Group-Lasso on Splines for Spectrum Cartography†

Juan A. Bazerque, Gonzalo Mateos and Georgios B. Giannakis (contact author)∗

Submitted: October 5, 2010

Abstract

The unceasing demand for continuous situational awareness calls for innovative and large-scale signal processing algorithms, complemented by collaborative and adaptive sensing platforms to accomplish the objectives of layered sensing and control. Towards this goal, the present paper develops a spline-based approach to field estimation, which relies on a basis expansion model of the field of interest. The model entails known bases, weighted by generic functions estimated from the field's noisy samples. A novel field estimator is developed based on a regularized variational least-squares (LS) criterion that yields finitely-parameterized (function) estimates spanned by thin-plate splines. Robustness considerations motivate well the adoption of an overcomplete set of (possibly overlapping) basis functions, while a sparsifying regularizer augmenting the LS cost endows the estimator with the ability to select a few of these bases that "better" explain the data. This parsimonious field representation becomes possible, because the sparsity-aware spline-based method of this paper induces a group-Lasso estimator for the coefficients of the thin-plate spline expansions per basis. A distributed algorithm is also developed to obtain the group-Lasso estimator using a network of wireless sensors, or, using multiple processors to balance the load of a single computational unit. The novel spline-based approach is motivated by a *spectrum cartography* application, in which a set of sensing cognitive radios collaborate to estimate the distribution of RF power in space and frequency. Simulated tests corroborate that the estimated power spectrum density atlas yields the desired RF state awareness, since the maps reveal spatial locations where idle frequency bands can be reused for transmission, even when fading and shadowing effects are pronounced.

Index Terms

Sparsity, splines, (group-)Lasso, field estimation, cognitive radio sensing, optimization.

† Work in this paper was supported by the NSF grants CCF-0830480 and ECCS-0824007. Part of the paper appeared in the *Proc. of the 43rd Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 1-4, 2009.

∗ The authors are with the Dept. of Electrical and Computer Engineering, University of Minnesota, 200 Union Street SE, Minneapolis, MN 55455. Tel/fax: (612)626-7781/625-4583; Emails: {bazer002,mate0058,georgios}@umn.edu

I. INTRODUCTION

Well-appreciated as a tool for field estimation, thin-plate (smoothing) splines find application in areas as diverse as climatology [27], image processing [9], and neurophysiology [21]. Spline-based field estimation involves approximating a deterministic map $g : \mathbb{R}^n \to \mathbb{R}$ from a finite number of its noisy data samples, by minimizing a variational least-squares (LS) criterion regularized with a smoothness-controlling functional. In the dilemma of trusting a model versus trusting the data, splines favor the latter since only a mild regularity condition is imposed on the derivatives of g, which is otherwise treated as a generic function. While this generality is inherent to the variational formulation, the smoothness penalty renders the estimated map unique and finitely parameterized [10, p. 85], [26, p. 31]. With the variational problem solution expressible by polynomials and specific kernels, the aforementioned map approximation task reduces to a parameter estimation problem. Consequently, thin-plate splines operate as a reproducing kernel Hilbert space (RKHS) learning machine in a suitably defined (Sobolev) space [26, p. 34].

Although splines emerge as variational LS estimators of *deterministic* fields, they are also connected to classes of estimators for *random* fields. The first class assumes that estimators are linearly related to the measured samples, while the second one assumes that fields are Gaussian distributed. The first corresponds to the Kriging method while the second to the Gaussian process model; but in both cases one deals with a best linear unbiased estimator (BLUE) [24]. Typically, wide sense stationarity is assumed for the field's spatial correlation needed to form the BLUE. The so-termed generalized covariance model adds a parametric nonstationary term comprising known functions specified a priori [17]. Inspection of the BLUE reveals that if the nonstationary part is selected to comprise polynomials, and the spatial correlation is chosen to be the splines kernel, then the Kriging, Gaussian process, and spline-based estimators coincide [26, p. 35].

Bearing in mind this unifying treatment of deterministic and random fields, the main subjects of this paper are spline-based estimation, and the practically motivated *sparse* (and thus parsimonious) description of the wanted field. Toward these goals, the following basis expansion model (BEM) is adopted for the target map

$$\Phi(\mathbf{x}, f) = \sum_{\nu=1}^{N_b} g_\nu(\mathbf{x}) b_\nu(f) \tag{1}$$

with $\mathbf{x} \in \mathbb{R}^2$, $f \in \mathbb{R}$, and the L_2-norms $\{||b_\nu(f)||_{L_2} = 1\}_{\nu=1}^{N_b}$ normalized to unity.

The bases $\{b_\nu(f)\}_{\nu=1}^{N_b}$ are preselected, and the functions $g_\nu(\mathbf{x})$ are to be estimated based on noisy samples of Φ. This way, the model-versus-data balance is calibrated by introducing a priori knowledge on the dependence of the map Φ with respect to (w.r.t.) variable f, or more generally a group of variables, while trusting the data to dictate the functions $g_\nu(\mathbf{x})$ of the remaining variables $\mathbf{x}$.

Consider selecting N_b basis functions using the *basis pursuit* approach [8], which entails an extensive set of bases thus rendering N_b overly large and the model overcomplete. This motivates augmenting the variational LS problem with a suitable sparsity-encouraging penalty, which endows the map estimator with the ability to discard factors $g_\nu(\mathbf{x})b_\nu(f)$ in (1), only keeping a few bases that "better" explain the data. This attribute is inherited because the novel sparsity-aware spline-based method of this paper induces a group-Lasso estimator for the coefficients of the optimal finitely-parameterized g_ν. Group-Lasso estimators are known to set groups of weak coefficients to zero (here the N_b groups associated with coefficients per g_ν), and outperform the sparsity-agnostic LS estimator by capitalizing on the sparsity present [29], [22]. An iterative group-Lasso algorithm is developed that yields closed-form estimates per iteration. A distributed version of this algorithm is also introduced for data collected by cooperating sensors, or, for computational load-balancing of multiprocessor architectures. A related approach to model selection in nonparametric regression is the component selection and smoothing operator (COSSO) [16]. Different from the approach followed here, COSSO is limited to smoothing-spline, analysis-of-variance models, where the target function is assumed to be expressible by a superposition of *orthogonal* component functions. Compared to the single group-Lasso estimate here, COSSO entails an iterative algorithm, which alternates through a sequence of smoothing spline [13, p. 151] and nonnegative garrote [7] subproblems.

The motivation behind the BEM in (1) comes from our interest in spectrum cartography for wireless *cognitive radio* (CR) networks, a *sensing* application that serves as an illustrating paradigm throughout the paper. CR technology holds great promise to address fruitfully the perceived dilemma of bandwidth under-utilization versus spectrum scarcity, which has rendered fixed-access communication networks inefficient. Sensing the ambient interference spectrum is of paramount importance to the operation of CR networks, since it enables spatial frequency reuse and allows for dynamic spectrum allocation; see, e.g., [11], [19] and references therein. Collaboration among CRs can markedly improve the sensing performance [23], and is key to revealing opportunities for spatial frequency reuse [20]. Pertinent existing approaches have mostly relied on detecting spectrum occupancy per radio, and do not account for spatio-temporal changes in the radio frequency (RF) ambiance, especially at intended receiver(s) which may reside several hops away from the sensed area.

The impact of this paper's novel field estimators to CR networks is a collaborative sensing scheme whereby receiving CRs cooperate to estimate the distribution of power in space $\mathbf{x}$ and frequency f, namely the power spectrum density (PSD) map $\Phi(\mathbf{x}, f)$ in (1), from local periodogram measurements. The estimator need not be extremely accurate, but precise enough to identify spectrum holes. This justifies adopting the known bases to capture the PSD frequency dependence in (1). As far as the spatial dependence is concerned,

the model must account for path loss, fading, mobility, and shadowing effects, all of which vary with the propagation medium. For this reason, it is prudent to let the data dictate the spatial component of (1). Knowing the spectrum at any location allows remote CRs to reuse dynamically idle bands. It also enables CRs to adapt their transmit-power so as to minimally interfere with licensed transmitters. The spline-based PSD map here provides an alternative to [4], where known bases are used both in space and frequency. Different from [1] and [4], the field estimator here does not presume a spatial covariance model or pathloss channel model. Moreover, it captures general propagation characteristics including both shadowing and fading; see also [15].

Notation: Bold uppercase letters will denote matrices, whereas bold lowercase letters will stand for column vectors. Operators $\otimes$, $(.)'$, tr(.), rank(.), bdiag(.), $E[\cdot]$ will denote Kronecker product, transposition, matrix trace, rank, block diagonal matrix and expectation, respectively; $|.|$ will be used for the cardinality of a set, and the magnitude of a scalar. The L_2 norm of function $b : \mathbb{R} \to \mathbb{R}$ is $||b||_{L_2}^2 := \int_{-\infty}^{\infty} b^2(f)df$, while the ℓ_p norm of vector $\mathbf{x} \in \mathbb{R}^p$ is $\|\mathbf{x}\|_p := (\sum_{i=1}^p |x_i|^p)^{1/p}$ for $p \geq 1$; and $\|\mathbf{M}\|_F := \sqrt{\text{tr}\,(\mathbf{M}\mathbf{M}')}$ is the matrix Frobenious norm. Positive definite matrices will be denoted by $\mathbf{M} \succ \mathbf{0}$. The $p \times p$ identity matrix will be represented by $\mathbf{I}_p$, while $\mathbf{0}_p$ will denote the $p \times 1$ vector of all zeros, and $\mathbf{0}_{p\times q} := \mathbf{0}_p\mathbf{0}_q'$. The i-th vector in the canonical basis for $\mathbb{R}^p$ will be denoted by $\mathbf{e}_{p,i}$, $i = 1, \ldots, p$.

II. BEM FOR SPECTRUM CARTOGRAPHY

Consider a set of N_s sources transmitting signals $\{u_s(t)\}_{s=1}^{N_s}$ using portions of the overall bandwidth B. The objective of revealing which of these portions (sub-bands) are available for new systems to transmit, suggests that the PSD estimate sought does not need to be super accurate. This motivates modeling the transmit-PSD of each $u_s(t)$ as

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

where the basis $b_\nu(f)$ is centered at frequency f_ν, $\nu = 1, \ldots, N_b$. The example depicted in Fig. 1 involves (generally *overlapping*) raised cosine bases with support $B_\nu = [f_\nu - (1+\rho)/2T_s, f_\nu + (1+\rho)/2T_s]$, where T_s is the symbol period, and ρ stands for the roll-off factor. Such bases can model transmit-spectra of e.g., multicarrier systems. In other situations, power spectral masks may dictate sharp transitions between contiguous sub-bands, cases in which non-overlapping rectangular bases may be more appropriate. All in all, the set of bases should be selected to accommodate a priori knowledge about the PSD.

The power transmitted by source s will propagate to the location $\mathbf{x} \in \mathbb{R}^2$ according to a generally unknown spatial loss function $l_s(\mathbf{x}) : \mathbb{R}^2 \to \mathbb{R}$. The propagation model $l_s(\mathbf{x})$ not only captures frequency-flat deterministic pathloss, but also stationary, block-fading and even frequency-selective Rayleigh channel

effects, since their statistical moments do not depend on the frequency variable. In this case, the following vanishing memory assumption is required on the transmitted signals for the spatial receive-PSD $\Phi(\mathbf{x}, f)$ to be factorizable as $l_s(\mathbf{x})\Phi_s(f)$; see [4] for further details.

(as) *Sources* $\{u_s(t)\}_{s=1}^{N_s}$ *are stationary, mutually uncorrelated, independent of the channels, and have vanishing correlation per coherence interval; i.e.,* $r_{ss}(\tau) := E[u_s(t+\tau)u_s(t)] = 0, \forall\, |\tau| > T_c - L$*, where* T_c *and* L *represent the coherence interval and delay spread of the channels, respectively.*

Under (as), the contribution of source s to the PSD at point $\mathbf{x}$ is $l_s(\mathbf{x}) \sum_{\nu=1}^{N_b} \theta_{s\nu} b_\nu(f)$; and the PSD due to all sources received at $\mathbf{x}$ will be given by $\Phi(\mathbf{x}, f) = \sum_{s=1}^{N_s} l_s(\mathbf{x}) \sum_{\nu=1}^{N_b} \theta_{s\nu} b_\nu(f)$. Such a model can be simplified by defining the function $g_\nu(\mathbf{x}) := \sum_{s=1}^{N_s} \theta_{s\nu} l_s(\mathbf{x})$. With this definition and upon exchanging the order of summation, the spatial PSD model takes the form in (1), where functions $\{g_\nu(\mathbf{x})\}_{\nu=1}^{N_b}$ are to be estimated. They represent the aggregate distribution of power across space corresponding to the frequencies spanned by the bases $\{b_\nu\}$. Observe that the sources are not explicitly present in (1). Even if this model could have been postulated directly for the cartography task at hand, the previous discussion justifies the factorization of the $\Phi(\mathbf{x}, f)$ map per band in factors depending on each of the variables $\mathbf{x}$ and f.

III. Cooperative Spline-Based PSD Field Estimation

The sensing strategy will rely on the periodogram estimate $\widehat{\phi}_{rn}(\tau)$ at a set of receiving (sampling) locations $\mathcal{X} := \{\mathbf{x}_r\}_{r=1}^{N_r} \in \mathbb{R}^2$, frequencies $\mathcal{F} := \{f_n\}_{n=1}^{N} \in B$, and time-slots $\{\tau\}_{\tau=1}^{T}$. In order to reduce the periodogram variance and mitigate fading effects, $\widehat{\phi}_{rn}(\tau)$ is averaged across a window of T time-slots [4], to obtain

$$\varphi_{rn} := \frac{1}{T} \sum_{\tau=1}^{T} \widehat{\phi}_{rn}(\tau). \tag{3}$$

Hence, the envisioned setup consists of N_r receiving CRs, which collaborate to construct the PSD map based on PSD observations $\{\varphi_{rn}\}$. The bulk of processing is performed centrally at a fusion center (FC), which is assumed to know the position vectors $\mathcal{X}$ of all CRs, and the sensed tones in $\mathcal{F}$. The FC receives over a dedicated control channel, the vector of samples $\boldsymbol{\varphi}_r := [\varphi_{r1}, \ldots, \varphi_{rN}]' \in \mathbb{R}^N$ taken by node r for all $r = 1, \ldots, N_r$.

While a BEM could be introduced for the spatial loss function $l_s(\mathbf{x})$ as well [4], the uncertainty on the source locations and obstructions in the propagation medium may render such a model imprecise. This will happen, e.g., when shadowing is present. The alternative approach followed here relies on estimating the functions $g_\nu(\mathbf{x})$ based on the data $\{\varphi_{rn}\}$. To capture the smooth portions of $\Phi(\mathbf{x}, f)$, the criterion for selecting $g_\nu(\mathbf{x})$ will be regularized using a so termed thin-plate penalty [26, p. 30]. This penalty extends to

$\mathbb{R}^2$ the one-dimensional roughness regularization used in smoothing spline models. Accordingly, functions $\{g_\nu\}_{\nu=1}^{N_b}$ are estimated as

$$\{\hat{g}_\nu\}_{\nu=1}^{N_b} := \arg\min_{\{g_\nu \in \mathcal{S}\}} \frac{1}{N_r N} \sum_{r=1}^{N_r} \sum_{n=1}^{N} \left(\varphi_{rn} - \sum_{\nu=1}^{N_b} g_\nu(\mathbf{x}_r) b_\nu(f_n) \right)^2 + \lambda \sum_{\nu=1}^{N_b} \int_{\mathbb{R}^2} ||\nabla^2 g_\nu(\mathbf{x})||_F^2 d\mathbf{x} \quad (4)$$

where $||\nabla^2 g_\nu||_F$ denotes the Frobenius norm of the Hessian of g_ν.

