in Proc. 44th Asilomar Conf. Signals, Systems, Computers, Pacific Grove, CA, Nov. 2010

A Lower Bound on the Estimator Variance for the Sparse Linear Model

Sebastian Schmutzhard[1], *Alexander Jung*[2], *Franz Hlawatsch*[2], *Zvika Ben-Haim*[3], and *Yonina C. Eldar*[3]

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

[2]Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology
A-1040 Vienna, Austria; e-mail: {ajung, fhlawats}@nt.tuwien.ac.at

[3]Technion—Israel Institute of Technology
Haifa 32000, Israel; e-mail: {zvikabh@tx, yonina@ee}.technion.ac.il

***Abstract*—We study the performance of estimators of a sparse nonrandom vector based on an observation which is linearly transformed and corrupted by white Gaussian noise. Using the framework of reproducing kernel Hilbert spaces, we derive a new lower bound on the estimator variance for a given differentiable bias function (including the unbiased case) and an almost arbitrary transformation matrix (including the underdetermined case considered in compressed sensing theory). For the special case of a sparse vector corrupted by white Gaussian noise—i.e., without a linear transformation—and unbiased estimation, our lower bound improves on a previously proposed bound.**

***Index Terms*—Sparsity, parameter estimation, sparse linear model, denoising, variance bound, reproducing kernel Hilbert space, RKHS.**

I. Introduction

We study the problem of estimating a nonrandom parameter vector $\mathbf{x}\in\mathbb{R}^N$ which is sparse, i.e., at most S of its entries are nonzero, where $1\leq S<N$ (typically $S\ll N$). We thus have

$$\mathbf{x} \in \mathcal{X}_S \triangleq \left\{\mathbf{x}'\in\mathbb{R}^N \,\middle|\, \|\mathbf{x}'\|_0 \leq S\right\}, \tag{1}$$

where $\|\mathbf{x}\|_0$ denotes the number of nonzero entries of $\mathbf{x}$. While the sparsity degree S is assumed to be known, the set of positions of the nonzero entries of $\mathbf{x}$ (denoted by $\operatorname{supp}(\mathbf{x})$) is unknown. The estimation of $\mathbf{x}$ is based on the observed vector $\mathbf{y}\in\mathbb{R}^M$ given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}\,, \tag{2}$$

with a known system matrix $\mathbf{H}\in\mathbb{R}^{M\times N}$ and white Gaussian noise $\mathbf{n}\sim\mathcal{N}(\mathbf{0},\sigma^2\mathbf{I})$ with known variance $\sigma^2>0$. The matrix $\mathbf{H}$ is arbitrary except that it is assumed to satisfy the standard requirement [1]

$$\operatorname{spark}(\mathbf{H}) > 2S\,, \tag{3}$$

where $\operatorname{spark}(\mathbf{H})$ denotes the minimum number of linearly dependent columns of $\mathbf{H}$ [2]. The assumption (3) is reasonable, since otherwise the statistics of $\mathbf{y}$ may be equal for two different parameter vectors $\mathbf{x}_1,\mathbf{x}_2\in\mathcal{X}_S$. However, for our development, the less restrictive assumption $\operatorname{spark}(\mathbf{H})>S$ would suffice.

The observation model (2) together with (1) will be referred to as the *sparse linear model* (SLM). Note that we also allow $M<N$ (this case is relevant to compressed sensing methods [2], [3]); however, condition (3) implies that $M\geq 2S$. The case of correlated Gaussian noise $\mathbf{n}$ with a known nonsingular correlation matrix can be reduced to the SLM by means of a noise whitening transformation. An important special case of the SLM is given by $\mathbf{H}=\mathbf{I}$ (so that $M=N$), i.e.,

$$\mathbf{y} = \mathbf{x} + \mathbf{n}\,, \tag{4}$$

where again $\mathbf{x}\in\mathcal{X}_S$ and $\mathbf{n}\sim\mathcal{N}(\mathbf{0},\sigma^2\mathbf{I})$. This will be referred to as the *sparse signal in noise model* (SSNM).

Bounds on the estimation variance for the SLM have been studied previously. In particular, the Cramér–Rao bound (CRB) for the SLM was derived in [1]. For the SSNM (4), lower and upper bounds on the minimum variance of unbiased estimators were derived in [4]. A remarkable property of the lower bounds in [1] and [4] is the fact that they are discontinuous when passing from the case $\|\mathbf{x}\|_0=S$ to the case $\|\mathbf{x}\|_0<S$.

In this paper, we use the mathematical framework of *reproducing kernel Hilbert spaces* (RKHS) [5]–[7] to derive a novel lower variance bound for the SLM. The RKHS framework allows pleasing geometric interpretations of existing bounds, including the CRB, the Hammersley-Chapman-Robbins bound [8], and the Barankin bound [9]. The bound we derive here holds for estimators with a given differentiable bias function. For the SSNM, in particular, we obtain a lower bound for unbiased estimators which is tighter than the bounds in [4] and, moreover, everywhere continuous. As we will show, RKHS theory relates the bound for the SLM to that obtained for the linear model without a sparsity assumption. We note that the RKHS framework has been previously applied to estimation [6], [7] but, to the best of our knowledge, not to the SLM.

This paper is organized as follows. In Section II, we review some fundamentals of parameter estimation. Relevant elements of RKHS theory are summarized in Section III. In Section IV, we use RKHS theory to derive a lower variance bound for the SLM. Section V considers the special case of unbiased

This work was supported by the FWF under Grants S10602 (Signal and Information Representation) and S10603 (Statistical Inference) within the National Research Network SISE, by the Israel Science Foundation under Grant 1081/07, and by the European Commission under the FP7 Network of Excellence in Wireless COMmunications NEWCOM++ (contract no. 216715).

estimation for the SSNM. Section VI presents a numerical comparison of the new bound with the variance of two established estimation schemes.

