Online Spectrum Cartography via Quantized Measurements

Daniel Romero
Dept. of Signal Theory & Comm.
University of Vigo, Spain
Email: dromero@gts.uvigo.es

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 interpolation task is formulated as a nonparametric regression problem in a reproducing kernel Hilbert space (RKHS) of vector-valued functions, and solved using a stochastic gradient descent iteration. 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.

The aim of the present work is to address these limitations through an estimator-interpolator approach that can afford low-cost, low-power sensing architectures. 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. A BEM is employed in the frequency domain to exploit spectral prior information and to interpret the spectrum contents in terms of the spatial power distribution of each channel. Spatial interpolation is accomplished using an online nonparametric regression technique that hinges on the framework of RKHSs of vector-valued functions [10], where the coefficients of the frequency basis functions are estimated as spatial vector fields.

This work was partially funded by the Spanish Government and the European Regional Development Fund (ERDF) under projects TACTICA, COMONSENS (CSD2008-00010) and COMPASS (TEC2013-47020-C2-1-R); by the Galician Government and ERDF under projects "Consolidation of Research Units" (GRC2013/009), REdTEIC (R2014/037) and AtlantTIC; by FPU Grant AP2010-0149; and by NSF grants 1202135, 1247885, and 1343248.

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. The main conclusions are summarized Sec. V.

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 [11], [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].

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]–[15]. 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. This operation can be represented as

$$\tilde{\boldsymbol{r}}_{\boldsymbol{x}}[b] = \tilde{\boldsymbol{G}}\boldsymbol{r}_{\boldsymbol{x}}[b] \tag{3}$$

where $\tilde{\boldsymbol{G}} \in \mathbb{C}^{\tilde{L}\times L}$ is the compression matrix. In order to guarantee the identifiability of the $l_m(\boldsymbol{x})$'s, this matrix must satisfy the criteria in [16].

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

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

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

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

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 $\phi_{\boldsymbol{x},m} := (B\tilde{L})^{-1}\,\mathrm{Tr}\left(\boldsymbol{G}\boldsymbol{\Sigma}_m\boldsymbol{G}^T\right)$, $\boldsymbol{l}(\boldsymbol{x}) := [l_1(\boldsymbol{x}), \ldots, l_M(\boldsymbol{x})]^T$, and plugging (5) into (4) yield

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

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 compressed observation of $\boldsymbol{l}(\boldsymbol{x})$ [16].

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

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

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 $y_{\boldsymbol{x}} := (2q_{\boldsymbol{x}}+1)\epsilon$,

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

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., [17])

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

Here, $\mathcal{S}$ is the space of functions, assumed to be a RKHS of vector-valued functions from $\mathbb{R}^d$ to $\mathbb{R}^M$ [10], and $\lambda > 0$ is a constant adjusted to attain the desired trade-off between fit and smoothness. The penalty J is chosen to be $J(\boldsymbol{l}) := \|\boldsymbol{l}\|_{\mathcal{S}}^2$, where $\|\boldsymbol{l}\|_{\mathcal{S}}$ is the norm induced by the inner product of $\mathcal{S}$.

The empirical risk R_{emp} is chosen to linearly penalize deviations from (8), using the so-called ϵ-insensitive loss $u_\epsilon(y) := \max(0, |y| - \epsilon)$ as in [18]

$$R_{\text{emp}}(\boldsymbol{l}; \{(\boldsymbol{x}, \boldsymbol{\phi}_{\boldsymbol{x}}, y_{\boldsymbol{x}})\}_{\boldsymbol{x}\in\mathcal{X}}) := \frac{1}{N}\sum_{\boldsymbol{x}\in\mathcal{X}} u_\epsilon(y_{\boldsymbol{x}} - \boldsymbol{\phi}_{\boldsymbol{x}}^T\boldsymbol{l}(\boldsymbol{x})). \tag{10}$$

This is a convex surrogate for the number of measurement errors, in the same way as the ℓ_1-norm is substituted for the ℓ_0-norm in the sparse regression literature. In other words, R_{emp} captures the sparsity present in the measurement errors.

The RKHS 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 [10]. A simple construction is $\boldsymbol{K}(\boldsymbol{z}, \boldsymbol{x}) = \mathrm{diag}\left\{k^{(1)}(\boldsymbol{z}, \boldsymbol{x}), \ldots, k^{(M)}(\boldsymbol{z}, \boldsymbol{x})\right\}$, where $k^{(m)}(\boldsymbol{z}, \boldsymbol{x})$ are valid scalar kernels.

