Distributed Spectrum Sensing for Cognitive Radio Networks by Exploiting Sparsity

Juan Andrés Bazerque, *Student Member, IEEE*, and Georgios B. Giannakis, *Fellow, IEEE*

***Abstract*—A cooperative approach to the sensing task of wireless cognitive radio (CR) networks is introduced based on a basis expansion model of the power spectral density (PSD) map in space and frequency. Joint estimation of the model parameters enables identification of the (un)used frequency bands at arbitrary locations, and thus facilitates spatial frequency reuse. The novel scheme capitalizes on two forms of sparsity: the first one introduced by the narrow-band nature of transmit-PSDs relative to the broad swaths of usable spectrum; and the second one emerging from sparsely located active radios in the operational space. An estimator of the model coefficients is developed based on the Lasso algorithm to exploit these forms of sparsity and reveal the unknown positions of transmitting CRs. The resultant scheme can be implemented via distributed online iterations, which solve quadratic programs locally (one per radio), and are adaptive to changes in the system. Simulations corroborate that exploiting sparsity in CR sensing reduces spatial and frequency spectrum leakage by 15 dB relative to least-squares (LS) alternatives.**

***Index Terms*—Cognitive radios, compressive sampling, cooperative systems, distributed estimation, parallel network processing, sensing, sparse models, spectral analysis.**

I. Introduction

SPECTRUM sensing is a critical prerequisite in envisioned applications of wireless cognitive radio (CR) networks which promise to resolve the perceived bandwidth scarcity versus under-utilization dilemma. Creating an interference map of the operational region plays an instrumental role in enabling spatial frequency reuse and allowing for dynamic spectrum allocation in a hierarchical access model comprising primary (licensed) and secondary (opportunistic) users [21], [22]. The non-coherent energy detector has been widely used to this end because it is simple and obviates the need for synchronization with unknown transmitted signals; see e.g., [11], [12], [14], and [17]. Power information (or other statistics [8], [9]) collected locally per CR is fused centrally by an access point in order to decide absence or presence of a primary user per frequency band. At the expense of commensurate communication overhead [12], these cooperative sensing and detection schemes have been shown to increase reliability, reduce the average detection time, cope with fading propagation effects, and improve throughput [9], [11], [14], [17]. Recently, the possibility of spatial reuse has received growing attention. It was noticed that even if a frequency band is occupied, there could be locations where the transmitted power is low enough so that these frequencies can be reused without suffering from or causing harmful interference to the primary system. These opportunities are discussed in [15], and a statistical model for the transmitters' spatial distribution is advocated in [16].

The present paper goes in the direction of sensing these reusable zones, by means of a collaborative scheme whereby receiving CRs cooperate to estimate the distribution of power in space and frequency as well as localize, as a byproduct, the positions of transmitting CRs. The main contribution is a distributed online approach to estimating a map of the power spectral density (PSD) at arbitrary locations in space. This is particularly useful in wide area ad-hoc networks, where the power transmitted by primary users reaches only a small subset of CRs. Knowing the spectrum at any location allows remote CRs to reuse dynamically idle bands. It also enables secondary users to adapt their transmit-power so as to minimally interfere with primary users. In this context, the threshold for deciding occupancy of a frequency band is not set according to the probability of false alarms, but through comparing PSD estimates with minimum power levels prescribed by the primary users.

The goal of estimating the power distribution in space and frequency is admittedly very ambitious. The PSD estimate sought however, does not need to be super accurate but precise enough to identify (un)used bands. This relaxed objective motivates the proposed PSD estimator using a parsimonious basis expansion model. The general setup includes N_r receiving CRs willing to cooperate in estimating the location of N_s transmitting radios as well as the frequency bands used for transmission. Upon constructing a basis expansion model of the PSD map $\Phi_r(f)$, in spatial location $\boldsymbol{x}_r$ and frequency f, the novel cooperative scheme amounts to estimating the basis expansion coefficients of $\Phi_r(f)$ based on PSD frequency samples f_k collected at receiving CRs located at positions $\boldsymbol{x}_r$. These coefficients are inherently sparse given the narrow-band individual transmissions compared to the overall band scanned, as well as the scarce distribution of active transmitters in the area. Sparsity is then exploited as prior information to improve estimation performance by suitably modifying the least-absolute shrinkage and selection operator (Lasso) in [20].

The novel *distributed* algorithm termed D-Lasso implements Lasso using an ad-hoc network of nodes. It does not require

Manuscript received January 12, 2009; accepted November 09, 2009. First published December 11, 2009; current version published February 10, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Daniel Palomar. Prepared through collaborative participation in the Communications and Networks Consortium sponsored by the U.S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon. Results from this paper were presented in the Forty-Second Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, October 26–29, 2008,.

The authors are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55414 USA (e-mail: bazer002@umn.edu; georgios@umn.edu).

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

Digital Object Identifier 10.1109/TSP.2009.2038417

1053-587X/$26.00 © 2010 IEEE

Authorized licensed use limited to: University of Minnesota. Downloaded on February 8, 2010 at 19:22 from IEEE Xplore. Restrictions apply.

coordination through a central unit but the local nodes (or CRs in the present application context) reach consensus on the coefficients by exchanging low-overhead messages with their single-hop neighbors. Global optimality is ensured in the sense that the distributed estimator approaches its centralized counterpart obtained when all observations are available at a central unit. The online version of D-Lasso is capable of tracking changes in the transmit-PSDs; e.g., when a transmitter joins or departs.

The overall sensing approach applies readily to hierarchical cognitive radio networks, where the transmitters belong to a primary system and the sensing receivers are secondary users. This is not compulsory however, since the resultant algorithms work as general sensing and localization tools.

The rest of the paper is organized as follows. Section II introduces the basis expansion model and describes the PSD data used for the model fitting approach. A centralized estimator is developed in Section III with the CR positions being unknown, case in which the PSD model becomes over-complete and sparsity in the vector of expansion coefficients is exploited to reveal the locations and frequency bands of the transmitting CRs. The distributed batch and online algorithms are derived in Section IV. Performance analysis of the expansion coefficient vector and resultant PSD estimators is provided in Section V. Numerical tests are presented in Section VI, and conclusions are drawn in Section VII. Detailed proofs can be found in the Appendixes.

II. PSD Basis Expansion Model

Consider N_s sources (transmitters) located at position vectors $\{\boldsymbol{x}_s\}_{s=1}^{N_s}$ with respect to (w.r.t.) a global reference system; and let $y_r(n) = \sum_{s=1}^{N_s}\sum_{l=0}^{L_{sr}-1} h_{sr}(n;l)u_s(n-l) + v(n)$ denote the received signal at position $\boldsymbol{x}_r$ representing the superposition of the transmitted signals $u_s(n)$, convolved with linear (possibly time-varying) finite-impulse response fading channels $\{h_{sr}(n;l)\}_{l=0}^{L_{sr}-1}$, and observed in the presence of additive white noise $v(n)$.

Received data are parsed in blocks each containing N samples, where N is chosen equal to the coherence interval of the system over which $h_{sr}(n;l)$ remains invariant w.r.t. n. These blocks are indexed by i so that $n = iN + \eta$, with $\eta = 0, 1, \ldots, N-1$.

The frequency-selective fading propagation model used throughout will obey the following assumption.

Assumption 1: *Channels $\{h_{sr}(n;l)\}$ are zero mean with frequency response $H_{sr}(n;f)$, and known gain $\gamma_{sr} := E|H_{sr}(n;f)|^2$; e.g., $\gamma_{sr} = \gamma(\|\boldsymbol{x}_s - \boldsymbol{x}_r\|)$, where γ is a known function of the source-receiver distance. In addition, channels $\{h_{sr}(n;l)\}$ are stationary w.r.t. n and uncorrelated across the block variable i, across the lag variable l, and across the spatial variables s and r.*

Uncorrelatedness of channels across space is realistic since sources are sufficiently apart relative to the high carrier frequencies for which wavelengths are short. Because $h_{sr}(n;l)$ is stationary and uncorrelated across lags l, its PSD $E|H_{sr}(n;f)|^2$ can be readily shown to be time and frequency invariant with $E|H_{sr}(n;f)|^2 = \sum_{l=0}^{L_{sr}-1}\sigma_{sr}^2(l) =: \gamma_{sr}$, where $\sigma_{sr}^2(l) := E|h_{sr}(n;l)|^2$.

Gain γ_{sr} in Assumption 1 can be acquired via training. However, this requires cooperation of the primary system and the ability to separate the sources. The alternative followed here is to adopt a pathloss model $\gamma_{sr} = \gamma(\|\boldsymbol{x}_r - \boldsymbol{x}_s\|)$. One possible choice for $\gamma(\cdot)$ is the inverse polynomial law for which $\gamma_{sr} = \min\{1, (d/d_0)^{-\alpha}\}$, where $d := \|\boldsymbol{x}_r - \boldsymbol{x}_s\|$ and d_0, α are preselected constants that depend on the propagation environment. It is worth stressing that the pathloss here does not apply to each realization $h_{sr}(n;l)$ but to the PSD; hence, the wireless fading channels in this paper are allowed to be frequency selective.

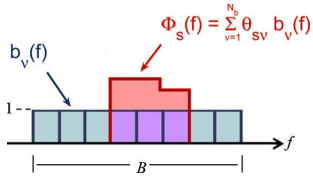

Fig. 1. Expansion with non-overlapping rectangular bases of unit height.

With regards to the transmitting radios, the following is assumed.

Assumption 2: *Sources $\{u_s(n)\}$ are stationary, mutually uncorrelated, independent of the channels $\{h_{sr}(n;l)\}$ with vanishing correlation per coherence interval; i.e., $\varphi_s(m) := E[u_s(n+m)u_s^*(n)] = 0$, $\forall |m| > N - L$, where $L := \max_{s,r} L_{sr}$. Furthermore, with N_b sufficiently large the PSD of each source is well approximated by the basis expansion model*

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

where $\{b_\nu(f)\}_{\nu=1}^{N_b}$ is a collection of known bases, and $\{\theta_{s\nu}\}$ denote the expansion coefficients to be estimated.

Note that although each receiving CR is assumed to know its location $\boldsymbol{x}_r$, the source locations $\{\boldsymbol{x}_s\}_{s=1}^{N_s}$ are not assumed known.

Possible choices of $\{b_\nu(f)\}_{\nu=1}^{N_b}$ include Gaussian bells, or, the set of non-overlapping rectangles of unit height spanning the bandwidth B of interest; see Fig. 1. For the rectangular basis, the coefficient $\theta_{s\nu}$ represents the power emitted by source s on the frequency band corresponding to the basis function $b_\nu(f)$.

With a single source received in the absence of noise, the assumption of uncorrelated channel taps independent of the stationary sources combined with the vanishing memory of $\varphi_s(m)$ allows one to express the received autocorrelation as $\varphi_r(m) = \varphi_s(m)\sum_{l=0}^{L-1} E|h_{sr}(n;l)|^2$; and hence the received PSD as $\Phi_r(f) = \gamma_{sr}\Phi_s(f)$. When N_s sources and receiver noise are present, it follows from the spatial uncorrelatedness of channels and sources that

$$\begin{aligned}\Phi_r(f) &= \sum_{s=1}^{N_s} \gamma_{sr}\Phi_s(f) + \sigma_r^2 \\ &= \sum_{s=1}^{N_s}\gamma_{sr}\sum_{\nu=1}^{N_b}\theta_{s\nu}b_\nu(f) + \sigma_r^2 = \boldsymbol{b}_r^T(f)\boldsymbol{\theta} + \sigma_r^2\end{aligned} \tag{2}$$

where the $P := N_b N_s \times 1$ vector $\boldsymbol{\theta}$ is formed by stacking the columns of the matrix with entries $\theta_{s\nu}$, and $\boldsymbol{b}_r^T(f)$ by concatenating the columns of the matrix with entries $\gamma_{sr}b_\nu(f)$. Note that the *noise-free* PSD in the linear model (2) is completely

Authorized licensed use limited to: University of Minnesota. Downloaded on February 8, 2010 at 19:22 from IEEE Xplore. Restrictions apply.

specified at any receiving point $\boldsymbol{x}_r$ and any frequency bin f, provided that $\boldsymbol{\theta}$ becomes available.

The sensing strategy aiming to obtain $\boldsymbol{\theta}$ (along with σ_r^2) will rely on the periodogram estimate of $\Phi_r(f)$. The fast Fourier transform of $y_r(n)$, namely $Y_{r,N}^{(i)}(f)$, and the periodogram $\hat{I}_{r,N}^{(i)}(f) := (1/N)|Y_{r,N}^{(i)}(f)|^2$ are computed per coherence block i. These are averaged across blocks to obtain $\hat{\Phi}_{r,N}^{(i)} := (1/i)\sum_{i'=0}^{i-1} \hat{I}_{r,N}^{(i')}(f)$. In the limit, this averaging procedure attains the expectation over the channels and transmitted signals. As a result, the averaged periodogram asymptotically yields [4, p. 123], $\lim_{N,i\longrightarrow\infty} \hat{\Phi}_{r,N}^{(i)} = \Phi_r(f)$ with probability one (w.p. 1).

To allow for tracking of slow-varying PSDs, the sample mean across coherence blocks will be replaced by an exponentially weighted moving average (EWMA) with exponent β, given by

$$\hat{\Phi}_{r,N}^{i,\beta}(f) := (1-\beta)\sum_{i'=0}^{i} \beta^{i-i'} \hat{I}_{r,N}^{(i')}(f). \tag{3}$$

The EWMA estimate in (3) weighs more recent values and "forgets" past values. It can be seen as an average modulated by a sliding window of equivalent length $N_\beta := \sum_{i=0}^{\infty} \beta^i = 1/1-\beta$, $|\beta| < 1$. This property allows one to track PSDs that can be considered stationary over a window of N_β coherence blocks. If the EWMA is used instead of the periodogram average, the following result is proved in Appendix A.

Proposition 1: *If Assumption 1 and Assumption 2 hold, then the exponentially weighted moving average periodogram in (3) approximates the received PSD at any point $\boldsymbol{x}_r$ and frequency f as*

$$\hat{\Phi}_{r,N}^{i,\beta}(f) = \Phi_r(f) + e_{r,N}^{i,\beta}(f) \tag{4}$$

with (asymptotic) variance bounded as follows: $\lim_{N\longrightarrow\infty} \mathrm{Var}[e_{r,N}^{i,\beta}(f)] \leq 3/2(1-\beta)\Phi_r^2(f)$.

The bound on the variance of the estimate in (4) reflects the trade-off present in selecting β between tracking ability, which increases with a shorter memory, and estimation accuracy which improves with a wider window.

III. Cooperative Sparse PSD Estimation

Based on (2) and (4), this section introduces an estimator of $\boldsymbol{\theta}$ using $\hat{\Phi}_{r,N}^{i,\beta}(f_k)$ at frequencies $\{f_k = 2\pi/Nk\}_{k=0}^{N-1}$ collected by a set of N_r cooperating CRs. Instead of LS applied to the linear model in (2), the criterion for estimating $\boldsymbol{\theta}$ (and thus the noise-free PSD anywhere) will exploit two forms of sparsity tacitly present.

Sparsity in $\boldsymbol{\theta}$ is manifested because the linear model in (2) is parsimonious both in frequency as well as in space. Indeed, relative to the possibly huge (over say 10 GHz) system bandwidth, individual transmissions typically occupy a small fraction (in the order of MHz). Likewise, active transmitters will be present only at a small fraction of the candidate source locations $\boldsymbol{x}_s$. If no source is active at point $\boldsymbol{x}_s$, then $\theta_{s\nu} = 0$ for all ν. But even when a transmitter is present at $\boldsymbol{x}_s$, the entries $\theta_{s\nu} = 0$ for all the bases $b_\nu(f)$ with support outside the bandwidth occupied by this radio. The net effect to be exploited when estimating $\boldsymbol{\theta}$ is that only a few (but unknown) entries of $\boldsymbol{\theta}$ are nonzero. Identifying those entries (a.k.a. support of $\boldsymbol{\theta}$) will reveal the positions of active transmitters and their transmission bands.