II. Basic Concepts

We first review some basic concepts of parameter estimation [10], [11]. Let $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^N$ be the nonrandom parameter vector to be estimated (which is known to belong to the fixed parameter set $\mathcal{X}$), $\mathbf{y} \in \mathbb{R}^M$ the observed vector, and $f(\mathbf{y};\mathbf{x})$ the probability density function (pdf) of $\mathbf{y}$, parameterized by $\mathbf{x}$. For the SLM, $\mathcal{X} = \mathcal{X}_S$ as defined in (1) and

$$f(\mathbf{y};\mathbf{x}) \,=\, \frac{1}{(2\pi\sigma^2)^{M/2}} \,\exp\!\left(-\,\frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2\right). \quad (5)$$

A. Minimum-Variance Estimators

The estimation error incurred by an estimator $\hat{\mathbf{x}}(\mathbf{y})$ can be quantified by the mean squared error (MSE) $\varepsilon(\hat{\mathbf{x}}(\cdot);\mathbf{x}) \triangleq \mathsf{E}_{\mathbf{x}}\big\{\|\hat{\mathbf{x}}(\mathbf{y})-\mathbf{x}\|_2^2\big\}$, where the notation $\mathsf{E}_{\mathbf{x}}\{\cdot\}$ indicates that the expectation is taken with respect to the pdf $f(\mathbf{y};\mathbf{x})$ parameterized by $\mathbf{x}$. Note that $\varepsilon(\hat{\mathbf{x}}(\cdot);\mathbf{x})$ depends on the true parameter value, $\mathbf{x}$. The MSE can be decomposed as

$$\varepsilon(\hat{\mathbf{x}}(\cdot);\mathbf{x}) \,=\, \|\mathbf{b}(\hat{\mathbf{x}}(\cdot);\mathbf{x})\|_2^2 + v(\hat{\mathbf{x}}(\cdot);\mathbf{x})\,, \quad (6)$$

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 $v(\hat{\mathbf{x}}(\cdot);\mathbf{x}) \triangleq \mathsf{E}_{\mathbf{x}}\big\{\big\|\hat{\mathbf{x}}(\mathbf{y}) - \mathsf{E}_{\mathbf{x}}\{\hat{\mathbf{x}}(\mathbf{y})\}\big\|_2^2\big\}$. A standard approach to defining an optimum estimator is to fix the bias, i.e., $\mathbf{b}(\hat{\mathbf{x}}(\cdot);\mathbf{x}) \overset{!}{=} \mathbf{c}(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}$, and minimize the variance $v(\hat{\mathbf{x}}(\cdot);\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}$ under this bias constraint. However, in many cases, such a "uniformly optimum" estimator does not exist. It is then natural to consider "locally optimum" estimators that minimize $v(\hat{\mathbf{x}}(\cdot);\mathbf{x}_0)$ only at a given parameter value $\mathbf{x} = \mathbf{x}_0 \in \mathcal{X}$. This approach is taken here. Note that it follows from (6) that once the bias is fixed, minimizing the variance is equivalent to minimizing the MSE $\varepsilon(\hat{\mathbf{x}}(\cdot);\mathbf{x}_0)$.

The bias constraint $\mathbf{b}(\hat{\mathbf{x}}(\cdot);\mathbf{x}) = \mathbf{c}(\mathbf{x})$ can be equivalently written as the mean constraint

$$\mathsf{E}_{\mathbf{x}}\{\hat{\mathbf{x}}(\mathbf{y})\} = \boldsymbol{\gamma}(\mathbf{x})\,, \quad \text{with } \boldsymbol{\gamma}(\mathbf{x}) \triangleq \mathbf{c}(\mathbf{x}) + \mathbf{x}\,.$$

Thus, we consider the constrained optimization problem

$$\hat{\mathbf{x}}_{\mathbf{x}_0}(\cdot) \,=\, \arg \min_{\hat{\mathbf{x}}(\cdot)\in\mathcal{B}_{\boldsymbol{\gamma}}} v(\hat{\mathbf{x}}(\cdot);\mathbf{x}_0)\,, \quad (7)$$

where $\mathcal{B}_{\boldsymbol{\gamma}} \triangleq \big\{\hat{\mathbf{x}}(\cdot)\big|\mathsf{E}_{\mathbf{x}}\{\hat{\mathbf{x}}(\mathbf{y})\} = \boldsymbol{\gamma}(\mathbf{x}), \forall\, \mathbf{x} \in \mathcal{X}\big\}$. The minimum variance achieved by the locally optimum estimator $\hat{\mathbf{x}}_{\mathbf{x}_0}(\cdot)$ at $\mathbf{x}_0$ will be denoted as

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

This is also known as the *Barankin bound* (for the prescribed mean $\boldsymbol{\gamma}(\mathbf{x})$) [9]. Using RKHS theory, it can be shown that $\hat{\mathbf{x}}_{\mathbf{x}_0}(\cdot)$ exists, i.e., there exists a unique minimum in (7), provided that there exists at least one estimator with mean $\boldsymbol{\gamma}(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}$ and finite variance at $\mathbf{x}_0$ (see also Section III). For unbiased estimation, i.e., $\boldsymbol{\gamma}(\mathbf{x}) \equiv \mathbf{x}$, $\hat{\mathbf{x}}_{\mathbf{x}_0}(\cdot)$ is called a *locally minimum variance unbiased* (LMVU) estimator. Unfortunately, $V_{\boldsymbol{\gamma}}(\mathbf{x}_0)$ is difficult to compute in many cases, including the case of the SLM. Lower bounds on $V_{\boldsymbol{\gamma}}(\mathbf{x}_0)$ are, e.g., the CRB and the Hammersley-Chapman-Robbins bound [8].

Let x_k, $\hat{x}_k(\mathbf{y})$, and $\gamma_k(\mathbf{x})$ denote the kth entries of $\mathbf{x}$, $\hat{\mathbf{x}}(\mathbf{y})$, and $\boldsymbol{\gamma}(\mathbf{x})$, respectively. We have $v(\hat{\mathbf{x}}(\cdot);\mathbf{x}) = \sum_{k=1}^{N} v(\hat{x}_k(\cdot);\mathbf{x})$ with $v(\hat{x}_k(\cdot);\mathbf{x}) \triangleq \mathsf{E}_{\mathbf{x}}\big\{\big[\hat{x}_k(\mathbf{y}) - \mathsf{E}_{\mathbf{x}}\{\hat{x}_k(\mathbf{y})\}\big]^2\big\}$. Thus, (7) is equivalent to the N scalar optimization problems

$$\hat{x}_{\mathbf{x}_0,k}(\cdot) \,=\, \arg \min_{\hat{x}_k(\cdot)\in\mathcal{B}_{\gamma_k}} v(\hat{x}_k(\cdot);\mathbf{x}_0)\,, \quad k = 1,\ldots,N\,, \quad (8)$$

where $\mathcal{B}_{\gamma_k} \triangleq \big\{\hat{x}(\cdot)\big|\mathsf{E}_{\mathbf{x}}\{\hat{x}(\mathbf{y})\} = \gamma_k(\mathbf{x}), \forall\, \mathbf{x} \in \mathcal{X}\big\}$. The minimum variance achieved by $\hat{x}_{\mathbf{x}_0,k}(\cdot)$ at $\mathbf{x}_0$ is denoted as

$$V_{\gamma_k}(\mathbf{x}_0) \,\triangleq\, v(\hat{x}_{\mathbf{x}_0,k}(\cdot);\mathbf{x}_0) \,=\, \min_{\hat{x}_k(\cdot)\in\mathcal{B}_{\gamma_k}} v(\hat{x}_k(\cdot);\mathbf{x}_0)\,. \quad (9)$$

