SPECTRUM CARTOGRAPHY USING QUANTIZED OBSERVATIONS

Daniel Romero[1]*, Seung-Jun Kim*[2]*, Roberto López-Valcarce*[1] *and Georgios B. Giannakis*[3]

[1]Dept. of Signal Theory and Communications, University of Vigo, Spain, {dromero,valcarce}@gts.uvigo.es.
[2]Dept. of Computer Sci. & Electrical Eng., University of Maryland, Baltimore County, USA, sjkim@umbc.edu.
[3]Dept. of ECE and Digital Technology Center, University of Minnesota, USA, georgios@umn.edu.

ABSTRACT

This work proposes a spectrum cartography algorithm used for learning the power spectrum distribution over a wide frequency band across a given geographic area. Motivated by low-complexity sensing hardware and stringent communication constraints, compressed and quantized measurements are considered. Setting out from a nonparametric regression framework, it is shown that a sensible approach leads to a support vector machine formulation. The simulated tests verify that accurate spectrum maps can be constructed using a simple sensing architecture with significant savings in the feedback.

1. INTRODUCTION

The importance of wireless spectrum situational awareness is critically recognized in the cognitive radio (CR) paradigm, where the goal is to improve the overall spectral efficiency through agile adaptation to spectrum availability via dynamic spectrum access (DSA) [1]. Acquiring information regarding how the radiated power from the interferers distributes across space and frequency is instrumental for ensuring the quality of service of CRs, as well as controlling interference to the incumbent primary systems. Medium access and network optimization may also benefit profoundly from such information.

Acquiring the spectrum landscape, termed as *spectrum cartography*, has received significant attention in the CR literature [2–8]. Existing techniques rely on the measurements reported by a network of sensors, which may be the CRs themselves, and apply regression procedures to estimate the signal power at arbitrary locations. Examples of such techniques include kriging [2], sparse regression [3–5], semiparametric regression [6], and dictionary learning [7,8]. Although some of these techniques do not take into account the frequency dimension, estimating power over both space and frequency can obtain numerous benefits since the frequency dimension allows to incorporate prior information and is capable of estimating metrics such as the signal-to-noise ratio [5]. Other intimately related techniques include [9, 10], where the goal is to interpolate the channel gains between arbitrary locations.

Developing a spectrum sensing strategy that takes into account location and propagation information may significantly improve the estimation performance, which consequently enhances the spectral efficiency of CRs by incorporating an additional degree of freedom to exploit through aggressive spatial reuse [11,12]. For this reason, spectrum cartography techniques provide an important improvement over existing cooperative spectrum sensing schemes, which traditionally neglect this information [13–15].

Spatial regression of power spectral density (PSD) was obtained via parametric and nonparametric techniques based on periodogram measurements [5, 6]. However, obtaining and reporting such statistics may be too complex and bandwidth-hungry for distributed sensors with limited hardware and resources. Indeed, acquiring periodogram estimates seems quite demanding on sensing hardware since the impinging waveforms must be sampled at the Nyquist rate, and long averaging windows have to be implemented in the frequency domain. Furthermore, the feedback of the high-resolution periodogram measurements through the control channels shared by a large number of sensor nodes may limit the speed of acquisition, hindering real-time update of PSD maps, which is essential for tracking time-varying interference spectrum. For these reasons, compressed spectrum measurements and low-rate communication protocols are well motivated [15, 16].

This work aims at alleviating those limitations by proposing a spectrum cartography algorithm where the map estimate is constructed from highly quantized versions of compressed measurements of wideband signals. Simple wideband converters such as the random filters [15, 16] or analog-to-information converters [17, 18] are considered. In particular, the latter allow sampling rates well below the Nyquist rate, which is especially convenient in wideband scenarios. On the other hand, low-rate feedback, where only a few bits are used to represent a measurement, allows efficient control channel implementations in multipoint-to-point medium access topologies, such as the one used in the IEEE 802.22 standard for DSA [19].

This work also establishes an interesting connection between the problem of spectrum cartography and machine learning. Motivated by the robustness considerations for quantized measurements, a natural nonparametric regression formulation for spectrum map construction transitions smoothly to a support vector machine (SVM) formulation. This perspective allows to exploit established efficient numerical methods and provides theoretical understanding of the resulting algorithms, since SVMs are universal approximators based on solid statistical learning foundations [20, 21].

Spatial regression of interference power was considered without taking into account the frequency domain in [2–4]. A map capturing both space and frequency domains was constructed in [5, 6], but the issue of sensing and communication complexity was not investigated. A simple sensor architecture was employed to construct a spectral distribution of interference in [15], but the spatial dimension was not incorporated. Our work addresses spatial and frequency domain spectrum cartography with minimal sensing and communication overhead. In order to capture the prior information while maintaining flexibility, the spatial field to be learned is modeled parametrically in the frequency dimension while nonparametric techniques are employed in space.

