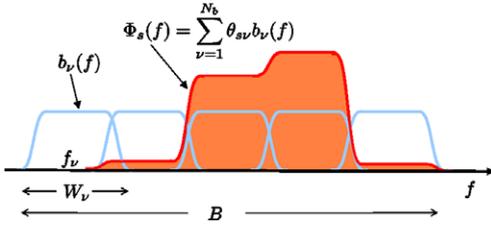


Fig. 1. Basis expansion model with overlapping raised cosine pulses.

monitored via cooperation of N_r CRs, also located in \mathcal{A} . Let $\mathcal{S} := \{\mathbf{x}_s \in \mathcal{A}\}_{s=1}^{N_s}$ denote the PU locations. The sensing objective is to localize the PU sources, and reveal available portions of B for the CRs to transmit opportunistically.

Let $u_s(t)$ be the (unknown) signal transmitted by PUs at time t . Then, the signal received at CR position \mathbf{x}_r at time t can be expressed as $y_r(t) := \sum_{s=1}^{N_s} \sum_{l=0}^{L_{sr}-1} h_{\mathbf{x}_s \rightarrow \mathbf{x}_r}(t; l) u_s(t-l) + v(t)$, where $h_{\mathbf{x}_s \rightarrow \mathbf{x}_r}(t; l)$ is the l -th tap of the time-varying channel impulse response of the link $\mathbf{x}_s \rightarrow \mathbf{x}_r$, and $v(t)$ denotes the additive white noise. Regarding signals $\{u_s(t)\}$, the following is assumed.

(As1) Sources $\{u_s(t)\}$ are stationary, mutually uncorrelated, independent of $\{h_{\mathbf{x}_s \rightarrow \mathbf{x}_r}(t; l)\}$, with vanishing correlation per channel coherence interval.

Consider approximating the transmit-PSD of PUs using the following basis expansion model [13]

$$\Phi_s(f) = \sum_{\nu=1}^{N_b} \theta_{s\nu} b_\nu(f), \quad s = 1, 2, \dots, N_s \quad (1)$$

where N_b is assumed sufficiently large, and $\{\theta_{s\nu}\}_{\nu=1}^{N_b}$ are nonnegative coefficients. Possible choices of $\{b_\nu(f)\}_{\nu=1}^{N_b}$ include the set of non-overlapping rectangles of unit height spanning the bandwidth B of interest. In this case, each $\theta_{s\nu}$ represents the power emitted by source s on the frequency band corresponding to the basis function $b_\nu(f)$. Alternatively, overlapping raised cosine bases can be employed with support $B_\nu = [f_\nu - (1 + \rho/2T_s), f_\nu + (1 + \rho/2T_s)]$, where ρ is the roll-off factor and T_s the symbol period, can be employed; see also Fig. 1.

Channel $\{h_{\mathbf{x}_s \rightarrow \mathbf{x}_r}(t; l)\}$ can be decomposed as [18] $h_{\mathbf{x}_s \rightarrow \mathbf{x}_r}(t; l) = (\gamma_{\mathbf{x}_s \rightarrow \mathbf{x}_r} s_{\mathbf{x}_s \rightarrow \mathbf{x}_r})^{\frac{1}{2}} f_{\mathbf{x}_s \rightarrow \mathbf{x}_r}(t; l)$, where $\gamma_{\mathbf{x}_s \rightarrow \mathbf{x}_r}$ stands for the path loss, $s_{\mathbf{x}_s \rightarrow \mathbf{x}_r}$ the temporally- and spatially-colored shadowing [19], and $\{f_{\mathbf{x}_s \rightarrow \mathbf{x}_r}(t; l)\}$ for the multi-path fast time-varying fading. The latter satisfies the following.

(As2) Variables $\{f_{\mathbf{x}_s \rightarrow \mathbf{x}_r}(t; l)\}$ are complex Gaussian with zero mean and variance $\sigma_{f, sr}^2$, stationary with respect to t , and uncorrelated across the lag variable l and the spatial variables \mathbf{x}_s and \mathbf{x}_r . Without loss of generality, assume that $\sum_{l=1}^{L_{sr}} \sigma_{f, sr}^2 = 1$ for every s and r .

Received samples $\{y_r(t)\}$ are parsed into N -dimensional blocks, where N is chosen equal to (or smaller than) the coherence interval of the small-scale fading, over which $h_{\mathbf{x}_s \rightarrow \mathbf{x}_r}(t; l)$ remains approximately invariant with respect to (wrt) t . These data blocks are hereafter indexed by n , so that $t = nN + m$, with $m = 0, 1, \dots, N - 1$.

Shadowing and small-scale fading are characterized by different dynamics. The following is assumed regarding

channel propagation and modeling; (see also, e.g., [19] and references therein).

(As3) The coherence interval of shadow fading exceeds that of $f_{\mathbf{x}_s \rightarrow \mathbf{x}_r}(t; l)$. Also, shadowing variations are sufficiently slower than the coherence interval of the PU signals.

Based on (As1)–(As3), it is possible to express the PSD measured at location \mathbf{x} due to N_s simultaneous PU transmissions as

$$\begin{aligned} \Phi_{\mathbf{x}}(f) &= \sum_{s=1}^{N_s} g_{\mathbf{x}_s \rightarrow \mathbf{x}} \Phi_s(f) + \sigma_v^2 \\ &= \sum_{s=1}^{N_s} g_{\mathbf{x}_s \rightarrow \mathbf{x}} \sum_{\nu=1}^{N_b} \theta_{s\nu} b_\nu(f) + \sigma_v^2 \end{aligned} \quad (2)$$

where σ_v^2 denotes noise variance at the receiver, and $g_{\mathbf{x}_s \rightarrow \mathbf{x}} := \gamma_{\mathbf{x}_s \rightarrow \mathbf{x}_r} s_{\mathbf{x}_s \rightarrow \mathbf{x}_r}$ the averaged channel gain.²

As neither the number of PU sources nor their locations are known to the CRs, a set of N_g candidate transmit-PUs is postulated on a grid of locations $\mathcal{G} := \{\mathbf{x}_g \in \mathcal{A}\}_{g=1}^{N_g}$. Without prior knowledge of the area(s) where PU activity is more likely, the set of candidate locations \mathcal{G} can be simply formed by discretizing \mathcal{A} to the set of grid points \mathcal{G} .

Define the $N_b \times 1$ vector $\theta_g := [\theta_{g1}, \dots, \theta_{gN_b}]^T$ collecting the basis coefficients that correspond to location \mathbf{x}_g , and let $\theta := [\theta_1^T, \dots, \theta_{N_g}^T]^T$; also, let $\mathbf{B}_{\mathbf{x}_r}$ be the $N \times N_b N_g$ matrix $\mathbf{B}_{\mathbf{x}_r} := [\mathbf{b}_{\mathbf{x}_r}(f_1), \dots, \mathbf{b}_{\mathbf{x}_r}(f_N)]^T$, with $\mathbf{b}_{\mathbf{x}_r}(f_k)$ having entries $\{g_{\mathbf{x}_g \rightarrow \mathbf{x}_r} \cdot b_\nu(f_k)\}$. Then, upon defining $\varphi_{\mathbf{x}_r} := [\Phi_{\mathbf{x}_r}(f_1), \dots, \Phi_{\mathbf{x}_r}(f_N)]^T$, the received PSD at CR location \mathbf{x}_r , sampled at frequencies $\{f_k\}_{k=1}^N$, can be compactly written as [cf. (2)]

$$\varphi_{\mathbf{x}_r} = \mathbf{B}_{\mathbf{x}_r} \theta + \sigma_v^2 \mathbf{1}_N. \quad (3)$$

The sensing objective of revealing PU locations and the available portions (sub-bands) of B is tantamount to estimating θ . To this end, CRs rely on the periodogram estimate of $\Phi_{\mathbf{x}_r}(f)$ at the sampling locations $\{\mathbf{x}_r\}_{r=1}^{N_r}$, and N frequency bins $\{f_k\}_{k=1}^N$. The fast Fourier transform of samples $\{y_r(t)\}$, namely $Y_{r,N}(n, f)$, and the periodogram $\hat{\varphi}_{\mathbf{x}_r, N}(n; f) := (1/N) |Y_{r,N}(n, f)|^2$ are computed per data block n . To average out small-scale fading effects, and allow for tracking of shadow fading as well as possible variations of the PUs' power spectra, the periodogram estimate at CR r is formed using an exponentially weighted moving average operation as

$$\hat{\Phi}_{\mathbf{x}_r}(\tau, f) := \sum_{n=1}^{\tau} \alpha^{\tau-n} \hat{\varphi}_{\mathbf{x}_r, N}(n, f) \quad (4)$$

with $\alpha \in (0, 1]$ denoting the so-called forgetting factor. As shown in [13], (4) gives an estimate of the PSD measured at point \mathbf{x}_r and frequency f as

$$\hat{\Phi}_{\mathbf{x}_r}(\tau; f) = \Phi_{\mathbf{x}_r}(\tau; f) + e_{\mathbf{x}_r, N}(\tau; f) \quad (5)$$

² Shadow fading is assumed to be frequency-invariant over the monitored PU band B ; however, frequency-selective shadowing can be readily incorporated.

with (asymptotic) variance bounded as $\lim_{N \rightarrow \infty} \text{var}[e_{r,N}(\tau; f)] \leq \frac{3}{2}(1 - \alpha) \Phi_{\mathbf{x}_r}^2(f)$. After dropping τ for notational brevity, let $\hat{\Phi}_{\mathbf{x}_r} := [\hat{\Phi}_{\mathbf{x}_r}(f_1), \dots, \hat{\Phi}_{\mathbf{x}_r}(f_N)]^T$.