We developed a *batch* algorithm to solve (9) in [19]. However, the cost that batch approaches incur can grow prohibitively as the number of measurements increases. This motivates well the *online* algorithm here, which can incrementally update the estimate $\boldsymbol{l}$ as new measurements arrive at the FC. As an additional advantage, online algorithms can track slow temporal variations of the fields of interest.

An elegant approach to derive online algorithms for kernel machines is based on stochastic gradient descent in the function space [20], [21]. First, one can define the instantaneous regularized error as

$$R_{\text{inst}}(\boldsymbol{l}, \boldsymbol{\phi}, \boldsymbol{x}, y) := u_\epsilon(y - \boldsymbol{\phi}^T\boldsymbol{l}(\boldsymbol{x})) + \lambda\|\boldsymbol{l}\|_{\mathcal{S}}^2. \tag{11}$$

Note that the objective of (9) is just the sample average of R_{inst}. Suppose that per 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.) The update rule is then given by

$$\boldsymbol{l}^{(t+1)} = \boldsymbol{l}^{(t)} - \kappa_t\partial_{\boldsymbol{l}}R_{\text{inst}}(\boldsymbol{l}^{(t)}, \boldsymbol{\phi}_{\boldsymbol{x}_t}, \boldsymbol{x}_t, y_{\boldsymbol{x}_t}) \tag{12}$$

where $\boldsymbol{l}^{(t)}$ is the estimate at time t, $\kappa_t > 0$ is the learning rate, and $\partial_{\boldsymbol{l}}$ is the subgradient operator with respect to $\boldsymbol{l}$. From [21, eq. (2)]

$$\begin{aligned}\partial_{\boldsymbol{l}} R_{\text{inst}}(\boldsymbol{l}, \boldsymbol{\phi}, \boldsymbol{x}, y) &= k_{\boldsymbol{x}} \partial_{\boldsymbol{z}} u_\epsilon(y - \boldsymbol{\phi}^T \boldsymbol{z})|_{\boldsymbol{z}=\boldsymbol{l}} + 2\lambda \boldsymbol{l} \\ &= -u'_\epsilon(y - \boldsymbol{\phi}^T \boldsymbol{l}) \cdot k_{\boldsymbol{x}} \boldsymbol{\phi} + 2\lambda \boldsymbol{l} \end{aligned} \quad (13)$$

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 $u'_\epsilon(z) = \frac{1}{2}(\text{sgn}(z-\epsilon) + \text{sgn}(z+\epsilon))$. Substituting (13) in (12) yields

$$\boldsymbol{l}^{(t+1)} = (1 - 2\kappa_t \lambda)\boldsymbol{l}^{(t)} + \kappa_t u'_\epsilon[y_{\boldsymbol{x}_t} - \boldsymbol{\phi}_{\boldsymbol{x}_t}^T \boldsymbol{l}^{(t)}(\boldsymbol{x}_t)] k_{\boldsymbol{x}_t} \boldsymbol{\phi}_{\boldsymbol{x}_t} \, .$$

Invoking the Representer Theorem [10, Thm. 5], one can see that, after one measurement is received from each sensor, the (batch) solution of (9) can be written as

$$\boldsymbol{l}^{(t)} = \sum_{\boldsymbol{x} \in \mathcal{X}} k_{\boldsymbol{x}} \boldsymbol{c}_{\boldsymbol{x}}^{(t)} \quad \text{for some } \boldsymbol{c}_{\boldsymbol{x}}^{(t)} \in \mathbb{R}^M. \quad (14)$$

