Minimum Variance Estimation of a Sparse Vector Within the Linear Gaussian Model: An RKHS Approach

Alexander Jung, *Member, IEEE*, Sebastian Schmutzhard, Franz Hlawatsch, *Fellow, IEEE*, Zvika Ben-Haim, *Member, IEEE*, and Yonina C. Eldar, *Fellow, IEEE*

***Abstract*— We consider minimum variance estimation within the *sparse linear Gaussian model* (SLGM). A sparse vector is to be estimated from a linearly transformed version embedded in Gaussian noise. Our analysis is based on the theory of reproducing kernel Hilbert spaces (RKHS). After a characterization of the RKHS associated with the SLGM, we derive a lower bound on the minimum variance achievable by estimators with a prescribed bias function, including the important special case of unbiased estimation. This bound is obtained via an orthogonal projection of the prescribed mean function onto a subspace of the RKHS associated with the SLGM. It provides an approximation to the minimum achievable variance (Barankin bound) that is tighter than any known bound. Our bound holds for an arbitrary system matrix, including the overdetermined and underdetermined cases. We specialize it to compressed sensing measurement matrices and express it in terms of the restricted isometry constant. For the special case of the SLGM given by the *sparse signal in noise model*, we derive closed-form expressions of the Barankin bound and of the corresponding locally minimum variance estimator. Finally, we compare our bound with the variance of several well-known estimators, namely, the maximum-likelihood estimator, the hard-thresholding estimator, and compressive reconstruction using orthogonal matching pursuit and approximate message passing.**

***Index Terms*— Sparsity, compressed sensing, unbiased estimation, denoising, RKHS, Cramér–Rao bound, Barankin bound, Hammersley–Chapman–Robbins bound, locally minimum variance unbiased estimator.**

Manuscript received April 17, 2013; revised January 31, 2014; accepted July 9, 2014. Date of publication August 8, 2014; date of current version September 11, 2014. This work was supported in part by the Austrian Science Fund under Grants S10602-N13 and S10603-N13 within the National Research Network Signal and Information Processing in Science and Engineering, in part by the Vienna Science and Technology Foundation under Grant MA 07-004 (SPORTS), in part by the Israel Science Foundation under Grant 1081/07, and in part by the European Commission through the FP7 Network of Excellence in Wireless Communications (NEWCOM++) under Contract 216715. This work was presented in part at the 44th Asilomar Conference on Signals, Systems and Computers, in 2010, and at the 2011 IEEE International Symposium on Information Theory.

A. Jung and F. Hlawatsch are with the Institute of Telecommunications, Vienna University of Technology, A-1040 Vienna, Austria (e-mail: ajung@nt.tuwien.ac.at; fhlawats@nt.tuwien.ac.at).

S. Schmutzhard is with the Numerical Harmonic Analysis Group, Faculty of Mathematics, University of Vienna, A-1090 Vienna, Austria (e-mail: sebastian.schmutzhard@univie.ac.at).

Z. Ben-Haim was with the Department of Electrical Engineering, Technion-Israel Institute of Technology, Haifa 32000, Israel. He is now with the Google Israel Research and Development Center, Haifa 31905, Israel (e-mail: zvika@google.com).

Y. C. Eldar is with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa 32000, Israel (e-mail: yonina@ee.technion.ac.il).

Communicated by G. Moustakides, Associate Editor for Detection and Estimation.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2014.2346508

I. Introduction

We study the problem of estimating the value $\mathbf{g}(\mathbf{x})$ of a known vector-valued function $\mathbf{g}(\cdot)$ evaluated at an unknown, nonrandom parameter vector $\mathbf{x} \in \mathbb{R}^N$. It is known that $\mathbf{x}$ is S-sparse, i.e., at most S of its entries are nonzero, where $S \in [N] \triangleq \{1, \ldots, N\}$ (typically $S \ll N$). While the sparsity degree S is known, the set of positions of the nonzero entries of $\mathbf{x}$, i.e., the *support* $\mathrm{supp}(\mathbf{x}) \subseteq [N]$, is unknown. The estimation of $\mathbf{g}(\mathbf{x})$ is based on an observed random vector $\mathbf{y} = \mathbf{H}\mathbf{x}+\mathbf{n} \in \mathbb{R}^M$, with a known system matrix $\mathbf{H} \in \mathbb{R}^{M\times N}$ and independent and identically distributed (i.i.d.) Gaussian noise $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ with known noise variance $\sigma^2 > 0$. We assume that any set of S columns of $\mathbf{H}$ is linearly independent.

We refer to the data model described above as the *sparse linear Gaussian model* (SLGM). This model is relevant, e.g., to sparse channel estimation [1], where the sparse parameter vector $\mathbf{x}$ represents the tap coefficients of a linear time-invariant channel and the system matrix $\mathbf{H}$ represents the training signal. More generally, the SLGM can be used to describe any type of sparse deconvolution [2]. The special case of the SLGM obtained for $\mathbf{H} = \mathbf{I}$ (so that $M = N$ and $\mathbf{y} = \mathbf{x} + \mathbf{n}$) will be termed the *sparse signal in noise model* (SSNM). The SSNM can be used, e.g., for sparse channel estimation [1] employing an orthogonal training signal [3] and for image denoising employing an orthonormal wavelet basis [4].

A fundamental question, to be considered in this work, is how to exploit the knowledge of the sparsity degree S. In contrast to compressed sensing (CS), where the sparsity is exploited for compression [5]–[8], here we investigate how much the sparsity assumption helps improve the accuracy of estimating $\mathbf{g}(\mathbf{x})$. Related questions have been previously addressed for the SLGM in [4] and [9]–[14]. In [9] and [10], bounds on the minimax risk and approximate minimax estimators whose worst-case risk is close to these bounds have been derived for the SLGM. An asymptotic analysis of minimax estimation for the SSNM has been given in the seminal work [4], [11]. In the context of minimum variance estimation (MVE), which is relevant to our present work, lower bounds on the minimum achievable variance for the SLGM have recently been derived. In particular, the Cramér–Rao bound (CRB) for the SLGM was developed and analyzed in [12] and [13]. In our previous work [14], we derived lower and upper bounds on the minimum achievable variance of unbiased estimators for the SSNM.

Here, we use the mathematical framework of *reproducing kernel Hilbert spaces* (RKHS) [15]–[17] to derive a lower bound on the minimum achievable variance that is tighter than previously proposed bounds. The contributions of this paper can be summarized as follows.

1) We characterize the RKHS associated with the SLGM. Using this characterization, we derive a new lower bound on the variance of estimators for the SLGM. Since this bound holds for any estimator with a prescribed mean function, it is also a lower bound on the minimum achievable variance (also known as *Barankin bound*) for the SLGM. The bound is tighter than any known bound, including the bounds presented in [12]–[14], and it has an appealing form in that it is a scaled version of the conventional CRB obtained for the nonsparse case [18], [19]. Furthermore, this bound holds for arbitrary system matrices $\mathbf{H}$, including the overdetermined and underdetermined cases. We note that our RKHS approach is quite different from the technique used in [14], which considered only the SSNM. By contrast, our approach applies to the general SLGM with arbitrary system matrix. Also, a shortcoming of the lower bounds presented in [12] and [14] is the fact that they exhibit a discontinuity when passing from the case $\|\mathbf{x}\|_0 = S$ (i.e., $\mathbf{x}$ has exactly S nonzero entries) to the case $\|\mathbf{x}\|_0 \leq S - 1$ (i.e., $\mathbf{x}$ has less than S nonzero entries). For unbiased estimation, our bound is a continuous function of $\mathbf{x}$ which exhibits a smooth transition between the two regimes given by $\|\mathbf{x}\|_0 = S$ and $\|\mathbf{x}\|_0 \leq S - 1$.
2) Based on the fact that the linear CS recovery problem is an instance of the SLGM, we specialize our lower bound to CS measurement matrices and express it in terms of the restricted isometry constant of these matrices.
3) For the SSNM, we derive the minimum achievable variance (Barankin bound) at a given parameter vector $\mathbf{x} = \mathbf{x}_0$ and the *locally minimum variance* (LMV) estimator, i.e., the (generally impractical) estimator achieving the minimum variance at $\mathbf{x}_0$. Simplified expressions of the minimum achievable variance and the LMV estimator are obtained for a certain subclass of "diagonal" bias functions (which includes the unbiased case).

Our bounds on the estimator variance may be useful in several ways. First, they provide interesting insights regarding the ease or difficulty of specific estimation scenarios, without a restriction to specific estimators. Second, they allow an assessment of the variance behavior of a given estimator. Finally, they can be used to derive bias functions resulting in a small variance and, in turn, a small overall estimation error.

A central aspect of this paper is the application of the mathematical framework of RKHS [15] to the SLGM. The RKHS framework has been previously applied to classical estimation in the seminal work reported in [16] and [17], and our present treatment is substantially based on that work. However, to the best of our knowledge, the RKHS approach has not been applied to the SLGM or, more generally, to the estimation of (functions of) S-sparse vectors. For $S < N$, the interior of the set of S-sparse vectors is empty, and thus there do not exist derivatives in every possible direction. This lack of a differentiable structure makes the characterization of the RKHS delicate.

The remainder of this paper is organized as follows. In Section II, we review some necessary fundamentals: formal statements of the SLGM and SSNM, basic elements of MVE, and RKHSs and their application to MVE. In Section III, we characterize and discuss the RKHS associated with the SLGM. We then use the RKHS framework to present formal characterizations of the class of bias functions allowing for finite-variance estimators for the SLGM, of the minimum achievable variance (Barankin bound), and of the LMV estimator. We also present a result on the shape of the Barankin bound. In Section IV, we reinterpret the sparse CRB of [12] from the RKHS perspective. Furthermore, we present a new lower variance bound for the SLGM and specialize it to CS measurement matrices. The important special case given by the SSNM is discussed in Section V, where we derive closed-form expressions of the Barankin bound and the corresponding LMV estimator. Finally, in Section VI, we present numerical results comparing our theoretical bounds with the actual variance of some popular estimation schemes.

Notation and Basic Definitions: The sets of real, nonnegative real, natural, and nonnegative integer numbers are denoted by $\mathbb{R}$, $\mathbb{R}_+$, $\mathbb{N} \triangleq \{1, 2, \ldots\}$, and $\mathbb{Z}_+ \triangleq \{0, 1, \ldots\}$, respectively. For $L \in \mathbb{N}$, we define $[L] \triangleq \{1, \ldots, L\}$. The space of all discrete-argument functions $f[\cdot] : \mathcal{T} \to \mathbb{R}$ (with $\mathcal{T} \subseteq \mathbb{Z}$) for which $\sum_{l \in \mathcal{T}} f^2[l] < \infty$ is denoted by $\ell^2(\mathcal{T})$, with associated norm $\|f[\cdot]\|_{\mathcal{T}} \triangleq \sqrt{\sum_{l \in \mathcal{T}} f^2[l]}$. The Kronecker delta $\delta_{k,l}$ is 1 if $k = l$ and 0 otherwise. Given an N-tuple of nonnegative integers (a "multi-index") $\mathbf{p} = (p_1 \cdots p_N)^T \in \mathbb{Z}_+^N$ [20], we define $\mathbf{p}! \triangleq \prod_{l \in [N]} p_l!$, $|\mathbf{p}| \triangleq \sum_{l \in [N]} p_l$, and $\mathbf{x}^{\mathbf{p}} \triangleq \prod_{l \in [N]} x_l^{p_l}$ (for $\mathbf{x} \in \mathbb{R}^N$). Given two multi-indices $\mathbf{p}_1, \mathbf{p}_2 \in \mathbb{Z}_+^N$, the inequality $\mathbf{p}_1 \leq \mathbf{p}_2$ is understood to hold entrywise, i.e., $p_{1,l} \leq p_{2,l}$ for all $l \in [N]$.

Lowercase (uppercase) boldface letters denote column vectors (matrices). The superscript ^{T} stands for transposition. The kth unit vector is denoted by $\mathbf{e}_k$, and the identity matrix by $\mathbf{I}$. For a rectangular matrix $\mathbf{H} \in \mathbb{R}^{M \times N}$, we denote by $\mathbf{H}^\dagger$ its Moore-Penrose pseudoinverse [21], by $\ker(\mathbf{H}) \triangleq \{\mathbf{x} \in \mathbb{R}^N | \mathbf{H}\mathbf{x} = \mathbf{0}\}$ its kernel (or null space), by $\mathrm{span}(\mathbf{H}) \triangleq \{\mathbf{y} \in \mathbb{R}^M | \exists \mathbf{x} \in \mathbb{R}^N : \mathbf{y} = \mathbf{H}\mathbf{x}\}$ its column span, and by $\mathrm{rank}(\mathbf{H})$ its rank. For a square matrix $\mathbf{H} \in \mathbb{R}^{N \times N}$, we denote by $\mathrm{tr}(\mathbf{H})$, $\det(\mathbf{H})$, and $\mathbf{H}^{-1}$ its trace, determinant, and inverse (if it exists), respectively. The kth entry of a vector $\mathbf{x}$ is denoted by $(\mathbf{x})_k = x_k$, and the entry in the kth row and lth column of a matrix $\mathbf{H}$ by $(\mathbf{H})_{k,l} = H_{k,l}$. The support (i.e., set of indices of all nonzero entries) and the number of nonzero entries of a vector $\mathbf{x}$ are denoted by $\mathrm{supp}(\mathbf{x})$ and $\|\mathbf{x}\|_0 = |\mathrm{supp}(\mathbf{x})|$, respectively. Given an index set $\mathcal{I} \subseteq [N]$, we denote by $\mathbf{x}^{\mathcal{I}} \in \mathbb{R}^N$ the vector obtained from $\mathbf{x} \in \mathbb{R}^N$ by zeroing all entries except those indexed by $\mathcal{I}$, and by $\mathbf{H}_{\mathcal{I}} \in \mathbb{R}^{M \times |\mathcal{I}|}$ the matrix formed by those columns of $\mathbf{H} \in \mathbb{R}^{M \times N}$ that are indexed by $\mathcal{I}$.

II. FUNDAMENTALS

A. The Sparse Linear Gaussian Model

Let $\mathbf{x} \in \mathbb{R}^N$ be an unknown, nonrandom parameter vector that is known to be S-sparse in the sense that at most S of its entries are nonzero, i.e., $\|\mathbf{x}\|_0 \leq S$, with a known sparsity degree $S \in [N]$ (typically $S \ll N$). We focus on strict sparsity instead of approximate sparsity because, as shown in [22, Sec. 5.6], this is necessary to allow for a smaller minimum achievable variance compared to the case without any sparsity constraints, i.e., the conventional linear Gaussian model. We express S-sparsity in terms of a parameter set $\mathcal{X}_S$, i.e.,

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

In the limiting case where S is equal to the dimension of $\mathbf{x}$, i.e., $S = N$, we have $\mathcal{X}_S = \mathbb{R}^N$. The support $\text{supp}(\mathbf{x}) \subseteq [N]$ is unknown. We observe a linearly transformed and noisy version of $\mathbf{x}$,

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \in \mathbb{R}^M, \tag{2}$$

where $\mathbf{H} \in \mathbb{R}^{M \times N}$ is a known matrix and $\mathbf{n} \in \mathbb{R}^M$ is i.i.d. Gaussian noise, i.e., $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$, with known noise variance $\sigma^2 > 0$. It follows that the probability density function (pdf) of the observation $\mathbf{y}$ for a specific value of $\mathbf{x}$ is given by

$$f_{\mathbf{H}}(\mathbf{y}; \mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{M/2}} \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2\right). \tag{3}$$

We assume that

$$\text{spark}(\mathbf{H}) > S, \tag{4}$$

where $\text{spark}(\mathbf{H})$ denotes the minimum number of linearly dependent columns of $\mathbf{H}$ [23], [24]. Note that we also allow $M < N$ (this case is relevant to CS methods as discussed in Section IV-C); however, condition (4) implies that $M \geq S$. Condition (4) is weaker than the standard condition $\text{spark}(\mathbf{H}) > 2S$ [12]. However, the standard condition is reasonable since otherwise one can find two different parameter vectors $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}_S$ such that $f_{\mathbf{H}}(\mathbf{y}; \mathbf{x}_1) = f_{\mathbf{H}}(\mathbf{y}; \mathbf{x}_2)$ for all $\mathbf{y}$, which implies that one cannot distinguish between $\mathbf{x}_1$ and $\mathbf{x}_2$ based on knowledge of $\mathbf{y}$.

We consider estimation of the function value $\mathbf{g}(\mathbf{x})$ from the observation $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$, where the *parameter function* $\mathbf{g}(\cdot) : \mathcal{X}_S \to \mathbb{R}^P$ is a known deterministic function. The *estimate* $\hat{\mathbf{g}} = \hat{\mathbf{g}}(\mathbf{y}) \in \mathbb{R}^P$ is derived from $\mathbf{y}$ via a deterministic *estimator* $\hat{\mathbf{g}}(\cdot) : \mathbb{R}^M \to \mathbb{R}^P$. We allow $\hat{\mathbf{g}} \in \mathbb{R}^P$ without constraining $\hat{\mathbf{g}}$ to be in $\mathbf{g}(\mathcal{X}_S) \triangleq \{\mathbf{g}(\mathbf{x}) | \mathbf{x} \in \mathcal{X}_S\}$, even though it is known that $\mathbf{x} \in \mathcal{X}_S$. The reason for not enforcing the sparsity constraint $\hat{\mathbf{g}} \in \mathbf{g}(\mathcal{X}_S)$ is twofold: first, it complicates the analysis; second, it typically results in a worse achievable estimator performance (in terms of mean squared error) since it restricts the class of allowed estimators. In particular, it has been shown that a sparsity constraint can increase the worst-case risk of the resulting estimators significantly [25].

Estimation of the parameter vector $\mathbf{x}$ itself is a special case obtained by choosing $\mathbf{g}(\mathbf{x}) = \mathbf{x}$, which implies $P = N$. Again, we allow $\hat{\mathbf{x}} \in \mathbb{R}^N$ and do not constrain $\hat{\mathbf{x}}$ to be in $\mathcal{X}_S$.

In what follows, it will be convenient to denote the SLGM estimation problem by the triple $\mathcal{E}_{\text{SLGM}} \triangleq \big(\mathcal{X}_S, f_{\mathbf{H}}(\mathbf{y}; \mathbf{x}), \mathbf{g}(\cdot)\big)$, where $f_{\mathbf{H}}(\mathbf{y}; \mathbf{x})$ is given by (3) and will be referred to as the *statistical model*. A related estimation problem is based on the *linear Gaussian model* (LGM) [18], [26]–[28], for which $\mathbf{x} \in \mathbb{R}^N$ rather than $\mathbf{x} \in \mathcal{X}_S$; this problem will be denoted by $\mathcal{E}_{\text{LGM}} \triangleq \big(\mathbb{R}^N, f_{\mathbf{H}}(\mathbf{y}; \mathbf{x}), \mathbf{g}(\cdot)\big)$. The SLGM shares with the LGM the observation model (2) and the statistical model (3); the two models coincide when $S = N$. Another important special case of the SLGM is given by the SSNM, in which $\mathbf{H} = \mathbf{I}$, $M = N$, and $\mathbf{y} = \mathbf{x} + \mathbf{n}$, where $\mathbf{x} \in \mathcal{X}_S$ and $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$. The SSNM estimation problem will be denoted as $\mathcal{E}_{\text{SSNM}} \triangleq \big(\mathcal{X}_S, f_{\mathbf{I}}(\mathbf{y}; \mathbf{x}), \mathbf{g}(\cdot)\big)$.

