Cooperative Spectrum Sensing for Cognitive Radios Using Kriged Kalman Filtering

Seung-Jun Kim, Emiliano Dall'Anese, *Student Member, IEEE*, and Georgios B. Giannakis, *Fellow, IEEE*

***Abstract*—A cooperative cognitive radio (CR) sensing problem is considered, where a number of CRs collaboratively detect the presence of primary users (PUs) by exploiting the novel notion of *channel gain (CG) maps*. The CG maps capture the propagation medium per frequency from *any* point in space and time to each CR user. They are updated in real-time using Kriged Kalman filtering (KKF), a tool with well-appreciated merits in geostatistics. In addition, the CG maps enable tracking the transmit-power and location of an unknown number of PUs, via a sparse regression technique. The latter exploits the sparsity inherent to the PU activities in a geographical area, using an ℓ_1-norm regularized, sparsity-promoting weighted least-squares formulation. The resulting sparsity-cognizant tracker is developed in both centralized and distributed formats, to reduce computational complexity and memory requirements of a batch alternative. Numerical tests demonstrate considerable performance gains achieved by the proposed algorithms.**

***Index Terms*—Channel estimation, cognitive radio, compressed sampling, distributed algorithms, Kalman filters.**

I. Introduction

In SPITE of the perceived scarcity of spectral resources for new wireless services, the licensed spectrum is grossly under-utilized in its space and time dimensions [1]. This has stimulated interest in the concept of cognitive radios (CRs), which aim to make opportunistic use of the unused bands, by identifying the "white spaces" in frequency, time, and/or space domains. In a hierarchical access model, the CR secondary users are allowed to transmit only when their transmissions do not interfere with the ongoing communication of the licensed primary users (PUs). Especially for this context, reliable *spectrum sensing* to detect the presence of PU signals is of paramount importance [2].

At the expense of increased communication overhead among CRs, *cooperative sensing* schemes can achieve significantly improved performance relative to single-CR sensing [3]–[5]. Through fusion of local measurements, cooperative sensing can collect the spatial diversity provided by different propagation distances and fading experienced by the PU-to-CR channels. In [6], a weighted sum of the power measurements was used as a test statistic at the fusion center (FC) in order to maximize the probability of detecting available bands. However, PU-to-CR channel gains were assumed independent, and thus correlation of observations across CR nodes was not accounted for. In [7], the sensing decisions made by individual CRs were combined at the FC using a linear-quadratic fusion rule, which takes into account correlated observations. However, the signal-to-noise-ratios (SNRs) of all channels were assumed to have identical means, which may not be realistic.

Recently, interest has grown in *spatial re-use* techniques for CR networks, where CR transmissions are permitted as long as the signal-to-interference-plus-noise-ratio (SINR) requirements of the PU receivers are satisfied thanks to attenuation in the propagation paths [2], [8]. Toward this direction, initial efforts have been made to construct *power spectral density (PSD) maps*, which capture the ambient RF power distribution in space, time, and frequency. The PSD maps provide vivid description of which region in the area of interest is "crowded" in terms of RF interference, and must hence be avoided by the CR transceivers. A spatial interpolation technique called Kriging [9] was employed by [10] and [11] to obtain the PSD maps, albeit the map estimation was performed offline. In [12], a smooth PSD map was computed using the method of splines. However, these works cannot cope with time-varying channel gains.

In this paper, a cooperative CR sensing scheme is developed, in which PU positions and transmit powers are tracked by exploiting a descriptive model of the propagation environment in the area where the CR network is operated. The novel concept of *channel gain (CG) maps* is introduced to capture per frequency the up-to-date channel gains from an arbitrary point in space to the individual CR receivers. Although beyond the scope of this paper, assuming reciprocity, the CG maps can even provide the range of power levels that can be used by the CR transmitters to prevent disruption to the existing PU links. In addition to coping with time-varying environments, the latter constitutes the major complementary information provided by CG maps that is not available with PSD maps, especially when dealing with shadowing propagation channels.

Manuscript received October 12, 2009; revised March 29, 2010; accepted June 04, 2010. Date of publication June 14, 2010; date of current version January 19, 2011. This work was supported by the National Science Foundation under Grants CCF 0830480 and CON 0824007 and also in part 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. Part of this work appeared in the Proceedings of the 43rd Asilomar Conference on Signal, Systems, and Computers, Pacific Grove, CA, November 2009, and in the Proceedings of the 3rd International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, Aruba, Dutch Antilles, December 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Andres Kwasinski.

S.-J. Kim and G. B. Giannakis are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: seungjun@umn.edu; georgios@umn.edu).

E. Dall'Anese is with the Department of Information Engineering (DEI), University of Padova, 35131 Padova, Italy (e-mail: edallane@dei.unipd.it).

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/JSTSP.2010.2053016

A key contribution of this work is a pragmatic model of the spatio-temporal channel correlation structure, so as to obtain an estimate of the entire map from a finite set of spatio-temporal samples. Kriged Kalman filtering (KKF) [13] (a.k.a. space–time Kalman filtering [14]) is a tool with well-appreciated merits in geostatistics and environmental science. Its novel adaption in the present context enables efficient tracking of the channel gains in space and time.

Once the CG map estimates are obtained, the spectrum sensing task is accomplished by solving a sparse regression problem. Sparsity emerges because in a given time instant and frequency band, only a few PU transmitters can be actively transmitting due to mutual interference. Motivated by recent advances in compressive sampling [15], this inherent sparsity was exploited in [16] to estimate transmission powers and the positions of an unknown number of active PUs using a least-squares (LS) criterion regularized with the ℓ_1-norm of the unknown vector, which promotes sparsity in the estimand.

Here, this approach is considerably broadened to track time-varying PU activities with the aid of a pragmatic fading model. To obtain an efficient tracker that can solve a sparse regression problem based on sequentially acquired observations, the recently developed online coordinate descent (OCD) Lasso algorithm [17] is adapted here to the time-varying spectrum sensing problem at hand. Moreover, a distributed version is developed, which converges to the solution of its centralized counterpart based only on local message-passing among neighboring nodes in the CR network.

The rest of the paper is organized as follows. In Section II, the problem statement is presented. In Section III, a KKF-based algorithm is developed for tracking CG maps. Section IV deals with estimation of the model parameters that are necessary for the operation of KKF. The centralized and distributed algorithms to track PU activities are the subjects of Section V. Simulations are presented in Section VI, followed by concluding remarks in Section VII.

II. Problem Statement

Consider an incumbent PU system, the activity of which is to be monitored in a geographical area $\mathcal{A} \subset \mathbb{R}^2$ via cooperation of N_r CRs that are also located in $\mathcal{A}$. The CR locations, denoted by position vectors $\{\mathbf{x}_r \in \mathcal{A}\}_{r=1}^{N_r}$ relative to a given reference coordinate system, are assumed known to the CR network. In lieu of prior information on the PU signal characteristics, and motivated by low-complexity considerations, the CRs employ non-coherent energy detectors to detect the presence of PU signals [18].

Suppose that N_s PU transmitters (sources) are located at $\{\mathbf{x}_s \in \mathcal{A}\}_{s=1}^{N_s}$. Neither the number of PU transmitters nor their positions are known to the CRs. Let $g_r(\mathbf{x}, t)$ denote the channel gain from an arbitrary position $\mathbf{x} \in \mathcal{A}$ to CR r at time t. Assuming that the PU signals are statistically independent, the *receive-power* in the band of interest measured by CR r at time t is given by

$$\pi_r(t) = \sum_{s=1}^{N_s} g_r(\mathbf{x}_s, t) p_s(t) + z_r(t), \quad r = 1, 2, \ldots, N_r \quad (1)$$

where $p_s(t)$ is the *transmit-power* of the PU located at $\mathbf{x}_s$ at time t, and $z_r(t)$ the receiver noise at CR r at time t with mean zero and variance σ_z^2.

The channel gain $g_r(\mathbf{x}, t)$ can be decomposed into path loss, shadowing, and small-scale fading factors. By averaging out the effect of small-scale fading, the (averaged) channel gain of the link $\mathbf{x} \to \mathbf{x}_r$ at time t, expressed in dB, can be written as

$$G_r(\mathbf{x}, t) = G_0 - 10\gamma \log_{10}(\|\mathbf{x} - \mathbf{x}_r\|_2) + s_r(\mathbf{x}, t) \quad (2)$$

where $G_r(\mathbf{x}, t) \triangleq 10 \log_{10} g_r(\mathbf{x}, t)$, G_0 is the path gain at the unit distance, γ the path loss exponent, and $s_r(\mathbf{x}, t)$ the shadow fading in dB. Note that frequency selectivity can be accounted for by dividing the spectrum into frequency-flat subbands, and processing them individually.[1]

In order to track the slowly time-varying $s_r(\mathbf{x}, t)$, training signals transmitted by the CRs will be exploited. The training signals are generally present in the CR packets to aid synchronization and channel estimation of the intended receivers. For simplicity, it is assumed that the CRs transmit unit-power training packets in a time-division multiple access (TDMA)-fashion when the PUs are detected to be silent. Thus, with CR $j \in \{1, \ldots, r-1, r+1, \ldots, N_r\}$ transmitting the training signal in a given TDMA slot at time t and the PUs being silent, CR r can acquire an estimate of $g_r(\mathbf{x}_j, t)$ (and thus of $G_r(\mathbf{x}_j, t)$ by translation to a dB-scale) simply by measuring the received power $\pi_r(t)$. Subsequently, given that CR locations are known, it is possible to obtain an estimate $\breve{s}_r(\mathbf{x}_j, t)$ of the shadow fading component $s_r(\mathbf{x}_j, t)$ by subtracting the known path loss from the estimate of $G_r(\mathbf{x}_j, t)$; i.e., $\breve{s}_r(\mathbf{x}_j, t) = 10 \log_{10} \pi_r(t) - G_0 + 10\gamma \log_{10}(\|\mathbf{x}_j - \mathbf{x}_r\|_2)$.

The noisy measurement $\breve{s}_r(\mathbf{x}_j, t)$ can be expressed as

$$\breve{s}_r(\mathbf{x}_j, t) = s_r(\mathbf{x}_j, t) + \epsilon_r(\mathbf{x}_j, t), \quad j \neq r \quad (3)$$

where $\epsilon_r(\mathbf{x}_j, t)$ denotes the measurement error modeled as Gaussian distributed with zero mean and variance σ_ϵ^2. Consider the vector $\breve{\mathbf{s}}_r(t) \in \mathbb{R}^{N_r - 1}$ containing the measurements collected by CR r at time t; that is $\breve{\mathbf{s}}_r(t) \triangleq [\breve{s}_r(\mathbf{x}_1, t), \ldots, \breve{s}_r(\mathbf{x}_{r-1}, t), \breve{s}_r(\mathbf{x}_{r+1}, t), \ldots, \breve{s}_r(\mathbf{x}_{N_r}, t)]^T$, where ^{T} denotes transposition.

The CR sensing scheme proposed in this paper proceeds in two steps:

S1) Based on $\{\breve{\mathbf{s}}_r(\tau)\}_{\tau=1}^t$ acquired during the training phase for the CR positions, maintain an up-to-date estimate of the CG map $G_r(\mathbf{x}, t)$ at each CR r for *arbitrary* positions $\mathbf{x} \in \mathcal{A}$;

[1]Henceforth, the analysis applies on a per frequency f basis; but f is dropped for notational brevity.

S2) Based on the estimates of $\{G_r(\mathbf{x},t)\}_{r=1}^{N_r}$ formed at all CRs in **S1**), and the received powers $\{\pi_r(\tau)\}_{\tau=1}^{t}$ measured at CRs during the operational phase, cooperatively track the PU locations $\{\mathbf{x}_s\}_{s=1}^{N_s}$ as well as the PU transmit-powers $\{p_s(t)\}_{s=1}^{N_s}$.

Under reciprocity, knowledge of $G_r(\mathbf{x},t)\ \forall \mathbf{x}\in\mathcal{A}$ in **S1**) and transmit-power per CR r enables cartography of the interference profile at any point in space and time.

III. Tracking of CG Maps via KKF

Given the cumulative measurements $\{\breve{\mathbf{s}}_r(\tau)\}_{\tau=1}^{t}$ available at the corresponding CR locations $\{\mathbf{x}_r\}_{r=1}^{N_r}$, the goal of KKF in the present application context is to estimate, that is, interpolate, $s_r(\mathbf{x},t)$ at any *arbitrary* point $\mathbf{x}\in\mathcal{A}$, and track it across time t. The critical step to this spatio-temporal interpolation and tracking task, is an appropriate state-space model for shadow fading—the subject dealt with in the next subsection.

A. State-Space Model

Shadow fading in linear scale is distributed according to a log-normal distribution [19, p. 104]; hence, $s_r(\mathbf{x},t)$ in dB follows a Gaussian distribution. Similar to [13] and [14], the spatio-temporal evolution of $s_r(\mathbf{x},t)$ for $t=1,2,\ldots$, is modeled here as

$$s_r(\mathbf{x},t)=\bar{s}_r(\mathbf{x},t)+\nu_r(\mathbf{x},t) \tag{4}$$

$$\bar{s}_r(\mathbf{x},t)=\int_{\mathcal{A}} w_r(\mathbf{x},\mathbf{u})\bar{s}_r(\mathbf{u},t-1)d\mathbf{u}+\eta_r(\mathbf{x},t) \tag{5}$$

where $\bar{s}_r(\mathbf{x},t)$ is a spatio-temporally colored component with the function $w_r(\mathbf{x},\mathbf{u})$ capturing the interaction between $\bar{s}_r$ at position $\mathbf{x}$ at time t and $\bar{s}_r$ at position $\mathbf{u}$ at time $(t-1)$; and $\nu_r(\mathbf{x},t)$ and $\eta_r(\mathbf{x},t)$ are spatially colored yet temporally white zero-mean Gaussian stationary random fields with separable covariance structure

$$\mathrm{cov}\{\nu_r(\mathbf{x},t),\nu_r(\mathbf{u},\tau)\}=C_{\nu_r}(\mathbf{x}-\mathbf{u})\delta(t-\tau) \tag{6}$$

$$\mathrm{cov}\{\eta_r(\mathbf{x},t),\eta_r(\mathbf{u},\tau)\}=C_{\eta_r}(\mathbf{x}-\mathbf{u})\delta(t-\tau) \tag{7}$$

where $\delta(\cdot)$ denotes the Dirac delta function.