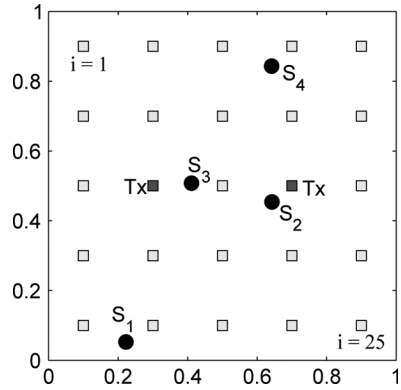

Fig. 2. Virtual network grid with $N_s = 25$ candidate locations, two transmitting sources, and $N_r = 4$ receiving CRs.

Substituting (2) into (4) yields $\hat{\Phi}_{r,N}^{i,\beta}(f_k) = \boldsymbol{b}_r^T(f_k)\boldsymbol{\theta} + \sigma_r^2 + e_{r,N}^{i,\beta}(f_k)$ per frequency f_k. Letting $\boldsymbol{\varphi}_r$ denote the $N \times 1$ vector with entries $\{\hat{\Phi}_{r,N}^{i,\beta}(f_k)\}_{k=0}^{N-1}$, and defining $\boldsymbol{e}_r$ likewise, we arrive at the local vector-matrix model

$$\boldsymbol{\varphi}_r = \boldsymbol{B}_r\boldsymbol{\theta} + \sigma_r^2\mathbf{1} + \boldsymbol{e}_r, \quad r = 1, 2, \ldots, N_r \tag{5}$$

where matrix $\boldsymbol{B}_r$ is formed to have rows $\boldsymbol{b}_r^T(f_k)$, and $\mathbf{1}$ denotes the $N \times 1$ vector of all ones.

The key enabler of cooperative PSD sensing is that $\boldsymbol{\theta}$ is common to all N_r receiving CRs. This allows estimation of $\boldsymbol{\theta}$ in the linear regression model (5) using the non-negative (NN), and thus non-linear, LS criterion

$$\min_{\boldsymbol{\theta}\geq\mathbf{0},\sigma_r^2\geq 0} \sum_{r=1}^{N_r} \left\| \boldsymbol{\varphi}_r - \boldsymbol{B}_r\boldsymbol{\theta} - \sigma_r^2\mathbf{1} \right\|^2 \tag{6}$$

where the non-negativity constraints are naturally imposed to prevent negative PSD estimates.

With position vectors $\boldsymbol{x}_s$ (and/or $\boldsymbol{x}_r$) unknown, even the model in (2) is nonlinear and the NN-LS optimization in (6) is rendered non-convex with multiple local minima. To bypass this challenge, the idea here relies on a virtual grid of *candidate* source locations depicted in Fig. 2. Vectors $\boldsymbol{x}_s$ in Fig. 2 no longer describe the actual positions of e.g., primary users but grid points with *known* spatial coordinates, where transmitting or receiving radios could be present. This virtual grid model was introduced in our preliminary CR sensing work [1], and also independently in [6] for the purpose of target localization. It removes the nonlinearity from the model which renders the estimation problem convex at the price of increasing the number of unknowns.

Aided by this virtual grid, one solution of (6) with unknown position vectors is possible via exhaustive search, as follows: Assume that only one transmitter is present and for each candidate location on the grid, estimate $\boldsymbol{\theta}$ using (6). Subsequently, assume that two transmitters are present and for each pair of candidate locations on the grid, obtain estimates of $\boldsymbol{\theta}$ via (6); and so on, until exhausting all grid points. Comparing the resultant LS errors and taking into account the model complexity (number of unknowns as in e.g., Akaike's information theoretic

Authorized licensed use limited to: University of Minnesota. Downloaded on February 8, 2010 at 19:22 from IEEE Xplore. Restrictions apply.

criterion) it is possible to procure the unknown parameters of the parsimonious basis expansion model.

The exhaustive search is clearly undesirable because it incurs combinatorial complexity in the number of grid points. Recent results in the area of compressed sensing on the other hand, see e.g., [5], [7] and [20], prompted us to avoid this search through the use of convex reformulations of (6) that are particularly suitable when the vector of unknowns is sparse, and the locations of the nonzero entries in $\boldsymbol{\theta}$ are unknown.

These methods share the idea of minimizing the ℓ_1 norm of the unknown vector in order to exploit the sparsity present. In particular, the least-absolute shrinkage and selection operator (Lasso) [20], a.k.a. de-noising basis pursuit [7], amounts to augmenting (6) with the ℓ_1 norm $\|\boldsymbol{\theta}\|_1 := \sum_{s=1}^{N_s}\sum_{\nu=1}^{N_b}|\theta_{s\nu}|$ weighted by a tuning parameter λ. The present setup entails the following NN-Lasso criterion

$$\min_{\boldsymbol{\theta}\geq\mathbf{0},\sigma_r^2\geq 0}\sum_{r=1}^{N_r}\left\|\boldsymbol{\varphi}_r-\boldsymbol{B}_r\boldsymbol{\theta}-\sigma_r^2\mathbf{1}\right\|^2+\lambda\mathbf{1}^T\boldsymbol{\theta} \tag{7}$$

modified by the PSD-imposed non-negativity constraints under which $\|\boldsymbol{\theta}\|_1=\mathbf{1}^T\boldsymbol{\theta}$.

How λ is selected trades off lower LS error for higher degree of sparsity in the solution. Indeed, setting $\lambda=0$ yields the NN-LS solution, while increasing λ pushes the solution towards the origin. The choice of λ for the standard Lasso can be found in [7], and for the NN-Lasso considered here in Appendix B. In this tradeoff, model over-fitting produces spurious coefficients at locations where no transmitter is present. This in turn compromises the generalization capability of the PSD model by rendering the resultant expansion inaccurate at arbitrary locations different from the positions of receiving CRs. Recovering the sparsity via Lasso will reveal the location of transmitters and their bands, and judicious selection of λ will minimize spatio-spectral leakage to spurious locations.

If the data $\{\boldsymbol{\varphi}_r\}_{r=1}^{N_r}$ from all N_r receiving CRs are available to a *central* processing unit, the minimization in (7) can be readily carried out using quadratic or second-order cone programming (SOCP); see, e.g., [3] and [19].

In the ensuing section, a distributed solver of (7) will be sought in lieu of a central unit—a task of paramount interest especially for cognitive sensing tasks pertaining to mobile ad-hoc CR networks. But before pursuing this goal, a remark is due on the consistency of the estimator $\hat{\boldsymbol{\theta}}$ obtained by solving (7).

Remark 1: Similar to LS, the identifiability and mean-square sense consistency of Lasso estimators depend on the rank properties of the overall regression matrix $\boldsymbol{B}:=\left[\boldsymbol{B}_1^T\ldots\boldsymbol{B}_{N_r}^T\right]^T$. Necessary and sufficient conditions for the consistency of the support and nonzero values of $\hat{\boldsymbol{\theta}}$ are available but incur combinatorial complexity to check beforehand [23], [24]. Alternatively, LS-weighted versions of Lasso estimators can be provably consistent for properly chosen values of λ [24]. Albeit related, establishing these conditions for NN-Lasso is challenging and goes beyond the scope of this paper. Extensive simulations however, suggest that NN-Lasso estimators are always consistent when periodogram data are collected from a number of receiving radios that sufficiently exceeds the number of transmitting radios. Intuition from source localization indicates that one must have $N_r\geq N_s$, but in the present setup additional factors affect consistency. Those include the channel gains, the chosen bases, and the resolution of the virtual grid selected.

IV. Distributed Lasso

Solutions of (7) yield $\hat{\boldsymbol{\theta}}$ estimates that enable estimation at arbitrary receiving points in space of the aggregate PSD comprising the superposition of signals emitted from transmitting radios positioned at unknown locations. However, the approach developed so far requires availability of the data $\{\boldsymbol{\varphi}_r\}_{r=1}^{N_r}$ at a designated central unit. The goal of this section is to develop a *distributed* solution of (7), implementable through cooperating CRs that exchange messages with one-hop neighbors over a dedicated control channel. The novel distributed Lasso (D-Lasso) scheme will be developed in two forms: off-line for batch operation and online for tracking operation.

A. Batch D-Lasso

Using the identity $x=(1/N_r)\sum_{r=1}^{N_r}x$, consider rewriting (7) as

$$\min_{\boldsymbol{\theta}\geq\mathbf{0},\sigma_r^2\geq 0}\sum_{r=1}^{N_r}\left\|\boldsymbol{\varphi}_r-\boldsymbol{B}_r\boldsymbol{\theta}-\sigma_r^2\mathbf{1}\right\|^2+\frac{\lambda}{N_r}\sum_{r=1}^{N_r}\mathbf{1}^T\boldsymbol{\theta}. \tag{8}$$

Note that data vector $\boldsymbol{\varphi}_r$ is available locally at CR r and matrix $\boldsymbol{B}_r$ depends only on γ_{sr}, which is also known locally because receiver r knows $\boldsymbol{x}_r$ and for each candidate source position, points $\boldsymbol{x}_s$ on the grid are known. This means that knowing N_r and λ locally, the data required for the problem are readily distributed.

However, the challenge arises because $\boldsymbol{\theta}$ is a global vector common to all radios, the very fact that enabled cooperation among radios as mentioned in Section III. Fortunately, it is possible to overcome this impasse by applying recent distributed optimization approaches based on consensus. Inspired by [18], the idea is to define local copies $\boldsymbol{\theta}_r$ of $\boldsymbol{\theta}$ and constrain them to coincide within the set $\mathcal{N}_r$ of single-hop neighboring CRs corresponding to each receiver r. To this end, extra constraints are added in (8) to arrive at

$$\begin{aligned}\min_{\boldsymbol{\theta}_r\geq\mathbf{0},\sigma_r^2\geq 0}\quad&\sum_{r=1}^{N_r}\left\|\boldsymbol{\varphi}_r-\boldsymbol{B}_r\boldsymbol{\theta}_r-\sigma_r^2\mathbf{1}\right\|^2+\frac{\lambda}{N_r}\sum_{r=1}^{N_r}\mathbf{1}^T\boldsymbol{\theta}_r\\ \text{s. to}\quad&\boldsymbol{\theta}_r=\boldsymbol{\theta}_\rho\quad\forall\rho\in\mathcal{N}_r.\end{aligned} \tag{9}$$

Problems (9) and (8) are equivalent provided that radios are connected in the sense that:

Assumption 3: *There always exists a (possibly multi-hop) path linking any two nodes of the connected CR network.*

Connectivity in Assumption 3 need not be strong, meaning that a CR must have just enough power to reach single-hop neighbors but not all nodes in the CR network. Nonetheless, Assumption 3 ensures that $\boldsymbol{\theta}_r=\boldsymbol{\theta}_\rho$ for all $r,\rho=1,2,\ldots,N_r$. In words, this means that not only the local variables coincide per neighborhood but across the entire network. With a common $\boldsymbol{\theta}$ replacing each $\boldsymbol{\theta}_r$ in (9) one readily arrives at (8); see also [18].

Although problems (8) and (9) are equivalent, only the latter turns out to be amenable to distributed implementation. Deferring proofs to Appendix C, the processing steps needed for the minimization in (9) are carried out in neighborhoods by iteratively solving (j denotes iteration index)

$$(\hat{\boldsymbol{\theta}}_r(j+1),\hat{\sigma}_r^2(j+1))=\arg\min_{\boldsymbol{\theta}_r\geq\mathbf{0},\sigma_r^2\geq 0}\mathcal{L}_\alpha(\boldsymbol{\theta}_r,\sigma_r^2;\cdot)$$

Authorized licensed use limited to: University of Minnesota. Downloaded on February 8, 2010 at 19:22 from IEEE Xplore. Restrictions apply.

$$\mathcal{L}_\alpha(\boldsymbol{\theta}_r,\sigma_r^2;\boldsymbol{\varphi}_r;\boldsymbol{q}_r(j);\{\hat{\boldsymbol{\theta}}_\rho(j)\}) := \left\|\boldsymbol{\varphi}_r - B_r\boldsymbol{\theta}_r - \sigma_r^2\mathbf{1}\right\|^2 + \frac{\lambda}{N_r}\mathbf{1}^T\boldsymbol{\theta}_r + 2\boldsymbol{q}_r^T(j)\boldsymbol{\theta}_r + \alpha\sum_{\rho\in\mathcal{N}_r}\left\|\boldsymbol{\theta}_r - \frac{1}{2}\left(\hat{\boldsymbol{\theta}}_r(j)+\hat{\boldsymbol{\theta}}_\rho(j)\right)\right\|^2 \quad (10)$$

and iteratively updating for $r = 1,2,\ldots,N_r$ the vector variables ($|\mathcal{N}_r|$ denotes the cardinality of $\mathcal{N}_r$)

$$\boldsymbol{q}_r(j+1) = \boldsymbol{q}_r(j) + \frac{\alpha}{2}\left[|\mathcal{N}_r|\hat{\boldsymbol{\theta}}_r(j+1) - \sum_{\rho\in\mathcal{N}_r}\hat{\boldsymbol{\theta}}_\rho(j+1)\right] \quad (11)$$

which represent the "price" of constraint violation, and are expressed analytically as the superposition of Lagrange multipliers associated with the consensus constraints (see Appendix C for details).

Iterations (10) and (11) constitute our batch D-Lasso scheme, which is tabulated as Algorithm 1. All CRs keep track of the local estimates $\hat{\boldsymbol{\theta}}_r(j)$ and $\sigma_r^2(j)$ along with the local $N_r \times 1$ price vector $\boldsymbol{q}_r(j)$. At the beginning of the j^{th} iteration, the CR r has collected $\hat{\boldsymbol{\theta}}_\rho(j)$ estimates from its neighbors $\rho \in \mathcal{N}_r$, and has locally available a vector of prices $\boldsymbol{q}_r(j)$ from the previous iteration. It proceeds by utilizing $\hat{\boldsymbol{\theta}}_\rho(j)$ and $\boldsymbol{q}_r(j)$ as parameters in the quadratic optimization problem (10) that allows it to obtain the updated local estimate $\hat{\boldsymbol{\theta}}_r(j+1)$ along with an updated estimate of the local noise power $\hat{\sigma}_r^2(j+1)$. Then the exchange phase of iteration j takes place in which the CR r sends its updated estimate $\hat{\boldsymbol{\theta}}_r(j+1)$ to all its neighbors $\rho \in \mathcal{N}_r$, and receives their local copies $\hat{\boldsymbol{\theta}}_\rho(j+1)$ as well. These new estimates are used by the CR r to adjust its local price vector $\boldsymbol{q}_r(j+1)$ via (11). Intuitively, the local copies will percolate across neighborhoods after several iterations, and the minimization of the last quadratic term in (10) together with the price update will ensure consensus across the network at equilibrium. Problem (10) is again solvable via quadratic programming or SOCP algorithms and incurs comparable complexity with the centralized solver on (7), but of course without the need for a central processing unit. The main analytical result of this section pertains to the convergence of D-Lasso, which is summarized in the next proposition proved in Appendix C.

Algorithm 1: Consensus-based D-Lasso

CR r initializes $\hat{\boldsymbol{\theta}}_r(0) = \mathbf{0}$, $\hat{\sigma}_r^2(0) = 0$, and $\boldsymbol{q}_r(0) = \mathbf{0}$, and locally runs

for $j = 0,1,\ldots$ **do**

S1. Update $\hat{\boldsymbol{\theta}}_r(j+1)$ and $\hat{\sigma}_r^2(j+1)$ via (10).
S2. Exchange $\hat{\boldsymbol{\theta}}_r(j+1)$ with all neighbors in $\mathcal{N}_r$.
S3. Update $\boldsymbol{q}_r(j+1)$ via (11).

end for

Proposition 2: *Under Assumption 3 and with local communications among single-hop neighbors, the iterates $\hat{\boldsymbol{\theta}}_r(j)$ in Algorithm 1 converge to the centralized solution of (7) for any selection of the constant step-size $\alpha > 0$.*

The D-Lasso algorithm exhibits a number of desirable features. As stated in Proposition 2, it converges for any constant step-size α to the centralized estimator solution of (7). It is fully distributed in the sense that each sensing CR only interacts with its neighbors. This makes it implementable in an ad-hoc network and also facilitates the incorporation of new CR users. Only one price is updated per node, even if there are several constraints (one per neighbor). Furthermore, these prices do not need to be exchanged. The only variables to be communicated are the local $\hat{\boldsymbol{\theta}}_r(j)$ estimates; and these are *highly sparse*, which translates to reduced overhead for cooperation. On the other hand, D-Lasso requires knowledge of the number of nodes N_r and the global parameter λ, which according to Appendix B, is chosen as $\lambda_{\max} = \max_{r=1,2,\ldots,N_r}\lambda_r$, with λ_r available locally.