B. Basic Elements of Minimum Variance Estimation

Let us consider[1] an arbitrary estimation problem $\mathcal{E} = \big(\mathcal{X}, f(\mathbf{y}; \mathbf{x}), \mathbf{g}(\cdot)\big)$ based on an arbitrary parameter set $\mathcal{X} \subseteq \mathbb{R}^N$ and an arbitrary statistical model $f(\mathbf{y}; \mathbf{x})$. The general goal in the design of an estimator $\hat{\mathbf{g}}(\cdot)$ is that $\hat{\mathbf{g}}(\mathbf{y})$ should be close to the true value $\mathbf{g}(\mathbf{x})$. A frequently used criterion for assessing the quality of an estimator is the mean squared error (MSE) defined as

$$\varepsilon \triangleq \mathsf{E}_{\mathbf{x}}\big\{\|\hat{\mathbf{g}}(\mathbf{y}) - \mathbf{g}(\mathbf{x})\|_2^2\big\} = \int_{\mathbb{R}^M} \|\hat{\mathbf{g}}(\mathbf{y}) - \mathbf{g}(\mathbf{x})\|_2^2 \, f(\mathbf{y}; \mathbf{x}) \, d\mathbf{y}.$$

Here, $\mathsf{E}_{\mathbf{x}}\{\cdot\}$ denotes the expectation operation with respect to the pdf $f(\mathbf{y}; \mathbf{x})$; the subscript in $\mathsf{E}_{\mathbf{x}}$ indicates the dependence on the parameter vector $\mathbf{x}$ parametrizing $f(\mathbf{y}; \mathbf{x})$. We will write $\varepsilon(\hat{\mathbf{g}}(\cdot); \mathbf{x})$ to indicate the dependence of the MSE on the estimator $\hat{\mathbf{g}}(\cdot)$ and the parameter vector $\mathbf{x}$. In general, there does not exist an estimator $\hat{\mathbf{g}}(\cdot)$ that minimizes the MSE simultaneously for all $\mathbf{x} \in \mathcal{X}$ [18], [30]. This follows from the fact that minimizing the MSE at a given parameter vector $\mathbf{x}_0 \in \mathcal{X}$ always yields zero MSE; this is achieved by the trivial estimator $\hat{\mathbf{g}}(\mathbf{y}) \equiv \mathbf{g}(\mathbf{x}_0)$, which ignores the observation $\mathbf{y}$.

A popular rationale for the design of good estimators is MVE. The MSE can be decomposed as

$$\varepsilon(\hat{\mathbf{g}}(\cdot); \mathbf{x}) = \|\mathbf{b}(\hat{\mathbf{g}}(\cdot); \mathbf{x})\|_2^2 + v(\hat{\mathbf{g}}(\cdot); \mathbf{x}), \tag{5}$$

with the bias $\mathbf{b}(\hat{\mathbf{g}}(\cdot); \mathbf{x}) \triangleq \mathsf{E}_{\mathbf{x}}\{\hat{\mathbf{g}}(\mathbf{y})\} - \mathbf{g}(\mathbf{x})$ and the variance $v(\hat{\mathbf{g}}(\cdot); \mathbf{x}) \triangleq \mathsf{E}_{\mathbf{x}}\big\{\|\hat{\mathbf{g}}(\mathbf{y}) - \mathsf{E}_{\mathbf{x}}\{\hat{\mathbf{g}}(\mathbf{y})\}\|_2^2\big\}$. In MVE, one specifies the bias on the entire parameter set $\mathcal{X}$, i.e., one requires that

$$\mathbf{b}(\hat{\mathbf{g}}(\cdot); \mathbf{x}) = \mathbf{c}(\mathbf{x}), \quad \text{for all } \mathbf{x} \in \mathcal{X},$$

with a *prescribed bias function* $\mathbf{c}(\cdot) : \mathcal{X} \to \mathbb{R}^P$, and one attempts to minimize the variance $v(\hat{\mathbf{g}}(\cdot); \mathbf{x})$ among all estimators with the given bias function $\mathbf{c}(\cdot)$. Fixing the bias is equivalent to fixing the estimator's mean function, i.e., $\mathsf{E}_{\mathbf{x}}\big\{\hat{\mathbf{g}}(\mathbf{y})\big\} = \boldsymbol{\gamma}(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}$, with the *prescribed mean function* $\boldsymbol{\gamma}(\mathbf{x}) \triangleq \mathbf{c}(\mathbf{x}) + \mathbf{g}(\mathbf{x})$. *Unbiased estimation* is obtained as a special case for $\mathbf{c}(\mathbf{x}) \equiv \mathbf{0}$ or equivalently $\boldsymbol{\gamma}(\mathbf{x}) \equiv \mathbf{g}(\mathbf{x})$. Constraining the bias can be viewed as a kind of "regularization" of the set of possible estimators [19], [30], since it excludes useless estimators such as $\hat{\mathbf{g}}(\mathbf{y}) \equiv \mathbf{g}(\mathbf{x}_0)$.

[1] This introductory section closely parallels [29, Sec. II-A]. We include it nevertheless because it constitutes an important basis for our subsequent discussion.

Another justification for considering a fixed bias function is that under mild conditions, for a large number L of i.i.d. observations $\{\mathbf{y}_i\}_{i\in[L]}$, the bias term dominates in the decomposition (5). Thus, in order to achieve a small MSE in that case, an estimator has to be at least asymptotically unbiased, i.e., one has to require that, for a large number of observations, $\mathbf{b}(\hat{\mathbf{g}}(\cdot);\mathbf{x}) \approx \mathbf{0}$ for all $\mathbf{x}\in\mathcal{X}$.

For an estimation problem $\mathcal{E} = \big(\mathcal{X}, f(\mathbf{y};\mathbf{x}), \mathbf{g}(\cdot)\big)$, a fixed parameter vector $\mathbf{x}_0\in\mathcal{X}$, and a prescribed bias function $\mathbf{c}(\cdot): \mathcal{X}\to\mathbb{R}^P$, we define the *set of allowed estimators* by

$$\mathcal{A}(\mathbf{c}(\cdot),\mathbf{x}_0) \triangleq \big\{\hat{\mathbf{g}}(\cdot)\,\big|\, v(\hat{\mathbf{g}}(\cdot);\mathbf{x}_0) < \infty, \\ \mathbf{b}(\hat{\mathbf{g}}(\cdot);\mathbf{x}) = \mathbf{c}(\mathbf{x})\ \forall \mathbf{x}\in\mathcal{X}\big\}.$$

We call a bias function $\mathbf{c}(\cdot)$ *valid* for the estimation problem $\mathcal{E}$ at $\mathbf{x}_0\in\mathcal{X}$ if the set $\mathcal{A}(\mathbf{c}(\cdot),\mathbf{x}_0)$ is nonempty, which means that there is at least one estimator $\hat{\mathbf{g}}(\cdot)$ that has finite variance at $\mathbf{x}_0$ and whose bias function equals $\mathbf{c}(\cdot)$, i.e., $\mathbf{b}(\hat{\mathbf{g}}(\cdot);\mathbf{x}) = \mathbf{c}(\mathbf{x})$ for all $\mathbf{x}\in\mathcal{X}$. For the SLGM, in particular, this definition trivially entails the following fact: If a bias function $\mathbf{c}(\cdot):\mathcal{X}_S\to\mathbb{R}^P$ is valid for $S=N$, it is also valid for $S<N$.

It follows from (5) that, for a fixed bias function $\mathbf{c}(\cdot)$, minimizing the MSE $\varepsilon(\hat{\mathbf{g}}(\cdot);\mathbf{x}_0)$ is equivalent to minimizing the variance $v(\hat{\mathbf{g}}(\cdot);\mathbf{x}_0)$. Let us denote the minimum (strictly speaking, infimum) variance at $\mathbf{x}_0$ for bias function $\mathbf{c}(\cdot)$ by

$$M(\mathbf{c}(\cdot),\mathbf{x}_0) \triangleq \inf_{\hat{\mathbf{g}}(\cdot)\in\mathcal{A}(\mathbf{c}(\cdot),\mathbf{x}_0)} v(\hat{\mathbf{g}}(\cdot);\mathbf{x}_0). \tag{6}$$

If $\mathcal{A}(\mathbf{c}(\cdot),\mathbf{x}_0)$ is empty, i.e., if $\mathbf{c}(\cdot)$ is not valid, we set $M(\mathbf{c}(\cdot),\mathbf{x}_0)\triangleq\infty$. Any estimator $\hat{\mathbf{g}}^{(\mathbf{c}(\cdot),\mathbf{x}_0)}(\cdot)\in\mathcal{A}(\mathbf{c}(\cdot),\mathbf{x}_0)$ that achieves the infimum in (6), i.e., for which

$$v\big(\hat{\mathbf{g}}^{(\mathbf{c}(\cdot),\mathbf{x}_0)}(\cdot);\mathbf{x}_0\big) = M(\mathbf{c}(\cdot),\mathbf{x}_0), \tag{7}$$

is called an *LMV estimator* at $\mathbf{x}_0$ for bias function $\mathbf{c}(\cdot)$ [16], [17], [19]. The corresponding minimum variance $M(\mathbf{c}(\cdot),\mathbf{x}_0)$ is called the *minimum achievable variance* at $\mathbf{x}_0$ for bias function $\mathbf{c}(\cdot)$. The minimization problem defined by (6) is referred to as a *minimum variance problem* (MVP). From its definition in (6), it follows that $M(\mathbf{c}(\cdot),\mathbf{x}_0)$ is a lower bound on the variance at $\mathbf{x}_0$ of any estimator with bias function $\mathbf{c}(\cdot)$, i.e., $\hat{\mathbf{g}}(\cdot)\in\mathcal{A}(\mathbf{c}(\cdot),\mathbf{x}_0)$ implies $v(\hat{\mathbf{g}}(\cdot);\mathbf{x}_0)\geq M(\mathbf{c}(\cdot),\mathbf{x}_0)$. This is sometimes referred to as the *Barankin bound* [17]; it is the tightest possible lower bound on the variance at $\mathbf{x}_0$ of estimators with bias function $\mathbf{c}(\cdot)$.

Finally, let $\hat{g}_k(\cdot)\triangleq\big(\hat{\mathbf{g}}(\cdot)\big)_k$ and $c_k(\cdot)\triangleq\big(\mathbf{c}(\cdot)\big)_k$. The variance of the vector estimator $\hat{\mathbf{g}}(\cdot)$ can be decomposed as

$$v(\hat{\mathbf{g}}(\cdot);\mathbf{x}) = \sum_{k\in[P]} v(\hat{g}_k(\cdot);\mathbf{x}), \tag{8}$$

where $v(\hat{g}_k(\cdot);\mathbf{x})\triangleq\mathsf{E}_{\mathbf{x}}\big\{\big[\hat{g}_k(\mathbf{y})-\mathsf{E}_{\mathbf{x}}\{\hat{g}_k(\mathbf{y})\}\big]^2\big\}$ is the variance of the kth estimator component $\hat{g}_k(\cdot)$. Furthermore, $\hat{\mathbf{g}}(\cdot)\in\mathcal{A}(\mathbf{c}(\cdot),\mathbf{x}_0)$ if and only if $\hat{g}_k(\cdot)\in\mathcal{A}(c_k(\cdot),\mathbf{x}_0)$ for all $k\in[P]$. This shows that the MVP (6) can be reduced to P separate *scalar* MVPs

$$M(c_k(\cdot),\mathbf{x}_0) \triangleq \inf_{\hat{g}_k(\cdot)\in\mathcal{A}(c_k(\cdot),\mathbf{x}_0)} v(\hat{g}_k(\cdot);\mathbf{x}_0), \quad k\in[P],$$

each requiring the optimization of a single scalar component $\hat{g}_k(\cdot)$ of $\hat{\mathbf{g}}(\cdot)$. Therefore, without loss of generality, we will hereafter assume that the parameter function $\mathbf{g}(\mathbf{x})$ is scalar-valued, i.e., $P=1$ and $\mathbf{g}(\mathbf{x})=g(\mathbf{x})$.

C. RKHS Fundamentals

As mentioned in Section I, the existing variance bounds for the SLGM are not maximally tight. Using the theory of RKHSs will allow us to derive a variance bound that is tighter than existing bounds. For the SSNM (see Section V), the RKHS approach even yields a closed-form characterization of the Barankin bound and of the accompanying LMV estimator. Next, we present a review (similar in part to [29, Sec. II-B]) of some fundamentals of the theory of RKHSs and of the application of RKHSs to MVE. These fundamentals will provide a framework for our analysis of the SLGM in later sections.

An RKHS is associated with a *kernel function* $R(\cdot,\cdot):\mathcal{X}\times\mathcal{X}\to\mathbb{R}$, where $\mathcal{X}$ is an arbitrary set. The defining properties of a kernel function are (i) symmetry, i.e., $R(\mathbf{x}_1,\mathbf{x}_2)=R(\mathbf{x}_2,\mathbf{x}_1)$ for all $\mathbf{x}_1,\mathbf{x}_2\in\mathcal{X}$, and (ii) positive semidefiniteness in the sense that, for every finite set $\{\mathbf{x}_1,\ldots,\mathbf{x}_D\}\subseteq\mathcal{X}$, the matrix $\mathbf{R}\in\mathbb{R}^{D\times D}$ with entries $R_{m,n}=R(\mathbf{x}_m,\mathbf{x}_n)$ is positive semidefinite. A fundamental result [15, p. 344] states that for any such kernel function R, there exists an RKHS $\mathcal{H}(R)$, which is a Hilbert space equipped with an inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}(R)}$ and satisfying the following two properties: (i) $R(\cdot,\mathbf{x})\in\mathcal{H}(R)$ for any $\mathbf{x}\in\mathcal{X}$ (here, $R(\cdot,\mathbf{x})$ denotes the function $f_{\mathbf{x}}(\mathbf{x}')=R(\mathbf{x}',\mathbf{x})$ for fixed $\mathbf{x}\in\mathcal{X}$). (ii) For any function $f(\cdot)\in\mathcal{H}(R)$ and any $\mathbf{x}\in\mathcal{X}$, $\big\langle f(\cdot), R(\cdot,\mathbf{x})\big\rangle_{\mathcal{H}(R)} = f(\mathbf{x})$. This reproducing property defines the inner product $\langle f_1,f_2\rangle_{\mathcal{H}(R)}$ for all $f_1(\cdot),f_2(\cdot)\in\mathcal{H}(R)$, because any $f(\cdot)\in\mathcal{H}(R)$ can be expanded into the set of functions $\{R(\cdot,\mathbf{x})\}_{\mathbf{x}\in\mathcal{X}}$. The induced norm is $\|f\|_{\mathcal{H}(R)}=\sqrt{\langle f,f\rangle_{\mathcal{H}(R)}}$.

RKHS theory provides a powerful mathematical framework for MVE [16]. Given an arbitrary estimation problem $\mathcal{E}=\big(\mathcal{X},f(\mathbf{y};\mathbf{x}),g(\cdot)\big)$ and a parameter vector $\mathbf{x}_0\in\mathcal{X}$ for which $f(\mathbf{y};\mathbf{x}_0)\neq 0$, a kernel function $R_{\mathcal{E},\mathbf{x}_0}(\cdot,\cdot)$ and, in turn, an RKHS $\mathcal{H}_{\mathcal{E},\mathbf{x}_0}$ can be defined as follows. We first define the *likelihood ratio*

$$\rho_{\mathbf{x}_0}(\mathbf{y},\mathbf{x}) \triangleq \frac{f(\mathbf{y};\mathbf{x})}{f(\mathbf{y};\mathbf{x}_0)}, \tag{9}$$

which is considered as a random variable (since it is a function of the random vector $\mathbf{y}$) that is parametrized by $\mathbf{x}\in\mathcal{X}$. Next, we define the Hilbert space $\mathcal{L}_{\mathcal{E},\mathbf{x}_0}$ as the closure of the linear span of the set of random variables $\big\{\rho_{\mathbf{x}_0}(\mathbf{y},\mathbf{x})\big\}_{\mathbf{x}\in\mathcal{X}}$. The inner product in $\mathcal{L}_{\mathcal{E},\mathbf{x}_0}$ is defined by

$$\big\langle\rho_{\mathbf{x}_0}(\mathbf{y},\mathbf{x}_1),\rho_{\mathbf{x}_0}(\mathbf{y},\mathbf{x}_2)\big\rangle_{\mathrm{RV}} \triangleq \mathsf{E}_{\mathbf{x}_0}\big\{\rho_{\mathbf{x}_0}(\mathbf{y},\mathbf{x}_1)\,\rho_{\mathbf{x}_0}(\mathbf{y},\mathbf{x}_2)\big\} = \mathsf{E}_{\mathbf{x}_0}\bigg\{\frac{f(\mathbf{y};\mathbf{x}_1)f(\mathbf{y};\mathbf{x}_2)}{f^2(\mathbf{y};\mathbf{x}_0)}\bigg\}.$$

(It can be shown that it is sufficient to define $\langle\cdot,\cdot\rangle_{\mathrm{RV}}$ for the random variables $\big\{\rho_{\mathbf{x}_0}(\mathbf{y},\mathbf{x})\big\}_{\mathbf{x}\in\mathcal{X}}$ [16].) From now on, we consider only estimation problems $\mathcal{E}=\big(\mathcal{X},f(\mathbf{y};\mathbf{x}),g(\cdot)\big)$ such that $\big\langle\rho_{\mathbf{x}_0}(\mathbf{y},\mathbf{x}_1),\rho_{\mathbf{x}_0}(\mathbf{y},\mathbf{x}_2)\big\rangle_{\mathrm{RV}}<\infty$ for all $\mathbf{x}_1,\mathbf{x}_2\in\mathcal{X}$, or, equivalently, $\mathsf{E}_{\mathbf{x}_0}\Big\{\frac{f(\mathbf{y};\mathbf{x}_1)f(\mathbf{y};\mathbf{x}_2)}{f^2(\mathbf{y};\mathbf{x}_0)}\Big\}<\infty$ for all $\mathbf{x}_1,\mathbf{x}_2\in\mathcal{X}$. Thus, $\langle\cdot,\cdot\rangle_{\mathrm{RV}}$ is well defined. We can interpret the inner product $\langle\cdot,\cdot\rangle_{\mathrm{RV}}:\mathcal{L}_{\mathcal{E},\mathbf{x}_0}\times\mathcal{L}_{\mathcal{E},\mathbf{x}_0}\to\mathbb{R}$ as a kernel function $R_{\mathcal{E},\mathbf{x}_0}(\cdot,\cdot):\mathcal{X}\times\mathcal{X}\to\mathbb{R}$:

$$R_{\mathcal{E},\mathbf{x}_0}(\mathbf{x}_1,\mathbf{x}_2) \triangleq \big\langle\rho_{\mathbf{x}_0}(\mathbf{y},\mathbf{x}_1),\rho_{\mathbf{x}_0}(\mathbf{y},\mathbf{x}_2)\big\rangle_{\mathrm{RV}}$$

$$= \mathsf{E}_{\mathbf{x}_0}\left\{\frac{f(\mathbf{y};\mathbf{x}_1)f(\mathbf{y};\mathbf{x}_2)}{f^2(\mathbf{y};\mathbf{x}_0)}\right\}. \quad (10)$$

The RKHS associated with the estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y};\mathbf{x}), g(\cdot))$ and the parameter vector $\mathbf{x}_0 \in \mathcal{X}$ is then defined to be the RKHS induced by the kernel function $R_{\mathcal{E},\mathbf{x}_0}(\cdot,\cdot)$. We will denote this RKHS as $\mathcal{H}_{\mathcal{E},\mathbf{x}_0}$, i.e., $\mathcal{H}_{\mathcal{E},\mathbf{x}_0} \triangleq \mathcal{H}(R_{\mathcal{E},\mathbf{x}_0})$. As shown in [16], the two Hilbert spaces $\mathcal{L}_{\mathcal{E},\mathbf{x}_0}$ and $\mathcal{H}_{\mathcal{E},\mathbf{x}_0}$ are isometric, and a specific congruence, i.e., isometric mapping $\mathsf{J}[\cdot]: \mathcal{H}_{\mathcal{E},\mathbf{x}_0} \to \mathcal{L}_{\mathcal{E},\mathbf{x}_0}$ is given by

$$\mathsf{J}[R_{\mathcal{E},\mathbf{x}_0}(\cdot,\mathbf{x})] = \rho_{\mathbf{x}_0}(\cdot,\mathbf{x}). \quad (11)$$