Process $\nu_r(\mathbf{x},t)$ in (4) represents the small-scale fluctuations of the shadow fading observed at CR r, which is uncorrelated with $\epsilon_r(\mathbf{u},\tau)$ in (3) for all $\mathbf{u}$ and τ; while $\eta_r(\mathbf{x},t)$ in (5) captures the system noise that is uncorrelated with $\nu_r(\mathbf{u},\tau)$ or $\epsilon_r(\mathbf{u},\tau)$, $\forall\mathbf{u},\tau$. Moreover, $\mathbb{E}\{\nu_r(\mathbf{x},t)\bar{s}_r(\mathbf{u},t)\}=\mathbb{E}\{\eta_r(\mathbf{x},t)\bar{s}_r(\mathbf{u},t-1)\}=0$, for all $\mathbf{x}$, $\mathbf{u}$ and t. For stability, it must further hold that $|\int_{\mathcal{A}} w_r(\mathbf{x},\mathbf{u})d\mathbf{u}|<1$.

Remark 1: A recent measurement campaign in the 2.4-GHz band for nomadic as well as mobile distributed channels confirmed that shadow fading can be accurately modeled by a log-normal-distributed first-order spatio-temporal autoregressive process [20]. One can draw analogy between the shadow fading model used here, and the model set forth in [20] as follows: $\bar{s}_r(\mathbf{x},t)$ captures the static shadowing component (both the obstruction-based as well as the multipath-based static shadowing), and the temporally correlated dynamic shadowing component; while $\nu_r(\mathbf{x},t)$ represents the dynamic shadowing component that has spatial structure, but no temporal correlation (e.g., due to a large inter-measurement spacing, compared to the time scale of the small-scale movements in the environment, which cause dynamic shadowing.) The spatial correlation model of the latter term is generic in the present framework, and the model parameters are estimated from training; see Section IV.

The state-space model in (4) and (5) is infinite-dimensional. A common approach to reduce its dimensionality is to employ a basis-expansion representation [14], [21]. Let $\{\psi_k(\cdot)\}_{k=1}^{\infty}$ denote a set of complete pre-specified orthonormal bases defined on $\mathcal{A}$. Then $\bar{s}_r(\cdot)$ and $w_r(\cdot)$ can be expressed as

$$\bar{s}_r(\mathbf{x},t)=\sum_{k=1}^{\infty}\alpha_{r,k}(t)\psi_k(\mathbf{x}) \tag{8}$$

$$w_r(\mathbf{x},\mathbf{u})=\sum_{k=1}^{\infty}\beta_{r,k}(\mathbf{x})\psi_k(\mathbf{u}) \tag{9}$$

where $\{\alpha_{r,k}(t)\}$ and $\{\beta_{r,k}(\mathbf{x})\}$ are the basis-expansion coefficients for $\bar{s}_r$ and w_r, respectively.

By retaining only the K dominant components of the expansion, and introducing the $K\times 1$ vectors $\boldsymbol{\alpha}_r(t)\triangleq[\alpha_{r,1}(t)\ldots\alpha_{r,K}(t)]^T$, $\boldsymbol{\beta}_r(\mathbf{x})\triangleq[\beta_{r,1}(\mathbf{x}),\ldots,\beta_{r,K}(\mathbf{x})]^T$, and $\boldsymbol{\psi}(\mathbf{x})\triangleq[\psi_1(\mathbf{x}),\ldots,\psi_K(\mathbf{x})]^T$, the state evolution in (5) can be rewritten as

$$\boldsymbol{\psi}^T(\mathbf{x})\boldsymbol{\alpha}_r(t)=\boldsymbol{\beta}_r^T(\mathbf{x})\boldsymbol{\alpha}_r(t-1)+\eta_r(\mathbf{x},t). \tag{10}$$

Upon defining

$$\mathbf{B}_r\triangleq[\boldsymbol{\beta}_r(\mathbf{x}_1),\ldots,\boldsymbol{\beta}_r(\mathbf{x}_{r-1}),\boldsymbol{\beta}_r(\mathbf{x}_{r+1}),\ldots,\boldsymbol{\beta}_r(\mathbf{x}_{N_r})]^T$$

$$\boldsymbol{\Psi}_r\triangleq[\boldsymbol{\psi}(\mathbf{x}_1),\ldots,\boldsymbol{\psi}(\mathbf{x}_{r-1}),\boldsymbol{\psi}(\mathbf{x}_{r+1}),\ldots,\boldsymbol{\psi}(\mathbf{x}_{N_r})]^T$$

$$\boldsymbol{\eta}_r(t)\triangleq[\eta_r(\mathbf{x}_1,t),\ldots,\eta_r(\mathbf{x}_{r-1},t),\eta_r(\mathbf{x}_{r+1},t),\ldots,\eta_r(\mathbf{x}_{N_r},t)]^T$$

and sampling (10) at positions $\{\mathbf{x}_1,\ldots,\mathbf{x}_{r-1},\mathbf{x}_{r+1},\ldots,\mathbf{x}_{N_r}\}$, the evolution of the state per CR r can be expressed as

$$\boldsymbol{\Psi}_r\boldsymbol{\alpha}_r(t)=\mathbf{B}_r\boldsymbol{\alpha}_r(t-1)+\boldsymbol{\eta}_r(t). \tag{11}$$

Consequently, assuming that $N_r-1\geq K$ and $\boldsymbol{\Psi}_r^T\boldsymbol{\Psi}_r$ is non-singular, the state equation

$$\boldsymbol{\alpha}_r(t)=\mathbf{T}_r\boldsymbol{\alpha}_r(t-1)+\boldsymbol{\Psi}_r^{\dagger}\boldsymbol{\eta}_r(t) \tag{12}$$

can be obtained, where $\boldsymbol{\Psi}_r^{\dagger}\triangleq(\boldsymbol{\Psi}_r^T\boldsymbol{\Psi}_r)^{-1}\boldsymbol{\Psi}_r^T$ is the pseudo-inverse of $\boldsymbol{\Psi}_r$ and $\mathbf{T}_r\triangleq\boldsymbol{\Psi}_r^{\dagger}\mathbf{B}_r$.

Substituting (11) into (4), the measurement equation in (3) can be expressed in vector-matrix form as

$$\breve{\mathbf{s}}_r(t)=\boldsymbol{\Psi}_r\boldsymbol{\alpha}_r(t)+\boldsymbol{\nu}_r(t)+\boldsymbol{\epsilon}_r(t) \tag{13}$$

where $\boldsymbol{\nu}_r(t)$ and $\boldsymbol{\epsilon}_r(t)$ are the appropriate vectorizations of $\{\nu_r(\mathbf{x}_j,t)\}_{j=1,j\neq r}^{N_r}$ and $\{\epsilon_r(\mathbf{x}_j,t)\}_{j=1,j\neq r}^{N_r}$, respectively.

From (12) and (13), a space–time Kalman filter can be derived as explained next to estimate $\boldsymbol{\alpha}_r(t)$, and thus $\bar{s}_r(\mathbf{x},t)$, $s_r(\mathbf{x},t)$, and hence $G_r(\mathbf{x},t)$ for any arbitrary position $\mathbf{x}$ at time t.

B. Spatio-Temporal KKF

Given the state (12) and the measurement (13), the minimum mean-square error (MMSE) estimate of the state vector $\boldsymbol{\alpha}_r(t)$ at time t can be obtained via ordinary Kalman filtering. Let $\mathbf{C}_{\nu_r}$

and $\mathbf{C}_{\epsilon_r}$ denote the covariances of $\boldsymbol{\nu}_r(t)$ and $\boldsymbol{\epsilon}_r(t)$, respectively, and define $\boldsymbol{\Sigma}_r \triangleq \mathbf{C}_{\nu_r} + \mathbf{C}_{\epsilon_r}$. Also, let matrix $\breve{\mathbf{S}}_r(t)$ denote the cumulative observations $\{\breve{\mathbf{s}}_r(1), \ldots, \breve{\mathbf{s}}_r(t)\}$. Then, the Kalman filtering equations are given by (see, e.g., [22, Ch. 3])

$$\begin{aligned}\hat{\boldsymbol{\alpha}}_r(t|t-1) &\triangleq \mathbb{E}\left\{\boldsymbol{\alpha}_r(t)|\breve{\mathbf{S}}_r(t-1)\right\} \\ &= \mathbf{T}_r\hat{\boldsymbol{\alpha}}_r(t-1|t-1) \qquad (14)\\ \mathbf{P}_r(t|t-1) &\triangleq \operatorname{cov}\left\{\boldsymbol{\alpha}_r(t)|\breve{\mathbf{S}}_r(t-1)\right\} \\ &= \mathbf{T}_r\mathbf{P}_r(t-1|t-1)\mathbf{T}_r^T + \boldsymbol{\Psi}_r^{\dagger}\mathbf{C}_{\eta_r}\boldsymbol{\Psi}_r^{\dagger T} \qquad (15)\\ \hat{\boldsymbol{\alpha}}_r(t|t) &\triangleq \mathbb{E}\left\{\boldsymbol{\alpha}_r(t)|\breve{\mathbf{S}}_r(t)\right\} \\ &= \hat{\boldsymbol{\alpha}}_r(t|t-1) + \mathbf{K}_r(t)\left[\breve{\mathbf{s}}_r(t) - \boldsymbol{\Psi}_r\hat{\boldsymbol{\alpha}}_r(t|t-1)\right] \qquad (16)\\ \mathbf{P}_r(t|t) &\triangleq \operatorname{cov}\left\{\boldsymbol{\alpha}_r(t)|\breve{\mathbf{S}}_r(t)\right\} \\ &= \mathbf{P}_r(t|t-1) + \mathbf{K}_r(t)\boldsymbol{\Psi}_r\mathbf{P}_r(t|t-1) \qquad (17)\end{aligned}$$

where the Kalman gain $\mathbf{K}_r(t)$ at time t is given by

$$\mathbf{K}_r(t) = \mathbf{P}_r(t|t-1)\boldsymbol{\Psi}_r^T\left(\boldsymbol{\Sigma}_r + \boldsymbol{\Psi}_r\mathbf{P}_r(t|t-1)\boldsymbol{\Psi}_r^T\right)^{-1}. \quad (18)$$

The Kalman filter-estimate $\hat{\boldsymbol{\alpha}}_r(t|t)$ can thus be used to track the temporally dynamic component $\bar{s}_r(\mathbf{x},t)$ of $s_r(\mathbf{x},t)$ in (4) through $\mathbb{E}\{\bar{s}_r(\mathbf{x},t)|\breve{\mathbf{S}}_r(t)\} = \boldsymbol{\psi}^T(\mathbf{x})\hat{\boldsymbol{\alpha}}_r(t|t)$. To capture the component corresponding to $\nu_r(\mathbf{x},t)$, which is temporally white yet spatially colored, one needs to augment the Kalman filter-based estimator with a linear spatial interpolator. The overall KKF estimator is described in the following proposition; see also [21, Lemma 4.2].

Proposition 1: Conditioned on the measurements $\breve{\mathbf{S}}_r(t)$, the shadow fading process $s_r(\mathbf{x},t)$ is Gaussian distributed with mean and variance given, respectively, by

$$\begin{aligned}\hat{s}_r(\mathbf{x},t) &\triangleq \mathbb{E}\left\{s_r(\mathbf{x},t)|\breve{\mathbf{S}}_r(t)\right\} \\ &= \boldsymbol{\psi}^T(\mathbf{x})\hat{\boldsymbol{\alpha}}_r(t|t) + \mathbf{c}_{\nu_r}^T(\mathbf{x})\boldsymbol{\Sigma}_r^{-1} \\ &\quad\times\left[\breve{\mathbf{s}}_r(t) - \boldsymbol{\Psi}_r\hat{\boldsymbol{\alpha}}_r(t|t)\right] \qquad (19)\\ \operatorname{var}\left\{s_r(\mathbf{x},t)|\breve{\mathbf{S}}_r(t)\right\} &= \sigma_{\nu_r}^2 - \mathbf{c}_{\nu_r}^T(\mathbf{x})\boldsymbol{\Sigma}_r^{-1}\mathbf{c}_{\nu_r}(\mathbf{x}) \\ &\quad+\left[\boldsymbol{\psi}^T(\mathbf{x}) - \mathbf{c}_{\nu_r}^T(\mathbf{x})\boldsymbol{\Sigma}_r^{-1}\boldsymbol{\Psi}_r\right]\mathbf{P}_r(t|t) \\ &\quad\times\left[\boldsymbol{\psi}(\mathbf{x}) - \boldsymbol{\Psi}_r^T\boldsymbol{\Sigma}_r^{-1}\mathbf{c}_{\nu_r}(\mathbf{x})\right] \qquad (20)\end{aligned}$$

where $\mathbf{c}_{\nu_r}(\mathbf{x}) \triangleq \mathbb{E}\{\nu_r(\mathbf{x},t)\boldsymbol{\nu}_r(t)\}$ and $\sigma_{\nu_r}^2 \triangleq \operatorname{var}\{\nu_r(\mathbf{x},t)\}$.

