in Proc. IEEE ISIT 2011, Saint Petersburg, Russia, July/Aug. 2011, pp. 124--128

Minimum Variance Estimation for the Sparse Signal in Noise Model

Sebastian Schmutzhard[1], *Alexander Jung*[2], *and Franz Hlawatsch*[2]

[1]NuHAG, Faculty of Mathematics, University of Vienna, A-1090 Vienna, Austria
e-mail: sebastian.schmutzhard@univie.ac.at

[2]Institute of Telecommunications, Vienna University of Technology, A-1040 Vienna, Austria
e-mail: {ajung, fhlawats}@nt.tuwien.ac.at

***Abstract*—We consider estimation of a sparse parameter vector from measurements corrupted by white Gaussian noise. Using the framework of reproducing kernel Hilbert spaces, we derive closed-form expressions of the Barankin bound, i.e., of the minimum locally achievable variance of any estimator with a prescribed bias function, including the unbiased case. We also derive the *locally minimum variance* (LMV) estimator that achieves the minimum variance, and a necessary and sufficient condition on the prescribed bias function for the existence of finite-variance estimators and, simultaneously, of the LMV estimator. Finally, we present a numerical comparison of the variance of the hard-thresholding estimator with the corresponding minimum achievable variance.**

***Index Terms*—Sparsity, denoising, reproducing kernel Hilbert space, RKHS, Barankin bound, minimum variance estimation, unbiased estimation.**

I. Introduction

We consider the problem of estimating an unknown but deterministic signal or parameter vector $\mathbf{x} \in \mathbb{R}^N$ based on a noisy observation $\mathbf{y} \in \mathbb{R}^N$, i.e.,

$$\mathbf{y} = \mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ is white Gaussian noise with known variance $\sigma^2 > 0$. The signal vector $\mathbf{x}$ is assumed S-sparse, i.e.,

$$\mathbf{x} \in \mathcal{X}_S \triangleq \left\{ \mathbf{x}' \in \mathbb{R}^N \,\middle|\, \|\mathbf{x}'\|_0 \leq S \right\}, \tag{2}$$

where $\|\mathbf{x}\|_0$ denotes the number of nonzero entries of $\mathbf{x}$. Whereas the sparsity degree S is assumed known, the set of positions of the nonzero entries of $\mathbf{x}$ (denoted by $\mathcal{S}(\mathbf{x})$; note that $|\mathcal{S}(\mathbf{x})| = \|\mathbf{x}\|_0 \leq S$) is unknown. We call the model defined by (1) and (2) the *sparse signal in noise model* (SSNM). Applications of the SSNM include channel estimation when the channel consists of only few significant taps and an orthogonal training signal is used, and image denoising using an orthonormal wavelet basis (see references in [1]).

For the SSNM, lower and upper bounds on the minimum achievable variance of unbiased estimators were derived in [1]. In [2, 3], a Cramér–Rao bound (CRB) was derived for a generalization of the SSNM called the *sparse linear model* (SLM). However, the CRB is discontinuous when passing from the case $\|\mathbf{x}\|_0 < S$ to the case $\|\mathbf{x}\|_0 = S$. One can conclude from this discontinuity that the CRB cannot be tight for all $\mathbf{x}$, i.e., there is a gap between the CRB and the minimum achievable variance (also known as the Barankin bound [4, 5]). Improved variance bounds for the SLM were derived in [6].

In this paper, we present a closed-form expression of the minimum variance at a given signal vector $\mathbf{x}_0 \in \mathcal{X}_S$ that can be achieved within the SSNM by any estimator $\hat{\mathbf{x}}(\cdot)$ that has a prescribed bias $\mathbf{c}(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}_S$. We also derive a closed-form expression of the *locally minimum variance* (LMV) estimator, i.e., the estimator achieving this minimum variance at $\mathbf{x}_0$ while satisfying the bias constraint. Finally, we present a necessary and sufficient condition on the bias function $\mathbf{c}(\cdot)$ for the existence of estimators with finite variance at $\mathbf{x}_0$ and, simultaneously, for the existence of the LMV estimator at $\mathbf{x}_0$.

Our main mathematical tool will be the theory of *reproducing kernel Hilbert spaces* (RKHS) [7]. The use of RKHS theory for minimum variance estimation has a long history [8, 9]. In [6], we applied RKHS theory to minimum variance estimation within the SLM and derived a lower variance bound that is tighter than the CRB in [2, 3]. In the present paper, we consider exclusively the SSNM, which is a special case of the SLM with a simpler structure. Our results will be stronger than those of [6] specialized to the SSNM.

This paper is organized as follows. In Section II, we review some elements of minimum variance estimation [4, 10]. Section III summarizes the RKHS approach to minimum variance estimation [8, 9]. In Section IV, we use RKHS theory to derive closed-form expressions of the minimum achievable variance and of the LMV estimator for the SSNM. The special case of bias functions with a particular "diagonal" structure is considered in Section V. In Section VI, we present numerical results that compare the minimum variance with the variance of the hard-thresholding estimator.

II. Review of Minimum Variance Estimation

A popular criterion for judging the performance of an estimator $\hat{\mathbf{x}}(\cdot)$ is the mean squared error (MSE)

$$\varepsilon(\hat{\mathbf{x}}(\cdot); \mathbf{x}) \triangleq \mathsf{E}_{\mathbf{x}}\left\{ \|\hat{\mathbf{x}}(\mathbf{y}) - \mathbf{x}\|_2^2 \right\},$$

This work was supported by the FWF under Grants S10602 (Signal and Information Representation) and S10603 (Statistical Inference) within the National Research Network SISE, and by the WWTF under Grant MA 07-004 (SPORTS).

where the notation $\mathsf{E}_{\mathbf{x}}\{\cdot\}$ indicates that the expectation is taken with respect to the probability density function (pdf) $f(\mathbf{y};\mathbf{x})$ parametrized by $\mathbf{x}$. For the SSNM,

$$f(\mathbf{y};\mathbf{x}) \,=\, \frac{1}{(2\pi\sigma^2)^{N/2}}\, \exp\!\bigg(\!-\frac{\|\mathbf{y}-\mathbf{x}\|_2^2}{2\sigma^2}\bigg). \tag{3}$$