An extra consensus protocol is thus needed to consent on $\lambda_{\max}$ across the network. This is relatively simple and also possible to perform in a distributed fashion using the steps S1 and S2 of Algorithm 2. Convergence of Algorithm 2 is summarized in the following proposition.

Algorithm 2 : Consensus on λ

CR r initializes with $\lambda_r^{(0)} = \lambda_r$ and locally runs

for $j = 0,1,\ldots$ **do**

S1. Exchange $\lambda_r^{(j-1)}$ with all neighbors in $\mathcal{N}_r$.
S2. Update $\lambda_r^{(j)} = \max_{\rho\in\mathcal{N}_r}\lambda_\rho^{(j-1)}$

end for

Proposition 3: *Under Assumption 3 and with local communications among one-hop neighbors, Algorithm 2 drives $\lambda_r(j)$ to $\lambda_{\max}$ after D iterations, where $D \leq N_r$ denotes the diameter of the CR network graph.*

Proof: Suppose w.l.o.g. that node r^* initializes with the maximum, that is $\lambda_{r^*}(0) = \lambda_{\max}$. Then at the j^{th} iteration, $\lambda_r = \lambda_{\max}$ for all nodes r located at most j hops away from r^*. Therefore, at most after D iterations $\lambda_{\max}$ reaches all CRs at distance D from the receiver r^*, and thus the information about the maximum percolates across the entire network. ■

Remark 2: A related consensus-based approach was also developed in [18] for distributed estimation based on general criteria using the alternating direction method of multipliers (ADMoM) [2]. However, to form the ADMoM cost in [18], a subset of sensors called bridge sensors is required. In turn, an algorithm to find the bridge sensor set is necessary to run whenever sensors become inactive or new sensors are activated. Setting all sensors as bridges is possible but the communication overhead is considerably increased. Compared to [18], D-Lasso does not require such a bridge sensor set, and in this sense it offers a fully distributed, low-overhead approach.

B. Online D-Lasso

The real-time requirements on the sensing radios and the convenience of an estimator that adapts to changes in the PSD are the motivating reasons behind the online D-Lasso algorithm of this subsection. In the off-line iteration (10), the periodogram observations pertain to those collected up to block i and the same $\boldsymbol{\varphi}_r$ is used to update $\hat{\boldsymbol{\theta}}_r(j)$ for the entire run of Algorithm 1. An online version of D-Lasso results if successive iterations are carried out *across* coherence blocks, and the current data $\boldsymbol{\varphi}_r(i)$ acquired during block i is incorporated in the estimator. This is possible after substituting $\boldsymbol{\varphi}_r(i)$ for $\boldsymbol{\varphi}_r$ in (10), and merging the iteration with coherence block indexes into a single index i. In addition, since λ depends on $\{\boldsymbol{\varphi}_r(i)\}_{r=1}^{N_r}$ (as detailed in Appendix B), the penalty parameter will in principle

Authorized licensed use limited to: University of Minnesota. Downloaded on February 8, 2010 at 19:22 from IEEE Xplore. Restrictions apply.

become block dependent, namely $\lambda(i)$. Taking these modifications into account, (10) and (11) are replaced by

$$(\hat{\boldsymbol{\theta}}_r(i+1), \hat{\sigma}_r^2(i+1)) = \arg \min_{\boldsymbol{\theta}_r \geq \mathbf{0}, \sigma_r^2 \geq 0} \mathcal{L}_\alpha(\boldsymbol{\theta}_r, \sigma_r^2; \cdot) \tag{12}$$

where

$$\mathcal{L}_\alpha(\boldsymbol{\theta}_r, \sigma_r^2; \boldsymbol{\varphi}_r(i); \boldsymbol{q}_r(i); \{\hat{\boldsymbol{\theta}}_\rho(i)\}) := \left\| \boldsymbol{\varphi}_r(i) - \boldsymbol{B}_r \boldsymbol{\theta}_r - \sigma_r^2 \mathbf{1} \right\|^2 + \frac{\lambda(i)}{N_r} \mathbf{1}^T \boldsymbol{\theta}_r + 2\boldsymbol{q}_r^T(i)\boldsymbol{\theta}_r + \alpha \sum_{\rho \in \mathcal{N}_r} \left\| \boldsymbol{\theta}_r - \frac{1}{2}\left(\hat{\boldsymbol{\theta}}_r(i) + \hat{\boldsymbol{\theta}}_\rho(i)\right) \right\|^2$$

and

$$\boldsymbol{q}_r(i+1) = \boldsymbol{q}_r(i) + \frac{\alpha}{2}\left[|\mathcal{N}_r| \hat{\boldsymbol{\theta}}_r(i+1) - \sum_{\rho \in \mathcal{N}_r} \hat{\boldsymbol{\theta}}_\rho(i+1) \right]. \tag{13}$$

The resulting distributed online iteration (online D-Lasso) is summarized next as Algorithm 3. In a stationary environment, the iterates $\hat{\boldsymbol{\theta}}(i)$ of (12) will converge to the solution of (9) even with constant $\lambda(i) = \lambda$, $\forall i$. Likewise, if the transmit-PSDs are slowly varying $\boldsymbol{\varphi}_r(i)$ will closely approach the true receive PSD, which is approximately time-invariant. In this case too, running (12) and (13) with λ remaining constant across several blocks will endow D-Lasso with tracking capability while minimally affecting estimation accuracy and avoiding the updates of λ per coherence block. Rigorous convergence analysis of Algorithm 3 goes beyond the scope of this work. Nevertheless, it will be illustrated by simulations.

Algorithm 3: Online-D-Lasso

CR r initializes $\hat{\boldsymbol{\theta}}_r(0) = \mathbf{0}$, $\hat{\sigma}_r^2(0) = 0$ and $\boldsymbol{q}_r(0) = \mathbf{0}$, and locally runs

for $i = 0, 1, \ldots$ **do**

S1. Update $\hat{\boldsymbol{\theta}}_r(i+1)$ and $\hat{\sigma}_r^2(i+1)$ via (12).
S2. Exchange $\hat{\boldsymbol{\theta}}_r(i+1)$ with all neighbors in $\mathcal{N}_r$.
S3. Update $\boldsymbol{q}_r(i+1)$ via (13).

end for

In a nutshell, this section introduced distributed Lasso algorithms for batch and adaptive operation. The contribution is twofold: a *distributed* solution to variable selection and compressive sampling approaches; and a rather neat application to the emerging area of cooperative sensing the ambient PSD map, which is instrumental for assessing the "interference temperature" in mobile ad-hoc CR networks.

V. Performance Analysis

The present section will analyze the large-sample mean-square error (MSE) performance of the Lasso parameter estimator $\hat{\boldsymbol{\theta}}$ obtained as the solution of the constrained minimization problem in (7). The MSE of $\hat{\boldsymbol{\theta}}$ will be subsequently used for MSE analysis of the *noise-free* PSD estimator $\hat{\Phi}_r(f) = \boldsymbol{b}_r^T(f)\hat{\boldsymbol{\theta}}$ at any receiving point $\boldsymbol{x}_r$ and any frequency f. As is usually the case with nonlinear estimators that are not expressible in closed form, the analysis will be valid asymptotically; that is, for a sufficiently large number of grid points N_s, frequencies N, and/or receiving points N_r. Two types of errors will be considered: those due the finite grid approximation, and those emerging from the estimation process.

A. Grid-Induced Errors

The gains γ_{sr} present in the regression matrix $\boldsymbol{B}_r$ were so far assumed known because the unknown source locations were taken to lie at known candidate positions $\boldsymbol{x}_s$ on the vertices of the given grid. If instead the sources lie in the interior of a grid cell at locations $\check{\boldsymbol{x}}_s \neq \boldsymbol{x}_s$, then the *true* noise-free PSD is $\check{\Phi}_r(f) = \check{\boldsymbol{b}}_r^T(f)\boldsymbol{\theta} = \sum_{s=1}^{N_s}\sum_{\nu=1}^{N_b} \check{\gamma}_{sr}\theta_{s\nu}b_\nu(f)$. Relative to the estimated *approximate* PSD with candidate sources assumed to lie on the vertices of the grid, which is given by $\hat{\Phi}_r(f) = \sum_{s=1}^{N_s}\sum_{\nu=1}^{N_b} \gamma_{sr}\hat{\theta}_{s\nu}b_\nu(f)$, the error can be expressed as

$$\hat{\Phi}_r(f) - \check{\Phi}_r(f) = \sum_{s=1}^{N_s}\gamma_{sr}\sum_{\nu=1}^{N_b} b_\nu(f)(\hat{\theta}_{s\nu} - \theta_{s\nu}) + \sum_{s=1}^{N_s}(\gamma_{sr} - \check{\gamma}_{sr})\sum_{\nu=1}^{N_b} b_\nu(f)\theta_{s\nu} \tag{14}$$

where in deriving (14) we added and subtracted $\gamma_{sr}\theta_{s\nu}$. Clearly, the first double sum in the right hand side (r.h.s.) of (14) corresponds to the estimation errors, while the second one is due to the finite grid approximation effects.

It will be argued that for an increasingly dense grid with $N_s \gg 1$ candidate locations, this second double sum is negligible. Indeed, if $\gamma(\cdot)$ is sufficiently smooth, a first-order Taylor's expansion around the point $\boldsymbol{x}_s$ yields

$$\gamma_{sr} - \check{\gamma}_{sr} := \gamma(\|\boldsymbol{x}_s - \boldsymbol{x}_r\|) - \gamma(\|\check{\boldsymbol{x}}_s - \boldsymbol{x}_r\|) \approx \frac{\dot{\gamma}(\|\check{\boldsymbol{x}}_s - \boldsymbol{x}_r\|)(\boldsymbol{x}_s - \boldsymbol{x}_r)^T(\boldsymbol{x}_s - \check{\boldsymbol{x}}_s)}{\|\boldsymbol{x}_s - \boldsymbol{x}_r\|}$$

where $\dot{\gamma}$ denotes the derivative of γ.

As the grid becomes increasingly dense (that is the candidate source locations $N_s \longrightarrow \infty$), it holds that $\boldsymbol{x}_s \longrightarrow \check{\boldsymbol{x}}_s$, which in turn implies based on the last approximation that $\gamma_{sr} - \check{\gamma}_{sr} \longrightarrow 0$. This formalizes the assertion that the second double sum in the r.h.s. of (14) will diminish gracefully as the density of the virtual grid increases.

B. Estimation Errors

It will be assumed henceforth, that the grid is sufficiently dense to induce no errors. Thus, skipping the second double sum in the r.h.s. of (14), and rewriting the first one in a vector form leads to

$$\hat{\Phi}_r(f) - \check{\Phi}_r(f) = \boldsymbol{b}_r^T(f)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}). \tag{15}$$

The MSE of the latter is expressed as

$$E\left[\left(\hat{\Phi}_r(f) - \check{\Phi}_r(f)\right)^2\right] = \boldsymbol{b}_r^T(f)\boldsymbol{C}_{\hat{\theta}}\boldsymbol{b}_r(f) \tag{16}$$

where $\boldsymbol{C}_{\hat{\theta}}$ denotes the covariance matrix of the Lasso parameter estimator in (7). Equation (16) shows that the MSE of the PSD estimator is readily obtained after $\boldsymbol{C}_{\hat{\theta}}$ is found.

Aiming at a tractable expression of $\boldsymbol{C}_{\hat{\theta}}$, consider concatenating the $N \times N_s N_b$ matrices $\boldsymbol{B}_r$ for $r = 1, 2, \ldots, N_r$ in (5) to form an $N_r N \times N_s N_b$ matrix $\boldsymbol{B}$; and similarly for the $N \times 1$

Authorized licensed use limited to: University of Minnesota. Downloaded on February 8, 2010 at 19:22 from IEEE Xplore. Restrictions apply.

vectors $\boldsymbol{\varphi}_r$ to form the $N_r N \times 1$ vector $\boldsymbol{\varphi}$. Likewise, collect variables σ_r^2 into the $N_r \times 1$ vector $\boldsymbol{\sigma}^2 := (\sigma_1^2, \ldots, \sigma_{N_r}^2)^T$. Since each variable σ_r^2 is repeated N times (cf. the vector $\mathbf{1}$ in (5)), it is convenient to introduce the repetition matrix $\boldsymbol{R} := \boldsymbol{I}_{N_r \times N_r} \otimes \mathbf{1}$, where $\boldsymbol{I}_{N_r \times N_r}$ denotes the identity matrix of range N_r and $\otimes$ the Kronecker product. With these notational conventions, the system of N_r linear equations in (5) can be compactly written as

$$\boldsymbol{\varphi} = \boldsymbol{B}\boldsymbol{\theta} + \boldsymbol{R}\boldsymbol{\sigma}^2 + \boldsymbol{e} \tag{17}$$

where the error vector $\boldsymbol{e}$ is zero-mean with covariance matrix $\boldsymbol{C}_e$. The mean of $\boldsymbol{e}$ is indeed zero provided that N_s is large enough and grid-induced errors are thus negligible.

For the linear regression model in (17), we wish to analyze the MSE performance of the estimator $\hat{\boldsymbol{\theta}}$ obtained as the solution of (7). Because the minimization problem in (7) is constrained and the resultant estimator is nonlinear, developing an expression for $\boldsymbol{C}_{\hat{\theta}}$ can become tractable if one focuses on the nonzero entries of $\hat{\boldsymbol{\theta}}$, which are denoted by the vector $\hat{\boldsymbol{\theta}}_1$. To this end, let $\boldsymbol{\theta}_1$ denote the entries of $\boldsymbol{\theta}$ corresponding to $\hat{\boldsymbol{\theta}}_1$, and $\boldsymbol{B}_1$ the corresponding sub-matrix of $\boldsymbol{B}$. To proceed, the following assumption is necessary.

Assumption 4: *The support of $\boldsymbol{\theta}$ is included in the support of the estimate $\hat{\boldsymbol{\theta}}$.*

As the number of data NN_r (size of $\boldsymbol{\varphi}$) in (17) grows large, consistency of the Lasso estimator ensures that the supports of $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$ coincide; and thus, Assumption 4 holds true asymptotically. But even for finite NN_r values, Assumption 4 is valid so long as the Lasso criterion in (7) does not produce false negatives, but only false positives in selecting the entries of $\hat{\boldsymbol{\theta}}$ to be set to zero. Note that even the ordinary linear LS estimator, which does not account for the sparsity present in $\boldsymbol{\theta}$, satisfies Assumption 4 too.

Under Assumption 4, the error in estimating the nonzero entries of the Lasso estimator can be expressed as summarized in the next lemma (see Appendix D for the proof).

Lemma 1: *If $\hat{\boldsymbol{\theta}}$ obeys Assumption 4, the estimation error of its nonzero entries can be expressed as*

$$\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1 = \left(\boldsymbol{B}_1^T \boldsymbol{D} \boldsymbol{B}_1\right)^{-1} \left(\boldsymbol{B}_1^T \boldsymbol{D} \boldsymbol{e} - \frac{1}{2}\boldsymbol{\lambda}_1\right) \tag{18}$$

where $\boldsymbol{\lambda}_1 := \lambda \mathbf{1}$, and $\boldsymbol{D} := \boldsymbol{I} - \boldsymbol{R}\boldsymbol{R}^T/N$.