The rest of the paper is organized as follows. In Section 2, the system model is described. The spectrum cartography problem and the corresponding learning algorithm are explained in Section 3. Results of numerical tests are presented in Section 4, whereas some conclusions are provided in Section 5.

D. Romero and R. López-Valcarce were supported by the European Regional Development Fund (ERDF) and the Spanish Government (TEC2010-21245-C02-02/TCM DYNACS, CONSOLIDER-INGENIO 2010 CSD2008-00010 COMONSENS), FPU Grant AP2010-0149, and the Galician Regional Government (CN 2012/260 AtlantTIC). S.-J. Kim and G. B. Giannakis were supported in part by NSF grants ECCS 1002180, ECCS 1202135 and AST 1247885.

2. MODEL DESCRIPTION

This section introduces a model for the PSD map, which hinges on a frequency-domain basis expansion model. Subsequently, the quantized observation model is put forth.

2.1. PSD map model

Suppose that $M-1$ interference sources, which may correspond to the transmitters of the primary system, are located in a geographical region. The m-th source transmits a signal $\sqrt{\gamma^{(m)}}s^{(m)}(t)$, where $\gamma^{(m)} > 0$ represents the transmit-power and $s^{(m)}(t)$ a (possibly complex) wide-sense stationary (WSS) signal normalized such that $\mathrm{E}\{|s^{(m)}(t)|^2\} = 1,\ \forall t$. The PSD of $s^{(m)}(t)$ is denoted as $\phi^{(m)}(f)$ and and satisfies $\int_0^1 \phi^{(m)}(f)df = 1$ due to the normalization. For simplicity, we assume that $\phi^{(m)}(f)$ is known, which is well motivated in many communication scenarios where the primary system obeys a transmission standard (e.g. DVB or ATSC in TV bands) and specific spectrum regulations, since in those cases bandwidths, carrier frequencies, transmission masks, roll-off factors, and so forth, are publicly available. If $\phi^{(m)}(f)$ is not known, the same scheme presented in this paper can be applied by introducing a basis expansion in the frequency domain (see, e.g., [5,6]).

The signal at position $\boldsymbol{x} \in \mathbb{R}^2$ resulting from the contributions of the $M-1$ uncorrelated sources can be expressed as

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

where $r_{\boldsymbol{x}}^{(m)}(t),\ m = 1,\ldots,M-1$, is the component propagating from the m-th transmitter and $r_{\boldsymbol{x}}^{(M)}(t)$ is noise with normalized PSD $\phi^{(M)}(f)$. Assuming frequency-flat channels, the PSD of $r_{\boldsymbol{x}}(t)$ can be written as

$$\psi_{\boldsymbol{x}}(f) = \sum_{m=1}^{M} l_{\boldsymbol{x}}^{(m)} \phi^{(m)}(f) \tag{2}$$

where $l_{\boldsymbol{x}}^{(M)}$ is the noise power at sensor $\boldsymbol{x}$, and the coefficients $l_{\boldsymbol{x}}^{(m)}$, $m = 1,\ldots,M-1$, subsume the transmit-power $\gamma^{(m)}$ and the propagation effects between the m-th source and the position $\boldsymbol{x}$. As explained in [5, Sec. II], this representation also applies to frequency selective cases where the impulse response of the channel is stationary and uncorrelated along the lag dimension.

We observe from (2) that the variable $l_{\boldsymbol{x}}^{(m)}$, $m = 1,\ldots,M-1$, represents the power of the signal of the m-th transmitter at the point $\boldsymbol{x}$. The function $\psi_{\boldsymbol{x}}(f)$ may be thought of as a map that represents the distribution of power over space and frequency and our goal is to estimate it. Since the *basis functions* $\phi^{(m)}(f)$ are known, estimating $\psi_{\boldsymbol{x}}(f)$ is tantamount to estimating $l_{\boldsymbol{x}}^{(m)}$ for all m, which will be seen as M functions of the spatial coordinate $\boldsymbol{x}$.

2.2. Observation model

The observation model proposed here captures a couple of simple sensing architectures. One of them, chosen here for the ease of exposition, is based on pseudo-random filters [15]. An alternative implementation performs compressive acquisition via the recently proposed analog-to-information converters [17, 18], which allow the acquisition of wideband signals with low power consumption and minimal hardware costs [22].