The MSE can be written as the sum of two nonnegative terms:

$$\varepsilon(\hat{\mathbf{x}}(\cdot);\mathbf{x}) \,=\, \|\mathbf{b}(\hat{\mathbf{x}}(\cdot);\mathbf{x})\|_2^2 \,+\, v(\hat{\mathbf{x}}(\cdot);\mathbf{x})\,,$$

with the estimator bias $\mathbf{b}(\hat{\mathbf{x}}(\cdot);\mathbf{x}) \triangleq \mathsf{E}_{\mathbf{x}}\{\hat{\mathbf{x}}(\mathbf{y})\} - \mathbf{x}$ and the estimator variance $v(\hat{\mathbf{x}}(\cdot);\mathbf{x}) \triangleq \mathsf{E}_{\mathbf{x}}\big\{\big\|\hat{\mathbf{x}}(\mathbf{y}) - \mathsf{E}_{\mathbf{x}}\{\hat{\mathbf{x}}(\mathbf{y})\}\big\|_2^2\big\}$. The MSE, bias, and variance depend on the underlying true signal vector $\mathbf{x}$.

A. Minimum Variance Estimation of $\mathbf{x}$

Requiring the estimator $\hat{\mathbf{x}}(\cdot)$ to minimize the MSE simultaneously for all $\mathbf{x}$ is not meaningful because such an estimator generally does not exist [10]. Therefore, a common approach is to fix the bias $\mathbf{b}(\hat{\mathbf{x}}(\cdot);\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}_S$ and look for the estimator(s) achieving minimum variance at a specific nominal signal vector $\mathbf{x}_0 \in \mathcal{X}_S$ under this bias constraint. Thus, hereafter we only consider estimators satisfying

$$\mathbf{b}(\hat{\mathbf{x}}(\cdot);\mathbf{x}) = \mathbf{c}(\mathbf{x})\,, \quad \forall\, \mathbf{x} \in \mathcal{X}_S\,, \tag{4}$$

for a prescribed bias function $\mathbf{c}(\cdot)\colon \mathcal{X}_S \to \mathbb{R}^N$. This is equivalent to the following constraint on the mean of the estimator:

$$\mathsf{E}_{\mathbf{x}}\{\hat{\mathbf{x}}(\mathbf{y})\} = \boldsymbol{\gamma}(\mathbf{x})\,, \;\; \forall\, \mathbf{x} \in \mathcal{X}_S\,, \quad \text{with}\; \boldsymbol{\gamma}(\mathbf{x}) \triangleq \mathbf{c}(\mathbf{x}) + \mathbf{x}\,. \tag{5}$$

The *LMV estimator at* $\mathbf{x}_0$, denoted $\hat{\mathbf{x}}^{(\mathbf{x}_0)}(\cdot)$, is then defined as the solution to the following optimization problem:

$$\hat{\mathbf{x}}^{(\mathbf{x}_0)}(\cdot) \,=\, \underset{\hat{\mathbf{x}}(\cdot)\in\mathcal{B}_{\boldsymbol{\gamma}(\cdot),\mathbf{x}_0}}{\operatorname{argmin}}\; v(\hat{\mathbf{x}}(\cdot);\mathbf{x}_0)\,, \tag{6}$$

where the constraint set $\mathcal{B}_{\boldsymbol{\gamma}(\cdot),\mathbf{x}_0}$ consists of all estimators whose bias is equal to $\mathbf{c}(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}_S$ and whose variance at $\mathbf{x}_0$ is finite:

$$\mathcal{B}_{\boldsymbol{\gamma}(\cdot),\mathbf{x}_0} \triangleq \big\{\hat{\mathbf{x}}(\cdot)\colon \mathbb{R}^N \!\to \mathbb{R}^N \,\big|\, \mathsf{E}_{\mathbf{x}}\{\hat{\mathbf{x}}(\mathbf{y})\} = \boldsymbol{\gamma}(\mathbf{x}) \;\forall \mathbf{x} \in \mathcal{X}_S\,, \\ v(\hat{\mathbf{x}}(\cdot);\mathbf{x}_0) < \infty\big\}.$$

Note that the estimators $\hat{\mathbf{x}}(\cdot)$ we consider are not constrained to be S-sparse. The prior information of S-sparsity enters merely by the fact that the bias constraint (4), (5) is formulated only for $\mathbf{x} \in \mathcal{X}_S$.

If an LMV estimator $\hat{\mathbf{x}}^{(\mathbf{x}_0)}(\cdot)$ exists, we call its variance,

$$L_{\boldsymbol{\gamma}(\cdot),\mathbf{x}_0} \triangleq \min_{\hat{\mathbf{x}}(\cdot)\in\mathcal{B}_{\boldsymbol{\gamma}(\cdot),\mathbf{x}_0}} v(\hat{\mathbf{x}}(\cdot);\mathbf{x}_0)\,,$$

the *minimum achievable variance* at $\mathbf{x}_0$ for the prescribed bias function $\mathbf{c}(\cdot)$. Furthermore, we call $\mathbf{c}(\cdot)$ a *valid* bias function at $\mathbf{x}_0$ if $\mathcal{B}_{\boldsymbol{\gamma}(\cdot),\mathbf{x}_0}$ is nonempty, i.e., if there exists at least one estimator with bias $\mathbf{c}(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}_S$ and with finite variance at $\mathbf{x}_0$.