Proof: See Appendix A. □

Note that (19) provides the MMSE estimate $\hat{s}_r(\mathbf{x},t)$ of the shadow fading process at any $\mathbf{x}$ and t, using only the preselected bases in $\boldsymbol{\Psi}_r$ and the estimates $\{\breve{\mathbf{s}}_r(\tau)\}_{\tau=1}^t$, acquired at the CRs during the training phase.

Given the KKF estimate $\hat{s}_r(\mathbf{x},t)$, the CG map estimate $\hat{G}_r(\mathbf{x},t)$ for an arbitrary $\mathbf{x} \in \mathcal{A}$ can be readily constructed by adding back the path loss component as [cf. (2)]

$$\hat{G}_r(\mathbf{x},t) = G_0 - 10\gamma\log_{10}\left(\|\mathbf{x}-\mathbf{x}_r\|_2\right) + \hat{s}_r(\mathbf{x},t). \quad (21)$$

C. KKF With Measurement Losses

As mentioned in Section II, the shadow fading measurements $\{\breve{\mathbf{s}}_r(t)\}$ can be obtained using training signals exchanged with the other CRs in the network. Since the training signals also need to respect the PU-CR hierarchy, such measurements cannot be acquired if the PUs are continuously active over an extended period of time. Measurement misses may also occur when the control channel is congested. In such cases, the KKF update must be performed open-loop without any measurements. Specifically, if the measurement is missing at time t, the prediction step of Kalman filtering is the same as in (14) and (15), but the correction step in (16) and (17) is replaced by

$$\begin{aligned}\hat{\boldsymbol{\alpha}}_r(t|t) &= \hat{\boldsymbol{\alpha}}_r(t|t-1) \qquad (22)\\ \mathbf{P}_r(t|t) &= \mathbf{P}_r(t|t-1). \qquad (23)\end{aligned}$$

The KKF estimate at time t then becomes

$$\begin{aligned}\hat{s}_r(\mathbf{x},t) &= \boldsymbol{\psi}^T(\mathbf{x})\hat{\boldsymbol{\alpha}}_r(t|t) \qquad (24)\\ \operatorname{var}\left\{s_r(\mathbf{x},t)|\breve{\mathbf{S}}_r(t)\right\} &= \sigma_{\nu_r}^2 + \boldsymbol{\psi}^T(\mathbf{x})\mathbf{P}_r(t|t)\boldsymbol{\psi}(\mathbf{x}). \qquad (25)\end{aligned}$$

It should be noted that under proper stability conditions, the measurement losses over an extended period will eventually bring the KKF estimate $\hat{s}_r(\mathbf{x},t)$ down to 0. In this case, the KKF-based model falls back to the path loss-only model [cf. (2)]. This is a nice safety feature ensuring that the proposed algorithms perform no worse than the alternatives which account only for path-loss effects. Note also that it is natural to initialize the CG map estimates with the path-loss-only map.

Remark 2: It should be noted that in the usual deployment scenarios for CRs, the probability of PU presence is very low. Moreover, shadowing typically varies very slowly compared to the coherence time of the PU activities. These considerations indicate that a prolonged outage of measurements should be a rare event.

IV. Estimation of Model Parameters

The KKF presented in Section III-B is optimal only when exact knowledge of the model covariances $\mathbf{C}_{\epsilon_r}, \mathbf{C}_{\nu_r}, \mathbf{C}_{\eta_r} \triangleq \operatorname{cov}\{\boldsymbol{\eta}_r\}$, and the state transition matrix $\mathbf{T}_r$ in (12) and (13) are all available. For the interpolation step, one also needs the cross-covariance $C_{\nu_r}(\mathbf{x},\mathbf{x}_0) \triangleq \operatorname{cov}\{\nu_r(\mathbf{x},t),\nu_r(\mathbf{x}_0,t)\}$ for arbitrary $\mathbf{x} \in \mathcal{A}$ and $\mathbf{x}_0 \in \{\mathbf{x}_j\}_{j=1,j\neq r}^{N_r}$ [cf. (19) and (20)].

In this paper, an empirical Bayesian approach is pursued to estimate the required parameters in (12) and (13) from the data, and use them in the KKF recursions. A fully Bayesian hierarchical method was advocated in [23], where the priors on the model parameters were estimated through a Gibbs sampler. In [14], the standard method of moments was employed to obtain the necessary estimates. This latter approach is adopted here for simplicity. To estimate $C_{\nu_r}(\cdot)$, however, a model-based approach is pursued to account for the fact that measurements are made at random positions in space.

A. Estimation of Model Covariances

Supposing that the measurement noise $\epsilon_r(\mathbf{x},t)$ is white in space and time, it follows that $\mathbf{C}_{\epsilon_r}=\sigma_\epsilon^2\mathbf{I}$, where σ_ϵ^2 can be obtained during the calibration process of the measurement device. The estimate $\hat{\mathbf{C}}_{\breve{s}_r}$ is readily found using the sample covariance of the data $\breve{s}_r(t)$. Then, the estimate of $\mathbf{C}_{s_r}$ is obtained as $\hat{\mathbf{C}}_{s_r}=\hat{\mathbf{C}}_{\breve{s}_r}-\mathbf{C}_{\epsilon_r}$. Since $\bar{s}_r(t)$ belongs to the subspace spanned by the columns of $\mathbf{\Psi}_r$, it is possible to estimate $\mathbf{C}_{\bar{s}_r}$ by a projection operation as

$$\hat{\mathbf{C}}_{\bar{s}_r}=\mathbf{\Psi}_r\mathbf{\Psi}_r^\dagger\hat{\mathbf{C}}_{s_r}\mathbf{\Psi}_r^{\dagger T}\mathbf{\Psi}_r^T. \tag{26}$$

Similarly, $\hat{\mathbf{C}}_{\nu_r}$ can be obtained as

$$\hat{\mathbf{C}}_{\nu_r}=\left(\mathbf{I}-\mathbf{\Psi}_r\mathbf{\Psi}_r^\dagger\right)\hat{\mathbf{C}}_{s_r}\left(\mathbf{I}-\mathbf{\Psi}_r\mathbf{\Psi}_r^\dagger\right)^T. \tag{27}$$

The covariance $C_{\nu_r}(\mathbf{x},\mathbf{x}_0)$ can be estimated for an arbitrary $\mathbf{x}\in\mathcal{A}$ and $\mathbf{x}_0\in\{\mathbf{x}_j\}_{j=1,j\neq r}^{N_r}$, using the exponential class of covariance functions, see also [24, p. 56]

$$C_{\nu_r}(\mathbf{x},\mathbf{x}_0)=\sigma_{\nu_r}^2e^{-\frac{\|\mathbf{x}-\mathbf{x}_0\|_2}{d_{\nu_r}}} \tag{28}$$

where $\sigma_{\nu_r}^2$ and d_{ν_r} are the variance and the coherence distance of ν_r, respectively. These parameters can be obtained in closed form by LS fitting the estimated $\hat{\mathbf{C}}_{\nu_r}$, as detailed in Appendix B.

Alternatively, the approach of [25] can be used, which can fit the covariances with negative values better. Specifically, a class of covariance functions with the form

$$C_{\nu_r}(\mathbf{x},\mathbf{x}_0)=\sum_{m=1}^M\cos\left(\|\mathbf{x}-\mathbf{x}_0\|_2u_m\right)v_m \tag{29}$$

is considered, where $u_m\geq 0$ and $v_m\geq 0$ are the parameters to be chosen. Stack the entries in the diagonal and the lower-triangular part of $\hat{\mathbf{C}}_{\nu_r}$ into a vector $\mathbf{c}\triangleq[c_1\ \ldots\ c_N]^T$ with dimension $N\triangleq N_r(N_r+1)/2$. Similarly, the diagonal and the lower-triangular part of a matrix $\mathbf{D}$ whose (i,j)th entry is $\|\mathbf{x}_i-\mathbf{x}_j\|_2$, is collected into a vector $\mathbf{d}\triangleq[d_1\ \ldots\ d_N]^T$. Given a choice of $\{u_m\}$, the values of $\{v_m\}$ are obtained by solving the constrained LS fitting problem

$$\min_{\{v_m\geq 0\}_{m=1}^M}\sum_{n=1}^N\left[c_n-\sum_{m=1}^M\cos(d_nu_m)v_m\right]^2 \tag{30}$$

$$\text{subject to}\ \sum_{m=1}^M u_mv_m\leq\varpi \tag{31}$$

where $\varpi>0$. The constraint in (31) ensures that the fitted covariance function is smooth by imposing a bound on its first derivative. For simplicity, $u_m=\Delta m$ for $m=1,2,\ldots,M$, will be used in the numerical tests with $\Delta>0$.

Before estimating $\mathbf{C}_{\eta_r}$, observe that the KKF recursion (15) requires only an estimate of the product $\mathbf{\Psi}_r^\dagger\mathbf{C}_{\eta_r}\mathbf{\Psi}_r^{\dagger T}$, and not an estimate of $\mathbf{C}_{\eta_r}$ in isolation. Define $\mathbf{C}_{\alpha_r}\triangleq\mathbb{E}\{\boldsymbol{\alpha}_r(t)\boldsymbol{\alpha}_r^T(t)\}$ and $\mathbf{C}_{\alpha_r}^{(1)}\triangleq\mathbb{E}\{\boldsymbol{\alpha}_r(t)\boldsymbol{\alpha}_r^T(t-1)\}$. Then, note that $\mathbf{C}_{\bar{s}_r}=\mathbf{\Psi}_r\mathbf{C}_{\alpha_r}\mathbf{\Psi}_r^T$, and that $\mathbf{C}_{\breve{s}_r}^{(1)}\triangleq\mathbb{E}\{\breve{\mathbf{s}}_r(t)\breve{\mathbf{s}}_r^T(t-1)\}$ is given by [cf. (13)]

$$\begin{aligned}\mathbf{C}_{\breve{s}_r}^{(1)}=&\mathbb{E}\left\{[\mathbf{\Psi}_r\boldsymbol{\alpha}_r(t)+\boldsymbol{\nu}_r(t)+\boldsymbol{\epsilon}_r(t)]\right.\\&\left.\times[\mathbf{\Psi}_r\boldsymbol{\alpha}_r(t-1)+\boldsymbol{\nu}_r(t-1)+\boldsymbol{\epsilon}_r(t-1)]^T\right\}\quad(32)\\=&\mathbf{\Psi}_r\mathbf{C}_{\alpha_r}^{(1)}\mathbf{\Psi}_r^T.\quad(33)\end{aligned}$$

Thus, $\mathbf{C}_{\alpha_r}=\mathbf{\Psi}_r^\dagger\mathbf{C}_{\bar{s}_r}\mathbf{\Psi}_r^{\dagger T}$ and $\mathbf{C}_{\alpha_r}^{(1)}=\mathbf{\Psi}_r^\dagger\mathbf{C}_{\breve{s}_r}^{(1)}\mathbf{\Psi}_r^{\dagger T}$ hold; hence, [cf. (12)]

$$\begin{aligned}\mathbf{\Psi}_r^\dagger\mathbf{C}_{\eta_r}\mathbf{\Psi}_r^{\dagger T}=&\mathbb{E}\left\{[\boldsymbol{\alpha}_r(t)-\mathbf{T}_r\boldsymbol{\alpha}_r(t-1)]\right.\\&\left.\times[\boldsymbol{\alpha}_r(t)-\mathbf{T}_r\boldsymbol{\alpha}_r(t-1)]^T\right\}\quad(34)\\=&\mathbf{\Psi}_r^\dagger\mathbf{C}_{\bar{s}_r}\mathbf{\Psi}_r^{\dagger T}-\mathbf{\Psi}_r^\dagger\mathbf{C}_{\breve{s}_r}^{(1)}\mathbf{\Psi}_r^{\dagger T}\mathbf{T}_r^T\\&-\mathbf{T}_r\mathbf{\Psi}_r^\dagger\mathbf{C}_{\breve{s}_r}^{(1)}\mathbf{\Psi}_r^\dagger+\mathbf{T}_r\mathbf{\Psi}_r^\dagger\mathbf{C}_{\bar{s}_r}\mathbf{\Psi}_r^{\dagger T}\mathbf{T}_r^T.\quad(35)\end{aligned}$$

Matrix $\mathbf{\Psi}_r^\dagger\mathbf{C}_{\eta_r}\mathbf{\Psi}_r^{\dagger T}$ can now be estimated by plugging into (35) the expression for $\hat{\mathbf{C}}_{\bar{s}_r}$ found as in (26), and the estimate $\hat{\mathbf{C}}_{\breve{s}_r}^{(1)}$ obtained via sample averaging.