B. CRB of the Linear Gaussian Model

In our further development, we will make use of the CRB for the *linear Gaussian model* (LGM) defined by

$$\mathbf{z} \,=\, \mathbf{A}\mathbf{s} + \mathbf{n}\,, \quad (10)$$

with the nonrandom parameter $\mathbf{s} \in \mathbb{R}^S$ (not assumed sparse), the observation $\mathbf{z} \in \mathbb{R}^M$, the known matrix $\mathbf{A} \in \mathbb{R}^{M\times S}$, and white Gaussian noise $\mathbf{n} \sim \mathcal{N}(\mathbf{0},\sigma^2\mathbf{I})$. As before, we assume that $M \geq 2S$; furthermore, we assume that $\mathbf{A}$ has full column rank, i.e., $\mathbf{A}^T\mathbf{A} \in \mathbb{R}^{S\times S}$ is nonsingular. The relationship of this model with the SLM, as well as the different notation and different dimension (S instead of N), will become clear in Section IV.

Consider estimators $\hat{s}_k(\mathbf{z})$ of the kth parameter component s_k whose bias is equal to some prescribed differentiable function $\tilde{c}_k(\mathbf{s})$, i.e., $b(\hat{s}_k(\cdot);\mathbf{s}) = \tilde{c}_k(\mathbf{s})$ or equivalently $\mathsf{E}_{\mathbf{s}}\big\{\hat{s}_k(\mathbf{z})\big\} = \tilde{\gamma}_k(\mathbf{s})$ with $\tilde{\gamma}_k(\mathbf{s}) \triangleq \tilde{c}_k(\mathbf{s}) + s_k$, for all $\mathbf{s} \in \mathbb{R}^S$. Let $V^{\mathrm{LGM}}_{\tilde{\gamma}_k}(\mathbf{s}_0)$ denote the minimum variance achievable by such estimators at a given true parameter $\mathbf{s}_0$. The CRB $C^{\mathrm{LGM}}_{\tilde{\gamma}_k}(\mathbf{s}_0)$ is the following lower bound on the minimum variance [10]:

$$V^{\mathrm{LGM}}_{\tilde{\gamma}_k}(\mathbf{s}_0) \,\geq\, C^{\mathrm{LGM}}_{\tilde{\gamma}_k}(\mathbf{s}_0) \,\triangleq\, \sigma^2\, \tilde{\mathbf{r}}_k^T(\mathbf{s}_0)\big(\mathbf{A}^T\mathbf{A}\big)^{-1}\tilde{\mathbf{r}}_k(\mathbf{s}_0)\,, \quad (11)$$

where $\tilde{\mathbf{r}}_k(\mathbf{s}) \triangleq \partial\tilde{\gamma}_k(\mathbf{s})/\partial\mathbf{s}$, i.e., $\tilde{\mathbf{r}}_k(\mathbf{s})$ is the vector of dimension S whose lth entry is $\partial\tilde{\gamma}_k(\mathbf{s})/\partial s_l$. We note that $V^{\mathrm{LGM}}_{\tilde{\gamma}_k}(\mathbf{s}_0) = C^{\mathrm{LGM}}_{\tilde{\gamma}_k}(\mathbf{s}_0)$ if $\tilde{\gamma}_k(\mathbf{s})$ is an affine function of $\mathbf{s}$. In particular, this includes the unbiased case ($\tilde{\gamma}_k(\mathbf{s}) = s_k$).

III. The RKHS Framework

In this section, we review some RKHS fundamentals which will provide a basis for our further development. Consider a set $\mathcal{X}$ (not necessarily a linear space) and a positive semidefinite[1] "kernel" function $R(\mathbf{x},\mathbf{x}') \colon \mathcal{X}\times\mathcal{X} \to \mathbb{R}$. For each fixed $\mathbf{x}' \in \mathcal{X}$, the function $f_{\mathbf{x}'}(\mathbf{x}) \triangleq R(\mathbf{x},\mathbf{x}')$ maps $\mathcal{X}$ into $\mathbb{R}$. The RKHS $\mathcal{H}(R)$ is a Hilbert space of functions $f : \mathcal{X} \to \mathbb{R}$ which is defined as the closure of the linear span of the set of functions $\{f_{\mathbf{x}'}(\mathbf{x}) = R(\mathbf{x},\mathbf{x}')\}_{\mathbf{x}'\in\mathcal{X}}$. This closure is taken with respect to the topology given by the scalar product $\langle\cdot,\cdot\rangle_{\mathcal{H}(R)}$ which is defined via the *reproducing property* [5]

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

This relation holds for all $f \in \mathcal{H}(R)$ and $\mathbf{x}' \in \mathcal{X}$. The associated norm is given by $\|f\|_{\mathcal{H}(R)} = \langle f, f\rangle^{1/2}_{\mathcal{H}(R)}$.

[1] That is, for any finite set $\{\mathbf{x}_k\}_{k=1,\ldots,P}$ with $\mathbf{x}_k \in \mathcal{X}$, the matrix $\mathbf{R} \in \mathbb{R}^{P\times P}$ with entries $(\mathbf{R})_{k,l} \triangleq R(\mathbf{x}_k,\mathbf{x}_l)$ is positive semidefinite.