Suppose that a sensor at position $\boldsymbol{x}$ passes $r_{\boldsymbol{x}}(t)$ through a linear time-invariant (LTI) filter with impulse response $g_{\boldsymbol{x}}(t)$ and measures the power of the output. The theoretical value of this power is given by

$$\eta_{\boldsymbol{x}} := \mathbb{E}\left\{|g_{\boldsymbol{x}}(t) \star r_{\boldsymbol{x}}(t)|^2\right\} = \int_0^1 |G_{\boldsymbol{x}}(f)|^2 \psi_{\boldsymbol{x}}(f)df \tag{3}$$

where $\star$ denotes convolution, and $G_{\boldsymbol{x}}(f)$ is the Fourier transform of $g_{\boldsymbol{x}}(t)$. From (2), it follows that

$$\eta_{\boldsymbol{x}} = \sum_{m=1}^{M} l_{\boldsymbol{x}}^{(m)} \int_0^1 |G_{\boldsymbol{x}}(f)|^2 \phi^{(m)}(f)df. \tag{4}$$

By defining $\phi_{\boldsymbol{x}}^{(m)} := \int_0^1 |G_{\boldsymbol{x}}(f)|^2 \phi^{(m)}(f)df$ and forming the vectors $\boldsymbol{\phi}_{\boldsymbol{x}} := [\phi_{\boldsymbol{x}}^{(1)},\ldots,\phi_{\boldsymbol{x}}^{(M)}]^T$ and $\boldsymbol{l}(\boldsymbol{x}) := [l_{\boldsymbol{x}}^{(1)},\ldots,l_{\boldsymbol{x}}^{(M)}]^T$, where the notation emphasizes that $\boldsymbol{l}(\boldsymbol{x})$ is a function of the spatial coordinate $\boldsymbol{x}$, we can rewrite $\eta_{\boldsymbol{x}}$ in (4) simply as

$$\eta_{\boldsymbol{x}} = \boldsymbol{\phi}_{\boldsymbol{x}}^T \boldsymbol{l}(\boldsymbol{x}) \tag{5}$$

where ^{T} denotes transposition.

The sensor computes an estimate of $\eta_{\boldsymbol{x}}$, denoted as $\hat{\eta}_{\boldsymbol{x}}$, and uniformly quantizes it according to a map $Q(\eta) = \lfloor \eta/(2\epsilon) \rfloor \in \mathbb{Z}$, where 2ϵ is the quantization step, and the result $\hat{q}_{\boldsymbol{x}} = Q(\hat{\eta}_{\boldsymbol{x}})$ is sent to the fusion center (FC) through a control channel. Depending on the quality of the estimate, either $\hat{q}_{\boldsymbol{x}} = q_{\boldsymbol{x}} := Q(\eta_{\boldsymbol{x}})$ and $\hat{q}_{\boldsymbol{x}} \neq q_{\boldsymbol{x}}$. The latter case is referred to as a *measurement error*.

For simplicity, the exposition made use of analog processing, but the operations can be implemented in digital as well. The filters $g_{\boldsymbol{x}}(t)$ can then be implemented using pseudo-random sequences with different random number seeds, so that multiple observations with different measurement vectors $\boldsymbol{\phi}_{\boldsymbol{x}}$ are reported.

3. CARTOGRAPHY LEARNING

The cartography problem can now be posed as estimating the spatial vector field $\boldsymbol{l} : \mathbb{R}^2 \to \mathbb{R}^M$ based on the knowledge of the set of sensor locations $\mathcal{X} = \{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N\}$, the measurement vectors $\{\boldsymbol{\phi}_{\boldsymbol{x}_1},\ldots,\boldsymbol{\phi}_{\boldsymbol{x}_N}\}$, and the quantized observations $\{\hat{q}_{\boldsymbol{x}_1},\ldots,\hat{q}_{\boldsymbol{x}_N}\}$. Note that this notation also accommodates the case where a sensor reports multiple measurements to the FC. In that case, the location of that sensor will be contained in $\mathcal{X}$ as many times as measurements it obtains. Throughout, the symbol $\boldsymbol{l}$ will represent the function itself, whereas $\boldsymbol{l}(\boldsymbol{x}) \in \mathbb{R}^M$ will refer to the result of evaluating $\boldsymbol{l}$ at point $\boldsymbol{x}$.

3.1. Nonparametric regression formulation

Let us first consider the case where no measurement errors have occurred. The presence of measurement errors will be addressed shortly. Upon receiving a measurement $\hat{q}_{\boldsymbol{x}} = q_{\boldsymbol{x}}$, the FC learns that

$$2\epsilon q_{\boldsymbol{x}} \leq \boldsymbol{\phi}_{\boldsymbol{x}}^T \boldsymbol{l}(\boldsymbol{x}) < 2\epsilon(q_{\boldsymbol{x}} + 1). \tag{6}$$