Based on the linear model (3), the sensing objective is to estimate θ from the received-PSD estimate $\hat{\Phi} := [\hat{\Phi}_{\mathbf{x}_1}, \dots, \hat{\Phi}_{\mathbf{x}_{N_r}}]^T$ gathered at CR locations $\{\mathbf{x}_r\}_{r=1}^{N_r}$.

3. Spectrum sensing via group sparse Lasso

The number of active PUs transmitting over the same spectral band in a given area is naturally limited by mutual interference. As a consequence, the number of PU sources (N_s) is far smaller than N_g , for a sufficiently dense grid. Absence of PU sources in most grid locations $\mathbf{x}_g \in \mathcal{G} \setminus \mathcal{S}$ gives rise to a *group sparsity* of the vector θ , since $\theta_g = \mathbf{0}_{N_b}$ for each of the locations \mathbf{x}_g that are not occupied by a PU transmitter. In addition to space, sparsity in the vector θ is also manifested in the frequency domain because of the parsimonious linear model (2). Compared to the possibly large swath of frequencies that the CRs can sense, individual PU transmissions typically occupy small portions of the spectrum (say, in the order of MHz). Sparsity in the frequency domain implies that individual entries within each group θ_g are zero.

The most popular criterion for estimating θ is the (non-negative) least-squares (LS) [20]. However, LS fails to provide a parsimonious model estimate involving only the prominent variables. The Lasso [21] and the so-called group Lasso [22] on the other hand, were proposed to overcome such a limitation of LS. In the Lasso criterion, the LS cost is augmented with the ℓ_1 -norm $\|\theta\|_1$ to encourage sparsity at the single-coefficient level; while in the group Lasso, the regularization term $\mathcal{R}_G(\theta) := \sum_{g=1}^{N_g} \|\theta_g\|_2$ enforces group sparsity.

Combining Lasso [21] with group Lasso [22], the so-called group sparse (GS-) Lasso [11,23] provides a parsimonious model estimate, where sparsity is accounted for both at the group- and at the single-coefficient levels. This hierarchical sparsity is possible by regularizing the conventional LS cost with the term $\mathcal{R}_G(\theta) := \sum_{g=1}^{N_g} \|\theta_g\|_2$ combined with $\|\theta\|_1$.

Taking also into account the non-negativity of PU power spectra, θ can be estimated by solving the following sparse regression problem:

$$\hat{\theta} = \arg \min_{\theta \geq \mathbf{0}_{N_b N_g}} \left[\frac{1}{2} \sum_{r=1}^{N_r} \|\hat{\Phi}_{\mathbf{x}_r} - \mathbf{B}_{\mathbf{x}_r} \theta - \sigma_v^2 \mathbf{1}_N\|_2^2 + \lambda_1 \|\theta\|_1 + \lambda_2 \mathcal{R}_G(\theta) \right] \quad (6)$$

where the coefficient $\lambda_1 \geq 0$ enforces sparsity at individual entries, whereas $\lambda_2 \geq 0$ promotes group sparsity. For $\lambda_1 = 0$ ($\lambda_2 = 0$), (6) reduces to the Lasso (group Lasso) based estimate. PU localization and PSD estimation was viewed as a sparse linear regression model in [13]; here, the formulation of [13] is considerably broadened by taking into account both individual and group sparsity.

To obtain regression matrices $\{\mathbf{B}_{\mathbf{x}_r}\}$, the channel gains $\{g_{\mathbf{x}_g \rightarrow \mathbf{x}_r}\}$ need to be estimated. To this end, CRs can simply neglect shadowing, and as in [13,14] resort to the distance-dependent path loss model $\hat{g}_{\mathbf{x}_g \rightarrow \mathbf{x}_r} = \min\{1, (\|\mathbf{x}_g - \mathbf{x}_r\|_2 / d_0)^{-\eta}\}$, where d_0 and η are preselected constants depending on the propagation environment. Alternatively, more sophisticated techniques can be employed [10,24]. Perturbations in the regression matrices $\{\mathbf{B}_{\mathbf{x}_r}\}$ arising due to inaccurate channel estimation and grid-mismatch effects will be dealt with in Section 4.

3.1. PSD atlas

It is worth re-iterating that identifying the support of the vector θ reveals not only the primary sub-bands occupied, but also the locations where the active PU transmitters reside. Complementing this information with either the PUs' channel gain maps [10] or a simple path loss-based propagation model, CRs can readily reconstruct the PSD atlas; that is, estimate PSD maps at *any* location of the monitored area as

$$\hat{\Phi}_{\mathbf{x}}(f) = \sum_{g=1}^{N_g} \hat{g}_{\mathbf{x}_g \rightarrow \mathbf{x}} \sum_{v=1}^{N_b} \hat{\theta}_{g,v} b_v(f), \quad \forall \mathbf{x} \in \mathcal{A} \quad (7)$$

with $\hat{g}_{\mathbf{x}_g \rightarrow \mathbf{x}}$ the estimate of $g_{\mathbf{x}_g \rightarrow \mathbf{x}}$ [13,10,14]. Having available estimates of the PSD map across space per frequency band (hence the term atlas), CRs can adjust their transmit power to prevent harmful interference inflicted to the PUs. In fact, the positions of potential PU receivers can be deduced from the PSD atlas [10]; and thus, CR transmission powers can be properly adapted [25].

3.2. ADMoM-based solver

In this section, a reduced-complexity algorithm attaining the optimal solution of GS-Lasso problems will be developed using the alternating direction method of multipliers (ADMoM). The crux is to show that (6) admits an equivalent reformulation that can be solved via ADMoM. Before doing so, the following lemmas are needed.

Lemma 1. Consider the following convex minimization problem in the variable $\mathbf{y} \in \mathbb{R}^N$

$$\mathbf{y}^* = \arg \min_{\mathbf{y}} \left[\frac{c}{2} \mathbf{y}^T \mathbf{y} - \mathbf{y}^T \mathbf{a} + \lambda \|\mathbf{y}\|_2 \right]. \quad (8)$$

Albeit non differentiable, (8) admits a closed-form solution. Specifically, the global minimizer \mathbf{y}^* is given by the following soft-thresholding vector operation expressed in terms of $[a]_+ := \max\{0, a\}$ as

$$\mathbf{y}^* = \frac{\mathbf{a}}{c \|\mathbf{a}\|_2} [\|\mathbf{a}\|_2 - \lambda]_+. \quad (9)$$

Proof. It will be argued that the solver of (8) takes the form $\mathbf{y} = z \mathbf{a}$ for some scalar $z \geq 0$. This is because among all \mathbf{y} with the same ℓ_2 -norm, the Cauchy-Schwarz inequality implies that the maximizer of $\mathbf{a}^T \mathbf{y}$ is collinear with (and in the same direction of) \mathbf{a} . Substituting $\mathbf{y} = z \mathbf{a}$ into (8) renders the problem scalar in $z \geq 0$, with solution

$z^* = (\|\mathbf{a}\| - \lambda)_+ / (c\|\mathbf{a}\|_2)$, which completes the proof. For completeness, note that the same result can be alternatively obtained by resorting to the subdifferential [26] of the cost in (8). \square

Lemma 2. Consider the following non-smooth convex problem in the vector variable $\mathbf{y} \in \mathbb{R}^N$

$$\mathbf{y}^* = \arg \min_{\mathbf{y}} \left[\frac{c}{2} \mathbf{y}^T \mathbf{y} - \mathbf{a}^T \mathbf{y} + \lambda \|\mathbf{y}\|_1 \right]. \quad (10)$$

Using operator $\mathcal{F}_\lambda(\cdot)$ defined as

$$\mathcal{F}_\lambda(\mathbf{a}) := [\text{sgn}(a_1)[|a_1| - \lambda]_+, \dots, \text{sgn}(a_N)[|a_N| - \lambda]_+]^T$$

the global minimized of (10) can be written as

$$\mathbf{y}^* = \frac{1}{c} \mathcal{F}_\lambda(\mathbf{a}). \quad (11)$$

If non-negativity of the entries in \mathbf{y} is imposed, the solution of (10) subject to $\mathbf{y} \geq \mathbf{0}_N$ is obtained by using the vector soft-thresholding operator

$$\mathcal{F}_\lambda^+(\mathbf{a}) := [\max\{0, a_1 - \lambda\}, \dots, \max\{0, a_N - \lambda\}]^T$$

as

$$\mathbf{y}^* = \frac{1}{c} \mathcal{F}_\lambda^+(\mathbf{a}). \quad (12)$$

Proof. Note first that (12) can be solved element-wise; specifically, each entry y_i of \mathbf{y} is found by solving the scalar problem $y_i^* = \arg \min_y C(y) := (\frac{c}{2}y^2 - a_i y + \lambda|y|)$, which has a non-differentiable cost. The necessary and sufficient condition for y_i^* to minimize $C(y)$ is [27, p. 92]

$$\begin{cases} |a_i| \leq \lambda, & \text{if } y_i^* = 0 \\ cy_i^* - a + \lambda \frac{y_i^*}{|y_i^*|} = 0, & \text{if } y_i^* \neq 0 \end{cases} \quad (13)$$

which is satisfied by $y_i^* = \text{sgn}(a_i)[|a_i| - \lambda]_+$; see also [21]. When \mathbf{y} is enforced to be non-negative, solution (12) can be easily derived from (13) element-wise. \square

Consider now the $N_b N_g \times 1$ auxiliary vector variables $\boldsymbol{\gamma}$ and $\boldsymbol{\xi}$, and neglect irrelevant terms to re-write the GS-Lasso problem as

$$\{\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\xi}}\} = \arg \min_{\boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\xi}} \left[\frac{1}{2} \boldsymbol{\theta}^T \mathbf{R} \boldsymbol{\theta} - \boldsymbol{\theta}^T \mathbf{r} + \lambda_1 \|\boldsymbol{\xi}\|_1 + \lambda_2 \mathcal{R}_G(\boldsymbol{\gamma}) \right] \quad (14)$$

$$\begin{aligned} \text{s.t. } \boldsymbol{\xi} &\geq \mathbf{0}_{N_b N_g} \\ \boldsymbol{\theta} &= \boldsymbol{\gamma}, \quad \boldsymbol{\theta} = \boldsymbol{\xi} \end{aligned}$$

where $\mathbf{R} := \sum_{r=1}^{N_r} \mathbf{B}_{\mathbf{x}_r}^T \mathbf{B}_{\mathbf{x}_r}$, $\mathbf{r} := \sum_{r=1}^{N_r} \mathbf{B}_{\mathbf{x}_r}^T \bar{\boldsymbol{\varphi}}_{\mathbf{x}_r}$, and $\bar{\boldsymbol{\varphi}}_{\mathbf{x}_r} := \hat{\boldsymbol{\varphi}}_{\mathbf{x}_r} - \sigma_v^2 \mathbf{1}_N$. For simplicity, σ_v^2 is assumed to be known; however, it could be incorporated in (14) and estimated as the intercept.