The optimization is over $\mathcal{S}$, the space of Sobolev functions, for which the penalty is well defined [10, p. 85]. The parameter $\lambda \geq 0$ controls the degree of smoothing. Specifically, for $\lambda = 0$ the estimates in (4) correspond to *rough* functions interpolating the data; while as $\lambda \to \infty$ the estimates yield linear functions (cf. $\nabla^2 \hat{g}_\nu(\mathbf{x}) \equiv \mathbf{0}_{2\times 2}$). A smoothing parameter in between these limiting values will be selected using a leave-one-out cross-validation (CV) approach, as discussed later.

A. Thin-plate splines solution

The optimization problem (4) is variational in nature, and in principle requires searching over the infinite-dimensional functional space $\mathcal{S}$. It turns out that (4) admits closed-form, finite dimensional minimizers $\hat{g}_\nu(\mathbf{x})$, as presented in the following proposition which provides a generalization of standard thin-plate splines results; see e.g., [26, p.31], to the multi-dimensional BEM (1).

Proposition 1: *The estimates* $\{\hat{g}_\nu\}_{\nu=1}^{N_b}$ *in* (4) *are thin-plate splines expressible in closed form as*

$$\hat{g}_\nu(\mathbf{x}) = \sum_{r=1}^{N_r} \beta_{\nu r} K(||\mathbf{x} - \mathbf{x}_r||_2) + \boldsymbol{\alpha}'_{\nu 1} \mathbf{x} + \alpha_{\nu 0} \quad (5)$$

where $K(\rho) := \rho^2 \log(\rho)$, *and* $\boldsymbol{\beta}_\nu := [\beta_{\nu 1}, \ldots, \beta_{\nu N_r}]'$ *is constrained to the linear subspace* $\mathcal{B} := \{\boldsymbol{\beta} \in \mathbb{R}^{N_r} : \sum_{r=1}^{N_r} \beta_r = 0, \sum_{r=1}^{N_r} \beta_r \mathbf{x}_r = \mathbf{0}_2,\ \mathbf{x}_r \in \mathcal{X}\}$ *for* $\nu = 1, \ldots, N_b$.

The proof of this proposition is given in Appendix A.

Remark 1 (Overlapping frequency basis). If the basis functions $\{b_\nu(f)\}$ have finite supports which do not overlap, then (4) decouples per g_ν, and thus the results in [10], [26] can be applied directly. The novelty of Proposition 1 is that the basis functions with spatial spline coefficients in (1) are allowed to be *overlapping*. The implication of Proposition 1 is finite parametrization of the PSD map [cf. (5)]. This is particularly important for non-FDMA based CR networks. In the forthcoming Section IV, an overcomplete set $\{b_\nu\}$ is adopted in (1), and overlapping bases naturally arise therein.

What is left to determine are the parameters $\alpha := [\alpha_{10}, \boldsymbol{\alpha}'_{11}, \ldots, \alpha_{N_b 0}, \boldsymbol{\alpha}'_{N_b 1}]' \in \mathbb{R}^{3N_b}$, and $\boldsymbol{\beta} := [\boldsymbol{\beta}'_1, \ldots, \boldsymbol{\beta}'_{N_b}]' \in \mathbb{R}^{N_r N_b}$ in (5). To this end, define the vector $\boldsymbol{\varphi} := [\varphi_{11}, \ldots, \varphi_{1N}, \ldots, \varphi_{N_r 1}, \ldots, \varphi_{N_r N}]' \in \mathbb{R}^{N_r N}$ containing the network-wide data obtained at all frequencies in $\mathcal{F}$. Three matrices are also introduced collecting the regression inputs: i) $\mathbf{T} \in \mathbb{R}^{N_r \times 3}$ with rth row $\mathbf{t}'_r := [1\, \mathbf{x}'_r]$ for $r = 1, \ldots, N_r$ and $\mathbf{x}_r \in \mathcal{X}$; ii)

$\mathbf{B} \in \mathbb{R}^{N \times N_b}$ with nth row $\mathbf{b}_n' := [b_1(f_n), \ldots, b_{N_b}(f_n)]$ for $n = 1, \ldots, N$; and iii) $\mathbf{K} \in \mathbb{R}^{N_r \times N_r}$ with ij-th entry $[\mathbf{K}]_{ij} := K(||\mathbf{x}_i - \mathbf{x}_j||)$ for $\mathbf{x}_i, \mathbf{x}_j \in \mathcal{X}$. Consider also the QR decompositions of $\mathbf{T} = [\mathbf{Q}_1\ \mathbf{Q}_2]\,[\mathbf{R}'\ \mathbf{0}]'$ and $\mathbf{B} = [\mathbf{\Omega}_1\ \mathbf{\Omega}_2]\,[\mathbf{\Gamma}'\ \mathbf{0}]'$.

Upon plugging (5) into (4), it is shown in Appendix B that the optimal $\{\boldsymbol{\alpha}, \boldsymbol{\beta}\}$ satisfy the following system of equations

$$(\mathbf{B} \otimes \mathbf{Q}_2')\boldsymbol{\varphi} = \left[(\mathbf{B}'\mathbf{B} \otimes \mathbf{Q}_2'\mathbf{K}\mathbf{Q}_2) + N_r N \lambda \mathbf{I}_{N_b(N_r-3)}\right] \hat{\boldsymbol{\gamma}} \tag{6}$$

$$[\mathbf{\Gamma} \otimes \mathbf{R}]\hat{\boldsymbol{\alpha}} = (\mathbf{\Omega}_1' \otimes \mathbf{Q}_1')\boldsymbol{\varphi} - (\mathbf{\Gamma} \otimes \mathbf{Q}_1'\mathbf{K}\mathbf{Q}_2)\hat{\boldsymbol{\gamma}} \tag{7}$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{I}_{N_b} \otimes \mathbf{Q}_2)\hat{\boldsymbol{\gamma}}. \tag{8}$$

Matrix $\mathbf{Q}_2'\mathbf{K}\mathbf{Q}_2$ is positive definite, and $\text{rank}(\mathbf{\Gamma} \otimes \mathbf{R}) = \text{rank}(\mathbf{\Gamma})\text{rank}(\mathbf{R})$; see e.g., [18]. It thus follows that (6)-(7) admit a unique solution if and only if $\mathbf{\Gamma}$ and $\mathbf{R}$ are invertible (correspondingly, $\mathbf{B}$ and $\mathbf{T}$ have full column rank). These conditions place practical constraints that should be taken into account at the system design stage. Specifically, $\mathbf{T}$ has full column rank if and only if the points in $\mathcal{X}$, i.e., the CR locations, are not aligned. Furthermore, $\mathbf{B}$ will have linearly independent columns provided the basis functions $\{b_\nu(f)\}_{\nu=1}^{N_b}$ comprise a linearly independent and complete set, i.e., $B \subseteq \bigcup_\nu B_\nu$. Note that completeness precludes all frequencies $\{f_n\}_{n=1}^N$ from falling outside the aggregate support of the basis set, hence preventing undesired all-zero columns in $\mathbf{B}$.

Remark 2 (Practicality of uniqueness conditions). The condition on $\mathcal{X}$ does not introduce an actual limitation as it can be easily satisfied in practice, especially when the CRs are randomly deployed. Likewise, the basis set is part of the system design, and can be chosen to satisfy the conditions on $\mathbf{B}$. Nonetheless, these conditions will be bypassed in Section IV by allowing for an overcomplete set of functions $\{b_\nu\}$.

The combined results in this section can be summarized in the following steps constituting the spline-based spectrum cartography algorithm, which amounts to estimating $\Phi(\mathbf{x}, f)$:

S1. Given $\boldsymbol{\varphi}$, solve (6)-(8) for $\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}$, after selecting λ as detailed in Appendix D.

S2. Substitute $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ into (5) to obtain $\{\hat{g}_\nu(\mathbf{x})\}_{\nu=1}^{N_b}$.

S3. Use $\{\hat{g}_\nu(\mathbf{x})\}_{\nu=1}^{N_b}$ in (1) to estimate $\Phi(\mathbf{x}, f)$.

B. PSD tracker

The real-time requirements on the sensing radios and the convenience of an estimator that adapts to changes in the spectrum map are the motivating reasons behind the PSD tracker introduced in this section. The spectrum map estimator will be henceforth denoted by $\Phi(\mathbf{x}, f, \tau)$, to make its time dependence explicit.

Define the vector $\widehat{\boldsymbol{\phi}}_n(\tau) := [\widehat{\phi}_{1n}(\tau), \ldots, \widehat{\phi}_{N_r n}(\tau)]'$ of periodogram samples taken at frequency f_n by all CRs, and form the supervector $\widehat{\boldsymbol{\phi}}(\tau) := [\widehat{\boldsymbol{\phi}}_1'(\tau), \ldots, \widehat{\boldsymbol{\phi}}_N'(\tau)]' \in \mathbb{R}^{N_r N}$. Per time-slot $\tau = 1, 2, \ldots$, the periodogram $\widehat{\boldsymbol{\phi}}(\tau)$ is averaged using the following adaptive counterpart of (3):

$$\boldsymbol{\varphi}(\tau) := \sum_{\tau'=1}^{\tau} \delta^{\tau-\tau'} \widehat{\boldsymbol{\phi}}(\tau') = \delta \boldsymbol{\varphi}(\tau-1) + \widehat{\boldsymbol{\phi}}(\tau) \tag{9}$$

which implements an exponentially weighted moving average operation with forgetting factor $\delta \in (0,1)$. For every τ, the online estimator $\Phi(\mathbf{x}, f, \tau)$ is obtained by plugging in (1) the solution $\{\hat{g}_\nu(\mathbf{x}, \tau)\}_{\nu=1}^{N_b}$ of (4), after replacing φ_{rn} with $\varphi_{rn}(\tau)$ [cf. the entries of the vector in (9)]. In addition to mitigating fading effects, this adaptive approach can track slowly time-varying PSDs because the averaging in (9) exponentially discards past data.

Suppose that per time-slot τ, the FC receives raw periodogram samples $\widehat{\boldsymbol{\phi}}(\tau)$ from the CRs in order to update $\Phi(\mathbf{x}, f, \tau)$. The results of Section III apply for every τ, meaning that $\{\hat{g}_\nu(\mathbf{x}, \tau)\}_{\nu=1}^{N_b}$ are given by (5), while the optimum coefficients $\{\hat{\boldsymbol{\alpha}}(\tau), \hat{\boldsymbol{\beta}}(\tau)\}$ are found after solving (6)-(8). Capitalizing on (9), straightforward manipulations of (6)-(8) show that $\{\hat{\boldsymbol{\alpha}}(\tau), \hat{\boldsymbol{\beta}}(\tau)\}$ are recursively given for all $\tau \geq 1$ by

$$\hat{\boldsymbol{\beta}}(\tau) = \delta \hat{\boldsymbol{\beta}}(\tau-1) + (\mathbf{I}_{N_b} \otimes \mathbf{Q}_2) \mathbf{G}_1 \widehat{\boldsymbol{\phi}}(\tau) \tag{10}$$

$$\hat{\boldsymbol{\alpha}}(\tau) = \delta \hat{\boldsymbol{\alpha}}(\tau-1) + \mathbf{G}_2 \widehat{\boldsymbol{\phi}}(\tau) \tag{11}$$

where the *time-invariant* matrices $\mathbf{G}_1$ and $\mathbf{G}_2$ are

$$\mathbf{G}_1 := \left[(\mathbf{B}'\mathbf{B} \otimes \mathbf{Q}_2' \mathbf{K} \mathbf{Q}_2) + N_r N \lambda \mathbf{I}_{N_b(N_r-3)} \right]^{-1} (\mathbf{B} \otimes \mathbf{Q}_2')$$

$$\mathbf{G}_2 := [\boldsymbol{\Gamma} \otimes \mathbf{R}]^{-1} \left[(\boldsymbol{\Omega}_1' \otimes \mathbf{Q}_1') - (\boldsymbol{\Gamma} \otimes \mathbf{Q}_1' \mathbf{K} \mathbf{Q}_2) \mathbf{G}_1 \right].$$

Recursions (10)-(11) provide a means to update $\Phi(\mathbf{x}, f, \tau)$ sequentially in time, by incorporating the newly acquired data from the CRs in $\widehat{\boldsymbol{\phi}}(\tau)$. There is no need to separately update $\boldsymbol{\varphi}(\tau)$ as in (9), yet the desired averaging takes place. Furthermore, matrices $\mathbf{G}_1$ and $\mathbf{G}_2$ need to be computed only once, during the startup phase of the network.

IV. Group-Lasso on Splines

An improved spline-based PSD estimator is developed in this section to fit the unknown spatial functions $\{g_\nu\}_{\nu=1}^{N_b}$ in the model $\Phi(\mathbf{x}, f) = \sum_{\nu=1}^{N_b} g_\nu(\mathbf{x}) b_\nu(f)$, with a large ($N_b \gg N_r N$), and a possibly overcomplete set of known basis functions $\{b_\nu\}_{\nu=1}^{N_b}$. These models are particularly attractive when there is an inherent uncertainty on the transmitters' parameters, such as central frequency and bandwidth of the pulse shapers; or, e.g., the roll-off factor when raised-cosine pulses are employed. In particular, adaptive communication

schemes rely on frequently adjusting these parameters [12, Ch. 9]. A sizeable collection of bases to effectively accommodate most of the possible cases provides the desirable robustness. Still, prior knowledge available on the incumbent communication technologies being sensed should be exploited to choose the most descriptive classes of basis functions; e.g., a large set of raised-cosine pulses. This knowledge justifies why known bases are selected to describe frequency characteristics of the PSD map, while a variational approach is preferred to capture spatial dependencies.

In this context, the envisioned estimation method should provide the CRs with the capability of selecting a *few* bases that "better explain" the actual transmitted signals. As a result, most functions g_ν are expected to be identically zero; hence, there is an inherent form of sparsity present that can be exploited to improve estimation. The rationale behind the proposed approach can be rooted in the *basis pursuit* principle, a term coined in [8] for finding the most parsimonious sparse signal expansion using an overcomplete basis set. A major differentiating aspect however, is that while the sparse coefficients in the basis expansions treated in [8] are scalars, model (1) here entails bases weighted by functions g_ν.