A sensible approach to construct the desired PSD map is to search for a smooth function $\boldsymbol{l}$, which is in a certain space of functions $\mathcal{S}$ and consistent with (6) for all $\boldsymbol{x} \in \mathcal{X}$. We assume that this space $\mathcal{S}$ is a *reproducing kernel Hilbert space* (RKHS) of vector-valued functions [23, 24]. This choice allows us to capture the smoothness of the function through the norm induced by the inner product. More specifically, the problem can be cast as

$$\begin{aligned} \underset{\boldsymbol{l}\in\mathcal{S}}{\text{minimize}} \quad & ||\boldsymbol{l}|| \\ \text{s.t.} \quad & 2\epsilon q_{\boldsymbol{x}} \leq \boldsymbol{\phi}_{\boldsymbol{x}}^T \boldsymbol{l}(\boldsymbol{x}) \leq 2\epsilon(q_{\boldsymbol{x}} + 1),\ \forall \boldsymbol{x} \in \mathcal{X}, \end{aligned} \tag{7}$$

which can be rewritten as

$$\begin{aligned} \underset{\boldsymbol{l}\in\mathcal{S}}{\text{minimize}} \quad & ||\boldsymbol{l}|| \\ \text{s.t.} \quad & |y_{\boldsymbol{x}} - \boldsymbol{\phi}_{\boldsymbol{x}}^T \boldsymbol{l}(\boldsymbol{x})| \leq \epsilon,\ \forall \boldsymbol{x} \in \mathcal{X} \end{aligned} \tag{8}$$

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

In the presence of a measurement error at $\boldsymbol{x}$, the true $\boldsymbol{l}$ will not satisfy the constraint $|y_{\boldsymbol{x}} - \boldsymbol{\phi}_{\boldsymbol{x}}^T \boldsymbol{l}(\boldsymbol{x})| \leq \epsilon$. Therefore, one must capitalize on the prior information on $\boldsymbol{l}$, rather than blindly trusting the data. Disciplined approaches to balance this trade-off are complexity control techniques in statistical learning. An example of those is the regularization framework, where an empirical loss augmented by a penalty term is minimized [20, 25]. Based on this approach, a smoothness enforcing term $\|\boldsymbol{l}\|^2$ is added to the empirical risk.

To choose a proper empirical loss function that can cope with measurement errors, it is noted that penalizing outliers in proportion to the deviation from the feasible set is widely appreciated in machine learning, as this yields fewer outliers. Thus, a suitable empirical loss function is the ϵ-insensitive loss, defined as [20, 21]:

$$u_\epsilon(y) = \max(0, |y| - \epsilon). \tag{9}$$

The loss function can be regarded as a convex surrogate of the number of measurement errors, the same way as the ℓ_1-norm-based regularizers are often used as a surrogate for ℓ_0-norm-based ones in sparse regression [20, Sec. 9.3]. Thus, our formulation is recast as

$$\underset{\boldsymbol{l} \in \mathcal{S}}{\text{minimize}} \quad \frac{1}{N} \sum_{\boldsymbol{x} \in \mathcal{X}} u_\epsilon(y_{\boldsymbol{x}} - \boldsymbol{\phi}_{\boldsymbol{x}}^T \boldsymbol{l}(\boldsymbol{x})) + \lambda ||\boldsymbol{l}||^2, \tag{10}$$

where $\lambda > 0$ is a parameter adjusted to attain the desired trade-off between empirical risk minimization and function smoothness.

3.2. Finite-dimensional parameterizaton

Solving (10) requires searching over a possibly infinite-dimensional space $\mathcal{S}$. Instrumental in this situation is a representer theorem, which establishes a finite-dimensional parameterization of the search space, converting (10) into a finite-dimensional optimization problem. Here, a representer theorem for vector-valued function spaces must be invoked [23, 24].

Let $\mathcal{S}$ be a RKHS of functions from $\mathbb{R}^2$ to $\mathbb{R}^M$ with an inner product $\langle \cdot, \cdot \rangle$ [26]. For any $\boldsymbol{l} \in \mathcal{S}$, $\boldsymbol{\phi}^T \boldsymbol{l}(\boldsymbol{x})$ can be represented as [23]

$$\boldsymbol{\phi}^T \boldsymbol{l}(\boldsymbol{x}) = \langle \boldsymbol{l}, k_{\boldsymbol{x}} \boldsymbol{\phi} \rangle \tag{11}$$

where $k_{\boldsymbol{x}} : \mathbb{R}^M \to \mathcal{S}$ is a linear operator associated with $\boldsymbol{x} \in \mathbb{R}^2$. If $\mathcal{S} := \times_{m=1}^M \bar{\mathcal{S}}_m$, where $\times$ represents the Cartesian product and $\bar{\mathcal{S}}_m$ a RKHS of functions from $\mathbb{R}^2$ to $\mathbb{R}$, containing $l^{(m)}$, the linearity of $k_{\boldsymbol{x}}$ implies that it can be regarded as a matrix

$$k_{\boldsymbol{x}} = \begin{bmatrix} k_{\boldsymbol{x}}^{(1,1)} & \dots & k_{\boldsymbol{x}}^{(1,M)} \\ \vdots & \ddots & \vdots \\ k_{\boldsymbol{x}}^{(M,1)} & \dots & k_{\boldsymbol{x}}^{(M,M)} \end{bmatrix} \tag{12}$$

