Stochastic Semiparametric Regression for Spectrum Cartography

Daniel Romero
Dept. of ECE & DTC
University of Minnesota, USA
Email: dromero@umn.edu

Seung-Jun Kim
Dept. of Computer Sci. & Electrical Eng.
University of Maryland, Baltimore County, USA
Email: sjkim@umbc.edu

Georgios B. Giannakis
Dept. of ECE & DTC
University of Minnesota, USA
Email: georgios@umn.edu

***Abstract*—An online spectrum cartography algorithm is proposed to reconstruct power spectral density (PSD) maps in space and frequency based on compressed and quantized sensor measurements. The emerging regression task is addressed by decomposing the PSD at every location into a linear combination of the power spectra (due to individual transmitters and background noise) scaled by attenuation functions capturing propagation effects. The attenuation functions are, in turn, postulated to be a sum of two terms: the first is a linear combination of a collection of basis functions whereas the second is an element of a reproducing kernel Hilbert space (RKHS) of vector-valued functions. A novel stochastic gradient descent algorithm is proposed to compute both components in an online fashion. Numerical tests verify the map estimation performance of the proposed technique.**

I. Introduction

The goal of power spectrum cartography is to characterize the RF signal power distribution over a geographical region by means of spectrum maps. The maps are instrumental for various management tasks in wireless networks, such as power control, interference mitigation, network planning, and in particular, in dynamic spectrum access (DSA) of cognitive radios, which aspire to opportunistically exploit underutilized spectral resources in time, space, and frequency [1].

Spatial interpolation of RF power measurements has been tackled using Kriging [2], [3], orthogonal matching pursuit [4], sparse Bayesian learning [5], and dictionary learning [6]. To account for the frequency dimension, a basis expansion model (BEM) was adopted in [7], [8]. However, these techniques require exchanging raw measurements, which imposes stringent requirements on the control channel. This issue was mitigated in [9], but the spatial dimension was not accounted for. In [10], we presented a nonparametric method to estimate spectrum maps from a collection of linearly compressed and heavily quantized observations, which alleviates all these difficulties. A stochastic algorithm to compute this solution in linear time was also proposed in [11].

Nonetheless, none of these approaches effectively exploit the partial knowledge on the propagation effects experienced by electromagnetic waves. In particular, the attenuation is typically decomposed into a pathloss term that models how the power decreases with distance plus a term that models shadowing/fading. The first one can be approximately known in practice, especially if the characteristics of the propagation environment (e.g. urban, suburban, etc) and locations of the transmitters are known. This is typically the case, for instance, in cellular networks or terrestrial television systems. On the other hand, the second term is usually highly unstructured and difficult to predict. The contribution of the present work is twofold: on the one hand, we propose a semiparametric technique to exploit this partial knowledge by assuming that the spatial fields representing how the radiated power distributes across space are the sum of a parametric component, which lies in the span of a set of basis functions, and a nonparametric component, which lies in a reproducing kernel Hilbert space (RKHS). On the other hand, we propose an online algorithm to efficiently compute the proposed estimate in linear time. To the best of our knowledge, this is the first work to consider online semiparametric regression in RKHSs.

In the frequency domain, a BEM is employed to exploit spectral prior information and to interpret the spectrum contents in terms of the spatial power distribution of each channel. As in [10], [11], wideband sensing is enabled with sub-Nyquist sampling, while the communication overhead is reduced via measurement compression and quantization. Robustness against measurement errors is ensured using machine learning techniques.

The rest of the paper is structured as follows. The system model and problem statement are introduced in Sec. II. Sec. III presents the online map estimation algorithm, whose performance is verified via simulations in Sec. IV.

II. Model and Problem Statement

Consider $M-1$ radios operating in a geographical region $\mathcal{R} \subset \mathbb{R}^d$ of interest, with d typically being 2 or 3. The m-th radio transmits a signal $\sqrt{\gamma_m} s_m(t)$, where γ_m represents its power and $s_m(t)$ a (possibly complex) wide-sense stationary random process, normalized so that $\mathbb{E}\left\{|s_m(t)|^2\right\} = 1 \ \forall t$. Let $\phi_m(f)$ denote the power spectral density (PSD) of $s_m(t)$, which for simplicity is assumed known since most primary systems obey transmission standards (e.g. DVB or ATSC in TV bands) and spectrum mask regulations, which fix the transmission parameters, such as bandwidth, carrier frequencies, and roll-off factors [12]. If the functions $\phi_m(f),\ m = 1, \ldots, M-1$, are not known, our approach can still be used by adopting a general BEM [7], [8].