We now consider the constrained optimization problem (8) for a given mean function $\gamma(\mathbf{x})$ (formerly denoted by $\gamma_k(\mathbf{x})$; we temporarily drop the subscript k for better readability). According to [6], [7], for certain pairs of parametrized pdf's $f(\mathbf{y};\mathbf{x})$ and parameter sets $\mathcal{X}$ (which include the Gaussian pdf in (5) together with the parameter set $\mathcal{X}_S$ in (1)), one can associate with this optimization problem an RKHS $\mathcal{H}(R_{\mathbf{x}_0})$ whose kernel $R_{\mathbf{x}_0}(\mathbf{x},\mathbf{x}')\colon \mathcal{X}\times\mathcal{X}\rightarrow\mathbb{R}$ is given by

$$R_{\mathbf{x}_0}(\mathbf{x},\mathbf{x}') \triangleq \mathsf{E}_{\mathbf{x}_0}\bigg\{\frac{f(\mathbf{y};\mathbf{x})}{f(\mathbf{y};\mathbf{x}_0)}\,\frac{f(\mathbf{y};\mathbf{x}')}{f(\mathbf{y};\mathbf{x}_0)}\bigg\} = \int_{\mathbb{R}^M}\frac{f(\mathbf{y};\mathbf{x})\,f(\mathbf{y};\mathbf{x}')}{f(\mathbf{y};\mathbf{x}_0)}\,d\mathbf{y}\,.$$

It can be shown [6], [7] that $\gamma(\mathbf{x})\in\mathcal{H}(R_{\mathbf{x}_0})$ if and only if there exists at least one estimator with mean $\gamma(\mathbf{x})$ for all $\mathbf{x}$ and finite variance at $\mathbf{x}_0$. Furthermore, under this condition, the minimum variance $V_\gamma(\mathbf{x}_0)$ in (9) is finite and allows the following expression involving the norm $\|\gamma\|_{\mathcal{H}(R_{\mathbf{x}_0})}$:

$$V_\gamma(\mathbf{x}_0) \,=\, \|\gamma\|^2_{\mathcal{H}(R_{\mathbf{x}_0})} - \gamma^2(\mathbf{x}_0)\,. \tag{12}$$

This is an RKHS formulation of the Barankin bound. Unfortunately, the norm $\|\gamma\|_{\mathcal{H}(R_{\mathbf{x}_0})}$ is often difficult to compute.

For the SLM in (1), (2), (5), $\mathcal{X}=\mathcal{X}_S$; the kernel here is a mapping $\mathcal{X}_S\times\mathcal{X}_S\rightarrow\mathbb{R}$ which is easily shown to be given by

$$R_{\mathbf{x}_0}(\mathbf{x},\mathbf{x}') \,=\, \exp\!\bigg(\frac{1}{\sigma^2}(\mathbf{x}-\mathbf{x}_0)^T\mathbf{H}^T\mathbf{H}\,(\mathbf{x}'-\mathbf{x}_0)\bigg), \tag{13}$$

where $\mathbf{x}_0\in\mathcal{X}_S$. An RKHS can also be defined for the LGM in (10). Here, $\mathcal{X}=\mathbb{R}^S$, and the kernel $R^{\text{LGM}}_{\mathbf{s}_0}(\mathbf{s},\mathbf{s}')$ with $\mathbf{s}_0\in\mathbb{R}^S$ is a mapping $\mathbb{R}^S\times\mathbb{R}^S\rightarrow\mathbb{R}$ given by

$$R^{\text{LGM}}_{\mathbf{s}_0}(\mathbf{s},\mathbf{s}') \,=\, \exp\!\bigg(\frac{1}{\sigma^2}(\mathbf{s}-\mathbf{s}_0)^T\mathbf{A}^T\mathbf{A}\,(\mathbf{s}'-\mathbf{s}_0)\bigg). \tag{14}$$

Note that these kernels differ in their domain, which is $\mathcal{X}_S\times\mathcal{X}_S$ for $R_{\mathbf{x}_0}(\mathbf{x},\mathbf{x}')$ and $\mathbb{R}^S\times\mathbb{R}^S$ for $R^{\text{LGM}}_{\mathbf{s}_0}(\mathbf{s},\mathbf{s}')$.

IV. A Lower Bound on the Estimator Variance

We now continue our treatment of the SLM estimation problem. In what follows, $V_\gamma(\mathbf{x}_0)$ will be understood to denote the bias-constrained minimum variance (9) *specifically for the SLM*. This means, in particular, that $\mathcal{X}=\mathcal{X}_S$, and hence the set of admissible estimators is given by

$$\mathcal{B}_\gamma = \big\{\hat{x}(\cdot)\,\big|\,\mathsf{E}_{\mathbf{x}}\{\hat{x}(\mathbf{y})\}=\gamma(\mathbf{x}),\forall\,\mathbf{x}\in\mathcal{X}_S\big\}\,. \tag{15}$$

We will next derive a lower bound on $V_\gamma(\mathbf{x}_0)$.

A. Relaxing the Bias Constraint

The first step in this derivation is to relax the bias constraint $\hat{x}(\cdot)\in\mathcal{B}_\gamma$. Let $\mathcal{K}\triangleq\{k_1,\ldots,k_S\}$ be a fixed set of S different indices $k_i\in\{1,\ldots,N\}$ (not related to $\text{supp}(\mathbf{x}_0)$), and let

$$\mathcal{X}_S^{\mathcal{K}} \triangleq \{\mathbf{x}\in\mathcal{X}_S\,|\,\text{supp}(\mathbf{x})\subseteq\mathcal{K}\}\,.$$

Clearly, $\mathcal{X}_S^{\mathcal{K}}\subseteq\mathcal{X}_S$; however, contrary to $\mathcal{X}_S$, $\mathcal{X}_S^{\mathcal{K}}$ is a linear subspace of $\mathbb{R}^N$. Let $\mathcal{B}_\gamma^{\mathcal{K}}$ be the set of all estimators with mean $\gamma(\mathbf{x})$ for all $\mathbf{x}\in\mathcal{X}_S^{\mathcal{K}}$ (but not necessarily for all $\mathbf{x}\in\mathcal{X}_S$), i.e.,

$$\mathcal{B}_\gamma^{\mathcal{K}} \triangleq \big\{\hat{x}(\cdot)\,\big|\,\mathsf{E}_{\mathbf{x}}\{\hat{x}(\mathbf{y})\}=\gamma(\mathbf{x}),\forall\,\mathbf{x}\in\mathcal{X}_S^{\mathcal{K}}\big\}\,.$$

Comparing with (15), we see that $\mathcal{B}_\gamma^{\mathcal{K}}\supseteq\mathcal{B}_\gamma$.

Let us now consider the minimum variance among all estimators in $\mathcal{B}_\gamma^{\mathcal{K}}$, i.e.,

$$V_\gamma^{\mathcal{K}}(\mathbf{x}_0) \,\triangleq \min_{\hat{x}(\cdot)\in\mathcal{B}_\gamma^{\mathcal{K}}} v(\hat{x}(\cdot);\mathbf{x}_0)\,. \tag{16}$$

Because $\hat{x}(\cdot)\in\mathcal{B}_\gamma^{\mathcal{K}}$ is a less restrictive constraint than $\hat{x}(\cdot)\in\mathcal{B}_\gamma$ used in the definition of $V_\gamma(\mathbf{x}_0)$, we have

$$V_\gamma(\mathbf{x}_0) \geq V_\gamma^{\mathcal{K}}(\mathbf{x}_0)\,, \tag{17}$$

i.e., $V_\gamma^{\mathcal{K}}(\mathbf{x}_0)$ is a lower bound on $V_\gamma(\mathbf{x}_0)$. A closed-form expression of $V_\gamma^{\mathcal{K}}(\mathbf{x}_0)$ for general mean functions $\gamma(\mathbf{x})$ is difficult to obtain. Therefore, we will use RKHS theory to derive a lower bound on $V_\gamma^{\mathcal{K}}(\mathbf{x}_0)$.