B. Estimation of the State Transition Matrix

To estimate $\mathbf{T}_r$, note first that [cf. (11) and (13)]

$$\begin{aligned}\mathbf{C}_{\breve{s}_r}^{(1)}=&\mathbb{E}\left\{[\mathbf{B}_r\boldsymbol{\alpha}_r(t-1)+\boldsymbol{\eta}_r(t)+\boldsymbol{\nu}_r(t)+\boldsymbol{\epsilon}_r(t)]\right.\\&\left.\times[\mathbf{\Psi}_r\boldsymbol{\alpha}_r(t-1)+\boldsymbol{\eta}_r(t-1)+\boldsymbol{\nu}_r(t-1)+\boldsymbol{\epsilon}_r(t-1)]^T\right\}\\=&\mathbf{B}_r\mathbf{\Psi}_r^\dagger\mathbf{C}_{\bar{s}_r}\mathbf{\Psi}_r^{\dagger T}\mathbf{\Psi}_r^T\end{aligned}\tag{36}$$

where the relation $\mathbf{C}_{\alpha_r}=\mathbf{\Psi}_r^\dagger\mathbf{C}_{\bar{s}_r}\mathbf{\Psi}_r^{\dagger T}$ is again used. Thus, an estimate of $\mathbf{B}_r$ is obtained as

$$\hat{\mathbf{B}}_r=\hat{\mathbf{C}}_{\breve{s}_r}^{(1)}\mathbf{\Psi}_r^{\dagger T}\left[\mathbf{\Psi}_r^\dagger\hat{\mathbf{C}}_{\bar{s}_r}\mathbf{\Psi}_r^{\dagger T}\right]^{-1} \tag{37}$$

and $\hat{\mathbf{T}}_r=\mathbf{\Psi}_r^\dagger\hat{\mathbf{B}}_r$.

Recall that $\mathbf{\Psi}_r$ (and hence $\mathbf{\Psi}_r^\dagger$) is formed using preselected basis functions. One such basis can be constructed easily for a rectangular area $\mathcal{A}$ using the set of Legendre polynomials in two variables [26, Ch. 2].

Table I summarizes the overall algorithm implementing the step **S1**) for tracking the CG map $G_r(\mathbf{x},t)$ at each CR r across time t, for any arbitrary point of interest $\mathbf{x}$ in the geographical area $\mathcal{A}$.

V. Sparsity-Aware Cooperative Spectrum Sensing

Based on the map estimates $\{\hat{G}_r(\mathbf{x},t)\}$ obtained via KKF, this section shows how to estimate locations and transmit-powers of the PUs in the geographical area $\mathcal{A}$. This task can be accomplished by solving a linear regression problem collaboratively among all the CRs, while exploiting the inherent sparsity present.

Consider a set of N_s *candidate* PU source locations $\mathcal{S}\triangleq\{\mathbf{x}_s\in\mathcal{A}\}_{s=1}^{N_s}$. Without prior knowledge of the potential PU positions, $\mathcal{S}$ can be formed simply by discretizing the area $\mathcal{A}$ into a set of grid points. Let $\mathbf{p}(t)\triangleq[p_1(t)\ \ldots\ p_{N_s}(t)]^T\in(\mathbb{R}^+)^{N_s}$, where $p_s(t)>0$ implies presence of an active PU transmitter at

TABLE I
SUMMARY OF THE MAP TRACKING ALGORITHM

Acquisition of observation $\check{\mathbf{s}}_r(t)$
During time interval $[t, t+1)$ in the j-th slot, CR $j \neq r$ transmits a training signal with power $p_j(t)$ in a TDMA fashion.
CR r measures the received power $\pi_r^{(j)}(t)$ in slot $j \neq r$ and computes the channel gain estimate by $\hat{g}_r(\mathbf{x}_j, t) = \pi_r^{(j)}(t)/p_j(t)$.
Compute $\hat{G}_r(\mathbf{x}_j, t) = 10 \log_{10} \hat{g}_r(\mathbf{x}_j, t)$.
Compute $\check{s}_r(\mathbf{x}_j, t) = \hat{G}_r(\mathbf{x}_j, t) - G_0 + 10\gamma \log_{10} \|\mathbf{x}_r - \mathbf{x}_j\|_2$.
Collect $\check{s}_r(\mathbf{x}_j, t)$, $j \in \{1, \ldots, r-1, r+1, \ldots, N_r\}$ and form $\check{\mathbf{s}}_r(t)$.
Model parameter estimation
For each $r \in \{1, \ldots, N_r\}$ and given $\mathbf{\Psi}_r$ and $\mathbf{C}_{\epsilon_r}$:
Obtain $\hat{\mathbf{C}}_{\check{s}_r}$ and $\hat{\mathbf{C}}_{\check{s}_r}^{(1)}$ from sample covariances.
Compute $\hat{\mathbf{C}}_{s_r} = \hat{\mathbf{C}}_{\check{s}_r} - \mathbf{C}_{\epsilon_r}$.
Obtain $\hat{\mathbf{C}}_{\bar{s}_r}$ from (26) and $\hat{\mathbf{C}}_{\nu_r}$ from (27).
Compute $\hat{\mathbf{c}}_{\nu_r}(\mathbf{x})$ by model fitting; see Sec. IV-A
Compute $\hat{\mathbf{B}}_r$ from (38) and set $\hat{\mathbf{T}}_r = \mathbf{\Psi}_r^{\dagger} \hat{\mathbf{B}}_r$.
Obtain an estimate of $\mathbf{\Psi}_r^{\dagger} \mathbf{C}_{\eta_r} \mathbf{\Psi}_r^{\dagger T}$ from (35).
Map tracking via KKF
1: Initialize $t = 0$
2: If PUs are active, update KF open-loop by (14)–(15) and (22)–(23). Compute the KKF estimate $\hat{s}_r(\mathbf{x}, t)$ by (24). Go to Step 4.
3: If PUs are inactive, acquire the measurement $\check{\mathbf{s}}_r(t)$. Update KF by (14)–(17). Compute the KKF estimate $\hat{s}_r(\mathbf{x}, t)$ by (19).
4: Increment t and go to Step 2.

location $\mathbf{x}_s$ at time t, while $p_s(t) = 0$ the absence of the same. With $\mathbf{g}_r(t) \triangleq [g_r(\mathbf{x}_1, t) \ \ldots \ g_r(\mathbf{x}_{N_s}, t)]^T$, the power measurement (1) can be compactly rewritten as

$$\pi_r(t) = \mathbf{g}_r^T(t)\mathbf{p}(t) + z_r(t), \quad r = 1, 2, \ldots, N_r. \tag{38}$$

Since the number of active transmitters over the same spectral band in a given geographical area is limited due to mutual interference, the number of nonzero entries of $\mathbf{p}(t)$ at a given time t is far smaller than N_s for large N_s; that is, the vector $\mathbf{p}(t)$ is *sparse*. Motivated by the recent advances in compressive sampling [15], the spectrum sensing problem has been formulated as a sparse regression problem with an ℓ_1-norm-based regularization term in [16]. Here, this formulation is considerably broadened to account not only for time-varying PU activities but also for spatio-temporal shadow fading propagation effects.

In order to track the time-varying PU activities, the following time-weighted non-negative Lasso formulation is considered:

$$\hat{\mathbf{p}}(t) = \arg\min_{\mathbf{p} \succeq \mathbf{0}} J_t(\mathbf{p})$$
$$J_t(\mathbf{p}) \triangleq \left[\frac{1}{2} \sum_{\tau=1}^{t} \mu^{t-\tau} \sum_{r=1}^{N_r} \left(\pi_r(\tau) - \mathbf{g}_r^T(\tau)\mathbf{p} \right)^2 + \lambda_t \|\mathbf{p}\|_1 \right] \tag{39}$$

where $\mu \in (0, 1]$ denotes the forgetting factor, $|\mathbf{p}||_1 \triangleq \sum_{s=1}^{N_s} |p_s|$, and $\lambda_t > 0$ is a tuning parameter that controls the sparsity of the solution. Albeit non-differentiable, the problem in (39) is convex; thus, it can be solved efficiently using standard quadratic programming iterations (e.g., the Sedumi solver [27]). However, solving it in a batch fashion would incur considerable overhead in terms of computational complexity and memory requirement. In the ensuing subsections, *recursive* algorithms are developed, both in centralized and distributed formats.

A. Centralized Algorithm for PU Tracking

In a centralized setup, the power measurements and the CG estimates from all cooperating CRs are collected at a central unit [a.k.a. fusion center (FC)], which may be either one of the CR nodes, or a separate control node. To track PU activities centrally at the FC, one can employ the adaptive Lasso algorithm of [17] that is termed online coordinate descent (OCD)-Lasso. OCD performs cyclic iterative minimization of the time-weighted Lasso cost, with respect to one entry of $\mathbf{p}$ per iteration. We employ the OCD-Lasso by properly adapting it to the vector observation case, and by imposing the non-negativity constraints on the PU transmission powers.

For notational convenience, let us express the time index as $t = iN_s + n$, with $n \in \{1, \ldots, N_s\}$ corresponding to the entry of $\hat{\mathbf{p}}(t)$ updated at time t, and $i = \lceil t/N_s \rceil - 1$ representing the number of times that the nth entry $\hat{p}_n(t)$ of $\hat{\mathbf{p}}(t)$ has been updated. With $J_t(\mathbf{p})$ denoting the objective function of the optimization problem in (39), the cyclic coordinate descent update can be expressed as

$$\hat{p}_n(t) = \arg\min_{\hat{p}_n \geq 0} J_t([\hat{p}_1(t-1), \ldots, \hat{p}_{n-1}(t-1), \hat{p}_n, \hat{p}_{n+1}(t-1), \ldots, \hat{p}_{N_s}(t-1)]) \tag{40}$$
$$\hat{p}_j(t) = \hat{p}_j(t-1) \qquad \forall j \neq n. \tag{41}$$

Now define

$$\mathbf{R}(t) \triangleq \sum_{\tau=1}^{t} \mu^{t-\tau} \sum_{r=1}^{N_r} \mathbf{g}_r(\tau)\mathbf{g}_r^T(\tau) = \mu\mathbf{R}(t-1) + \sum_{r=1}^{N_r} \mathbf{g}_r(t)\mathbf{g}_r^T(t) \tag{42}$$

$$\mathbf{r}(t) \triangleq \sum_{\tau=1}^{t} \mu^{t-\tau} \sum_{r=1}^{N_r} \pi_r(\tau)\mathbf{g}_r(\tau) = \mu\mathbf{r}(t-1) + \sum_{r=1}^{N_r} \pi_r(t)\mathbf{g}_r(t). \tag{43}$$

Then, the update in (40) is equivalent to

$$\hat{p}_n(t) = \arg\min_{\hat{p}_n \geq 0} \frac{1}{2} R_{n,n}(t)\hat{p}_n^2 - \tilde{r}_n(t)\hat{p}_n + \lambda_t |\hat{p}_n| \tag{44}$$

where $R_{n,j}(t)$ and $r_n(t)$ denote the (n, j)th entry of $\mathbf{R}(t)$, and the nth entry of $\mathbf{r}(t)$, respectively, and

$$\tilde{r}_n(t) \triangleq r_n(t) - \sum_{j=1, j \neq n}^{N_s} R_{n,j}(t)\hat{p}_j(t-1). \tag{45}$$

The solution of (44) is obtained in closed form as

$$\hat{p}_n(t) = \frac{[\tilde{r}_n(t) - \lambda_t]_+}{R_{n,n}(t)} \tag{46}$$

where $[a]_+ \triangleq \max\{a, 0\}$.

B. Distributed Algorithm for PU Tracking

In certain cases, a distributed algorithm may be more desirable than a centralized implementation, due to scalability and robustness issues. A distributed algorithm does not require all the measurements to be collected at a single processor,

but rather performs consensus-based in-network processing, which requires only local message-passing among single-hop neighboring nodes in the CR network [28].

Let $\mathcal{N}_r \subset \{1,\ldots,r-1,r+1,\ldots,N_r\}$ denote the set of one-hop neighbors of CR r. It is assumed that the links in the CR network are bidirectional; i.e., $\varrho \in \mathcal{N}_r$ implies $r \in \mathcal{N}_\varrho$. The idea is to define local copies $\mathbf{p}_r(t) \triangleq [p_{r,1}(t),\ldots,p_{r,N_s}(t)]^T$, $r=1,\ldots,N_r$, of the global $\mathbf{p}(t)$ and solve a constrained optimization problem where local copies in the one-hop neighborhood are enforced to be coherent; i.e., constant. Under the assumption that the network is connected, i.e., that there are paths from any node in the network to any other nodes, the localized problem is equivalent to the original centralized problem.