The next step is to take expectations of the error vector and its norm in (18), to assess the bias and MSE in estimating the nonzero entries of $\hat{\boldsymbol{\theta}}$. This is challenging though, because the sub-matrix $\boldsymbol{B}_1$ is selected in accordance with the nonzero entries of the random vector $\hat{\boldsymbol{\theta}}$; hence, $\boldsymbol{B}_1$ is itself random. To simplify the analysis, we will approximately treat $\boldsymbol{B}_1$ as deterministic, which basically amounts to considering that the bias, covariance matrix, and MSE of $\hat{\boldsymbol{\theta}}_1$ are largely affected by the number of spurious nonzero entries of $\hat{\boldsymbol{\theta}}$, while the positions those entries occur have minimal effect.

Under this approximation, taking expectations in (18) shows that even when the grid effects are negligible ($\boldsymbol{e}$ is indeed zero mean), the Lasso estimator remains biased. The bias is given by $(1/2)(\boldsymbol{B}_1^T \boldsymbol{D} \boldsymbol{B}_1)^{-1}\boldsymbol{\lambda}_1$. With regards to the covariance matrix of $\hat{\boldsymbol{\theta}}_1$, the following proposition can be readily established from (18).

Proposition 4: *If Assumption 4 holds and $\boldsymbol{e}$ is zero mean, the covariance matrix of $\hat{\boldsymbol{\theta}}_1$ is approximately given by*

$$\boldsymbol{C}_{\hat{\theta}_1} = (\boldsymbol{B}_1^T \boldsymbol{D} \boldsymbol{B}_1)^{-1} \boldsymbol{B}_1^T \boldsymbol{D} \boldsymbol{C}_{\boldsymbol{e}} \boldsymbol{D} \boldsymbol{B}_1 (\boldsymbol{B}_1^T \boldsymbol{D} \boldsymbol{B}_1)^{-1} + \frac{1}{4}(\boldsymbol{B}_1^T \boldsymbol{D} \boldsymbol{B}_1)^{-1} \boldsymbol{\lambda}_1 \boldsymbol{\lambda}_1^T (\boldsymbol{B}_1^T \boldsymbol{D} \boldsymbol{B}_1)^{-1}. \tag{19}$$

The trace of $\boldsymbol{C}_{\hat{\theta}_1}$ in Proposition 4 yields not only the MSE of $\hat{\boldsymbol{\theta}}_1$ but also the MSE of $\hat{\boldsymbol{\theta}}$, since the entries of $\hat{\boldsymbol{\theta}}$ not in $\hat{\boldsymbol{\theta}}_1$ are zero. Formally stated, the following corollary can be easily established.

Corollary 1: *If Assumption 4 holds, the MSE of $\hat{\boldsymbol{\theta}}$ obtained as the solution of (7) is approximately given by*

$$E\left[\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2\right] = \operatorname{trace}\left\{(\boldsymbol{B}_1^T \boldsymbol{D} \boldsymbol{B}_1)^{-2} \boldsymbol{B}_1^T \boldsymbol{D} \boldsymbol{C}_{\boldsymbol{e}} \boldsymbol{D} \boldsymbol{B}_1\right\} + \frac{1}{4}\boldsymbol{\lambda}_1^T (\boldsymbol{B}_1^T \boldsymbol{D} \boldsymbol{B}_1)^{-2} \boldsymbol{\lambda}_1. \tag{20}$$

In order to gain intuition and confidence on the approximate expressions (19) and (20), it is instructive to look at the following special case.

Special Case: Suppose that the parameters σ_r^2 in (7) are known so that the term $\boldsymbol{R}\boldsymbol{\sigma}^2$ can be removed from (17), and hence $\boldsymbol{D} = \boldsymbol{I}$ in (18). Furthermore, suppose that matrix $\boldsymbol{B}$ (and thus $\boldsymbol{B}_1$) is orthonormal, and $\boldsymbol{e}$ is white with $\boldsymbol{C}_e = \sigma_e^2 \boldsymbol{I}$. Since $\boldsymbol{B}_1^T \boldsymbol{B}_1 = \boldsymbol{I}$, it is straightforward to show from (20) that $E[\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2] = (\sigma_{\boldsymbol{e}}^2 + \lambda^2/4)P_1$, where P_1 denotes the size of $\hat{\boldsymbol{\theta}}_1$. In comparison, the MSE of the ordinary LS estimator for the same setup is $\sigma_e^2 P$, where $P = N_s N_b$. While P can be large, P_1 is typically in the order of the length of $\boldsymbol{\theta}_1$ which, given that $\boldsymbol{\theta}$ is sparse, is much smaller than P.

At least this special case confirms that the Lasso yields an "oracle" estimator, which approximately deciphers the support of $\boldsymbol{\theta}$ at the cost of adding bias $(\lambda^2/4)P_1$ relative to the "oracle" LS estimator which knows perfectly the support of $\boldsymbol{\theta}$. Lasso's improved MSE performance over the support-agnostic LS estimator is the result of effectively exploiting the additional information available, namely the sparsity. This allows Lasso to reduce the over-fitting that the LS estimator incurs because it returns non-zero values in all entries of $\hat{\boldsymbol{\theta}}$.

Remark 3: The oracle properties of the Lasso estimator are formally reviewed in [24], where a weighted version of Lasso is proved to ensure asymptotically unbiased, MSE- and support-consistent estimators. The weighted Lasso (termed adaptive Lasso in [24]) has the ℓ_1 norm in (7) weighted by the inverse of the LS estimator entries, and applies readily to the present context as well. The main reason for not adopting it here is simplicity in computation and exposition.

Having available an approximate expression for the covariance matrix of the Lasso estimator, the MSE for the PSD estimator follows immediately after substituting (19) into (16). The resultant MSE is given next as a corollary.

Corollary 2: *If Assumption 4 holds, the MSE of $\hat{\Phi}_r(f)$ at an arbitrary location $\boldsymbol{x}_r$ is approximately given by*

$$E\left[\left(\hat{\Phi}_r(f) - \check{\Phi}_r(f)\right)^2\right] = \boldsymbol{b}_{1,r}^T(f) \boldsymbol{C}_{\hat{\theta}_1} \boldsymbol{b}_{1,r}(f) \tag{21}$$

where $\boldsymbol{b}_{1,r}(f)$ is a sub-vector of $\boldsymbol{b}_r(f)$ obtained by selecting the entries, where $\hat{\boldsymbol{\theta}}$ is strictly positive.

Authorized licensed use limited to: University of Minnesota. Downloaded on February 8, 2010 at 19:22 from IEEE Xplore. Restrictions apply.

The covariance matrix $\boldsymbol{C}_e$ needed to find $\boldsymbol{C}_{\hat{\theta}_1}$ in (21) can be estimated by sample averaging multiple realizations of $\boldsymbol{\varphi}$. If those are not available, a single realization based approximate $\boldsymbol{C}_e$ is suggested in Appendix A [c.f (26) and (27)] as

$$\boldsymbol{C}_e \approx (1-\beta)[\boldsymbol{C}_{e,1} \odot \boldsymbol{M} + \mathrm{diag}(\boldsymbol{C}_{e,1} + \boldsymbol{C}_{e,2})] \qquad (22)$$

where $\boldsymbol{C}_{e,1} := (\boldsymbol{\varphi} - \hat{\boldsymbol{\sigma}} \otimes \mathbf{1})(\boldsymbol{\varphi} - \hat{\boldsymbol{\sigma}} \otimes \mathbf{1})^T$, $\boldsymbol{C}_{e,2} := \boldsymbol{\varphi}\boldsymbol{\varphi}^T$, $\boldsymbol{M} := \boldsymbol{I}_{N_r \times N_r} \otimes (\mathbf{1}\mathbf{1}^T)$, and $\odot$ denotes the Hadamard (entry-wise) product. The rationale behind this selection is that $\boldsymbol{\varphi}$ contains PSD estimates; hence, its entries can be substituted for $\Phi_r(f_k) = \sum_{s=1}^{N_s} \gamma_{sr}\Phi_s(f_k) + \sigma_r^2$ in (26) and (27). Likewise, the entries of $\boldsymbol{\varphi} - \hat{\boldsymbol{\sigma}} \otimes \mathbf{1}$ can replace $\Phi_r(f_k) - \sigma_r^2 = \sum_{s=1}^{N_s} \gamma_{sr}\Phi_s(f_k)$. Note also that the Hadamard product $\boldsymbol{C}_{e,1} \odot \boldsymbol{M}$ sets entries corresponding to different receivers to zero in order to account for the uncorrelated channel gains and thus periodograms received at distinct locations.

VI. Simulated Tests

The simulations described here to validate and compare the algorithms of this paper, are performed with reference to the CR network depicted in Fig. 3 (top). The setup includes $N_r = 4$ CRs that cooperate to estimate the PSD map in space and frequency, generated by two sources located at unknown positions on a grid of $N_s = 25$ candidate locations. The cooperating CRs scan $N = 128$ frequencies from 15 to 30 MHz, and adopt the basis expansion model in (1) over this band comprising $N_b = 16$ rectangles as frequency bases. The average gains of the network links obey an inverse polynomial pathloss model for γ_{sr} with $d_0 = 100$ m and $\alpha = 3$.

For coherence blocks $i = 1, \ldots, 650$, the transmit-PSDs of the sources are spanned by three bases each, as shown in Fig. 3 (bottom). At coherence block $i = 650$, the high-frequency source shuts off, and only one source remains. This corresponds to a true parameter vector $\boldsymbol{\theta}$ of size $N_s N_b = 400$ with only six non-zero entries set to one for $i = 1, \ldots, 650$; and only three non-zero entries remaining equal to one from $i = 651$ to $i = 1000$.

The first test compares the centralized Lasso estimator (7) with the NNLS solution of (6). The batch D-Lasso algorithm is also tested as a solver of (7) using an EWMA snapshot at $i = 500$. In addition, the online D-Lasso iterations are run to check its ability to track the change in the topology after $i = 650$. Subsequently, a test point at location not coinciding with any of the four receiving CR locations, is selected to assess the generalization (prediction) capabilities of the aforementioned PSD estimators. Finally, the performance of the Lasso estimator is studied by comparing Monte Carlo MSE estimates against the approximate MSE expressions of Section V.

The data are generated using the transmit-PSDs described earlier, a Rayleigh channel model with $L = 6$ taps, and additive white Gaussian receiver noise at 0 dB. The channel deviates are generated per coherence block to acquire snapshots of the eight frequency-selective instantaneous channel gains $|H_{sr}(i; f)|^2$ between the two transmitters and the four receiving CRs. With these receive-PSDs per coherence block i, frequency f_k, and CR r, the mean of an exponentially distributed random variable is subsequently used to simulate the (asymptotic) behavior

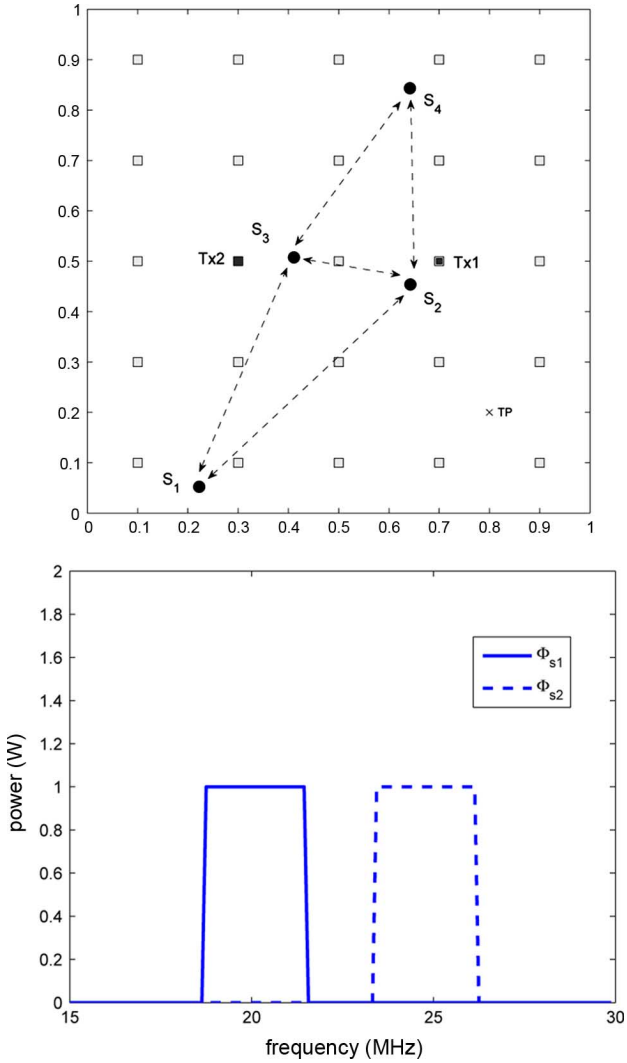

Fig. 3. (top) Simulated CR network with two sources, four receiving CRs, and one test point; (bottom) transmit-PSDs of the two sources.

of the periodogram estimates $\hat{I}_{r,N}^{(i)}$. Those are then averaged across coherence blocks according to (3) with $\beta = 0.99$ (corresponding to a window of 100 coherence blocks). Across coherence blocks $i = 1, \ldots, 1000$, the resulting NN_r frequency samples are collected to form the data vector $\boldsymbol{\varphi}$.

A. Centralized Lasso Algorithm

At each coherence block i from $i = 1$ to $i = 1000$ the batch Lasso estimator of (7) is run assuming that all the data are available at a central unit. The trajectories of successive parameter estimates $\hat{\boldsymbol{\theta}}(i)$ for $i = 1, \ldots, 1000$ are compared against the true vector $\boldsymbol{\theta}$, and the norm of their difference is normalized by $||\boldsymbol{\theta}||$. The evolution of this normalized difference is depicted in Fig. 4 (top) along with that of the NNLS algorithm (6). It is seen that the ℓ_1 norm in (7) is effective in exploiting the sparsity present in the model—the *a priori* information not exploited by the NNLS estimator. The price paid by the sparsity-agnostic NNLS is in excess of 15 dB of error in the steady state.

An alternative visualization of the improvement effected when exploiting sparsity is provided by Fig. 4 (bottom), which

Authorized licensed use limited to: University of Minnesota. Downloaded on February 8, 2010 at 19:22 from IEEE Xplore. Restrictions apply.

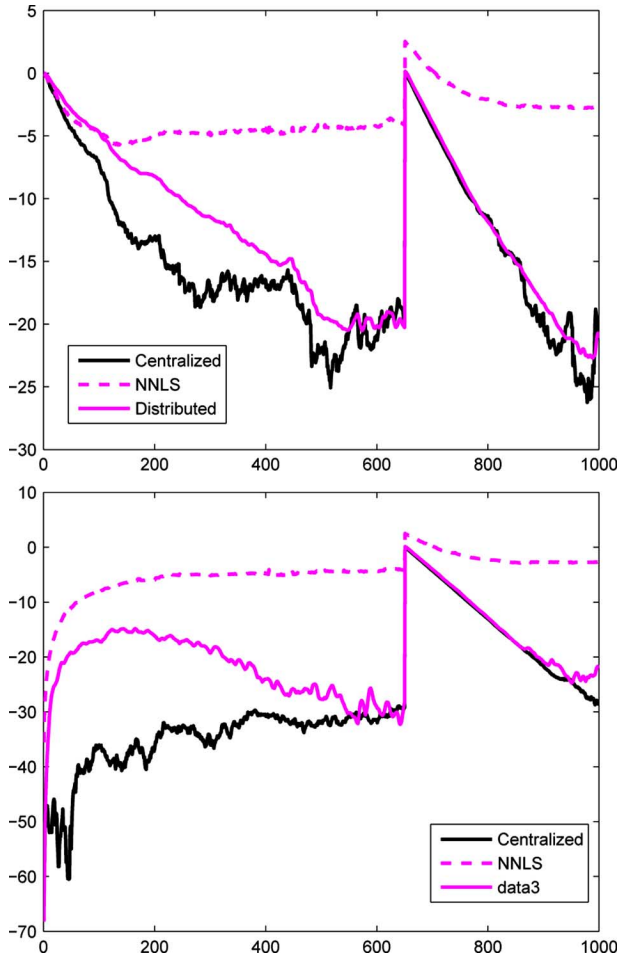

Fig. 4. (top) Comparison of the $||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}||$ errors normalized by $||\boldsymbol{\theta}||$; (bottom) aggregate spurious power of PSD estimates.