The proposed approach to sparsity-aware spline-based field estimation from the space-frequency power spectrum measurements φ_{rn} [cf. (3)], is to obtain $\{\hat{g}_\nu\}_{\nu=1}^{N_b}$ as

$$\{\hat{g}_\nu\}_{\nu=1}^{N_b} := \arg\min_{\{g_\nu \in \mathcal{S}\}} \quad \frac{1}{N_r N}\sum_{r=1}^{N_r}\sum_{n=1}^{N}\left(\varphi_{rn} - \sum_{\nu=1}^{N_b} g_\nu(\mathbf{x}_r) b_\nu(f_n)\right)^2 + \lambda \sum_{\nu=1}^{N_b}\int_{\mathbb{R}^2} ||\nabla^2 g_\nu(\mathbf{x})||_F^2 d\mathbf{x} + \mu \sum_{\nu=1}^{N_b} \left\| [g_\nu(\mathbf{x}_1), \ldots, g_\nu(\mathbf{x}_{N_r})]' \right\|_2 . \tag{12}$$

Relative to (4), the cost here is augmented with an additional regularization term weighted by a tuning parameter $\mu \geq 0$. Clearly, if $\mu = 0$ then (12) boils down to (4). To appreciate the role of the new penalty term, note that the minimization of $\left\| [g_\nu(\mathbf{x}_1), \ldots, g_\nu(\mathbf{x}_{N_r})]' \right\|_2$ intuitively shrinks all pointwise functional values $\{g_\nu(\mathbf{x}_1), \ldots, g_\nu(\mathbf{x}_{N_r})\}$ to zero for sufficiently large μ. Interestingly, it will be shown in the ensuing section that this is enough to guarantee that $\hat{g}_\nu(\mathbf{x}) \equiv 0 \ \forall \mathbf{x}$, for μ large enough.

A. Estimation using the group-Lasso

Consider the classical problem of linear regression; see, e.g. [13, p. 11], where a vector $\mathbf{y} \in \mathbb{R}^n$ of observations is available, along with a matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$ of inputs. The group Lasso estimate for the vector of features $\boldsymbol{\zeta} := [\boldsymbol{\zeta}_1', \ldots, \boldsymbol{\zeta}_{N_b}']' \in \mathbb{R}^p$ is defined as the solution to [3], [29]

$$\min_{\boldsymbol{\zeta}} \frac{1}{2}\|\mathbf{y} - \mathbf{X}\boldsymbol{\zeta}\|_2^2 + \mu \sum_{\nu=1}^{N_b} \|\boldsymbol{\zeta}_\nu\|_2. \tag{13}$$

This criterion achieves model selection by retaining relevant factors $\boldsymbol{\zeta}_\nu \in \mathbb{R}^{p/N_b}$ in which the features are grouped. In other words, group-Lasso encourages sparsity at the factor level, either by shrinking to zero all variables within a factor, or by retaining them altogether depending on the value of the tuning parameter $\mu \geq 0$. As μ is increased, more sub-vector estimates $\boldsymbol{\zeta}_\nu$ become zero, and the corresponding factors drop out of the model. It can be shown from the Karush-Kuhn-Tucker optimality conditions that only for $\mu \geq \mu_{\max} := \max_i \|\mathbf{X}_i'\mathbf{y}\|_2$ it holds that $\boldsymbol{\zeta}_1 = \ldots = \boldsymbol{\zeta}_{N_b} = \mathbf{0}_{p/N_b}$, so that the values of interest are $\mu \in [0, \mu_{\max}]$ [2].

The connection between (13) and the spline-based field estimator (12) builds on Proposition 1, which still holds in this context. That is, even though criteria (4) and (12) purposely differ, their respective solutions $\hat{g}_\nu(\mathbf{x})$ have the same form in (5). Indeed, the adaptation of the proof in Appendix A to the new case is straightforward, since the additional penalty term in (12) depends on g_ν evaluated at the knots. The essential difference manifested by this penalty is revealed when estimating the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in (5), as presented in the following proposition.

Proposition 2: *The spline-based field estimator* (12) *is equivalent to group-Lasso* (13)*, under the identities*

$$\mathbf{y} := \frac{1}{\sqrt{N_r N}}[\boldsymbol{\varphi}',\ \mathbf{0}]', \quad \mathbf{X} := \frac{1}{\sqrt{N_r N}} \begin{bmatrix} \mathbf{B} \otimes \mathbf{I}_{N_r} \\ \mathbf{I}_{N_b} \otimes \left\{ bdiag((N_r N \lambda \mathbf{Q}_2'\mathbf{K}\mathbf{Q}_2)^{1/2}, \mathbf{0})[\mathbf{K}\mathbf{Q}_2 \ \ \mathbf{T}]^{-1} \right\} \end{bmatrix} \tag{14}$$

with their respective solutions related by

$$\hat{g}_\nu(\mathbf{x}) = \sum_{r=1}^{N_r} \beta_{\nu r} K(||\mathbf{x} - \mathbf{x}_r||_2) + \boldsymbol{\alpha}_{\nu 1}'\mathbf{x} + \alpha_{\nu 0} \tag{15}$$

$$\left[\boldsymbol{\beta}_\nu', \boldsymbol{\alpha}_\nu'\right]' = bdiag(\mathbf{Q}_2, \mathbf{I}_3)[\mathbf{K}\mathbf{Q}_2 \ \ \mathbf{T}]^{-1}\hat{\boldsymbol{\zeta}}_\nu \tag{16}$$

where $\boldsymbol{\beta}_\nu := [\beta_{\nu 1}, \ldots, \beta_{\nu N_r}]'$ *and* $\boldsymbol{\alpha}_\nu := [\alpha_{\nu 0}, \boldsymbol{\alpha}_{\nu 1}']'$.

The factors $\{\boldsymbol{\zeta}_\nu\}_{\nu=1}^{N_b}$ in (13) are in one-to-one correspondence with the vectors $\{[\boldsymbol{\beta}_\nu', \boldsymbol{\alpha}_\nu']'\}_{\nu=1}^{N_b}$ through the linear mapping (16). This implies that whenever a factor $\boldsymbol{\zeta}_\nu$ is dropped from the linear regression model obtained after solving (13), then $\hat{g}_\nu(\mathbf{x}) \equiv 0$, and the term corresponding to $b_\nu(f)$ does not contribute to (1). Hence, by appropriately selecting the value of μ, criterion (12) has the potential of retaining only the most significant terms in $\Phi(\mathbf{x}, f) = \sum_{\nu=1}^{N_b} g_\nu(\mathbf{x}) b_\nu(f)$, and thus yields parsimonious PSD map estimates. All in all, the motivation behind the variational problem (12) is now unravelled. The additional penalty term not present in (4) renders (12) equivalent to a group-Lasso problem. This enforces sparsity in the parameters of the splines expansion for $\Phi(\mathbf{x}, f)$ at a factor level, which is exactly what is needed to potentially null the less descriptive functions g_ν.

Remark 3 (Comparison with the PSD map estimator in Section III). The sparsity-agnostic LS problem

(4) will not give rise to identically zero vectors $\{\boldsymbol{\alpha}_\nu, \boldsymbol{\beta}_\nu\}$, for any ν. Even when N_b is not large, a sparsity-aware estimator will perform better if the underlying PSD is generated by a few basis functions. This is expected since the out-of-band residual error will increase when all basis functions enter the model (1); see also [4] for a related assessment. What is more, when the number of bases is sufficiently large ($N_b \gg N_r N$) matrix $\mathbf{B}$ is fat, and the approach in Section III is not applicable . On the other hand, it is admittedly more complex computationally to solve (13) than the system of linear equations (6)-(8). Because (12) is not a linear smoother, a leave-one-out (bi-) CV approach to select the tuning parameters λ and μ does not enjoy the computational savings detailed in Appendix D. K-fold CV can be utilized instead, with typical choices of $K = 5$ or 10, as suggested in [13, p. 242].

The group-Lassoed splines-based approach to spectrum cartography developed in this section can be summarized in the following steps to estimate the global PSD map $\Phi(\mathbf{x}, f)$:

S1. Given $\boldsymbol{\varphi}$ and utilizing any group Lasso solver, obtain $\hat{\boldsymbol{\zeta}} := [\hat{\boldsymbol{\zeta}}_1', \ldots, \hat{\boldsymbol{\zeta}}_{N_b}']'$ by solving (13).

S2. Form the estimates $\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}$ using the change of variables $[\hat{\boldsymbol{\beta}}_\nu', \hat{\boldsymbol{\alpha}}_\nu']' = \text{bdiag}(\mathbf{Q}_2, \mathbf{I}_3)[\mathbf{K}\mathbf{Q}_2 \ \mathbf{T}]^{-1}\hat{\boldsymbol{\zeta}}_\nu$ for $\nu = 1, \ldots, N_b$.

S3. Substitute $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ into (15) to obtain $\{\hat{g}_\nu(\mathbf{x})\}_{\nu=1}^{N_b}$.

S4. Use $\{\hat{g}_\nu(\mathbf{x})\}_{\nu=1}^{N_b}$ in (1) to estimate $\Phi(\mathbf{x}, f)$.

Implementing S1-S4 presumes that CRs communicate their local PSD estimates to a fusion center, which uses their aggregation in $\boldsymbol{\varphi}$ to estimate the field. But what if an FC is not available for centrally running S1-S4? In certain cases, forgoing with an FC is reasonable when the designer wishes to avoid an isolated point of failure, or, aims at a network topology which scales well with an increasing number of CRs based on power considerations (CRs located far away from the FC will drain their batteries more to reach the FC). These reasons motivate well a fully distributed counterpart of S1-S4, which is pursued next.

V. Distributed Group-Lasso for In-Network Spectrum Cartography

Consider N_r networked CRs that are capable of sensing the ambient RF spectrum, performing some local computations, as well as exchanging messages among neighbors via dedicated control channels. In lieu of a fusion center, the CR network is naturally modeled as an undirected graph $\mathcal{G}(\mathcal{R}, \mathcal{E})$, where the vertex set $\mathcal{R} := \{1, \ldots, N_r\}$ corresponds to the sensing radios, and the edges in $\mathcal{E}$ represent pairs of CRs that can communicate. Radio $r \in \mathcal{R}$ communicates with its single-hop neighbors in $\mathcal{N}_r$, and the size of the neighborhood is denoted by $|\mathcal{N}_r|$. The locations $\{\mathbf{x}_r\}_{r=1}^{N_r} := \mathcal{X}$ of the sensing radios are assumed known to the CR network. To ensure that the measured data from an arbitrary CR can eventually percolate throughout the entire network, it is assumed that the graph $\mathcal{G}$ is *connected*; i.e., there exists a (possibly)

multi-hop communication path connecting any two CRs.

For the purpose of estimating an unknown vector $\boldsymbol{\zeta} := \left[\boldsymbol{\zeta}_1', \ldots, \boldsymbol{\zeta}_{N_b}'\right]' \in \mathbb{R}^p$, each radio $r \in \mathcal{R}$ has available a local vector of observations $\mathbf{y}_r \in \mathbb{R}^{n_r}$ as well as its own matrix of inputs $\mathbf{X}_r \in \mathbb{R}^{n_r \times p}$. Radios collaborate to form the wanted group-Lasso estimator (13) in a distributed fashion, using

$$\hat{\boldsymbol{\zeta}}_{\text{glasso}} = \arg\, \min_{\boldsymbol{\zeta}} \frac{1}{2} \sum_{r=1}^{N_r} \|\mathbf{y}_r - \mathbf{X}_r \boldsymbol{\zeta}\|_2^2 + \mu \sum_{\nu=1}^{N_b} \|\boldsymbol{\zeta}_\nu\|_2 \tag{17}$$

where $\mathbf{y} := [\mathbf{y}_1', \ldots, \mathbf{y}_{N_r}']' \in \mathbb{R}^{n \times 1}$ with $n := \sum_{r=1}^{N_r} n_r$, and $\mathbf{X} := [\mathbf{X}_1', \ldots, \mathbf{X}_{N_r}']' \in \mathbb{R}^{n \times p}$. The motivation behind developing a distributed solver of (17) is to tackle (12) based on in-network processing of the local observations $\boldsymbol{\varphi}_r := [\varphi_{r1}, \ldots, \varphi_{rN}]'$ available per radio [cf. (3)]. Indeed, it readily follows that (17) can be used instead of (13) in Proposition 2 when

$$\mathbf{y}_r := \frac{1}{\sqrt{N_r N}} \begin{bmatrix} \boldsymbol{\varphi}_r \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{X}_\mathbf{r} := \frac{1}{\sqrt{N_r N}} \begin{bmatrix} \mathbf{B} \otimes \mathbf{e}_{N_r, r}' \\ \mathbf{I}_{N_b} \otimes \left[\text{bdiag}((N_r N \lambda \mathbf{Q}_2' \mathbf{K} \mathbf{Q}_2)^{1/2}, \mathbf{0}) [\mathbf{K}\mathbf{Q}_2 \;\; \mathbf{T}]^{-1}\right] \end{bmatrix}, \quad r \in \mathcal{R}$$

corresponding to the identifications $n_r = N \; \forall r \in \mathcal{R}$, $p = N_b N_r$. Note that because the locations $\{\mathbf{x}_r\}$ are assumed known to the entire network, CR r can form matrices $\mathbf{T}$, $\mathbf{K}$, and thus, the local regression matrix $\mathbf{X}_r$.

A. Consensus-based reformulation of the group-Lasso

To distribute the cost in (17), replace the *global* variable $\boldsymbol{\zeta}$ which couples the per-agent summands with *local* variables $\{\boldsymbol{\zeta}_r\}_{r=1}^{N_r}$ representing candidate estimates of $\boldsymbol{\zeta}$ per sensing radio. It is now possible to reformulate (17) as a convex *constrained* minimization problem

$$\left\{\hat{\boldsymbol{\zeta}}_r\right\}_{r=1}^{N_r} = \arg\, \min_{\{\boldsymbol{\zeta}_r\}} \frac{1}{2} \sum_{r=1}^{N_r} \left[\|\mathbf{y}_r - \mathbf{X}_r \boldsymbol{\zeta}_r\|_2^2 + \frac{2\mu}{N_r} \sum_{\nu=1}^{N_b} \|\boldsymbol{\zeta}_{r\nu}\|_2\right] \tag{18}$$

$$\text{s. t.} \qquad \boldsymbol{\zeta}_r = \boldsymbol{\zeta}_{r'},\; r \in \mathcal{R},\; r' \in \mathcal{N}_r,\; \boldsymbol{\zeta}_r := \left[\boldsymbol{\zeta}_{r1}', \ldots, \boldsymbol{\zeta}_{rN_b}'\right]'.$$

The equality constraints directly effect local agreement across each CR's neighborhood. Since the communication graph $\mathcal{G}$ is assumed connected, these constraints also ensure *global* consensus a fortiori, meaning that $\boldsymbol{\zeta}_r = \boldsymbol{\zeta}_{r'}$, $\forall r, r' \in \mathcal{R}$. Indeed, let $P(r, r') : r, r_1, r_2, \ldots, r_n, r'$ denote a path on $\mathcal{G}$ that joins an arbitrary pair of CRs (r, r'). Because contiguous radios in the path are neighbors by definition, the corresponding chain of equalities $\boldsymbol{\zeta}_r = \boldsymbol{\zeta}_{r_1} = \boldsymbol{\zeta}_{r_2} = \ldots = \boldsymbol{\zeta}_{r_n} = \boldsymbol{\zeta}_{r'}$ dictated by the constraints in (18) imply $\boldsymbol{\zeta}_r = \boldsymbol{\zeta}_{r'}$, as desired. Thus, the constraints can be eliminated by replacing all the $\{\boldsymbol{\zeta}_r\}$ with a common $\boldsymbol{\zeta}$, in which case the cost in (18) reduces to the one in (17). This argument establishes the following result.