Letting $\boldsymbol{\eta}$ and $\boldsymbol{\mu}$ denote the Lagrange multipliers associated with the equality constraints $\boldsymbol{\theta} = \boldsymbol{\gamma}$ and $\boldsymbol{\theta} =$

$\boldsymbol{\xi}$, respectively, the quadratically augmented Lagrangian function of problem (14) is

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\xi}, \boldsymbol{\eta}, \boldsymbol{\mu}) &= \frac{1}{2} \boldsymbol{\theta}^T \mathbf{R} \boldsymbol{\theta} - \boldsymbol{\theta}^T \mathbf{r} + \lambda_1 \|\boldsymbol{\xi}\|_1 \\ &\quad + \lambda_2 \mathcal{R}_G(\boldsymbol{\gamma}) + \boldsymbol{\eta}^T (\boldsymbol{\theta} - \boldsymbol{\gamma}) \\ &\quad + \boldsymbol{\mu}^T (\boldsymbol{\theta} - \boldsymbol{\xi}) + \frac{c_1}{2} \|\boldsymbol{\theta} - \boldsymbol{\gamma}\|_2^2 \\ &\quad + \frac{c_2}{2} \|\boldsymbol{\theta} - \boldsymbol{\xi}\|_2^2 \end{aligned} \quad (15)$$

where $c_1, c_2 > 0$ are arbitrary constants. Then, for any initial vectors $\boldsymbol{\gamma}^{(0)}, \boldsymbol{\xi}^{(0)}, \boldsymbol{\eta}^{(0)}, \boldsymbol{\mu}^{(0)}$, the ADMoM algorithm entails the following primal-dual iterative updates

$$\boldsymbol{\theta}^{(j)} = \arg \min_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\gamma}^{(j-1)}, \boldsymbol{\xi}^{(j-1)}, \boldsymbol{\eta}^{(j-1)}, \boldsymbol{\mu}^{(j-1)}) \quad (16a)$$

$$\{\boldsymbol{\gamma}^{(j)}, \boldsymbol{\xi}^{(j)}\} = \arg \min_{\boldsymbol{\gamma}, \boldsymbol{\xi} \geq \mathbf{0}} \mathcal{L}(\boldsymbol{\theta}^{(j)}, \boldsymbol{\gamma}, \boldsymbol{\xi}, \boldsymbol{\eta}^{(j-1)}, \boldsymbol{\mu}^{(j-1)}) \quad (16b)$$

$$\boldsymbol{\eta}^{(j)} = \boldsymbol{\eta}^{(j-1)} + c_1 (\boldsymbol{\theta}^{(j)} - \boldsymbol{\gamma}^{(j)}) \quad (16c)$$

$$\boldsymbol{\mu}^{(j)} = \boldsymbol{\mu}^{(j-1)} + c_2 (\boldsymbol{\theta}^{(j)} - \boldsymbol{\xi}^{(j)}) \quad (16d)$$

where $j = 1, 2, \dots$ is the iteration index. The first step updates the primal vector $\boldsymbol{\theta}^{(j)}$ by using the values of the auxiliary variables and the Lagrange multipliers obtained at the previous iteration $j - 1$; since $\mathcal{L}(\cdot)$ is quadratic in $\boldsymbol{\theta}$, the convex optimization problem (16a) can be solved in closed form as

$$\begin{aligned} \boldsymbol{\theta}^{(j)} &= (\mathbf{R} + (c_1 + c_2) \mathbf{I}_{N_b N_g})^{-1} \\ &\quad \times (\mathbf{r} + c_1 \boldsymbol{\gamma}^{(j-1)} + c_2 \boldsymbol{\xi}^{(j-1)} - \boldsymbol{\eta}^{(j-1)} - \boldsymbol{\mu}^{(j-1)}). \end{aligned} \quad (17)$$

Next, variables $\boldsymbol{\gamma}$ and $\boldsymbol{\xi}$ can be updated using the newly computed vector $\boldsymbol{\theta}^{(j)}$, with the Lagrange multipliers fixed from the previous iteration. Inspection of the function $\mathcal{L}(\boldsymbol{\theta}^{(j)}, \boldsymbol{\gamma}, \boldsymbol{\xi}, \boldsymbol{\eta}^{(j-1)}, \boldsymbol{\mu}^{(j-1)})$ reveals that (16b) can be split into two sub-problems, where minimization over $\boldsymbol{\gamma}$ and $\boldsymbol{\xi}$ can be performed separately. After neglecting irrelevant terms, minimization of (16b) wrt $\boldsymbol{\gamma}$ reduces to the following non-differentiable convex problem

$$\boldsymbol{\gamma}^{(j)} = \arg \min_{\boldsymbol{\gamma}} \left[\frac{c_1}{2} \boldsymbol{\gamma}^T \boldsymbol{\gamma} - \boldsymbol{\gamma}^T (c_1 \boldsymbol{\theta}^{(j)} + \boldsymbol{\eta}^{(j-1)}) + \lambda_2 \mathcal{R}_G(\boldsymbol{\gamma}) \right] \quad (18)$$

which, in turn, can be separated in the following N_g sub-problems

$$\begin{aligned} \boldsymbol{\gamma}_g^{(j)} &= \arg \min_{\boldsymbol{\gamma}_g} \left[\frac{c_1}{2} \boldsymbol{\gamma}_g^T \boldsymbol{\gamma}_g - \boldsymbol{\gamma}_g^T (c_1 \boldsymbol{\theta}_g^{(j)} \right. \\ &\quad \left. + \boldsymbol{\eta}_g^{(j-1)}) + \lambda_2 \|\boldsymbol{\gamma}_g\|_2 \right], \quad g = 1, \dots, N_g \end{aligned} \quad (19)$$

where $\boldsymbol{\gamma}_g$ and $\boldsymbol{\eta}_g$ are $N_b \times 1$ sub-vectors of $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$, respectively, collecting elements $\{\gamma_n\}_{n=N_g(g-1)+1}^{N_g g}$ and $\{\eta_n\}_{n=N_g(g-1)+1}^{N_g g}$.

From Lemma 1, the global minimizer of each sub-problem (19) is given by

$$\boldsymbol{\gamma}_g^{(j)} = (\boldsymbol{\theta}_g^{(j)} + c_1^{-1} \boldsymbol{\eta}_g^{(j-1)}) \frac{\left[\|\mathbf{c}_1 \boldsymbol{\theta}_g^{(j)} + \boldsymbol{\eta}_g^{(j-1)}\|_2 - \lambda_2 \right]_+}{\|\mathbf{c}_1 \boldsymbol{\theta}_g^{(j)} + \boldsymbol{\eta}_g^{(j-1)}\|_2}. \quad (20)$$

Upon neglecting constant terms, minimization of (16b) wrt the non-negative variable ξ can be obtained after solving the following non-smooth convex problem

$$\xi^{(j)} = \arg \min_{\xi \geq 0} \left[\frac{c_2}{2} \xi^T \xi - \xi^T (\mu^{(j-1)} + c_2 \theta^{(j)}) + \lambda_1 \|\xi\|_1 \right] \quad (21)$$

which, from Lemma 2, admits the following closed-form solution

$$\xi^{(j)} = \frac{1}{c_2} \mathcal{T}_{\lambda_1}^+ (c_2 \theta^{(j)} + \mu^{(j-1)}). \quad (22)$$

The overall ADMoM-based solver for GS-Lasso problems is tabulated as Algorithm 1.

Algorithm 1 ADMoM-based GS-Lasso solver

```

Initialize  $\gamma^{(0)} = \mathbf{0}_{N_b N_g}$ ,  $\xi^{(0)} = \mathbf{0}_{N_b N_g}$ ,  $\eta^{(0)} = \mathbf{0}_{N_b N_g}$ , and  $\mu^{(0)} = \mathbf{0}_{N_b N_g}$ 
Form  $\mathbf{R}$  and  $\mathbf{r}$ 
for  $j = 0, 1, \dots$  do
    Update  $\theta^{(j)}$  via (17)
    Update  $\gamma_g^{(j)}$  via (20) for all  $g = 1, \dots, N_g$ 
    Update  $\xi^{(j)}$  via (22)
    Update  $\eta^{(j)} = \eta^{(j-1)} + c_1 (\theta^{(j)} - \gamma^{(j)})$ 
    Update  $\mu^{(j)} = \mu^{(j-1)} + c_2 (\theta^{(j)} - \xi^{(j)})$ 
end for
    
```

The distinct feature of the proposed ADMoM-based algorithm for solving the GS-Lasso problem (14) is its computationally affordable implementation, offered by the closed-form expressions for the primal variable updates; as well as the simple updates of the dual variables $\eta^{(j)}$ and $\mu^{(j)}$. Furthermore, since ADMoM has provable convergence to the global minimizer when the considered problem is convex, convergence of the proposed algorithm to $\hat{\theta}$ in (6) is ensured as stated next.