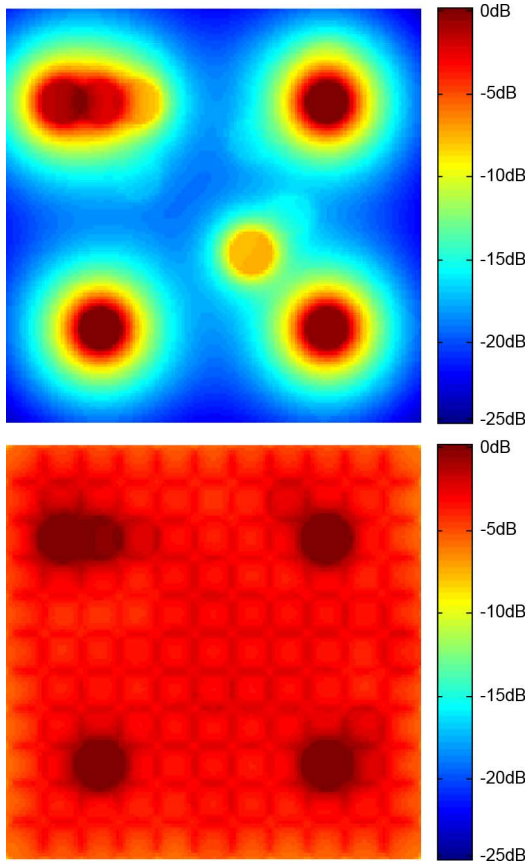

Fig. 5. (top) Estimated power map via Lasso; (bottom) via NNLS. True spatial PSD generated by 5 sources.

compares the aggregate spurious power each estimate places in frequency bands that are actually free, or, in candidate locations where there is no transmitter. The spurious power is compared with the aggregate transmit-power (in dBs), and demonstrates that the centralized Lasso estimator outperforms the NNLS both in localizing the free bands and also in positioning the sources.

When the high-frequency source becomes inactive after $i = 650$, the memory of the EWMA produces traces of the disappearing source in $\boldsymbol{\varphi}$ that require an interval of approximately $T_\beta = 1/(1-\beta)$ coherence blocks to vanish. This inertia manifests itself as a jump in the error and spurious power estimates at $i = 650$. The jump is magnified when normalizing with $||\boldsymbol{\theta}||$, which becomes smaller. On the bright side, the EWMA is responsible for the ability of the online Lasso to overcome this transition, and settle down to a lower error. The linear decaying trend of the error measured in dB reflects the exponential evolution of the EWMA.

To highlight the localization capabilities of Lasso relative to LS, we tested a setup with five sources transmitting over non-overlapping frequency bands, a virtual grid with $N_s = 121$ candidate locations, and $N_r = 64$ sensing CRs at receive-SNR $= -3$ dB. The estimated maps of the spatial PSDs (superimposed over all 5 frequency bands) are plotted using a mesh format for the Lasso [Fig. 5 (top)], and the NNLS [Fig. 5 (bottom)]. Relative to NNLS, the Lasso map is more accurate, separating signal from noise and revealing the position of the five sources (only four sources are "seen" by the NNLS solution).

B. Batch D-Lasso Algorithm

The distributed off-line iterations (10), (11) are tested at $i = 500$. The sensing CRs communicate with their neighbors as depicted in Fig. 3 (top), where CRs 2 and 3 can reach all other radios while CRs 1 and 4 do not communicate with each other. Consensus of the local iterates $\hat{\boldsymbol{\theta}}_r(j)$, $j = 1, \ldots, 1000$ is achieved rapidly. Fig. 6 (top) illustrates that the errors between $\hat{\boldsymbol{\theta}}_1(j)$ and estimates of its neighbors decay to -40 dB after ten iterations. Fig. 6 (bottom) compares the error between the estimate $\hat{\boldsymbol{\theta}}_1(j)$, which is taken as representative of the local off-line iterates, and the centralized estimate $\hat{\boldsymbol{\theta}}(i)$ at $i = 500$, normalized by the norm of the latter, confirming the convergence of Algorithms 1 and 2. The knee in the error plot occurs when the estimates transition from a first phase in which $\boldsymbol{\theta}_r(j)$ grows from zero to the power level being transmitted, to a second phase in which fine tuning takes place.

C. Online D-Lasso Algorithm

When the online iterations (12) and (13) are simulated, the local iterates reach consensus within ten iterations, after which

Authorized licensed use limited to: University of Minnesota. Downloaded on February 8, 2010 at 19:22 from IEEE Xplore. Restrictions apply.

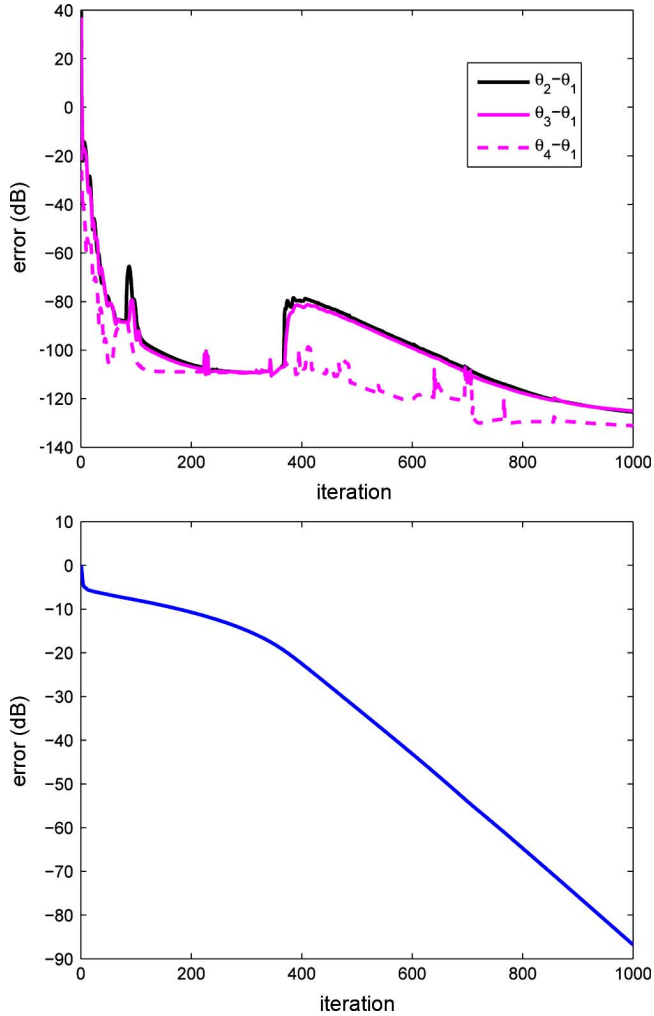

Fig. 6. (top) Batch D-Lasso iterations based on $i = 500$ blocks reach consensus; (bottom) comparison of batch D-Lasso with the centralized Lasso estimate.

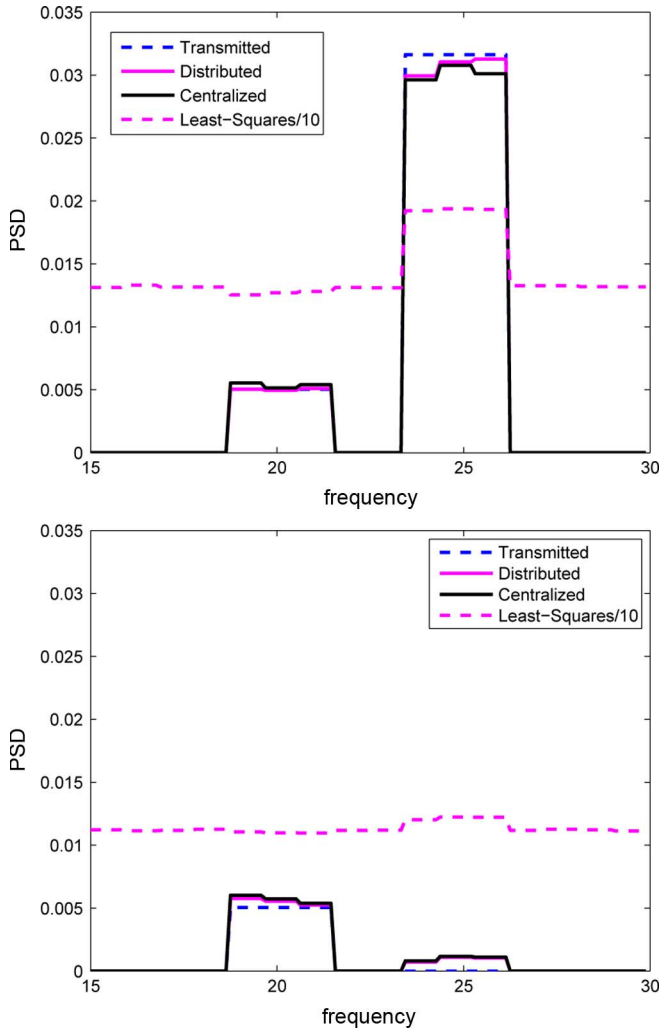

Fig. 7. (top) PSD estimate at the test point after $i = 500$ blocks; (bottom) after $i = 1000$ blocks.

their differences stay below -30 dB in absolute value. Again, $\hat{\boldsymbol{\theta}}_1(i)$ is taken as representative of the four local estimates, and is compared with the "true" sequence $\{\boldsymbol{\theta}(i)\}_{i=1}^{1000}$ adopted for the simulations. The error trajectory and the amount of spurious power estimates are shown in Fig. 4 along with the centralized Lasso estimate. This comparison confirms that the online iterates follow closely the trajectory generated through successive runs of the centralized batch estimator. As in the centralized case, the online D-Lasso algorithm adapts to the changes in the ambient PSD produced at $i = 650$ by the disappearing source after the EWMA adjusts to the updated network topology.

D. Generalization Capability

A test point $\boldsymbol{x}_r$ placed in the area sensed by the four receiving CRs is marked with a cross in Fig. 3 (top). The PSD estimate at the test point is obtained per coherence block using the online D-Lasso parameter estimate $\hat{\boldsymbol{\theta}}_1(i)$, and the adopted model (2). Fig. 7 depicts PSD estimates obtained by the different methods after $i = 500$ and after $i = 1000$ coherence blocks, and compares them against the true PSDs corresponding to the transmit-PSDs generated by the source(s), and the average gains of the channels (from each source to the test point) computed using the pathloss model described earlier. The estimate generated by the NNLS algorithm is shrunk by a factor of 10 and shown together with the centralized and online D-Lasso solutions.

The ability of D-Lasso to exploit sparsity is the main reason it outperforms NNLS in estimating accurately the PSD at the test point. In this experiment, the NNLS solution overfits the data and places power in all candidate locations in order to match the observations as faithfully as possible in the LS sense. In addition, NNLS interprets flat PSDs corresponding to receive noise spectra as transmit-PSDs emitted by sources occupying the bands. With power placed in wrong locations and in free bands, the NNLS estimate produces an erroneous PSD map at locations other than the sensing points. The (D-) Lasso instead succeeds in finding an estimate of the PSD map whose support approximates the support of the PSD generated by the "true" $\boldsymbol{\theta}(i)$. This translates to accurately localizing the transmitting sources and correctly declaring the (un)occupied bands at arbitrary points in space and frequency.

Authorized licensed use limited to: University of Minnesota. Downloaded on February 8, 2010 at 19:22 from IEEE Xplore. Restrictions apply.

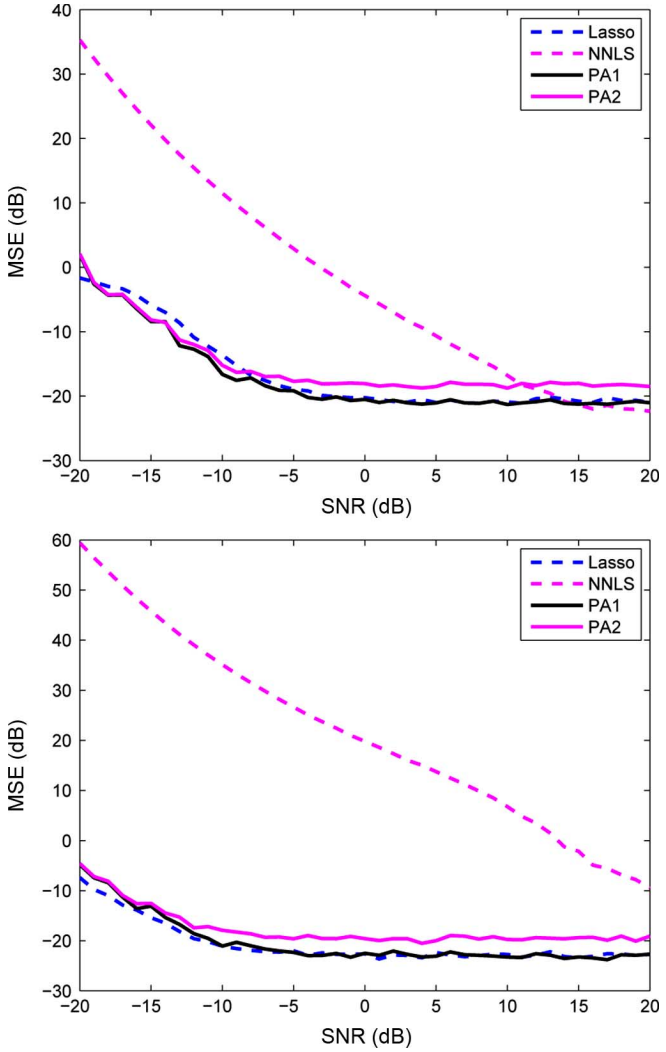

Fig. 8. (top) MSE of the parameter estimator; (bottom) MSE of the PSD estimator $\hat{\Phi}_r(f)$.

E. Performance Analysis

In this test case, the approximate expressions for the MSE of the parameter estimator $\hat{\boldsymbol{\theta}}$ in (20), and the MSE of the PSD estimator $\hat{\Phi}_r(f)$ (at the test point) in (21), are validated. Forty one SNR values (from -20 to 20 dBs) are tested, with all CRs having the same SNR every time, and 200 samples of the data vector $\boldsymbol{\varphi}_r|_{i=500}$ generated per SNR value. The first 100 samples are used to estimate the aforementioned MSEs by averaging over Monte Carlo runs (both estimates are labelled as Lasso in Fig. 8). The Lasso criterion in (7) is used per run with all data available centrally. The second part of the sample is used to estimate $\boldsymbol{C}_e$, which is subsequently used in (20) and (21) to obtain the line labelled as PA1. The line PA2 corresponds to the approximation of $\boldsymbol{C}_e$ in (22). These results are shown in Fig. 8 along with the MSE of the NNLS estimator obtained by averaging over Monte Carlo realizations. All MSEs exhibit a floor at high SNR values. This is because the error in the model is proportional to the PSD square [see (4)] that does not vanish with the background noise. In order to push this error to zero, the EWMA window must be opened but this trades off the ability to track changing environments.

VII. Concluding Remarks

The key challenge in developing cognitive wireless transceivers is enabling them to sense the ambient power spectral density at arbitrary locations in space. The present paper addressed this challenging task through a parsimonious basis expansion model of the PSD in frequency and space. This model reduces the sensing task to estimating a sparse vector of unknown parameters. As a byproduct, sparsity also facilitates localization of transmitting radios even under multipath fading. The associated estimators rely on the Lasso algorithm, which here enforces sparsity in the solution to reveal the position and frequency bands of transmitting radios. Once these become available, the model characterizes how power is distributed in frequency and space—the major step enabling spatial frequency reuse.