A fundamental relation of the RKHS $\mathcal{H}_{\mathcal{E},\mathbf{x}_0}$ with MVE is established by the following central result [16], [17]. Consider an estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y};\mathbf{x}), g(\cdot))$, a fixed parameter vector $\mathbf{x}_0 \in \mathcal{X}$, and a prescribed bias function $c(\cdot): \mathcal{X} \to \mathbb{R}$, corresponding to the mean function $\gamma(\cdot) = c(\cdot) + g(\cdot)$. Then, the following holds:

1) The bias function $c(\cdot)$ is valid for $\mathcal{E}$ at $\mathbf{x}_0$ if and only if $\gamma(\cdot)$ belongs to the RKHS $\mathcal{H}_{\mathcal{E},\mathbf{x}_0}$, i.e.,

$$c(\cdot) \text{ valid for } \mathcal{E} \text{ at } \mathbf{x}_0 \iff c(\cdot) = \gamma(\cdot) - g(\cdot) \text{ for some } \gamma(\cdot) \in \mathcal{H}_{\mathcal{E},\mathbf{x}_0}. \quad (12)$$

2) If the bias function $c(\cdot)$ is valid for $\mathcal{E}$ at $\mathbf{x}_0$, then the minimum achievable variance at $\mathbf{x}_0$ (Barankin bound) is given by

$$M(c(\cdot),\mathbf{x}_0) = \|\gamma(\cdot)\|^2_{\mathcal{H}_{\mathcal{E},\mathbf{x}_0}} - \gamma^2(\mathbf{x}_0), \quad (13)$$

and the LMV estimator at $\mathbf{x}_0$ is

$$\hat{g}^{(c(\cdot),\mathbf{x}_0)}(\cdot) = \mathsf{J}[\gamma(\cdot)]. \quad (14)$$

According to the first result, the RKHS $\mathcal{H}_{\mathcal{E},\mathbf{x}_0}$ can be interpreted as the set of mean functions $\gamma(\mathbf{x}) = \mathsf{E}_{\mathbf{x}}\{\hat{g}(\mathbf{y})\}$ of all estimators $\hat{g}(\cdot)$ with a finite variance at $\mathbf{x}_0$, i.e., $v(\hat{g}(\cdot);\mathbf{x}_0) < \infty$. Furthermore, we can use (13) to establish a large class of lower bounds on the minimum achievable variance $M(c(\cdot),\mathbf{x}_0)$. Indeed, let $\mathcal{U} \subseteq \mathcal{H}_{\mathcal{E},\mathbf{x}_0}$ be an arbitrary subspace of $\mathcal{H}_{\mathcal{E},\mathbf{x}_0}$ and let $\mathsf{P}_{\mathcal{U}}\gamma(\cdot)$ denote the orthogonal projection of $\gamma(\cdot)$ onto $\mathcal{U}$. We then have $\|\gamma(\cdot)\|^2_{\mathcal{H}_{\mathcal{E},\mathbf{x}_0}} \geq \|\mathsf{P}_{\mathcal{U}}\gamma(\cdot)\|^2_{\mathcal{H}_{\mathcal{E},\mathbf{x}_0}}$ [31, Ch. 4] and thus, from (13),

$$M(c(\cdot),\mathbf{x}_0) \geq \|\mathsf{P}_{\mathcal{U}}\gamma(\cdot)\|^2_{\mathcal{H}_{\mathcal{E},\mathbf{x}_0}} - \gamma^2(\mathbf{x}_0). \quad (15)$$

Some well-known lower bounds on the estimator variance, such as the Cramér–Rao and Bhattacharya bounds, are obtained from (15) by specific choices of the subspace $\mathcal{U}$ [29].

III. RKHS-Based Analysis of Minimum Variance Estimation for the SLGM

In this section, we apply the RKHS framework to the SLGM estimation problem $\mathcal{E}_{\text{SLGM}} = (\mathcal{X}_S, f_{\mathbf{H}}(\mathbf{y};\mathbf{x}), g(\cdot))$. Thus, the parameter set is the set of S-sparse vectors, $\mathcal{X} = \mathcal{X}_S \subseteq \mathbb{R}^N$ in (1), and the statistical model is given by $f(\mathbf{y};\mathbf{x}) = f_{\mathbf{H}}(\mathbf{y};\mathbf{x})$ in (3). More specifically, we consider MVE at a given parameter vector $\mathbf{x}_0 \in \mathcal{X}_S$, for a prescribed bias function $c(\cdot): \mathcal{X}_S \to \mathbb{R}$. We recall that the set of allowed estimators, $\mathcal{A}(c(\cdot),\mathbf{x}_0)$, consists of all estimators $\hat{g}(\cdot)$ with finite variance at $\mathbf{x}_0$, i.e., $v(\hat{g}(\cdot);\mathbf{x}_0) < \infty$, whose bias function equals $c(\cdot)$, i.e., $b(\hat{g}(\cdot);\mathbf{x}) = c(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}_S$.

Our results can be summarized as follows. We characterize the RKHS associated with the SLGM and employ it to analyze the associated MVP. Using this characterization, we provide conditions on the bias function $c(\cdot)$ such that the minimum achievable variance is finite, i.e., we characterize the set of valid bias functions (cf. Section II-B). Furthermore, we present expressions of the Barankin bound $M_{\text{SLGM}}(c(\cdot),\mathbf{x}_0)$ and of the corresponding LMV estimator $\hat{g}^{(c(\cdot),\mathbf{x}_0)}(\cdot)$ for an arbitrary valid bias function $c(\cdot)$. Since these expressions are difficult to evaluate in general, we finally derive a lower bound on the minimum achievable variance. This lower bound is also a lower bound on the variance of any estimator with the prescribed bias function. Finally, we present a result on the shape of the Barankin bound.

In our analysis of the SLGM, the RKHS associated with the LGM will play an important role. Consider $\mathcal{X} = \mathbb{R}^N$ and $f(\mathbf{y};\mathbf{x}) = f_{\mathbf{H}}(\mathbf{y};\mathbf{x})$ as defined in (3), where the system matrix $\mathbf{H} \in \mathbb{R}^{M\times N}$ is *not* required to satisfy condition (4). The likelihood ratio (9) for $f(\mathbf{y};\mathbf{x}) = f_{\mathbf{H}}(\mathbf{y};\mathbf{x})$ is obtained as

$$\rho_{\text{LGM},\mathbf{x}_0}(\mathbf{y},\mathbf{x}) = \frac{f_{\mathbf{H}}(\mathbf{y};\mathbf{x})}{f_{\mathbf{H}}(\mathbf{y};\mathbf{x}_0)} = \exp\left(-\frac{1}{2\sigma^2}\left[2\mathbf{y}^T\mathbf{H}(\mathbf{x}_0-\mathbf{x}) + \|\mathbf{H}\mathbf{x}\|_2^2 - \|\mathbf{H}\mathbf{x}_0\|_2^2\right]\right). \quad (16)$$

Furthermore, from (10), the kernel associated with the LGM follows as

$$R_{\text{LGM},\mathbf{x}_0}(\cdot,\cdot): \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R};$$
$$R_{\text{LGM},\mathbf{x}_0}(\mathbf{x}_1,\mathbf{x}_2) = \exp\left(\frac{1}{\sigma^2}(\mathbf{x}_2-\mathbf{x}_0)^T\mathbf{H}^T\mathbf{H}(\mathbf{x}_1-\mathbf{x}_0)\right). \quad (17)$$

A. The RKHS Associated With the SLGM

Let us consider the SLGM estimation problem $\mathcal{E}_{\text{SLGM}} = (\mathcal{X}_S, f_{\mathbf{H}}(\mathbf{y};\mathbf{x}), g(\cdot))$ and the corresponding LGM estimation problem $\mathcal{E}_{\text{LGM}} = (\mathbb{R}^N, f_{\mathbf{H}}(\mathbf{y};\mathbf{x}), g(\cdot))$ with the same system matrix $\mathbf{H} \in \mathbb{R}^{M\times N}$ satisfying condition (4) and with the same noise variance σ^2. For an S-sparse parameter vector $\mathbf{x}_0 \in \mathcal{X}_S$, let $\mathcal{H}_{\text{SLGM},\mathbf{x}_0}$ and $\mathcal{H}_{\text{LGM},\mathbf{x}_0}$ denote the RKHSs associated with the estimation problems $\mathcal{E}_{\text{SLGM}}$ and $\mathcal{E}_{\text{LGM}}$, respectively. In what follows, we will use the thin singular value decomposition (SVD) of the system matrix $\mathbf{H}$, i.e., $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, where $\mathbf{U} \in \mathbb{R}^{M\times D}$ with $\mathbf{U}^T\mathbf{U} = \mathbf{I}$, $\mathbf{V} \in \mathbb{R}^{N\times D}$ with $\mathbf{V}^T\mathbf{V} = \mathbf{I}$, and $\mathbf{\Sigma} \in \mathbb{R}^{D\times D}$ is a diagonal matrix with positive diagonal entries $(\mathbf{\Sigma})_{k,k} > 0$ [21]. Here, $D = \text{rank}(\mathbf{H})$. We also define $\widetilde{\mathbf{H}} \triangleq \mathbf{V}\mathbf{\Sigma}^{-1}$.

Using (10) and (3), the kernel underlying $\mathcal{H}_{\text{SLGM},\mathbf{x}_0}$ is obtained as

$$R_{\text{SLGM},\mathbf{x}_0}(\cdot,\cdot): \mathcal{X}_S \times \mathcal{X}_S \to \mathbb{R};$$
$$R_{\text{SLGM},\mathbf{x}_0}(\mathbf{x}_1,\mathbf{x}_2) = \exp\left(\frac{1}{\sigma^2}(\mathbf{x}_2-\mathbf{x}_0)^T\mathbf{H}^T\mathbf{H}(\mathbf{x}_1-\mathbf{x}_0)\right). \quad (18)$$

Comparing with the kernel $R_{\text{LGM},\mathbf{x}_0}(\cdot,\cdot)$ in (17), we conclude that $R_{\text{SLGM},\mathbf{x}_0}(\cdot,\cdot)$ is the restriction of $R_{\text{LGM},\mathbf{x}_0}(\cdot,\cdot)$ to the subdomain $\mathcal{X}_S \times \mathcal{X}_S \subseteq \mathbb{R}^N \times \mathbb{R}^N$. This suggests, in turn, that the two RKHSs $\mathcal{H}_{\mathrm{SLGM},\mathbf{x}_0}$ and $\mathcal{H}_{\mathrm{LGM},\mathbf{x}_0}$ are closely related. In fact, using results from [15, p. 351] yields the following characterization of $\mathcal{H}_{\mathrm{SLGM},\mathbf{x}_0}$:

Theorem III.1: Let $\mathbf{x}_0 \in \mathcal{X}_S$.

1) *The RKHS $\mathcal{H}_{\mathrm{SLGM},\mathbf{x}_0}$ consists of the restrictions of all functions $\widetilde{f}(\cdot) : \mathbb{R}^N \to \mathbb{R}$ contained in $\mathcal{H}_{\mathrm{LGM},\mathbf{x}_0}$ to the subdomain $\mathcal{X}_S \subseteq \mathbb{R}^N$, i.e.,*

$$\mathcal{H}_{\mathrm{SLGM},\mathbf{x}_0} = \left\{ f(\cdot) = \widetilde{f}(\cdot)\big|_{\mathcal{X}_S} \,\middle|\, \widetilde{f}(\cdot) \in \mathcal{H}_{\mathrm{LGM},\mathbf{x}_0} \right\}. \quad (19)$$

The functions of the RKHS $\mathcal{H}_{\mathrm{LGM},\mathbf{x}_0}$ can be expressed as

$$\widetilde{f}(\mathbf{x}) = \widetilde{f}_a(\mathbf{x}) = \exp\left(\frac{1}{2\sigma^2} \|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma^2} \mathbf{x}^T \mathbf{H}^T \mathbf{H} \mathbf{x}_0 \right) \times \sum_{\mathbf{p} \in \mathbb{Z}_+^D} \frac{a[\mathbf{p}]}{\sqrt{\mathbf{p}!}} \left(\frac{1}{\sigma} \widetilde{\mathbf{H}}^\dagger \mathbf{x} \right)^{\mathbf{p}}, \quad (20)$$

with some coefficient sequence $a[\mathbf{p}] \in \ell^2(\mathbb{Z}_+^D)$.

2) *The norm of a function $f(\cdot) \in \mathcal{H}_{\mathrm{SLGM},\mathbf{x}_0}$ is equal to the minimum of the norms of all functions $\widetilde{f}_a(\cdot) \in \mathcal{H}_{\mathrm{LGM},\mathbf{x}_0}$ whose restriction to $\mathcal{X}_S$ equals $f(\cdot)$, i.e.,*

$$\|f(\cdot)\|_{\mathcal{H}_{\mathrm{SLGM},\mathbf{x}_0}} = \min_{\substack{\widetilde{f}_a(\cdot) \in \mathcal{H}_{\mathrm{LGM},\mathbf{x}_0} \\ \widetilde{f}_a(\cdot)|_{\mathcal{X}_S} = f(\cdot)}} \|\widetilde{f}_a(\cdot)\|_{\mathcal{H}_{\mathrm{LGM},\mathbf{x}_0}}. \quad (21)$$

The norm of a function $\widetilde{f}_a(\cdot) \in \mathcal{H}_{\mathrm{LGM},\mathbf{x}_0}$ (with coefficient sequence $a[\mathbf{p}]$) is given by

$$\|\widetilde{f}_a(\cdot)\|_{\mathcal{H}_{\mathrm{LGM},\mathbf{x}_0}} = \|a[\cdot]\|_{\ell^2(\mathbb{Z}_+^D)}, \quad (22)$$

where $\|a[\cdot]\|^2_{\ell^2(\mathbb{Z}_+^D)} \triangleq \sum_{\mathbf{p} \in \mathbb{Z}_+^D} a^2[\mathbf{p}]$.

Proof: see Appendix A.

An immediate consequence of Theorem III.1 is the obvious fact that the minimum achievable variance for the SLGM can never exceed that for the LGM (if the prescribed bias function for the SLGM is the restriction of the prescribed bias function for the LGM). Indeed, prescribing the bias for all $\mathbf{x} \in \mathbb{R}^N$ (as is done within the LGM), instead of prescribing it only for the sparse vectors $\mathbf{x} \in \mathcal{X}_S$ (as is done within the SLGM) can only result in a higher (or equal) minimum achievable variance. More formally, let $c(\cdot) : \mathbb{R}^N \to \mathbb{R}$ be the LGM bias function and $\gamma(\cdot) = c(\cdot) + g(\cdot)$ the corresponding mean function, and recall that $\mathbf{x}_0 \in \mathcal{X}_S$. Then we have

$$\begin{aligned} M_{\mathrm{SLGM}}\big(c(\cdot)\big|_{\mathcal{X}_S}, \mathbf{x}_0\big) &\overset{(13)}{=} \big\|\gamma(\cdot)\big|_{\mathcal{X}_S}\big\|^2_{\mathcal{H}_{\mathrm{SLGM},\mathbf{x}_0}} - \gamma^2(\mathbf{x}_0) \\ &\overset{(21)}{\leq} \|\gamma(\cdot)\|^2_{\mathcal{H}_{\mathrm{LGM},\mathbf{x}_0}} - \gamma^2(\mathbf{x}_0) \overset{(13)}{=} M_{\mathrm{LGM}}(c(\cdot), \mathbf{x}_0). \end{aligned}$$

B. The Class of Valid Bias Functions

The class of valid bias functions for the SLGM estimation problem $\mathcal{E}_{\mathrm{SLGM}} = \big(\mathcal{X}_S, f_{\mathbf{H}}(\mathbf{y}; \mathbf{x}), g(\cdot)\big)$ at $\mathbf{x}_0 \in \mathcal{X}_S$ is characterized by the following corollary of Theorem III.1 (see [22, Th. 5.3.1]):

Corollary III.2: A bias function $c(\cdot) : \mathcal{X}_S \to \mathbb{R}$ is valid for $\mathcal{E}_{\mathrm{SLGM}} = \big(\mathcal{X}_S, f_{\mathbf{H}}(\mathbf{y}; \mathbf{x}), g(\cdot)\big)$ at $\mathbf{x}_0 \in \mathcal{X}_S$ if and only if it can be expressed as

$$c(\mathbf{x}) = \exp\left(\frac{1}{2\sigma^2} \|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma^2} \mathbf{x}^T \mathbf{H}^T \mathbf{H} \mathbf{x}_0 \right) \times \sum_{\mathbf{p} \in \mathbb{Z}_+^D} \frac{a[\mathbf{p}]}{\sqrt{\mathbf{p}!}} \left(\frac{1}{\sigma} \widetilde{\mathbf{H}}^\dagger \mathbf{x} \right)^{\mathbf{p}} - g(\mathbf{x}), \quad \mathbf{x} \in \mathcal{X}_S, \quad (23)$$

with some coefficient sequence $a[\mathbf{p}] \in \ell^2(\mathbb{Z}_+^D)$.

Proof: According to (12), any valid bias function $c(\cdot)$ is given by $c(\cdot) = \gamma(\cdot) - g(\cdot)$ with some $\gamma(\cdot) \in \mathcal{H}_{\mathrm{SLGM},\mathbf{x}_0}$. The statement then follows by the characterization (20) of the functions belonging to $\mathcal{H}_{\mathrm{SLGM},\mathbf{x}_0}$. □

Corollary III.2 implies that the mean function $\gamma(\cdot) = c(\cdot) + g(\cdot)$ corresponding to a bias function $c(\cdot)$ that is valid for $\mathcal{E}_{\mathrm{SLGM}}$ at $\mathbf{x}_0 \in \mathcal{X}_S$ is of the form

$$\gamma(\mathbf{x}) = \exp\left(\frac{1}{2\sigma^2} \|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma^2} \mathbf{x}^T \mathbf{H}^T \mathbf{H} \mathbf{x}_0 \right) \times \sum_{\mathbf{p} \in \mathbb{Z}_+^D} \frac{a[\mathbf{p}]}{\sqrt{\mathbf{p}!}} \left(\frac{1}{\sigma} \widetilde{\mathbf{H}}^\dagger \mathbf{x} \right)^{\mathbf{p}}, \quad \mathbf{x} \in \mathcal{X}_S, \quad (24)$$

with some coefficient sequence $a[\mathbf{p}] \in \ell^2(\mathbb{Z}_+^D)$. Note that the condition $a[\mathbf{p}] \in \ell^2(\mathbb{Z}_+^D)$ requires the coefficients $a[\mathbf{p}]$ to decay sufficiently fast with increasing $\|\mathbf{p}\|_\infty$. The function on the right-hand side in (24) is *analytic* on the domain $\mathcal{X}_S$ in the sense[2] that it can be locally represented at any point $\mathbf{x} \in \mathcal{X}_S$ by a convergent power series. Thus, in particular, the mean function $\gamma(\mathbf{x}) = \mathsf{E}_{\mathbf{x}}\{\hat{g}(\mathbf{y})\}$ of any finite-variance estimator $\hat{g}(\mathbf{y})$ is necessarily an analytic function. This agrees with the general result about the mean function of estimators for exponential families presented in [32, Lemma 2.8]. (Note that the statistical model of the SLGM is an exponential family.)