where $k_{\boldsymbol{x}}^{(m,m')} \in \bar{\mathcal{S}}_m \ \forall m$ are set to capture the prior knowledge on the wireless propagation laws in our context. The *reproducing kernel* K is the result of evaluating $k_{\boldsymbol{x}}$, seen as a function on $\mathbb{R}^2$, at a point $\boldsymbol{z} \in \mathbb{R}^2$, that is, $K(\boldsymbol{z}, \boldsymbol{x}) := k_{\boldsymbol{x}}(\boldsymbol{z}) \in \mathbb{R}^{M \times M}$. Due to the properties of the RKHS, we have $K(\boldsymbol{z}, \boldsymbol{x}) = K(\boldsymbol{x}, \boldsymbol{z})^T$ and $K(\boldsymbol{x}, \boldsymbol{x})$ is a positive semi-definite matrix. Moreover, K satisfies

$$\boldsymbol{\phi}_1^T K(\boldsymbol{z}, \boldsymbol{x}) \boldsymbol{\phi}_2 = \boldsymbol{\phi}_1^T (k_{\boldsymbol{x}} \boldsymbol{\phi}_2)(\boldsymbol{z}) = \langle k_{\boldsymbol{x}} \boldsymbol{\phi}_2, k_{\boldsymbol{z}} \boldsymbol{\phi}_1 \rangle \tag{13}$$

which follows from (11) by noting that $k_{\boldsymbol{x}} \boldsymbol{\phi}_2$ is itself an element of $\mathcal{S}$. By virtue of the representer theorem [23, Theorem 5], for $\lambda > 0$, the solution to (10) can be written as

$$\boldsymbol{l} = \sum_{\boldsymbol{x} \in \mathcal{X}} k_{\boldsymbol{x}} \boldsymbol{c}_{\boldsymbol{x}} \tag{14}$$

for some vectors $\{\boldsymbol{c}_{\boldsymbol{x}} \in \mathbb{R}^M\}_{\boldsymbol{x} \in \mathcal{X}}$.

Now, (10) can be written in terms of $\{\boldsymbol{c}_{\boldsymbol{x}} \in \mathbb{R}^M\}_{\boldsymbol{x} \in \mathcal{X}}$. To see this, first note that

$$\boldsymbol{\phi}^T \boldsymbol{l}(\boldsymbol{x}) = \sum_{\boldsymbol{x}' \in \mathcal{X}} \langle k_{\boldsymbol{x}'} \boldsymbol{c}_{\boldsymbol{x}'}, k_{\boldsymbol{x}} \boldsymbol{\phi} \rangle = \sum_{\boldsymbol{x}' \in \mathcal{X}} \boldsymbol{\phi}^T K(\boldsymbol{x}, \boldsymbol{x}') \boldsymbol{c}_{\boldsymbol{x}'} \tag{15}$$

due to the linearity of inner products and (13). Note also that

$$||\boldsymbol{l}||^2 = \langle \boldsymbol{l}, \boldsymbol{l} \rangle = \sum_{\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}} \langle k_{\boldsymbol{x}} \boldsymbol{c}_{\boldsymbol{x}}, k_{\boldsymbol{x}'} \boldsymbol{c}_{\boldsymbol{x}'} \rangle = \sum_{\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}} \boldsymbol{c}_{\boldsymbol{x}'}^T K(\boldsymbol{x}', \boldsymbol{x}) \boldsymbol{c}_{\boldsymbol{x}}. \tag{16}$$

Define the matrix $\boldsymbol{K} \in \mathbb{R}^{MN \times MN}$ as

$$\boldsymbol{K} := \sum_{i,j=1}^N (\boldsymbol{e}_i \boldsymbol{e}_j^T) \otimes K(\boldsymbol{x}_i, \boldsymbol{x}_j) \tag{17}$$

and the vector $\boldsymbol{c} := \sum_{i=1}^N \boldsymbol{e}_i \otimes \boldsymbol{c}_{\boldsymbol{x}_i} \in \mathbb{R}^{MN}$, where $\otimes$ represents the Kronecker product. Then, one can write $||\boldsymbol{l}||^2 = \boldsymbol{c}^T \boldsymbol{K} \boldsymbol{c}$ and $\boldsymbol{\phi}_{\boldsymbol{x}_i}^T \boldsymbol{l}(\boldsymbol{x}_i) = \check{\boldsymbol{\phi}}_{\boldsymbol{x}_i}^T \boldsymbol{K} \boldsymbol{c}$, where $\check{\boldsymbol{\phi}}_{\boldsymbol{x}_i} := \boldsymbol{e}_i \otimes \boldsymbol{\phi}_{\boldsymbol{x}_i}$. Substituting these into (10) yields

$$\underset{\boldsymbol{c} \in \mathbb{R}^{MN}}{\text{minimize}} \quad \frac{1}{N} \sum_{i=1}^N u_\epsilon(y_{\boldsymbol{x}_i} - \check{\boldsymbol{\phi}}_{\boldsymbol{x}_i}^T \boldsymbol{K} \boldsymbol{c}) + \lambda \boldsymbol{c}^T \boldsymbol{K} \boldsymbol{c}. \tag{18}$$