Specifically, consider the coordinate descent update at time t for the nth coordinate of $\hat{\mathbf{p}}(t)$ given in (40). Define the per CR (i.e., local) cost function $\mathcal{J}_{r,t}(\mathbf{p}_r)$ as [cf. (39)]

$$\mathcal{J}_{r,t}(\mathbf{p}_r) \triangleq \frac{1}{2}\sum_{\tau=1}^{t}\mu^{t-\tau}\left(\pi_r(\tau)-\mathbf{g}_r^T(\tau)\mathbf{p}_r\right)^2+\frac{\lambda_t}{N_r}||\mathbf{p}_r||_1. \quad (47)$$

Then, the following formulation, which is equivalent to (40), is amenable to distributed implementation:

$$\begin{aligned}&\{\hat{p}_{r,n}(t)\}_{r=1}^{N_r}\\ &= \arg\min_{\substack{\hat{p}_{r,n}\geq 0,\\ r=1,\ldots,N_r}}\sum_{r=1}^{N_r}\mathcal{J}_{r,t}\Big([\hat{p}_{r,1}(t-1),\ldots,\hat{p}_{r,n-1}(t-1),\\ &\qquad \hat{p}_{r,n},\hat{p}_{r,n+1}(t-1),\ldots,\hat{p}_{r,N_s}(t-1)]^T\Big)\\ &\text{subject to } \hat{p}_{r,n}=\hat{p}_{\varrho,n},\ \forall \varrho\in\mathcal{N}_r,\ \ r=1,2,\ldots,N_r. \quad (48)\end{aligned}$$

Using the provably convergent alternating direction method of multipliers (ADMoM) [29, p. 253], we will develop a solver of (48), which converges to $\{\hat{p}_{r,n}(t)\}_{r=1}^{N_r}$ using only local message-passing. To this end, and similar to the centralized algorithm, define

$$\begin{aligned}\mathbf{R}_r(t) &\triangleq \sum_{\tau=1}^{t}\mu^{t-\tau}\hat{\mathbf{g}}_r(\tau)\hat{\mathbf{g}}_r^T(\tau)\\ &= \mu\mathbf{R}_r(t-1)+\hat{\mathbf{g}}_r(t)\hat{\mathbf{g}}_r^T(t) \quad (49)\\ \mathbf{r}_r(t) &\triangleq \sum_{\tau=1}^{t}\mu^{t-\tau}\pi_r(\tau)\hat{\mathbf{g}}_r(\tau)\\ &= \mu\mathbf{r}_r(t-1)+\pi_r(\tau)\hat{\mathbf{g}}_r(t) \quad (50)\\ \tilde{r}_{r,n}(t) &\triangleq r_{r,n}(t)-\sum_{\substack{n'=1\\ n'\neq n}}^{N_s}R_{r,n,n'}(t)\hat{p}_{r,n'}(t-1) \quad (51)\end{aligned}$$

where $R_{r,n,n'}(t)$ and $r_{r,n}(t)$ denote the (n,n')th element of $\mathbf{R}_r(t)$ and the nth entry of $\mathbf{r}_r(t)$, respectively. Note that the computation of $\mathbf{R}_r(t)$, $\mathbf{r}_r(t)$ and $\tilde{r}_{r,n}$ is performed locally, and does not require any message passing between nodes. Then, at time t, each CR r performs the following updates iteratively:

$$\hat{p}_{r,n}^{(j+1)}(t) = \frac{\left[\tilde{r}_{r,n}(t)-\frac{\lambda_t}{N_r}-\zeta_r^{(j)}+\kappa\left(|\mathcal{N}_r|\hat{p}_{r,n}^{(j)}(t)+\sum_{\varrho\in\mathcal{N}_r}\hat{p}_{\varrho,n}^{(j)}(t)\right)\right]_+}{R_{r,n,n}(t)+2\kappa|\mathcal{N}_r|} \quad (52)$$

TABLE II
SUMMARY OF THE SPECTRUM SENSING ALGORITHMS

0: Initialize $t=0$.
1: Acquire measurements $\{\pi_r(t)\}$ at CR $r\in\{1,\ldots,N_r\}$.
2: **[Centralized Algorithm]**
  Collect the measurements $\{\pi_r(t)\}_{r=1}^{N_r}$ at an FC.
  Update $\mathbf{R}(t)$ and $\mathbf{r}(t)$ by (43) and (44).
  Compute $\hat{p}_n(t)$ by (47).
2': **[Distributed Algorithm]**
  Set $\hat{p}_{r,n}^{(0)}(t)=\hat{p}_{r,n}(t)$ for all r.
  Update $\mathbf{R}_r(t)$ and $\mathbf{r}_r(t)$ by (50) and (51), for all r.
  For $j=0,1,\ldots,N_I-1$
    For each CR $r\in\{1,2,\ldots,N_r\}$
      Collect from the neighbors $\hat{p}_{\varrho,n}^{(j)}(t)$, $\varrho\in\mathcal{N}_r$.
      Compute $\hat{p}_{r,n}^{(j+1)}(t)$ from (53).
    Next r
  Next j
  Set $\hat{p}_{r,n}(t)=\hat{p}_{r,n}^{(N_I)}(t)$ for all r.
3: Increment t and go to Step 1.

$$\zeta_r^{(j+1)} = \zeta_r^{(j)}+\kappa\left(|\mathcal{N}_r|\hat{p}_{r,n}^{(j+1)}(t)-\sum_{\varrho\in\mathcal{N}_r}\hat{p}_{\varrho,n}^{(j+1)}(t)\right) \quad (53)$$

where j is the iteration index, $\kappa>0$ is a given constant, and $|\mathcal{N}_r|$ denotes the cardinality of the set $\mathcal{N}_r$. To compute $\hat{p}_{r,n}^{(j+1)}(t)$, CR r must collect from its one-hop neighbors, the power estimates $\hat{p}_{\varrho,n}^{(j)}(t)$, $\varrho\in\mathcal{N}_r$, of the previous iteration. These messages are exchanged over a dedicated control channel.

The following proposition states the convergence property of the distributed algorithm.

Proposition 2: By performing the updates in (52) and (53) at each CR $r\in\{1,2,\ldots,N_r\}$ per iteration $j=0,1,\ldots$, the local copies $\hat{p}_{r,n}^{(j)}(t)$ for all $r\in\{1,2,\ldots,N_r\}$ converge and coincide with $\hat{p}_n(t)$ of (40) as $j\to\infty$, for any positive κ and any initialization for $\{\hat{p}_{r,n}^{(0)}(t)\}$, provided that the network is connected, and the links in the network are bidirectional.

Proof: See Appendix C. □

The overall spectrum sensing algorithms that address step **S2**) in Section II are tabulated in Table II.

VI. NUMERICAL TESTS

Numerical tests were performed to verify the performance of the proposed algorithms. A CR network of $N_r=40$ nodes was considered. The CR nodes, indicated by the circles in Fig. 1, are uniformly distributed over an area of 200 m $\times$ 200 m. Two PUs are involved at positions marked by the triangles in Fig. 1. The square dots represent $N_s=36$ grid points. It is assumed that the active PUs transmit at a constant power of 1 W. The path loss parameters were set to $G_0=0$ dB and $\gamma=3$. The measurement noise variance σ_ϵ^2 was set to 0.1, and $K=15$ orthonormal polynomials were used in $\boldsymbol{\psi}(\mathbf{x})$.

To generate the space-time shadow fading field for each CR, the model proposed in [30] was adopted. Specifically, a time-varying zero-mean Gaussian "spatial loss" field $\ell(\mathbf{x},t)$ with a stationary, isotropic, and exponentially decaying covariance structure was generated first. The time-variance was insinuated by an AR(1) perturbation model. Then, the shadow

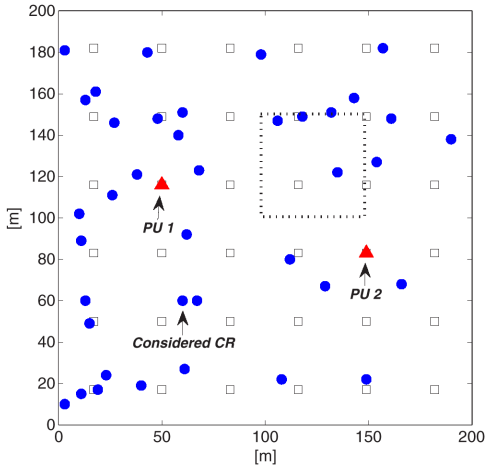

Fig. 1. Simulation setup. The CRs are indicated by the circles, and the PUs by the triangles. The squares represent the grid points.

fading experienced in the link $\mathbf{x} \rightarrow \mathbf{x}_r$ at time t was generated by means of a line integral

$$s_r(\mathbf{x},t) = \frac{1}{\|\mathbf{x}-\mathbf{x}_r\|_2^{1/2}} \int_{\mathbf{x}}^{\mathbf{x}_r} \ell(\mathbf{u},t)\mathrm{d}\mathbf{u}. \tag{54}$$

The effect of a building depicted by a dashed box in the upper right part of Fig. 1 was also simulated by setting $\ell(\mathbf{x},t) = -40$ dB in the region. The generated shadow fading has standard deviation $\sigma_s = 10$ dB, coherence distance 78.6 m, and coherence time corresponding to 99.5 KKF iterations. It should be emphasized that we do not generate the simulated shadow fading map from the model set forth in Section II, but through an unrelated mechanism tested experimentally in [30].

The CG map and the PU state tracking are performed in two different time scales, since the shadow fading evolves very slowly with the coherence time on the order of seconds to minutes [19], while agility is desired in detecting the changes in spectrum occupancy. In the experiments, $40N_s$ Lasso iterations were performed between consecutive map updates; i.e., the Lasso iteration index $t^{\text{Lasso}} = 40N_s t$, where t is the KKF iteration index. However, if the spectrum is sensed to be occupied, the training signals cannot be transmitted; in this case, the map was updated open-loop, as described in Section III-C.

Remark 3: Actual design of the PU detection strategies goes beyond the scope this work, and all the numerical tests assume that the detection is practically error-free. There are a couple of justifications for such an idealization. In CR systems, the miss detection, which is the event of not identifying active PU transmissions, must be strictly regulated to protect the licensed PU systems. In other words, no matter which detector is selected, it has to be designed to yield very low probability of miss detection. Therefore, the implication of miss detection to the map tracking performance would be negligible. On the other hand, false alarms, which are the cases where the detector erroneously reports the presence of active PUs, may affect the map tracking performance through missed measurements. However, as was discussed in Section III-C, the performance degradation is lower-bounded by the schemes based on the path-loss-only map. In fact, the PU tracking performance is hardly affected even if the path-loss-only map is used, when there are no active PUs; see also Fig. 5.

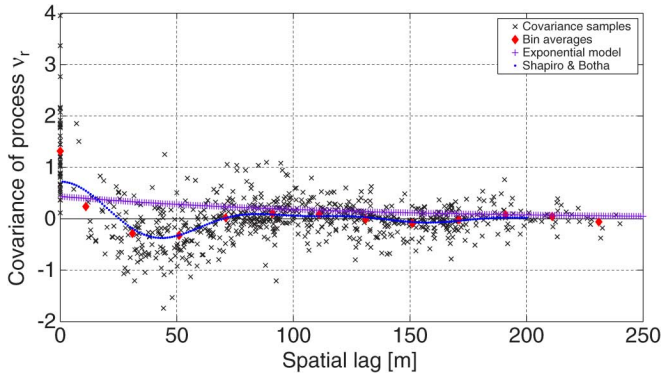

Fig. 2. Model fitting of covariance function $C_{\nu_r}(\cdot)$. The crosses represent the estimated covariance samples from $\hat{\mathbf{C}}_{\nu_r}$, and the diamond dots the bin averages over a bin width of 20 m.

A. Performance of CG Map Tracking

Consider the particular CR r located at (60, 60). In order to perform the map tracking, the model parameters for KKF must be estimated from the data as explained in Section IV. In particular, the covariance function $C_{\nu_r}(\mathbf{x},\mathbf{x}_0)$ must be identified from a model fitting procedure. Fig. 2 shows the fitted covariance functions using the models in (28) and (29). The values of $\hat{\mathbf{C}}_{\nu_r}$ are plotted with "x" markers, and the bin averages of the estimated covariance samples with a bin width of 20 m are also shown in diamond dots. To fit the data to the model in (29), various values of M, Δ, and ϖ were experimented. The curve in small dots in Fig. 2 represents the fitted covariance with $M = 30$, $\Delta = 0.01$ and $\varpi = 1$. The curve with "+" markers corresponds to the covariance fitted to the exponential class in (28). Since the model in (29) obviously characterizes the data better than the exponential model, the numerical tests presented subsequently will be based on $\hat{C}_{\nu_r}(\cdot)$ fitted to this model.

Fig. 3 shows the true and the estimated CG and shadow fading maps for CR r. Fig. 3(a) depicts the true CG map $G_r(\mathbf{x},t)$ at $t = 200$. Clearly, the channel gain peaks at the location of CR r due to the path loss effect. However, the spatially inhomogeneous shadowing component $s_r(\mathbf{x},t)$ depicted in Fig. 3(b) renders the overall CG map non-isotropic. Fig. 3(c) and 3(d) show the estimated CG and shadow fading maps, respectively. It can be seen that the spatial interpolation can effectively predict the values of the shadow fading even in the locations where measurements were not made. The standard deviation of the estimation error of the shadow fading map averaged over all the grid points and over 100 independent realizations of shadowing was found to be 3.72 dB.