Proposition 1. For any $c_1, c_2 > 0$ and any initializing vectors $\gamma^{(0)}, \xi^{(0)}, \eta^{(0)}$ and $\mu^{(0)}$, the iterates (17) for $\theta^{(j)}$, (20) for $\{\gamma_g^{(j)}\}$, (21) for $\xi^{(j)}$, and (16c)–(16d) for $\eta^{(j)}$ and $\mu^{(j)}$, respectively, are convergent. Also, $\theta^{(j)}$ converges to the solution of the GS-Lasso (6); i.e., $\lim_{j \rightarrow +\infty} \theta^{(j)} = \hat{\theta}$. □

A couple of remarks are now in order.

Remark 1. Shadow fading as well as possible slow temporal variations of the PU transmit-PSDs lead to time-varying $\{\Phi_{x_r}(f)\}$. Following the lines of [20], time-varying PSDs can be tracked by employing the following time-weighted version of the GS-Lasso [cf. (6)]

$$\hat{\theta}(t) = \arg \min_{\theta \geq \mathbf{0}_{N_b N_g}} \left[\frac{1}{2} \sum_{\tau=1}^t \beta_{\tau,t} \sum_{r=1}^{N_r} \|\bar{\varphi}_{x_r}(\tau) - \mathbf{B}_{x_r}(\tau)\theta\|_2^2 + \lambda_1 \|\theta\|_1 + \lambda_2 \mathcal{R}_G(\theta) \right] \quad (23)$$

where $\beta_{\tau,t} \in (0, 1]$ is the so-called forgetting factor, and index $\tau = 1, \dots, t$ emphasizes the temporal variability of channels and received PSDs. Also, to address the need for real-time processing, the estimation of θ in (23) can be performed *on-line* [20], where each iteration of the ADMoM algorithm is performed after acquiring new estimates $\{\hat{\varphi}_{x_r}(\tau)\}$. In this case, the ADMoM iteration index j coincides with the temporal index τ . □

Remark 2. Algorithm 1 is centralized, meaning that the whole set of PSD estimates $\{\hat{\varphi}_{x_r}\}_{r=1}^{N_r}$ must be available at either an FC or a CR cluster head. To reduce the considerable message-passing overhead associated with globally sharing PSD measurements across CRs, and to address scalability and robustness concerns (FC constitutes an isolated point of failure), a distributed counterpart of Algorithm 1 can be derived along the lines of [9,28]. □

4. Spectrum sensing under channel uncertainties

4.1. Group sparse total least-squares

Uncertainty in the matrices $\{\mathbf{B}_{x_r}\}$ is manifested because of (i) errors in the estimates of $\{g_{x_g \rightarrow x_r}\}$ (with or without accounting for shadowing [13,10,14]); (ii) grid offsets when PUs are located between grid points; and, (iii) basis expansion approximation errors. To cope with these perturbations, a robust version of the GS-Lasso is developed in this section.

TLS is the workhorse used for estimating non-sparse vectors obeying an over-determined linear system of equations with uncertainty present in both the regression matrix and the observations (fully-perturbed model) [29]. Sparsity in the estimate was taken into account in [16], where the TLS framework was extended to solve sparse under-determined fully-perturbed linear systems. The sparse TLS approach is broadened here to account for sparsity present both at individual entries, and also at groups of entries.

Define $\mathbf{B} := [\mathbf{B}_{x_1}^T, \dots, \mathbf{B}_{x_{N_g}}^T]^T$, and let $\bar{\varphi} := [\bar{\varphi}_{x_1}^T, \dots, \bar{\varphi}_{x_{N_g}}^T]^T$, and \mathbf{E} a $NN_r \times N_g N_b$ matrix capturing perturbations corrupting the matrix \mathbf{B} . Consider now estimating θ as follows

$$\{\hat{\theta}, \hat{\mathbf{E}}\} = \arg \min_{\substack{\mathbf{E} \\ \theta \geq \mathbf{0}_{N_b N_g}}} \left[\frac{1}{2} \|\bar{\varphi} - (\mathbf{B} + \mathbf{E})\theta\|_2^2 + \frac{1}{2} \|\mathbf{E}\|_F^2 + \lambda_1 \|\theta\|_1 + \lambda_2 \mathcal{R}_G(\theta) \right]. \quad (24)$$

Relative to the classical TLS [29], the cost in the group sparse (GS-)TLS problem (24) is augmented with the regularization terms accounting for the two forms of sparsity inherent to θ . Compared to [16], problem (24) includes also the term $\lambda_2 \mathcal{R}_G(\theta)$.

Problem (24) is generally non-convex due to the presence of the product $\mathbf{E}\theta$; thus, it is in general difficult to obtain a globally-optimal solution. However, a novel reduced-complexity algorithm with provable convergence to a stationary point of (24) will be developed in the ensuing section.

4.2. Alternating descent solver

The cost in (24) will be optimized here iteratively using a block coordinate descent algorithm, which cyclically minimizes it wrt \mathbf{E} (keeping θ fixed), and wrt θ after fixing \mathbf{E} [30]. Specifically, the following two steps are performed at the i -th iteration:

(i1) Fix $\mathbf{E} = \hat{\mathbf{E}}^{(i-1)}$, and update $\hat{\theta}^{(i)}$ as

$$\hat{\theta}^{(i)} = \arg \min_{\theta \succeq \mathbf{0}_{N_b N_g}} \left[\frac{1}{2} \left\| \bar{\varphi} - \left(\mathbf{B} + \hat{\mathbf{E}}^{(i-1)} \right) \theta \right\|_2^2 + \lambda_1 \|\theta\|_1 + \lambda_2 \mathcal{R}_G(\theta) \right]. \quad (25)$$

(i2) Fix $\theta = \hat{\theta}^{(i)}$, and obtain $\hat{\mathbf{E}}^{(i)}$ as

$$\hat{\mathbf{E}}^{(i)} = \arg \min_{\mathbf{E}} \frac{1}{2} \left\| \bar{\varphi} - \mathbf{B} \hat{\theta}^{(i)} - \mathbf{E} \hat{\theta}^{(i)} \right\|_2^2 + \frac{1}{2} \|\mathbf{E}\|_F^2. \quad (26)$$

By fixing $\hat{\mathbf{E}}^{(i-1)}$, (24) boils down to a GS-Lasso problem; thus, $\hat{\theta}^{(i)}$ can be computed by using the ADMoM-based solver of Section 3.2 after replacing (17) with the following update, where j still represents the index for the (inner) ADMoM iterations:

$$\hat{\theta}^{(i,j)} = \left[\left(\mathbf{B} + \hat{\mathbf{E}}^{(i-1)} \right)^T \left(\mathbf{B} + \hat{\mathbf{E}}^{(i-1)} \right) + (c_1 + c_2) \mathbf{I}_{N_b N_g} \right]^{-1} \times \left[\left(\mathbf{B} + \hat{\mathbf{E}}^{(i-1)} \right)^T \bar{\varphi} + c_1 \boldsymbol{\gamma}^{(j-1)} + c_2 \xi^{(j-1)} - \eta^{(j-1)} - \mu^{(j-1)} \right]. \quad (27)$$

The quadratic convex problem (26) admits the following closed-form solution

$$\hat{\mathbf{E}}^{(i)} = (1 + \|\hat{\theta}^{(i)}\|_2^2)^{-1} \left[\bar{\varphi} - \mathbf{B} \hat{\theta}^{(i)} \right] \hat{\theta}^{(i)T} \quad (28)$$

which can be obtained after equating the derivative of the cost in (26) with zero. The overall solver for GS-TLS is tabulated as Algorithm 2.

Algorithm 2 GS-TLS

```

Initialize  $\hat{\mathbf{E}}^{(0)} = \mathbf{0}_{N_b N_g \times N_b N_g}$ 
while Stopping criterion is not satisfied ( $i$  iteration index) do
  Initialize  $\boldsymbol{\gamma}^{(0)} = \mathbf{0}_{N_b N_g}$ ,  $\xi^{(0)} = \mathbf{0}_{N_b N_g}$ ,  $\eta^{(0)} = \mathbf{0}_{N_b N_g}$ ,  $\mu^{(0)} = \mathbf{0}_{N_b N_g}$ 
  while Stopping criterion is not satisfied ( $j$  iteration index) do
    Update  $\theta^{(i,j)}$  via (27)
    Update  $\boldsymbol{\gamma}_g^{(j)}$  via (20) with  $\theta^{(i,j)}$  in place of  $\theta^{(j)}$ , for all  $g = 1, \dots, N_g$ 
    Update  $\xi^{(j)}$  via (22) with  $\theta^{(i,j)}$  in place of  $\theta^{(j)}$ 
    Update  $\eta^{(j)} = \eta^{(j-1)} + c_1 (\theta^{(i,j)} - \boldsymbol{\gamma}^{(j)})$ 
    Update  $\mu^{(j)} = \mu^{(j-1)} + c_2 (\theta^{(i,j)} - \xi^{(j)})$ 
  end while
  Update  $\hat{\theta}^{(i)} = \theta^{(i,j)}$ 
  Update  $\hat{\mathbf{E}}^{(i)} = (1 + \|\hat{\theta}^{(i)}\|_2^2)^{-1} \left[ \bar{\varphi} - \mathbf{B} \hat{\theta}^{(i)} \right] \hat{\theta}^{(i)T}$ 
end while

```

Under certain conditions, the block coordinate descent algorithm is known to converge (at least) to a local optimum point, as asserted next.