Supported by the Spanish Government and the European Regional Development Fund (ERDF) (project TEC2013-47020- C2-1-R COMPASS, FPU Grant AP2010-0149), by the Galician Regional Government and ERDF (projects GRC2013/009, R2014/037 and AtlantTIC), and by NSF grants 1202135, 1247885, and 1343248.

The received waveform at position $\boldsymbol{x} \in \mathcal{R}$, due to the $M-1$ uncorrelated transmitters, can be expressed as

$$r_{\boldsymbol{x}}(t) = \sum_{m=1}^{M-1} r_{\boldsymbol{x},m}(t) + r_{\boldsymbol{x},M}(t) = \sum_{m=1}^{M} r_{\boldsymbol{x},m}(t) \qquad (1)$$

where $r_{\boldsymbol{x},M}(t)$ captures additive receiver noise and $r_{\boldsymbol{x},m}(t)$, $m = 1, 2, \ldots, M-1$, corresponds to transmitter m. Assuming frequency-flat channels, the PSD of $r_{\boldsymbol{x}}(t)$ can be written as

$$P(\boldsymbol{x}, f) = \sum_{m=1}^{M} l_m(\boldsymbol{x}) \phi_m(f) \qquad (2)$$

where $\phi_M(f)$ is the noise PSD, normalized such that $\int_{-\infty}^{\infty} \phi_M(f) df = 1$, $l_M(\boldsymbol{x})$ is the noise power at $\boldsymbol{x}$, and the coefficients $l_m(\boldsymbol{x})$, $m = 1, 2, \ldots, M-1$, subsume the transmit power γ_m and the propagation effects between transmitter m and $\boldsymbol{x}$. This representation, which can be seen as a BEM per position $\boldsymbol{x}$, also applies to certain frequency-selective cases [7, Sec. II]. Since $\phi_m(f)$ is normalized, $l_m(\boldsymbol{x})$, $m = 1, 2, \ldots, M-1$, represents the power of the signal due to the m-th transmitter at position $\boldsymbol{x}$.

A network of N sensors is deployed at positions collected in the set $\mathcal{X} := \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \mathcal{R}$ to estimate the PSD map (2). In order to allow for estimation in wide frequency bands at low hardware cost and power consumption, acquisition via analog-to-information converters (AICs) is considered [13]. For each input block $\boldsymbol{r}_{\boldsymbol{x}}[b] := [r_{\boldsymbol{x}}[bL], \ldots, r_{\boldsymbol{x}}[bL + (L-1)]]^T$ of L samples $r_{\boldsymbol{x}}[k] := r_{\boldsymbol{x}}(kT_s)$, where $1/T_s$ is the Nyquist rate, an AIC produces a linearly compressed block $\tilde{\boldsymbol{r}}_{\boldsymbol{x}}[b]$ of $\tilde{L}$ ($< L$) samples:

$$\tilde{\boldsymbol{r}}_{\boldsymbol{x}}[b] = \tilde{\boldsymbol{G}} \boldsymbol{r}_{\boldsymbol{x}}[b]. \qquad (3)$$

Here, $\tilde{\boldsymbol{G}} \in \mathbb{C}^{\tilde{L} \times L}$ is the compression matrix. In order to guarantee the identifiability of the $l_m(\boldsymbol{x})$'s, this matrix must satisfy the criteria in [14].

The whole observation frame $\boldsymbol{r}_{\boldsymbol{x}} := [\boldsymbol{r}_{\boldsymbol{x}}^T[0], \ldots, \boldsymbol{r}_{\boldsymbol{x}}^T[B-1]]^T$ therefore produces $\tilde{\boldsymbol{r}}_{\boldsymbol{x}} = \boldsymbol{G} \boldsymbol{r}_{\boldsymbol{x}}$, where $\tilde{\boldsymbol{r}}_{\boldsymbol{x}} := [\tilde{\boldsymbol{r}}_{\boldsymbol{x}}^T[0], \ldots, \tilde{\boldsymbol{r}}_{\boldsymbol{x}}^T[B-1]]^T$ and $\boldsymbol{G} := \boldsymbol{I}_B \otimes \tilde{\boldsymbol{G}}$ ($\otimes$ denotes the Kronecker product). Note that this is merely a convenient abstraction, since an AIC does not go through the intermediate step of obtaining the Nyquist samples, but directly obtains the compressed observations $\tilde{\boldsymbol{r}}_{\boldsymbol{x}}$.