Lemma 1: *If $\mathcal{G}$ is a connected graph,* (17) *and* (18) *are equivalent optimization problems, in the sense that $\hat{\boldsymbol{\zeta}}_{glasso} = \hat{\boldsymbol{\zeta}}_r$, $\forall\, r \in \mathcal{R}$.*

Problem (18) will be modified further for the purpose of reducing the computational complexity of the resulting algorithm. To this end, for a given $\mathbf{a} \in \mathbb{R}^p$ consider the problem

$$\min_{\boldsymbol{\zeta}} \frac{1}{2}||\boldsymbol{\zeta}||_2^2 - \mathbf{a}'\boldsymbol{\zeta} + \mu \sum_{\nu=1}^{N_b} \|\boldsymbol{\zeta}_\nu\|_2, \qquad \boldsymbol{\zeta} := [\boldsymbol{\zeta}_1', \ldots, \boldsymbol{\zeta}_{N_b}']' \tag{19}$$

and notice that it is separable in the N_b subproblems

$$\min_{\boldsymbol{\zeta}_\nu} \frac{1}{2}||\boldsymbol{\zeta}_\nu||_2^2 - \mathbf{a}_\nu'\boldsymbol{\zeta}_\nu + \mu\|\boldsymbol{\zeta}_\nu\|_2, \qquad \mathbf{a} := [\mathbf{a}_1', \ldots, \mathbf{a}_{N_b}']'. \tag{20}$$

Interestingly, each of these subproblems admits a closed-form solution as given in the following lemma.

Lemma 2: *The minimizer $\boldsymbol{\zeta}_\nu^\star$ of* (20) *is obtained via the vector soft-thresholding operator $\mathcal{T}_\mu(\cdot)$ defined by*

$$\boldsymbol{\zeta}_\nu^\star = \mathcal{T}_\mu(\mathbf{a}_\nu) := \frac{\mathbf{a}_\nu}{\|\mathbf{a}_\nu\|_2} \left(\|\mathbf{a}_\nu\|_2 - \mu\right)_+ \tag{21}$$

where $(\cdot)_+ := \max\{\cdot, 0\}$.

Problem (19) is an instance of the group-Lasso (13) when $\mathbf{X}'\mathbf{X} = \mathbf{I}_p$, and $\mathbf{a} := \mathbf{X}'\mathbf{y}$. As such, result (21) can be viewed as a particular case of the operators in [22] and [28]. However it is worth to prove Lemma 2 directly, since in this case the special form of (20) renders the proof neat in its simplicity.

Proof: It will be argued that the solver of (20) takes the form $\boldsymbol{\zeta}_\nu^\star = t\mathbf{a}_\nu$ for some scalar $t \geq 0$. This is because among all $\boldsymbol{\zeta}_\nu$ with the same ℓ_2-norm, the Cauchy-Schwarz inequality implies that the maximizer of $\mathbf{a}_\nu'\boldsymbol{\zeta}_\nu$ is colinear with (and in the same direction of) $\mathbf{a}_\nu$. Substituting $\boldsymbol{\zeta}_\nu = t\mathbf{a}_\nu$ into (20) renders the problem scalar in $t \geq 0$, with solution $t^\star = \left(\|\mathbf{a}_\nu\| - \mu\right)_+ / \left(2\|\mathbf{a}_\nu\|\right)$, which completes the proof. ■

In order to take advantage of Lemma 2, auxiliary variables $\boldsymbol{\gamma}_r,\ r = 1, \ldots, N_r$ are introduced as copies of $\boldsymbol{\zeta}_r$. Upon introducing appropriate constraints $\boldsymbol{\gamma}_r = \boldsymbol{\zeta}_r$ that guarantee the equivalence of the formulations along the lines of Lemma 1, problem (18) can be recast as

$$\begin{aligned} \left\{\hat{\boldsymbol{\zeta}}_r\right\}_{r=1}^{N_r} = \ & \arg\min_{\{\boldsymbol{\zeta}_r, \boldsymbol{\gamma}_r, \boldsymbol{\gamma}_r^{r'}\}} \frac{1}{2} \sum_{r=1}^{N_r} \left[\|\mathbf{y}_r - \mathbf{X}_r\boldsymbol{\gamma}_r\|_2^2 + \frac{2\mu}{N_r} \sum_{\nu=1}^{N_b} \|\boldsymbol{\zeta}_{r\nu}\|_2 \right] \\ & \text{s. to} \qquad \boldsymbol{\zeta}_r = \boldsymbol{\gamma}_r^{r'} = \boldsymbol{\zeta}_{r'},\ r \in \mathcal{R},\ r' \in \mathcal{N}_r \\ & \qquad\quad \boldsymbol{\gamma}_r = \boldsymbol{\zeta}_r,\ r \in \mathcal{R}. \end{aligned} \tag{22}$$

The dummy variables $\boldsymbol{\gamma}_r^{r'}$ are inserted for technical reasons that will become apparent in the ensuing section, and will be eventually eliminated.

B. Distributed group-Lasso algorithm

The distributed group-Lasso algorithm is constructed by optimizing (22) using the alternating direction method of multipliers (AD-MoM) [6]. In this direction, associate Lagrange multipliers $\mathbf{v}_r, \bar{\mathbf{v}}_r^{r'}$ and $\check{\mathbf{v}}_r^{r'}$ with

the constraints $\boldsymbol{\gamma}_r = \boldsymbol{\zeta}_r$, $\boldsymbol{\zeta}_{r'} = \boldsymbol{\gamma}_r^{r'}$ and $\boldsymbol{\zeta}_r = \boldsymbol{\gamma}_r^{r'}$, respectively, and consider the augmented Lagrangian with parameter $c > 0$

$$\begin{aligned}\mathcal{L}_c\left[\{\boldsymbol{\zeta}_r\}, \boldsymbol{\gamma}, \boldsymbol{v}\right] =& \frac{1}{2}\sum_{r=1}^{N_r}\left[\|\mathbf{y}_r - \mathbf{X}_r\boldsymbol{\gamma}_r\|_2^2 + \frac{2\mu}{N_r}\sum_{\nu=1}^{N_b}\|\boldsymbol{\zeta}_{r\nu}\|_2\right] + \sum_{r=1}^{N_r}\mathbf{v}_r'(\boldsymbol{\zeta}_r - \boldsymbol{\gamma}_r) + \frac{c}{2}\sum_{r=1}^{N_r}\|\boldsymbol{\zeta}_r - \boldsymbol{\gamma}_r\|_2^2 \\ &+ \sum_{r=1}^{N_r}\sum_{r'\in\mathcal{N}_r}\left[(\breve{\mathbf{v}}_r^{r'})'(\boldsymbol{\zeta}_r - \boldsymbol{\gamma}_r^{r'}) + (\bar{\mathbf{v}}_r^{r'})'(\boldsymbol{\zeta}_{r'} - \boldsymbol{\gamma}_r^{r'})\right] + \frac{c}{2}\sum_{r=1}^{N_r}\sum_{r'\in\mathcal{N}_r}\left[\|\boldsymbol{\zeta}_r - \boldsymbol{\gamma}_r^{r'}\|_2^2 + \|\boldsymbol{\zeta}_{r'} - \boldsymbol{\gamma}_r^{r'}\|_2^2\right]\end{aligned} \tag{23}$$

where for notational convenience we group the variables $\boldsymbol{\gamma} := \{\boldsymbol{\gamma}_r, \{\boldsymbol{\gamma}_r^{r'}\}_{r'\in\mathcal{N}_r}\}_{r\in\mathcal{R}}$, and multipliers $\boldsymbol{v} := \{\mathbf{v}_r, \{\breve{\mathbf{v}}_r^{r'}\}_{r'\in\mathcal{N}_r}, \{\bar{\mathbf{v}}_r^{r'}\}_{r'\in\mathcal{N}_r}\}_{r\in\mathcal{R}}$.

Application of the AD-MoM to the problem at hand consists of a cycle of $\mathcal{L}_c$ minimizations in a block-coordinate fashion w.r.t. $\{\boldsymbol{\zeta}_r\}$ firstly, and $\boldsymbol{\gamma}$ secondly, together with an update of the multipliers per iteration $k = 0, 1, 2, \ldots$. Deferring the details to Appendix E, the four main properties of this procedure that are instrumental to the resulting algorithm can be highlighted as follows.

i) Thanks to the introduction of the local copies $\boldsymbol{\zeta}_r$ and the dummy variables $\boldsymbol{\gamma}_r^{r'}$, the minimizations of $\mathcal{L}_c$ w.r.t. both $\{\boldsymbol{\zeta}_r\}$ and $\boldsymbol{\gamma}$ decouple per CR r, thus enabling distribution of the algorithm. Moreover, the constraints in (22) involve variables of neighboring CRs only, which allows the required communications to be local within each CR's neighborhood.

ii) Introduction of the variables $\boldsymbol{\gamma}_r$ separates the quadratic cost $\|\mathbf{y}_r - \mathbf{X}_r\boldsymbol{\gamma}_r\|_2^2$ from the group-Lasso penalty $\sum_{\nu=1}^{N_b}\|\boldsymbol{\zeta}_{r\nu}\|_2$. As a result, minimization of (23) w.r.t. $\boldsymbol{\zeta}_r$ takes the form of (19), which admits a closed-form solution via the vector soft-thresholding operator $\mathcal{T}_\mu(\cdot)$ in Lemma 2.

iii) Minimization of (23) w.r.t. $\boldsymbol{\gamma}$ consists of an unconstrained quadratic problem, which can also be solved in closed form. In particular, the optimal $\boldsymbol{\gamma}_r^{r'}$ at iteration k takes the value $\boldsymbol{\gamma}_r^{r'}(k) = \left(\boldsymbol{\zeta}_r(k) + \boldsymbol{\zeta}_{r'}(k)\right)/2$, and thus can be eliminated.

iv) It turns out that it is not necessary to carry out updates of the Lagrange multipliers $\{\bar{\mathbf{v}}_r^{r'},\ \breve{\mathbf{v}}_r^{r'}\}_{r'\in\mathcal{N}_r}$ separately, but only of their sums which are henceforth denoted by $\mathbf{p}_r := \sum_{r'\in\mathcal{N}_r}\left(\bar{\mathbf{v}}_r^{r'} + \breve{\mathbf{v}}_r^{r'}\right)$. Hence, there is one price $\mathbf{p}_r$ per CR $r = 1, \ldots, N_r$, which can be updated locally.

Building on these four features, it is established in Appendix E that the proposed AD-MoM scheme boils down to four parallel recursions run locally per CR:

$$\mathbf{p}_r(k) = \mathbf{p}_r(k-1) + c\sum_{r'\in\mathcal{N}_r}\left[\boldsymbol{\zeta}_r(k) - \boldsymbol{\zeta}_{r'}(k)\right] \tag{24}$$

$$\mathbf{v}_r(k) = \mathbf{v}_r(k-1) + c[\boldsymbol{\zeta}_r(k) - \boldsymbol{\gamma}_r(k)] \tag{25}$$

Algorithm 1 : DGLasso

All radios $r \in \mathcal{R}$ initialize $\{\boldsymbol{\zeta}_r(0), \boldsymbol{\gamma}_r(0), \mathbf{p}_r(-1), \mathbf{v}_r(-1)\}$ to zero, and locally run:

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

Transmit $\boldsymbol{\zeta}_r(k)$ to neighbors in $\mathcal{N}_r$.

Update $\mathbf{p}_r(k)$ via $\mathbf{p}_r(k) = \mathbf{p}_r(k-1) + c\sum_{r' \in \mathcal{N}_r}[\boldsymbol{\zeta}_r(k) - \boldsymbol{\zeta}_{r'}(k)]$.

Update $\mathbf{v}_r(k)$ via $\mathbf{v}_r(k) = \mathbf{v}_r(k-1) + c[\boldsymbol{\zeta}_r(k) - \boldsymbol{\gamma}_r(k)]$.

Update $\boldsymbol{\zeta}_r(k+1)$ using (26).

Update $\boldsymbol{\gamma}_r(k+1)$ using (27).

end for

$$\boldsymbol{\zeta}_{r\nu}(k+1) = \frac{\mathcal{T}_\mu\left(N_r\left(c\boldsymbol{\gamma}_{r\nu}(k) + c\sum_{r' \in \mathcal{N}_r}[\boldsymbol{\zeta}_{r\nu}(k) + \boldsymbol{\zeta}_{r'\nu}(k)] - \mathbf{p}_{r\nu}(k) - \mathbf{v}_{r\nu}(k)\right)\right)}{cN_r(2|\mathcal{N}_r| + 1)}, \ \nu = 1, \ldots, N_b \quad (26)$$

$$\boldsymbol{\gamma}_r(k+1) = \left[c\mathbf{I}_p + \mathbf{X}_r'\mathbf{X}_r\right]^{-1}\left(\mathbf{X}_r'\mathbf{y}_r + c\boldsymbol{\zeta}_r(k+1) + \mathbf{v}_r(k)\right). \quad (27)$$

Recursions (24)-(27) comprise the novel DGLasso algorithm, tabulated as Algorithm 1.

The algorithm entails the following steps. During iteration $k+1$, CR r receives the local estimates $\{\boldsymbol{\zeta}_{r'}(k)\}_{r' \in \mathcal{N}_r}$ from the neighboring CRs and plugs them into (24) to evaluate the dual price vector $\mathbf{p}_r(k)$. The new multiplier $\mathbf{v}_r(k)$ is then obtained using the locally available vectors $\{\boldsymbol{\gamma}_r(k), \boldsymbol{\zeta}_r(k)\}$. Subsequently, vectors $\{\mathbf{p}_r(k), \mathbf{v}_r(k)\}$ are jointly used along with $\{\boldsymbol{\zeta}_{r'}(k)\}_{r' \in \mathcal{N}_r}$ to obtain $\boldsymbol{\zeta}_r(k+1)$ via N_b parallel vector soft-thresholding operations $\mathcal{T}_\mu(\cdot)$ as in (21). Finally, the updated $\boldsymbol{\gamma}_r(k+1)$ is obtained from (27), and requires the previously updated quantities along with the vector of local observations $\mathbf{y}_r$ and regression matrix $\mathbf{X}_r$. The $(k+1)$st iteration is concluded after CR r broadcasts $\boldsymbol{\zeta}_r(k+1)$ to its neighbors. Even if an arbitrary initialization is allowed, the sparse nature of the estimator sought suggests the all-zero vectors as a natural choice. Three additional remarks are now in order.

Remark 4 (Distributed Lasso algorithm as a special case). When $N_b = p$ and there are as many groups as entries of $\boldsymbol{\zeta}$, then the sum $\sum_{\nu=1}^{N_b}\|\boldsymbol{\zeta}_\nu\|$ becomes the ℓ_1-norm of $\boldsymbol{\zeta}$, and group-Lasso reduces to Lasso. In this case, DGLasso offers a distributed algorithm to solve Lasso that coincides with the one in [5].