The novel cooperative sensing approach, D-Lasso, is designed to be implemented in an ad-hoc network where the radios exchange information locally only with their one-hop neighbors, eliminating the need for a fusion center, and with guaranteed convergence to the globally optimum solution. Simulations corroborated that CRs reach consensus on their PSD estimates and the online implementation of D-Lasso adapts to changes in the transmit-PSDs. Additional tests demonstrated how online-D-Lasso succeeds to estimate the free and occupied bands at an arbitrary location not coinciding with the receiving CRs. Finally, the simulations confirmed that exploiting sparsity is well justified in distributed sensing because LS alternatives incur 15 dB higher leakage of power across space and frequency.

A number of intriguing directions open up for future research. In the CR network sensing front, the present approach accounts for pathloss in the fading links but not for shadowing effects. One approach to cope with shadowing is to complement the distance-only dependent propagation functions with a non-parametric model that can be learned from the data. Preliminary results in this direction can be found in [13]. As far as the distributed Lasso, it will be interesting to explore online coordinate descent type solvers to further lower the complexity of local iterations. Finally, it will be nice to develop algorithms with quantifiable performance for distributed localization in the presence of spatial inhomogeneities and test their application in multi-target and jammer identification settings.

Appendix A
Proof of Proposition 1

Since the average periodogram is (asymptotically) unbiased and the EWMA does not affect its mean, the error $e^{i,\beta}_{r,N}(f)$ in (4) is zero-mean. Thus, the variance of $e^{i,\beta}_{r,N}(f)$ is equal to that of $\hat{\Phi}^{i,\beta}_{r,N}(f)$ as $N \to \infty$. In order to prove that the bound holds for the variance of the latter, recall that $\hat{I}^{(i)}_{r,N}(f)$ is per coherence block i an inconsistent estimator of $\Phi^{(i)}_r(f) := \sum_{s=1}^{N_s} |H_{rs}(i;f)|^2 \Phi_s(f) + \sigma_r^2$, with asymptotic variance $[\Phi^{(i)}_r(f)]^2$ [4, p. 125]. But since the periodograms and channels are uncorrelated across coherence blocks, it follows under Assumption 1 and Assumption 2 that [see (23), shown

Authorized licensed use limited to: University of Minnesota. Downloaded on February 8, 2010 at 19:22 from IEEE Xplore. Restrictions apply.

at the bottom of this page], where we used the identities $\mathrm{Var}_{S,H}(X) = E_H(\mathrm{Var}_{S|H}(X)) + \mathrm{Var}_H(E_{S|H}(X))$ and $(1-\beta)^2\sum_{i'=0}^{i}\beta^{2(i-i')} = (1-\beta)^2(1-\beta^{2i+2})/(1-\beta^2) \simeq (1-\beta)/2$ for i large and $\beta \simeq 1$. In establishing the last approximation it was further assumed that the variances and expectations in the sum do not depend on i', which follows from stationarity and is corroborated next. As for the interchange of limits with Var_H or E_H, the distributions of random fading channels H encountered in practice (e.g., Rayleigh ones) have finite moments; thus, the periodogram moments with respect to H are finite, which is sufficient. Focusing on the expectation of the squared sum in (23), it holds that

$$\begin{aligned}
&E_H\Big(\Big[\sum_{s=1}^{N_s}|H_{rs}(i;f)|^2\Phi_s(f)+\sigma_r^2\Big]^2\Big)\\
&=\sum_{s=1}^{N_s}\Phi_s^2(f)E\Big(|H_{sr}(i;f)|^4\Big)\\
&+\sum_{s=1}^{N_s}\sum_{s'=1,s'\neq s}^{N_s}\Phi_s(f)\Phi_{s'}(f)E\Big(|H_{sr}(i;f)|^2\Big)E\Big(|H_{s'r}(i;f)|^2\Big)\\
&+2\sigma_r^2\sum_{s=1}^{N_s}\Phi_s(f)E\Big(|H_{sr}(i;f)|^2\Big)+\sigma_r^4.
\end{aligned}$$

As the channel gains are square amplitudes of complex Gaussian random variables, they are χ_2^2 (exponentially) distributed. Hence, after recalling that $\gamma_{sr} := E(|H_{sr}(i;f)|^2)$, it follows that $E(|H_{sr}(i;f)|^4) = 2E^2(|H_{sr}(i;f)|^2) = 2\gamma_{sr}^2$, and the last expression reduces to

$$\begin{aligned}
E_H\Big(\Big[\sum_{s=1}^{N_s}|H_{rs}(i;f)|^2\Phi_s(f)+\sigma_r^2\Big]^2\Big) &= 2\sum_{s=1}^{N_s}\Phi_s^2(f)\gamma_{sr}^2\\
+\sum_{s=1}^{N_s}\sum_{s'=1s'\neq s}^{N_s}\Phi_s(f)\Phi_{s'}(f)\gamma_{sr}\gamma_{s'r}&+2\sigma_r^2\sum_{s=1}^{N_s}\Phi_s(f)\gamma_{sr}+\sigma_r^4\\
&=\sum_{s=1}^{N_s}\Phi_s^2(f)\gamma_{sr}^2+\Big[\sum_{s=1}^{N_s}\Phi_s(f)\gamma_{sr}+\sigma_r^2\Big]^2. \qquad (24)
\end{aligned}$$

As for the variance term in (23), it holds that

$$\begin{aligned}
&\mathrm{Var}_H\Big(\sum_{s=1}^{N_s}|H_{rs}(i;f)|^2\Phi_s(f)+\sigma_r^2\Big)\\
&=\sum_{s=1}^{N_s}\Phi_s^2(f)\mathrm{Var}_H\Big(|H_{rs}(i;f)|^2\Big)\\
&=\sum_{s=1}^{N_s}\Phi_s^2(f)\gamma_{sr}^2. \qquad (25)
\end{aligned}$$

Putting (23)–(25) together one arrives at

$$\begin{aligned}
\lim_{N\to\infty}\mathrm{Var}\Big(\hat{\Phi}_{r,N}^{i,\beta}(f)\Big) &= \frac{1-\beta}{2}\Big(2\sum_{s=1}^{N_s}\Phi_s^2(f)\gamma_{sr}^2\Big)\\
&\quad+\frac{1-\beta}{2}\Big(\sum_{s=1}^{N_s}\Phi_s(f)\gamma_{sr}+\sigma_r^2\Big)^2\\
&\le \frac{1-\beta}{2}\Big[2\Big(\sum_{s=1}^{N_s}\Phi_s(f)\gamma_{sr}\Big)^2+\Big(\sum_{s=1}^{N_s}\Phi_s(f)\gamma_{sr}+\sigma_r^2\Big)^2\Big] \qquad (26)
\end{aligned}$$

which in light of (2) completes the proof.

A similar procedure can be used for deriving an expression for the covariance shown in the equation at the bottom of the next page, where the identity $\mathrm{Cov}_{SH}(X_1,X_2) = E_H[\mathrm{Cov}_{S|H}(X_1,X_2)]+\mathrm{Cov}_H[E_{S|H}(X_1),E_{S|H}(X_2)]$ was used along with the asymptotic uncorrelatedness of periodogram values at distinct frequencies, that is $\lim_{N\to\infty}\mathrm{Cov}_{S|H}[I_{r,N}^{i,\beta}(f_1),\ I_{r,N}^{i,\beta}(f_2)] = 0$, and the asymptotic unbiasedness of the periodogram to obtain $\lim_{N\to\infty}E_{S|H}\Big[I_{r,N}^{i,\beta}(f_1)\Big] = \sum_{s=1}^{N_s}|H_{rs}(i';f_1)|^2\Phi_s(f_2)+\sigma_r^2$. This last expression requires knowledge of the channel

$$\begin{aligned}
\lim_{N\to\infty}\mathrm{Var}\Big(\hat{\Phi}_{r,N}^{i,\beta}(f)\Big) &= (1-\beta)^2\sum_{i'=0}^{i}\beta^{2(i-i')}\lim_{N\to\infty}\mathrm{Var}\Big(\hat{I}_{r,N}^{i',\beta}(f)\Big)\\
&=(1-\beta)^2\sum_{i'=0}^{i}\beta^{2(i-i')}\Big[E_H\Big(\lim_{N\to\infty}\mathrm{Var}_{S|H}\Big[I_{r,N}^{i',\beta}(f)\Big]\Big)+\mathrm{Var}_H\Big(\lim_{N\to\infty}E_{S|H}\Big[I_{r,N}^{i',\beta}(f)\Big]\Big)\Big]\\
&=(1-\beta)^2\sum_{i'=0}^{i}\beta^{2(i-i')}E_H\Big(\Big[\sum_{s=1}^{N_s}|H_{rs}(i';f)|^2\Phi_s(f)+\sigma_r^2\Big]^2\Big)\\
&\quad+(1-\beta)^2\sum_{i'=0}^{i}\beta^{2(i-i')}\mathrm{Var}_H\Big(\sum_{s=1}^{N_s}|H_{rs}(i';f)|^2\Phi_s(f)+\sigma_r^2\Big)\\
&\simeq\frac{1-\beta}{2}E_H\Big(\sum_{s=1}^{N_s}|H_{rs}(i;f)|^2\Phi_s(f)+\sigma_r^2\Big)^2+\frac{1-\beta}{2}\mathrm{Var}_H\Big(\sum_{s=1}^{N_s}|H_{rs}(i;f)|^2\Phi_s(f)+\sigma_r^2\Big) \qquad (23)
\end{aligned}$$

Authorized licensed use limited to: University of Minnesota. Downloaded on February 8, 2010 at 19:22 from IEEE Xplore. Restrictions apply.

correlation across frequencies. If this information is not available, the Cauchy-Schwartz inequality can be used for a simpler bound: $\mathrm{Cov}(X_1,X_2) \leq \sqrt{E(X_1^2)E(X_1^2)} - E(X_1)E(X_2)$, which for $X_1 := |H_{rs}(i;f_1)|^2$ and $X_2 := |H_{rs}(i;f_2)|^2$ leads to $\mathrm{Cov}_H\Big(|H_{rs}(i;f_1)|^2, |H_{rs}(i;f_2)|^2\Big) \leq \gamma_{sr}^2$, and hence

$$\lim_{N\longrightarrow\infty} \mathrm{Cov}\Big[\hat{\Phi}_{r,N}^{i,\beta}(f_1), \Phi_{r,N}^{i,\beta}(f_2)\Big] \leq \frac{1-\beta}{2}\sum_{s=1}^{N_s}\Phi_s(f_1)\Phi_s(f_2)\gamma_{sr}^2 \leq \frac{1-\beta}{2}\sum_{s=1}^{N_s}\Phi_s(f_1)\gamma_{sr}\sum_{s=1}^{N_s}\Phi_s(f_2)\gamma_{sr}. \quad (27)$$

APPENDIX B
SELECTION OF λ

Concatenating (5) for $r = 1, 2, \ldots, N_r$ yields the aggregate model $\boldsymbol{\varphi} = \boldsymbol{B\theta} + \mathbf{1}\otimes\boldsymbol{\sigma}^2 + \boldsymbol{e}$, where $\otimes$ represents the Kronecker product, and $\boldsymbol{\sigma}^2 := [\sigma_1^2 \ldots \sigma_{N_r}^2]^T$. If the $P := N_sN_b$ columns of $\boldsymbol{B}$ are orthonormal and the noise $\boldsymbol{e}$ is white, the optimum Lasso estimator for this linear regression model can be found in closed form [10]. Exploiting the latter and its asymptotic (as $NN_r \longrightarrow \infty$) minimax optimality, [7] advocated choosing $\lambda = \sigma_e\sqrt{2\log(P)}$, even when the columns of $\boldsymbol{B}$ are not necessarily orthogonal. Our selection of λ here will follow these guidelines, accounting for the norm of the columns of $\boldsymbol{B}$, and also for the color of the noise $\boldsymbol{e}$ present in our context.

B1. Equivalent Noise Level:

The averaged periodogram estimates comprising our data vector $\boldsymbol{\varphi}$, are known to be asymptotically unbiased with asymptotic variance proportional to the square of the PSD [4, p. 125]. The asymptotic unbiasedness implies that the covariance matrix of $\boldsymbol{e}$ coincides with that of $\boldsymbol{\varphi}$; while the asymptotic variance dictates that the error at a frequency f of an occupied band will have higher variance than that of a free band. This implies that the entries in the diagonal of the covariance matrix of $\boldsymbol{e}$ are not identical. Given that, a reasonable choice for a scalar equivalent noise power is

$$\bar{\sigma}_e^2 = (1-\beta)\max_{r=1,\ldots,N_r}\frac{3}{2N}\|\boldsymbol{\varphi}_r\|^2. \quad (28)$$

To justify the choice in (28), recall that $\boldsymbol{\varphi}_r$ provides a PSD estimate formed at CR r. Furthermore, the variance of $\boldsymbol{\varphi}_r$ averaged across frequencies is bounded by $3(1-\beta)\|\boldsymbol{\varphi}_r\|^2/(2N)$ (cf. Proposition 1). Clearly, selecting the maximum noise variance across receiving CRs represents a conservative worst-case choice for the "equivalent noise" power.

B2. Normalized λ:

The next item needed in deciding the proper λ in our context is an (at least approximate) expression for the norm of the columns of $\boldsymbol{B}$.

Let $\boldsymbol{c}_p$, $p = \nu + (s-1)N_b$, denote the p^{th} column of $\boldsymbol{B}$ corresponding to the transmitter s and the basis $b_\nu(f)$. With k denoting the k^{th} frequency of the periodogram and r the receiving CR, we have $\|\boldsymbol{c}_p\|^2 = \sum_{r=1}^{N_r}\sum_{k=1}^{N} b_\nu^2(f_k)\gamma_{sr}^2$.

If the frequency bases $b_\nu(f_k)$ are non-overlapping boxes as in Fig. 1, then $b_\nu(f_k) = 1$ for N/N_b frequencies f_k and zero for the remaining $N(N_b-1)/N_b$ frequencies. Hence, $\|\boldsymbol{c}_p\|^2 = (N/N_b)\sum_{r=1}^{N_r}\gamma_{sr}^2$ which for N_r sufficiently large can be well approximated by

$$\|\boldsymbol{c}_p\|^2 \simeq \frac{NN_r}{N_b}E_{\boldsymbol{x}_r}[\gamma_{sr}^2]. \quad (29)$$

If receivers are uniformly distributed over a region $\mathcal{A}$ of area A, and $\gamma_{sr} = \min\{1, (d/d_0)^{-\alpha}\}$ is adopted as the PSD loss model, the expectation in (29) can be approximated as

$$E_{\boldsymbol{x}_r}[\gamma_{sr}^2] \simeq \int_{\mathcal{A}}\gamma_{sr}^2\frac{1}{A}d\boldsymbol{x}_r \simeq \frac{1}{A}\int_{\mathbb{R}^2}\min\left\{1, \left(\frac{d}{d_0}\right)^{-\alpha}\right\}d\boldsymbol{x}_r = \frac{\pi d_0^2\alpha}{A(\alpha-1)}. \quad (30)$$

Substituting (30) into (29) it follows that

$$\|\boldsymbol{c}_p\|^2 \simeq \frac{\pi d_0^2\alpha NN_r}{2A(\alpha-1)N_b} =: \bar{c}^2. \quad (31)$$

To finalize the selection of λ for the model with white-noise-equivalent variance given by (28) and regressor norms approximated as in (31), consider the Lasso cost for the aggregate regression model:

$$\min_{\boldsymbol{\theta}\geq\mathbf{0},\boldsymbol{\sigma}^2\geq\mathbf{0}}\|\boldsymbol{B\theta} + \mathbf{1}\otimes\boldsymbol{\sigma}^2 - \boldsymbol{\varphi}_r\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1. \quad (32)$$

With change of variables $\boldsymbol{\theta}' = \boldsymbol{\theta}/\bar{c}$, the problem (32) is clearly equivalent to

$$\min_{\boldsymbol{\theta}'\geq\mathbf{0},\boldsymbol{\sigma}^2\geq\mathbf{0}}\left\|\frac{\boldsymbol{B\theta}'}{\bar{c}} + \mathbf{1}\otimes\boldsymbol{\sigma}^2 - \boldsymbol{\varphi}\right\|_2^2 + \lambda\left\|\frac{\boldsymbol{\theta}'}{\bar{c}}\right\|_1. \quad (33)$$

$$\begin{aligned}\lim_{N\longrightarrow\infty}\mathrm{Cov}\Big(\hat{\Phi}_{r,N}^{i,\beta}(f_1), \Phi_{r,N}^{i,\beta}(f_2)\Big) &= \frac{(1-\beta)}{2}\Big[E_H\Big(\lim_{N\longrightarrow\infty}\mathrm{Cov}_{S|H}\Big[I_{r,N}^{i,\beta}(f_1), I_{r,N}^{i,\beta}(f_2)\Big]\Big) \\ &\quad + \mathrm{Cov}_H\Big(\lim_{N\longrightarrow\infty}E_{S|H}\Big[I_{r,N}^{i,\beta}(f_1)\Big], \lim_{N\longrightarrow\infty}E_{S|H}\Big[I_{r,N}^{i,\beta}(f_2)\Big]\Big)\Big] \\ &= \frac{1-\beta}{2}\mathrm{Cov}_H\Big(\sum_{s=1}^{N_s}|H_{rs}(i;f_1)|^2\Phi_s(f_1), \sum_{s=1}^{N_s}|H_{rs}(i;f_2)|^2\Phi_s(f_2)\Big) \\ &= \frac{(1-\beta)}{2}\sum_{s=1}^{N_s}\Phi_s(f_1)\Phi_s(f_2)\mathrm{Cov}_H\Big(|H_{rs}(i;f_1)|^2, |H_{rs}(i;f_2)|^2\Big)\end{aligned}$$

Authorized licensed use limited to: University of Minnesota. Downloaded on February 8, 2010 at 19:22 from IEEE Xplore. Restrictions apply.