3.3. SVM-based solution

Since K is positive semi-definite [23, eq. (2.5)], problem (18) is convex, and can in fact be formulated as a quadratic program. However, careful inspection reveals that it is also reduced to a standard SVM problem without a bias term through the change of variables $\tilde{\boldsymbol{c}} = \boldsymbol{K}^{\frac{1}{2}} \boldsymbol{c}$, where $\boldsymbol{K}^{\frac{1}{2}}$ represents a symmetric square root of $\boldsymbol{K}$. Therefore, it is known that it can be efficiently solved in the dual domain. For example, sequential minimal optimization (SMO) can be applied for efficient solution [27], although in the present case, further simplification can be effected due to the lack of the bias term [28]. For moderate problem sizes, say, with $MN < 5000$, the interior point method is often used due to its reliability [25, Ch. 10].

By noting that $u_\epsilon(z) = \max(0, z - \epsilon) + \max(0, -z - \epsilon)$ for $\epsilon \geq 0$, and introducing the slack variables $\xi_{\boldsymbol{x}}, \xi_{\boldsymbol{x}}^*$, (18) can be rewritten as

$$\begin{aligned} \underset{\boldsymbol{c} \in \mathbb{R}^{MN}, \xi_{\boldsymbol{x}}, \xi_{\boldsymbol{x}}^*}{\text{minimize}} \quad & \sum_{\boldsymbol{x} \in \mathcal{X}} [\xi_{\boldsymbol{x}} + \xi_{\boldsymbol{x}}^*] + \lambda N \boldsymbol{c}^T \boldsymbol{K} \boldsymbol{c} \\ \text{s.t.} \quad & \xi_{\boldsymbol{x}} \geq y_{\boldsymbol{x}} - \check{\boldsymbol{\phi}}_{\boldsymbol{x}}^T \boldsymbol{K} \boldsymbol{c} - \epsilon \\ & \xi_{\boldsymbol{x}}^* \geq -y_{\boldsymbol{x}} + \check{\boldsymbol{\phi}}_{\boldsymbol{x}}^T \boldsymbol{K} \boldsymbol{c} - \epsilon \\ & \xi_{\boldsymbol{x}}, \xi_{\boldsymbol{x}}^* \geq 0. \end{aligned} \tag{19}$$

The Lagrangian dual of (19) is given by

$$\begin{aligned} \underset{\boldsymbol{\alpha}, \boldsymbol{\alpha}^*}{\text{minimize}} \quad & \frac{1}{4N\lambda} (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)^T (\boldsymbol{I}_N \odot \boldsymbol{\Phi})^T \boldsymbol{K} (\boldsymbol{I}_N \odot \boldsymbol{\Phi}) (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) \\ & - (\boldsymbol{y} - \epsilon \boldsymbol{1}_N)^T \boldsymbol{\alpha} + (\boldsymbol{y} + \epsilon \boldsymbol{1}_N)^T \boldsymbol{\alpha}^* \\ \text{s.t.} \quad & \boldsymbol{0}_N \leq \boldsymbol{\alpha} \leq \boldsymbol{1}_N, \\ & \boldsymbol{0}_N \leq \boldsymbol{\alpha}^* \leq \boldsymbol{1}_N \end{aligned} \tag{20}$$

where $\boldsymbol{y} := [y_{\boldsymbol{x}_1}, \dots, y_{\boldsymbol{x}_N}]^T$, $\boldsymbol{\Phi} := [\boldsymbol{\phi}_{\boldsymbol{x}_1}, \dots, \boldsymbol{\phi}_{\boldsymbol{x}_N}]$ and $\odot$ is the Khatri-Rao product defined as the column-wise application of the Kronecker product. Finally, $\boldsymbol{c}$ can be recovered as

$$\boldsymbol{c} = \frac{1}{2\lambda N} (\boldsymbol{I}_N \odot \boldsymbol{\Phi}) (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*). \tag{21}$$