Algorithm 3 Robust GS-TLS

```

Initialize  $\hat{\theta}^{(0)} = \mathbf{0}_{N_b N_g}$  and  $\hat{\mathbf{E}}^{(0)} = \mathbf{0}_{N_b N_g \times N_b N_g}$ 
while Stopping criterion is not satisfied ( $i$  iteration index) do
  Initialize  $\boldsymbol{\gamma}^{(0)} = \mathbf{0}_{N_b N_g}$ ,  $\xi^{(0)} = \mathbf{0}_{N_b N_g}$ ,  $\eta^{(0)} = \mathbf{0}_{N_b N_g}$ ,  $\mu^{(0)} = \mathbf{0}_{N_b N_g}$ 
  while Stopping criterion is not satisfied ( $j$  iteration index) do
    Update  $\theta^{(i,j)}$  via (37)
    Update  $\boldsymbol{\gamma}_g^{(j)}$  via (20) with  $\theta^{(i,j)}$  in place of  $\theta^{(j)}$ , for all  $g = 1, \dots, N_g$ 
    Update  $\mathbf{o}^{(i,j)}$  via (38)
    Update  $\xi^{(j)}$  via (22) with  $\theta^{(i,j)}$  in place of  $\theta^{(j)}$ 
    Update  $\eta^{(j)} = \eta^{(j-1)} + c_1 (\theta^{(i,j)} - \boldsymbol{\gamma}^{(j)})$ 
    Update  $\mu^{(j)} = \mu^{(j-1)} + c_2 (\theta^{(i,j)} - \xi^{(j)})$ 
  end while
  Update  $\hat{\theta}^{(i)} = \theta^{(i,j)}$  and  $\hat{\mathbf{E}}^{(i)} = \mathbf{o}^{(i,j)}$ 
  Update  $\hat{\mathbf{E}}^{(i)} = (1 + \|\hat{\theta}^{(i)}\|_2^2)^{-1} \left[ \bar{\varphi} - \mathbf{B} \hat{\theta}^{(i)} + \hat{\mathbf{o}}^{(i)} \right] \hat{\theta}^{(i)T}$ 
end while

```

Proposition 2. For any initialization $\{\hat{\theta}^{(0)}, \hat{\mathbf{E}}^{(0)}\}$, the iterates $\{\hat{\theta}^{(i)}, \hat{\mathbf{E}}^{(i)}\}$ in (25)–(26) converge monotonically to a stationary point of problem (24).

Proof. The proof uses the result of [16, Prop. 3]. Concisely, the cost function in (24) satisfies the Assumptions (B1)–(B3) and (C2) in [30], thus ensuring convergence of $\{\hat{\theta}^{(i)}, \hat{\mathbf{E}}^{(i)}\}$ to a minimum point of the cost as proved in [30, Thm. 5.1]. \square

5. Outlier-resilient spectrum sensing

5.1. Robust GS-TLS

The problem dealt with in the previous section accounts for uncertainty in the entries of the regression matrix \mathbf{B} . However, due to particularly abrupt local shadow fading, failures of the sensing modules, or unexpected narrow-band impulsive noise and/or interference, CRs observations may be affected by abundant errors. This section develops schemes for discerning and removing the observations that largely deviate from the underlying model (a.k.a. outliers) [31,17].

A simple heuristic to detect unreliable data could be to estimate θ via (24), compute the residuals and, then, reject the PSD observations whose residuals exceed a certain threshold. A systematic method that accounts for possible outliers can be found in [32], where the underlying linear regression model is augmented by an auxiliary outlier vector. Using this model, the receiver PSD at the CR locations $\bar{\varphi}$ can be expressed as

$$\bar{\varphi} = (\mathbf{B} + \mathbf{E})\theta + \mathbf{o} + \mathbf{e} \quad (29)$$

where the nonzero entries of the $NN_r \times 1$ real vector \mathbf{o} capture outliers; and \mathbf{e} is a proper vectorization of the periodogram estimation errors [cf. (5)]. Since few outliers are expected compared to the total number of data collected in $\bar{\varphi}$, the vector \mathbf{o} is sparse.

Capitalizing on the three forms of sparsity emerging from (i) the grid-based model (group sparsity), (ii) the PSD basis expansion (single-coefficient sparsity) and (iii) the

outliers (single-coefficient sparsity), and accounting for perturbations in the regression matrix, the following robust GS-TLS is considered:

$$\begin{aligned} \{\hat{\boldsymbol{\theta}}, \hat{\mathbf{E}}, \hat{\mathbf{o}}\} = \arg \min_{\substack{\boldsymbol{\theta} \geq \mathbf{0}_{N_b N_g} \\ \mathbf{E}}} & \left[\frac{1}{2} \|\bar{\boldsymbol{\varphi}} - (\mathbf{B} + \mathbf{E}) \boldsymbol{\theta} + \mathbf{o}\|_2^2 \right. \\ & \left. + \frac{1}{2} \|\mathbf{E}\|_F^2 + \lambda_1 \|\boldsymbol{\theta}\|_1 + \lambda_2 \mathcal{R}_G(\boldsymbol{\theta}) + \lambda_3 \|\mathbf{o}\|_1 \right] \end{aligned} \quad (30)$$

where $\lambda_3 \geq 0$ promotes the (single-coefficient) sparsity of vector $\hat{\mathbf{o}}$. The nature of the perturbations captured in \mathbf{E} and \mathbf{o} is in general different: \mathbf{o} collects unmodeled errors, whereas \mathbf{E} describes (small) perturbations. It is also worth noticing that the support of $\hat{\mathbf{o}}$ reveals the unreliable CR data.

5.2. Alternating descent algorithm

Although (30) is generally a non-convex problem, a block coordinate descent algorithm can still be employed. In this case, the cost in (30) will be iteratively minimized wrt \mathbf{E} and $\{\boldsymbol{\theta}, \mathbf{o}\}$; that is, the following two updates are performed at the i -th iteration:

(i1) Fix $\mathbf{E} = \hat{\mathbf{E}}^{(i-1)}$ and solve

$$\begin{aligned} \{\hat{\boldsymbol{\theta}}^{(i)}, \hat{\mathbf{o}}^{(i)}\} = \arg \min_{\boldsymbol{\theta} \geq \mathbf{0}_{N_b N_g}} & \left[\frac{1}{2} \|\bar{\boldsymbol{\varphi}} - (\mathbf{B} + \hat{\mathbf{E}}^{(i-1)}) \boldsymbol{\theta} + \mathbf{o}\|_2^2 \right. \\ & \left. + \lambda_1 \|\boldsymbol{\theta}\|_1 + \lambda_2 \mathcal{R}_G(\boldsymbol{\theta}) + \lambda_3 \|\mathbf{o}\|_1 \right]. \end{aligned} \quad (31)$$

(i2) Fix $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(i)}$ and $\mathbf{o} = \hat{\mathbf{o}}^{(i)}$ and update $\hat{\mathbf{E}}^{(i)}$ as

$$\hat{\mathbf{E}}^{(i)} = \arg \min_{\mathbf{E}} \frac{1}{2} \|\bar{\boldsymbol{\varphi}} - \mathbf{B}\hat{\boldsymbol{\theta}}^{(i)} + \mathbf{E}\hat{\boldsymbol{\theta}}^{(i)} + \hat{\mathbf{o}}^{(i)}\|_2^2 + \frac{1}{2} \|\mathbf{E}\|_F^2. \quad (32)$$

The quadratic problem (32) can be solved in closed form, to obtain

$$\hat{\mathbf{E}}^{(i)} = (1 + \|\hat{\boldsymbol{\theta}}^{(i)}\|_2^2)^{-1} \left[\bar{\boldsymbol{\varphi}} - \mathbf{B}\hat{\boldsymbol{\theta}}^{(i)} + \hat{\mathbf{o}}^{(i)} \right] \hat{\boldsymbol{\theta}}^{(i)T}. \quad (33)$$

As for (31), the ADMoM can be employed to find its optimal solution. To this end, (31) can be re-formulated as [cf. (14)]

$$\begin{aligned} \{\hat{\boldsymbol{\theta}}^{(i)}, \boldsymbol{\gamma}, \boldsymbol{\xi}, \hat{\mathbf{o}}^{(i)}\} = \arg \min_{\substack{\boldsymbol{\theta} \geq \mathbf{0}_{N_b N_g} \\ \boldsymbol{\gamma}, \boldsymbol{\xi}, \mathbf{o}}} & \left[\frac{1}{2} \|\bar{\boldsymbol{\varphi}} - (\mathbf{B} + \hat{\mathbf{E}}^{(i-1)}) \boldsymbol{\theta} \right. \\ & \left. + \|\mathbf{o}\|_2^2 + \lambda_1 \|\boldsymbol{\xi}\|_1 + \lambda_2 \mathcal{R}_G(\boldsymbol{\gamma}) + \lambda_3 \|\mathbf{o}\|_1 \right] \\ & \text{s.t. } \boldsymbol{\xi} \geq \mathbf{0}_{N_b N_g} \\ & \boldsymbol{\theta} = \boldsymbol{\gamma}, \quad \boldsymbol{\theta} = \boldsymbol{\xi} \end{aligned} \quad (34)$$

with $\boldsymbol{\gamma}$ and $\boldsymbol{\xi}$ denoting auxiliary vector variables. Letting again $\boldsymbol{\eta}$ and $\boldsymbol{\mu}$ denote the Lagrange multipliers associated with the constraints $\boldsymbol{\theta} = \boldsymbol{\gamma}$ and $\boldsymbol{\theta} = \boldsymbol{\xi}$, respectively,

the quadratically augmented Lagrangian function (34) is given by

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\xi}, \mathbf{o}, \boldsymbol{\eta}, \boldsymbol{\mu}) = & \frac{1}{2} \|\bar{\boldsymbol{\varphi}} - (\mathbf{B} + \hat{\mathbf{E}}^{(i-1)}) \boldsymbol{\theta} + \mathbf{o}\|_2^2 \\ & + \lambda_1 \|\boldsymbol{\xi}\|_1 + \lambda_2 \mathcal{R}_G(\boldsymbol{\gamma}) + \lambda_3 \|\mathbf{o}\|_1 \\ & + \boldsymbol{\eta}^T (\boldsymbol{\theta} - \boldsymbol{\gamma}) + \boldsymbol{\mu}^T (\boldsymbol{\theta} - \boldsymbol{\xi}) \\ & + \frac{c_1}{2} \|\boldsymbol{\theta} - \boldsymbol{\gamma}\|_2^2 + \frac{c_2}{2} \|\boldsymbol{\theta} - \boldsymbol{\xi}\|_2^2. \end{aligned} \quad (35)$$