Upon expanding the product $\boldsymbol{B}\boldsymbol{\theta}'$ and taking $\bar{c}$ out of the ℓ_1 norm, (33) reduces to

$$\min_{\boldsymbol{\theta}'\geq\mathbf{0},\boldsymbol{\sigma}^2\geq\mathbf{0}} \left\| \frac{\sum_{p=1}^{P} \boldsymbol{c}_p\theta'_p}{\bar{c}} + \mathbf{1}\otimes\boldsymbol{\sigma}^2 - \boldsymbol{\varphi} \right\|_2^2 + \left(\frac{\lambda}{\bar{c}}\right)\|\boldsymbol{\theta}'\|_1. \quad (34)$$

For this problem the regressors $\boldsymbol{c}_p/\bar{c}$ have unit norm, which allows one to select the tradeoff factor $\lambda' = \lambda/\bar{c} = \sigma_e\sqrt{2\log P}$. For the original problem, this implies $\lambda = \bar{c}\sigma_e\sqrt{2\log P}$. Combining the latter with (31), we arrive at

$$\lambda = \sigma_e d_0 \sqrt{\frac{2\pi\alpha N N_r \log P}{(\alpha-1)AN_b}}. \quad (35)$$

Plugging (28) into (35), the chosen trade-off parameter takes the form of

$$\lambda = \max_{r=1,\ldots,N_r} \lambda_r,\ \lambda_r^2 := 3(1-\beta)\|\boldsymbol{\varphi}_r\|^2 \frac{\pi\alpha d_0^2 N_r \log(P)}{(\alpha-1)AN_b}. \quad (36)$$

Appendix C
Proof Proposition 2

The goal here is to prove that iterations (10) and (11) converge to the solution of (9). But since (9) is equivalent to (7), this will automatically establish that the iterate $\hat{\boldsymbol{\theta}}(j)$ converges to the solution of (7) as well. The proof amounts to showing that these iterations can be put in the form of the alternating direction method of multipliers (ADMoM), which is known to converge [2, pp. 249-260].

Consider all pairs of neighboring CRs (r,ρ) included in the constraints of (9). For each one of these pairs consider the two unidirectional links $r\longrightarrow\rho$ and $\rho\longrightarrow r$, and define the auxiliary variables $\boldsymbol{z}_{r\rho}$ and $\boldsymbol{z}_{\rho r}$. These variables will enforce consensus indirectly, and also decouple the problem (9) w.r.t. the variables $\boldsymbol{\theta}_r$. Specifically, using these variables (9) can be equivalently written as

$$\min_{\substack{\boldsymbol{\theta}_r\geq\mathbf{0},\ \sigma_r^2\geq 0\\ \{\boldsymbol{z}_{r\rho}\}_{\rho\in\mathcal{N}_r}}} \sum_{r=1}^{N_r} F_r(\boldsymbol{\theta}_r,\sigma_r^2)$$
$$\text{s. to}\quad \boldsymbol{\theta}_r = \boldsymbol{z}_{r\rho},\ \boldsymbol{\theta}_\rho = \boldsymbol{z}_{r\rho},\ \forall r=1,2,\ldots,N_r,\ \rho\in\mathcal{N}_r \quad (37)$$

where for notational brevity we defined

$$F_r(\boldsymbol{\theta}_r,\sigma_r^2) := \left\|\boldsymbol{\varphi}_r - \boldsymbol{B}_r\boldsymbol{\theta}_r - \sigma_r^2\right\|^2 + (\lambda/N_r)\sum_{r=1}^{N_r}\mathbf{1}^T\boldsymbol{\theta}_r.$$

Letting $\boldsymbol{p}_{r\rho 1}$ and $\boldsymbol{p}_{r\rho 2}$ denote the Lagrange multipliers associated with the constraints $\boldsymbol{\theta}_r = \boldsymbol{z}_{r\rho}$ and $\boldsymbol{\theta}_\rho = \boldsymbol{z}_{r\rho}$, respectively, the augmented Lagrangian corresponding to (37) is

$$\begin{aligned}&\mathcal{L}_\alpha(\{\boldsymbol{\theta}_r\},\{\sigma_r^2\},\{\boldsymbol{z}_{r\rho}\},\{\boldsymbol{p}_{r\rho 1}\},\{\boldsymbol{p}_{r\rho 2}\})\\ &=\sum_{r=1}^{N_r} F_r(\boldsymbol{\theta}_r,\sigma_r^2)+\sum_{r=1}^{N_r}\sum_{\rho\in\mathcal{N}_r}\left[\boldsymbol{p}_{r\rho 1}^T(\boldsymbol{\theta}_r-\boldsymbol{z}_{r\rho})+\boldsymbol{p}_{r\rho 2}^T(\boldsymbol{\theta}_\rho-\boldsymbol{z}_{r\rho})\right]\\ &\quad+\frac{\alpha}{2}\sum_{r=1}^{N_r}\sum_{\rho\in\mathcal{N}_r}\left[\left\|\boldsymbol{\theta}_r-\boldsymbol{z}_{r\rho}\right\|^2+\left\|\boldsymbol{\theta}_\rho-\boldsymbol{z}_{r\rho}\right\|^2\right] \quad (38)\end{aligned}$$

where the curly brackets $\{\cdot\}=\{\cdot,\ r=1,2,\ldots N_r,\ \rho\in\mathcal{N}_r\}$ denote the set of all the variables included.

Based on the augmented Lagrangian, the ADMoM comprises the following iterations:

$$\{\hat{\boldsymbol{\theta}}_r(j+1),\hat{\sigma}_r^2(j+1)\} = \arg\min_{\substack{\{\boldsymbol{\theta}_r\geq\mathbf{0}\}\\ \{\sigma_r^2\geq 0\}}} \mathcal{L}_\alpha(\cdot) \quad (39)$$

$$\mathcal{L}_\alpha(\cdot)=\mathcal{L}_\alpha(\{\boldsymbol{\theta}_r\},\{\sigma_r^2\},\{\boldsymbol{z}_{r\rho}(j)\},\{\boldsymbol{p}_{r\rho 1}\}(j),\{\boldsymbol{p}_{r\rho 2}(j)\})$$
$$\{\boldsymbol{z}_{r\rho}(j+1)\} = \arg\min_{\boldsymbol{z}_{r\rho}}\mathcal{L}_\alpha(\cdot) \quad (40)$$

$$\mathcal{L}_\alpha(\cdot)=\mathcal{L}_\alpha(\{\hat{\boldsymbol{\theta}}_r(j+1)\},\{\hat{\sigma}_r^2(j+1)\},\{\boldsymbol{z}_{r\rho}\},\{\boldsymbol{p}_{r\rho 1}(j)\},\{\boldsymbol{p}_{r\rho 2}(j)\})$$
$$\boldsymbol{p}_{r\rho 1}(j+1)=\boldsymbol{p}_{r\rho 1}(j)+\alpha\left[\hat{\boldsymbol{\theta}}_r(j+1)-\boldsymbol{z}_{r\rho}(j+1)\right] \quad (41)$$

$$\boldsymbol{p}_{r\rho 2}(j+1)=\boldsymbol{p}_{r\rho 2}(j)+\alpha\left[\hat{\boldsymbol{\theta}}_\rho(j+1)-\boldsymbol{z}_{r\rho}(j+1)\right] \quad (42)$$

where α denotes the step-size.

The ADMoM iterates converge to the minimizer of the original problem for any constant α[2, pp. 253-260]. For the case studied in this appendix, this implies that the iterates given by (39)–(42) converge to the minimizer of (37). The remainder of the proof aims at showing that these iterations are equivalent with the D-Lasso ones in (10) and (11).

To this end, notice that (38) is quadratic and unconstrained in the variables $\boldsymbol{z}_{r\rho}$; hence, (40) can be solved in closed form. Furthermore, (40) decouples for each variable $\boldsymbol{z}_{r\rho}$ into sub-problems of the form

$$\begin{aligned}\boldsymbol{z}_{r\rho}(j+1)=\arg\min_{\boldsymbol{z}_{r\rho}}&\boldsymbol{p}_{r\rho 1}^T(\hat{\boldsymbol{\theta}}_r(j+1)-\boldsymbol{z}_{r\rho})+\boldsymbol{p}_{r\rho 2}^T(\hat{\boldsymbol{\theta}}_\rho(j+1)-\boldsymbol{z}_{r\rho})\\ &+\frac{\alpha}{2}\left[\left\|\hat{\boldsymbol{\theta}}_r(j+1)-\boldsymbol{z}_{r\rho}\right\|^2+\left\|\hat{\boldsymbol{\theta}}_\rho(j+1)-\boldsymbol{z}_{r\rho}\right\|^2\right] \quad (43)\end{aligned}$$

which are solved in closed form as

$$\begin{aligned}\boldsymbol{z}_{r\rho}(j+1)=&\frac{1}{2}\left(\hat{\boldsymbol{\theta}}_r(j+1)+\hat{\boldsymbol{\theta}}_\rho(j+1)\right)\\ &+\frac{1}{2\alpha}\left(\boldsymbol{p}_{r\rho 1}(j)+\boldsymbol{p}_{r\rho 2}(j)\right). \quad (44)\end{aligned}$$

Substituting (44) into (41) and (42), yields

$$\begin{aligned}\boldsymbol{p}_{r\rho 1}(j+1)=&\boldsymbol{p}_{r\rho 1}(j)+\frac{\alpha}{2}\left(\hat{\boldsymbol{\theta}}_r(j+1)-\hat{\boldsymbol{\theta}}_\rho(j+1)\right)\\ &-\frac{1}{2}\left(\boldsymbol{p}_{r\rho 1}(j)+\boldsymbol{p}_{r\rho 2}(j)\right) \quad (45)\\ \boldsymbol{p}_{r\rho 2}(j+1)=&\boldsymbol{p}_{r\rho 2}(j)+\frac{\alpha}{2}\left(\hat{\boldsymbol{\theta}}_\rho(j+1)-\hat{\boldsymbol{\theta}}_r(j+1)\right)\\ &-\frac{1}{2}\left(\boldsymbol{p}_{r\rho 1}(j)+\boldsymbol{p}_{r\rho 2}(j)\right). \quad (46)\end{aligned}$$

If $\boldsymbol{p}_{r\rho 1}(0)=\boldsymbol{p}_{r\rho 2}(0)=\mathbf{0}$, it is easy to recognize by inspection that $\boldsymbol{p}_{r\rho 1}(1)=-\boldsymbol{p}_{r\rho 2}(1)$. Arguing by induction, it further follows readily that $\boldsymbol{p}_{r\rho 1}(j)=-\boldsymbol{p}_{r\rho 2}(j)\ \forall\ j$; thus, (45) and (46) reduce to

$$\boldsymbol{p}_{r\rho 1}(j+1)=\boldsymbol{p}_{r\rho 1}(j)+\frac{\alpha}{2}\left(\hat{\boldsymbol{\theta}}_r(j+1)-\hat{\boldsymbol{\theta}}_\rho(j+1)\right) \quad (47)$$

$$\boldsymbol{p}_{r\rho 2}(j+1)=\boldsymbol{p}_{r\rho 2}(j)+\frac{\alpha}{2}\left(\hat{\boldsymbol{\theta}}_\rho(j+1)-\hat{\boldsymbol{\theta}}_r(j+1)\right). \quad (48)$$

Consider now interchanging subscripts r and ρ in (47). Comparing the resulting recursion with (47), it follows that $\boldsymbol{p}_{\rho r2}(j)=\boldsymbol{p}_{r\rho 1}(j)\ \forall j$.

Authorized licensed use limited to: University of Minnesota. Downloaded on February 8, 2010 at 19:22 from IEEE Xplore. Restrictions apply.

Turning attention to (39), observe that it can also be decoupled for each pair of variables $\left(\boldsymbol{\theta}_r, \sigma_r^2\right)$, into N_r sub-problems of the form [see (49) at the bottom of this page]. Recall that $\boldsymbol{p}_{\rho r 2}(j) = \boldsymbol{p}_{r\rho 1}(j)$ and $\boldsymbol{p}_{r\rho 1}(j) = -\boldsymbol{p}_{r\rho 2}(j)$, which implies that (c.f (44)): $\boldsymbol{z}_{r\rho}(j) = \left(\hat{\boldsymbol{\theta}}_r(j) + \hat{\boldsymbol{\theta}}_\rho(j)\right)/2$. Using these identities to eliminate $\boldsymbol{z}_{r\rho}(j)$, $\boldsymbol{z}_{\rho r}(j)$, and $\boldsymbol{p}_{r\rho 2}(j)$ from (49), we arrive at

$$(\hat{\boldsymbol{\theta}}_r(j+1), \hat{\sigma}_r^2(j+1)) = \arg \min_{\substack{\boldsymbol{\theta}_r \geq \mathbf{0} \\ \sigma_r^2 \geq 0}} F_r(\boldsymbol{\theta}_r, \sigma_r^2) + 2 \sum_{\rho \in \mathcal{N}_r} \boldsymbol{p}_{r\rho 1}^T(j) \boldsymbol{\theta}_r + \alpha \sum_{\rho \in \mathcal{N}_r} \left|\left| \boldsymbol{\theta}_r - \frac{1}{2}\left(\hat{\boldsymbol{\theta}}_r(j) + \hat{\boldsymbol{\theta}}_\rho(j)\right) \right|\right|^2. \quad (50)$$

Equation (50) evidences that CR r does not need to update a separate price vector $\boldsymbol{p}_{r\rho 1}(j)$ per neighbor, but only the sum of them denoted by $\boldsymbol{q}_r(j) = \sum_{\rho \in \mathcal{N}_r} \boldsymbol{p}_{r\rho 1}(j)$. Using $\boldsymbol{q}_r(j)$, problem (50) is written as

$$\left(\hat{\boldsymbol{\theta}}_r(j+1), \hat{\sigma}_r^2(j+1)\right) = \arg \min_{\substack{\boldsymbol{\theta}_r \geq \mathbf{0} \\ \sigma_r^2 \geq 0}} F_r(\boldsymbol{\theta}_r, \sigma_r^2) + 2\boldsymbol{q}_r^T(j)\boldsymbol{\theta}_r + \alpha \sum_{\rho \in \mathcal{N}_r} \left|\left| \boldsymbol{\theta}_r - \frac{1}{2}\left(\hat{\boldsymbol{\theta}}_r(j) + \hat{\boldsymbol{\theta}}_\rho(j)\right) \right|\right|^2 \quad (51)$$

and $\boldsymbol{q}_r(j)$ is updated as in (47) after summing w.r.t. $\rho \in \mathcal{N}_r$; that is,

$$\boldsymbol{q}_r(j+1) = \boldsymbol{q}_r(j) + \frac{\alpha}{2}\left(|\mathcal{N}_r| \hat{\boldsymbol{\theta}}_r(j+1) - \sum_{\rho \in \mathcal{N}_r} \hat{\boldsymbol{\theta}}_\rho(j+1) \right). \quad (52)$$

Since the convergent ADMoM iterations (39)–(42) can be rewritten as in (VII)–(52), it follows readily that the D-Lasso iterations (10) and (11) are also convergent, which concludes the proof of the proposition.