B. Minimum Variance Estimation of x_k

The variance can be decomposed as

$$v(\hat{\mathbf{x}}(\cdot);\mathbf{x}) \,=\, \sum_{k\in[N]} v(\hat{x}_k(\cdot);\mathbf{x})\,,$$

where $[N] \triangleq \{1,\ldots,N\}$ and $v(\hat{x}_k(\cdot);\mathbf{x}) \triangleq \mathsf{E}_{\mathbf{x}}\big\{\big[\hat{x}_k(\mathbf{y}) - \mathsf{E}_{\mathbf{x}}\{\hat{x}_k(\mathbf{y})\}\big]^2\big\}$ is the variance of the component estimator $\hat{x}_k(\cdot) = \big(\hat{\mathbf{x}}(\cdot)\big)_k$. Furthermore, the estimator $\hat{\mathbf{x}}(\cdot)$ has mean $\boldsymbol{\gamma}(\mathbf{x}) = \mathbf{c}(\mathbf{x}) + \mathbf{x}$ if and only if the mean of each component $\hat{x}_k(\cdot)$ equals $\gamma_k(\mathbf{x}) = c_k(\mathbf{x}) + x_k$, where $\gamma_k(\mathbf{x})$, $c_k(\mathbf{x})$, and x_k denote the kth components of $\boldsymbol{\gamma}(\mathbf{x})$, $\mathbf{c}(\mathbf{x})$, and $\mathbf{x}$, respectively. It is then easily verified that solving the problem (6) is equivalent to separately solving for each $k \in [N]$ the scalar problem

$$\hat{x}_k^{(\mathbf{x}_0)}(\cdot) \,=\, \underset{\hat{x}_k(\cdot)\in\mathcal{B}_{\gamma_k(\cdot),\mathbf{x}_0}}{\operatorname{argmin}}\; v(\hat{x}_k(\cdot);\mathbf{x}_0)\,, \tag{7}$$

with the constraint set

$$\mathcal{B}_{\gamma_k(\cdot),\mathbf{x}_0} \triangleq \big\{\hat{x}_k(\cdot)\colon \mathbb{R}^N \!\to \mathbb{R} \,\big|\, \mathsf{E}_{\mathbf{x}}\{\hat{x}_k(\mathbf{y})\} = \gamma_k(\mathbf{x}) \;\forall \mathbf{x} \in \mathcal{X}_S\,, \\ v(\hat{x}_k(\cdot);\mathbf{x}_0) < \infty\big\}.$$

Therefore, hereafter we will consider the scalar problem (7). We will call $\hat{x}_k^{(\mathbf{x}_0)}(\cdot)$ (if it exists) the LMV estimator of x_k at $\mathbf{x}_0$, and its variance, $L_{\gamma_k(\cdot),\mathbf{x}_0} \triangleq \min_{\hat{x}_k(\cdot)\in\mathcal{B}_{\gamma_k(\cdot),\mathbf{x}_0}} v(\hat{x}_k(\cdot);\mathbf{x}_0)$, the minimum achievable variance at $\mathbf{x}_0$ for the prescribed bias function $c_k(\cdot)$. Furthermore, we will call $c_k(\cdot) : \mathcal{X}_S \to \mathbb{R}$ a valid bias function at $\mathbf{x}_0$ if $\mathcal{B}_{\gamma_k(\cdot),\mathbf{x}_0}$ is nonempty. If all $c_k(\cdot)$, $k \in [N]$ are valid bias functions at $\mathbf{x}_0$, then so is $\mathbf{c}(\cdot)$, and we have $L_{\boldsymbol{\gamma}(\cdot),\mathbf{x}_0} = \sum_{k\in[N]} L_{\gamma_k(\cdot),\mathbf{x}_0}$.

Using the RKHS framework of [8], it can be shown that if $c_k(\cdot)$ is a valid bias function at $\mathbf{x}_0$, then there exists a unique LMV estimator $\hat{x}_k^{(\mathbf{x}_0)}(\cdot)$. Thus, a necessary and sufficient condition for the existence of a unique LMV estimator is that there is at least one estimator $\hat{x}_k(\cdot)$ with finite variance at $\mathbf{x}_0$ whose mean $\mathsf{E}_{\mathbf{x}}\{\hat{x}_k(\mathbf{y})\}$ is equal to $\gamma_k(\mathbf{x}) = c_k(\mathbf{x}) + x_k$ for all $\mathbf{x} \in \mathcal{X}_S$. It can also be shown that if an LMV estimator exists, it is unique [4].

If there exists a single estimator $\hat{x}_k(\cdot)$ which is the LMV estimator for *all* $\mathbf{x}_0$, i.e., which solves (7) for all $\mathbf{x}_0$ simultaneously, then that estimator is called a *uniformly minimum variance* (UMV) estimator. For the special case of unbiased estimation, i.e., $c_k(\mathbf{x}) \equiv 0$ or equivalently $\gamma_k(\mathbf{x}) \equiv x_k$, the LMV and UMV estimators are termed *locally minimum variance unbiased* (LMVU) and *uniformly minimum variance unbiased* (UMVU) estimators, respectively. However, for the SSNM with sparsity $S < N$, it was shown in [11] that, under mild technical conditions, there does not exist a UMVU estimator. For the nonsparse case, i.e., $S = N$, a UMVU estimator is given by the trivial estimator $\hat{x}_k(\mathbf{y}) = y_k$.

III. The RKHS Framework

The RKHS framework of classical estimation was introduced in [8], further developed in [9], and specialized to the SLM and SSNM in [6, 12]. This framework is based on two Hilbert spaces. The first, denoted $\mathcal{L}$, consists of functions $a(\cdot)\colon \mathbb{R}^N \to \mathbb{R}$. For the SSNM, it is defined as the closed linear span of the set $\{\rho_{\mathbf{x}}(\cdot)\}_{\mathbf{x}\in\mathcal{X}_S}$, where

$$\rho_{\mathbf{x}}(\mathbf{y}) \triangleq \frac{f(\mathbf{y};\mathbf{x})}{f(\mathbf{y};\mathbf{x}_0)} = \exp\left(\frac{2\mathbf{y}^T(\mathbf{x}-\mathbf{x}_0) - \|\mathbf{x}\|_2^2 + \|\mathbf{x}_0\|_2^2}{2\sigma^2}\right)$$

(cf. (3)), with a fixed $\mathbf{x}_0 \in \mathcal{X}_S$, is the likelihood ratio for the parameter vector $\mathbf{x}$. The inner product in $\mathcal{L}$ (which is also required mathematically for the closure operation defining $\mathcal{L}$) is given by $\langle a, b\rangle_{\mathrm{RV}} \triangleq \mathsf{E}_{\mathbf{x}_0}\{a(\mathbf{y})b(\mathbf{y})\}$, with induced norm $\|a\|_{\mathrm{RV}} = \sqrt{\langle a, a\rangle_{\mathrm{RV}}} = \sqrt{\mathsf{E}_{\mathbf{x}_0}\{a^2(\mathbf{y})\}}$. Note that these quantities explicitly depend on $\mathbf{x}_0$.