Remark 5 (Centralized Group-Lasso algorithm as a special case). For $N_r = 1$, the network consists of a single CR. In this case, DGLasso yields a novel algorithm for the centralized group-Lasso estimator (17), which is specified as Algorithm 2. Notice that the thresholding operator $\mathcal{T}_\mu$ in GLasso sets the entire sub-vector $\boldsymbol{\zeta}_\nu(k+1)$ to zero whenever $\|c\boldsymbol{\gamma}_\nu(k) - \mathbf{v}_\nu(k)\|_2$ does not exceed μ, in par with the group-sparsifying property of group-Lasso. Different from [29], GLasso can handle a general (not orthonormal) regression matrix $\mathbf{X}$. Compared to the block-coordinate algorithm of [22], GLasso does not require an inner Newton-Raphson recursion per iteration. If in addition $N_b = p$, then GLasso yields the Lasso estimator.

Algorithm 2 : GLasso

Initialize $\{\boldsymbol{\zeta}(0), \boldsymbol{\gamma}(0), \mathbf{v}(-1)\}$ to zero, and run:

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

Update $\mathbf{v}(k) = \mathbf{v}(k-1) + c[\boldsymbol{\zeta}(k) - \boldsymbol{\gamma}(k)]$.

Update $\boldsymbol{\zeta}_\nu(k+1) = (1/c)\mathcal{T}_\mu\left(c\boldsymbol{\gamma}_\nu(k) - \mathbf{v}_\nu(k)\right),\ \nu = 1, \ldots, N_b$.

Update $\boldsymbol{\gamma}(k+1) = \left[c\mathbf{I}_p + \mathbf{X}'\mathbf{X}\right]^{-1}\left(\mathbf{X}'\mathbf{y} + c\boldsymbol{\zeta}(k+1) + \mathbf{v}(k)\right)$.

end for

Remark 6 (Computational load balancing). Update (27) involves inversion of the $p \times p$ matrix $c\mathbf{I}_p + \mathbf{X}_r'\mathbf{X}_r$, that may be computationally demanding for sufficiently large p. Fortunately, this operation can be carried out offline before running the algorithm. More importantly, the matrix inversion lemma can be invoked to obtain $\left[c\mathbf{I}_p + \mathbf{X}_r'\mathbf{X}_r\right]^{-1} = c^{-1}\left[\mathbf{I}_p - \mathbf{X}_r'\left(c\mathbf{I}_{n_r} + \mathbf{X}_r\mathbf{X}_r'\right)^{-1}\mathbf{X}_r\right]$. In this new form, the dimensionality of the matrix to invert becomes $n_r \times n_r$, where n_r is the number of locally acquired data. For highly underdetermined cases $(n_r \ll p)$, (D)GLasso enjoys considerable computational savings through the aforementioned matrix inversion identity. One also recognizes that the distributed operation parallelizes the numerical computation across CRs: if GLasso is run at a central unit with all network-wide data available centrally, then the matrix to invert has dimension $n = \sum_{r\in\mathcal{R}} n_r$, which increases linearly with the network size N_r. Beyond a networked scenario, DGLasso provides an attractive alternative for computational load balancing in contemporary multi-processor architectures.

To close this section, it is useful to mention that convergence of Algorithm 1, and thus of Algorithm 2 as well, is ensured by the convergence of the AD-MoM [6]. This result is formally stated next.

Proposition 3: *Let $\mathcal{G}$ be a connected graph, and consider recursions* (24)-(27) *that comprise the DGLasso algorithm. Then, for any value of the step-size $c > 0$, the iterates $\boldsymbol{\zeta}_r(k)$ converge to the group-Lasso solution [cf.* (17)*] as $k \to \infty$, i.e.,*

$$\lim_{k\to\infty} \boldsymbol{\zeta}_r(k) = \hat{\boldsymbol{\zeta}}_{glasso},\ \forall\, r \in \mathcal{R}. \tag{28}$$

In words, all local estimates $\boldsymbol{\zeta}_r(k)$ achieve consensus asymptotically, converging to a common vector that coincides with the desired estimator $\hat{\boldsymbol{\zeta}}_{\text{glasso}}$. Formally, if the number of parameters p exceeds the number of data n, then a unique solution of (13) is not guaranteed for a general design matrix $\mathbf{X}$. Proposition 3 remains valid however, if the right-hand side of (28) is replaced by the set of minima; that is,

$$\lim_{k\to\infty} \boldsymbol{\zeta}_r(k) \in \arg\min_{\boldsymbol{\zeta}} \frac{1}{N_r}\sum_{r=1}^{N_r} \|\mathbf{y}_r - \mathbf{X}_r\boldsymbol{\zeta}\|_2^2 + \mu\sum_{\nu=1}^{N_b} \|\boldsymbol{\zeta}_\nu\|_2.$$

VI. NUMERICAL TESTS

Consider a set of $N_r = 100$ CRs uniformly distributed in an area of $1\mathrm{Km}^2$, cooperating to estimate the PSD map generated by $N_s = 5$ licensed users (sources) located as in Fig. 2 (left). The five transmitted signals are raised cosine pulses with roll-off factors $\rho \in \{0, 1\}$, and bandwidths $W \in \{10, 20, 30\}$ MHz. They share the frequency band $B = [100, 260]$ MHz with spectra centered at frequencies $f_c = 105,\ 140,\ 185,\ 215$, and 240 MHz, respectively. Fig. 2 (right) depicts the PSD generated by the active transmitters.

The PSD generated by source s experiences fading and shadowing effects in its propagation from $\mathbf{x}_s$ to any location $\mathbf{x}$, where it can be measured in the presence of noise. A 6-tap Rayleigh model is adopted for the multipath channel $H_s(f, \tau, \mathbf{x})$ between $\mathbf{x}_s$ and $\mathbf{x}$, whose expected gain adheres to the path-loss law $E(|H_s|^2) = \exp(-||\mathbf{x}_s - \mathbf{x}||_2^2/\Delta^2)$, with $\Delta = 0.8$. A deterministic shadowing effect is generated by a 18m-high and 500m-wide wall represented by the black segment in Fig. 2 (left). It produces a knife-edge effect on the power emitted by the antennas at a height of 20m. The simulated tests presented here account for the shadowing at ground level.

A. Spectrum cartography

When designing the basis functions in (1), it is known a priori that the transmitted signals are indeed normalized raised cosine pulses with roll-off factors $\rho \in \{0, 1\}$, and bandwidths $W \in \{10, 20, 30\}$ MHz. However, the actual combination of bandwidths and roll-off factors used can be unknown, which justifies why an overcomplete set of bases becomes handy. Transmitted signals with bandwidth $W = 10$ MHz are searched over a grid of 16 evenly spaced center frequencies f_c in B. Likewise, for $W = 20$ and 30 MHz, 15 and 14 center frequencies are considered, respectively. This amounts to $2 \times (16 + 15 + 14) = 90$ possible combinations for ρ, W, and f_c, thus $N_b = 90$ bases are adopted.

Each CR computes periodogram samples $\hat{\phi}_{rn}(\tau)$ at $N = 64$ frequencies with $SNR = -5$ dB, and averages them across $T = 100$ time-slots to form $\varphi_{rn},\ n = 1, \ldots, 64$ as in (3). These network-wide observations at $T = 100$ are collected in $\boldsymbol{\varphi}$, and following steps S1-S4 at the end of Section IV, the spline-based estimator (12), and thus the PSD map $\hat{\Phi}(\mathbf{x}, f)$ is formed. This map is summed across frequencies, and the result is shown in Fig. 3 (left) which depicts the positions of transmitting CRs, as well as the radially-decaying spectra of four of them (those not affected by the obstacle). It also identifies the effect of the wall by "flattening" the spectrum emitted by the fifth source at the top-left corner. Inspection of the estimate $\hat{\Phi}(\mathbf{x}, f)$ across frequency confirms that group-Lasso succeeds in selecting the candidate bases. Fig. 4 (left) shows points representing $\|\hat{\boldsymbol{\zeta}}_\nu\|_2$, $\nu = 1, \ldots, N_b$, where $\hat{\boldsymbol{\zeta}}_\nu$ is the sub-vector in the solution of the

group-Lasso estimator (13) associated with $g_\nu(\mathbf{x})$ and $b_\nu(f)$. They peak at indexes $\nu = 1,\ 28,\ 46,\ 51$, and 70 (circled in red), which correspond to the "ground-truth" model, since bases $b_1,\ b_{28},\ b_{46},\ b_{51}$, and b_{70} match the spectra of the transmitted signals. Even though approximately 75% of the variables drop out of the model, some spurious coefficients are retained and their norms are markedly smaller than those of the "ground-truth" bases. This is expected because based on finite samples there is no guarantee that group-Lasso will recover the exact support, in general. Nevertheless, the effectiveness of group-Lasso in revealing the transmitted bases is apparent when compared to other regularization alternatives. Fig. 4 (right) depicts the counterpart of Fig. 4 (left) when using a sparsity-agnostic ridge regression scheme instead of (13). In this case, no basis selection takes place, and the spurious factors are magnified up to a level comparable to three of the "true" basis function $b_\nu(f)$. To the best of our knowledge, no other basis selection methods in the literature are applicable to the nonparametric model (1) considered here. In particular, COSSO in [16] is not applicable since it does not provide a basis selection method and relies on orthogonality assumptions.

In summary, this test case demonstrates that the spline-based estimator can reveal which frequency bands are (un)occupied at each point in space, thus allowing for spatial reuse of the idle bands. For instance, transmitter TX_5 at the top-right corner is associated with the basis function $b_{46}(f)$, the only one of the transmitted five that occupies the $230 - 260$ MHz sub-band. Therefore, this sub-band can be reused at locations $\mathbf{x}$ away from the transmission range of TX_5, which is revealed in Fig. 3 (left).

The group-Lasso estimator in S1 was obtained via the GLasso algorithm developed in Section V (cf. Algorithm 2). The GLasso output at iteration $k = 1,000$ is compared to previous iterates $\boldsymbol{\zeta}(k)$ in Fig. 3 (right), which demonstrates the monotone decay of their difference, and thus corroborates convergence to a limit point. Then, it is verified numerically that $\boldsymbol{\zeta}(1000)$ satisfies the necessary and sufficient conditions for optimality of (17), as given in [29]. These two tests together provide numerical confirmation of Proposition 3 on the convergence of GLasso, and the optimality of the limit point.

B. Tuning parameters via cross-validation

Results in Figs. 3 (left) and 4 depend on the judicious selection of parameters λ and μ in (12). Parameter λ affects smoothness, which translates to congruence among PSD samples, allowing the CRs to recover the radial aspect of the transmit-power. Parameter μ controls the sparsity in the solution, which dictates the number of bases, and thus transmission schemes that the estimator considers active.

To select λ and μ jointly so that both smoothness and sparsity are properly accounted for, one could consider a two-dimensional grid of candidate pairs, and minimize the CV error over this grid. However, this is computationally demanding, especially because the nondifferentiable cost in (13) renders the shortcuts in

Appendix D not applicable (see also Remark 3). A three-step alternative is followed here. First, estimator (12) is obtained using an arbitrarily small value of $\lambda = 1 \times 10^{-6}$, and selecting $\mu = 0.1\mu_{\max}$, where $\mu_{\max}$ is given in subsection IV-A. In the second step, only the surviving bases are kept, and the sparsifying penalty is no longer considered, thus reducing the estimator to that of Section III. If the reduced matrix $\mathbf{B}$, built from the surviving bases, is full rank (otherwise repeat the first step with a larger value of μ), the procedure in Appendix D is followed to adjust the value of λ via leave-one-out CV. The result of this step is illustrated in Fig. 5 (left), where the minimizer $\lambda_{CV} = 7.9433 \times 10^{-6}$ of the OCV cost is selected. The final step consists of reconsidering the sparsity enforcing penalty in (12), and selecting μ using 5-fold CV. The minimizer of the CV error $\mu_{CV} = 0.0078\mu_{\max}$ corresponding to this step is depicted in Fig. 5 (right). Using the λ_{CV} and μ_{CV} so obtained, the PSD map plotted in Fig. 3 (left) was constructed. The rationale behind this approach is that it corresponds to a single step of a coordinate descent algorithm for minimizing the CV error $CV(\lambda, \mu)$. Function $CV(\lambda, \mu)$ is typically unimodal, with much higher sensitivity on μ than on λ, a geometric feature leading the first coordinate descent update to be close to the optimum.

The importance of an appropriate μ value becomes evident when inspecting how many bases are retained by the estimator as μ decreases from $\mu_{\max}$ to $1 \times 10^{-4}\mu_{\max}$. The N_b lines in Fig. 6 (left) link points representing $\|\hat{\boldsymbol{\zeta}}_\nu(\mu)\|_2$, as μ takes on 20 evenly spaced values on a logarithmic scale, comprising the so-termed group-Lasso *path of solutions*. When $\mu = \mu_{\max}$ is selected, by definition the estimator forces all $\hat{\boldsymbol{\zeta}}_\nu$ to zero, thus discarding all bases. As μ tends to zero all bases become relevant and eventually enter the model, which confirms the premise that LS estimators suffer from overfitting when the underlying model is overcomplete. The cross-validated value μ_{CV} is indicated with a dashed vertical line that crosses the path of solutions at the values of $\|\hat{\boldsymbol{\zeta}}_\nu\|_2$. At this point, five sub-vectors corresponding to the factors $\nu = 1,\ 28,\ 46,\ 51$, and 70 are considerably far away from zero hence showing strong effects, in par with the results depicted in Fig. 4 (left). Certainly interesting would be to corroborate the effectiveness of the proposed PSD map estimator on real data comprising spatially distributed RF measurements. Upon availability of such dataset, this direction will be pursued and reported elsewhere.

C. Example with real data

The goal of this section is to demonstrate that the GLasso algorithm in Section V can be useful for applications other than the spline-based BEM for spectrum cartography dealt with in Sections III and IV. This demonstration will rely on the birthweight dataset from [14], considered also by the seminal group-Lasso work of [29]. The objective is to predict the human birthweight from $p = 8$ factors including the mother's `age`, `weight`, `race`, `smoke` habits, number of previous `premature` labors, history of

`hypertension`, uterine `irritability`, and number of physician `visits` during the first trimester of pregnancy. Third-order polynomials were considered to model nonlinear effects of the `age` and `weight` on the response, augmenting the model size to $p = 12$ by grouping the polynomial coefficients in two subsets of three variables.

GLasso is run under this setup, over the set of $N = 189$ samples, with μ selected via 7-fold CV. Fig. 6 (right) depicts the evolution of the factors' strength measured by $\|\zeta_\nu\|_2^2$, which – as expected – converge to the same prediction model as in [29]. Additionally, GLasso is capable of determining that the eighth factor (`visits`) is not significant even from the first iterations, allowing for early model selection.