B. Performance of PU State Tracking

Centralized Algorithm: To track the PU state centrally at the FC, the forgetting factor $\mu = 0.9$ was used, and weighting factor λ_t for the ℓ_1-penalty term was set as suggested in [17] to

$$\lambda_{t^{\text{Lasso}}} = \sigma_z \sqrt{2 \cdot (\log_{10} N_s) \sum_{\tau=1}^{t^{\text{Lasso}}} \mu^{2(t^{\text{Lasso}}-\tau)}} \tag{55}$$

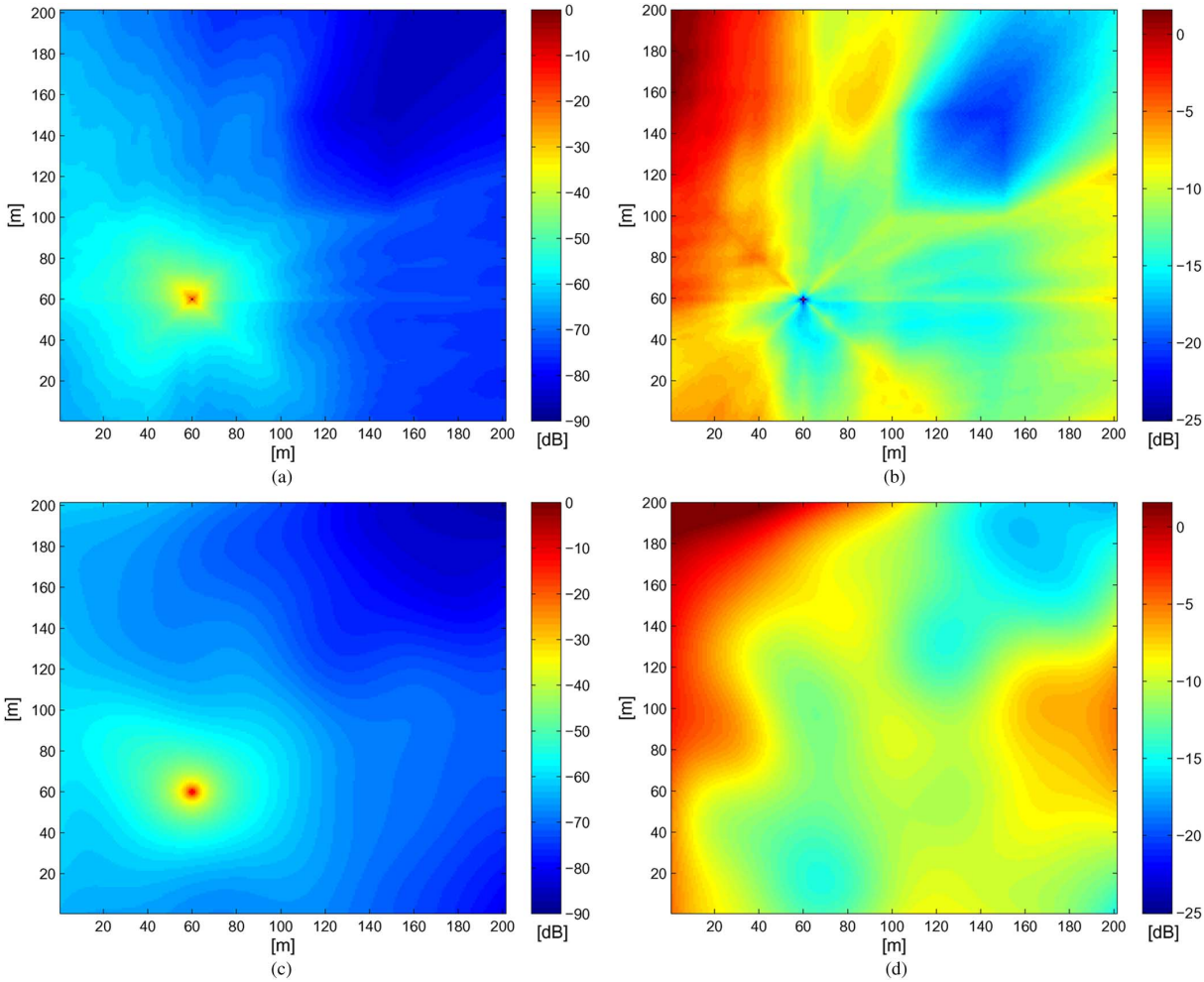

Fig. 3. CG maps and shadow fading maps at $t = 200$. (a) True CG map. (b) True shadow fading map. (c) Estimated CG map. (d) Estimated shadow fading map.

where t^{Lasso} is the iteration index for the OCD-Lasso algorithm, and $\sigma_z^2 = 10^{-10}$ was used.

In Fig. 4, the trajectories of PU state estimates are shown for three different cases: 1) perfect knowledge of CG maps, i.e., $\hat{G}_r(\mathbf{x}, t) = G_r(\mathbf{x}, t), \forall \mathbf{x}, t, r$; 2) using the CG maps estimated via KKF; and 3) by resorting to the path loss-only model, which sets $\hat{s}_r(\mathbf{x}, t) = 0, \forall \mathbf{x}, t, r$. The top panel of Fig. 4 corresponds to the power emitted from the position of "PU 1," and the bottom panel to "PU 2." It is seen that the KKF-based algorithm correctly identifies the presence of both PU transmitters, while the path loss-based scheme essentially misses "PU 2." Moreover, the transmission power of "PU 1" is clearly overestimated in the path loss-based scheme. By exploiting the estimated CG maps, the estimation performance of the PU power levels is significantly improved and approaches that achieved through perfect channel knowledge. Failing to accurately estimate the PU power levels will have negative impact on spatial reuse techniques.

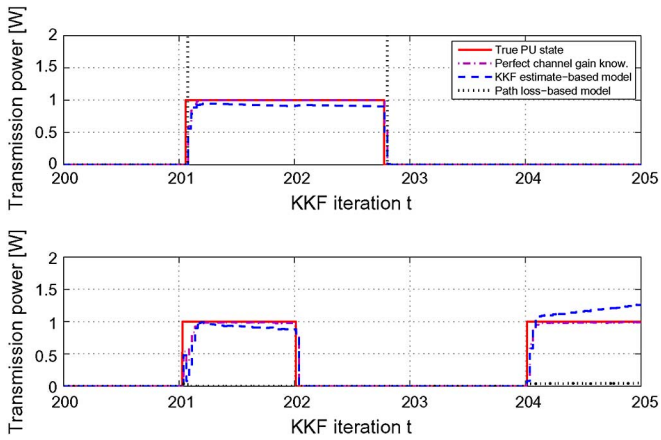

Fig. 4. Power levels at PU 1's location (top) and at PU 2's location (bottom) estimated by the centralized algorithm.

In Fig. 5, the average performance of the centralized algorithm is shown. The mean-square error (MSE) curves plotted in Fig. 5(a) were computed by averaging over 20 independent shadow fading realizations. The true PU state evolution is the same as in Fig. 4. To highlight the merits of the sparsity-exploiting technique, the MSE curves corresponding to the plain recursive least-squares (RLS) are also shown in Fig. 5(a). The plain RLS, which does not exploit sparsity of the estimand is seen to yield much higher MSE. The error is more pronounced when both PUs are silent. In this case, the RLS algorithm essentially reports the presence of low-power PUs, which may cause false alarms. Fig. 5(b) plots the estimated power averaged only at those grid points where PUs are absent, and thus represents the spurious power that should ideally be zero. Again,

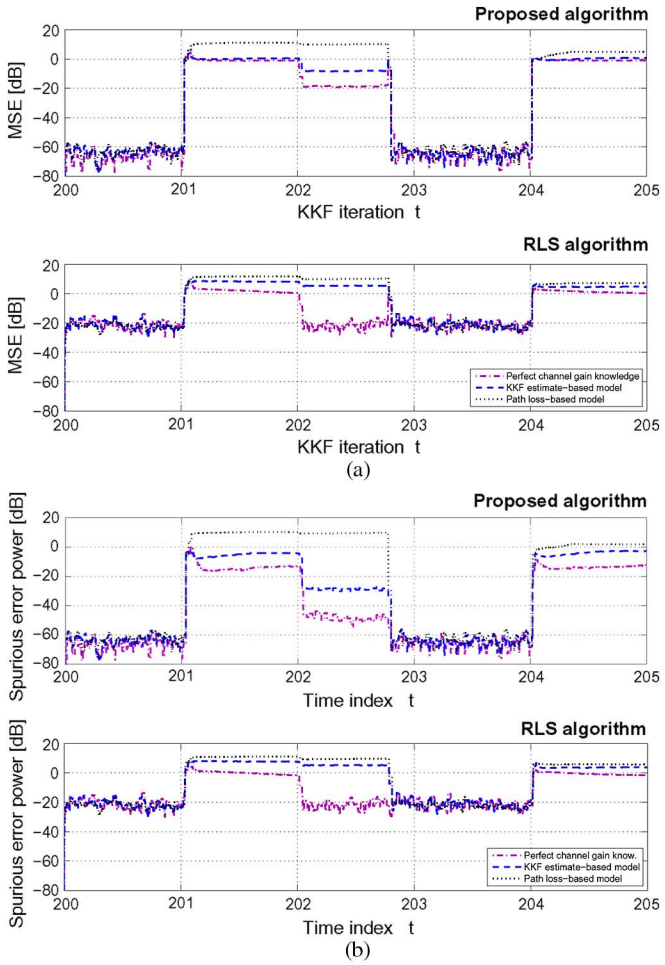

Fig. 5. Average performance of the centralized algorithm (top panels) and the plain RLS algorithm (bottom panels). (a) MSE summed over all grid points. (b) Spurious estimated power at the grid points where the PU is absent.

the KKF-based method exhibits performance close to that of the perfect channel knowledge case, while the path loss-based method yields large spurious power, indicating the positions of the RF emitters are not correctly identified.

The last point is more evident in Fig. 6, where the MSEs at the individual grid points at time instant $t = 201.5$ are plotted, when both PUs are active. For the path loss-based scheme, considerable MSE is observed at three more grid points besides the actual positions of the PUs. This is in contrast to the KKF-based scheme, which is reliable at identifying the correct positions of the PUs, although the power estimates themselves at the correct PU positions are seen to contain some errors.

Fig. 7 shows the average MSE and mean spurious power performance of PU state tracking using the perfect, KKF-based, and pathloss-based CG maps, when the number of simultaneously active PUs is varied. The random positions of the N_r CRs and the PUs are uniformly distributed. It is seen that the KKF-based map tracking offers a clear advantage in terms of PU tracking performance, although the performance degrades as the number of PUs increases.

Distributed Algorithm: In Fig. 8, a sample path of the PU power estimate using the distributed algorithm is presented. The parameter $\kappa = 0.1$ and the number of iterations to reach consensus on the power estimates was set to $N_I = 50$. Comparing Fig. 8 with Fig. 4, it is clear that the distributed algorithm achieves tracking performance very close to that of its centralized counterpart.

Fig. 9 shows the MSE curves for the distributed algorithm when: 1) the channel is perfectly known; 2) the path loss-only model is adopted; and 3) the KKF-based CG map estimates are utilized. The MSE is seen to be a bit larger than that of the centralized case. This is because the number of consensus iterations N_I is sometimes insufficient to bring the local estimates to convergence. In other words, there is a tradeoff between message-passing overhead and tracking performance in the dis-

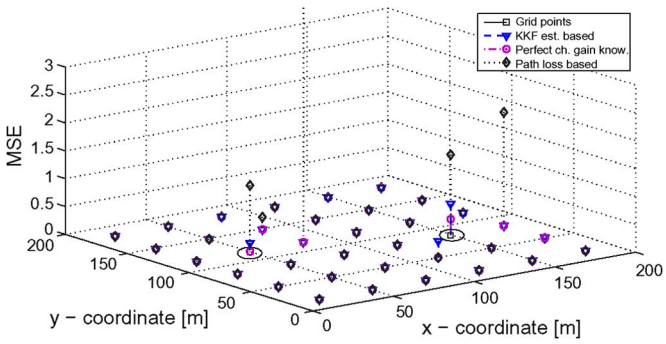

Fig. 6. MSE at each grid location at time $t = 201.5$.

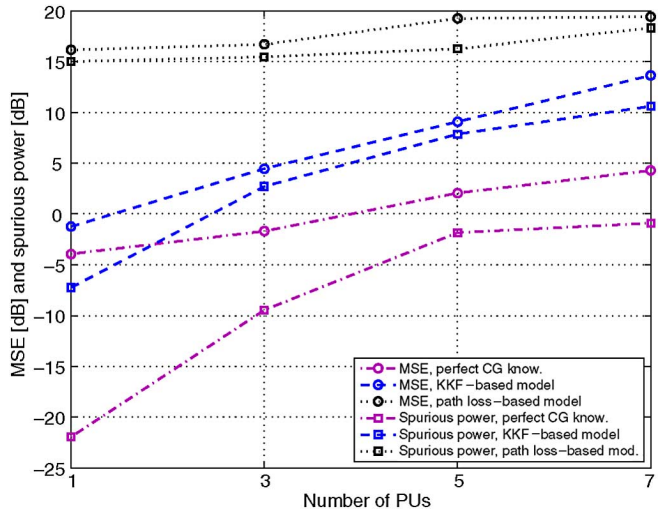

Fig. 7. MSE and mean spurious power versus number of PUs.

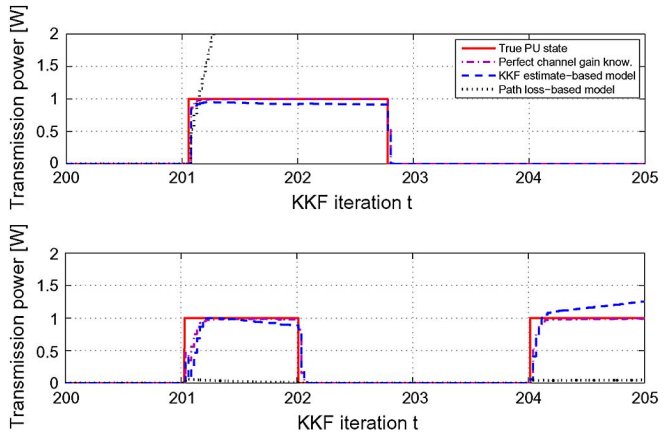

Fig. 8. Power levels at PU 1's location (top) and at PU 2's location (bottom) estimated by the centralized algorithm.