B. Two Isometric RKHSs

An RKHS for the SLM can also be defined on $\mathcal{X}_S^{\mathcal{K}}$, using a kernel $R^{\mathcal{K}}_{\mathbf{x}_0}\colon\mathcal{X}_S^{\mathcal{K}}\times\mathcal{X}_S^{\mathcal{K}}\rightarrow\mathbb{R}$ that is given by the right-hand side of (13) but whose arguments $\mathbf{x},\mathbf{x}'$ are assumed to be in $\mathcal{X}_S^{\mathcal{K}}$ and not just in $\mathcal{X}_S$ (note, however, that $\mathbf{x}_0\notin\mathcal{X}_S^{\mathcal{K}}$ in general). This RKHS will be denoted $\mathcal{H}(R^{\mathcal{K}}_{\mathbf{x}_0})$. The minimum variance $V_\gamma^{\mathcal{K}}(\mathbf{x}_0)$ in (16) can then be expressed as (cf. (12))

$$V_\gamma^{\mathcal{K}}(\mathbf{x}_0) \,=\, \|\gamma\|^2_{\mathcal{H}(R^{\mathcal{K}}_{\mathbf{x}_0})} - \gamma^2(\mathbf{x}_0)\,. \tag{18}$$

In order to develop this expression, we define some notation. Consider an index set $\mathcal{I}=\{k_1,\ldots,k_{|\mathcal{I}|}\}\subseteq\{1,\ldots,N\}$. We denote by $\mathbf{H}_{\mathcal{I}}\in\mathbb{R}^{M\times|\mathcal{I}|}$ the submatrix of our matrix $\mathbf{H}\in\mathbb{R}^{M\times N}$ whose ith column is given by the k_ith column of $\mathbf{H}$. Furthermore, for a vector $\mathbf{x}\in\mathbb{R}^N$, we denote by $\mathbf{x}^{\mathcal{I}}\in\mathbb{R}^{|\mathcal{I}|}$ the subvector whose ith entry is the k_ith entry of $\mathbf{x}$.

We now introduce a second RKHS. Consider the LGM in (10) with matrix $\mathbf{A}=\mathbf{H}_{\mathcal{K}}\in\mathbb{R}^{M\times S}$, and let $\mathcal{H}(R^{\text{LGM}}_{\mathbf{s}_0})$ with $\mathbf{s}_0\in\mathbb{R}^S$ denote the RKHS for that LGM as defined by the kernel $R^{\text{LGM}}_{\mathbf{s}_0}\colon\mathbb{R}^S\times\mathbb{R}^S\rightarrow\mathbb{R}$ in (14). Exploiting the linear-subspace structure of $\mathcal{X}_S^{\mathcal{K}}$, it can be shown that the RKHS $\mathcal{H}(R^{\mathcal{K}}_{\mathbf{x}_0})$ for a given $\mathbf{x}_0$ is *isometric* to $\mathcal{H}(R^{\text{LGM}}_{\mathbf{s}_0})$ with $\mathbf{s}_0$ chosen as

$$\mathbf{s}_0 \,=\, \mathbf{H}_{\mathcal{K}}^{\dagger}\mathbf{H}\,\mathbf{x}_0\,. \tag{19}$$

Here, $\mathbf{H}_{\mathcal{K}}^{\dagger}\triangleq(\mathbf{H}_{\mathcal{K}}^T\mathbf{H}_{\mathcal{K}})^{-1}\mathbf{H}_{\mathcal{K}}^T\in\mathbb{R}^{S\times M}$ is the pseudo-inverse of $\mathbf{H}_{\mathcal{K}}$ (recall that $M\geq 2S$, and note that $(\mathbf{H}_{\mathcal{K}}^T\mathbf{H}_{\mathcal{K}})^{-1}$ is guaranteed to exist because of our assumption (3)). More specifically, the isometry $\mathsf{J}:\mathcal{H}(R^{\mathcal{K}}_{\mathbf{x}_0})\rightarrow\mathcal{H}(R^{\text{LGM}}_{\mathbf{s}_0})$ mapping each $f\in\mathcal{H}(R^{\mathcal{K}}_{\mathbf{x}_0})$ to an $\tilde{f}\in\mathcal{H}(R^{\text{LGM}}_{\mathbf{s}_0})$ is given by

$$\mathsf{J}\{f(\mathbf{x})\} = \tilde{f}(\mathbf{x}^{\mathcal{K}}) = \beta_{\mathbf{x}_0}f(\mathbf{x})\,,\quad \mathbf{x}\in\mathcal{X}_S^{\mathcal{K}}\,, \tag{20}$$

where

$$\beta_{\mathbf{x}_0} \triangleq \exp\!\bigg(-\frac{1}{2\sigma^2}\big\|(\mathbf{I}-\mathbf{P}_{\mathcal{K}})\,\mathbf{H}\mathbf{x}_0\big\|_2^2\bigg). \tag{21}$$

Here, $\mathbf{P}_{\mathcal{K}}\triangleq\mathbf{H}_{\mathcal{K}}\mathbf{H}_{\mathcal{K}}^{\dagger}$ is the orthogonal projection matrix on the range of $\mathbf{H}_{\mathcal{K}}$. The factor $\beta_{\mathbf{x}_0}$ can be interpreted as a measure of the distance between the point $\mathbf{H}\mathbf{x}_0$ and the subspace $\mathcal{X}_S^{\mathcal{K}}$.

We can write (20) as

$$\tilde{f}(\mathbf{s}) = \beta_{\mathbf{x}_0} f(\mathbf{x}(\mathbf{s}))\,, \quad \mathbf{s}\in\mathbb{R}^S,$$

where $\mathbf{x}(\mathbf{s})$ denotes the $\mathbf{x}\in\mathcal{X}_S^{\mathcal{K}}$ for which $\mathbf{x}^{\mathcal{K}}=\mathbf{s}$ (i.e., the S entries of $\mathbf{s}$ appear in $\mathbf{x}(\mathbf{s})$ at the appropriate positions within $\mathcal{K}$, and the $N-S$ remaining entries of $\mathbf{x}(\mathbf{s})$ are zero).