The power of $\tilde{\boldsymbol{r}}_{\boldsymbol{x}}$ can then be expressed as

$$\eta_{\boldsymbol{x}} = \frac{1}{B\tilde{L}} \mathbb{E}\left\{\tilde{\boldsymbol{r}}_{\boldsymbol{x}}^T \tilde{\boldsymbol{r}}_{\boldsymbol{x}}\right\} = \frac{1}{B\tilde{L}} \operatorname{Tr}\left(\boldsymbol{G} \boldsymbol{\Sigma} \boldsymbol{G}^T\right) \qquad (4)$$

where $\boldsymbol{\Sigma} := \mathbb{E}\left\{\boldsymbol{r}_{\boldsymbol{x}} \boldsymbol{r}_{\boldsymbol{x}}^T\right\}$. The latter can be written as

$$\boldsymbol{\Sigma} = \sum_{m=1}^{M} l_m(\boldsymbol{x}) \boldsymbol{\Sigma}_m \qquad (5)$$

where the matrix $\boldsymbol{\Sigma}_m := \mathbb{E}\left\{\boldsymbol{s}_m \boldsymbol{s}_m^T\right\}$; while the vector $\boldsymbol{s}_m := [s_m[0], \ldots, s_m[BL-1]]^T$ is formed with the corresponding Nyquist samples of $s_m(t)$, and can be obtained from the inverse Fourier transform of $\phi_m(f)$. Upon defining $\boldsymbol{\phi}_{\boldsymbol{x}} := [\phi_{\boldsymbol{x},1}, \ldots, \phi_{\boldsymbol{x},M}]^T$ with $\phi_{\boldsymbol{x},m} := (B\tilde{L})^{-1} \operatorname{Tr}\left(\boldsymbol{G} \boldsymbol{\Sigma}_m \boldsymbol{G}^T\right)$, $\boldsymbol{l}(\boldsymbol{x}) := [l_1(\boldsymbol{x}), \ldots, l_M(\boldsymbol{x})]^T$, and plugging (5) into (4) yield

$$\eta_{\boldsymbol{x}} = \boldsymbol{\phi}_{\boldsymbol{x}}^T \boldsymbol{l}(\boldsymbol{x}). \qquad (6)$$

To sum up, the power of the signal at the output of the AIC is just a linear combination of the power received from each transmitter. In other words, by measuring $\eta_{\boldsymbol{x}}$ we obtain a linearly compressed observation of $\boldsymbol{l}(\boldsymbol{x})$ [14].

After estimating $\eta_{\boldsymbol{x}}$ based on its own measurements, the sensor at position $\boldsymbol{x}$ uniformly quantizes its estimate $\hat{\eta}_{\boldsymbol{x}}$ using

$$q_{\boldsymbol{x}} := Q(\hat{\eta}_{\boldsymbol{x}}) := \lfloor \hat{\eta}_{\boldsymbol{x}} / (2\epsilon) \rfloor \in \mathbb{Z} \qquad (7)$$

where 2ϵ is the quantization step size, and sends the result to the fusion center (FC) through a control channel. Depending on the quality of the estimate, either $Q(\eta_{\boldsymbol{x}}) = Q(\hat{\eta}_{\boldsymbol{x}})$ or $Q(\eta_{\boldsymbol{x}}) \neq Q(\hat{\eta}_{\boldsymbol{x}})$ is possible. The latter case will be termed as a *measurement error*.

The problem addressed in this paper is that of estimating the PSD map $P(\boldsymbol{x}, f)$. Since the functions $\phi_m(f)$ are known, this is tantamount to estimating $l_m(\boldsymbol{x})$ for all m, seen as M functions of the spatial coordinate $\boldsymbol{x}$. The information available comprises the measurement vectors $\{\boldsymbol{\phi}_{\boldsymbol{x}}\}_{\boldsymbol{x} \in \mathcal{X}}$, the quantized observations $\{q_{\boldsymbol{x}}\}_{\boldsymbol{x} \in \mathcal{X}}$, and the locations of the N sensors $\mathcal{X} := \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \mathcal{R}$. Note that unlike $\{q_{\boldsymbol{x}}\}_{\boldsymbol{x} \in \mathcal{X}}$, $\{\boldsymbol{\phi}_{\boldsymbol{x}}\}_{\boldsymbol{x} \in \mathcal{X}}$ and $\mathcal{X}$ are assumed known *a priori* to the FC, and thus need not be communicated.