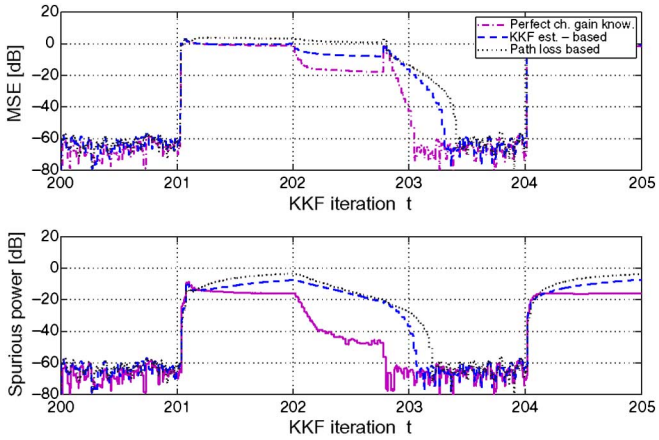

Fig. 9. MSE curves of the power estimates obtained using the distributed algorithm, summed over all grid points (top) and over the grid points unoccupied by the PUs (bottom).

tributed algorithm. It is also observed that after $t = 202.7$, there is difference in the convergence speed of the three cases. For example, it is seen that it takes longer for the path loss-based scheme than for the KKF-based scheme, to detect the channel availability, for the same number of iterations N_I.

VII. Concluding Remarks

Novel spectrum sensing algorithms were developed in this paper for cooperative CR networks. In order to improve the sensing performance, a descriptive spatio-temporal shadow fading model was employed. Based on the model, the optimal filtering technique known as KKF was introduced to track the temporal variation of the shadow fading, as well as to spatially interpolate the channel gains from arbitrary points in the area of interest to the individual CR positions. Analysis and simulated tests confirmed that the KKF can accurately construct the CG maps of the entire area using measurements acquired via training signals exchanged among CRs. Exploiting the CG map information, a sparsity-exploiting linear regression was formulated to estimate the PU locations and transmit-power levels in real-time. Both centralized and distributed implementations of recursive sparse regression algorithms were developed. Compared to the crude path loss-only model-based scheme, the CG map-assisted PU tracking schemes exhibited considerable improvement in estimating not only the positions of the PUs but also their power levels. Such estimates are potentially very useful for spatial reuse in CR networks—a challenging subject with high priority in our future research agenda.

Appendix A
Proof of Proposition 1

It can be easily verified from the modeling assumptions that $\breve{\mathbf{S}}_r(t)$ and $s_r(\mathbf{x},t) = \boldsymbol{\psi}^T(\mathbf{x})\boldsymbol{\alpha}_r(t) + \nu_r(\mathbf{x},t)$ are jointly Gaussian. Thus, conditioned on $\breve{\mathbf{S}}_r(t)$, the process $s_r(\mathbf{x},t)$ is Gaussian distributed.

To derive the conditional mean, note that

$$\mathbb{E}\left\{s_r(\mathbf{x},t)|\breve{\mathbf{S}}_r(t)\right\} = \mathbb{E}\left\{\mathbb{E}\left\{s_r(\mathbf{x},t)|\breve{\mathbf{S}}_r(t),\boldsymbol{\alpha}_r(t)\right\}|\breve{\mathbf{S}}_r(t)\right\} \tag{56}$$

where the outer expectation is with respect to $\boldsymbol{\alpha}_r(t)|\breve{\mathbf{S}}_r(t)$. The inner expectation is easily evaluated as

$$\mathbb{E}\left\{s_r(\mathbf{x},t)|\breve{\mathbf{S}}_r(t),\boldsymbol{\alpha}_r(t)\right\} = \boldsymbol{\psi}^T(\mathbf{x})\boldsymbol{\alpha}_r(t) + \mathbf{c}_{\nu_r}^T(\mathbf{x})\boldsymbol{\Sigma}_r^{-1} \times \left[\breve{\mathbf{s}}_r(t) - \boldsymbol{\Psi}_r\boldsymbol{\alpha}_r(t)\right] \tag{57}$$

from the fact that $\nu_r(\mathbf{x},t)$ is temporally white, and jointly Gaussian with $\breve{\mathbf{s}}r(t)$. Plugging (57) into (56), and using the definition $\hat{\boldsymbol{\alpha}}_r(t|t) \triangleq \mathbb{E}\{\boldsymbol{\alpha}_r(t)|\breve{\mathbf{S}}_r(t)\}$, the desired equation in (19) is obtained.

To verify the conditional variance in (20), recall first the variance decomposition formula

$$\mathrm{var}\left\{s_r(\mathbf{x},t)|\breve{\mathbf{S}}_r(t)\right\} = \mathbb{E}\left\{\mathrm{var}\left\{s_r(\mathbf{x},t)|\breve{\mathbf{S}}_r(t),\boldsymbol{\alpha}_r(t)\right\}|\breve{\mathbf{S}}_r(t)\right\} + \mathrm{var}\left\{\mathbb{E}\left\{s_r(\mathbf{x},t)|\breve{\mathbf{S}}_r(t),\boldsymbol{\alpha}_r(t)\right\}|\breve{\mathbf{S}}_r(t)\right\}. \tag{58}$$

The first term on the right-hand side of (58) can be rewritten as

$$\begin{aligned}
&\mathbb{E}\left\{\mathrm{var}\left\{s_r(\mathbf{x},t)|\breve{\mathbf{S}}_r(t),\boldsymbol{\alpha}_r(t)\right\}|\breve{\mathbf{S}}_r(t)\right\} \\
&= \mathbb{E}\left\{\mathrm{var}\left\{\boldsymbol{\psi}^T(\mathbf{x})\boldsymbol{\alpha}_r(t) + \nu_r(\mathbf{x},t)|\breve{\mathbf{S}}_r(t),\boldsymbol{\alpha}_r(t)\right\}|\breve{\mathbf{S}}_r(t)\right\} && (59)\\
&= \mathbb{E}\left\{\mathrm{var}\left\{\nu_r(\mathbf{x},t)|\breve{\mathbf{S}}_r(t),\boldsymbol{\alpha}_r(t)\right\}|\breve{\mathbf{S}}_r(t)\right\} && (60)\\
&= \mathbb{E}\left\{\sigma_{\nu_r}^2 - \mathbf{c}_{\nu_r}^T(\mathbf{x})\boldsymbol{\Sigma}_r^{-1}\mathbf{c}_{\nu_r}(\mathbf{x})|\breve{\mathbf{S}}_r(t)\right\} && (61)\\
&= \sigma_{\nu_r}^2 - \mathbf{c}_{\nu_r}^T(\mathbf{x})\boldsymbol{\Sigma}_r^{-1}\mathbf{c}_{\nu_r}(\mathbf{x}) && (62)
\end{aligned}$$

where (61) follows from the fact that $\nu_r(\mathbf{x},t)$ is jointly Gaussian with $\breve{\mathbf{s}}r(t)$, and uncorrelated with $\boldsymbol{\alpha}_r(t)$. The second term on the right-hand side of (58) is found to be

$$\begin{aligned}
&\mathrm{var}\left\{\mathbb{E}\left\{s_r(\mathbf{x},t)|\breve{\mathbf{S}}_r(t),\boldsymbol{\alpha}_r(t)\right\}|\breve{\mathbf{S}}_r(t)\right\} && (63)\\
&= \mathrm{var}\left\{\left[\boldsymbol{\psi}^T(\mathbf{x}) - \mathbf{c}_{\nu_r}^T(\mathbf{x})\boldsymbol{\Sigma}_r^{-1}\boldsymbol{\Psi}_r\right]\boldsymbol{\alpha}_r(t) + \mathbf{c}_{\nu_r}^T(\mathbf{x})\boldsymbol{\Sigma}_r^{-1}\breve{\mathbf{s}}_r(t)|\breve{\mathbf{S}}_r(t)\right\} && (64)\\
&= \left[\boldsymbol{\psi}^T(\mathbf{x}) - \mathbf{c}_{\nu_r}^T(\mathbf{x})\boldsymbol{\Sigma}_r^{-1}\boldsymbol{\Psi}_r\right]\mathbf{P}_r(t|t) \times \left[\boldsymbol{\psi}(\mathbf{x}) - \boldsymbol{\Psi}_r^T\boldsymbol{\Sigma}_r^{-1}\mathbf{c}_{\nu_r}(\mathbf{x})\right]. && (65)
\end{aligned}$$

Plugging (62) and (65) into (58) yields (20), and completes the proof of the proposition.

Appendix B
LS Fitting of Exponential Covariances

Stack the entries in the diagonal and the lower-triangular portion of $\hat{\mathbf{C}}_{\nu_r}$ into a vector $\mathbf{c}$ with dimension $N \triangleq N_r(N_r+1)/2$. Likewise, collect in a vector $\mathbf{d}$ the diagonal and the lower-triangular part of a matrix $\mathbf{D}$ whose (i,j)th entry is $\|\mathbf{x}_i - \mathbf{x}_j\|_2$. To fit the estimated covariance of $\nu_r(\mathbf{x},t)$ to the model in (28) in the least-squares sense, solve the following optimization problem:

$$\min_{d_{\nu_r}\geq 0,\ \sigma_{\nu_r}^2} \left\|\check{\mathbf{c}} - \left(\log\sigma_{\nu_r}^2\right)\mathbf{1} + \frac{1}{d_{\nu_r}}\mathbf{d}\right\|_2^2 \tag{66}$$

where $\mathbf{1}$ denotes the $N\times 1$ vector with all entries equal to 1, and $\tilde{\mathbf{c}}$ is a vector with its ith entry equal to the log of the ith entry of $\mathbf{c}$. Solving the optimization problem in (66) using the Lagrange multiplier method yields

$$\hat{d}_{\nu_r} = \frac{\|\mathbf{d}\|^2 - (\mathbf{d}^T\mathbf{1})^2/N}{\tilde{\mathbf{c}}^T\mathbf{1}\mathbf{d}^T\mathbf{1}/N - \tilde{\mathbf{c}}^T\mathbf{d}} \tag{67}$$

$$\hat{\sigma}^2_{\nu_r} = \exp\left[\frac{1}{N}\left(\tilde{\mathbf{c}}^T\mathbf{1} + \frac{\mathbf{d}^T\mathbf{1}}{\hat{d}_{\nu_r}}\right)\right]. \tag{68}$$

Appendix C
Proof of Proposition 2

It is clear that the solution $\{\hat{p}_{r,n}(t)\}_{r=1}^{N_r}$ to the optimization problem (48) will coincide with the solution $\hat{p}_n(t)$ to the optimization problem (40) when the network is connected. Thus, it suffices to show that the proposed distributed algorithm yields the sequences of $\{\hat{p}^{(j)}_{r,n}(t)\}$, $j=0,1,\ldots$, that converge to the solution of (48).

Consider the following equivalent reformulation of (48):

$$\begin{aligned}\min_{\substack{\hat{p}_{r,n}\geq 0,\ q_{r,\varrho}\\ \varrho\in\mathcal{N}_r,\ r=1,2,\ldots,N_r}} &\sum_{r=1}^{N_r}\tilde{\mathcal{J}}_{r,t}(\hat{p}_{r,n})\\ \text{subject to } \hat{p}_{r,n} &= q_{r,\varrho} \quad \forall \varrho\in\mathcal{N}_r,\quad r=1,2,\ldots,N_r\\ q_{r,\varrho} &= q_{\varrho,r} \quad \forall \varrho\in\mathcal{N}_r,\quad r=1,2,\ldots,N_r\end{aligned} \tag{69}$$

where $\tilde{\mathcal{J}}_{r,t}(\hat{p}_{r,n}) \triangleq \mathcal{J}_{r,t}([\hat{p}_{r,1}(t-1),\ldots,\hat{p}_{r,n-1}(t-1),\hat{p}_{r,n},\hat{p}_{r,n+1}(t-1),\ldots,\hat{p}_{r,N_s}(t-1)]^T)$, and $\{q_{r,\varrho}\}$ are auxiliary optimization variables. Problem (69) is in the form to which the alternating direction method of multipliers (ADMoM) can be applied [29, p. 253].

Specifically, consider the augmented (partial) Lagrangian given by

$$\begin{aligned}\mathcal{L}(\{\hat{p}_{r,n}\},\{q_{r,\varrho}\},\{\xi_{r,\varrho}\}) &= \sum_{r=1}^{N_r}\tilde{\mathcal{J}}_{r,t}(\hat{p}_{r,n})\\ &+\sum_{r=1}^{N_r}\sum_{\varrho\in\mathcal{N}_r}\xi_{r,\varrho}(\hat{p}_{r,n}-q_{r,\varrho}) + \kappa\sum_{r=1}^{N_r}\sum_{\varrho\in\mathcal{N}_r}(\hat{p}_{r,n}-q_{r,\varrho})^2\end{aligned} \tag{70}$$

where $\{\xi_{r,\varrho}\}$ are the Lagrange multipliers, and κ is a positive constant. The ADMoM procedure updates the primal variables $\{\hat{p}_{r,n}\},\{q_{r,\varrho}\}$ and the dual variables $\{\xi_{r,\varrho}\}$, alternately, as

$$\left\{\hat{p}^{(j+1)}_{r,n}\right\} = \arg\min_{\{\hat{p}_{r,n}\geq 0\}}\mathcal{L}\left(\{\hat{p}_{r,n}\},\left\{q^{(j)}_{r,\varrho}\right\},\left\{\xi^{(j)}_{r,\varrho}\right\}\right) \tag{71}$$

$$\left\{q^{(j+1)}_{r,\varrho}\right\} = \arg\min_{\substack{q_{r,\varrho}=q_{\varrho,r},\\ \varrho\in\mathcal{N}_r,\ r=1,2,\ldots,N_r}}\mathcal{L}\left(\left\{\hat{p}^{(j+1)}_{r,n}\right\},\ \{q_{r,\varrho}\},\left\{\xi^{(j)}_{r,\varrho}\right\}\right) \tag{72}$$

$$\xi^{(j+1)}_{r,\varrho} = \xi^{(j)}_{r,\varrho} + 2\kappa\left(\hat{p}^{(j+1)}_{r,n} - q^{(j+1)}_{r,\varrho}\right),\quad \varrho\in\mathcal{N}_r,\quad r=1,\ldots,N_r. \tag{73}$$

Proposition 4.2 in [29] implies that the sequence $\hat{p}^{(j)}_{r,n}$, $j=1,2,\ldots$, generated by (71)–(73) converges to the optimal solution $\hat{p}_{r,n}(t)$ of (69), for each r, for an arbitrary initial $\hat{p}^{(0)}_{r,n}$ and any positive constant κ.