Appendix D
Proof of Lemma 1

Let $\boldsymbol{\mu}$ and $\boldsymbol{\rho}$ denote the Lagrange multipliers corresponding to the non-negativity constrains on $\boldsymbol{\theta}$ and $\boldsymbol{\sigma}^2$, respectively; and $\boldsymbol{\lambda} := \lambda \mathbf{1}$, where $\mathbf{1}$ is the $P \times 1$ vector of all ones. Using (17), the Lagrangian function corresponding to (7) is given by

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\sigma}^2, \boldsymbol{\mu}, \boldsymbol{\rho}) = ||\boldsymbol{B}\boldsymbol{\theta} + \boldsymbol{R}\boldsymbol{\sigma}^2 - \boldsymbol{\varphi}||^2 + \boldsymbol{\lambda}^T\boldsymbol{\theta} + \boldsymbol{\mu}^T\boldsymbol{\theta} + \boldsymbol{\rho}^T\boldsymbol{\sigma}^2. \quad (53)$$

The pair $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\sigma}}^2$ solving (7) satisfies the following equations that result from setting to zero the gradients of the Lagrangian respect to $\boldsymbol{\theta}$ and $\boldsymbol{\sigma}^2$, that is

$$\boldsymbol{B}^T\boldsymbol{B}\hat{\boldsymbol{\theta}} + \boldsymbol{B}^T\boldsymbol{R}\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{B}^T\boldsymbol{\varphi} + \frac{1}{2}(\boldsymbol{\lambda} + \boldsymbol{\mu}) = \mathbf{0} \quad (54)$$

$$\boldsymbol{R}^T\boldsymbol{R}\hat{\boldsymbol{\sigma}}^2 + \boldsymbol{R}^T\boldsymbol{B}\hat{\boldsymbol{\theta}} - \boldsymbol{R}^T\boldsymbol{\varphi} + \frac{1}{2}\boldsymbol{\rho} = \mathbf{0}. \quad (55)$$

Let $\hat{\boldsymbol{\theta}}_1$ (resp. $\hat{\boldsymbol{\theta}}_0$) denote the sub-vector of $\hat{\boldsymbol{\theta}}$ with nonzero (resp. zero) entries. The entries of $\boldsymbol{\mu}$ are collected in $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_0$, those of $\boldsymbol{\theta}$ in $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$, and the columns of $\boldsymbol{B}$ in the sub-matrices $\boldsymbol{B}_0$ and $\boldsymbol{B}_1$, always using the partition defined by the separation of $\hat{\boldsymbol{\theta}}$ into $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_0$.

The equations in (54) can be separated accordingly in the following two sets

$$\begin{aligned} \boldsymbol{B}_0^T\boldsymbol{B}\hat{\boldsymbol{\theta}} + \boldsymbol{B}_0^T\boldsymbol{R}\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{B}_0^T\boldsymbol{\varphi} + \frac{1}{2}(\boldsymbol{\lambda}_0 + \boldsymbol{\mu}_0) &= \mathbf{0} \\ \boldsymbol{B}_1^T\boldsymbol{B}\hat{\boldsymbol{\theta}} + \boldsymbol{B}_1^T\boldsymbol{R}\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{B}_1^T\boldsymbol{\varphi} + \frac{1}{2}(\boldsymbol{\lambda}_1 + \boldsymbol{\mu}_1) &= \mathbf{0}. \end{aligned} \quad (56)$$

Henceforth, only the second set of equations will be used. By the complementary slackness conditions, $\boldsymbol{\mu}_1 = \mathbf{0}$ and (56) reduces to

$$\boldsymbol{B}_1^T\boldsymbol{B}_1\hat{\boldsymbol{\theta}}_1 + \boldsymbol{B}_1^T\boldsymbol{R}\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{B}_1^T\boldsymbol{\varphi} + \frac{1}{2}\boldsymbol{\lambda}_1 = \mathbf{0} \quad (57)$$

where we used the fact that $\boldsymbol{B}\hat{\boldsymbol{\theta}} = \boldsymbol{B}_0\hat{\boldsymbol{\theta}}_0 + \boldsymbol{B}_1\hat{\boldsymbol{\theta}}_1 = \boldsymbol{B}_1\hat{\boldsymbol{\theta}}_1$, because $\hat{\boldsymbol{\theta}}_0$ is null by definition.

Replacing $\boldsymbol{\varphi}$ in (57) from (17), yields

$$\boldsymbol{B}_1^T\boldsymbol{B}_1\hat{\boldsymbol{\theta}}_1 + \boldsymbol{B}_1^T\boldsymbol{R}\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{B}_1^T\left(\boldsymbol{B}\boldsymbol{\theta} + \boldsymbol{R}\boldsymbol{\sigma}^2 + \boldsymbol{e}\right) + \frac{1}{2}\boldsymbol{\lambda}_1 = \mathbf{0}$$

and upon rearranging terms we arrive at

$$\boldsymbol{B}_1^T\boldsymbol{B}_1\hat{\boldsymbol{\theta}}_1 - \boldsymbol{B}_1^T\boldsymbol{B}\boldsymbol{\theta} + \boldsymbol{B}_1^T\boldsymbol{R}(\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}^2) - \boldsymbol{B}_1^T\boldsymbol{e} + \frac{1}{2}\boldsymbol{\lambda}_1 = \mathbf{0}. \quad (58)$$

Using Assumption 4, it follows that $\boldsymbol{B}^T\boldsymbol{\theta} = \boldsymbol{B}_1^T\boldsymbol{\theta}_1$, and (58) becomes

$$\boldsymbol{B}_1^T\boldsymbol{B}_1(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1) + \boldsymbol{B}_1^T\boldsymbol{R}(\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}^2) - \boldsymbol{B}_1^T\boldsymbol{e} + \frac{1}{2}\boldsymbol{\lambda}_1 = \mathbf{0}. \quad (59)$$

The same steps of substituting $\boldsymbol{\varphi} = \boldsymbol{B}\boldsymbol{\theta} + \boldsymbol{R}\boldsymbol{\sigma}^2 + \boldsymbol{e}$, $\boldsymbol{B}\hat{\boldsymbol{\theta}} = \boldsymbol{B}_1\hat{\boldsymbol{\theta}}_1$, and $\boldsymbol{B}\boldsymbol{\theta} = \boldsymbol{B}_1\boldsymbol{\theta}_1$, transform (55) into

$$\boldsymbol{R}^T\boldsymbol{R}(\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}^2) + \boldsymbol{R}^T\boldsymbol{B}_1(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}) - \boldsymbol{R}^T\boldsymbol{e} + \boldsymbol{\rho} = \mathbf{0}. \quad (60)$$

But the estimate $\hat{\boldsymbol{\sigma}}^2$ is positive, because otherwise $\hat{\boldsymbol{\theta}}$ must increase in size and amplitude to fit the noise power $\boldsymbol{\sigma}^2$; this in turn will increase the ℓ_1 penalty term in the Lasso cost (7). Hence, the complementary slackness conditions force $\boldsymbol{\rho}$ to be null. In addition, $\boldsymbol{R}$ satisfies by definition $\boldsymbol{R}^T\boldsymbol{R} = N\boldsymbol{I}$, where $\boldsymbol{I}$ denotes the $N_r \times N_r$ identity matrix. Thus, (60) becomes

$$N(\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}^2) + \boldsymbol{R}^T\boldsymbol{B}_1(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}) - \boldsymbol{R}^T\boldsymbol{e} = \mathbf{0}. \quad (61)$$

$$\left(\hat{\boldsymbol{\theta}}_r(j+1), \hat{\sigma}_r^2(j+1)\right) = \arg \min_{\substack{\boldsymbol{\theta}_r \geq \mathbf{0} \\ \sigma_r^2 \geq 0}} F_r(\boldsymbol{\theta}_r, \sigma_r^2) + \sum_{\rho \in \mathcal{N}_r} \boldsymbol{p}_{r\rho 1}^T(j)\boldsymbol{\theta}_r + \sum_{\rho \in \mathcal{N}_r} \boldsymbol{p}_{\rho r 2}^T(j)\boldsymbol{\theta}_r + \frac{\alpha}{2} \sum_{\rho \in \mathcal{N}_r} \left|\left| \boldsymbol{\theta}_r - \boldsymbol{z}_{r\rho}(j) \right|\right|^2 + \frac{\alpha}{2} \sum_{\rho \in \mathcal{N}_r} \left|\left| \boldsymbol{\theta}_r - \boldsymbol{z}_{\rho r}(j) \right|\right|^2. \quad (49)$$

Authorized licensed use limited to: University of Minnesota. Downloaded on February 8, 2010 at 19:22 from IEEE Xplore. Restrictions apply.

Substituting $\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}^2$ from (61) into (59), we obtain

$$\boldsymbol{B}_1^T\boldsymbol{B}_1(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1) - \boldsymbol{B}_1^T\frac{\boldsymbol{R}\boldsymbol{R}^T}{N}\boldsymbol{B}_1(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1) + \boldsymbol{B}_1^T\frac{\boldsymbol{R}\boldsymbol{R}^T}{N}\boldsymbol{e} - \boldsymbol{B}_1^T\boldsymbol{e} + \frac{1}{2}\boldsymbol{\lambda}_1 = \boldsymbol{0}$$

from which the error $\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1$ can be expressed as

$$\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1 = \left(\boldsymbol{B}_1^T\left[\frac{\boldsymbol{I} - \boldsymbol{R}\boldsymbol{R}^T}{N}\right]\boldsymbol{B}_1\right)^{-1}\left(\boldsymbol{B}_1^T\left[\frac{\boldsymbol{I} - \boldsymbol{R}\boldsymbol{R}^T}{N}\right]\boldsymbol{e} - \frac{1}{2}\boldsymbol{\lambda}_1\right)$$

and the results follows.

REFERENCES

[1] J. A. Bazerque and G. B. Giannakis, "Distributed spectrum sensing for cognitive radios by exploiting sparsity," presented at the 42nd Asilomar Conf. Signals, Syst., Comput., Pacific Grove, CA, Oct. 26–29, 2008.

[2] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, 2nd ed. Belmont, MA: Athena Scientific, 1999.

[3] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[4] D. R. Brillinger, *Time Series: Data Analysis and Theory*. San Francisco, CA: Holden Day, 1981.

[5] E. J. Candès and Y. Plan, "Near-ideal model selection by L1 minimization," *ArXiv E-prints, Eprint 0801.0345*, 2008.

[6] V. Cevher, M. F. Duarte, and R. G. Baraniuk, "Distributed target localization via spatial sparsity," presented at the 16th Eur. Signal Process. Conf., Lausanne, Switzerland, Aug. 25–29, 2008.

[7] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1999.

[8] R. Chen, J. M. Park, and K. Bian, "Robust distributed spectrum sensing in cognitive radio networks," in *Proc. 27th Conf. Comput. Commun.*, Phoenix, AZ, Apr. 13–18, 2008, pp. 1876–1884.

[9] C. R. C. da Silva, B. Choi, and K. Kim, "Distributed spectrum sensing for cognitive radio systems," in *Proc. Workshop Inf. Theory Appl.*, San Diego, CA, Feb. 2, 2007, pp. 120–123.

[10] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.

[11] G. Ganesan, Y. Li, B. Bing, and S. Li, "Spatiotemporal sensing in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 26, pp. 5–12, Jan. 2006.

[12] A. Ghasemi and E. S. Sousa, "Spectrum sensing in cognitive radio networks: The cooperation-processing tradeoff," *Wireless Commun. Mobile Comput.*, vol. 7, no. 9, pp. 1049–1060, 2007.

[13] G. Mateos, J.-A. Bazerque, and G. B. Giannakis, "Spline-based spectrum cartography for cognitive radios," presented at the 43rd Asilomar Conf. Signals, Syst., Comput., Pacific Grove, CA, Nov. 1–4, 2009.

[14] S. M. Mishra, A. Sahai, and R. W. Brodersen, "Cooperative sensing among cognitive radios," in *Proc. 42nd Int. Conf. Commun.*, Istanbul, Turkey, Jun. 11–15, 2006, pp. 1658–1663.

[15] K. Nishimori, R. D. Taranto, H. Yomo, P. Popovski, Y. Takatori, R. Prasad, and S. Kubota, "Spatial opportunity for cognitive radio systems with heterogeneous path loss conditions," in *Proc. 65th Veh. Technol. Conf.*, Dublin, Ireland, Apr. 22–25, 2007, pp. 2631–2635.

[16] J. Riihijärvi and P. Mähönen, "Exploiting spatial statistics of primary and secondary users towards improved cognitive radio networks," in *Proc. 3rd Int. Conf. Cognitive Radio Oriented Wireless Netw. Commun.*, Singapore, May 15–17, 2008, pp. 1–7.

[17] Z. Quan, S. Cui, A. Sayed, and H. V. Poor, "Optimal multiband joint detection for spectrum sensing in cognitive radio networks," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1128–1140, Mar. 2009.

[18] I. D. Schizas, A. Ribeiro, and G. B. Giannakis, "Consensus in Ad Hoc WSNs with noisy links—Part I: Distributed estimation of deterministic signals," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 350–364, Jan. 2008.

[19] J. F. Stürm, "Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones," *Optim. Methods. Softw.*, vol. 12, no. 11, pp. 625–653, 1999.

[20] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. Roy. Stat. Soc., Series B*, vol. 58, no. 1, pp. 267–288, 1996.

[21] Q. Zhao and B. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 79–89, May 2007.

[22] Q. Zhao and A. Swami, "A survey of dynamic spectrum access: Signal processing and networking perspectives," in *Proc. 55th Int. Conf. Acoust., Speech, Signal Process.*, Honolulu, HI, Apr. 15–20, 2007, pp. 1349–1352.

[23] P. Zhao and B. Yu, "On model selection consistency of Lasso," *J. Mach. Learn. Res.*, vol. 7, pp. 2541–2563, Nov. 2006.

[24] H. Zou, "The adaptive Lasso and its oracle properties," *J. Amer. Stat. Assoc.*, vol. 101, no. 476, pp. 1418–1429, Dec. 2006.

Juan Andrés Bazerque (S'06) received the B.Sc. degree in electrical engineering from the 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 decentralized algorithms for in-network processing, cooperative wireless communications, cognitive radios, compressive sampling, and sparsity-aware statistical models.

Mr. Bazerque is the recipient of the UofM's Distinguished Master's Thesis Award 2009–2010 and is corecipient of the Best Student Paper Award at the Second International Conference on Cognitive Radio Oriented Wireless Networks and Communication 2007.

Georgios B. Giannakis (F'97) received the Diploma degree in electrical engineering from the National Technical University of Athens, Greece, in 1981 and the MSc. degree in electrical engineering, the M.Sc. degree in mathematics, and the Ph.D. degree in electrical engineering from the University of Southern California (USC) in 1983, 1986 and 1986, respectively.

Since 1999, he has been a Professor with the University 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 285 journal papers, 485 conference papers, two edited books, and two research monographs. Current research focuses on compressive sensing, cognitive radios, network coding, cross-layer designs, mobile ad hoc networks, wireless sensor, and social networks.

Dr. Giannakis is the (co)recipient of seven paper awards from the IEEE Signal Processing (SP) and Communications Societies, including the G. Marconi Prize 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 a Fellow of EURASIP, has served the IEEE in a number of posts, and is also as a Distinguished Lecturer for the IEEE-SP Society.

Authorized licensed use limited to: University of Minnesota. Downloaded on February 8, 2010 at 19:22 from IEEE Xplore. Restrictions apply.