Thus, due to the linear independence of the kernel operators $\{k_{\boldsymbol{x}}\}_{\boldsymbol{x} \in \mathcal{X}}$, the updating rule (12) to be applied every time an observation $(y_{\boldsymbol{z}}, \boldsymbol{\phi}_{\boldsymbol{z}})$ is received from the sensor at location $\boldsymbol{z} \in \mathcal{X}$ can be written as

$$\boldsymbol{c}_{\boldsymbol{x}}^{(t+1)} = (1 - 2\kappa_t \lambda)\boldsymbol{c}_{\boldsymbol{x}}^{(t)} \text{ for } \boldsymbol{x} \in \mathcal{X}, \ \boldsymbol{x} \neq \boldsymbol{z} \quad (15)$$

$$\boldsymbol{c}_{\boldsymbol{z}}^{(t+1)} = (1 - 2\kappa_t \lambda)\boldsymbol{c}_{\boldsymbol{z}}^{(t)} + \kappa_t u'_\epsilon(y_{\boldsymbol{z}} - \boldsymbol{\phi}_{\boldsymbol{z}}^T \boldsymbol{l}^{(t)}(\boldsymbol{z}))\boldsymbol{\phi}_{\boldsymbol{z}} \quad (16)$$

with the initialization $\boldsymbol{c}_{\boldsymbol{x}}^{(1)} = \boldsymbol{0}$.

The convergence of this algorithm can be established along the lines in [21, Thm. 1]:

Theorem 1: If $\lambda_{\max}(\boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x})) < \mu < \infty$ for all $\boldsymbol{x}$, $||\boldsymbol{\phi}_{\boldsymbol{x}_t}||_2 \leq S$ for all t, and $\kappa_t := \kappa t^{-1/2}$ with $\kappa\lambda < 1$, then the sequence $\{\boldsymbol{l}^{(t)}\}_t$ satisfies

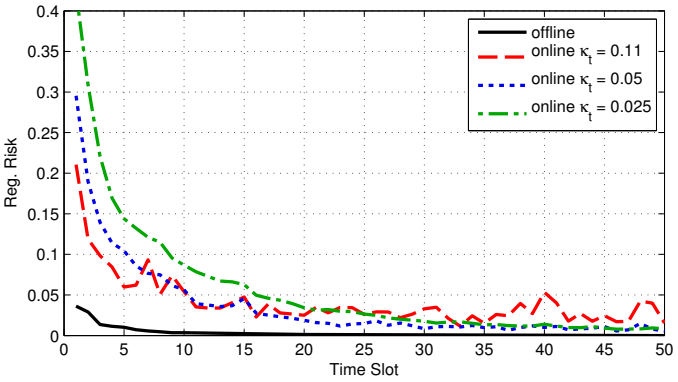
$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^{T} R_{\text{inst}}(\boldsymbol{l}^{(t)}, \boldsymbol{\phi}_{\boldsymbol{x}_t}, \boldsymbol{x}_t, y_{\boldsymbol{x}_t}) \leq \\ &\inf_{\boldsymbol{l} \in \mathcal{S}} \left[\frac{1}{T}\sum_{t=1}^{T} R_{\text{inst}}(\boldsymbol{l}, \boldsymbol{\phi}_{\boldsymbol{x}_t}, \boldsymbol{x}_t, y_{\boldsymbol{x}_t})\right] + \frac{\alpha}{\sqrt{T}} + \frac{\beta}{T}\end{aligned} \quad (17)$$

where $\beta = \mu^2 S^2/(8\lambda^2\kappa)$ and $\alpha = 4(\mu^2 S^2 \kappa + \beta)$.

Proof: The proof of [21, Thm. 1] generalizes the proof of [20, Thm. 4] to the vector-valued function case. To prove Theorem 1, one can follow the same approach with two small differences. First, the proof in [21] involves a gradient, rather than a *subgradient* in (12). It can be seen that this subtlety requires no modification since the vector version of the argument used to establish [20, eq. (49)] holds for both gradients and subgradients, and this is the only part in the proof that requires the properties of gradients/subgradients.

Secondly, in the proof in [21], the fitting term of the instantaneous regularized error (in our case the first term of (11)) seen as a function of the vector $\boldsymbol{l}(\boldsymbol{x})$, say $f(\boldsymbol{v}) = u_\epsilon(y - \boldsymbol{\phi}^T \boldsymbol{v})$, is not dependent on t. In our case, this dependence exists through $\boldsymbol{\phi}_{\boldsymbol{x}_t}$ and $y_{\boldsymbol{x}_t}$. However, it can be seen that the proof of [21] can easily accommodate this situation, provided that the functions share a common Lipschitz constant. That is, for $f_t(\boldsymbol{v}) = u_\epsilon(y_{\boldsymbol{x}_t} - \boldsymbol{\phi}_{\boldsymbol{x}_t}^T \boldsymbol{v})$, there exists a c such that

$$|f_t(\boldsymbol{v}_1) - f_t(\boldsymbol{v}_2)| \leq c||\boldsymbol{v}_1 - \boldsymbol{v}_2||_2 \quad (18)$$