C. Minimum Achievable Variance (Barankin Bound) and LMV Estimator

Let us consider the MVP (6) at a given parameter vector $\mathbf{x}_0 \in \mathcal{X}_S$ for an SLGM estimation problem $\mathcal{E}_{\mathrm{SLGM}} = \big(\mathcal{X}_S, f_{\mathbf{H}}(\mathbf{y}; \mathbf{x}), g(\cdot)\big)$ and for a prescribed bias function $c(\cdot) : \mathcal{X}_S \to \mathbb{R}$, which is known to be valid. The Barankin bound at $\mathbf{x}_0$, denoted $M_{\mathrm{SLGM}}(c(\cdot), \mathbf{x}_0)$ (cf. (6)), and the corresponding LMV estimator $\hat{g}^{(c(\cdot),\mathbf{x}_0)}(\cdot)$ (cf. (7)) are characterized by the following corollary of Theorem III.1 [22, Th. 5.3.1].

Corollary III.3: Consider an SLGM estimation problem $\mathcal{E}_{\mathrm{SLGM}} = \big(\mathcal{X}_S, f_{\mathbf{H}}(\mathbf{y}; \mathbf{x}), g(\cdot)\big)$ and a prescribed bias function $c(\cdot) : \mathcal{X}_S \to \mathbb{R}$ that is valid for $\mathcal{E}_{\mathrm{SLGM}}$ at $\mathbf{x}_0 \in \mathcal{X}_S$. Then:

1) *The minimum achievable variance at $\mathbf{x}_0$ is given by*

$$M_{\mathrm{SLGM}}(c(\cdot), \mathbf{x}_0) = \min_{a[\cdot] \in \mathcal{C}(c)} \|a[\cdot]\|^2_{\ell^2(\mathbb{Z}_+^D)} - \gamma^2(\mathbf{x}_0), \quad (25)$$

where $\gamma(\cdot) = c(\cdot) + g(\cdot)$ and $\mathcal{C}(c) \subseteq \ell^2(\mathbb{Z}_+^D)$ denotes the set of coefficient sequences $a[\mathbf{p}] \in \ell^2(\mathbb{Z}_+^D)$ that are consistent with (23).

2) *The function $\hat{g}(\cdot) : \mathbb{R}^M \to \mathbb{R}$ given by*

$$\hat{g}(\mathbf{y}) = \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{H}\mathbf{x}_0\|_2^2 \right) \sum_{\mathbf{p} \in \mathbb{Z}_+^D} \frac{a[\mathbf{p}]}{\sqrt{\mathbf{p}!}} \chi_{\mathbf{p}}(\mathbf{y}), \quad (26)$$

[2]Note that a function with domain $\mathcal{X}_S$, with $S < N$, cannot be analytic in the conventional sense since the domain of an analytic function has to be open by definition [20, Definition 2.2.1].

with an arbitrary coefficient sequence $a[\cdot] \in \mathcal{C}(c)$ *and*

$$\chi_{\mathbf{p}}(\mathbf{y}) \triangleq \frac{\partial^{\mathbf{p}}\big[\rho_{\mathrm{LGM},\mathbf{x}_0}(\mathbf{y},\sigma\widetilde{\mathbf{H}}\mathbf{z})\exp\big(\frac{1}{\sigma}\mathbf{x}_0^T\mathbf{H}^T\mathbf{H}\widetilde{\mathbf{H}}\mathbf{z}\big)\big]}{\partial \mathbf{z}^{\mathbf{p}}}\bigg|_{\mathbf{z}=\mathbf{0}},$$

where $\rho_{\mathrm{LGM},\mathbf{x}_0}(\mathbf{y},\mathbf{x})$ *is given by* (16)*, is an allowed estimator at* $\mathbf{x}_0$ *for* $c(\cdot)$*, i.e.,* $\hat{g}(\cdot) \in \mathcal{A}(c(\cdot),\mathbf{x}_0)$.

3) *The LMV estimator at* $\mathbf{x}_0$*,* $\hat{g}^{(c(\cdot),\mathbf{x}_0)}(\cdot)$*, is given by* (26) *using the unique coefficient sequence*

$$a_0[\mathbf{p}] = \underset{a[\cdot]\in\mathcal{C}(c)}{\operatorname{argmin}} \|a[\cdot]\|_{\ell^2(\mathbb{Z}_+^D)}.$$

Proof: see Appendix B.

The kernel $R_{\mathrm{SLGM},\mathbf{x}_0}(\cdot,\cdot)$ given by (18) is pointwise continuous with respect to the parameter $\mathbf{x}_0$, i.e., $\lim_{\mathbf{x}_0'\to\mathbf{x}_0} R_{\mathrm{SLGM},\mathbf{x}_0'}(\mathbf{x}_1,\mathbf{x}_2) = R_{\mathrm{SLGM},\mathbf{x}_0}(\mathbf{x}_1,\mathbf{x}_2)$ for all $\mathbf{x}_0,\mathbf{x}_1,\mathbf{x}_2 \in \mathcal{X}_S$. Therefore, applying [29, Th. III.6] to the SLGM yields the following result.

Corollary III.4: Consider the SLGM with parameter function $g(\mathbf{x}) = x_k$ *and a prescribed bias function* $c(\cdot): \mathcal{X}_S \to \mathbb{R}$ *that is valid for* $\mathcal{E}_{\mathrm{SLGM}} = \big(\mathcal{X}_S, f_{\mathbf{H}}(\mathbf{y};\mathbf{x}), g(\mathbf{x}) = x_k\big)$ *at each parameter vector* $\mathbf{x}_0 \in \mathcal{X}_S$*. If* $c(\cdot)$ *is continuous, then the minimum achievable variance* $M_{\mathrm{SLGM}}(c(\cdot),\mathbf{x}_0)$ *is a lower semi-continuous*[3] *function of* $\mathbf{x}_0$.

Proof: see Appendix C.

In Section IV-A, we will use Corollary III.4 to show that the sparse CRB derived in [12] cannot be tight in general.

IV. Lower Variance Bounds for the SLGM

While Corollary III.3 provides a mathematically complete characterization of the minimum achievable variance and the LMV estimator, the corresponding expressions are somewhat difficult to evaluate in general. Therefore, we next present lower bounds on the minimum achievable variance $M_{\mathrm{SLGM}}(c(\cdot),\mathbf{x}_0)$ for the estimation problem $\mathcal{E}_{\mathrm{SLGM}} = \big(\mathcal{X}_S, f_{\mathbf{H}}(\mathbf{y};\mathbf{x}), g(\mathbf{x}) = x_k\big)$ with some $k \in [N]$. These bounds are easier to evaluate. As mentioned before, they are also lower bounds on the variance of any estimator having the prescribed bias function. Our assumption that $g(\mathbf{x}) = x_k$ is no restriction because, according to [22, Th. 2.3.1], the MVP for a given parameter function $g(\mathbf{x})$ and prescribed bias function $c(\mathbf{x})$ is equivalent to the MVP for parameter function $g'(\mathbf{x}) = x_k$ and prescribed bias function $c'(\mathbf{x}) = c(\mathbf{x}) + g(\mathbf{x}) - x_k$. In particular,[4] $c'(\mathbf{x})$ is valid for $g'(\mathbf{x}) = x_k$ if and only if $c(\mathbf{x}) = c'(\mathbf{x}) - g(\mathbf{x}) + x_k$ is valid for $g(\mathbf{x})$. Therefore, any MVP can be reduced to an equivalent MVP with $g(\mathbf{x}) = x_k$ and an appropriately modified prescribed bias function.

We assume that the prescribed bias function $c(\cdot)$ is valid for $\mathcal{E}_{\mathrm{SLGM}} = \big(\mathcal{X}_S, f_{\mathbf{H}}(\mathbf{y};\mathbf{x}), g(\mathbf{x}) = x_k\big)$. This validity assumption is no real restriction either, since our lower bounds are finite and therefore are lower bounds also if $M_{\mathrm{SLGM}}(c(\cdot),\mathbf{x}_0) = \infty$, which, by our definition in Section II-B, is the case if $c(\cdot)$ is not valid.

The lower bounds to be presented are based on the generic lower bound (15), i.e., they are of the form

$$M_{\mathrm{SLGM}}(c(\cdot),\mathbf{x}_0) \geq \|\mathsf{P}_{\mathcal{U}}\gamma(\cdot)\|^2_{\mathcal{H}_{\mathrm{SLGM},\mathbf{x}_0}} - \gamma^2(\mathbf{x}_0), \qquad (27)$$

for some subspace $\mathcal{U} \subseteq \mathcal{H}_{\mathrm{SLGM},\mathbf{x}_0}$. Here, the prescribed mean function $\gamma(\cdot): \mathcal{X}_S \to \mathbb{R}$, given by $\gamma(\mathbf{x}) = c(\mathbf{x}) + x_k$, is an element of $\mathcal{H}_{\mathrm{SLGM},\mathbf{x}_0}$ since $c(\cdot)$ is assumed valid.

A. The Sparse CRB

The first bound is an adaptation of the CRB [18], [19] [27], [29] to the sparse setting and has been previously derived in a slightly different form in [12]. Consider the estimation problem $\mathcal{E}_{\mathrm{SLGM}} = \big(\mathcal{X}_S, f_{\mathbf{H}}(\mathbf{y};\mathbf{x}), g(\mathbf{x}) = x_k\big)$ with a system matrix $\mathbf{H} \in \mathbb{R}^{M\times N}$ satisfying (4). Let $\mathbf{x}_0 \in \mathcal{X}_S$. If the prescribed bias function $c(\cdot): \mathcal{X}_S \to \mathbb{R}$ is such that the partial derivatives $\frac{\partial c(\mathbf{x})}{\partial x_l}\big|_{\mathbf{x}=\mathbf{x}_0}$ exist for all $l \in [N]$, then

$$M_{\mathrm{SLGM}}(c(\cdot),\mathbf{x}_0) \geq \begin{cases} \sigma^2\mathbf{b}^T(\mathbf{H}^T\mathbf{H})^{\dagger}\mathbf{b}, & \text{if } \|\mathbf{x}_0\|_0 \leq S-1 \\ \sigma^2\mathbf{b}_{\mathbf{x}_0}^T(\mathbf{H}_{\mathbf{x}_0}^T\mathbf{H}_{\mathbf{x}_0})^{\dagger}\mathbf{b}_{\mathbf{x}_0}, & \text{if } \|\mathbf{x}_0\|_0 = S. \end{cases} \qquad (28)$$

Here, in the case $\|\mathbf{x}_0\|_0 \leq S-1$, $\mathbf{b} \in \mathbb{R}^N$ is given by $b_l \triangleq \delta_{k,l} + \frac{\partial c(\mathbf{x})}{\partial x_l}\big|_{\mathbf{x}=\mathbf{x}_0}$, $l \in [N]$, and in the case $\|\mathbf{x}_0\|_0 = S$, $\mathbf{b}_{\mathbf{x}_0} \in \mathbb{R}^S$ and $\mathbf{H}_{\mathbf{x}_0} \in \mathbb{R}^{M\times S}$ consist of those entries of $\mathbf{b}$ and columns of $\mathbf{H}$, respectively that are indexed by $\mathrm{supp}(\mathbf{x}_0) \equiv \{k_1,\ldots,k_S\}$, i.e., $(\mathbf{b}_{\mathbf{x}_0})_i = b_{k_i}$ and $(\mathbf{H}_{\mathbf{x}_0})_{m,i} = (\mathbf{H})_{m,k_i}$, $i \in [S]$.

As shown in [22, Th. 5.4.1], the bound (28) for $\|\mathbf{x}_0\|_0 \leq S-1$ is obtained from the generic bound (27) using the subspace $\mathcal{U} = \mathrm{span}\big\{u_0(\cdot), \{u_l(\cdot)\}_{l\in[N]}\big\}$, where

$$u_0(\cdot) \triangleq R_{\mathrm{SLGM},\mathbf{x}_0}(\cdot,\mathbf{x}_0),$$
$$u_l(\cdot) \triangleq \frac{\partial R_{\mathrm{SLGM},\mathbf{x}_0}(\cdot,\mathbf{x}_2)}{\partial(\mathbf{x}_2)_l}\bigg|_{\mathbf{x}_2=\mathbf{x}_0}, \quad l \in [N],$$

with $R_{\mathrm{SLGM},\mathbf{x}_0}(\cdot,\cdot)$ given by (18) (note that $R_{\mathrm{SLGM},\mathbf{x}_0}(\cdot,\cdot)$ is differentiable), and the bound (28) for $\|\mathbf{x}_0\|_0 = S$ is obtained from (27) using the subspace $\mathcal{U} = \mathrm{span}\big\{u_0(\cdot), \{u_l(\cdot)\}_{l\in\mathrm{supp}(\mathbf{x}_0)}\big\}$. This establishes a new, RKHS-based interpretation of the bound in [12] in terms of the projection of the prescribed mean function $\gamma(\mathbf{x}) = c(\mathbf{x}) + x_k$ onto an RKHS-related subspace $\mathcal{U}$. We note that the bound in [12] was formulated as a bound on the variance $v(\hat{\mathbf{x}}(\cdot);\mathbf{x}_0)$ of a vector-valued estimator $\hat{\mathbf{x}}(\cdot)$ of $\mathbf{x}$, and not only of the kth entry x_k. Consistent with (8), that bound can be reobtained by summing our bound in (28) (with $c(\cdot) = c_k(\cdot)$) over all $k \in [N]$. Thus, the two bounds are equivalent.

An important aspect of the bound in (28) is the fact that it is not a continuous function of $\mathbf{x}_0$ on $\mathcal{X}_S$ in general. Indeed, for the case $\mathbf{H} = \mathbf{I}$ and $c(\cdot) \equiv 0$, which has been considered in [14], it can be verified that the bound is a strictly upper semi-continuous function of $\mathbf{x}_0$: for example, for $M = N = 2$, $\mathbf{H} = \mathbf{I}$, $c(\cdot) \equiv 0$, $S = 1$, $k = 2$, and $\mathbf{x}_0 = a\cdot(1\ 0)^T$ with $a \in \mathbb{R}_+$, the bound is equal to 1 for $a = 0$ (case of $\|\mathbf{x}_0\|_0 \leq S-1$) but equal to 0 for all $a > 0$ (case of $\|\mathbf{x}_0\|_0 = S$).

[3] A definition of lower semi-continuity is provided in Appendix C-B.

[4] Indeed, if $c'(\mathbf{x})$ is valid at $\mathbf{x}_0$ for the MVP with parameter function x_k, there exists a finite-variance estimator $\hat{g}(\cdot)$ with mean function $\mathsf{E}_{\mathbf{x}}\{\hat{g}(\mathbf{y})\} = c'(\mathbf{x}) + x_k$. For the MVP with parameter function $g(\cdot)$, that estimator $\hat{g}(\cdot)$ has the bias function

$$b(\hat{g}(\cdot),\mathbf{x}) = \mathsf{E}_{\mathbf{x}}\{\hat{g}(\mathbf{y})\} - g(\mathbf{x}) = c'(\mathbf{x}) + x_k - g(\mathbf{x}) = c(\mathbf{x}).$$

Thus, there exists a finite-variance estimator with bias function $c(\mathbf{x}) = c'(\mathbf{x}) - g(\mathbf{x}) + x_k$, which implies that the bias function $c(\cdot)$ is valid for the MVP with parameter function $g(\cdot)$. This reasoning can also be performed in the reverse direction. Note that, trivially, the estimator $\hat{g}(\cdot)$ has the same variance at $\mathbf{x}_0$ for both MVPs.

However, by Corollary III.4, the minimum achievable variance $M_{\mathrm{SLGM}}(c(\cdot),\mathbf{x}_0)$ is a lower semi-continuous function of $\mathbf{x}_0$. Since a function cannot be simultaneously lower semi-continuous and strictly upper semi-continuous, it follows that the sparse CRB in (28) cannot be tight in general, i.e., it cannot be equal to $M_{\mathrm{SLGM}}(c(\cdot),\mathbf{x}_0)$ for all $\mathbf{x}_0 \in \mathcal{X}_S$. This means that we have a strict inequality in (28) at least for some $\mathbf{x}_0 \in \mathcal{X}_S$.

Let us finally consider the special case where $M \geq N$ and $\mathbf{H} \in \mathbb{R}^{M\times N}$ has full rank, i.e., $\mathrm{rank}(\mathbf{H}) = N$. The least-squares (LS) estimator [18], [19], [27] of x_k is given by $\hat{x}_{\mathrm{LS},k}(\mathbf{y}) = \mathbf{e}_k^T \mathbf{H}^{\dagger}\mathbf{y}$; it is unbiased and its variance is

$$v(\hat{x}_{\mathrm{LS},k}(\cdot);\mathbf{x}_0) = \sigma^2 \mathbf{e}_k^T(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{e}_k. \tag{29}$$

On the other hand, for unbiased estimation, i.e., $c(\cdot)\equiv 0$, our lower bound for $\|\mathbf{x}_0\|_0 \leq S-1$ in (28) becomes $M_{\mathrm{SLGM}}(c(\cdot)\equiv 0,\mathbf{x}_0) \geq \sigma^2\mathbf{b}^T(\mathbf{H}^T\mathbf{H})^{\dagger}\mathbf{b} = \sigma^2\mathbf{e}_k^T(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{e}_k$. Comparing with (29), we conclude that our bound is tight and the minimum achievable variance is in fact

$$M_{\mathrm{SLGM}}(c(\cdot)\equiv 0,\mathbf{x}_0) = \sigma^2\mathbf{e}_k^T(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{e}_k,$$

which is achieved by the LS estimator. Thus, for $M \geq N$ and $\mathrm{rank}(\mathbf{H}) = N$, the LS estimator is the[5] LMV unbiased estimator for the SLGM at each parameter vector $\mathbf{x}_0 \in \mathcal{X}_S$ with $\|\mathbf{x}_0\|_0 \leq S$–1. It is interesting to note that the LS estimator does not exploit the sparsity information expressed by the parameter set $\mathcal{X}_S$, i.e., the knowledge that $\|\mathbf{x}\|_0 \leq S$, and that it has the constant variance (29) for each $\mathbf{x}_0 \in \mathcal{X}_S$ (in fact, even for $\mathbf{x}_0 \in \mathbb{R}^N$). We also note that the LS estimator is not an LMV unbiased estimator for the case $\|\mathbf{x}_0\|_0 = S$; therefore, it is not a uniformly minimum variance (UMV) estimator [16], [17], [19] on $\mathcal{X}_S$ (i.e., an unbiased estimator with minimum variance at each $\mathbf{x}_0 \in \mathcal{X}_S$). In fact, as shown in [14] and [22], there does not exist a UMV unbiased estimator for the SLGM in general.

B. A New CRB-Type Lower Variance Bound

A new lower bound on $M_{\mathrm{SLGM}}(c(\cdot),\mathbf{x}_0)$ is stated in the following theorem [33]. This bound follows from the generic lower bound (27) by using the subspace $\mathcal{U} = \mathrm{span}\{u_0(\cdot),\{\tilde{u}_i(\cdot)\}_{i\in[|\mathcal{K}|]}\}$, with

$$u_0(\cdot) \triangleq R_{\mathrm{SLGM},\mathbf{x}_0}(\cdot,\mathbf{x}_0),$$
$$\tilde{u}_i(\cdot) \triangleq \frac{\partial R_{\mathrm{SLGM},\mathbf{x}_0}(\cdot,\mathbf{x}_2)}{\partial(\mathbf{x}_2)_{l_i}}\bigg|_{\mathbf{x}_2=\tilde{\mathbf{x}}_0}, \quad i\in[|\mathcal{K}|], \tag{30}$$

where $\tilde{\mathbf{x}}_0 \in \mathbb{R}^N$ and $\mathcal{K} = \{l_1,\ldots,l_{|\mathcal{K}|}\} \subseteq [N]$ are defined in the theorem.