The second Hilbert space, denoted $\mathcal{H}(R)$, is an RKHS [7] consisting of functions $f(\cdot) : \mathcal{X}_S \to \mathbb{R}$. It is defined via the kernel function $R(\cdot,\cdot) : \mathcal{X}_S \times \mathcal{X}_S \to \mathbb{R}$ given by

$$R(\mathbf{x}_1, \mathbf{x}_2) \triangleq \langle \rho_{\mathbf{x}_1}, \rho_{\mathbf{x}_2}\rangle_{\mathrm{RV}} = \exp\left(\frac{(\mathbf{x}_1-\mathbf{x}_0)^T(\mathbf{x}_2-\mathbf{x}_0)}{\sigma^2}\right).$$

More precisely, $\mathcal{H}(R)$ is defined as the closure of the linear span of the set of functions $\{f_{\mathbf{x}'}(\mathbf{x}) = R(\mathbf{x}, \mathbf{x}')\}_{\mathbf{x}'\in\mathcal{X}_S}$ (i.e., these functions are obtained from $R(\mathbf{x}, \mathbf{x}')$ by fixing $\mathbf{x}'$ and using $\mathbf{x}$ as the function argument). This closure is taken with respect to the topology that is given by the inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}(R)}$ defined via the *reproducing property* [7]

$$\big\langle f(\cdot), R(\cdot, \mathbf{x}')\big\rangle_{\mathcal{H}(R)} = f(\mathbf{x}')\,.$$

The above relation holds for all $f(\cdot) \in \mathcal{H}(R)$ and $\mathbf{x}' \in \mathcal{X}_S$. The induced norm is $\|f\|_{\mathcal{H}(R)} = \sqrt{\langle f, f\rangle_{\mathcal{H}(R)}}$. As shown in [8], $\mathcal{H}(R)$ is isometric to $\mathcal{L}$; there exists a congruence[1] $\mathsf{J}[\cdot] : \mathcal{H}(R) \to \mathcal{L}$, which is completely specified by the relation

$$\mathsf{J}[R(\cdot, \mathbf{x})] = \rho_{\mathbf{x}}(\cdot)\,.$$

The following facts, proven in [8] in the general context of minimum variance estimation, are the basis for our results. As before, we consider the SSNM together with a prescribed bias function $c_k(\cdot)$—equivalently, a prescribed mean function $\gamma_k(\mathbf{x}) = c_k(\mathbf{x}) + x_k$—and a fixed parameter vector $\mathbf{x}_0 \in \mathcal{X}_S$.

1.) There exists a finite minimum in (7), i.e., $c_k(\cdot)$ is valid at $\mathbf{x}_0$, if and only if $\gamma_k(\cdot) \in \mathcal{H}(R)$.

2.) If $\gamma_k(\cdot) \in \mathcal{H}(R)$, then the minimum variance at $\mathbf{x}_0$ achievable by an estimator $\hat{x}_k(\cdot)$ with prescribed bias function $c_k(\cdot)$ is given by

$$L_{\gamma_k(\cdot),\mathbf{x}_0} = \|\gamma_k\|^2_{\mathcal{H}(R)} - \gamma_k^2(\mathbf{x}_0)\,. \tag{8}$$

3.) The unique LMV estimator $\hat{x}_k^{(\mathbf{x}_0)}(\cdot)$ that achieves this minimum variance is given by

$$\hat{x}_k^{(\mathbf{x}_0)}(\cdot) = \mathsf{J}[\gamma_k(\cdot)]\,. \tag{9}$$

IV. Minimum Variance Estimation for the SSNM

According to (8) and (9), finding the minimum achievable variance and the LMV estimator amounts to evaluating the squared RKHS norm $\|\gamma_k\|^2_{\mathcal{H}(R)}$ and the image $\mathsf{J}[\gamma_k(\cdot)]$ of the prescribed mean function $\gamma_k(\mathbf{x}) = c_k(\mathbf{x}) + x_k$.

A. An Isometric RKHS

To perform these tasks, it will be convenient to work not directly in the RKHS $\mathcal{H}(R)$ but in another RKHS $\mathcal{H}(R')$ of functions $f(\cdot) : \mathcal{X}_S \to \mathbb{R}$. This RKHS is defined by the kernel $R'(\cdot,\cdot) : \mathcal{X}_S \times \mathcal{X}_S \to \mathbb{R}$ given by

$$R'(\mathbf{x}_1, \mathbf{x}_2) \triangleq \exp\left(\mathbf{x}_1^T\mathbf{x}_2\right).$$

The two RKHSs $\mathcal{H}(R)$ and $\mathcal{H}(R')$ are isometric; a congruence $\mathsf{K}[\cdot] : \mathcal{H}(R) \to \mathcal{H}(R')$ is provided by

$$\mathsf{K}[f(\cdot)] = f(\sigma\mathbf{x})\,\nu_{\mathbf{x}_0}(\mathbf{x})\,, \quad \mathbf{x}\in\mathcal{X}_S\,,\ f(\cdot)\in\mathcal{H}(R)\,, \tag{10}$$

with the weight function $\nu_{\mathbf{x}_0}(\cdot) : \mathcal{X}_S \to \mathbb{R}$ defined as

$$\nu_{\mathbf{x}_0}(\mathbf{x}) \triangleq \exp\left(-\frac{\|\mathbf{x}_0\|_2^2}{2\sigma^2} + \frac{\mathbf{x}_0^T\mathbf{x}}{\sigma}\right). \tag{11}$$

Due to (10), $R'(\mathbf{x}_1, \mathbf{x}_2) = R(\sigma\mathbf{x}_1, \sigma\mathbf{x}_2)\,\nu_{\mathbf{x}_0}(\mathbf{x}_1)\nu_{\mathbf{x}_0}(\mathbf{x}_2)$.