III. PSD Map Estimation

Upon receiving $q_{\boldsymbol{x}}$, the FC learns that the true $\boldsymbol{l}$ satisfies $2\epsilon q_{\boldsymbol{x}} \leq \boldsymbol{\phi}_{\boldsymbol{x}}^T \boldsymbol{l}(\boldsymbol{x}) < 2\epsilon(q_{\boldsymbol{x}} + 1)$ unless a measurement error has occurred [cf. (7)]. Thus, upon defining $y_{\boldsymbol{x}} := (2q_{\boldsymbol{x}} + 1)\epsilon$,

$$|y_{\boldsymbol{x}} - \boldsymbol{\phi}_{\boldsymbol{x}}^T \boldsymbol{l}(\boldsymbol{x})| \leq \epsilon. \qquad (8)$$

The problem becomes that of choosing a member of the family of functions $\boldsymbol{l} : \mathbb{R}^d \to \mathbb{R}^M$ satisfying (8) for all non-erroneous $y_{\boldsymbol{x}}$. Since it is not known a priori which measurements are erroneous, one can resort to the *regularization inductive principle*, where a linear combination of a fitting cost R_{emp}, called empirical risk, and a smoothness-enforcing penalty J, is minimized as (see, e.g., [15])

$$\underset{\boldsymbol{l} \in \mathcal{S}}{\text{minimize}} \quad R_{\text{emp}}(\boldsymbol{l}; \{(\boldsymbol{x}, \boldsymbol{\phi}_{\boldsymbol{x}}, y_{\boldsymbol{x}})\}_{\boldsymbol{x} \in \mathcal{X}}) + \lambda J(\boldsymbol{l}). \qquad (9)$$

Here, $\lambda > 0$ is a constant adjusted to attain the desired trade-off between fit and smoothness and $\mathcal{S}$ is a space of functions of the form $\boldsymbol{l} = \boldsymbol{l}_{\text{par}} + \boldsymbol{l}_{\text{npar}}$, where $\boldsymbol{l}_{\text{npar}}$ is in an RKHS $\mathcal{S}'$ of vector-valued functions from $\mathbb{R}^d$ to $\mathbb{R}^M$ [16] and $\boldsymbol{l}_{\text{par}}$ can be written as

$$\boldsymbol{l}_{\text{par}} = \sum_{b=1}^{B-1} \boldsymbol{\Psi}_b \boldsymbol{d}_b \qquad (10)$$

where the $\boldsymbol{\Psi}_b$ can be thought of as basis functions of the form $\boldsymbol{\Psi}_b : \mathbb{R}^d \to \mathbb{R}^{M \times M}$. We will compute the solution of (9) by independently optimizing with respect to these two components.

The penalty J is chosen to be $J(\boldsymbol{l}) := \|\boldsymbol{l}_{\text{npar}}\|_{\mathcal{S}'}^2$, where $\|\cdot\|_{\mathcal{S}'}$ is the norm induced by the inner product of $\mathcal{S}'$. The empirical risk R_{emp} linearly penalizes deviations from (8) using the so-called ϵ-insensitive loss $u_\epsilon(y) := \max(0, |y| - \epsilon)$ [15]

$$R_{\text{emp}}(\boldsymbol{l}; \{(\boldsymbol{x}, \boldsymbol{\phi}_{\boldsymbol{x}}, y_{\boldsymbol{x}})\}_{\boldsymbol{x} \in \mathcal{X}}) := \frac{1}{N} \sum_{\boldsymbol{x} \in \mathcal{X}} u_\epsilon(y_{\boldsymbol{x}} - \boldsymbol{\phi}_{\boldsymbol{x}}^T \boldsymbol{l}(\boldsymbol{x})). \qquad (11)$$

The sum on the right-hand side is a convex surrogate for the number of measurement errors, in the same way as the ℓ_1-norm is typically substituted for the ℓ_0-norm in sparse regression. In other words, R_{emp} captures the sparsity present in the measurement errors.

The RKHS $\mathcal{S}'$ can be specified using a reproducing kernel, which is a function $\boldsymbol{K}(\boldsymbol{z},\boldsymbol{x}) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^{M\times M}$. As with scalar RKHSs, $\boldsymbol{K}(\boldsymbol{z},\boldsymbol{x})$ represents a valid kernel provided certain requirements are met [16]. A simple construction is $\boldsymbol{K}(\boldsymbol{z},\boldsymbol{x}) = \text{diag}\left\{k^{(1)}(\boldsymbol{z},\boldsymbol{x}),\ldots,k^{(M)}(\boldsymbol{z},\boldsymbol{x})\right\}$, where $k^{(m)}(\boldsymbol{z},\boldsymbol{x})$ are valid scalar kernels.

Although the problem (9) could be solved extending the batch technique from [10], in practice online approaches are preferable for two reasons. First, the computational cost of batch approaches grows prohibitively as the number of measurements increases. *Online* algorithms, on the other hand, keep the complexity linear by incrementally updating their estimates as new measurements arrive at the FC. Second, online algorithms can track slow temporal variations of the fields of interest.