Theorem IV.1: Consider the estimation problem $\mathcal{E}_{\mathrm{SLGM}} = (\mathcal{X}_S, f_{\mathbf{H}}(\mathbf{y};\mathbf{x}), g(\mathbf{x}) = x_k)$ *with a system matrix* $\mathbf{H} \in \mathbb{R}^{M\times N}$ *satisfying* (4). *Let* $\mathbf{x}_0 \in \mathcal{X}_S$, *and consider an arbitrary index set* $\mathcal{K} = \{l_1,\ldots,l_{|\mathcal{K}|}\} \subseteq [N]$ *consisting of no more than S indices, i.e.,* $|\mathcal{K}| \leq S$. *If the prescribed bias function* $c(\cdot): \mathcal{X}_S \to \mathbb{R}$ *is such that the partial derivatives* $\frac{\partial c(\mathbf{x})}{\partial x_{l_i}}\big|_{\mathbf{x}=\mathbf{x}_0}$ *exist for all* $i \in [|\mathcal{K}|]$, *then*[6]

$$M_{\mathrm{SLGM}}(c(\cdot),\mathbf{x}_0) \geq \exp\Big(-\frac{1}{\sigma^2}\|(\mathbf{I}-\mathbf{P})\mathbf{H}\mathbf{x}_0\|_2^2\Big) \times\sigma^2\mathbf{b}_{\mathbf{x}_0}^T\big(\mathbf{H}_{\mathcal{K}}^T\mathbf{H}_{\mathcal{K}}\big)^{-1}\mathbf{b}_{\mathbf{x}_0}. \tag{31}$$

Here, $\mathbf{P} \triangleq \mathbf{H}_{\mathcal{K}}(\mathbf{H}_{\mathcal{K}})^{\dagger} \in \mathbb{R}^{M\times M}$ *and* $\mathbf{b}_{\mathbf{x}_0} \in \mathbb{R}^{|\mathcal{K}|}$ *is defined entrywise as* $(\mathbf{b}_{\mathbf{x}_0})_i \triangleq \delta_{k,l_i} + \frac{\partial c(\mathbf{x})}{\partial x_{l_i}}\big|_{\mathbf{x}=\tilde{\mathbf{x}}_0}$ *for* $i \in [|\mathcal{K}|]$, *where* $\tilde{\mathbf{x}}_0 \in \mathbb{R}^N$ *is the unique (due to* (4)*) vector with* $\mathrm{supp}(\tilde{\mathbf{x}}_0) \subseteq \mathcal{K}$ *solving* $\mathbf{H}\tilde{\mathbf{x}}_0 = \mathbf{P}\mathbf{H}\mathbf{x}_0$.

Proof: see Appendix D.

The bound (31) has an intuitively appealing interpretation in terms of a scaled CRB for an LGM. Indeed, the quantity $\sigma^2\mathbf{b}_{\mathbf{x}_0}^T\big(\mathbf{H}_{\mathcal{K}}^T\mathbf{H}_{\mathcal{K}}\big)^{-1}\mathbf{b}_{\mathbf{x}_0}$ appearing in (31) can be interpreted as the CRB [18] for the LGM with parameter dimension $N = |\mathcal{K}|$, parameter function $g(\mathbf{x}) = x_k$, and prescribed bias function $c(\cdot)$. For a discussion of the scaling factor $\exp\big(-\frac{1}{\sigma^2}\|(\mathbf{I}-\mathbf{P})\mathbf{H}\mathbf{x}_0\|_2^2\big)$, we will consider the following two complementary cases:

1) Consider the case where $k \notin \mathrm{supp}(\mathbf{x}_0)$ and $\|\mathbf{x}_0\|_0 = S$. Let us choose $\mathcal{K} = \mathcal{L}\cup\{k\}$, where $\mathcal{L}$ comprises the indices of the $S-1$ largest (in magnitude) entries of $\mathbf{x}_0$. We then obtain $\|(\mathbf{I}-\mathbf{P})\mathbf{H}\mathbf{x}_0\|_2^2 = \xi_0^2\|(\mathbf{I}-\mathbf{P})\mathbf{H}\mathbf{e}_{j_0}\|_2^2$, where ξ_0 and j_0 denote the value and index, respectively, of the smallest (in magnitude) nonzero entry of $\mathbf{x}_0$.[7] Typically,[8] $\|(\mathbf{I}-\mathbf{P})\mathbf{H}\mathbf{e}_{j_0}\|_2^2 > 0$ and therefore, as ξ_0 becomes larger (in magnitude), the bound (31) transitions from a "low signal-to-noise ratio (SNR)" regime, where $\exp\big(-\frac{1}{\sigma^2}\|(\mathbf{I}-\mathbf{P})\mathbf{H}\mathbf{x}_0\|_2^2\big) \approx 1$, to a "high-SNR" regime, where $\exp\big(-\frac{1}{\sigma^2}\|(\mathbf{I}-\mathbf{P})\mathbf{H}\mathbf{x}_0\|_2^2\big) \approx 0$. In the low-SNR regime, the bound (31) is approximately equal to $\sigma^2\mathbf{b}_{\mathbf{x}_0}^T\big(\mathbf{H}_{\mathcal{K}}^T\mathbf{H}_{\mathcal{K}}\big)^{-1}\mathbf{b}_{\mathbf{x}_0}$, i.e., to the CRB for the LGM with $N = |\mathcal{K}|$. In the high-SNR regime, the bound becomes approximately equal to 0; this suggests that the zero entries x_k with $k \notin \mathrm{supp}(\mathbf{x})$ can be estimated with small variance. Note that for increasing ξ_0, the transition from the low-SNR regime to the high-SNR regime exhibits an exponential decay.
2) On the other hand, in the complementary case where either $k \in \mathrm{supp}(\mathbf{x}_0)$ or $\|\mathbf{x}_0\|_0 \leq S-1$ (or both), the factor $\exp\big(-\frac{1}{\sigma^2}\|(\mathbf{I}-\mathbf{P})\mathbf{H}\mathbf{x}_0\|_2^2\big)$ can be made equal to 1 by choosing $\mathcal{K} = \mathrm{supp}(\mathbf{x}_0)\cup\{k\}$.

We note that the bound presented in [33] is obtained by maximizing (31) with respect to the index set $\mathcal{K}$; this gives the tightest possible bound of the type (31).

The matrix $\mathbf{P}$ appearing in (31) is the orthogonal projection matrix [21] on the subspace $\mathcal{H}_{\mathcal{K}} \triangleq \mathrm{span}(\mathbf{H}_{\mathcal{K}})$

[5]If an LMV estimator exists, then it is unique [19].

[6]Note that $\big(\mathbf{H}_{\mathcal{K}}^T\mathbf{H}_{\mathcal{K}}\big)^{-1}$ exists because of (4).

[7]Indeed, we have $\mathbf{H}\mathbf{x}_0 = \mathbf{H}_{\mathcal{L}}\mathbf{x}_{0,\mathcal{L}} + \xi_0\mathbf{H}\mathbf{e}_{j_0}$, where $\mathbf{H}_{\mathcal{L}}$ and $\mathbf{x}_{0,\mathcal{L}}$ denote, respectively, the restriction of the matrix $\mathbf{H}$ and the vector $\mathbf{x}_0$ to the columns and entries indexed by $\mathcal{L}$. The component $\mathbf{H}_{\mathcal{L}}\mathbf{x}_{0,\mathcal{L}}$ belongs to the subspace $\mathrm{span}(\mathbf{H}_{\mathcal{L}})$. Because $\mathrm{span}(\mathbf{H}_{\mathcal{L}}) \subseteq \mathrm{span}(\mathbf{H}_{\mathcal{K}})$, the orthogonal projection matrix $\mathbf{I}-\mathbf{P}$ suppresses any vector component belonging to $\mathrm{span}(\mathbf{H}_{\mathcal{L}})$. It then follows that

$$(\mathbf{I}-\mathbf{P})\mathbf{H}\mathbf{x}_0 = (\mathbf{I}-\mathbf{P})(\mathbf{H}_{\mathcal{L}}\mathbf{x}_{0,\mathcal{L}} + \xi_0\mathbf{H}\mathbf{e}_{j_0}) = \xi_0(\mathbf{I}-\mathbf{P})\mathbf{H}\mathbf{e}_{j_0}.$$

[8]Note that, for the case $k \notin \mathrm{supp}(\mathbf{x}_0)$ and $\|\mathbf{x}_0\|_0 = S$ considered, $j_0 \notin \mathcal{K}$ with $|\mathcal{K}| = S$. For a system matrix $\mathbf{H}$ satisfying (4), we then have $\|(\mathbf{I}-\mathbf{P})\mathbf{H}\mathbf{e}_{j_0}\|_2^2 > 0$ if and only if the submatrix $\mathbf{H}_{\mathcal{K}\cup\{j_0\}}$ has full column rank.

$\subseteq \mathbb{R}^M$, i.e., the subspace spanned by those columns of $\mathbf{H}$ whose indices are in $\mathcal{K}$. Consequently, $\mathbf{I}-\mathbf{P}$ is the orthogonal projection matrix on the orthogonal complement of $\mathcal{H}_{\mathcal{K}}$, and the norm $\|(\mathbf{I}-\mathbf{P})\mathbf{H}\mathbf{x}_0\|_2$ thus represents the distance between the point $\mathbf{H}\mathbf{x}_0$ and the subspace $\mathcal{H}_{\mathcal{K}}$ [31]. Therefore, the factor $\exp\big(-\frac{1}{\sigma^2}\|(\mathbf{I}-\mathbf{P})\mathbf{H}\mathbf{x}_0\|_2^2\big)$ appearing in the bound (31) can be interpreted as a measure of the distance between $\mathbf{H}\mathbf{x}_0$ and $\mathcal{H}_{\mathcal{K}}$. In general, the bound (31) is tighter (i.e., higher) if $\mathcal{K}$ is chosen such that the distance $\|(\mathbf{I}-\mathbf{P})\mathbf{H}\mathbf{x}_0\|_2$ is smaller.

For the special case given by the SSNM, i.e., $\mathbf{H} = \mathbf{I}$, and unbiased estimation, i.e., $c(\cdot) \equiv 0$, the bound (31) is a continuous function of $\mathbf{x}_0$ on $\mathcal{X}_S$. This is an important difference from the bound in (28) and, also, from the bound to be given in (50) below. Furthermore, still for $\mathbf{H} = \mathbf{I}$ and $c(\cdot) \equiv 0$, the bound (31) can be shown [33], [22, p. 106] to be tighter (higher) than the bounds given by (28) and (50).

C. The SLGM View of Compressed Sensing

The lower bounds for the SLGM presented in Sections IV-A and IV-B are also relevant to the linear CS recovery problem, which can be viewed as an instance of the SLGM estimation problem. In this section, we express the new lower bound in Theorem IV.1 in terms of the restricted isometry constant of the system matrix (CS measurement matrix) $\mathbf{H}$.

The compressive measurement process within a CS problem is often modeled as [2], [7], [8], [23], [34], [35]

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}. \qquad (32)$$

Here, $\mathbf{y} \in \mathbb{R}^M$ denotes the compressive measurements; $\mathbf{H} \in \mathbb{R}^{M\times N}$, where $M \leq N$ and typically $M \ll N$, denotes the CS measurement matrix; $\mathbf{x} \in \mathcal{X}_S \subseteq \mathbb{R}^N$ is an unknown S-sparse signal or parameter vector, with known sparsity degree S (typically $S \ll N$); and $\mathbf{n}$ represents additive measurement noise. We assume that $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ and that the columns $\{\mathbf{h}_j\}_{j\in[N]}$ of $\mathbf{H}$ are normalized, i.e., $\|\mathbf{h}_j\|_2 = 1$ for all $j \in [N]$. The CS measurement model (32) is then identical to the SLGM observation model (2). Any CS recovery method, such as basis pursuit (BP) [34], [36] or orthogonal matching pursuit (OMP) [23], [37], can be interpreted as an estimator $\hat{\mathbf{x}}(\mathbf{y})$ that estimates the sparse vector $\mathbf{x}$ from the observation $\mathbf{y}$.

Due to the typically large dimension of the measurement matrix $\mathbf{H}$, a complete characterization of the properties of $\mathbf{H}$ (e.g., via its SVD) is often infeasible. Useful incomplete characterizations are provided by the (mutual) coherence and the restricted isometry property [7], [23], [34], [35]. A matrix $\mathbf{H} \in \mathbb{R}^{M\times N}$ is said to satisfy the *restricted isometry property* (RIP) of order K if there is a constant $\delta'_K \in \mathbb{R}_+$ such that for every index set $\mathcal{I} \subseteq [N]$ of size $|\mathcal{I}| = K$,

$$(1-\delta'_K)\|\mathbf{z}\|_2^2 \;\leq\; \|\mathbf{H}_{\mathcal{I}}\mathbf{z}\|_2^2 \;\leq\; (1+\delta'_K)\|\mathbf{z}\|_2^2, \quad \text{for all } \mathbf{z} \in \mathbb{R}^K. \qquad (33)$$

The smallest δ'_K for which (33) holds—hereafter denoted δ_K—is called the *RIP constant* of $\mathbf{H}$. Condition (4) is necessary for a matrix $\mathbf{H}$ to satisfy the RIP of order S with a RIP constant $\delta_S < 1$.[9] It can be easily verified that $\delta_{K'} \geq \delta_K$ for $K' \geq K$.

We now specialize the bound (31) on the minimum achievable variance for $\mathcal{E}_{\text{SLGM}}$ to the CS scenario, i.e., to the SLGM with sparsity degree S and a system matrix $\mathbf{H}$ that is a CS measurement matrix (i.e., $M \leq N$) with known RIP constant $\delta_S < 1$. Note that $\delta_S < 1$ implies that condition (4) is satisfied. The following result was presented in [22, Th. 5.7.2].

Corollary IV.2: Consider the SLGM estimation problem $\mathcal{E}_{\text{SLGM}} = \big(\mathcal{X}_S, f_{\mathbf{H}}(\mathbf{y};\mathbf{x}), g(\mathbf{x}) = x_k\big)$, where $\mathbf{H} \in \mathbb{R}^{M\times N}$ with $M \leq N$ satisfies the RIP of order S with RIP constant $\delta_S < 1$. Let $\mathbf{x}_0 \in \mathcal{X}_S$, and consider an arbitrary index set $\mathcal{K} \subseteq [N]$ consisting of no more than S indices, i.e., $|\mathcal{K}| \leq S$. If the first-order partial derivatives $\frac{\partial c(\mathbf{x})}{\partial x_l}\big|_{\mathbf{x}=\mathbf{x}_0}$ of the prescribed bias function $c(\cdot): \mathcal{X}_S \to \mathbb{R}$ exist for all $l \in \mathcal{K}$, then

$$M_{\text{SLGM}}(c(\cdot),\mathbf{x}_0) \geq \exp\!\left(-\frac{1+\delta_S}{\sigma^2}\,\big\|\mathbf{x}_0^{\text{supp}(\mathbf{x}_0)\setminus\mathcal{K}}\big\|_2^2\right)\frac{\sigma^2\|\mathbf{b}_{\mathbf{x}_0}\|_2^2}{1+\delta_S}, \qquad (34)$$

with $\mathbf{b}_{\mathbf{x}_0} \in \mathbb{R}^{|\mathcal{K}|}$ as defined in Theorem IV.1.

Proof: see Appendix E.

For a comparison of the actual variance behavior of a given CS recovery scheme (or, estimator) $\hat{x}_k(\cdot)$ with the variance bound (34), we set $c(\cdot)$ in (34) equal to the estimator's bias function, i.e., $c(\mathbf{x}) = \mathsf{E}_{\mathbf{x}}\{\hat{x}_k(\mathbf{y})\} - x_k$. We note that the first-order partial derivatives of $c(\mathbf{x})$, which occur in Corollary IV.2, are given by [32, Corollary 2.6]

$$\frac{\partial c(\mathbf{x})}{\partial x_l} = \delta_{k,l} + \frac{1}{\sigma^2}\mathsf{E}_{\mathbf{x}}\big\{\hat{x}_k(\mathbf{y})(\mathbf{y}-\mathbf{H}\mathbf{x})^T\mathbf{H}\mathbf{e}_l\big\}. \qquad (35)$$

For a "good" CS measurement matrix—i.e., a matrix with small δ_S—the bound (34) is very close to a bound for the SSNM (i.e., for $\mathbf{H}=\mathbf{I}$) that will be presented in Section V-C (see (49) below). This means that, conversely, in terms of a lower bound on the achievable variance, relative to the SSNM (case $\mathbf{H}=\mathbf{I}$), no loss of information is incurred by multiplying $\mathbf{x}$ by the CS measurement matrix $\mathbf{H} \in \mathbb{R}^{M\times N}$ and thereby reducing the signal dimension from N to M, where typically $M \ll N$. This agrees with the fact that for small δ_S, one can recover—e.g., by using BP—the sparse parameter vector $\mathbf{x} \in \mathcal{X}_S$ from the compressed observation $\mathbf{y} = \mathbf{H}\mathbf{x}+\mathbf{n}$ up to an error that is typically very small (and whose norm is almost independent of $\mathbf{H}$ and solely determined by the measurement noise $\mathbf{n}$ [7], [38]).

V. RKHS-Based Analysis of Minimum Variance Estimation for the SSNM

Next, we specialize our RKHS-based MVE analysis to the SSNM, i.e., to the case given by $\mathbf{H} = \mathbf{I}$ (which implies $M = N$ and $\mathbf{y} = \mathbf{x} + \mathbf{n}$). We note that the SLGM with a system matrix $\mathbf{H} \in \mathbb{R}^{M\times N}$ having orthonormal columns, i.e., satisfying $\mathbf{H}^T\mathbf{H} = \mathbf{I}$, is equivalent to the SSNM [14].

[9] Indeed, assume that $\text{spark}(\mathbf{H}) \leq S$. This means that there exists an index set $\mathcal{I} \subseteq [N]$ consisting of S indices such that the columns of $\mathbf{H}_{\mathcal{I}}$ are linearly dependent. This, in turn, implies that there is a nonzero coefficient vector $\mathbf{z} \in \mathbb{R}^S$ such that $\mathbf{H}_{\mathcal{I}}\mathbf{z} = \mathbf{0}$ and consequently $\|\mathbf{H}_{\mathcal{I}}\mathbf{z}\|_2^2 = 0$. Therefore, there cannot exist a constant $\delta'_K < 1$ for which (33) holds.

Specializing the kernel $R_{\text{SLGM},\mathbf{x}_0}(\cdot,\cdot)$ (see (18)) to $\mathbf{H}=\mathbf{I}$, we obtain

$$R_{\text{SSNM},\mathbf{x}_0}(\mathbf{x}_1,\mathbf{x}_2) = \exp\left(\frac{1}{\sigma^2}(\mathbf{x}_2-\mathbf{x}_0)^T(\mathbf{x}_1-\mathbf{x}_0)\right),\ \mathbf{x}_0,\mathbf{x}_1,\mathbf{x}_2\in\mathcal{X}_S. \quad (36)$$

The corresponding RKHS, $\mathcal{H}(R_{\text{SSNM},\mathbf{x}_0})$, will be briefly denoted by $\mathcal{H}_{\text{SSNM},\mathbf{x}_0}$.

A. Valid Bias Functions, Minimum Achievable Variance, and LMV Estimator

Since the SSNM is a special case of the SLGM, we can characterize the class of valid bias functions, the Barankin bound, and the corresponding LMV estimator by Corollary III.2 and Corollary III.3 with $\mathbf{H}=\mathbf{I}$. However, a more convenient characterization [22, Th. 5.5.2] can be obtained by exploiting the specific structure of $\mathcal{H}_{\text{SSNM},\mathbf{x}_0}$ that is induced by the choice $\mathbf{H}=\mathbf{I}$.