We now present a characterization of the RKHS $\mathcal{H}(R')$ in the sense of [9], i.e., we will specify an orthonormal basis for $\mathcal{H}(R')$ together with a condition that allows one to judge if a given function $f(\cdot) : \mathcal{X}_S \to \mathbb{R}$ belongs to $\mathcal{H}(R')$. It is one of the appealing properties of $\mathcal{H}(R')$ that an orthonormal basis can be readily constructed. In what follows, let $\mathbb{N}_S \triangleq \mathbb{Z}_+^N \cap \mathcal{X}_S$ (with $\mathbb{Z}_+ \triangleq \{0, 1, \ldots\}$ the set of nonnegative integers) be the set of all S-sparse N-dimensional multi-indices $\mathbf{p} = (p_1, \ldots, p_N)$. Furthermore, let $\mathbf{p}! \triangleq \prod_{k\in[N]} p_k!$, $\mathbf{x}^{\mathbf{p}} \triangleq \prod_{k\in[N]} x_k^{p_k}$, and $\frac{\partial^{\mathbf{p}} f(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}}} \triangleq \left(\prod_{k\in[N]} \frac{\partial^{p_k}}{\partial x_k^{p_k}}\right) f(\mathbf{x})$. The following results are presented without proof because of space restrictions; a detailed proof is provided in [12].

Theorem 1: 1.) For any $\mathbf{p} \in \mathbb{N}_S$, the function

$$g^{(\mathbf{p})}(\mathbf{x}) \triangleq \frac{1}{\sqrt{\mathbf{p}!}} \frac{\partial^{\mathbf{p}} R'(\mathbf{x}, \mathbf{x}')}{\partial \mathbf{x}'^{\mathbf{p}}}\bigg|_{\mathbf{x}'=\mathbf{0}} = \frac{1}{\sqrt{\mathbf{p}!}}\,\mathbf{x}^{\mathbf{p}}$$

is an element of $\mathcal{H}(R')$.

2.) The functions $\{g^{(\mathbf{p})}(\cdot)\}_{\mathbf{p}\in\mathbb{N}_S}$ are orthonormal, i.e.,

$$\langle g^{(\mathbf{p})}, g^{(\mathbf{p}')}\rangle_{\mathcal{H}(R')} = \begin{cases} 1, & \mathbf{p} = \mathbf{p}' \\ 0, & \mathbf{p} \neq \mathbf{p}'. \end{cases}$$

3.) The inner product of any $f(\cdot) \in \mathcal{H}(R')$ with $g^{(\mathbf{p})}(\cdot)$ is given by

$$\langle f, g^{(\mathbf{p})}\rangle_{\mathcal{H}(R')} = \frac{1}{\sqrt{\mathbf{p}!}} \frac{\partial^{\mathbf{p}} f(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}}}\bigg|_{\mathbf{x}=\mathbf{0}}.$$

4.) The set $\{g^{(\mathbf{p})}(\cdot)\}_{\mathbf{p}\in\mathbb{N}_S}$ forms an orthonormal basis for $\mathcal{H}(R')$, i.e., any $f(\cdot) \in \mathcal{H}(R')$ can be expanded as

$$f(\cdot) = \sum_{\mathbf{p}\in\mathbb{N}_S} \langle f, g^{(\mathbf{p})}\rangle_{\mathcal{H}(R')}\, g^{(\mathbf{p})}(\cdot)\,,$$

where the sum converges in the RKHS norm $\|\cdot\|_{\mathcal{H}(R')}$ and also pointwise.

[1] A linear mapping $\mathsf{J}[\cdot] : \mathcal{H}_1 \to \mathcal{H}_2$ between two Hilbert spaces $\mathcal{H}_1$, $\mathcal{H}_2$ is said to be a *congruence* if it is bijective and preserves inner products, i.e., $\langle f, g\rangle_{\mathcal{H}_1} = \langle \mathsf{J}[f], \mathsf{J}[g]\rangle_{\mathcal{H}_2}$ for all $f(\cdot), g(\cdot) \in \mathcal{H}_1$. Two Hilbert spaces are said to be *isometric* if there exists at least one congruence between them.

5.) A function $f(\cdot)\colon \mathcal{X}_S \to \mathbb{R}$ is an element of $\mathcal{H}(R')$ if and only if it can be written pointwise as a series

$$f(\mathbf{x}) = \sum_{\mathbf{p}\in\mathbb{N}_S} a[\mathbf{p}]\, g^{(\mathbf{p})}(\mathbf{x})\,,$$

with a coefficient sequence $a[\mathbf{p}] \in \ell^2(\mathbb{N}_S)$, where $\ell^2(\mathbb{N}_S)$ denotes the set of all square summable sequences $a[\cdot]\colon \mathbb{N}_S \to \mathbb{R}$.

Based on Theorem 1, we will next characterize for the SSNM the set of valid bias functions, the minimum achievable variance, and the LMV estimator.

B. Valid Bias Functions

Consider a bias function $c_k(\cdot)$ and the associated mean function $\gamma_k(\mathbf{x}) = c_k(\mathbf{x}) + x_k$. According to Section III, $c_k(\cdot)$ is valid at $\mathbf{x}_0$ if and only if $\gamma_k(\cdot) \in \mathcal{H}(R)$. It is then further shown in [12] that $c_k(\cdot)$ is valid at $\mathbf{x}_0$ if and only if there exists a coefficient sequence $a[\mathbf{p}] \in \ell^2(\mathbb{N}_S)$ such that

$$\gamma_k(\sigma\mathbf{x})\,\nu_{\mathbf{x}_0}(\mathbf{x}) = \sum_{\mathbf{p}\in\mathbb{N}_S} a[\mathbf{p}]\,\frac{\mathbf{x}^{\mathbf{p}}}{\sqrt{\mathbf{p}!}}\,, \quad \forall\, \mathbf{x}\in\mathcal{X}_S\,, \tag{12}$$

with $\nu_{\mathbf{x}_0}(\mathbf{x})$ defined in (11). In the unbiased case $c_k(\cdot) \equiv 0$, we have $\gamma_k(\sigma\mathbf{x}) = \sigma x_k$; it can here be verified easily that condition (12) is satisfied, and hence $c_k(\cdot) \equiv 0$ is a valid bias function.