Our approach is based on computing the nonparametric component $\boldsymbol{l}_{\text{npar}}$ via stochastic gradient descent in $\mathcal{S}'$ [17], [18] and the parametric component $\boldsymbol{l}_{\text{par}}$ via stochastic gradient descent in the Euclidean space. First, one can define the instantaneous regularized error as

$$\begin{aligned} &R_{\text{inst}}(\boldsymbol{l}_{\text{npar}},\boldsymbol{d},\boldsymbol{\phi},\boldsymbol{x},y) \\ &\quad := u_\epsilon(y-\boldsymbol{\phi}^T\boldsymbol{l}_{\text{npar}}(\boldsymbol{x})-\boldsymbol{\phi}^T\boldsymbol{l}_{\text{par}}(\boldsymbol{x}))+\lambda\|\boldsymbol{l}_{\text{npar}}\|^2_{\mathcal{S}'}, \end{aligned} \tag{12}$$

where $\boldsymbol{d} := [\boldsymbol{d}_1^T,\ldots,\boldsymbol{d}_{B-1}^T]^T$. Note that the objective of (9) is just the sample average of R_{inst}. Suppose that, at every time slot $t = 1, 2, \ldots$, an observation $y_{\boldsymbol{x}_t}$ is received from the sensor at $\boldsymbol{x}_t$ (multiple measurements associated with different $\boldsymbol{\phi}_{\boldsymbol{x}_t}$ received from each sensor can be accommodated without modification). We propose the following update rule to be applied every time an observation $(y_{\boldsymbol{z}},\boldsymbol{\phi}_{\boldsymbol{z}})$ is received from the sensor at location $\boldsymbol{z}\in\mathcal{X}$:

$$\begin{aligned} \boldsymbol{l}_{\text{npar}}^{(t+1)} &= \boldsymbol{l}_{\text{npar}}^{(t)} - \kappa^{(t)}\partial_{\boldsymbol{l}_{\text{npar}}} R_{\text{inst}}(\boldsymbol{l}_{\text{npar}},\boldsymbol{d}^{(t)},\boldsymbol{\phi}_{\boldsymbol{x}_t},\boldsymbol{x}_t,y_{\boldsymbol{x}_t})|_{\boldsymbol{l}_{\text{npar}}=\boldsymbol{l}_{\text{npar}}^{(t)}} \\ \boldsymbol{d}^{(t+1)} &= \boldsymbol{d}^{(t)} - \mu^{(t)}\partial_{\boldsymbol{d}} R_{\text{inst}}(\boldsymbol{l}_{\text{npar}}^{(t)},\boldsymbol{d},\boldsymbol{\phi}_{\boldsymbol{x}_t},\boldsymbol{x}_t,y_{\boldsymbol{x}_t})|_{\boldsymbol{d}=\boldsymbol{d}^{(t)}} \end{aligned} \tag{13}$$

Here, $\boldsymbol{l}^{(t)}$, $\boldsymbol{l}_{\text{npar}}^{(t)}$ and $\boldsymbol{d}^{(t)}$ correspond to the estimate at time t; $\kappa^{(t)},\mu^{(t)}$ are the (positive) learning rates, and ∂ denotes subgradient. On the one hand, from [18, eq. (2)]

$$\begin{aligned} &\partial_{\boldsymbol{l}_{\text{npar}}} R_{\text{inst}}(\boldsymbol{l}_{\text{npar}},\boldsymbol{d}^{(t)},\boldsymbol{\phi},\boldsymbol{x},y) \\ &\qquad = -u_\epsilon'(y-\boldsymbol{\phi}^T\boldsymbol{l}(\boldsymbol{x}))\cdot k_{\boldsymbol{x}}\boldsymbol{\phi} + 2\lambda\boldsymbol{l}_{\text{npar}} \end{aligned} \tag{14}$$

where $k_{\boldsymbol{x}}$ is the kernel operator of $\mathcal{S}'$, and u_ϵ' is a subgradient of u_ϵ, which can be taken as $u_\epsilon'(z) = \frac{1}{2}(\text{sgn}(z-\epsilon)+\text{sgn}(z+\epsilon))$. On the other hand, if $\boldsymbol{\Psi} := [\boldsymbol{\Psi}_1,\ldots,\boldsymbol{\Psi}_{B-1}]$, we have that

$$\begin{aligned} &\partial_{\boldsymbol{d}} R_{\text{inst}}(\boldsymbol{l}_{\text{npar}},\boldsymbol{d},\boldsymbol{\phi},\boldsymbol{x},y) \\ &\qquad = -u_\epsilon'\left(y-\boldsymbol{\phi}^T\boldsymbol{l}(\boldsymbol{x})\right)\boldsymbol{\Psi}^T(\boldsymbol{x})\boldsymbol{\phi} \end{aligned} \tag{15}$$