Theorem V.1: Consider the SSNM estimation problem $\mathcal{E}_{\text{SSNM}}=\big(\mathcal{X}_S, f_{\mathbf{I}}(\mathbf{y};\mathbf{x}), g(\cdot)\big)$.

1) *A prescribed bias function $c(\cdot):\mathcal{X}_S\to\mathbb{R}$ is valid for $\mathcal{E}_{\text{SSNM}}$ at $\mathbf{x}_0\in\mathcal{X}_S$ if and only if the corresponding mean function $\gamma(\cdot)=c(\cdot)+g(\cdot)$ can be expressed as*

$$\gamma(\mathbf{x}) = \nu_{\mathbf{x}_0}(\mathbf{x})\sum_{\mathbf{p}\in\mathbb{Z}_+^N\cap\mathcal{X}_S}\frac{a[\mathbf{p}]}{\sqrt{\mathbf{p}!}}\left(\frac{\mathbf{x}}{\sigma}\right)^{\mathbf{p}},\quad \mathbf{x}\in\mathcal{X}_S, \quad (37)$$

with

$$\nu_{\mathbf{x}_0}(\mathbf{x}) \triangleq \exp\left(\frac{1}{2\sigma^2}\|\mathbf{x}_0\|_2^2-\frac{1}{\sigma^2}\mathbf{x}^T\mathbf{x}_0\right) \quad (38)$$

and with a coefficient sequence $a[\mathbf{p}]\in\ell^2(\mathbb{Z}_+^N\cap\mathcal{X}_S)$. This coefficient sequence is unique for a given $c(\cdot)$.

2) *Let the prescribed bias function $c(\cdot):\mathcal{X}_S\to\mathbb{R}$ be valid for $\mathcal{E}_{\text{SSNM}}$ at $\mathbf{x}_0\in\mathcal{X}_S$. Then:*
   a) *The minimum achievable variance at $\mathbf{x}_0\in\mathcal{X}_S$ is given by*

$$M_{\text{SSNM}}(c(\cdot),\mathbf{x}_0) = \sum_{\mathbf{p}\in\mathbb{Z}_+^N\cap\mathcal{X}_S} a_{\mathbf{x}_0}^2[\mathbf{p}]-\gamma^2(\mathbf{x}_0), \quad (39)$$

   with

$$a_{\mathbf{x}_0}[\mathbf{p}] \triangleq \frac{1}{\sqrt{\mathbf{p}!}}\frac{\partial^{\mathbf{p}}\big(\gamma(\sigma\mathbf{x})/\nu_{\mathbf{x}_0}(\sigma\mathbf{x})\big)}{\partial\mathbf{x}^{\mathbf{p}}}\bigg|_{\mathbf{x}=\mathbf{0}}. \quad (40)$$

   b) *The LMV estimator at $\mathbf{x}_0$ is given by*

$$\hat{g}^{(c(\cdot),\mathbf{x}_0)}(\mathbf{y}) = \sum_{\mathbf{p}\in\mathbb{Z}_+^N\cap\mathcal{X}_S}\frac{a_{\mathbf{x}_0}[\mathbf{p}]}{\sqrt{\mathbf{p}!}}\frac{\partial^{\mathbf{p}}\psi_{\mathbf{x}_0}(\mathbf{x},\mathbf{y})}{\partial\mathbf{x}^{\mathbf{p}}}\bigg|_{\mathbf{x}=\mathbf{0}}, \quad (41)$$

   with

$$\psi_{\mathbf{x}_0}(\mathbf{x},\mathbf{y}) \triangleq \exp\left(\frac{\mathbf{y}^T(\sigma\mathbf{x}-\mathbf{x}_0)}{\sigma^2}+\frac{\mathbf{x}_0^T\mathbf{x}}{\sigma}-\frac{\|\mathbf{x}\|_2^2}{2}\right). \quad (42)$$

Proof: see Appendix F.

Note that the statement of Theorem V.1 is stronger than that of Corollary III.3, because it contains explicit expressions of the minimum achievable variance $M_{\text{SSNM}}(c(\cdot),\mathbf{x}_0)$ and the corresponding LMV estimator $\hat{g}^{(c(\cdot),\mathbf{x}_0)}(\mathbf{y})$.

The expression (39) nicely shows the influence of the sparsity constraints on the minimum achievable variance. Indeed, consider a bias function $c(\cdot):\mathbb{R}^N\to\mathbb{R}$ that is valid for the SSNM with $S=N$, and therefore also for the SSNM with $S<N$. Let us denote by M_N and M_S the minimum achievable variance $M(c(\cdot),\mathbf{x}_0)$ for the degenerate SSNM without sparsity ($S=N$) and for the SSNM with sparsity ($S<N$), respectively. Note that in the nonsparse case $S=N$, the SSNM coincides with the LGM with $\mathbf{H}=\mathbf{I}$. It then follows from (39) that $M_N=\sum_{\mathbf{p}\in\mathbb{Z}_+^N}a_{\mathbf{x}_0}^2[\mathbf{p}]-\gamma^2(\mathbf{x}_0)$ and

$$M_N - M_S = \sum_{\mathbf{p}\in\mathbb{Z}_+^N\setminus\mathcal{X}_S} a_{\mathbf{x}_0}^2[\mathbf{p}].$$

Clearly, if $\mathbf{x}$ is more sparse, i.e., if the sparsity degree S is smaller, the number of (nonnegative) terms in the above sum is larger. This implies a larger difference M_N-M_S and, thus, a stronger reduction of the minimum achievable variance due to the sparsity information.

We mention the obvious fact that a UMV estimator for $\mathcal{E}_{\text{SSNM}}=\big(\mathcal{X}_S, f_{\mathbf{I}}(\mathbf{y};\mathbf{x}), g(\cdot)\big)$ and bias function $c(\cdot)$ exists if and only if the LMV estimator $\hat{g}^{(c(\cdot),\mathbf{x}_0)}(\cdot)$ given by (41) does not depend on $\mathbf{x}_0$.

Consider the SSNM with parameter function $g(\mathbf{x})=x_k$, i.e., $\mathcal{E}_{\text{SSNM}}=\big(\mathcal{X}_S, f_{\mathbf{I}}(\mathbf{y};\mathbf{x}), g(\mathbf{x})=x_k\big)$, for some $k\in[N]$. Because the specific estimator $\hat{g}(\mathbf{y})=y_k$ has finite variance and zero bias at each $\mathbf{x}\in\mathcal{X}_S$, the bias function $c_u(\mathbf{x})\equiv 0$ must be valid for $\mathcal{E}_{\text{SSNM}}$ at each $\mathbf{x}_0\in\mathcal{X}_S$. Therefore, according to Corollary III.4, the minimum achievable variance for unbiased estimation within the SSNM with parameter function $g(\mathbf{x})=x_k$, $M_{\text{SSNM}}(c_u(\cdot),\mathbf{x}_0)$, is a lower semi-continuous function of $\mathbf{x}_0$ on its domain, i.e., on $\mathcal{X}_S$. (Note that this remark is not related to Theorem V.1.)

Finally, we note that the explicit expression (41) for the LMV estimator is mainly of theoretical interest, since it depends on the unknown parameter $\mathbf{x}_0$ and therefore cannot be implemented in practice. However, this expression might give some intuition on how to design practical estimators whose performance comes close to that of the LMV estimator.

B. Diagonal Bias Functions

In this subsection, we consider the SSNM estimation problem[10] $\mathcal{E}_{\text{SSNM}}=\big(\mathcal{X}_S, f_{\mathbf{I}}(\mathbf{y};\mathbf{x}), g(\mathbf{x})=x_k\big)$, for some $k\in[N]$, and study a specific class of bias functions. Let us call a bias function $c(\cdot):\mathcal{X}_S\to\mathbb{R}$ *diagonal* if $c(\mathbf{x})$ depends only on the kth entry of the parameter vector $\mathbf{x}$, i.e., the specific scalar parameter x_k to be estimated. That is, $c(\mathbf{x})=\tilde{c}(x_k)$, with some function $\tilde{c}(\cdot):\mathbb{R}\to\mathbb{R}$ that may depend on k. Similarly, we say that an estimator $\hat{x}_k(\mathbf{y})$ is diagonal if it depends only on the kth entry of $\mathbf{y}$, i.e., $\hat{x}_k(\mathbf{y})=\hat{x}_k(y_k)$ (with an abuse of notation). Clearly, the bias function $b(\hat{x}_k(\cdot);\mathbf{x})$ of a diagonal

[10] We recall that the assumption $g(\mathbf{x})=x_k$ is no restriction, because the MVP for any given parameter function $g(\cdot)$ is equivalent to the MVP for the parameter function $g'(\mathbf{x})=x_k$ and the modified prescribed bias function $c'(\mathbf{x})=c(\mathbf{x})+g(\mathbf{x})-x_k$.

estimator $\hat{x}_k(\cdot)$ is diagonal, i.e., $b(\hat{x}_k(\cdot); \mathbf{x}) = b(\hat{x}_k(\cdot); x_k)$. Well-known examples of diagonal estimators are the hard- and soft-thresholding estimators described in [2], [11], and [39] and the LS estimator, $\hat{x}_{\mathrm{LS},k}(\mathbf{y}) = y_k$. The maximum likelihood estimator for the SSNM is not diagonal, and neither is its bias function [14].

The following theorem [22, Th. 5.5.4], which can be regarded as a specialization of Theorem V.1 to the case of diagonal bias functions, provides a characterization of the class of valid diagonal bias functions, as well as of the minimum achievable variance and LMV estimator for a prescribed diagonal bias function. In the theorem, we will use the lth order (probabilists') Hermite polynomial $H_l(\cdot) : \mathbb{R} \to \mathbb{R}$ defined as [40]

$$H_l(x) \triangleq (-1)^l e^{x^2/2} \frac{d^l}{dx^l} e^{-x^2/2}.$$

Furthermore, in the case $\|\mathbf{x}_0\|_0 = S$, the support of $\mathbf{x}_0$ will be denoted as $\mathrm{supp}(\mathbf{x}_0) = \{k_1, \ldots, k_S\}$.

Theorem V.2: Consider the SSNM estimation problem $\mathcal{E}_{\mathrm{SSNM}} = \big(\mathcal{X}_S, f_{\mathbf{I}}(\mathbf{y}; \mathbf{x}), g(\mathbf{x}) = x_k\big)$, $k \in [N]$, *at* $\mathbf{x}_0 \in \mathcal{X}_S$. *Furthermore consider a bias function* $c(\cdot) : \mathcal{X}_S \to \mathbb{R}$ *that is diagonal and such that the corresponding mean function* $\gamma(\mathbf{x}) = c(\mathbf{x}) + x_k$ *can be written as a convergent power series centered at* $\mathbf{x}_0$, *i.e.,*

$$\gamma(\mathbf{x}) = \sum_{l \in \mathbb{Z}_+} \frac{m_l}{l!} (x_k - x_{0,k})^l, \tag{43}$$

with suitable coefficients m_l. *(Note, in particular, that* $m_0 = \gamma(\mathbf{x}_0)$.*) In what follows, let*

$$B_c \triangleq \sum_{l \in \mathbb{Z}_+} \frac{m_l^2 \sigma^{2l}}{l!}.$$

1) *The bias function* $c(\cdot)$ *is valid at* $\mathbf{x}_0$ *if and only if* $B_c < \infty$.
2) *Assume that* $B_c < \infty$, *i.e.,* $c(\cdot)$ *is valid. Then:*
   a) *The minimum achievable variance at* $\mathbf{x}_0$ *is*

   $$M_{\mathrm{SSNM}}(c(\cdot), \mathbf{x}_0) = B_c \, \phi(\mathbf{x}_0) - \gamma^2(\mathbf{x}_0),$$

   with

   $$\phi(\mathbf{x}_0) \triangleq \begin{cases} 1, & \text{if } |\mathrm{supp}(\mathbf{x}_0) \cup \{k\}| \leq S \\ \displaystyle\sum_{i \in [S]} \exp\!\left(-\frac{x_{0,k_i}^2}{\sigma^2}\right) \prod_{j \in [i-1]} \left[1 - \exp\!\left(-\frac{x_{0,k_j}^2}{\sigma^2}\right)\right] < 1, & \text{if } |\mathrm{supp}(\mathbf{x}_0) \cup \{k\}| = S+1. \end{cases} \tag{44}$$

   (Recall that $\mathrm{supp}(\mathbf{x}_0) = \{k_i\}_{i=1}^S$ *in the case* $|\mathrm{supp}(\mathbf{x}_0) \cup \{k\}| = S+1$.*)*

   b) *The LMV estimator at* $\mathbf{x}_0$ *is*

   $$\hat{x}_k^{(c(\cdot),\mathbf{x}_0)}(\mathbf{y}) = \psi(\mathbf{y}, \mathbf{x}_0) \sum_{l \in \mathbb{Z}_+} \frac{m_l \sigma^l}{l!} H_l\!\left(\frac{y_k - x_{0,k}}{\sigma}\right),$$

   with

   $$\psi(\mathbf{y}, \mathbf{x}_0) \triangleq \begin{cases} 1, & \text{if } |\mathrm{supp}(\mathbf{x}_0) \cup \{k\}| \leq S \\ \displaystyle\sum_{i \in [S]} \exp\!\left(-\frac{x_{0,k_i}^2 + 2 y_{k_i} x_{0,k_i}}{2\sigma^2}\right) \times \prod_{j \in [i-1]} \left[1 - \exp\!\left(-\frac{x_{0,k_j}^2 + 2 y_{k_j} x_{0,k_j}}{2\sigma^2}\right)\right], & \text{if } |\mathrm{supp}(\mathbf{x}_0) \cup \{k\}| = S+1. \end{cases} \tag{45}$$

3) *Finally, assume that the prescribed bias function* $c(\cdot)$ *is the actual bias function of a diagonal estimator* $\hat{x}_k(\mathbf{y}) = \hat{x}_k(y_k)$, *i.e.,* $c(\mathbf{x}) = b(\hat{x}_k(\cdot); \mathbf{x})$. *Then, the minimum achievable variance is*

   $$M_{\mathrm{SSNM}}(c(\cdot), \mathbf{x}_0) = v(\hat{x}_k(\cdot); \mathbf{x}_0)\phi(\mathbf{x}_0) + [\phi(\mathbf{x}_0) - 1]\, \gamma^2(\mathbf{x}_0), \tag{46}$$

   and the corresponding LMV estimator is

   $$\hat{x}_k^{(c(\cdot),\mathbf{x}_0)}(\mathbf{y}) = \hat{x}_k(y_k)\, \psi(\mathbf{y}, \mathbf{x}_0). \tag{47}$$

Proof: The proof is based on applying Theorem V.1 to a diagonal prescribed bias function such that the corresponding mean function is of the form (43). In particular, one relates the coefficients m_l in (43) with the coefficients $a[\mathbf{p}]$ in the expansion (37). The relations (46) and (47) are obtained by recognizing that the set $\big\{\frac{1}{\sqrt{l!}} H_l\big(\frac{y_k - x_{0,k}}{\sigma}\big)\big\}_{l \in \mathbb{Z}_+}$ forms an orthonormal basis for the Hilbert space of diagonal estimator functions with finite variance at $\mathbf{x}_0$ and inner product given by $\langle \hat{g}_1(y_k), \hat{g}_2(y_k) \rangle = \int_{\mathbb{R}} \hat{g}_1(y_k)\, \hat{g}_2(y_k) \exp\big(-\frac{1}{2\sigma^2}(y_k - x_{0,k})^2\big)\, dy_k$ [41]. For a detailed proof, we refer to [22, Appendix A]. □

Regarding the case distinction in Theorem V.2, we note that $|\mathrm{supp}(\mathbf{x}_0) \cup \{k\}| \leq S$ either if $\|\mathbf{x}\|_0 \leq S-1$ or if both $\|\mathbf{x}\|_0 = S$ and $k \in \mathrm{supp}(\mathbf{x}_0)$, and $|\mathrm{supp}(\mathbf{x}_0) \cup \{k\}| = S+1$ if both $\|\mathbf{x}\|_0 = S$ and $k \notin \mathrm{supp}(\mathbf{x}_0)$.

Remarkably, as shown by (47), the LMV estimator can be obtained by multiplying the diagonal estimator $\hat{x}_k(\mathbf{y})$—which is arbitrary except for the condition that its variance at $\mathbf{x}_0$ is finite—by the "correction factor" $\psi(\mathbf{y}, \mathbf{x}_0)$ in (45). It can be easily verified that $\psi(\mathbf{y}, \mathbf{x}_0)$ does not depend on y_k. According to (45), the following two cases have to be distinguished:

1) For $k \in [N]$ such that $|\mathrm{supp}(\mathbf{x}_0) \cup \{k\}| \leq S$, we have $\psi(\mathbf{y}, \mathbf{x}_0) = 1$, and therefore the LMV estimator is obtained from (47) as $\hat{x}_k^{(c(\cdot),\mathbf{x}_0)}(\mathbf{y}) = \hat{x}_k(y_k) = \hat{x}_k(\mathbf{y})$. Thus, in that case, it follows that every diagonal estimator $\hat{x}_k(\cdot) : \mathbb{R}^N \to \mathbb{R}$ for the SSNM that has finite variance at $\mathbf{x}_0$ is necessarily an LMV estimator. In particular, the variance $v(\hat{x}_k(\cdot); \mathbf{x}_0)$ equals the minimum achievable variance $M_{\mathrm{SSNM}}(c(\cdot), \mathbf{x}_0)$, i.e., the Barankin bound. Furthermore, the sparsity information cannot be leveraged for improved MVE, because the estimator $\hat{x}_k(\cdot)$ is an LMV estimator for the parameter set $\mathcal{X}_S$ with arbitrary S, including the nonsparse case $\mathcal{X} = \mathbb{R}^N$.
2) For $k \in [N]$ such that $|\mathrm{supp}(\mathbf{x}_0) \cup \{k\}| = S+1$, it follows from (46) and (44) that there exist estimators (in particular, the LMV estimator $\hat{x}_k^{(c(\cdot),\mathbf{x}_0)}(\mathbf{y})$) with the same bias function as $\hat{x}_k(\cdot)$ but with a smaller variance at $\mathbf{x}_0$. Indeed, in this case, we have $\phi(\mathbf{x}_0) < 1$ in (44),

and by (46) it thus follows that $M_{\mathrm{SSNM}}(c(\cdot), \mathbf{x}_0) < v(\hat{x}_k(\cdot); \mathbf{x}_0)$.

Let us for the moment make the (mild) assumption that the given diagonal estimator $\hat{x}_k(\cdot)$ has finite variance at every parameter vector $\mathbf{x} \in \mathbb{R}^N$. It can then be shown that the LMV estimator $\hat{x}_k^{(c(\cdot),\mathbf{x}_0)}(\cdot)$ is robust to deviations from the nominal parameter $\mathbf{x}_0$ in the sense that its bias and variance depend continuously on $\mathbf{x}_0$. Furthermore, $\hat{x}_k^{(c(\cdot),\mathbf{x}_0)}(\cdot)$ has finite bias and finite variance at every parameter vector $\mathbf{x} \in \mathbb{R}^N$, i.e., $\big|b\big(\hat{x}_k^{(c(\cdot),\mathbf{x}_0)}(\cdot); \mathbf{x}\big)\big| < \infty$ and $v\big(\hat{x}_k^{(c(\cdot),\mathbf{x}_0)}(\cdot); \mathbf{x}\big) < \infty$ for all $\mathbf{x} \in \mathbb{R}^N$.

We finally note that (46) and (47) also apply to unbiased estimation, i.e., prescribed bias function $c(\cdot) \equiv 0$ (equivalently, $\gamma(\mathbf{x}) = x_k$). This is because $c(\cdot) \equiv 0$ is the actual bias function of the LS estimator $\hat{x}_{\mathrm{LS},k}(\mathbf{y}) = y_k$. Clearly, the LS estimator is diagonal and has finite variance at $\mathbf{x}_0$. Thus, it can be used as the given diagonal estimator $\hat{x}_k(\mathbf{y})$ in (46) and (47).