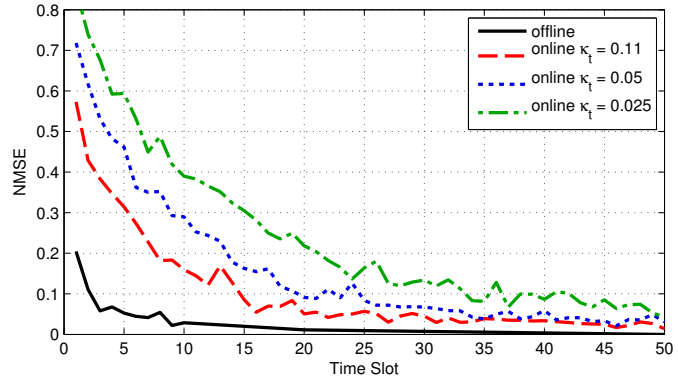


Fig. 1: Influence of the step size in the convergence of the online algorithm.

for all $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathbb{R}^M$ and $t = 1, \ldots, T$. In our case, it can be noted that

$$\begin{aligned}|f_t(\boldsymbol{v}_1) - f_t(\boldsymbol{v}_2)| &= |u_\epsilon(y_{\boldsymbol{x}_t} - \boldsymbol{\phi}_{\boldsymbol{x}_t}^T \boldsymbol{v}_1) - u_\epsilon(y_{\boldsymbol{x}_t} - \boldsymbol{\phi}_{\boldsymbol{x}_t}^T \boldsymbol{v}_2)| \\ &= |\boldsymbol{\phi}_{\boldsymbol{x}_t}^T(\boldsymbol{v}_2 - \boldsymbol{v}_1)| \leq ||\boldsymbol{\phi}_{\boldsymbol{x}_t}||_2 ||\boldsymbol{v}_1 - \boldsymbol{v}_2||_2 \end{aligned} \quad (19)$$

where the second equality is due to the fact that $u_\epsilon(z)$ is Lispchitz function of z with constant 1. Thus, (18) holds for any $c \geq \max_{1 \leq t \leq T} ||\boldsymbol{\phi}_{\boldsymbol{x}_t}||_2$. ■

In summary, the averaged instantaneous error from the online algorithm converges to the to the regularized empirical error of the batch solution.

IV. Numerical Tests

In our preliminary tests, the performance of the proposed algorithm is illustrated when $\mathcal{R} = [0, 1]$ (hence $d = 1$). The propagation model is given by the standard inverse polynomial law $l_m(\boldsymbol{x}) = \delta A_m/(\delta + \|\boldsymbol{x} - \boldsymbol{z}_m\|^\gamma)$, $m = 1, 2, \ldots, M-1$, where $\delta = 10^{-3}$ is a small constant ensuring that the denominator does not vanish, $\gamma = 3$ is the pathloss exponent, and the parameters A_m and $\boldsymbol{z}_m$ representing the transmit-power and the source location, respectively, are given by $\boldsymbol{z}_1 = 0.1$, $\boldsymbol{z}_2 = 0.2$, $\boldsymbol{z}_3 = 0.4$, $\boldsymbol{z}_4 = 0.8$, $A_1 = 0.8$, $A_2 = 0.9$, $A_3 = 0.8$, and $A_4 = 0.7$. The noise power was set to $l_M(\boldsymbol{x}) = 0.75$.

Our test used $N = 30$ sensors deployed uniformly at random, reporting one measurement to the FC per time

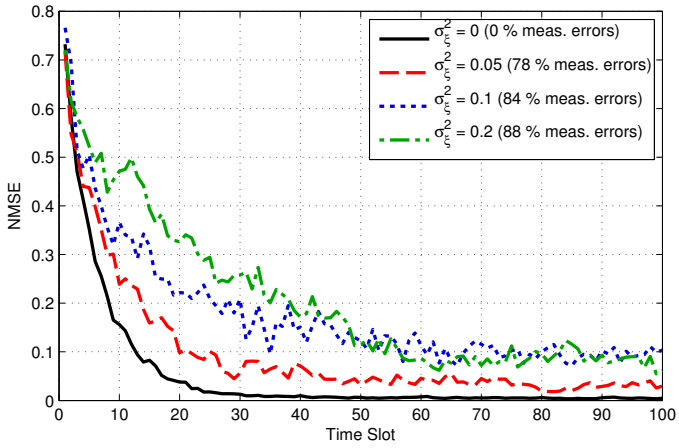


Fig. 2: Influence of the measurement errors in the convergence of the online algorithm.