Starting from any initial vectors $\mathbf{o}^{(0)}, \boldsymbol{\gamma}^{(0)}, \boldsymbol{\xi}^{(0)}, \boldsymbol{\eta}^{(0)}, \boldsymbol{\mu}^{(0)}$, each iteration j of the ADMoM (within each iteration i of the block coordinate descent) proceeds in these steps:

$$\boldsymbol{\theta}^{(i,j)} = \arg \min_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\gamma}^{(i,j-1)}, \boldsymbol{\xi}^{(i,j-1)}, \mathbf{o}^{(i,j-1)}, \boldsymbol{\eta}^{(i,j-1)}, \boldsymbol{\mu}^{(i,j-1)}) \quad (36a)$$

$$\boldsymbol{\gamma}^{(j)} = \arg \min_{\boldsymbol{\gamma}} \mathcal{L}(\boldsymbol{\theta}^{(i,j)}, \boldsymbol{\gamma}, \boldsymbol{\xi}^{(i,j-1)}, \mathbf{o}^{(i,j-1)}, \boldsymbol{\eta}^{(i,j-1)}, \boldsymbol{\mu}^{(i,j-1)}) \quad (36b)$$

$$\boldsymbol{\xi}^{(j)} = \arg \min_{\boldsymbol{\xi} \geq \mathbf{0}} \mathcal{L}(\boldsymbol{\theta}^{(i,j)}, \boldsymbol{\gamma}^{(j)}, \boldsymbol{\xi}, \mathbf{o}^{(i,j-1)}, \boldsymbol{\eta}^{(i,j-1)}, \boldsymbol{\mu}^{(i,j-1)}) \quad (36c)$$

$$\mathbf{o}^{(i,j)} = \arg \min_{\mathbf{o}} \mathcal{L}(\boldsymbol{\theta}^{(i,j)}, \boldsymbol{\gamma}, \boldsymbol{\xi}^{(i,j-1)}, \mathbf{o}, \boldsymbol{\eta}^{(i,j-1)}, \boldsymbol{\mu}^{(i,j-1)}) \quad (36d)$$

$$\boldsymbol{\eta}^{(j)} = \boldsymbol{\eta}^{(j-1)} + c_1 (\boldsymbol{\theta}^{(i,j)} - \boldsymbol{\gamma}^{(j)}) \quad (36e)$$

$$\boldsymbol{\mu}^{(j)} = \boldsymbol{\mu}^{(j-1)} + c_2 (\boldsymbol{\theta}^{(i,j)} - \boldsymbol{\xi}^{(j)}). \quad (36f)$$

Since (34) is convex and satisfies the requirements for the ADMoM to be convergent [12], iterates $\{\boldsymbol{\theta}^{(i,j)}, \mathbf{o}^{(i,j)}\}$ will converge to the solution $\{\hat{\boldsymbol{\theta}}^{(i)}, \hat{\mathbf{o}}^{(i)}\}$ of (31).

Problem (36a) admits the closed-form solution given by

$$\begin{aligned} \boldsymbol{\theta}^{(i,j)} = & \left[(\mathbf{B} + \hat{\mathbf{E}}^{(i-1)})^T (\mathbf{B} + \hat{\mathbf{E}}^{(i-1)}) + (c_1 + c_2) \mathbf{I}_{N_b N_g} \right]^{-1} \\ & \times \left[(\mathbf{B} + \hat{\mathbf{E}}^{(i-1)}) (\bar{\boldsymbol{\varphi}} + \mathbf{o}^{(i,j-1)}) \right. \\ & \left. + c_1 \boldsymbol{\gamma}^{(j-1)} + c_2 \boldsymbol{\xi}^{(i,j-1)} - \boldsymbol{\eta}^{(j-1)} - \boldsymbol{\mu}^{(i,j-1)} \right] \end{aligned} \quad (37)$$

where $\boldsymbol{\gamma}^{(j)}$ and $\boldsymbol{\xi}^{(j)}$ are still computed via (20) and (22). Using Lemma 2, it is possible to show that the solution of (36d) is computed via soft-thresholding as

$$\mathbf{o}^{(i,j)} = \mathcal{T}_{\lambda_3} \left(\bar{\boldsymbol{\varphi}} - (\mathbf{B} + \hat{\mathbf{E}}^{(i-1)}) \boldsymbol{\theta}^{(i,j)} \right). \quad (38)$$

At each step of the ADMoM algorithm the soft-thresholding in (38) tags as outliers the (current) residuals $\bar{\boldsymbol{\varphi}} - (\mathbf{B} + \hat{\mathbf{E}}^{(i-1)}) \boldsymbol{\theta}^{(i,j)}$ that exceed λ_3 . The overall solver is tabulated as Algorithm 3.

With arguments similar to Proposition 2, the following result can be asserted.

Proposition 3. For any initialization $\{\hat{\boldsymbol{\theta}}^{(0)}, \hat{\mathbf{o}}^{(0)}, \hat{\mathbf{E}}^{(0)}\}$, the iterates $\{\hat{\boldsymbol{\theta}}^{(i)}, \hat{\mathbf{o}}^{(i)}, \hat{\mathbf{E}}^{(i)}\}$ in (32)–(31) converge monotonically to a stationary point of problem (30).

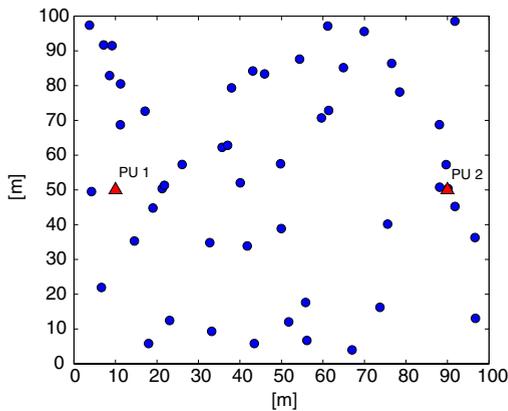


Fig. 2. CR network topology.

6. Simulations

Consider a set of $N_r = 50$ CRs uniformly distributed in an area of $100\text{ m} \times 100\text{ m}$, cooperating to localize $N_s = 2$ active PUs and estimate their PSD map. CRs and PU sources are marked with blue circles and red triangles, respectively, in Fig. 2. PUs transmit raised cosine pulses with unitary amplitude (0dB), roll-off factor $\rho = 0.5$, and bandwidth $W = 10$ MHz. They share the band $B = [100, 200]$ MHz with spectra centered at frequencies $f_c = 115$ and 175 MHz for “PU 1” and “PU 2”, respectively. CRs adopt a path loss-only model to accomplish the sensing task.

Transmitted signals are searched over a grid of $N_b = 10$ evenly spaced center frequencies $f_c = 95 + \nu W$, $\nu \in \{1, \dots, 10\}$. Each CR computes periodogram samples at $N = 64$ frequencies at signal-to-noise-ratio (SNR) -5 dB, and averages them across $\tau = 100$ time-slots to form $\hat{\Phi}_{x_r}(\tau, f_k)$, $k = 1, \dots, 64$, as in (4).

In the first experiment, the PSD generated by PU s experiences only small-scale fading in its propagation from \mathbf{x}_s to any location \mathbf{x} , where it is measured in the presence of noise with variance $\sigma_v^2 = 0.1$. To simulate small-scale fading $\{h_{\mathbf{x}_s \rightarrow \mathbf{x}}\}$, a 6-tap Rayleigh model with exponential power delay profile is adopted. Since the expected gain adheres to a path loss propagation law, the regression matrix

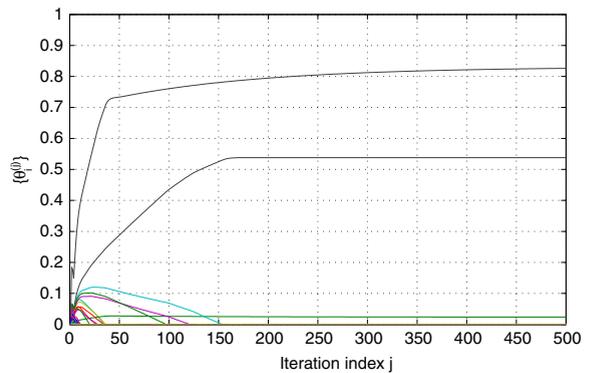


Fig. 4. Evolution of the entries of $\theta^{(j)}$.

is in this case perfectly known. Fig. 3(a) depicts the true PSD map summed across frequencies, which peaks at the active PU locations. To localize and estimate the transmit-PSDs, a grid of $N_g = 100$ equidistant points is used. The map obtained by using the GS-Lasso-based sensing algorithm is shown in Fig. 3(b), which also depicts the estimated positions of the transmitting PUs along with their transmission powers (in dB) represented by the coefficients of the normalized raised-cosines. One can readily notice that “PU 1” is perfectly localized and its transmit-PSD is estimated accurately. As for “PU 2”, its location is revealed although spurious power is also leaked to an adjacent grid point. The sparsity-promoting parameters λ_1 and λ_2 are set to $\lambda_1 = 30 \cdot \max\{\mathbf{B}^T \hat{\phi}\}$ and $\lambda_2 = 10 \cdot \max_r\{\|\mathbf{B}_{x_r}^T \hat{\phi}\|_2\}$, respectively [9]. Fig. 4 corroborates the convergence of the GS-Lasso solver by showing the evolution of the elements of $\theta^{(j)}$ [cf. Fig. 3(b)].