C. Lower Variance Bounds

Finally, we complement the exact expressions of the minimum achievable variance $M_{\mathrm{SSNM}}(c(\cdot), \mathbf{x}_0)$ presented above by simple lower bounds. The following bound is obtained by specializing the sparse CRB in (28) to the SSNM ($\mathbf{H} = \mathbf{I}$). Consider the estimation problem $\mathcal{E}_{\mathrm{SSNM}} = \big(\mathcal{X}_S, f_{\mathbf{I}}(\mathbf{y}; \mathbf{x}), g(\mathbf{x}) = x_k\big)$, and let $\mathbf{x}_0 \in \mathcal{X}_S$. If the bias function $c(\cdot) : \mathcal{X}_S \to \mathbb{R}$ is such that the partial derivatives $\frac{\partial c(\mathbf{x})}{\partial x_l}\big|_{\mathbf{x}=\mathbf{x}_0}$ exist for all $l \in [N]$, then

$$M_{\mathrm{SSNM}}(c(\cdot), \mathbf{x}_0) \geq \begin{cases} \sigma^2 \|\mathbf{b}\|_2^2, & \text{if } \|\mathbf{x}_0\|_0 \leq S-1 \\ \sigma^2 \|\mathbf{b}_{\mathbf{x}_0}\|_2^2, & \text{if } \|\mathbf{x}_0\|_0 = S. \end{cases} \tag{48}$$

Here, in the case $\|\mathbf{x}_0\|_0 \leq S-1$, $\mathbf{b} \in \mathbb{R}^N$ is given by $b_l \triangleq \delta_{k,l} + \frac{\partial c(\mathbf{x})}{\partial x_l}\big|_{\mathbf{x}=\mathbf{x}_0}$, $l \in [N]$, and in the case $\|\mathbf{x}_0\|_0 = S$, $\mathbf{b}_{\mathbf{x}_0} \in \mathbb{R}^S$ consists of those entries of $\mathbf{b}$ that are indexed by $\mathrm{supp}(\mathbf{x}_0) = \{k_1, \ldots, k_S\}$, i.e., $(\mathbf{b}_{\mathbf{x}_0})_i = b_{k_i}$, $i \in [S]$.

Next, we specialize the bound in Theorem IV.1 to the SSNM. To this end, note that because $\mathbf{H} = \mathbf{I}$, we have $\mathbf{P} = \mathbf{H}_{\mathcal{K}}(\mathbf{H}_{\mathcal{K}})^{\dagger} = \mathbf{I}_{\mathcal{K}}(\mathbf{I}_{\mathcal{K}})^{\dagger} = \sum_{l \in \mathcal{K}} \mathbf{e}_l \mathbf{e}_l^T$. Therefore, multiplying $\mathbf{x}_0$ by $\mathbf{I}$–$\mathbf{P}$ simply zeros all entries of $\mathbf{x}_0$ whose indices belong to $\mathcal{K}$, i.e., $(\mathbf{I}-\mathbf{P})\mathbf{x}_0 = \mathbf{x}_0^{\mathrm{supp}(\mathbf{x}_0) \setminus \mathcal{K}}$.

Corollary V.3: Consider the estimation problem $\mathcal{E}_{\mathrm{SSNM}} = \big(\mathcal{X}_S, f_{\mathbf{I}}(\mathbf{y}; \mathbf{x}), g(\mathbf{x}) = x_k\big)$. *Let* $\mathbf{x}_0 \in \mathcal{X}_S$, *and consider an arbitrary index set* $\mathcal{K} = \{l_1, \ldots, l_{|\mathcal{K}|}\} \subseteq [N]$ *consisting of no more than S indices, i.e.,* $|\mathcal{K}| \leq S$. *If the bias function* $c(\cdot) : \mathcal{X}_S \to \mathbb{R}$ *is such that the partial derivatives* $\frac{\partial c(\mathbf{x})}{\partial x_{l_i}}\big|_{\mathbf{x}=\mathbf{x}_0}$ *exist for all* $i \in [|\mathcal{K}|]$, *then*

$$M_{\mathrm{SSNM}}(c(\cdot), \mathbf{x}_0) \geq \exp\bigg(-\frac{1}{\sigma^2} \big\|\mathbf{x}_0^{\mathrm{supp}(\mathbf{x}_0)\setminus\mathcal{K}}\big\|_2^2\bigg) \sigma^2 \|\mathbf{b}_{\mathbf{x}_0}\|_2^2. \tag{49}$$

Here, $\mathbf{b}_{\mathbf{x}_0} \in \mathbb{R}^{|\mathcal{K}|}$ *is defined entrywise as* $(\mathbf{b}_{\mathbf{x}_0})_i \triangleq \delta_{k,l_i} + \frac{\partial c(\mathbf{x})}{\partial x_{l_i}}\big|_{\mathbf{x}=\mathbf{x}_0^{\mathcal{K}}}$ *for* $i \in [|\mathcal{K}|]$.

For unbiased estimation ($c(\cdot) \equiv 0$), the following lower bound on $M_{\mathrm{SSNM}}(c(\cdot) \equiv 0, \mathbf{x}_0)$ is based on the Hammersley-Chapman-Robbins bound (HCRB) [19], [29], [42]. This bound has been previously derived in a slightly different form in [14]. Consider the estimation problem $\mathcal{E}_{\mathrm{SSNM}} = \big(\mathcal{X}_S, f_{\mathbf{I}}(\mathbf{y}; \mathbf{x}), g(\mathbf{x}) = x_k\big)$ with $k \in [N]$ and the bias function $c(\cdot) \equiv 0$. Let $\mathbf{x}_0 \in \mathcal{X}_S$. Then,

$$M_{\mathrm{SSNM}}(c(\cdot), \mathbf{x}_0) \geq \begin{cases} \sigma^2, & \text{if } |\mathrm{supp}(\mathbf{x}_0) \cup \{k\}| \leq S \\ \sigma^2 \dfrac{N-S-1}{N-S} \exp(-\xi_0^2/\sigma^2), & \\ & \text{if } |\mathrm{supp}(\mathbf{x}_0) \cup \{k\}| = S+1, \end{cases} \tag{50}$$

where ξ_0 denotes the S-largest (in magnitude) entry of $\mathbf{x}_0$.

In [22, Th. 5.4.2], it is shown that the bound (50) for $|\mathrm{supp}(\mathbf{x}_0) \cup \{k\}| \leq S$ is obtained as a limit of the generic bound (27) for the sequence of subspaces $\mathcal{U} = \mathcal{U}^{(t)} \triangleq \mathrm{span}\big\{u_0(\cdot), \{u_l^{(t)}(\cdot)\}_{l \in [N]}\big\}$ as $t \to 0$. Here, $u_0(\cdot) \triangleq R_{\mathrm{SSNM},\mathbf{x}_0}(\cdot, \mathbf{x}_0)$ and

$$u_l^{(t)}(\cdot) \triangleq \begin{cases} R_{\mathrm{SSNM},\mathbf{x}_0}(\cdot, \mathbf{x}_0 + t\mathbf{e}_l) - R_{\mathrm{SSNM},\mathbf{x}_0}(\cdot, \mathbf{x}_0), \\ \qquad\qquad \text{if } l \in \mathrm{supp}(\mathbf{x}_0) \\ R_{\mathrm{SSNM},\mathbf{x}_0}(\cdot, \mathbf{x}_0 - \xi_0 \mathbf{e}_{j_0} + t\mathbf{e}_l) - R_{\mathrm{SSNM},\mathbf{x}_0}(\cdot, \mathbf{x}_0), \\ \qquad\qquad \text{if } l \in [N] \setminus \mathrm{supp}(\mathbf{x}_0), \end{cases}$$

for $l \in [N]$, where j_0 denotes the index of the S-largest (in magnitude) entry of $\mathbf{x}_0$. Similarly, the bound (50) for $|\mathrm{supp}(\mathbf{x}_0) \cup \{k\}| = S+1$ is obtained as a limit of (27) for $\mathcal{U} = \widetilde{\mathcal{U}}^{(t)} \triangleq \mathrm{span}\big\{u_0(\cdot), u^{(t)}(\cdot)\big\}$ as $t \to 0$, where $u^{(t)}(\cdot) \triangleq R_{\mathrm{SSNM},\mathbf{x}_0}(\cdot, \mathbf{x}_0 + t\mathbf{e}_k) - R_{\mathrm{SSNM},\mathbf{x}_0}(\cdot, \mathbf{x}_0)$. (An expression of $R_{\mathrm{SSNM},\mathbf{x}_0}(\cdot, \cdot)$ was given in (36).) In [14], an equivalent bound on the MSE (equivalently, on the variance, because $c(\cdot) \equiv 0$) was formulated for a vector-valued estimator $\hat{\mathbf{x}}(\cdot)$; that bound can be obtained by summing (50) over all $k \in [N]$.

It can be shown that the HCRB-type bound (50) is tighter (higher) than the CRB (48) specialized to $c(\cdot) \equiv 0$. For $|\mathrm{supp}(\mathbf{x}_0) \cup \{k\}| = S+1$ (which is true if both $\|\mathbf{x}\|_0 = S$ and $k \notin \mathrm{supp}(\mathbf{x}_0)$), the HCRB-type bound (50) is a strictly upper semi-continuous function of $\mathbf{x}_0$, just as the CRB (48). Hence, it again follows from Corollary III.4 that the bound cannot be tight, i.e., in general, we have a strict inequality in (50). However, for $|\mathrm{supp}(\mathbf{x}_0) \cup \{k\}| \leq S$ (which is true either if $\|\mathbf{x}\|_0 \leq S-1$ or if both $\|\mathbf{x}\|_0 = S$ and $k \in \mathrm{supp}(\mathbf{x}_0)$), the bound (50) is tight since it is achieved by the LS estimator $\hat{x}_{\mathrm{LS},k}(\mathbf{y}) = y_k$.

According to (50), the MSE of any unbiased estimator $\hat{\mathbf{x}}(\mathbf{y})$ for the parameter vector $\mathbf{x}$ is larger than or equal to $N\sigma^2$ at all $\mathbf{x}$ with $\|\mathbf{x}\|_0 \leq S$–1. However, as shown in [43] and [44], there exist biased estimators for the SSNM whose MSE is on the order of $S\sigma^2 \log(N)$, which can be significantly smaller than the corresponding bound for unbiased estimators. This implies that, in general, unbiased estimators are not well suited for the estimation of sparse vectors. One possible use of the lower bounds presented in this paper is as a tool for finding bias functions that allow for a small variance and, in turn, a small MSE for all sparse parameter vectors $\mathbf{x} \in \mathcal{X}_S$.

VI. Numerical Results

In this section, we compare the lower variance bounds presented in Sections IV and V with the actual variance behavior of some well-known estimators. We consider the SLGM estimation problem $\mathcal{E}_{\mathrm{SLGM}} = \big(\mathcal{X}_S, f_{\mathbf{H}}(\mathbf{y}; \mathbf{x}), g(\mathbf{x}) = x_k\big)$ for $k \in [N]$. In what follows, we will denote the lower bounds (28) and (31) by $B_k^{(1)}(c(\cdot),\mathbf{x}_0)$ and $B_k^{(2)}(c(\cdot),\mathbf{x}_0)$, respectively. We recall that the latter bound depends on an index set $\mathcal{K}\subseteq[N]$ with $|\mathcal{K}|\leq S$, which can be chosen freely.

Let $\hat{\mathbf{x}}(\cdot)$ be an estimator of $\mathbf{x}$ with bias function $\mathbf{c}(\cdot)$. Because of (8), a lower bound on the estimator variance $v(\hat{\mathbf{x}}(\cdot);\mathbf{x}_0)$ can be obtained by summing with respect to $k\in[N]$ the "scalar bounds" $B_k^{(1)}(c_k(\cdot),\mathbf{x}_0)$ or $B_k^{(2)}(c_k(\cdot),\mathbf{x}_0)$, where $c_k(\cdot)\triangleq\big(\mathbf{c}(\cdot)\big)_k$, i.e.,

$$v(\hat{\mathbf{x}}(\cdot);\mathbf{x}_0)\;\geq\;B^{(1/2)}(\mathbf{c}(\cdot),\mathbf{x}_0)\;\triangleq\;\sum_{k\in[N]}B_k^{(1/2)}(c_k(\cdot),\mathbf{x}_0).\quad(51)$$

Here, the index sets $\mathcal{K}_k$ used in $B_k^{(2)}(c_k(\cdot),\mathbf{x}_0)$ can be chosen differently for different k.

A. An SLGM View of Fourier Analysis

Our first example is inspired by [18, Example 4.2]. We consider the SLGM with N even, i.e., $N=2L$, and $\sigma^2=1$. The system matrix $\mathbf{H}\in\mathbb{R}^{M\times 2L}$ is given by $H_{m,l}=\cos\big(\theta_l(m-1)\big)$ for $m\in[M]$ and $l\in[L]$ and $H_{m,l}=\sin\big(\theta_l(m-1)\big)$ for $m\in[M]$ and $l\in\{L+1,\ldots,2L\}$. Here, the normalized angular frequencies θ_l are uniformly spaced according to $\theta_l=\theta_0+\big[(l-1)\bmod L\big]\Delta\theta$, $l\in[N]$. The multiplication of $\mathbf{x}$ by $\mathbf{H}$ then corresponds to an inverse discrete Fourier transform that maps $2L$ spectral samples (the entries of $\mathbf{x}$) to M temporal samples (the entries of $\mathbf{Hx}$). In our simulation, we chose $M=128$, $L=8$ (hence, $N=16$), $S=4$, $\theta_0=0.2$, and $\Delta\theta=3.9\cdot10^{-3}$. The frequency spacing $\Delta\theta$ is about half the nominal DFT frequency resolution, which is $1/128\approx7.8\cdot10^{-3}$.

We consider the OMP estimator $\hat{\mathbf{x}}_{\text{OMP}}(\cdot)$ that is obtained by applying the OMP [23], [37] with $S=4$ iterations to the observation $\mathbf{y}$. We used Monte Carlo simulation with randomly generated noise $\mathbf{n}\sim\mathcal{N}(\mathbf{0},\mathbf{I})$ to estimate the variance $v(\hat{\mathbf{x}}_{\text{OMP}}(\cdot);\mathbf{x}_0)$ of $\hat{\mathbf{x}}_{\text{OMP}}(\cdot)$. The parameter vector was chosen as $\mathbf{x}_0=\sqrt{\text{SNR}}\,\tilde{\mathbf{x}}_0$, where $\tilde{\mathbf{x}}_0\in\{0,1\}^{16}$, $\text{supp}(\tilde{\mathbf{x}}_0)=\{3,6,11,14\}$, and SNR varies between 10^{-2} and 10^4. Thus, the observation $\mathbf{y}=\mathbf{Hx}_0+\mathbf{n}$ is a noisy superposition of four sinusoidal components with identical amplitudes; two of them are cosine and sine components with frequency $\theta_3=\theta_{11}=\theta_0+2\Delta\theta$, and two are cosine and sine components with frequency $\theta_6=\theta_{14}=\theta_0+5\Delta\theta$. In Fig. 1, we plot $v(\hat{\mathbf{x}}_{\text{OMP}}(\cdot);\mathbf{x}_0)$ versus SNR. For comparison, we also plot the lower bounds $B^{(1)}(\mathbf{c}_{\text{OMP}}(\cdot),\mathbf{x}_0)$ and $B^{(2)}(\mathbf{c}_{\text{OMP}}(\cdot),\mathbf{x}_0)$ in (51), with $\mathbf{c}_{\text{OMP}}(\mathbf{x})\triangleq\mathbf{b}(\hat{\mathbf{x}}_{\text{OMP}}(\cdot);\mathbf{x})$ being the actual bias function of the OMP estimator $\hat{\mathbf{x}}_{\text{OMP}}(\cdot)$. To evaluate these bounds, we computed the first-order partial derivatives of the bias functions $c_{\text{OMP},k}(\mathbf{x})$ (see (28) and Theorem IV.1) by means of (35) using the Monte Carlo method outlined in [28, Sec. 3]. The index sets $\mathcal{K}_k$ in the bound $B^{(2)}(\mathbf{c}_{\text{OMP}}(\cdot),\mathbf{x}_0)$ were chosen as $\mathcal{K}_k=\text{supp}(\mathbf{x}_0)$ for $k\in\text{supp}(\mathbf{x}_0)$ and $\mathcal{K}_k=\{k\}$ for $k\notin\text{supp}(\mathbf{x}_0)$. This is the simplest nontrivial choice of the $\mathcal{K}_k$ for which $B^{(2)}(\mathbf{c}_{\text{OMP}}(\cdot),\mathbf{x}_0)$ is tighter than the state-of-the-art bound $B^{(1)}(\mathbf{c}_{\text{OMP}}(\cdot),\mathbf{x}_0)$ (the sparse CRB, which was originally presented in [12]). Finally, Fig. 1 also shows the "oracle CRB," which is defined as the CRB for known $\text{supp}(\mathbf{x}_0)$. This is simply the CRB for a linear Gaussian model

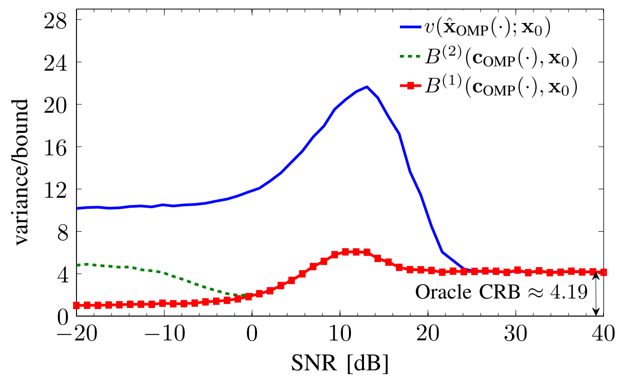

Fig. 1. Variance of the OMP estimator and corresponding lower bounds versus SNR, for the SLGM with $N=16$, $M=128$, $S=4$, and $\sigma^2=1$.

with system matrix $\mathbf{H}_{\text{supp}(\mathbf{x}_0)}$ and is thus given by $\text{tr}\big((\mathbf{H}^T_{\text{supp}(\mathbf{x}_0)}\mathbf{H}_{\text{supp}(\mathbf{x}_0)})^{-1}\big)\approx4.19$ [18] for all values of SNR (recall that we set $\sigma^2=1$).

As can be seen in Fig. 1, for SNR below 20 dB, $v(\hat{\mathbf{x}}_{\text{OMP}}(\cdot);\mathbf{x}_0)$ is significantly higher than the three lower bounds. This suggests that there might exist estimators with the same bias as that of the OMP estimator but a smaller variance; however, a positive statement regarding the existence of such estimators cannot be based on our analysis. In particular, our analysis does not rule out the possibility that the OMP estimator already achieves the minimum achievable variance. For SNR larger than about 15 dB, the three lower bounds coincide. Furthermore, for SNR larger than about 12 dB, $v(\hat{\mathbf{x}}_{\text{OMP}}(\cdot);\mathbf{x}_0)$ quickly converges toward the lower bounds. This is because for high SNR, the OMP estimator is able to correctly detect $\text{supp}(\mathbf{x}_0)$ with very high probability.