VII. Concluding Summary

A basis expansion approach was introduced in this paper to estimate a multi-dimensional field, whose dependence on a subset of its variables is modeled through preselected (and generally *overlapping*) basis functions weighted by unknown coefficient-functions of the remaining variables. The unknown coefficient functions can be estimated from the field's noisy samples, by solving a variational LS problem which admits infinite solutions. Without extra constraints, the estimated field interpolates perfectly the data samples, at the price of severely overfitting the true field elsewhere. The first contribution was to regularize this variational LS cost by a smoothing term, which can afford a unique finite-parameter spline-based solution. The latter is expressed in terms of radial kernels and polynomials whose parameters were estimated in closed form. A recursive PSD tracker was also developed for slowly time-varying spectra.

The second main contribution pertains to a robust variant of the function estimator, when an overcomplete set of bases is adopted to effectively accommodate model uncertainties. The novel estimator here minimizes the variational LS cost regularized by a term that performs *basis selection*, and thus yields a parsimonious description of the field by retaining those few members of the basis that "better" explain the data. This attribute is achieved because the added penalty induces a group (G)Lasso estimator on the parameters of the kernels and polynomials. Even though the number of unknowns increases with overcomplete bases, most coefficients are zero, meaning that the complexity remains at an affordable level using the sparsity-promoting GLasso. Notwithstanding, (group-) Lasso here is introduced to effect (group-) sparsity in the space of smooth functions.

The third contribution is a provably convergent GLasso estimator developed based on AD-MoM iterations. It entails parallel *closed-form* updates, which involve simple vector soft-thresholding operations per factor. Its fully-distributed counterpart was also developed for use by a network of wireless sensors, or, multiple processors to balance the load of a computational cluster. It is worth stressing that both GLassso and

DGLasso are standalone tools for sparse linear regression, applicable to a gamut of problems that go beyond the field estimation context of this paper.

The fourth contribution is in the context of wireless CR network sensing (our overarching practical motivation), where the overcomplete estimated field enables cartographing the space-frequency distribution of power generated by RF sources whose transmit-PSDs are shaped by, e.g., raised-cosine pulses with possibly different roll-off factors, center frequencies, and bandwidths. Using periodogram samples collected by spatially distributed CRs, the sparsity-aware spline-based estimator yields an atlas of *PSD maps* (one map per frequency). As corroborated by simulations, the atlas enables localizing the sources and discerning their transmission parameters, even in the presence of frequency-selective Rayleigh fading and pronounced shadowing effects due to e.g., an obstructing wall. Simulated tests also illustrated the convergence of Glasso, and confirmed that the sparsity-promoting regularization is effective in selecting those basis functions that strongly influence the field, when the tuning parameters are cross-validated properly.

Given the existing connections between splines and classical estimators for both random and deterministic field models, the spline-based methods developed in this paper provide a unifying framework suitable for both paradigms. The model and the resultant (parsimonious) estimates can thus be used in more general statistical inference and localization problems, whenever the data admit a basis expansion over a proper subset of its dimensions. Furthermore, results in this paper extend to kernels other than radial basis functions, whenever the smoothing penalty is replaced by a norm induced from an RKHS. Also of interest is to quantify the number of data required to attain a prescribed approximation error, in light of the existing connections between spline-based reconstruction and Shannon's sampling theory [25].

ACNOWLEDGMENTS

The authors would like to thank Prof. Hui Zou (School of Statistics, University of Minnesota) for his feedback which improved the exposition of ideas in this paper.

APPENDIX

A. Proof of Proposition 1: Rewrite (4) as

$$\min_{\{g_\nu\in\mathcal{S}\}_{\nu=2}^{N_b}}\left[\min_{g_1\in\mathcal{S}}\sum_{r=1}^{N_r}\sum_{n=1}^{N}\left(\varphi_{rn}^{(-1)}-g_1(\mathbf{x}_r)b_1(f_n)\right)^2+\lambda\int_{\mathbb{R}^2}||\nabla^2 g_1(\mathbf{x})||_F^2 d\mathbf{x}\right]+\lambda\sum_{\nu=2}^{N_b}\int_{\mathbb{R}^2}||\nabla^2 g_\nu(\mathbf{x})||_F^2 d\mathbf{x} \tag{29}$$

with $\varphi_{rn}^{(-1)} := \varphi_{rn}-\sum_{\nu=2}^{N_b} g_\nu(\mathbf{x}_r)b_\nu(f_n)$. Focusing on the inner minimization w.r.t. g_1, fix the set of functions $\{g_\nu\}_{\nu=2}^{N_b}$, and note that only the first two terms are relevant (those within the square brackets). It follows from [10, Theorem 4 bis] that $\hat{g}_1$ takes the form in (5), with coefficients $\boldsymbol{\beta}_1, \boldsymbol{\alpha}_{11}$ and α_{10} that

depend on $G^{(-1)} := \{g_\nu(x_r),\ r = 1, \ldots, N_r,\ \nu = 2, \ldots, N_b\}$ through $\varphi_{rn}^{(-1)}$. The next step is to minimize (29) w.r.t. g_2 but with $\{g_\nu\}_{\nu=3}^{N_b}$ fixed, which amounts to

$$\min_{g_2 \in \mathcal{S}} \sum_{r=1}^{N_r} \sum_{n=1}^{N} \left(\varphi_{rn}^{(-2)} - \hat{g}_1(\mathbf{x}_r) b_1(f_n) - g_2(\mathbf{x}_r) b_2(f_n)\right)^2 + \lambda \int_{\mathbb{R}^2} ||\nabla^2 \hat{g}_1(\mathbf{x})||_F^2 d\mathbf{x} + \lambda \int_{\mathbb{R}^2} ||\nabla^2 g_2(\mathbf{x})||_F^2 d\mathbf{x} \tag{30}$$

where $\varphi_{rn}^{(-2)} := \varphi_{rn} - \sum_{\nu=3}^{N_b} g_\nu(\mathbf{x}_r) b_\nu(f_n)$. In the first two summands of the cost in (30), $\hat{g}_1$ depends on g_2 via $G^{(-1)}$. Because $G^{(-1)}$ only involves evaluating g_2 on $\mathcal{X}$, [10, Theorem 4 bis] can be applied again, and the optimal solution $\hat{g}_2$ takes the from (5). The same argument carries over to subsequent minimization steps for $\nu = 3, \ldots, N_b$, establishing that all $\{\hat{g}_\nu(\mathbf{x})\}$ are thin-plate splines as in (5).

B. Proof of (6)-(8): Upon substituting (5) into (4), it will shown next that the optimal coefficients $\{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}\}$ specifying $\{\hat{g}_\nu(\mathbf{x})\}_{\nu=1}^{N_b}$ are obtained as solutions to the following constrained, regularized LS problem

$$\begin{aligned} \min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \frac{1}{N_r N} \|\boldsymbol{\varphi} - (\mathbf{B} \otimes \mathbf{K})\boldsymbol{\beta} - (\mathbf{B} \otimes \mathbf{T})\boldsymbol{\alpha}\|_2^2 + \lambda \boldsymbol{\beta}'(\mathbf{I}_{N_b} \otimes \mathbf{K})\boldsymbol{\beta} \\ \text{s. t.} \quad (\mathbf{I}_{N_b} \otimes \mathbf{T}')\boldsymbol{\beta} = \mathbf{0}_{3N_b}. \end{aligned} \tag{31}$$

Observe first that the constraints $\boldsymbol{\beta}_\nu \in \mathcal{B}$ in Proposition 1 can be expressed as $\mathbf{T}'\boldsymbol{\beta}_\nu = \mathbf{0}_3$ for each $\nu = 1, \ldots, N_b$, or jointly as $(\mathbf{I}_{N_b} \otimes \mathbf{T}')\boldsymbol{\beta} = \mathbf{0}_{3N_b}$. For the optimization objective in (31), note from (5) that $\hat{g}_\nu(\mathbf{x}_r) = \mathbf{k}_r'\boldsymbol{\beta}_\nu + \mathbf{t}_r'\boldsymbol{\alpha}_\nu$, where $\mathbf{k}_r'$ and $\mathbf{t}_r'$ are the rth rows of $\mathbf{K}$ and $\mathbf{T}$, respectively. The first term in the cost of (4) can be expressed (up to a factor $(N_r N)^{-1}$) as

$$\begin{aligned} \sum_{n=1}^{N} \sum_{r=1}^{N_r} \left(\varphi_{rn} - \sum_{\nu=1}^{N_b} b_\nu(f_n)[\mathbf{k}_r'\boldsymbol{\beta}_\nu + \mathbf{t}_r'\boldsymbol{\alpha}_\nu]\right)^2 &= \sum_{n=1}^{N} \sum_{r=1}^{N_r} \left(\varphi_{rn} - (\mathbf{b}_n \otimes \mathbf{k}_r)'\boldsymbol{\beta} - (\mathbf{b}_n \otimes \mathbf{t}_r)'\boldsymbol{\alpha}\right)^2 \\ &= \sum_{n=1}^{N} \left\|\boldsymbol{\varphi}_n - (\mathbf{b}_n' \otimes \mathbf{K})\boldsymbol{\beta} - (\mathbf{b}_n' \otimes \mathbf{T})\boldsymbol{\alpha}\right\|_2^2 \\ &= \|\boldsymbol{\varphi} - (\mathbf{B} \otimes \mathbf{K})\boldsymbol{\beta} - (\mathbf{B} \otimes \mathbf{T})\boldsymbol{\alpha}\|_2^2. \end{aligned}$$

Consider next the penalty term in the cost of (4). Substituting into (5), it follows that $\int_{\mathbb{R}^2} ||\nabla^2 \hat{g}_\nu(\mathbf{x})||_F^2 d\mathbf{x} = \boldsymbol{\beta}_\nu' \mathbf{K} \boldsymbol{\beta}_\nu$ [26, p. 33]. It thus holds that

$$\lambda \sum_{\nu=1}^{N_b} \int_{\mathbb{R}^2} ||\nabla^2 \hat{g}_\nu(\mathbf{x})||_F^2 d\mathbf{x} = \lambda \sum_{\nu=1}^{N_b} \boldsymbol{\beta}_\nu' \mathbf{K} \boldsymbol{\beta}_\nu = \lambda \boldsymbol{\beta}'(\mathbf{I}_{N_b} \otimes \mathbf{K})\boldsymbol{\beta}$$

from which (31) follows readily.

Now that the equivalence between (4) and (31) has been established, the latter must be solved for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Even though $\mathbf{K}$ (hence $\mathbf{I}_{N_b} \otimes \mathbf{K}$) is not positive definite, it is still possible to show that $\boldsymbol{\beta}'(\mathbf{I}_{N_b} \otimes \mathbf{K})\boldsymbol{\beta} > 0$ for any $\boldsymbol{\beta}$ such that $(\mathbf{I}_{N_b} \otimes \mathbf{T}')\boldsymbol{\beta} = \mathbf{0}_{3N_b}$ [10, p. 85], implying that (31) is convex. Proceeding along the lines of [26, p. 33], note first that the constraint $(\mathbf{I}_{N_b} \otimes \mathbf{T}')\boldsymbol{\beta} = \mathbf{0}_{3N_b}$ implies the existence of a vector

$\boldsymbol{\gamma} \in \mathbb{R}^{N_b(N_r-3)}$ satisfying (8). After this change of variables, (31) is transformed into an unconstrained quadratic program, which can be solved in closed form for $\{\boldsymbol{\alpha}, \boldsymbol{\gamma}\}$. Hence, setting both gradients w.r.t. $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}\}$ to zero yields (6) and (7).

C. Proof of Proposition 2: After substituting (15) into (12), one finds the optimal $\{\boldsymbol{\alpha}, \boldsymbol{\beta}\}$ specifying $\{\hat{g}_\nu(\mathbf{x})\}_{\nu=1}^{N_b}$ in (15), as solutions to the following constrained, regularized LS problem

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \frac{1}{N_r N} \|\boldsymbol{\varphi} - (\mathbf{B}\otimes\mathbf{K})\boldsymbol{\beta} - (\mathbf{B}\otimes\mathbf{T})\boldsymbol{\alpha}\|_2^2 + \lambda\boldsymbol{\beta}'(\mathbf{I}_{N_b}\otimes\mathbf{K})\boldsymbol{\beta} + \mu\sum_{\nu=1}^{N_b}\|\mathbf{K}\boldsymbol{\beta}_\nu + \mathbf{T}\boldsymbol{\alpha}_\nu\|_2$$
$$\text{s. t.} \quad (\mathbf{I}_{N_b}\otimes\mathbf{T}')\boldsymbol{\beta} = \mathbf{0}_{3N_b}. \tag{32}$$

With reference to (32), consider grouping and reordering the variables $\{\boldsymbol{\alpha}, \boldsymbol{\beta}\}$ in the vector $\mathbf{u} := [\mathbf{u}_1', \ldots, \mathbf{u}_{N_b}']'$, where $\mathbf{u}_\nu := [\boldsymbol{\beta}_\nu' \;\; \boldsymbol{\alpha}_\nu']'$. As argued in Section III-A, the constraints $\mathbf{T}'\boldsymbol{\beta}_\nu = \mathbf{0}$ can be eliminated through the change of variables $\mathbf{u}_\nu = \text{bdiag}(\mathbf{Q}_2, \mathbf{I}_3)\mathbf{v}_\nu$ for $\nu = 1, \ldots, N_b$; or compactly as $\mathbf{u} = (\mathbf{I}_{N_b}\otimes\text{bdiag}(\mathbf{Q}_2, \mathbf{I}_3))\mathbf{v}$. The next step is to express the three summands in the cost of (32) in terms of the new vector optimization variable $\mathbf{v}$. Noting that $\mathbf{k}_r'\boldsymbol{\beta}_\nu + \mathbf{t}_r'\boldsymbol{\alpha}_\nu = [\mathbf{k}_r' \;\; \mathbf{t}_r']\mathbf{u}_\nu$, and mimicking the steps in Appendix A, the first summand is

$$\begin{aligned}\frac{1}{N_r N} \|\boldsymbol{\varphi} - (\mathbf{B}\otimes\mathbf{K})\boldsymbol{\beta} - (\mathbf{B}\otimes\mathbf{T})\boldsymbol{\alpha}\|_2^2 &= \frac{1}{N_r N} \|\boldsymbol{\varphi} - (\mathbf{B}\otimes[\mathbf{K} \;\; \mathbf{T}])\mathbf{u}\|_2^2 \\ &= \frac{1}{N_r N} \|\boldsymbol{\varphi} - (\mathbf{B}\otimes[\mathbf{K}\mathbf{Q}_2 \;\; \mathbf{T}])\mathbf{v}\|_2^2. \end{aligned} \tag{33}$$

The second summand due to the thin-plate penalty can be expressed as

$$\begin{aligned}\lambda\sum_{\nu=1}^{N_b}\boldsymbol{\beta}_\nu'\mathbf{K}\boldsymbol{\beta}_\nu = \lambda\sum_{\nu=1}^{N_b}\mathbf{u}_\nu'\text{bdiag}(\mathbf{K}, \mathbf{0})\mathbf{u}_\nu &= \lambda\sum_{\nu=1}^{N_b}\mathbf{v}_\nu'\text{bdiag}(\mathbf{Q}_2'\mathbf{K}\mathbf{Q}_2, \mathbf{0})\mathbf{v}_\nu \\ &= \lambda\mathbf{v}'(\mathbf{I}_{N_b}\otimes\text{bdiag}(\mathbf{Q}_2'\mathbf{K}\mathbf{Q}_2, \mathbf{0}))\mathbf{v} \end{aligned} \tag{34}$$

while the last term is $\mu\sum_{\nu=1}^{N_b}\|\mathbf{K}\boldsymbol{\beta}_\nu + \mathbf{T}\boldsymbol{\alpha}_\nu\|_2 = \mu\sum_{\nu=1}^{N_b}\|[\mathbf{K} \;\; \mathbf{T}]\mathbf{u}_\nu\|_2 = \mu\sum_{\nu=1}^{N_b}\|[\mathbf{K}\mathbf{Q}_2 \;\; \mathbf{T}]\mathbf{v}_\nu\|_2$. Combining (33) with (34) by completing the squares, problem (32) is equivalent to

$$\min_{\mathbf{v}} \frac{1}{N_r N} \left\| \begin{bmatrix} \boldsymbol{\varphi} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{B}\otimes[\mathbf{K}\mathbf{Q}_2 \;\; \mathbf{T}] \\ \mathbf{I}_{N_b}\otimes\text{bdiag}((N_r N\lambda\mathbf{Q}_2'\mathbf{K}\mathbf{Q}_2)^{1/2}, \mathbf{0}) \end{bmatrix} \mathbf{v} \right\|_2^2 + \mu\sum_{\nu=1}^{N_b}\|[\mathbf{K}\mathbf{Q}_2 \;\; \mathbf{T}]\mathbf{v}_\nu\|_2 \tag{35}$$

and becomes (13) under the identities (14), and after the change of variables $\boldsymbol{\zeta} := [\boldsymbol{\zeta}_1', \ldots, \boldsymbol{\zeta}_{N_b}']' = (\mathbf{I}_{N_b}\otimes[\mathbf{K}\mathbf{Q}_2 \, \mathbf{T}])\mathbf{v}$. By definition of $\mathbf{u}$, $\mathbf{v}$, and $\boldsymbol{\zeta}$, the original variables can be recovered through the transformation in (16).