Based on condition (12), it can be shown that for specific choices of σ, S, and N, there are bias functions $c_k(\cdot)$ that are valid at some $\mathbf{x}_0$ but not at all $\mathbf{x}_0 \in \mathcal{X}_S$ [12]. We note that if $c_k(\cdot)$ is the actual bias function of an existing estimator with finite variance at all $\mathbf{x}_0 \in \mathcal{X}_S$ (e.g., the hard-thresholding estimator, cf. Section VI), then it trivially follows by the very definition of a valid bias function in Section II-B that $c_k(\cdot)$ is valid at all $\mathbf{x}_0 \in \mathcal{X}_S$. In particular, the validity of $c_k(\cdot) \equiv 0$ also follows trivially from the fact that the specific estimator $\hat{x}_k(\mathbf{y}) = y_k$ is unbiased and has a finite variance at all $\mathbf{x}_0$.

C. Minimum Achievable Variance and LMV Estimator

The following further results are shown in [12]. If the prescribed bias $c_k(\cdot)$ is valid at $\mathbf{x}_0$, then it follows from (8), with (12), that the minimum achievable variance at $\mathbf{x}_0$ is given by

$$\begin{aligned} L_{\gamma_k(\cdot),\mathbf{x}_0} &= \sum_{\mathbf{p}\in\mathbb{N}_S} \frac{1}{\mathbf{p}!}\bigg(\frac{\partial^{\mathbf{p}}[\gamma_k(\sigma\mathbf{x})\,\nu_{\mathbf{x}_0}(\mathbf{x})]}{\partial\mathbf{x}^{\mathbf{p}}}\bigg|_{\mathbf{x}=\mathbf{0}}\bigg)^2 - \gamma_k^2(\mathbf{x}_0) \\ &= \sum_{\mathbf{p}\in\mathbb{N}_S} a^2[\mathbf{p}] - \gamma_k^2(\mathbf{x}_0)\,. \end{aligned} \tag{13}$$

Furthermore, the LMV estimator at $\mathbf{x}_0$—i.e., the estimator whose variance at $\mathbf{x}_0$ equals $L_{\gamma_k(\cdot),\mathbf{x}_0}$ in (13)—is given by

$$\begin{aligned} \hat{x}_k^{(\mathbf{x}_0)}(\mathbf{y}) &= \sum_{\mathbf{p}\in\mathbb{N}_S} \frac{1}{\sqrt{\mathbf{p}!}}\bigg[\frac{\partial^{\mathbf{p}}[\gamma_k(\sigma\mathbf{x})\,\nu_{\mathbf{x}_0}(\mathbf{x})]}{\partial\mathbf{x}^{\mathbf{p}}}\,\frac{\partial^{\mathbf{p}}\psi_{\mathbf{x}_0}(\mathbf{x};\mathbf{y})}{\partial\mathbf{x}^{\mathbf{p}}}\bigg]_{\mathbf{x}=\mathbf{0}} \\ &= \sum_{\mathbf{p}\in\mathbb{N}_S} a[\mathbf{p}]\bigg[\frac{\partial^{\mathbf{p}}\psi_{\mathbf{x}_0}(\mathbf{x};\mathbf{y})}{\partial\mathbf{x}^{\mathbf{p}}}\bigg]_{\mathbf{x}=\mathbf{0}}, \end{aligned}$$

with

$$\psi_{\mathbf{x}_0}(\mathbf{x};\mathbf{y}) \triangleq \exp\bigg(\frac{\mathbf{y}^T(\sigma\mathbf{x}-\mathbf{x}_0)}{\sigma^2} + \frac{\mathbf{x}_0^T\mathbf{x}}{\sigma} - \frac{\|\mathbf{x}\|_2^2}{2}\bigg).$$

The expression (13) nicely demonstrates the reduction of the minimum achievable variance due to the sparsity constraint (2). For simplicity, we consider the unbiased case, i.e., $c_k(\cdot) \equiv 0$ or $\gamma_k(\mathbf{x}) = x_k$. It can here be shown that the difference between the minimum achievable variance without a sparsity constraint (this would be obtained for $S = N$) and the minimum achievable variance for the actual sparsity S is given by

$$\Delta L_{\mathbf{x}_0} = \sum_{\mathbf{p}\in\mathbb{Z}_+^N\setminus\mathbb{N}_S} a^2[\mathbf{p}]\,.$$

For decreasing S, $\Delta L_{\mathbf{x}_0}$ increases because the set $\mathbb{Z}_+^N \setminus \mathbb{N}_S$ becomes larger. Thus, when $\mathbf{x}_0$ becomes more sparse, the minimum achievable variance decreases.

V. Diagonal Bias Function

We now consider the SSNM for the special case of a bias function $c_k(\mathbf{x})$ that is "diagonal" in the sense that it depends only on the entry x_k, i.e., $c_k(\mathbf{x}) = \tilde{c}_k(x_k)$ with some function $\tilde{c}_k(\cdot) : \mathbb{R} \to \mathbb{R}$. We make the weak assumption that $\tilde{c}_k(\cdot)$ can be represented by a power series centered at $x_{0,k} \triangleq (\mathbf{x}_0)_k$. Equivalently, the mean function $\gamma_k(\mathbf{x}) = \tilde{c}_k(x_k) + x_k$ can be represented by a power series centered at $x_{0,k}$, i.e.,

$$\gamma_k(\mathbf{x}) = \sum_{l=0}^{\infty} \frac{\gamma_{k,l}}{l!}(x_k - x_{0,k})^l\,, \tag{14}$$

with suitable coefficients $\gamma_{k,l}$. Examples where the diagonal power series representation (14) applies include unbiased estimation and the mean functions of the hard- and soft-thresholding estimators (cf. Section VI).