Consider now the image of $\gamma(\mathbf{x})$ under the mapping J,

$$\tilde{\gamma}(\mathbf{s}) \triangleq \mathsf{J}\{\gamma(\mathbf{x})\} = \beta_{\mathbf{x}_0}\gamma(\mathbf{x}(\mathbf{s}))\,, \quad \mathbf{s}\in\mathbb{R}^S. \tag{22}$$

Since J is an isometry, we have $\|\tilde{\gamma}\|^2_{\mathcal{H}(R^{\mathrm{LGM}}_{\mathbf{s}_0})} = \|\gamma\|^2_{\mathcal{H}(R^{\mathcal{K}}_{\mathbf{x}_0})}$. Combining this identity with (18), we obtain

$$V_\gamma^{\mathcal{K}}(\mathbf{x}_0) \,=\, \|\tilde{\gamma}\|^2_{\mathcal{H}(R^{\mathrm{LGM}}_{\mathbf{s}_0})} - \gamma^2(\mathbf{x}_0)\,. \tag{23}$$

C. Lower Bound on $V_\gamma^{\mathcal{K}}(\mathbf{x}_0)$

We will now use expression (23) to derive a lower bound on $V_\gamma^{\mathcal{K}}(\mathbf{x}_0)$ in terms of the CRB for the LGM in (11). Consider the minimum estimator variance for the LGM under the constraint of the prescribed mean function $\tilde{\gamma}(\mathbf{s})$, $V_{\tilde{\gamma}}^{\mathrm{LGM}}(\mathbf{s}_0)$, still for $\mathbf{A}=\mathbf{H}_{\mathcal{K}}$ and for $\mathbf{s}_0$ given by (19). We have (cf. (12))

$$V_{\tilde{\gamma}}^{\mathrm{LGM}}(\mathbf{s}_0) \,=\, \|\tilde{\gamma}\|^2_{\mathcal{H}(R^{\mathrm{LGM}}_{\mathbf{s}_0})} - \tilde{\gamma}^2(\mathbf{s}_0)\,.$$

Combining with (23), we obtain the relation

$$V_\gamma^{\mathcal{K}}(\mathbf{x}_0) \,=\, V_{\tilde{\gamma}}^{\mathrm{LGM}}(\mathbf{s}_0) + \tilde{\gamma}^2(\mathbf{s}_0) - \gamma^2(\mathbf{x}_0)\,.$$

Using the CRB $V_{\tilde{\gamma}}^{\mathrm{LGM}}(\mathbf{s}_0) \geq C_{\tilde{\gamma}}^{\mathrm{LGM}}(\mathbf{s}_0)$ (see (11)) yields

$$V_\gamma^{\mathcal{K}}(\mathbf{x}_0) \geq L_\gamma^{\mathcal{K}}(\mathbf{x}_0)\,, \tag{24}$$

with

$$L_\gamma^{\mathcal{K}}(\mathbf{x}_0) \,\triangleq\, C_{\tilde{\gamma}}^{\mathrm{LGM}}(\mathbf{s}_0) + \tilde{\gamma}^2(\mathbf{s}_0) - \gamma^2(\mathbf{x}_0)\,. \tag{25}$$

Finally, using (22) and the implied CRB relation $C_{\tilde{\gamma}}^{\mathrm{LGM}}(\mathbf{s}_0) = \beta^2_{\mathbf{x}_0} C^{\mathrm{LGM}}_{\gamma(\mathbf{x}(\mathbf{s}))}(\mathbf{s}_0)$, the lower bound (25) can be reformulated as

$$L_\gamma^{\mathcal{K}}(\mathbf{x}_0) \,=\, \beta^2_{\mathbf{x}_0}\big[C^{\mathrm{LGM}}_{\gamma(\mathbf{x}(\mathbf{s}))}(\mathbf{s}_0) + \gamma^2(\mathbf{x}(\mathbf{s}_0))\big] - \gamma^2(\mathbf{x}_0)\,. \tag{26}$$

Here, $C^{\mathrm{LGM}}_{\gamma(\mathbf{x}(\mathbf{s}))}(\mathbf{s}_0)$ denotes the CRB for prescribed mean function $\gamma'(\mathbf{s}) = \gamma(\mathbf{x}(\mathbf{s}))$, which is given by (see (11))

$$C^{\mathrm{LGM}}_{\gamma(\mathbf{x}(\mathbf{s}))}(\mathbf{s}_0) \,=\, \sigma^2\,\mathbf{r}^T(\mathbf{s}_0)\big(\mathbf{H}_{\mathcal{K}}^T\mathbf{H}_{\mathcal{K}}\big)^{-1}\mathbf{r}(\mathbf{s}_0)\,, \tag{27}$$

where $\mathbf{r}(\mathbf{s}) \triangleq \partial\gamma(\mathbf{x}(\mathbf{s}))/\partial\mathbf{s}$ and $\mathbf{s}_0$ is related to $\mathbf{x}_0$ via (19).

To summarize, we have the following chain of lower bounds on the bias-constrained variance at $\mathbf{x}_0$:

$$v(\hat{x}(\cdot);\mathbf{x}_0) \overset{(9)}{\geq} V_\gamma(\mathbf{x}_0) \overset{(17)}{\geq} V_\gamma^{\mathcal{K}}(\mathbf{x}_0) \overset{(24)}{\geq} L_\gamma^{\mathcal{K}}(\mathbf{x}_0)\,. \tag{28}$$

While $L_\gamma^{\mathcal{K}}(\mathbf{x}_0)$ is the loosest of these bounds, it is attractive because of its closed-form expression in (26) (together with (27) and (19)). We note that the inequality (24) becomes an equality if $\gamma(\mathbf{x})$ is the restriction of an affine function of $\mathbf{x}$ to $\mathcal{X}_S$ (note that $\gamma(\mathbf{x})$ is only defined for $\mathbf{x}\in\mathcal{X}_S$). In particular, this includes the unbiased case ($\gamma(\mathbf{x})\equiv x$).

Recalling that $v(\hat{\mathbf{x}}(\cdot);\mathbf{x}_0) = \sum_{k=1}^N v(\hat{x}_k(\cdot);\mathbf{x}_0)$ (we now reintroduce the subscript k), a lower bound on $v(\hat{\mathbf{x}}(\cdot);\mathbf{x}_0)$ is obtained from (28) as $v(\hat{\mathbf{x}}(\cdot);\mathbf{x}_0) \geq \sum_{k=1}^N L_{\gamma_k}^{\mathcal{K}_k}(\mathbf{x}_0)$. For a high lower bound, the index sets $\mathcal{K}_k$ should in general be chosen such that the respective factors $\beta^2_{\mathbf{x}_0,k}$ in (26) are large. (This means that the "distances" between $\mathbf{H}\mathbf{x}_0$ and $\mathcal{X}_S^{\mathcal{K}_k}$ are small, see (21).) Formally using the optimum $\mathcal{K}_k$ for each k, we arrive at the main result of this paper.