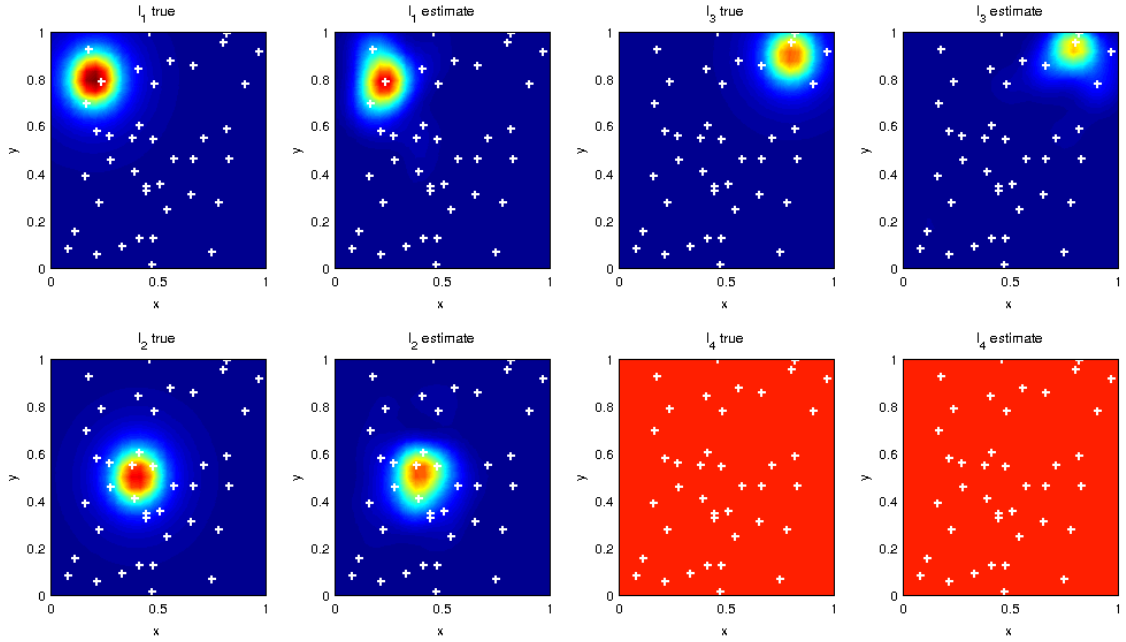

Fig. 1: Comparison of the true and estimated maps for 40 sensors, 6 measurements per sensor, 8 bits per measurement and $\lambda = 10^{-5}$.

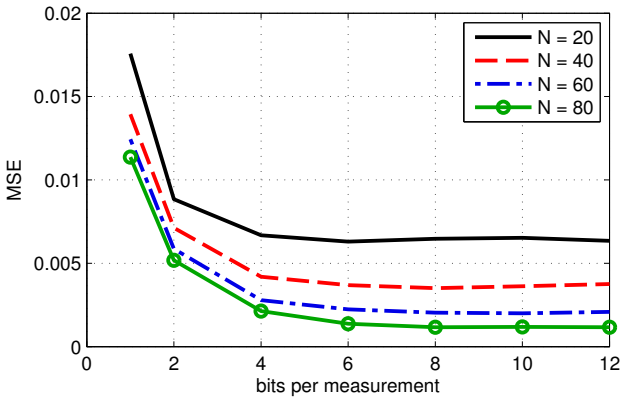

Fig. 2: Mean square error of the estimate with respect to the true map.

4. SIMULATIONS

The operation of the proposed technique is illustrated by means of simple numerical examples. Consider the case of $M = 4$, corresponding to 3 interference sources with additive noise, which is constant across the space. The deployment region is set to $[0, 1] \times [0, 1] \subset \mathbb{R}^2$. The true $\boldsymbol{l}$ postulated to model the attenuation is given by

$$l_{\boldsymbol{x}}^{(m)} = \frac{\delta A_m}{\delta + ||\boldsymbol{x} - \boldsymbol{y}_m||^{\gamma}} \tag{22}$$

where $\delta = 10^{-3}$ is a small constant to ensure that the denominator does not vanish and $\gamma = 3$ is the pathloss exponent. Parameters A_M and $\boldsymbol{y}_M$ denote the transmit-powers and the source locations. In the experiment, they were set as $A_1 = 0.9$, $A_2 = 0.8$, $A_3 = 0.7$, $\boldsymbol{y}_1 = (0.2, 0.8)$, $\boldsymbol{y}_2 = (0.4, 0.5)$ and $\boldsymbol{y}_3 = (0.8, 0.9)$. The noise power was set to $l_{\boldsymbol{x}}^{(M)} = 0.75$ for all $\boldsymbol{x}$.