Substituting (14) and (15) in (13) yields

$$\begin{aligned} \boldsymbol{l}_{\text{npar}}^{(t+1)} &= (1-2\kappa^{(t)}\lambda)\boldsymbol{l}_{\text{npar}}^{(t)} + \kappa^{(t)}u_\epsilon'[y_{\boldsymbol{x}_t}-\boldsymbol{\phi}_{\boldsymbol{x}_t}^T\boldsymbol{l}^{(t)}(\boldsymbol{x}_t)]k_{\boldsymbol{x}_t}\boldsymbol{\phi}_{\boldsymbol{x}_t} \\ \boldsymbol{d}^{(t+1)} &= \boldsymbol{d}^{(t)} + \mu^{(t)}u_\epsilon'\left(y_{\boldsymbol{x}_t}-\boldsymbol{\phi}_{\boldsymbol{x}_t}^T\boldsymbol{l}^{(t)}(\boldsymbol{x}_t)\right)\boldsymbol{\Psi}^T(\boldsymbol{x}_t)\boldsymbol{\phi}_{\boldsymbol{x}_t} \end{aligned}$$

Although this gives a simple update rule for $\boldsymbol{d}^{(t)}$, $\boldsymbol{l}_{\text{npar}}^{(t)}$ cannot be manipulated in its present form. However, invoking the Representer Theorem [16, Thm. 5], one can see that, after at least one measurement is received from each sensor, the solution of (9) satisfies

$$\boldsymbol{l}_{\text{npar}}^{(t)} = \sum_{\boldsymbol{x}\in\mathcal{X}} k_{\boldsymbol{x}}\boldsymbol{c}_{\boldsymbol{x}}^{(t)} \quad \text{for some } \boldsymbol{c}_{\boldsymbol{x}}^{(t)}\in\mathbb{R}^M. \tag{16}$$

Thus, one can simply operate on the vectors $\boldsymbol{c}_{\boldsymbol{x}}^{(t)}$, which can be updated as

$$\boldsymbol{c}_{\boldsymbol{x}}^{(t+1)} = (1-2\kappa^{(t)}\lambda)\boldsymbol{c}_{\boldsymbol{x}}^{(t)} \text{ for } \boldsymbol{x}\in\mathcal{X},\ \boldsymbol{x}\neq\boldsymbol{z} \tag{17}$$

$$\boldsymbol{c}_{\boldsymbol{z}}^{(t+1)} = (1-2\kappa^{(t)}\lambda)\boldsymbol{c}_{\boldsymbol{z}}^{(t)} + \kappa^{(t)}u_\epsilon'(y_{\boldsymbol{z}}-\boldsymbol{\phi}_{\boldsymbol{z}}^T\boldsymbol{l}^{(t)}(\boldsymbol{z}))\boldsymbol{\phi}_{\boldsymbol{z}} \tag{18}$$

which follows from the linear independence of the kernel operators $\{k_{\boldsymbol{x}}\}_{\boldsymbol{x}\in\mathcal{X}}$. As a possible initialization, one may choose $\boldsymbol{c}_{\boldsymbol{x}}^{(1)} = \boldsymbol{0}$ and $\boldsymbol{d}^{(1)} = \boldsymbol{0}$.

IV. Numerical Tests

To visually analyze the operation of the proposed semi-parametric algorithm, we consider a simple scenario where $\mathcal{R} = [0,1]$ (hence $d = 1$). The propagation model assumes that, for $m = 1, 2, 3$,

$$10\log_{10} l_m(\boldsymbol{x}) = 10\log_{10} A_m + \psi_m(\boldsymbol{x}) + u_m(\boldsymbol{x}) \tag{19}$$

where the parameters $A_1 = 0.6$, $A_2 = 0.2$, and $A_3 = 0.6$ capture the transmitted power; $\psi_m(\boldsymbol{x})$ relates to the propagation loss, and $u_m(\boldsymbol{x})$ represents shadowing. The propagation loss follows the Okumura-Hata model [19, Sec. 9.1.2], where

$$\psi_m(\boldsymbol{x}) = \gamma[\log_{10}\delta - \log_{10}(\delta + \|\boldsymbol{x}-\boldsymbol{z}_m\|)]. \tag{20}$$

Here, $\delta = 0.1$ is a small constant ensuring that the arguments of the logarithms do not vanish, $\gamma = 10$ relates to how fast the radiated power decays with distance, and $\boldsymbol{z}_1 = 0.1$, $\boldsymbol{z}_2 = 0.3$, $\boldsymbol{z}_3 = 0.8$, represent the location of the transmitters. The lognormal shadowing processes $u_m(\boldsymbol{x})$ are generated as described in [20]. Finally, the noise PSD is $l_4(\boldsymbol{x}) = A_4 = 0.75$.