It is now shown that the procedure (71)–(73) can be simplified to (52), (53). First, it is noted that (72) can be rewritten as

$$\begin{aligned}\left\{q^{(j+1)}_{r,\varrho}\right\} &= \arg\min_{\substack{q_{r,\varrho}=q_{\varrho,r},\\ \varrho\in\mathcal{N}_r,\ r=1,2,\ldots,N_r}} -\sum_{r=1}^{N_r}\sum_{\varrho\in\mathcal{N}_r}\xi^{(j)}_{r,\varrho}q_{r,\varrho}\\ &\quad+\kappa\sum_{r=1}^{N_r}\sum_{\varrho\in\mathcal{N}_r}\left(\hat{p}^{(j+1)}_{r,n}-q_{r,\varrho}\right)^2 \qquad (74)\\ &= \arg\min_{\{q_{r,\varrho}\}}\sum_{\{(r,\varrho)|\varrho\in\mathcal{N}_r,r>\varrho\}}\left\{-\left(\xi^{(j)}_{r,\varrho}+\xi^{(j)}_{\varrho,r}\right)q_{r,\varrho}\right.\\ &\quad\left.+\kappa\left[\left(\hat{p}^{(j+1)}_{r,n}-q_{r,\varrho}\right)^2+\left(\hat{p}^{(j+1)}_{\varrho,n}-q_{r,\varrho}\right)^2\right]\right\} \qquad (75)\end{aligned}$$

where in the last step the conditions $\{q_{r,\varrho}=q_{\varrho,r}\}$ and the assumption that the links are bidirectional are used. It is clear from (75) that $\{q^{(j+1)}_{r,\varrho}\}$ can be obtained as

$$q^{(j+1)}_{r,\varrho} = \frac{1}{2}\left(\hat{p}^{(j+1)}_{r,n}+\hat{p}^{(j+1)}_{\varrho,n}\right)+\frac{1}{4\kappa}\left(\xi^{(j)}_{r,\varrho}+\xi^{(j)}_{\varrho,r}\right),\quad \varrho\in\mathcal{N}_r,\quad r=1,2,\ldots,N_r. \tag{76}$$

By substituting (76) into (73), one obtains

$$\xi^{(j+1)}_{r,\varrho} = \kappa\left(\hat{p}^{(j+1)}_{r,n}-\hat{p}^{(j+1)}_{\varrho,n}\right)+\frac{1}{2}\left(\xi^{(j)}_{r,\varrho}-\xi^{(j)}_{\varrho,r}\right),\quad \varrho\in\mathcal{N}_r,\quad r=1,2,\ldots,N_r. \tag{77}$$

Thus, it can be verified that

$$\xi^{(j)}_{r,\varrho}+\xi^{(j)}_{\varrho,r}=0,\quad \varrho\in\mathcal{N}_r,\quad r=1,2,\ldots,N_r. \tag{78}$$

From (78) and (76), (77), it follows readily that

$$q^{(j+1)}_{r,\varrho} = \frac{1}{2}\left(\hat{p}^{(j+1)}_{r,n}+\hat{p}^{(j+1)}_{\varrho,n}\right),\quad \varrho\in\mathcal{N}_r,\quad r=1,2,\ldots,N_r \tag{79}$$

$$\xi^{(j+1)}_{r,\varrho} = \xi^{(j)}_{r,\varrho}+\kappa\left(\hat{p}^{(j+1)}_{r,n}-\hat{p}^{(j+1)}_{\varrho,n}\right),\quad \varrho\in\mathcal{N}_r,\quad r=1,2,\ldots,N_r. \tag{80}$$

Now, consider (71), which can be rewritten as

$$\begin{aligned}&\left\{\hat{p}^{(j+1)}_{r,n}\right\}\\ &= \arg\min_{\{\hat{p}_{r,n}\geq 0\}}\sum_{r=1}^{N_r}\left[\tilde{\mathcal{J}}_{r,t}(\hat{p}_{r,n})+\sum_{\varrho\in\mathcal{N}_r}\xi^{(j)}_{r,\varrho}\hat{p}_{r,n}+\kappa\sum_{\varrho\in\mathcal{N}_r}\left(\hat{p}_{r,n}-q^{(j)}_{r,\varrho}\right)^2\right] \qquad (81)\\ &= \arg\min_{\{\hat{p}_{r,n}\geq 0\}}\sum_{r=1}^{N_r}\left[\left(-\tilde{r}_{r,n}+\frac{\lambda_t}{N_r}+\sum_{\varrho\in\mathcal{N}_r}\xi^{(j)}_{r,\varrho}-2\kappa\sum_{\varrho\in\mathcal{N}_r}q^{(j)}_{r,\varrho}\right)\hat{p}_{r,n}\right.\\ &\quad\left.+\left(\frac{1}{2}R_{r,n,n}+\kappa|\mathcal{N}_r|\right)\hat{p}^2_{r,n}\right]. \qquad (82)\end{aligned}$$

Clearly, $\{\hat{p}^{(j+1)}_{r,n}\}$ can be obtained in closed-form as

$$\hat{p}^{(j+1)}_{r,n} = \frac{\left[\tilde{r}_{r,n}-\frac{\lambda_t}{N_r}-\sum_{\varrho\in\mathcal{N}_r}\xi^{(j)}_{r,\varrho}+2\kappa\sum_{\varrho\in\mathcal{N}_r}q^{(j)}_{r,\varrho}\right]_+}{R_{r,n,n}+2\kappa|\mathcal{N}_r|},\quad r=1,2,\ldots,N_r. \tag{83}$$

Upon defining $\zeta_r^{(j)} \triangleq \sum_{\varrho \in \mathcal{N}_r} \xi_{r,\varrho}^{(j)}$, and substituting (79)to (83), (52) is obtained. Finally, (53) follows from the definition of $\zeta_r^{(j)}$ and (80).

REFERENCES

[1] A. Sahai, N. Hoven, M. Mishra, and R. Tandra, "Fundamental tradeoffs in robust spectrum sensing for opportunistic frequency reuse,' Univ. of Calif., Berkeley, CA, 2006, Tech. Rep..

[2] X. Hong, C. Wang, H. Chen, and Y. Zhang, "Secondary spectrum access networks," *IEEE Veh. Technol. Mag.*, vol. 4, no. 2, pp. 36–43, Jun. 2009.

[3] E. Visotsky, S. Kuffner, and R. Peterson, "On collaborative detection of TV transmissions in support of dynamic spectrum sharing," in *Proc. DySPAN Conf.*, Baltimore, MD, Nov. 2005, pp. 338–345.

[4] G. Ganesan, Y. Li, B. Bing, and S. Li, "Spatio-temporal sensing in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 5–12, Jan. 2008.

[5] Z. Quan, S. Cui, H. V. Poor, and A. H. Sayed, "Collaborative wideband sensing for cognitive radios," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 60–73, Nov. 2008.

[6] J. Ma, G. Zhao, and Y. Li, "Soft combination and detection for cooperative spectrum sensing in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 11, pp. 4502–4507, Nov. 2008.

[7] J. Unnikrishnan and V. V. Veeravalli, "Cooperative sensing for primary detection in cognitive radio," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 18–27, Feb. 2008.

[8] B. Mark and A. Nasif, "Estimation of maximum interference-free power level for opportunistic spectrum access," *IEEE Trans. Wireless Commun.*, vol. 8, no. 5, pp. 2505–2513, May 2009.

[9] N. Cressie, "The origin of Kriging," *Math. Geol.*, vol. 22, no. 3, pp. 239–252, 1990.

[10] A. B. H. Alaya-Feki, S. B. Jemaa, B. Sayrac, P. Houze, and E. Moulines, "Informed spectrum usage in cognitive radio networks: Interference cartography," in *Proc. PIMRC Conf.*, Cannes, France, Sep. 2008, pp. 1–5.

[11] M. Wellens, J. Riihijärvi, M. Gordziel, and P. Mähönen, "Spatial statistics of spectrum usage: From measurements to spectrum models," in *Proc. ICC Conf.*, Dresden, Germany, Jun. 2009, pp. 1–6.

[12] G. Mateos, J.-A. Bazerque, and G. B. Giannakis, "Spline-based spectrum cartography for cognitive radios," in *Proc. 43rd Asilomar Conf. Signal, Syst., Comput.*, Pacific Grove, CA, Nov. 2009.

[13] K. V. Mardia, C. Goodall, E. J. Redfern, and F. J. Alonso, "The Kriged Kalman filter," *Test*, vol. 7, no. 2, pp. 217–285, Dec. 1998.

[14] C. K. Wikle and N. Cressie, "A dimension-reduced approach to space-time Kalman filtering," *Biometrika*, vol. 86, no. 4, pp. 815–829, 1999.

[15] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. R. Statist. Soc.*, vol. 58, no. 1, pp. 267–288, 1996.

[16] J.-A. Bazerque and G. B. Giannakis, "Distributed spectrum sensing for cognitive radio networks by exploiting sparsity," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1847–1862, Mar. 2010.

[17] D. Angelosante, J.-A. Bazerque, and G. B. Giannakis, "Online coordinate descent for adaptive estimation of sparse signals," in *Proc. IEEE Workshop Statist. Signal Process.*, Cardiff, Wales, U.K., Aug.–Sep. 2009, pp. 369–372.

[18] D. Cabric, A. Tkachenko, and R. W. Brodersen, "Spectrum sensing measurements of pilot, energy, and collaborative detection," in *Proc. MILCOM Conf.*, Washington, D.C., Oct. 2006, pp. 1–7.

[19] T. S. Rappaport, *Wireless Communications: Principles & Practice*. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[20] C. Oestges, N. Czink, B. Bandemer, P. Castiglione, F. Kaltenberger, and A. J. Paulraj, "Experimental characterization and modeling of outdoor-to-indoor and indoor-to-indoor distributed channels," *IEEE Trans. Veh. Technol.* vol. 59, no. 5, pp. 2253–2265, Jun. 2010.

[21] J. Cortés, "Distributed Kriged Kalman filter for spatial estimation," *IEEE Trans. Autom. Control*, vol. 54, no. 12, pp. 2816–2827, Dec. 2009.

[22] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, ser. Prentice-Hall Information and System Sciences Series. Englewood Cliffs, NJ: Prentice-Hall, 1979.

[23] C. K. Wikle, L. M. Berliner, and N. Cressie, "Hierarchical Bayesian space-time models," *Environ. Ecol. Statist.*, vol. 5, no. 2, pp. 117–154, Jun. 1998.

[24] B. D. Ripley, *Spatial Statistics*. New York: Wiley, 1981.

[25] A. Shapiro and J. D. Botha, "Variogram fitting with a general class of conditionally nonnegative definite functions," *Computat. Statist. Data Anal.*, vol. 11, no. 1, pp. 87–96, Jan. 1991.

[26] P. K. Suetin, *Orthogonal Polynomials in Two Variables*. Boca Raton, FL: CRC, 1999.

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

[28] H. Zhu, G. B. Giannakis, and A. Cano, "Distributed in-network channel decoding," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3970–3983, Oct. 2009.

[29] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

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

Seung-Jun Kim received the B.S. and M.S. degrees from Seoul National University, Seoul, Korea, in 1996 and 1998, respectively, and the Ph.D. degree from the University of California, Santa Barbara, in 2005, all in electrical engineering.

From 2005 to 2008, he worked for NEC Laboratories America, Princeton, NJ, initially as a Visiting Researcher, and then as a Research Staff Member. Since 2008, he has been with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, where he is currently a Research Assistant Professor. His research interests lie in wireless communications and networking, statistical signal processing, and optimization.

Emiliano Dall'Anese (S'08) was born in Belluno, Italy, on 10 March, 1983. He received the Laurea Triennale (B.Sc. degree) and the Laurea Specialistica (M.Sc. degree) in telecommunication engineering from the University of Padova, Padova, Italy, in 2005 and 2007, respectively. He is currently working towards the Ph.D. in information engineering at the Department of Information Engineering (DEI), University of Padova.

Since January 2009, he has been a Visiting Scholar at the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis. His research interests lie in the areas of communication theory and signal processing. His current research focuses on wireless cognitive radio systems, multi-antenna systems, sensor networks, and distributed signal processing.

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

Since 1999, he has been a Professor with the University of Minnesota, Minneapolis, where he now holds an ADC Chair in Wireless Telecommunications in the Electrical and Computer Engineering Department, and serves as Director of the Digital Technology Center. His general interests span the areas of communications, networking, and statistical signal processing—subjects on which he has published more than 300 journal papers, 500 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 sensors, and social networks. He is the inventor or co-inventor of 17 patents issued.

Prof. Giannakis is the recipient or corecipient 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 also a Fellow of EURASIP, and has served the IEEE in a number of posts, including that of a Distinguished Lecturer for the IEEE-SP Society.