In Fig. 5, the transmit-PSDs undergo not only small-scale fading but also log-normal shadowing. As the CRs employ a path loss-based model, shadowing here perturbs the regression matrix. Expressed in logarithmic scale, the shadowing process has zero mean and standard deviation 6dB. The estimated PSD maps obtained by using the “plain” GS-Lasso and the GS-TLS-based sensing algorithm of Section 4.1 are compared in Fig. 6.

Fig. 6(a) illustrates that the GS-Lasso is unable to localize the two PUs, as clouds of PU sources are falsely

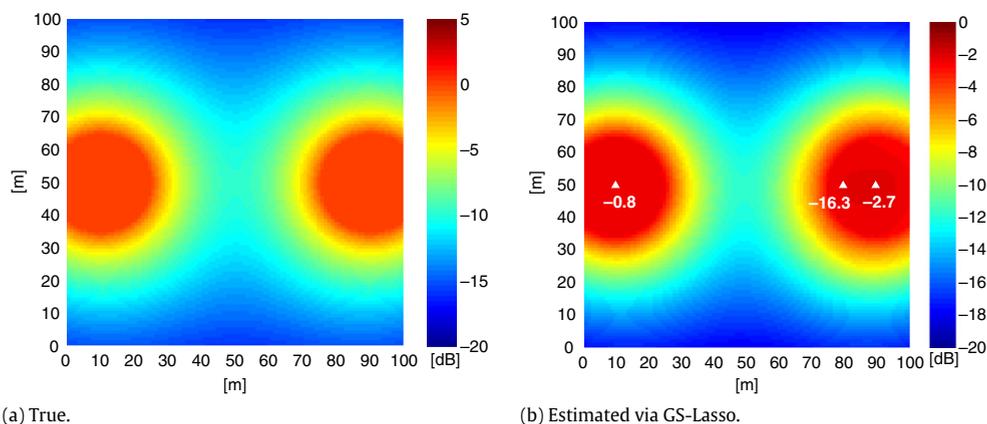


Fig. 3. PU PSD maps (path loss-only propagation).

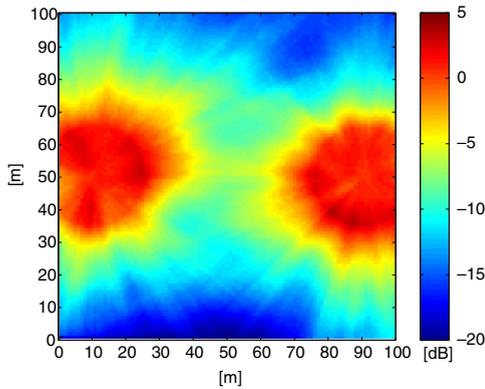


Fig. 5. True PSD map (shadowing propagation effects).

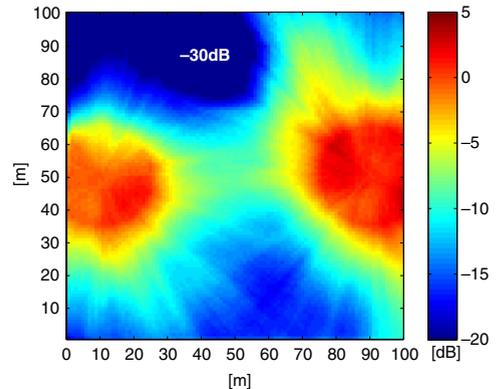


Fig. 7. True PSD map (deep shadowing caused by an obstacle).

revealed around the actual locations of “PU 1” and “PU 2”. Also, it does not accurately estimate their PSDs. On the other hand, the GS-TLS algorithm reveals the exact location of both PUs, although a rather small amount of spurious power is leaked to a grid point close to “PU 1”. Note also that the transmit-powers are estimated with considerably higher accuracy. Numerical experiments have shown that only a few (5–10) iterations suffice for the alternating descent algorithm to converge.

The enhanced localization and power estimation capabilities impact also the subsequent CR power allocation task, which relies on the estimated coverage region of the PU-transmitters to re-use the licensed bands without causing harmful interference to any potential PU receiver [25]. GS-Lasso will be preferable if a coarse description of the “interference-heavy” areas in terms of PU activity is desired over say accurate localization and transmit-power estimation of the PUs. Such a coarse can be useful for e.g., temporal (rather than spatio-temporal) frequency re-use purposes.

As mentioned in Section 5, abrupt local shadow fading may severely compromise the PSD estimates at CRs, and thus degrade the sensing performance. This is the case considered in Fig. 7, where an obstacle positioned in the upper-left part of the monitored area causes deep fades of the receive-power at some CRs. Fig. 8(a) demonstrates that

“plain” GS-Lasso fails to localize the two PU transmitters. Activity is revealed around the actual location of “PU 1”, and shadowing causes the false detection of a third low-power PU in position $\mathbf{x}_g = (80, 70)$ transmitting over the same band of “PU 2”. This false-detection event is not present in Fig. 8(b), where the robust GS-TLS algorithm of Section 5 is used. In fact, “PU 2” is well-localized and its transmit-PSD is estimated accurately. A small amount of power is still dribbled on an adjacent grid point of “PU 2”. With $\lambda_3 = 30 \cdot \max\{\mathbf{B}^T \hat{\phi}\}$, further analysis of the data reveals that 15% of the periodogram samples, specifically those collected by the “faded CRs”, was declared unreliable, and was thus discarded.

7. Concluding remarks

Spatio-temporal and dynamic re-use of the licensed bands calls for collaborative CR network sensing algorithms able to portray the ambient power spectral density at arbitrary locations in space, frequency, and time. The present paper addressed this ambitious task through a parsimonious model of the PSD in frequency and space, which reduces the sensing task to estimating a sparse vector of unknown parameters. An estimator of the model parameters was developed based on the GS-Lasso, and

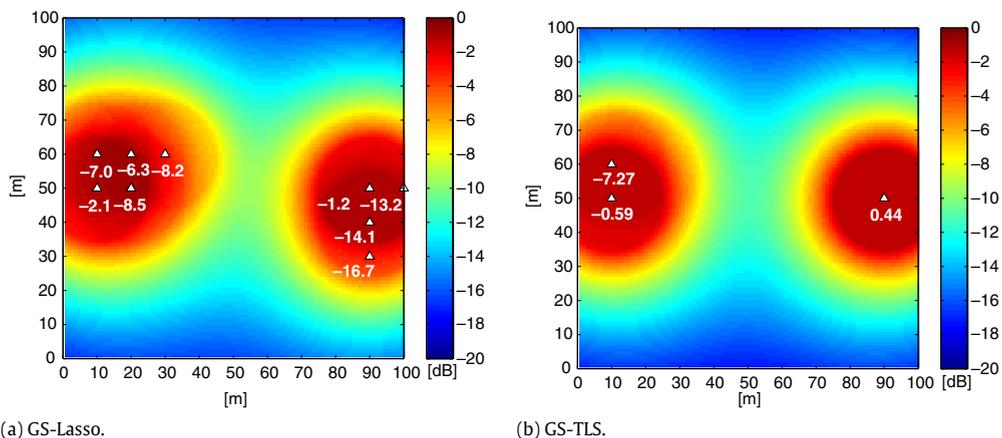


Fig. 6. Estimated PSD maps of Fig. 5.

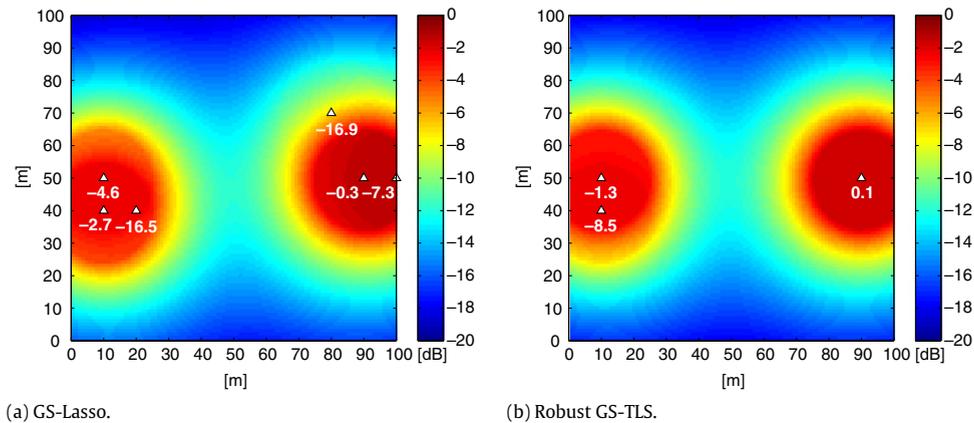


Fig. 8. Estimated PSD maps of Fig. 7.

a low-complexity solver based on the ADMM was presented. The location and transmit-PSD information conveyed by the estimated model parameters, complemented with either a path loss-based or more elaborated propagation models was shown to allow CRs to accurately reconstruct the PSD atlas of the primary system. To cope with uncertainty in the regression matrix, a provably convergent sensing algorithm was introduced which combines the merits of the TLS framework with the hierarchical sparsity inherent to the network-level sensing problem. To account for outliers, a robust algorithm able to discern and reject unreliable PSD data was also developed. The novel robust GS-TLS approach capitalizes on the sparsity of the unknown parameters and the outliers, and offers systematic estimation of the spectrum holes jointly in space, frequency, and time while taking into account channel uncertainties and unmodeled errors.