Our test used $N = 30$ sensors deployed uniformly at random, reporting one measurement to the FC per time slot. Uniform quantization to 3 bits per measurement was implemented, where ϵ was set so that the probability of clipping $\text{P}\{\eta_{\boldsymbol{x}}^{(t)} > r_R\} \approx 10^{-3}$. A Gaussian diagonal kernel with diagonal entries $K(\boldsymbol{z},\boldsymbol{x})|_{i,i} = \exp\left\{-||\boldsymbol{z}-\boldsymbol{x}||^2/\sigma_i^2\right\}$ was adopted with $\sigma_i = 0.022$, and the regularization constant was set to $\lambda = 0.005$. The parametric part of the proposed method is defined by $\boldsymbol{\Psi}_1 = \text{diag}\{\psi_1,\ldots,\psi_4\}$ where the ψ_m, $m = 1, 2, 3$, are defined in (20), and $\psi_4(\boldsymbol{x}) = 1\ \forall\boldsymbol{x}$. The learning rates were set to $\kappa^{(t)} = \mu^{(t)} = 0.1$.

As mentioned in Sec. II, the result of quantizing the sensor estimate $\hat{\eta}_{\boldsymbol{x}}$ may differ from $Q(\eta_{\boldsymbol{x}})$ since the observation windows are finite. In order to model these measurement errors, we set $\hat{\eta}_{\boldsymbol{x}}^{(t)} = \eta_{\boldsymbol{x}}^{(t)} + \xi_{\boldsymbol{x}}^{(t)}$, where $\xi_{\boldsymbol{x}}^{(t)}$ is a normally distributed random variable with zero mean and variance σ_ξ^2 that captures estimation errors. In our simulations, σ_ξ was set to $\sigma_\xi = 0.05$, which results in a probability of measurement error approximately of 0.2.

Fig. 1 represents a realization of the channel and the result of executing the proposed algorithm for 30 time slots, where an observation is received from each sensor. It is observed

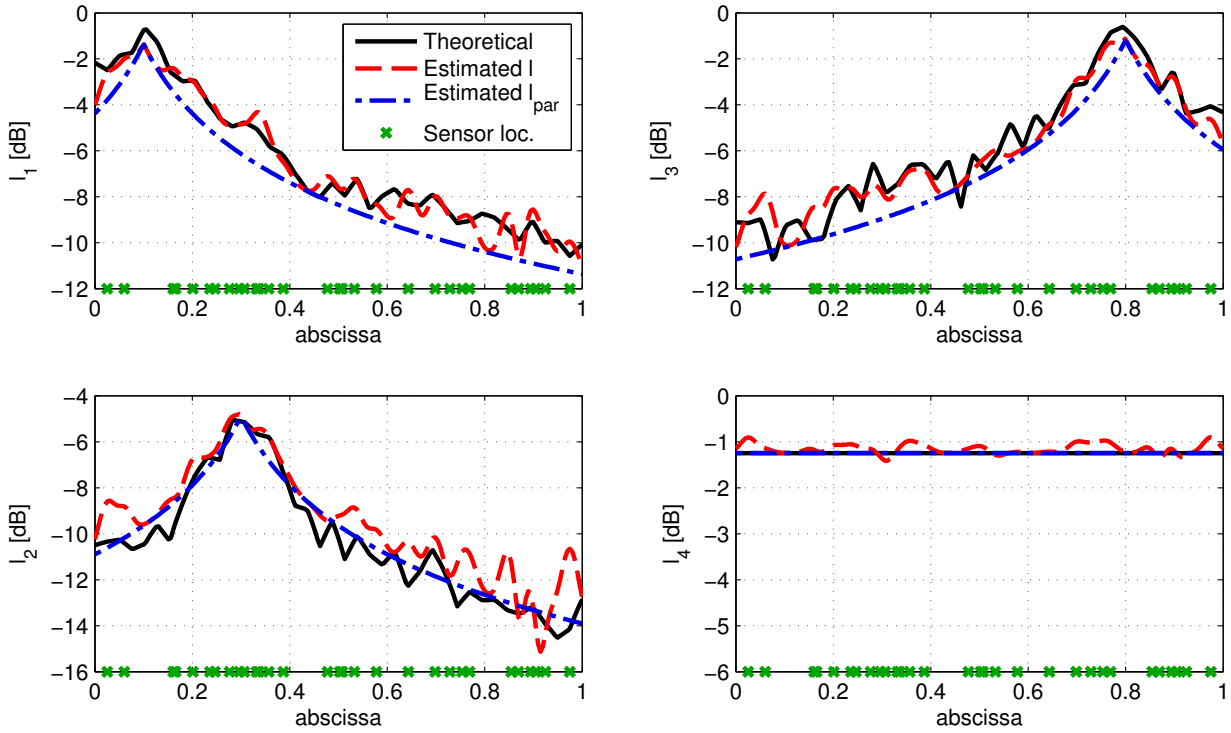

Fig. 1: Channel realization and the resulting estimate.

that the parametric part accounts for most of the energy of the estimated $\boldsymbol{l}$, thus capturing the effects of attenuation due to the distance, whereas the nonparametric part fits the effects of shadowing. Clearly, the quality of the estimate is poorer in those areas where no sensors have been deployed or where the received power is low, which is a consequence of quantization. However, the overall results are satisfactory, especially noting that barely 340 bytes have been used for the estimate.