The following results can now be shown [12]. A diagonal bias function is valid if and only if the coefficients $\gamma_{k,l}$ satisfy $P_k < \infty$, with $P_k \triangleq \sum_{l=0}^{\infty} \gamma_{k,l}^2 \sigma^{2l}/l!$. Furthermore, if the bias function is valid, the minimum achievable variance at $\mathbf{x}_0$ is obtained as

$$L_{\gamma_k(\cdot),\mathbf{x}_0} = g(\mathbf{x}_0)\,P_k - \gamma_k^2(\mathbf{x}_0)\,, \tag{15}$$

where

$$g(\mathbf{x}_0) \triangleq \sum_{j\in[S]} \exp\bigg(-\frac{x_{0,i_j}^2}{\sigma^2}\bigg) \prod_{j'\in[j-1]} \bigg[1 - \exp\bigg(-\frac{x_{0,i_{j'}}^2}{\sigma^2}\bigg)\bigg] \tag{16}$$

if $|\mathcal{S}(\mathbf{x}_0) \cup \{k\}| = S+1$ and $g(\mathbf{x}_0) \triangleq 1$ otherwise. As to the indices i_j in (16), we note that the case $|\mathcal{S}(\mathbf{x}_0) \cup \{k\}| = S+1$ implies that $\mathbf{x}_0$ has exactly S nonzero entries; the corresponding positions are denoted by $\mathcal{S}(\mathbf{x}_0) = \{i_1, \ldots, i_S\}$. Finally, the LMV estimator at $\mathbf{x}_0$ is given by

$$\hat{x}_k^{(\mathbf{x}_0)}(\mathbf{y}) = h(\mathbf{y},\mathbf{x}_0) \sum_{l=0}^{\infty} \frac{\gamma_{k,l}\,\sigma^l}{l!}\, H_l\Big(\frac{y_k}{\sigma}\Big)\,, \tag{17}$$

where

$$\begin{aligned} h(\mathbf{y},\mathbf{x}_0) \triangleq \sum_{j\in[S]} &\exp\bigg(-\frac{x_{0,i_j}^2 + 2y_{i_j}x_{0,i_j}}{2\sigma^2}\bigg) \\ &\times \prod_{j'\in[j-1]} \bigg[1 - \exp\bigg(-\frac{x_{0,i_{j'}}^2 + 2y_{i_{j'}}x_{0,i_{j'}}}{2\sigma^2}\bigg)\bigg] \end{aligned}$$

if $|\mathcal{S}(\mathbf{x}_0) \cup \{k\}| = S+1$ and $h(\mathbf{y},\mathbf{x}_0) \triangleq 1$ otherwise. Here,

$H_l(\cdot): \mathbb{R} \rightarrow \mathbb{R}$ is the lth-order (probabilists') Hermite polynomial, i.e., $H_l(x) \triangleq (-1)^l e^{x^2/2} \frac{d^l}{dx^l} e^{-x^2/2}$ [13].

Some comments can now be made. 1.) The LMVU estimator at $\mathbf{x}_0$ is given by (17) with $\gamma_{k,0} = x_{0,k}$, $\gamma_{k,1} = 1$, and $\gamma_{k,l} = 0$ for $l \geq 2$. One can show that its bias and variance remain finite at $\mathbf{x} \neq \mathbf{x}_0$; this is also true if $\mathbf{x} \notin \mathcal{X}_S$, i.e., if $\mathbf{x}$ violates the sparsity constraint (2).

2.) If $P_k = \sum_{l=0}^{\infty} \gamma_{k,l}^2 \sigma^{2l}/l!$ appearing in the expression (15) of $L_{\gamma_k(\cdot),\mathbf{x}_0}$ is replaced by a partial sum $\sum_{l \in \mathcal{T}} \gamma_{k,l}^2 \sigma^{2l}/l!$, with arbitrary $\mathcal{T} \subseteq \mathbb{Z}_+$, then the resulting expression provides a lower bound on $L_{\gamma_k(\cdot),\mathbf{x}_0}$ (and, in turn, on the variance at $\mathbf{x}_0$ of any estimator with mean equal to $\gamma_k(\cdot)$).

3.) Finally, consider a valid bias function $\mathbf{c}(\cdot)$ such that every component $\gamma_k(\cdot)$ of $\boldsymbol{\gamma}(\mathbf{x}) = \mathbf{c}(\mathbf{x}) + \mathbf{x}$ is of the diagonal form (14). It can then be shown [12] that the minimum achievable overall variance at $\mathbf{x}_0$, $L_{\boldsymbol{\gamma}(\cdot),\mathbf{x}_0} = \sum_{k \in [N]} L_{\gamma_k(\cdot),\mathbf{x}_0}$, is given by

$$L_{\boldsymbol{\gamma}(\cdot),\mathbf{x}_0} = \sum_{k \in \mathcal{S}(\mathbf{x}_0)} P_k + \sum_{k \in \overline{\mathcal{S}(\mathbf{x}_0)}} g(\mathbf{x}_0) P_k - \|\boldsymbol{\gamma}(\mathbf{x}_0)\|_2^2 \,, \quad (18)$$

where $\overline{\mathcal{S}(\mathbf{x}_0)} \triangleq [N] \backslash \mathcal{S}(\mathbf{x}_0)$.

VI. Numerical Results

We study the minimum achievable variance $L_{\boldsymbol{\gamma}(\cdot),\mathbf{x}_0}$ for prescribed mean functions $\boldsymbol{\gamma}(\cdot)$ that are the actual mean functions of given estimators. We consider the family of hard-thresholding (HT) estimators $\hat{x}_{\mathrm{HT},k}(\cdot)$ given by $\hat{x}_{\mathrm{HT},k}(\mathbf{y}) = y_k$ when $|y_k| \geq T$ and $\hat{x}_{\mathrm{HT},k}(\mathbf{y}) = 0$ otherwise, where T is a parameter. In the limiting case $T = 0$, the HT estimator equals the least squares (LS) estimator given by $\hat{x}_{\mathrm{LS},k}(\mathbf{y}) = y_k$, which is unbiased. The mean function of $\hat{x}_{\mathrm{HT},k}(\mathbf{y})$ can be shown to be of the form (14). For $T = 0$ (LS estimator), (14) simplifies because $\gamma_{k,0} = x_{0,k}$, $\gamma_{k,1} = 1$, and $\gamma_{k,l} = 0$ for $l \geq 2$. It can also be shown that, for any T, $\mathsf{E}_{\mathbf{x}_0}\{\hat{x}_{\mathrm{HT},k}(\mathbf{y})\} = \gamma_{k,0}$ and $v(\hat{x}_{\mathrm{HT},k}(\cdot); \mathbf{x}_0) = \sum_{l=1}^{\infty} \gamma_{k,l}^2 \sigma^{2l}/l! = P_k - \gamma_{k,0}^2$.