Theorem. *Let $\hat{\mathbf{x}}(\cdot)$ be an estimator for the SLM* (2), (1) *whose mean equals $\boldsymbol{\gamma}(\mathbf{x})$ for all $\mathbf{x}\in\mathcal{X}_S$. Then the variance of $\hat{\mathbf{x}}(\cdot)$ at a given parameter vector $\mathbf{x}=\mathbf{x}_0\in\mathcal{X}_S$ satisfies*

$$v(\hat{\mathbf{x}}(\cdot);\mathbf{x}_0) \geq \sum_{k=1}^N L^*_{\gamma_k}(\mathbf{x}_0)\,, \tag{29}$$

where $L^*_{\gamma_k}(\mathbf{x}_0) \triangleq \max_{\mathcal{K}_k:|\mathcal{K}_k|=S} L_{\gamma_k}^{\mathcal{K}_k}(\mathbf{x}_0)$, with $L_{\gamma_k}^{\mathcal{K}_k}(\mathbf{x}_0)$ given by (26) *together with* (27) *and* (19).

V. Special Case: Unbiased Estimation for the SSNM

The SSNM in (4) is a special case of the SLM with $\mathbf{H}=\mathbf{I}$. We now consider unbiased estimation (i.e., $\boldsymbol{\gamma}(\mathbf{x})\equiv\mathbf{x}$) for the SSNM. Since an unbiased estimator with uniformly minimum variance does not exist [4], we are interested in a lower variance bound at a fixed $\mathbf{x}_0\in\mathcal{X}_S$. We denote by $\xi(\mathbf{x}_0)$ and $j(\mathbf{x}_0)$ the value and index, respectively, of the S-largest (in magnitude) entry of $\mathbf{x}_0$; note that this is the smallest (in magnitude) nonzero entry of $\mathbf{x}_0$ if $\|\mathbf{x}_0\|_0=S$, and zero if $\|\mathbf{x}_0\|_0<S$.

Consider an unbiased estimator $\hat{x}_k(\cdot)$. For $k\in\mathrm{supp}(\mathbf{x}_0)$, using the lower bound $L_{\gamma_k}^{\mathcal{K}_k}(\mathbf{x}_0)$ in (26) with any index set $\mathcal{K}_k$ of size $|\mathcal{K}_k|=S$ such that $\mathrm{supp}(\mathbf{x}_0)\subseteq\mathcal{K}_k$, one can show that

$$v(\hat{x}_k(\cdot);\mathbf{x}_0) \geq \sigma^2, \quad k\in\mathrm{supp}(\mathbf{x}_0)\,. \tag{30}$$

This bound is actually the minimum variance (i.e., the variance of the LMVU estimator) since it is achieved by the specific unbiased estimator $\hat{x}_k(\mathbf{y})=y_k$ (which is the LMVU estimator for $k\in\mathrm{supp}(\mathbf{x}_0)$). On the other hand, for $k\notin\mathrm{supp}(\mathbf{x}_0)$, the lower bound $L_{\gamma_k}^{\mathcal{K}_k}(\mathbf{x}_0)$ with $\mathcal{K}_k = \big(\mathrm{supp}(\mathbf{x}_0)\backslash\{j(\mathbf{x}_0)\}\big)\cup\{k\}$ can be shown to lead to the inequality

$$v(\hat{x}_k(\cdot);\mathbf{x}_0) \geq \sigma^2 e^{-\xi^2(\mathbf{x}_0)/\sigma^2}, \quad k\notin\mathrm{supp}(\mathbf{x}_0)\,. \tag{31}$$

Combining (30) and (31), a lower bound on the overall variance $v(\hat{\mathbf{x}}(\cdot);\mathbf{x}_0) = \sum_{k=1}^N v(\hat{x}_k(\cdot);\mathbf{x}_0)$ is obtained as

$$v(\hat{\mathbf{x}}(\cdot);\mathbf{x}_0) \,\geq \sum_{k\in\mathrm{supp}(\mathbf{x}_0)}\sigma^2 + \sum_{k\notin\mathrm{supp}(\mathbf{x}_0)}\sigma^2 e^{-\xi^2(\mathbf{x}_0)/\sigma^2}. \tag{32}$$

Thus, noting that $v(\hat{\mathbf{x}}(\cdot);\mathbf{x}_0) = \varepsilon(\hat{\mathbf{x}}(\cdot);\mathbf{x}_0)$ for unbiased estimators, we arrive at the following result.

Corollary. *Let $\hat{\mathbf{x}}(\cdot)$ be an unbiased estimator for the SSNM in* (4). *Then the MSE of $\hat{\mathbf{x}}(\cdot)$ at a given $\mathbf{x}=\mathbf{x}_0\in\mathcal{X}_S$ satisfies*

$$\varepsilon(\hat{\mathbf{x}}(\cdot);\mathbf{x}_0) \,\geq\, \big[S + (N-S)\,e^{-\xi^2(\mathbf{x}_0)/\sigma^2}\big]\,\sigma^2. \tag{33}$$

(Note that (32) is equivalent to (33) even when $|\mathrm{supp}(\mathbf{x}_0)|<S$, since in that case $\xi^2(\mathbf{x}_0)=0$.)

The lower bound (33) is tighter (i.e., higher) than the lower bound derived in [4]. Furthermore, in contrast to the bound in [4], it is a continuous function of $\mathbf{x}_0$. This fact is theoretically pleasing since (under mild technical conditions) the MSE of any estimator is a continuous function of $\mathbf{x}_0$ [11].