N sensors were deployed randomly over a 200×200 uniform grid. The measurements were generated as $\hat{q}_{\boldsymbol{x}} = Q(|\eta_{\boldsymbol{x}} + z_{\boldsymbol{x}}|) = Q(|\boldsymbol{\phi}_{\boldsymbol{x}}^T \boldsymbol{l}(\boldsymbol{x}) + z_{\boldsymbol{x}}|)$, where $z_{\boldsymbol{x}} \sim \mathcal{N}(0, 10^{-3}\,\mathrm{Var}\{y_{\boldsymbol{x}}\})$ is the measurement noise. The vectors $\boldsymbol{\phi}_{\boldsymbol{x}}$ were randomly generated with uniformly distributed components. A Gaussian diagonal kernel was adopted, which is defined as

$$K(\boldsymbol{z}, \boldsymbol{x})|_{i,i} = \exp\left\{-\frac{||\boldsymbol{z} - \boldsymbol{x}||^2}{\sigma_i^2}\right\}, \tag{23}$$

where $\sigma_m^2 = 0.1$ for $m = 1, 2, 3$ and σ_4^2 is a very large constant.

Fig. 1 presents the true and the estimated fields $\boldsymbol{l}(\boldsymbol{x})$, where the crosses denote the sensor locations. Each sensor transmits 6 measurements, each produced with a different $\boldsymbol{\phi}_{\boldsymbol{x}}$ and quantized to 8 bits. In total, a sensor transmits only 48 bits. Yet, it is observed that the constructed PSD maps match well with the true ones for all three sources as well as the background noise.

To see the performance depending on the number of quantization levels, Fig. 2 depicts the reconstruction mean square error (MSE) obtained by Monte Carlo simulations. To capture only the quantization effects, no measurement errors were considered. Each sensor reports 8 measurements to the FC. Two observations are made. First, the MSE does not go to zero even when the number of bits per measurement increases. This is reasonable considering the limited number of sensors. Figuratively speaking, the situation is like trying to reconstruct a band-limited signal with samples taken below the Nyquist rate. Secondly, the MSE becomes flat after certain number of bits, offering optimal quantization levels at around 4 bits per measurement.

5. CONCLUSIONS

A novel cartography scheme has been introduced to estimate PSD maps across frequency and space by relying on a small number of bits reported by a set of inexpensive sensors. A nonparametric regression problem for vector-valued functions was formulated, which was seen to be equivalent to SVM formulations. Future research directions include exploring full implications of this as well as obtaining online algorithms capable of tracking time variations.

6. REFERENCES

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

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

[3] 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 *Int. Conf. Commun.*, Jun. 2013, pp. 5657–5661.

[4] 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.*, 2014, to appear.

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

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

[7] S.-J. Kim, N. Jain, G. Giannakis, and P. Forero, "Joint link learning and cognitive radio sensing," in *45th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 2011.

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

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

[10] E. Dall'Anese, S.-J. Kim, and G. B. Giannakis, "Channel gain map tracking via distributed kriging," *IEEE Trans. Veh. Technol.*, vol. 60, no. 3, pp. 1205–1211, 2011.

[11] E. Axell, G. Leus, and E. G. Larsson, "Overview of spectrum sensing for cognitive radio," in *Int. Workshop Cognitive Inform. Process.*, 2010, pp. 322–327.

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

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

[14] D. D. Arianandа, D. Romero, and G. Leus, "Cooperative compressive power spectrum estimation," in *IEEE Sensor Array Multichannel Signal Process. Workshop*, A Corunha, Spain, Jun. 2014.

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

[16] O. Mehanna, "Frugal sensing and estimation over wireless networks," *PhD thesis, University of Minnesota, MN*, 2014.

[17] 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.

[18] S. Becker, *Practical compressed sensing: modern data acquisition and signal processing*, Ph.D. thesis, California Institute of Technology, 2011.

[19] "IEEE 802.22 standard: Cognitive wireless RAN medium access control (MAC) and physical layer (PHY) specifications: policies and procedures for operation in the TV bands," Jul. 2011.

[20] V. Cherkassky and F. M. Mulier, *Learning from data: concepts, theory, and methods*, John Wiley & Sons, 2007.

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

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

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

[24] C. Carmeli, E. De Vito, A. Toigo, and V. Umanita, "Vector valued reproducing kernel Hilbert spaces and universality," *Anal. Appl.*, vol. 8, no. 01, pp. 19–61, 2010.

[25] B. Schölkopf and A. J. Smola, *Learning with kernels: Support vector machines, regularization, optimization, and beyond*, the MIT Press, 2001.

[26] A. Berlinet and C. Thomas-Agnan, *Reproducing kernel Hilbert spaces in probability and statistics*, Springer, 2004.

[27] J. C. Platt, "Fast training of support vector machines using sequential minimal optimization," in *Adv. Kernel Methods – Support Vector Learning*, chapter 12. The MIT Press, Cambridge, MA, 2000.

[28] V. Kecman, M. Vogt, and T. M. Huang, "On the equality of kernel AdaTron and sequential minimal optimization in classification and regression tasks and alike algorithms for kernel machines," in *European Symp. Artificial Neural Netw.*, Apr. 2003, pp. 215–222.