In what follows, we choose $N = 5$ and $\|\mathbf{x}_0\|_0 = S = 1$. In Fig. 1, we show the variance $v(\hat{\mathbf{x}}_{\mathrm{HT}}(\cdot); \mathbf{x}_0)$ (obtained by numerical integration) versus the "signal-to-noise ratio" (SNR) ξ_0^2/σ^2, where ξ_0 denotes the single nonzero entry of $\mathbf{x}_0$ and $\sigma^2 = 1$, for different choices of T. For comparison, along with each variance curve, we also show the corresponding minimum achievable variance (Barankin bound) $L_{\boldsymbol{\gamma}(\cdot),\mathbf{x}_0}$ calculated according to (18). Here, "corresponding" means that $\boldsymbol{\gamma}(\cdot)$ is chosen equal to the mean of the respective estimator, which was calculated by numerical integration. It is seen that for small T (in particular, for $T = 0$ where the HT estimator reduces to the LS estimator), the Barankin bound is significantly below the corresponding variance curve. However, as T increases, the gap between variance and Barankin bound becomes smaller; in particular, the two curves are already indistinguishable for $T = 4$. For high SNR, the Barankin bound converges to $S\sigma^2 = 1$ for any value of T; this equals the variance of an oracle estimator that knows the support of $\mathbf{x}_0$.

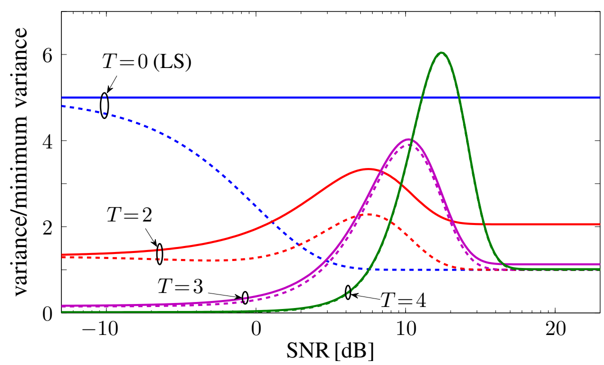

Fig. 1. Variance of the HT estimator, $v(\hat{\mathbf{x}}_{\mathrm{HT}}(\cdot); \mathbf{x}_0)$, for different T (solid lines) and corresponding minimum achievable variance (Barankin bound) $L_{\boldsymbol{\gamma}(\cdot),\mathbf{x}_0}$ (dashed lines) versus the SNR ξ_0^2/σ^2, for $N = 5$, $S = 1$, and $\sigma^2 = 1$.

VII. Conclusion

We applied the mathematical framework of reproducing kernel Hilbert spaces to minimum variance estimation within the sparse signal in noise model. This provided a necessary and sufficient condition on the prescribed bias function for the existence of estimators with finite variance and, simultaneously, for the existence of the *locally minimum variance* (LMV) estimator. We also derived closed-form expressions of the minimum locally achievable variance of any estimator with a prescribed bias function (Barankin bound), and of the LMV estimator that achieves the minimum variance. Finally, using numerical simulation, we analyzed how far the variance of the hard-thresholding estimator exceeds the corresponding minimum achievable variance (Barankin bound).

References

[1] A. Jung, Z. Ben-Haim, F. Hlawatsch, and Y. C. Eldar, "On unbiased estimation of sparse vectors corrupted by Gaussian noise," in *Proc. IEEE ICASSP-2010*, Dallas, TX, March 2010, pp. 3990–3993.

[2] Z. Ben-Haim and Y. C. Eldar, "Performance bounds for sparse estimation with random noise," in *Proc. IEEE-SP Workshop Statist. Signal Process.*, Cardiff, Wales (UK), Aug. 2009, pp. 225–228.

[3] ——, "The Cramér–Rao bound for estimating a sparse parameter vector," *IEEE Trans. Signal Processing*, vol. 58, pp. 3384–3389, June 2010.

[4] E. W. Barankin, "Locally best unbiased estimates," *Ann. Math. Statist.*, vol. 20, no. 4, pp. 477–501, 1949.

[5] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Trans. Inf. Theory*, vol. 36, no. 6, pp. 1285–1301, Nov. 1990.

[6] S. Schmutzhard, A. Jung, F. Hlawatsch, Z. Ben-Haim, and Y. C. Eldar, "A lower bound on the estimator variance for the sparse linear model," in *Proc. 44th Asilomar Conf. Signals, Systems, Computers*, Pacific Grove, CA, Nov. 2010.

[7] N. Aronszajn, "Theory of reproducing kernels," *Trans. Am. Math. Soc.*, vol. 68, no. 3, pp. 337–404, May 1950.

[8] E. Parzen, "Statistical inference on time series by Hilbert space methods, I." Appl. Math. Stat. Lab., Stanford University, Stanford, CA, Tech. Rep. 23, Jan. 1959.

[9] D. D. Duttweiler and T. Kailath, "RKHS approach to detection and estimation problems – Part V: Parameter estimation," *IEEE Trans. Inf. Theory*, vol. 19, no. 1, pp. 29–37, Jan. 1973.

[10] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[11] A. Jung, Z. Ben-Haim, F. Hlawatsch, and Y. C. Eldar, "Unbiased estimation of a sparse vector in white Gaussian noise," *submitted to IEEE Trans. Inf. Theory*, 2010.

[12] A. Jung, "An RKHS Approach to Estimation with Sparsity Constraints," Ph.D. dissertation, Vienna University of Technology, 2011.

[13] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions*. New York: Dover, 1965.