Let us consider the special case of $S=1$. Here, $\xi(\mathbf{x}_0)$ and $j(\mathbf{x}_0)$ are simply the value and index, respectively, of the single nonzero entry of $\mathbf{x}_0$. Using RKHS theory, one can show that the estimator $\hat{\mathbf{x}}(\cdot)$ given componentwise by

$$\hat{x}_k(\mathbf{y}) = \begin{cases} y_{j(\mathbf{x}_0)}\,, & k = j(\mathbf{x}_0) \\ \alpha(\mathbf{y};\mathbf{x}_0)\, y_k\,, & \text{else}\,, \end{cases}$$

with $\alpha(\mathbf{y};\mathbf{x}_0) \triangleq \exp\big(-\frac{1}{2\sigma^2}[2y_{j(\mathbf{x}_0)}\xi(\mathbf{x}_0) + \xi^2(\mathbf{x}_0)]\big)$, is the LMVU estimator at $\mathbf{x}_0$. That is, the estimator $\hat{\mathbf{x}}(\cdot)$ is unbiased and its MSE achieves the lower bound (33). This also means that the bound (33) is actually the minimum MSE (achieved by the LMVU estimator). While $\hat{\mathbf{x}}(\cdot)$ is not very practical since it explicitly involves the unknown true parameter $\mathbf{x}_0$, its existence demonstrates the tightness of the bound (33).

VI. Numerical Results

For the SSNM in (4), we will compute the lower variance bound $\sum_{k=1}^{N} L^{*}_{\gamma_k}(\mathbf{x}_0)$ (see (29)) and compare it with the variance of two established estimators, namely, the maximum likelihood (ML) estimator and the hard-thresholding (HT) estimator. The ML estimator is given by

$$\hat{\mathbf{x}}_{\text{ML}}(\mathbf{y}) \triangleq \arg\max_{\mathbf{x}' \in \mathcal{X}_S} f(\mathbf{y};\mathbf{x}') = \mathsf{P}_S(\mathbf{y})\,,$$

where the operator P_S retains the S largest (in magnitude) entries and zeros out all others. The HT estimator $\hat{\mathbf{x}}_{\text{HT}}(\mathbf{y})$ is given by

$$\hat{x}_{\text{HT},k}(\mathbf{y}) = \begin{cases} y_k\,, & |y_k| \geq T \\ 0\,, & \text{else}, \end{cases} \qquad (34)$$

where T is a fixed threshold.

For simplicity, we consider the SSNM for $S=1$. In this case, the bound (29) can be shown to be

$$v(\hat{\mathbf{x}}(\cdot);\mathbf{x}_0) \geq L^{\mathcal{K}_j}_{\gamma_j}(\mathbf{x}_0) + (N-1)\, e^{-\xi^2(\mathbf{x}_0)/\sigma^2} L^{\mathcal{K}_i}_{\gamma_i}(\mathbf{x}_0)\,, \quad (35)$$

where $j \triangleq j(\mathbf{x}_0)$, i is any index different from $j(\mathbf{x}_0)$ (it can be shown that all such indices equally maximize the lower bound), $\mathcal{K}_j \triangleq \{j(\mathbf{x}_0)\}$, and $\mathcal{K}_i \triangleq \{i\}$. (We note that (35) simplifies to (32) for the special case of an unbiased estimator.) Since we compare the bound (35) to the ML and HT estimators, $\boldsymbol{\gamma}(\mathbf{x})$ is set equal to the mean of the respective estimator (ML or HT).

For a numerical evaluation, we generated parameter vectors $\mathbf{x}_0$ with $N=5$, $S=1$, $j(\mathbf{x}_0)=1$, and different $\xi(\mathbf{x}_0)$. (The fixed choice $j(\mathbf{x}_0)=1$ is justified by the fact that neither the variances of the ML and HT estimators nor the corresponding variance bounds depend on $j(\mathbf{x}_0)$.) In Fig. 1, we plot the variances $v(\hat{\mathbf{x}}_{\text{ML}}(\cdot);\mathbf{x}_0)$ and $v(\hat{\mathbf{x}}_{\text{HT}}(\cdot);\mathbf{x}_0)$ (the latter for three different choices of T in (34)) along with the corresponding bounds (35), as a function of the signal-to-noise ratio (SNR) $\xi^2(\mathbf{x}_0)/\sigma^2$. It is seen that for SNR larger than about 18 dB, all variances and bounds are effectively equal (for the HT estimator, this is true if T is not too small). However, in the medium-SNR range, the

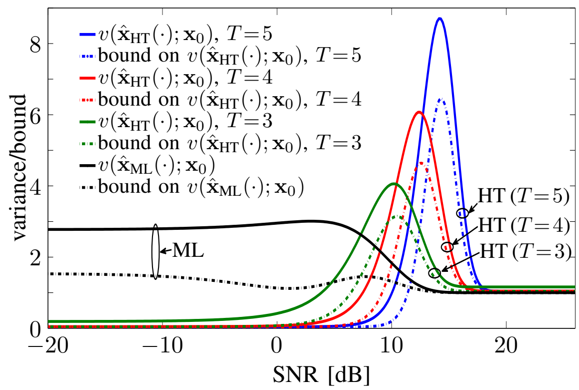

Figure 1. Variance of the ML and HT estimators and corresponding lower bounds versus the SNR $\xi^2(\mathbf{x}_0)/\sigma^2$, for the SSNM with $N=5$ and $S=1$.

variances of the ML and HT estimators are significantly higher than the corresponding lower bounds. We can conclude that there *might* exist estimators with the same mean as that of the ML or HT estimator but a smaller variance. Note, however, that a positive statement regarding the existence of such estimators cannot be based on our analysis.

VII. Conclusion

Using the mathematical framework of reproducing kernel Hilbert spaces, we derived a novel lower bound on the variance of estimators of a sparse vector under a bias constraint. The observed vector was assumed to be a linearly transformed and noisy version of the sparse vector to be estimated. This setup includes the underdetermined case relevant to compressed sensing. In the special case of unbiased estimation of a noise-corrupted sparse vector, our bound improves on the best known lower bound. A comparison with the variance of two established estimators showed that there might exist estimators with the same bias but a smaller variance.

References

[1] Z. Ben-Haim and Y. C. Eldar, "The Cramér–Rao bound for estimating a sparse parameter vector," *IEEE Trans. Signal Processing*, vol. 58, pp. 3384–3389, June 2010.

[2] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, pp. 2231–2242, Oct. 2004.

[3] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, April 2006.

[4] 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, pp. 3990–3993, March 2010.

[5] N. Aronszajn, "Theory of reproducing kernels," *Trans. Am. Math. Soc.*, vol. 68, pp. 337–404, May 1950.

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

[7] D. D. Duttweiler and T. Kailath, "RKHS approach to detection and estimation problems – Part V: Parameter estimation," *IEEE Trans. Inf. Theory*, vol. 19, pp. 29–37, Jan. 1973.

[8] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Trans. Inf. Theory*, vol. 36, pp. 1285–1301, Nov. 1990.

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

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

[11] E. L. Lehmann and G. Casella, *Theory of Point Estimation*. New York: Springer, 2003.