slot. Uniform quantization to 4 bits per measurement was implemented, where ϵ was set so that the probability of clipping $\mathbb{P}\{\eta_{\boldsymbol{x},p} > r_R\}$ was approximately 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 $\sigma_i = 0.1$, and the regularization constant was set to $\lambda = 10^{-6}$.

Fig. 1 depicts the performance of the online algorithm for several learning rates κ_t along with the offline (batch) algorithm from [19], which was executed at each time slot with all the data received up to that time slot. No measurement errors were introduced. The top panel shows the regularized empirical risks (evaluated at each time slot using the whole set of observations) for different learning rates κ_t. As usually occurs in gradient methods with constant step size, a larger value of κ_t results in a faster convergence, but the residual error is larger. The bottom panel depicts the time evolution of the normalized mean-square error: NMSE $= \mathbb{E}\left\{\|\boldsymbol{l}(\boldsymbol{x}) - \hat{\boldsymbol{l}}(\boldsymbol{x})\|_2^2\right\}/\mathbb{E}\left\{\|\boldsymbol{l}(\boldsymbol{x})\|_2^2\right\}$, where the expectation was taken with respect to uniformly distributed $\boldsymbol{x}$.

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, one can model $\hat{\eta}_{\boldsymbol{x}}$ as

$$\hat{\eta}_{\boldsymbol{x}} = \eta_{\boldsymbol{x}} + \xi_{\boldsymbol{x}} \tag{20}$$

where $\xi_{\boldsymbol{x}}$ is a normally distributed random variable with zero mean and variance σ_ξ^2 that captures estimation errors. Fig. 2 represents the evolution of the normalized mean-square error for several values of σ_ξ^2. It is seen that a larger σ_ξ^2 results in a slower convergence and a larger residual error. However, even for large values of σ_ξ^2 (e.g. for $\sigma_\xi^2 = 0.2$, 88 % of the measurements contain an error), the algorithm is capable of converging.

V. Conclusions

An online stochastic gradient descent method is proposed to perform nonparametric regression in a RKHS of vector-valued functions and applied to the problem of power spectrum map estimation. A convergence result and a simulation experiment were provided to illustrate the performance of the proposed algorithm.

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. Proc. 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. of the IEEE 19th Intl. Symp. Personal, Indoor Mobile Radio Commun. (PIMRC)*, 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. Proc.*, 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. Intl. Conf. Commun. (ICC)*, 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.*, 2015, to appear.

[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. Proc.*, 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. Proc.*, 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. Proc.*, vol. 61, no. 10, pp. 2693–2703, May 2013.

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

[11] G. Vázquez-Vilar and R. López-Valcarce, "Spectrum sensing exploiting guard bands and weak channels," *IEEE Trans. Sig. Proc.*, vol. 59, no. 12, pp. 6045–6057, 2011.

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

[13] S. Becker, "Practical compressed sensing: Modern data acquisition and signal processing," Ph.D. dissertation, California Institute of Technology, 2011.

[14] 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. Info. Theory*, vol. 56, no. 1, pp. 520–544, Jan. 2010.

[15] J. Yoo, S. Becker, M. Monge, M. Loh, E. Candes, and A. Emami-Neyestanak, "Design and implementation of a fully integrated compressed-sensing signal acquisition system," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2012, pp. 5325–5328.

[16] D. Romero, R. López-Valcarce, and G. Leus, "Compression limits for random vectors with linearly parameterized second-order statistics," *IEEE Trans. Info. Theory*, vol. 61, no. 3, pp. 1–16, Mar. 2015.

[17] B. Schölkopf and A. J. Smola, *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, 2001.

[18] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004.

[19] D. Romero, S.-J. Kim, R. López-Valcarce, and G. B. Giannakis, "Spectrum cartography using quantized observations," *IEEE Int. Conf. Acoust., Speech, Signal Process. (Accepted)*, 2015.

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

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