References

- [1] Q. Zhao, B.M. Sadler, A survey of dynamic spectrum access, *Signal Process. Mag. IEEE* 24 (3) (2007) 79–89.
- [2] X. Hong, C. Wang, H. Chen, Y. Zhang, Secondary spectrum access networks, *IEEE Veh. Technol. Mag.* 4 (2) (2009) 36–43.
- [3] G. Ganesan, Y. Li, B. Bing, S. Li, Spatio-temporal sensing in cognitive radio networks, *IEEE J. Sel. Areas Commun.* 26 (1) (2008) 5–12.
- [4] Z. Quan, S. Cui, H.V. Poor, A.H. Sayed, Collaborative wideband sensing for cognitive radios, *IEEE Signal Process. Mag.* 25 (6) (2008) 60–73.
- [5] J. Ma, G. Zhao, Y. Li, Soft combination and detection for cooperative spectrum sensing in cognitive radio networks, *IEEE Trans. Wireless Commun.* 7 (11) (2008) 4502–4507.
- [6] J. Unnikrishnan, V.V. Veeravalli, Cooperative sensing for primary detection in cognitive radio, *IEEE J. Sel. Top. Signal Process.* 2 (1) (2008) 18–27.
- [7] S.-J. Kim, G.B. Giannakis, Sequential and cooperative sensing for multichannel cognitive radios, *IEEE Trans. Signal Process.* 58 (8) (2010) 4239–4253.
- [8] A.B.H. Alaya-Feki, S.B. Jemaa, B. Sayrac, P. Houze, E. Moulines, Informed spectrum usage in cognitive radio networks: interference cartography. In *Proc. of the PIMRC Conf.*, 1–5, Cannes, France, Sep. 2008.
- [9] J.A. Bazerque, G. Mateos, G.B. Giannakis, Group-lasso on splines for spectrum cartography, *IEEE Trans. Signal Process.* 59 (10) (2011) 4648–4663.
- [10] E. Dall'Anese, S.-J. Kim, G.B. Giannakis, Channel gain map tracking via distributed kriging, *IEEE Trans. Veh. Technol.* 60 (3) (2011) 1205–1211.
- [11] J. Friedman, T. Hastie, R. Tibshirani, A note on the group Lasso and a sparse group Lasso Preprint, 2010. Downloadable from <http://www-stat.stanford.edu/~tibs>.
- [12] D.P. Bertsekas, J.N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Prentice-Hall, Englewood Cliffs, NJ, 1989.
- [13] J.-A. Bazerque, G.B. Giannakis, Distributed spectrum sensing for cognitive radio networks by exploiting sparsity, *IEEE Trans. Signal Process.* 58 (3) (2010) 1847–1862.
- [14] B. Mark, A. Nasif, Estimation of maximum interference-free power level for opportunistic spectrum access, *IEEE Trans. Wireless Commun.* 8 (5) (2009) 2505–2513.
- [15] M.A. Herman, T. Ströhmer, General deviants: an analysis of perturbations in compressive sensing, *IEEE J. Sel. Top. Signal Process.* 4 (2010) 342–349.
- [16] H. Zhu, G. Leus, G.B. Giannakis, Sparsity-cognizant total least-squares for perturbed compressive sampling, *IEEE Trans. Signal Process.* 59 (2011).
- [17] Y. Zhang, N. Meratnia, P. Havinga, Outlier detection techniques for wireless sensor networks: a survey, *IEEE Commun. Surveys Tuts.* 12 (2) (2010) 159–170.
- [18] T.S. Rappaport, *Wireless Communications: Principles & Practice*, Prentice Hall, 1996.
- [19] C. Oestges, N. Czink, B. Bandemer, P. Castiglione, F. Kaltenberger, A.J. Paulraj, Experimental characterization and modeling of outdoor-to-indoor and indoor-to-indoor distributed channels, *IEEE Trans. Veh. Technol.* 59 (5) (2010) 2253–2265.
- [20] D. Angelosante, J.A. Bazerque, G.B. Giannakis, Online adaptive estimation of sparse signals: where RLS meets the ℓ_1 -norm, *IEEE Trans. Signal Process.* 58 (7) (2010) 3436–3447.
- [21] R. Tibshirani, Regression shrinkage and selection via the Lasso, *J. R. Stat. Soc.* 58 (1) (1996) 267–288.
- [22] M. Yuan, Y. Lin, Model selection and estimation in regression with grouped variables, *J. R. Stat. Soc.* 68 (2006) 49–67.
- [23] P. Sprechmann, I. Ramírez, G. Sapiro, Y.C. Eldar, C-HiLasso: A collaborative hierarchical sparse modeling framework, *IEEE Transactions on Signal Processing* 9 (59) (2011) 4183–4198.
- [24] S.-J. Kim, E. Dall'Anese, G.B. Giannakis, Cooperative spectrum sensing for cognitive radios using kriged Kalman filtering, *IEEE J. Sel. Top. Signal Process.* 5 (2011) 24–36.
- [25] E. Dall'Anese, S.-J. Kim, G.B. Giannakis, S. Pupolin, Power allocation for cognitive radio networks under channel uncertainty, in: *Proc. of the Intl. Conf. on Communications*, Kyoto, Japan, Jun. 2011.
- [26] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [27] A. Ruszczyński, *Nonlinear Optimization*, Princeton Univ. Press, Princeton, NJ, 2006.
- [28] G. Mateos, J.-A. Bazerque, G.B. Giannakis, Distributed sparse linear regression, *IEEE Trans. Sig. Proc.* 58 (10) (2010) 5262–5276.
- [29] S. Van Huffel, J. Vandewalle, *The total least-squares problem: computational aspects and analysis*, in: *ser. Frontier in Applied Mathematics*, SIAM, Philadelphia, 1991.
- [30] P. Tseng, Convergence of block coordinate descent method for nondifferentiable maximization, *J. Optim. Theory Appl.* 109 (3) (2001) 475–494.

- [31] P.J. Huber, E.M. Ronchetti, *Robust Statistics*, Wiley, New York, 2006.
 [32] J.J. Fuchs, A new approach to robust linear regression. in: Proc. of 14th IFAC world congress, pp. 427–432, Beijing, 1999.



Emiliano Dall'Anese (S'08, M'10) was born in Belluno, Italy, on 10 March, 1983. He received the Laurea Triennale (B.Sc degree) and the Laurea Specialistica (M.Sc degree) in telecommunication engineering from the University of Padova, Italy, in 2005 and 2007, respectively, and the Ph.D in Information Engineering at the Department of Information Engineering (DEI), University of Padova, Italy, in 2011. From January 2009 to September 2010, he was a visiting scholar at the Department of Electrical and

Computer Engineering, University of Minnesota, USA. He is currently a post-doctoral associate at the Department of Electrical and Computer Engineering, University of Minnesota, USA.

His research interests lie in the areas of communication theory and statistical signal processing. His current research focuses on wireless cognitive radio systems, sensor networks, distributed signal processing, and smart grid.



Juan Andrés Bazerque (S'06) received his B.Sc. degree in Electrical Engineering from Universidad de la República (UdelaR), Montevideo, Uruguay in 2003. Since August 2006 he has been a research assistant at the University of Minnesota (UofM), Minneapolis, where he received his M.Sc. in Electrical Engineering in August 2009, and continues working towards the Ph.D. degree. From 2000 to 2006 he was a teaching assistant with the Department of Mathematics and Statistics, and with the Department of Electrical

Engineering (UdelaR). From 2003 to 2006 he worked as a telecommunications engineer at the Uruguayan company Uniotel S.A. developing applications for Voice over IP.

His broad research interests lie in the general areas of networking, communications, and signal processing. His current research focuses on distributed signal processing, cooperative wireless communications, compressive sampling, sparsity-aware statistical models, and gene expression networks.

Juan Andrés Bazerque is the recipient of the UofM's Distinguished Master's Thesis Award 2009-2010, and co-recipient of the best student paper award at the 2nd International Conference on Cognitive Radio Oriented Wireless Networks and Communication 2007.



Georgios B. Giannakis (Fellow'97) received his Diploma in Electrical Engr. from the Ntl. Tech. Univ. of Athens, Greece, 1981. From 1982 to 1986, he was with the Univ. of Southern California (USC), where he received his MSc. in Electrical Engineering, 1983, MSc. in Mathematics, 1986, and Ph.D. in Electrical Engr., 1986. Since 1999, he has been a professor with the Univ. of Minnesota, where he now holds an ADC Chair in Wireless Telecommunications in the ECE Department, and serves as director of

the Digital Technology Center.

His general interests span the areas of communications, networking and statistical signal processing—subjects on which he has published more than 300 journal papers, 550 conference papers, two edited books and two research monographs. Recent and current research focuses on distributed and sparsity-aware signal processing, machine learning, compressed sensing, network coding, cognitive radios, cross-layer designs for wireless mobile networks, and social networks. He is the (co-) inventor of 21 patents issued, and the (co-) recipient of seven paper awards from the IEEE Signal Processing (SP) and Communications Societies, including the G. MarconiPrize Paper Award in Wireless Communications. He also received Technical Achievement Awards from the SP Society (2000), from EURASIP (2005), a Young Faculty Teaching Award, and the G. W. Taylor Award for Distinguished Research from the University of Minnesota. He is also a Fellow of EURASIP, and has served the IEEE in a number of posts, including that of a Distinguished Lecturer for the IEEE-SP Society.