D. Selection of the smoothing parameter in (4): The method to be developed builds on the so-termed leave-one-out CV, which proceeds as follows; see e.g., [26, Ch. 4]. Consider removing a single data point φ_{rn}

from the collection of N_rN measurements available to the sensing radios. For a given λ, let $\hat{\Phi}_{\lambda}^{(-rn)}(\mathbf{x}, f)$ denote the *leave-one-out* estimated PSD map, obtained by solving (4) following steps S1-S3 in Section III-A, using the N_rN-1 remaining data points. The aforementioned estimation procedure is repeated N_rN times by leaving out each of the data points φ_{rn}, $r=1,\ldots,N_r$ and $n=1,\ldots,N$, one at a time. The leave-one-out or ordinary CV (OCV) [13, p. 242], [26, p. 47], for the problem at hand is given by

$$\text{OCV}(\lambda)=\frac{1}{N_rN}\sum_{r=1}^{N_r}\sum_{n=1}^{N}\left(\varphi_{rn}-\hat{\Phi}_{\lambda}^{(-rn)}(\mathbf{x}_r,f_n)\right)^2 \tag{36}$$

while the optimum λ is selected as the minimizer of $\text{OCV}(\lambda)$, over a grid of values $\lambda\in[0,\lambda_{\max}]$. Function (36) constitutes an average of the squared prediction errors over all data points; hence, its minimization offers a natural criterion. The method is quite computationally demanding though, since the system of linear equations (6)-(8) has to be solved N_rN times for each value of λ on the grid. Fortunately, this computational burden can be significantly reduced for the spline-based PSD map estimator considered here.

Recall the vector $\boldsymbol{\varphi}$ collecting all data points measured at locations $\mathcal{X}$ and frequencies $\mathcal{F}$. Define next a similar vector $\hat{\boldsymbol{\varphi}}$ containing the respective predicted values at the given locations and frequencies, which is obtained after solving (4) with all the data in $\boldsymbol{\varphi}$ and a given value of λ. The following lemma establishes that the PSD map estimator is a *linear smoother*, which means that the predicted values are linearly related to the measurements, i.e., $\hat{\boldsymbol{\varphi}}=\mathbf{S}_{\lambda}\boldsymbol{\varphi}$ for a λ-dependent matrix $\mathbf{S}_{\lambda}$ to be determined. Common examples of linear smoothers are ridge regressors and smoothing splines; further details are in [13, p. 153]. For linear smoothers, by virtue of the leave-one-out lemma [26, p. 50] it is possible to rewrite (36) as

$$\text{OCV}(\lambda)=\frac{1}{N_rN}\sum_{r=1}^{N_r}\sum_{n=1}^{N}\left(\frac{\varphi_{rn}-\hat{\Phi}_{\lambda}(\mathbf{x}_r,f_n)}{1-[\mathbf{S}_{\lambda}]_{ii}}\right)^2 \tag{37}$$

where $\hat{\Phi}_{\lambda}(\mathbf{x},f)$ stands for the estimated PSD map when all data in $\boldsymbol{\varphi}$ are utilized in (4). The beauty of the leave-one-out lemma stems from the fact that given λ and the main diagonal of matrix $\mathbf{S}_{\lambda}$, the right-hand side of (37) indicates that fitting a single model (rather than N_rN of them) suffices to evaluate $\text{OCV}(\lambda)$. The promised lemma stated next specifies the value of $\mathbf{S}_{\lambda}$ necessary to evaluate (37).

Lemma 3: *The PSD map estimator in* (4) *is a linear smoother, with smoothing matrix given by*

$$\begin{aligned}\mathbf{S}_{\lambda}=&(\mathbf{B}\otimes\{\mathbf{K}\mathbf{Q}_2-\mathbf{T}\mathbf{R}^{-1}\mathbf{Q}_1'\mathbf{K}\mathbf{Q}_2\})[(\mathbf{B}'\mathbf{B}\otimes\mathbf{Q}_2'\mathbf{K}\mathbf{Q}_2)+N_rN\lambda\mathbf{I}]^{-1}(\mathbf{B}'\otimes\mathbf{Q}_2')\\&+(\mathbf{B}\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}_1^{-1}\otimes\mathbf{T}\mathbf{R}^{-1}\mathbf{Q}_1').\end{aligned} \tag{38}$$

Proof: Reproduce the structure of $\boldsymbol{\varphi}$ in Section III-A to form the supervector $\hat{\boldsymbol{\varphi}} := [\hat{\boldsymbol{\varphi}}_1', \ldots, \hat{\boldsymbol{\varphi}}_N']' \in \mathbb{R}^{N_r N}$, by stacking each vector $\hat{\boldsymbol{\varphi}}_n := [\hat{\Phi}_\lambda(\mathbf{x}_1, f_n), \ldots, \hat{\Phi}_\lambda(\mathbf{x}_{N_r}, f_n)]'$ corresponding to the spatial PSD predictions at frequency $f_n \in \mathcal{F}$. From (5), it follows that $\hat{\Phi}_\lambda(\mathbf{x}_r, f_n) = (\mathbf{b}_n \otimes \mathbf{k}_r)'\hat{\boldsymbol{\beta}} - (\mathbf{b}_n \otimes \mathbf{t}_r)'\hat{\boldsymbol{\alpha}}$, where $\mathbf{b}_n'$, $\mathbf{k}_r'$ and $\mathbf{t}_r'$ are the nth and rth rows of $\mathbf{B}$, $\mathbf{K}$ and $\mathbf{T}$, respectively. By stacking the PSD map estimates, it follows that $\hat{\boldsymbol{\varphi}}_n = (\mathbf{b}_n' \otimes \mathbf{K})\hat{\boldsymbol{\beta}} - (\mathbf{b}_n' \otimes \mathbf{T})\hat{\boldsymbol{\alpha}}$, which readily yields

$$\hat{\boldsymbol{\varphi}} = (\mathbf{B} \otimes \mathbf{K})\hat{\boldsymbol{\beta}} - (\mathbf{B} \otimes \mathbf{T})\hat{\boldsymbol{\alpha}}. \tag{39}$$

Because the estimates $\{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}\}$ are linearly related to the measurements $\boldsymbol{\varphi}$ [cf. (6)-(8)], so is $\hat{\boldsymbol{\varphi}}$ as per (39), establishing that the PSD map estimator in (4) is indeed a linear smoother. Next, solve explicitly for $\{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}\}$ in (6)-(8) and substitute the results in (39), to unveil the structure of the smoothing matrix $\mathbf{S}_\lambda$ such that $\hat{\boldsymbol{\varphi}} = \mathbf{S}_\lambda \boldsymbol{\varphi}$. Simple algebraic manipulations lead to the expression (38). ■

The effectiveness of the leave-one-out CV approach is corroborated via simulations in Section VI.

E. Proof of (24)-(27)*:* Recall the augmented Lagrangian function in (23), and let $\boldsymbol{\zeta} := \{\boldsymbol{\zeta}_r\}_{r \in \mathcal{R}}$ for notational brevity. When used to solve (22), the three steps in the AD-MoM are given by:

[S1] Local estimate updates:

$$\boldsymbol{\zeta}(k+1) = \arg \min_{\boldsymbol{\zeta}} \mathcal{L}_c \left[\boldsymbol{\zeta}, \boldsymbol{\gamma}(k), \boldsymbol{v}(k)\right]. \tag{40}$$

[S2] Auxiliary variable updates:

$$\boldsymbol{\gamma}(k+1) = \arg \min_{\boldsymbol{\gamma}} \mathcal{L}_c \left[\boldsymbol{\zeta}(k+1), \boldsymbol{\gamma}, \boldsymbol{v}(k)\right]. \tag{41}$$

[S3] Multiplier updates:

$$\mathbf{v}_r(k+1) = \mathbf{v}_r(k) + c[\boldsymbol{\zeta}_r(k+1) - \boldsymbol{\gamma}_r(k+1)] \tag{42}$$

$$\breve{\mathbf{v}}_r^{r'}(k+1) = \breve{\mathbf{v}}_r^{r'}(k) + c[\boldsymbol{\zeta}_r(k+1) - \boldsymbol{\gamma}_r^{r'}(k+1)] \tag{43}$$

$$\bar{\mathbf{v}}_r^{r'}(k+1) = \bar{\mathbf{v}}_r^{r'}(k) + c[\boldsymbol{\zeta}_{r'}(k+1) - \boldsymbol{\gamma}_r^{r'}(k+1)]. \tag{44}$$

Focusing first on [S2], observe that (23) is separable across the collection of variables $\{\boldsymbol{\gamma}_j\}$ and $\{\boldsymbol{\gamma}_r^{r'}\}$ that comprise $\boldsymbol{\gamma}$. The minimization w.r.t. the latter group reduces to

$$\begin{aligned}
\boldsymbol{\gamma}_r^{r'}(k+1) &= \arg \min_{\boldsymbol{\gamma}_r^{r'}} c\|\boldsymbol{\gamma}_r^{r'}\|^2 - c\Big(\boldsymbol{\zeta}_r(k+1) + \boldsymbol{\zeta}_{r'}(k+1)\Big)\boldsymbol{\gamma}_r^{r'} - \Big(\bar{\mathbf{v}}_r^{r'}(k) + \breve{\mathbf{v}}_r^{r'}(k)\Big)\boldsymbol{\gamma}_r^{r'} \\
&= \frac{1}{2}\Big(\boldsymbol{\zeta}_r(k+1) + \boldsymbol{\zeta}_{r'}(k+1)\Big) + \frac{1}{2c}\Big(\bar{\mathbf{v}}_r^{r'}(k) + \breve{\mathbf{v}}_r^{r'}(k)\Big) \\
&= \frac{1}{2}\Big(\boldsymbol{\zeta}_r(k+1) + \boldsymbol{\zeta}_{r'}(k+1)\Big).
\end{aligned} \tag{45}$$

The result in (45) assumes that $\bar{\mathbf{v}}_r^{r'}(k)+\breve{\mathbf{v}}_r^{r'}(k)=\mathbf{0},\ \forall k$. A simple inductive argument over (43), (44) and (45) shows that this is indeed true if the multipliers are initialized such that $\bar{\mathbf{v}}_r^{r'}(0)+\breve{\mathbf{v}}_r^{r'}(0)=\mathbf{0}$.

The remaining minimization in (41) with respect to $\{\boldsymbol{\gamma}_r\}$ decouples into N_r quadratic sub-problems [cf. (23)], that is

$$\boldsymbol{\gamma}_r(k+1)=\arg\min_{\boldsymbol{\gamma}_r}\frac{1}{2}\|\mathbf{y}_r-\mathbf{X}_r\boldsymbol{\gamma}_r\|_2^2-\mathbf{v}_r'(k)\boldsymbol{\gamma}_r+\frac{c}{2}\|\boldsymbol{\zeta}_r(k+1)-\boldsymbol{\gamma}_r\|_2^2$$

which admit the closed-form solutions in (27).

In order to obtain the update (24) for the prices $\mathbf{p}_r$, consider their definition together with (43), (44) and (45) to obtain

$$\begin{aligned}\mathbf{p}_r(k+1)&=\sum_{r'\in\mathcal{N}_r}\Big(\breve{\mathbf{v}}_r^{r'}(k+1)+\bar{\mathbf{v}}_{r'}^{r}(k+1)\Big)\\&=\sum_{r'\in\mathcal{N}_r}\Big(\breve{\mathbf{v}}_r^{r'}(k)+\bar{\mathbf{v}}_{r'}^{r}(k)\Big)+\sum_{r'\in\mathcal{N}_r}c\Big(2\boldsymbol{\zeta}_r(k+1)-\boldsymbol{\gamma}_r^{r'}(k)-\boldsymbol{\gamma}_{r'}^{r}(k)\Big)\\&=\mathbf{p}_r(k)+c\sum_{r'\in\mathcal{N}_r}\left(\boldsymbol{\zeta}_r(k+1)-\boldsymbol{\zeta}_{r'}(k+1)\right)\end{aligned}$$

which coincides with (24).

Towards obtaining the updates for the local variables in $\boldsymbol{\zeta}$, the optimization (40) in [S1] can be also split into N_r sub-problems, namely

$$\begin{aligned}\boldsymbol{\zeta}_r(k+1)=\arg\min_{\boldsymbol{\zeta}_r}&\left\{\frac{\mu}{N_r}\sum_{\nu=1}^{N_b}\|\boldsymbol{\zeta}_{r\nu}\|_2+\mathbf{v}_r'(k)\boldsymbol{\zeta}_r+\frac{c}{2}\|\boldsymbol{\zeta}_r-\boldsymbol{\gamma}_r(k)\|_2^2+\sum_{r'\in\mathcal{N}_r}\Big[\breve{\mathbf{v}}_r^{r'}(k)+\bar{\mathbf{v}}_{r'}^{r}(k)\Big]'\boldsymbol{\zeta}_r\right.\\&\left.+\frac{c}{2}\sum_{r'\in\mathcal{N}_r}\Big[\|\boldsymbol{\zeta}_r-\boldsymbol{\gamma}_r^{r'}(k)\|_2^2+\|\boldsymbol{\zeta}_r-\boldsymbol{\gamma}_{r'}^{r}(k)\|_2^2\Big]\right\}\\=\arg\min_{\boldsymbol{\zeta}_r}&\left\{\frac{\mu}{N_r}\sum_{\nu=1}^{N_b}\|\boldsymbol{\zeta}_{r\nu}\|_2-\left(c\sum_{r'\in\mathcal{N}_r}\Big(\boldsymbol{\zeta}_r(k)+\boldsymbol{\zeta}_{r'}(k)\Big)+c\boldsymbol{\gamma}_r(k)-\mathbf{p}_r(k)-\mathbf{v}_r(k)\right)'\boldsymbol{\zeta}_r\right.\\&\left.+\frac{c}{2}(1+2|\mathcal{N}_r|)\|\boldsymbol{\zeta}_r\|_2^2\right\}.\end{aligned}$$

Upon dividing by $c(1+2|\mathcal{N}_r|)$ each subproblem becomes identical to problem (19), and thus by Proposition 2 takes the form of (26).