REFERENCES

[1] E. Axell, G. Leus, E. G. Larsson, and H. V. Poor, "Spectrum sensing for cognitive radio: State-of-the-art and recent advances," *IEEE Sig. Process. Mag.*, vol. 29, no. 3, pp. 101–116, May 2012.

[2] A. Alaya-Feki, S. B. Jemaa, B. Sayrac, P. Houze, and E. Moulines, "Informed spectrum usage in cognitive radio networks: Interference cartography," in *Proc. IEEE Int. Symp. Personal, Indoor Mobile Radio Commun.*, 2008, pp. 1–5.

[3] S.-J. Kim, E. Dall'Anese, and G. B. Giannakis, "Cooperative spectrum sensing for cognitive radios using kriged Kalman filtering," *IEEE J. Sel. Topics Sig. Process.*, vol. 5, no. 1, pp. 24–36, Feb 2011.

[4] B. A. Jayawickrama, E. Dutkiewicz, I. Oppermann, G. Fang, and J. Ding, "Improved performance of spectrum cartography based on compressive sensing in cognitive radio networks," in *Proc. Int. Conf. Commun.*, Budapest, Hungary, Jun. 2013, pp. 5657–5661.

[5] D.-H. Huang, S.-H. Wu, W.-R. Wu, and P.-H. Wang, "Cooperative radio source positioning and power map reconstruction: A sparse Bayesian learning approach," *IEEE Trans. Veh. Technol.*, vol. 64, no. 6, pp. 2318–2332, Jun. 2015.

[6] S.-J. Kim and G. B. Giannakis, "Cognitive radio spectrum prediction using dictionary learning," in *Proc. IEEE Global Commun. Conf.*, Atlanta, GA, Dec. 2013.

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

[8] J.-A. Bazerque, G. Mateos, and G. B. Giannakis, "Group-lasso on splines for spectrum cartography," *IEEE Trans. Sig. Process.*, vol. 59, no. 10, pp. 4648–4663, Oct. 2011.

[9] O. Mehanna and N. Sidiropoulos, "Frugal sensing: Wideband power spectrum sensing from few bits," *IEEE Trans. Sig. Process.*, vol. 61, no. 10, pp. 2693–2703, May 2013.

[10] D. Romero, S.-J. Kim, R. López-Valcarce, and G. B. Giannakis, "Spectrum cartography using quantized observations," in *Proc. IEEE Int. Conf. Acoust., Speech, Sig. Process.*, Brisbane, Australia, Apr. 2015, pp. 3252 – 3256.

[11] D. Romero, S.-J. Kim, and G. B. Giannakis, "Online spectrum cartography via quantized measurements," in *Proc. Conf. Inf. Sci. Syst.*, Baltimore, MD, Mar. 2015.

[12] D. Romero and G. Leus, "Wideband spectrum sensing from compressed measurements using spectral prior information," *IEEE Trans. Sig. Process.*, vol. 61, no. 24, pp. 6232–6246, 2013.

[13] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, "Beyond Nyquist: Efficient sampling of sparse bandlimited signals," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 520–544, Jan. 2010.

[14] D. Romero, R. López-Valcarce, and G. Leus, "Compression limits for random vectors with linearly parameterized second-order statistics," *IEEE Trans. Inf. Theory*, vol. 61, no. 3, pp. 1410–1425, Mar. 2015.

[15] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, 2002.

[16] C. A. Micchelli and M. Pontil, "On learning vector-valued functions," *Neural Comput.*, vol. 17, no. 1, pp. 177–204, 2005.

[17] J. Kivinen, A. J. Smola, and R. C. Williamson, "Online learning with kernels," *IEEE Trans. Sig. Process.*, vol. 52, no. 8, pp. 2165–2176, 2004.

[18] J. Audiffren and H. Kadri, "Online learning with multiple operator-valued kernels," Preprint at http://arXiv.org/abs/1311.0222, 2013.

[19] M. C. Jeruchim, P. Balaban, and K. S. Shanmugan, *Simulation of communication systems: modeling, methodology, and techniques*. Springer, 2000.

[20] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *Electron. Letters*, vol. 27, no. 23, pp. 2145–2146, 1991.