REFERENCES

[1] A. Alaya-Feki, S. B. Jemaa, B. Sayrac, P. Houze, and E. Moulines, "Informed spectrum usage in cognitive radio networks: Interference cartography," in *Proc. of 19th Intl. Symp. on Personal, Indoor and Mobile Radio Comms.*, Cannes, France, Aug./Sep. 2008, pp. 1–5.

[2] D. Angelosante and G. B. Giannakis, "Group Lasso for catching changes in piecewise-stationary autoregressive spectra," *IEEE Trans. on Signal Processing*, 2010 (submitted).

[3] S. Bakin, "Adaptive regression and model selection in data mining problems," Ph.D. dissertation, Australian National University, Canberra, 1999.

[4] J. A. Bazerque and G. B. Giannakis, "Distributed spectrum sensing for cognitive radio networks by exploiting sparsity," *IEEE Trans. on Signal Processing*, vol. 58, pp. 1847–1862, Mar. 2010.

[5] J. A. Bazerque, G. Mateos, and G. B. Giannakis, "Distributed Lasso for in-network linear regression," in *Proc. of Intl. Conf. on Acoustics, Speech and Signal Processing*, Dallas, TX, Mar. 2010.

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

[7] L. Breiman, "Better subset regression using the nonnegative garrote," *Technometrics*, vol. 37, pp. 373–384, Nov. 1995.

[8] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, 1998.

[9] H. Chui and A. Rangarajan, "A new algorithm for non-rigid point matching," in *Proc. of Conf. on Computer Vision and Pattern Recognition*, Hilton Head, SC, Jun. 2000, pp. 44–51.

[10] J. Duchon, *Splines Minimizing Rotation-Invariant Semi-norms in Sobolev Spaces*. Springer-Verlag, 1977.

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

[12] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2006.

[13] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. Springer, 2009.

[14] D. Hosmer and S. Lemeshow, *Applied Logistic Regression*. Wiley, NY, 1989.

[15] S.-J. Kim, E. Dall'Anese, and G. B. Giannakis, "Spectrum sensing for cognitive radios using kriged kalman filtering," in *Proc. of 3rd Intl. Workshop on Comp. Advances in Multi-Sensor Adapt. Proces.*, Aruba Island, Dec. 2009, pp. 392–395.

[16] Y. Lin and H. H. Zhang, "Component selection and smoothing in multivariate nonparametric regression," *Annals of Statistics*, vol. 34, pp. 2272–2297, May 2006.

[17] G. Matheron, "The intrinsic random functions and their application," *Adv. Appl. Prob.*, vol. 5, pp. 439–468, 1973.

[18] T. Minka, "Old and new matrix algebra useful for statistics," Dec. 2000. [Online]. Available: http://research.microsoft.com/en-us/um/people/minka/papers/matrix/

[19] S. M. Mishra, A. Sahai, and R. W. Brodersen, "Cooperative sensing among cognitive radios," in *Proc. of 42nd Intl. Conf. on Communications*, Istanbul, Turkey, Jun. 2006, pp. 1658–1663.

[20] 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. of 65th Vehicular Technology Conference*, Dublin, Ireland, Apr. 2007, pp. 2631–2635.

[21] F. Perrin, O. Bertrand, and J. Pernier, "Scalp current density mapping: Value and estimation from potential data," *IEEE Trans. on Biomedical Engr.*, vol. 34, no. 4, pp. 283–288, Apr. 1987.

[22] A. T. Puig, A. Wiesel, and A. O. Hero, "A multidimensional shrinkage-thresholding operator," in *Proc. of 15th Workshop on Statistical Signal Processing*, Cardiff, Wales, Aug./Sep. 2009, pp. 113 –116.

[23] Z. Quan, S. Cui, V. H. Poor, and A. H. Sayed, "Collaborative wideband sensing for cognitive radios," *IEEE Signal Processing Magazine*, vol. 25, pp. 60–73, Nov. 2008.

[24] M. L. Stein, *Interpolation of Spatial Data*. Springer, 1999.

[25] M. Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal Processing Magazine*, vol. 16, pp. 22–38, Nov. 1999.

[26] G. Wahba, *Spline Models for Observational Data*. SIAM, 1990.

[27] G. Wahba and J. Wendelberger, "Some new mathematical methods for variational objective analysis using splines and cross validation," *Monthly Weather Review*, vol. 108, pp. 1122–1145, 1980.

[28] S. J. Wright, R. D. Novak, and M. A. T. Figueredo, "Sparse reconstruction by separable approximation," *IEEE Trans. on Signal Processing*, vol. 57, pp. 2479–2493, Jul. 2009.

[29] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Royal. Statist. Soc B*, vol. 68, pp. 49–67, 2006.

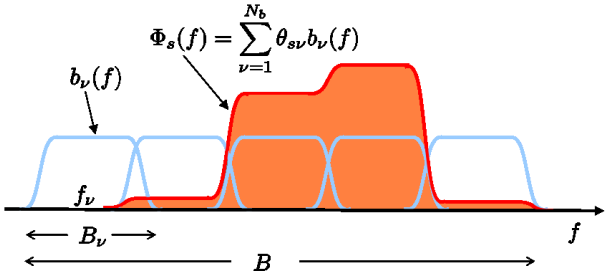

Fig. 1. Expansion with overlapping raised cosine pulses.

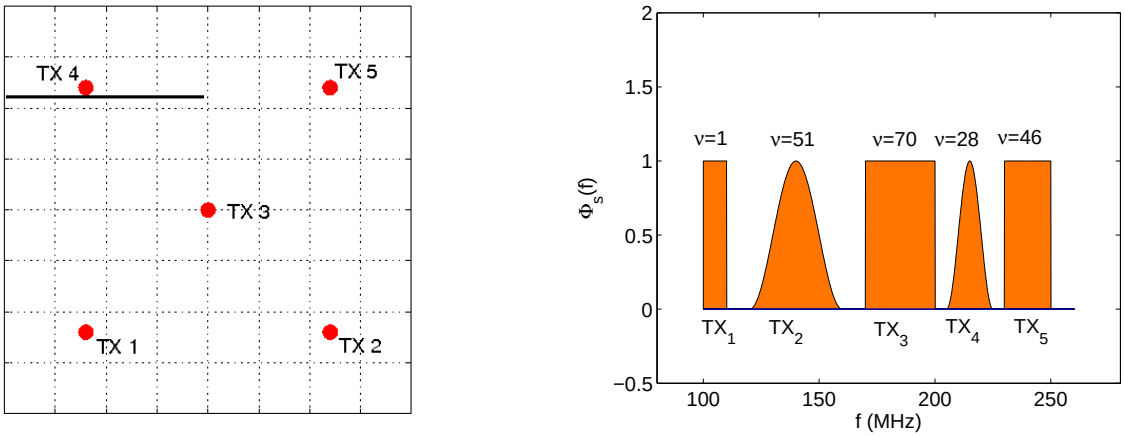

Fig. 2. (left) Position of sources and obstructing wall; (right) PSD generated by the active transmitters.

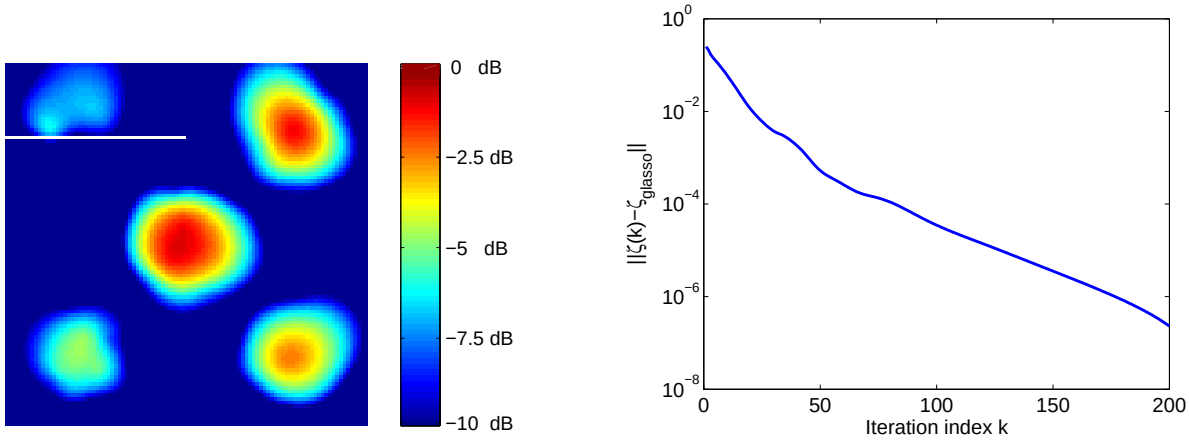

Fig. 3. (left) Aggregate map estimate in dB; (right) error evolution of the GLasso updates

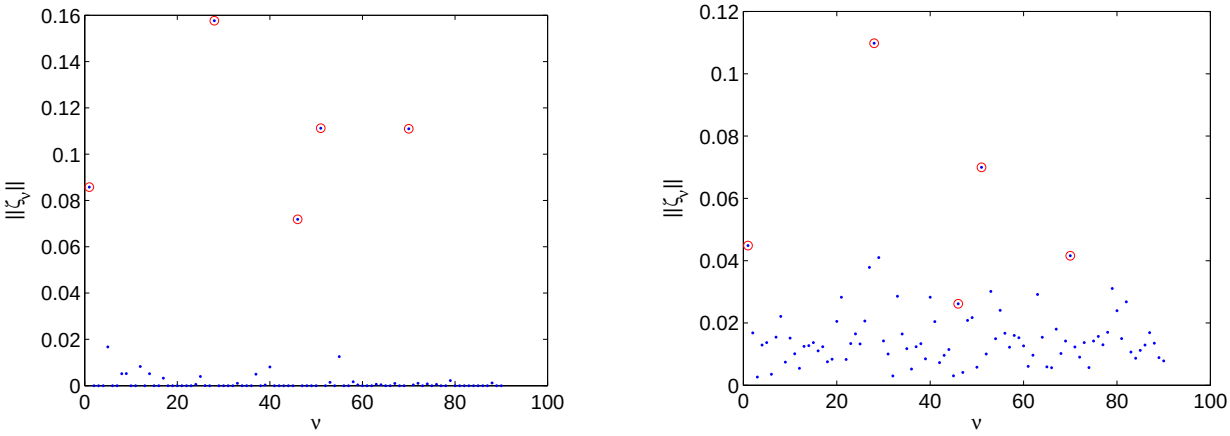

Fig. 4. (left) Frequency bases selected by the group-Lassoed spline-based estimator; (right) and by ridge regression.

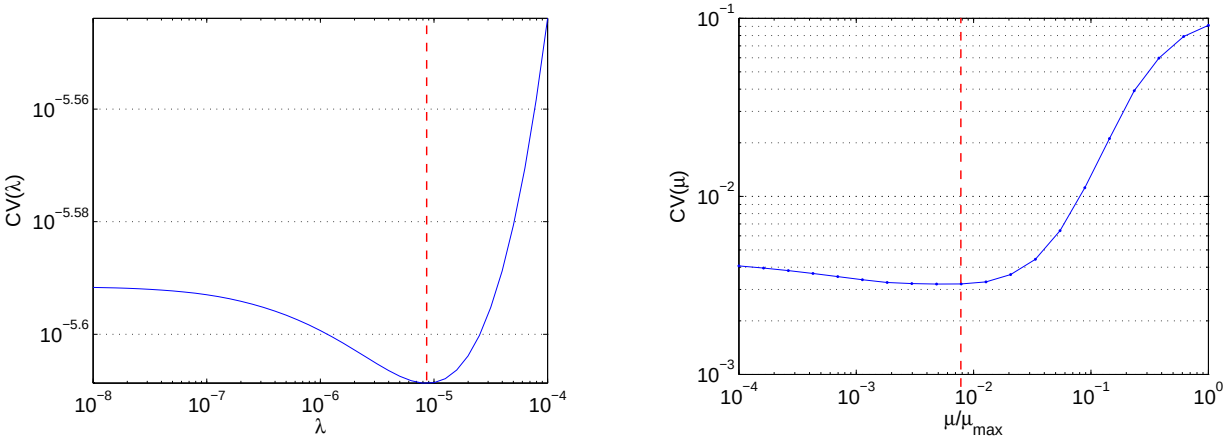

Fig. 5. (left) Minimization of the CV error over λ; (right) and over μ.

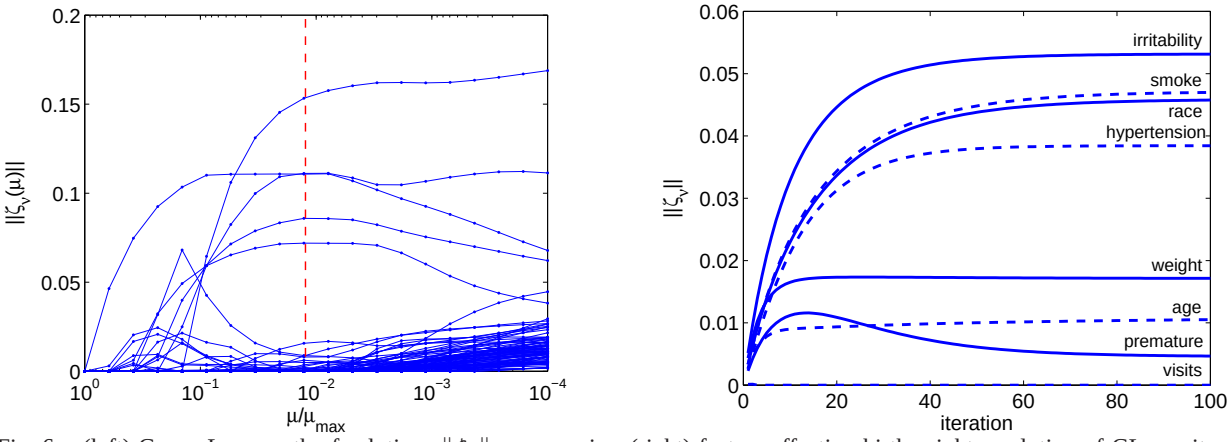

Fig. 6. (left) Group-Lasso path of solutions $\|\zeta_\nu\|_2$ as μ varies; (right) factors affecting birthweight, evolution of GLasso iterates.