B. Minimum Variance Analysis for the SSNM

Next, we consider the maximum likelihood (ML) estimator, an estimator based on the approximate message passing (AMP) algorithm, and the hard-thresholding (HT) estimator for the SSNM, i.e., for $M=N$ and $\mathbf{H}=\mathbf{I}$, with $N=50$, $S=5$, and $\sigma^2=1$. The ML estimator is given by

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

where the operator P_S retains the S largest (in magnitude) entries and zeros all other entries. Closed-form expressions of the mean and variance of the ML estimator were derived in [14]. For the AMP estimator $\hat{\mathbf{x}}_{\text{AMP}}(\cdot)$, we used the implementation [45, Algorithm 1] with the choice $\lambda=2.5$. The HT estimator $\hat{\mathbf{x}}_{\text{HT}}(\cdot)$ is given by

$$\hat{x}_{\text{HT},k}(\mathbf{y})\;=\;\hat{x}_{\text{HT},k}(y_k)\;=\;\begin{cases}y_k, & |y_k|\geq T\\ 0, & \text{else},\end{cases}\qquad k\in[N],\quad(52)$$

where T is a fixed threshold. Note that in the limiting case $T=0$, the HT estimator coincides with the LS estimator $\hat{\mathbf{x}}_{\text{LS}}(\mathbf{y})=\mathbf{y}$. The mean and variance of the HT estimator are given by

$$\mathsf{E}_{\mathbf{x}}\big\{\hat{x}_{\text{HT},k}(\mathbf{y})\big\}=\frac{1}{\sqrt{2\pi\sigma^2}}\int_{\mathbb{R}\setminus[-T,T]}y\,\exp\!\Big(-\frac{1}{2\sigma^2}(y-x_k)^2\Big)dy\quad(53)$$

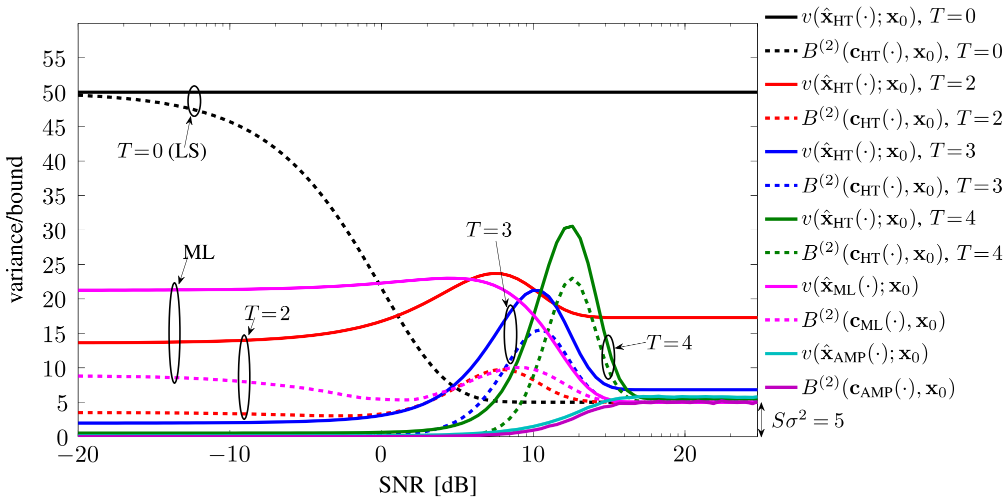

Fig. 2. Variance of the ML, AMP, and HT estimators and corresponding lower bounds versus SNR, for the SSNM with $N = 50$, $S = 5$, and $\sigma^2 = 1$.

$$v(\hat{x}_{\mathrm{HT},k}(\cdot);\mathbf{x}) = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{\mathbb{R}\setminus[-T,T]} y^2 \exp\left(-\frac{1}{2\sigma^2}(y-x_k)^2\right)dy - \left(\mathsf{E}_{\mathbf{x}}\{\hat{x}_{\mathrm{HT},k}(\mathbf{y})\}\right)^2. \quad (54)$$

We calculated the variances $v(\hat{\mathbf{x}}_{\mathrm{ML}}(\cdot);\mathbf{x}_0)$, $v(\hat{\mathbf{x}}_{\mathrm{AMP}}(\cdot);\mathbf{x}_0)$, and $v(\hat{\mathbf{x}}_{\mathrm{HT}}(\cdot);\mathbf{x}_0)$ at parameter vectors $\mathbf{x}_0 = \sqrt{\mathrm{SNR}}\,\tilde{\mathbf{x}}_0$, where $\tilde{\mathbf{x}}_0 \in \{0,1\}^{50}$, $\mathrm{supp}(\tilde{\mathbf{x}}_0) = [S] = [5]$, and SNR varies between 10^{-2} and 10^2. (The fixed choice $\mathrm{supp}(\mathbf{x}_0) = [S]$ is justified by the fact that neither the variances of the ML, AMP, and HT estimators nor the corresponding variance bounds depend on the location of $\mathrm{supp}(\mathbf{x}_0)$.) In particular, $v(\hat{\mathbf{x}}_{\mathrm{HT}}(\cdot);\mathbf{x}_0)$ was calculated by numerical evaluation of the integrals (53) and (54). The variance $v(\hat{\mathbf{x}}_{\mathrm{AMP}}(\cdot);\mathbf{x}_0)$ of the AMP estimator was estimated through Monte Carlo simulation. Fig. 2 shows $v(\hat{\mathbf{x}}_{\mathrm{ML}}(\cdot);\mathbf{x}_0)$, $v(\hat{\mathbf{x}}_{\mathrm{AMP}}(\cdot);\mathbf{x}_0)$, and $v(\hat{\mathbf{x}}_{\mathrm{HT}}(\cdot);\mathbf{x}_0)$—the last for four different choices of T in (52)—versus SNR. Also shown are the lower bounds $B^{(2)}(\mathbf{c}_{\mathrm{ML}}(\cdot),\mathbf{x}_0)$, $B^{(2)}(\mathbf{c}_{\mathrm{AMP}}(\cdot),\mathbf{x}_0)$, and $B^{(2)}(\mathbf{c}_{\mathrm{HT}}(\cdot),\mathbf{x}_0)$ (cf. (51)), with $\mathbf{c}_{\mathrm{ML}}(\cdot)$, $\mathbf{c}_{\mathrm{AMP}}(\cdot)$, and $\mathbf{c}_{\mathrm{HT}}(\cdot)$ being the actual bias functions of $\hat{\mathbf{x}}_{\mathrm{ML}}(\cdot)$, $\hat{\mathbf{x}}_{\mathrm{AMP}}(\cdot)$, and $\hat{\mathbf{x}}_{\mathrm{HT}}(\cdot)$, respectively. The index sets underlying the bounds were chosen as $\mathcal{K}_k = \mathrm{supp}(\mathbf{x}_0)$ for $k \in \mathrm{supp}(\mathbf{x}_0)$ and $\mathcal{K}_k = \{k\} \cup \{\mathrm{supp}(\mathbf{x}_0) \setminus \{j_S\}\}$ for $k \notin \mathrm{supp}(\mathbf{x}_0)$, where j_S denotes the index of the S-largest (in magnitude) entry of $\mathbf{x}_0$. The first-order partial derivatives of the bias functions $c_{\mathrm{ML},k}(\mathbf{x})$ involved in the bound $B^{(2)}(\mathbf{c}_{\mathrm{ML}}(\cdot),\mathbf{x}_0)$ were approximated by a finite-difference quotient [28], i.e., $\frac{\partial c_{\mathrm{ML},k}(\mathbf{x})}{\partial x_l} = \delta_{k,l} + \frac{\partial \mathsf{E}_{\mathbf{x}}\{\hat{x}_{\mathrm{ML},k}(\mathbf{y})\}}{\partial x_l}$ with

$$\frac{\partial \mathsf{E}_{\mathbf{x}}\{\hat{x}_{\mathrm{ML},k}(\mathbf{y})\}}{\partial x_l} \approx \frac{\mathsf{E}_{\mathbf{x}+\Delta\mathbf{e}_l}\{\hat{x}_{\mathrm{ML},k}(\mathbf{y})\} - \mathsf{E}_{\mathbf{x}}\{\hat{x}_{\mathrm{ML},k}(\mathbf{y})\}}{\Delta},$$

where $\Delta > 0$ is a small stepsize and the expectations were calculated using the closed-form expressions presented in [14, Appendix I]. The first-order partial derivatives of the bias functions $c_{\mathrm{AMP},k}(\mathbf{x})$ involved in the bound $B^{(2)}(\mathbf{c}_{\mathrm{AMP}}(\cdot),\mathbf{x}_0)$ were calculated by means of (35) using the Monte Carlo method outlined in [28, Sec. 3]. The first-order partial derivatives of the bias functions $c_{\mathrm{HT},k}(\mathbf{x})$ involved in the bound $B^{(2)}(\mathbf{c}_{\mathrm{HT}}(\cdot),\mathbf{x}_0)$ were calculated by means of (35) using numerical integration.

It can be seen in Fig. 2 that for SNR larger than about 15 dB, the variances of the ML, AMP, and HT estimators and the corresponding bounds are effectively equal (for the HT estimator, this is true if T is not too small). Also, all bounds are close to $S\sigma^2 = 5$; this equals the variance of an oracle estimator that knows $\mathrm{supp}(\mathbf{x}_0)$ and is given by $\hat{x}_k(\mathbf{y}) = y_k$ for $k \in \mathrm{supp}(\mathbf{x}_0)$ and $\hat{x}_k(\mathbf{y}) = 0$ otherwise. However, in the medium-SNR range, the variances of the ML and HT estimators are significantly higher than the corresponding lower bounds. It is also seen that the AMP estimator has a smaller variance than the ML and HT estimators on the entire displayed SNR range. Moreover, comparing with the corresponding lower bound, we conclude that the AMP estimator nearly yields the minimum variance among all estimators whose bias equals that of the AMP estimator.

The fact that, in the medium-SNR range, the variances of the ML and HT estimators are higher than the corresponding lower bounds suggests that there might exist estimators with the same bias as that of the ML or HT estimator but a smaller variance. However, in general, a positive statement regarding the existence of such estimators cannot be based on our analysis. On the other hand, for the special case of diagonal estimators such as the HT estimator, Theorem V.2 makes a positive statement about the existence of estimators that have *locally* a smaller variance than the HT estimator. In particular, we can use (47) and (46) to obtain the LMV estimator and corresponding minimum achievable variance at a parameter vector $\mathbf{x}_0 \in \mathcal{X}_S$ for the given bias function of the HT estimator, $\mathbf{c}_{\mathrm{HT}}(\cdot)$. In Fig. 3, we plot the variance $v(\hat{\mathbf{x}}_{\mathrm{HT}}(\cdot);\mathbf{x}_0)$ for four different choices of T versus SNR. We also plot the corresponding minimum achievable variance (Barankin bound) $M_{\mathrm{HT}}(\mathbf{x}_0) \triangleq \sum_{k\in[N]} M_{\mathrm{SSNM}}(c_{\mathrm{HT},k}(\cdot),\mathbf{x}_0)$, where $M_{\mathrm{SSNM}}(c_{\mathrm{HT},k}(\cdot),\mathbf{x}_0)$ is obtained from (46). (Note that (46) is applicable because the estimator $\hat{x}_{\mathrm{HT},k}(\mathbf{y})$ is diagonal and has finite variance at all $\mathbf{x}_0 \in \mathcal{X}_S$.) It is seen that for small T (including $T=0$, where the HT estimator reduces to the LS estimator) and for SNR above 0 dB, $v(\hat{\mathbf{x}}_{\mathrm{HT}}(\cdot);\mathbf{x}_0)$ is significantly higher than $M_{\mathrm{HT}}(\mathbf{x}_0)$. As T increases, the gap between the $v(\hat{\mathbf{x}}_{\mathrm{HT}}(\cdot);\mathbf{x}_0)$ and $M_{\mathrm{HT}}(\mathbf{x}_0)$ curves becomes smaller; in particular, the two curves are almost

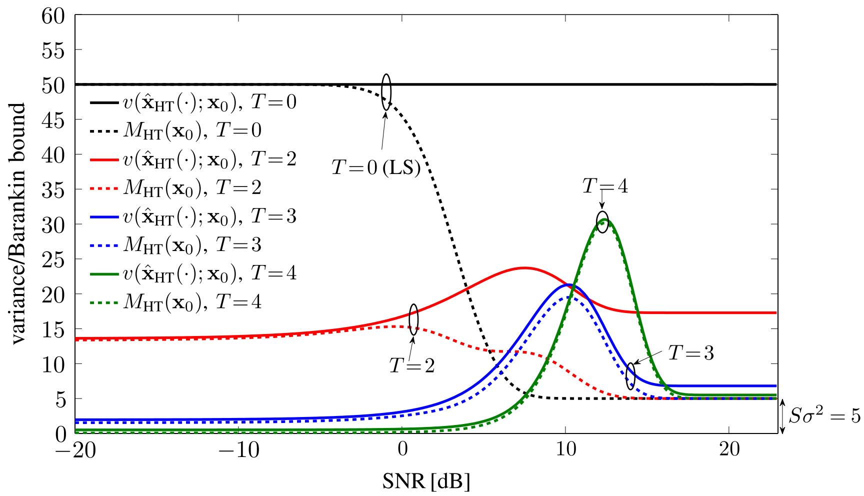

Fig. 3. Variance of the HT estimator for different T and corresponding minimum achievable variance (Barankin bound) versus SNR, for the SSNM with $N=50$, $S=5$, and $\sigma^2=1$.

indistinguishable already for $T=4$. For high SNR, $M_{\rm HT}(\mathbf{x}_0)$ approaches the oracle variance $S\sigma^2 = 5$ for any value of T. However, note that, in contrast to the HT estimator, the LMV estimator achieving $M_{\rm HT}(\mathbf{x}_0)$ cannot be implemented in practice since it depends on the unknown parameter vector $\mathbf{x}_0$.

VII. Conclusion

We used RKHS theory to analyze the minimum variance estimation (MVE) problem within the sparse linear Gaussian model (SLGM). In this model, the unknown parameter vector to be estimated is assumed to be sparse with a known sparsity degree, and the observed vector is a linearly transformed version of the parameter vector that is corrupted by i.i.d. Gaussian noise with known variance. The RKHS framework allowed us to establish a geometric interpretation of existing lower bounds on the estimator variance and to derive new lower bounds on the estimator variance, in both cases under a bias constraint. These bounds were obtained by an orthogonal projection of the prescribed mean function onto a subspace of the RKHS associated with the SLGM. Viewed as functions of the SNR, the bounds were observed to vary between two extreme regimes. In the low-SNR regime, the nonzero entries of the true parameter vector are small compared with the noise variance. Here, our bounds predict that if the estimator bias is approximately zero, then the *a priori* sparsity information does not help much in the estimation; however, if the bias is allowed to be nonzero, the estimator variance can be reduced by the sparsity information. On the other hand, in the high-SNR regime, where the nonzero entries of the true parameter vector are large compared with the noise variance, our bounds coincide with the Cramér–Rao bound of an associated conventional linear Gaussian model in which the support of the unknown parameter vector is known. Our bounds exhibit a steep transition between these two regimes. In general, this transition has an exponential decay with respect to the SNR.

For the special case of the SLGM that corresponds to the recovery problem in a linear compressed sensing scheme, we expressed our lower bounds in terms of the restricted isometry constant of the measurement matrix. Furthermore, for the special case of the SLGM given by the sparse signal in noise model (SSNM), we derived closed-form expressions of the minimum achievable variance (Barankin bound) and the corresponding LMV estimator. These latter results include closed-form expressions of the (unbiased) Barankin bound and the (unbiased) LMV estimator for the SSNM. Simplified expressions of the Barankin bound and the LMV estimator were presented for the subclass of "diagonal" bias functions.

The comparison of our bounds with the actual variance of established estimators for the SLGM and SSNM (maximum likelihood estimator, hard thresholding estimator, orthogonal matching pursuit, and AMP estimator) showed that there might exist estimators with the same bias but a smaller variance.

An interesting direction for future investigations is the search for (classes of) estimators that have a prescribed bias function and asymptotically approach our lower variance bounds when the estimation is based on an increasing number of i.i.d. observation vectors $\mathbf{y}_i$. While [46] presents an estimator for the SLGM whose MSE attains the sparse CRB (28) asymptotically, the estimator bias is not constrained in our sense. For the popular class of M-estimators or penalized maximum likelihood estimators, a characterization of the asymptotic behavior is available [30], [47], [48]. Under mild conditions, M-estimators allow an efficient implementation via convex optimization techniques. Furthermore, it would be interesting to characterize, using our variance bounds, the prescribed bias functions that allow for estimators having a nearly optimum minimax risk.

Finally, it may be worthwhile to generalize our results to the case of block or group sparsity [49]–[51]. This could be useful, e.g., for sparse channel estimation in the case of clustered scatterers and delay-Doppler leakage [52] and for the estimation of structured sparse spectra (extending sparsity-exploiting spectral estimation as proposed in [53]–[57]).

Appendix A
Proof of Theorem III.1

The relations (19) and (21) follow directly from [15, p. 351] since the kernel of the RKHS $\mathcal{H}_{\mathrm{SLGM},\mathbf{x}_0}$ is the restriction of the kernel of the RKHS $\mathcal{H}_{\mathrm{LGM},\mathbf{x}_0}$ to the subdomain $\mathcal{X}_S \times \mathcal{X}_S \subseteq \mathbb{R}^N \times \mathbb{R}^N$ (cf. (18) and (17)). To show (20) and (22), we will characterize the RKHS $\mathcal{H}_{\mathrm{LGM},\mathbf{x}_0}$ by using a congruence with another RKHS.

Lemma A.1: The RKHS $\mathcal{H}_{\mathrm{LGM},\mathbf{x}_0}$ is isometric to the RKHS $\mathcal{H}(R_{\mathrm{G}})$ whose kernel $R_{\mathrm{G}}(\cdot,\cdot)\colon \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}$ is given by

$$R_{\mathrm{G}}(\mathbf{z}_1, \mathbf{z}_2) = \exp\big(\mathbf{z}_1^T \mathbf{z}_2\big), \quad \mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^D. \tag{55}$$

A congruence from $\mathcal{H}(R_{\mathrm{G}})$ to $\mathcal{H}_{\mathrm{LGM},\mathbf{x}_0}$ is constituted by the mapping $\mathsf{K}_{\mathrm{G}}[\cdot]\colon \mathcal{H}(R_{\mathrm{G}}) \to \mathcal{H}_{\mathrm{LGM},\mathbf{x}_0}$ given by

$$\begin{aligned}\mathsf{K}_{\mathrm{G}}[f(\cdot)] &= \widetilde{f}(\mathbf{x}) \\ &\triangleq f\left(\frac{1}{\sigma}\widetilde{\mathbf{H}}^{\dagger}\mathbf{x}\right)\exp\left(\frac{1}{2\sigma^2}\|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma^2}\mathbf{x}^T\mathbf{H}^T\mathbf{H}\mathbf{x}_0\right), \\ &\qquad \mathbf{x} \in \mathbb{R}^N,\end{aligned} \tag{56}$$

for all $f(\cdot) \in \mathcal{H}(R_{\mathrm{G}})$, and a congruence from $\mathcal{H}_{\mathrm{LGM},\mathbf{x}_0}$ to $\mathcal{H}(R_{\mathrm{G}})$ is constituted by the inverse mapping $\mathsf{K}_{\mathrm{G}}^{-1}[\cdot]$: $\mathcal{H}_{\mathrm{LGM},\mathbf{x}_0} \to \mathcal{H}(R_{\mathrm{G}})$ given by

$$\begin{aligned}\mathsf{K}_{\mathrm{G}}^{-1}[\widetilde{f}(\cdot)] &= f(\mathbf{z}) \\ &\triangleq \widetilde{f}\big(\sigma\widetilde{\mathbf{H}}\mathbf{z}\big)\exp\left(-\frac{1}{2\sigma^2}\|\mathbf{H}\mathbf{x}_0\|_2^2 + \frac{1}{\sigma}\mathbf{z}^T\widetilde{\mathbf{H}}^{\dagger}\mathbf{x}_0\right), \\ &\qquad \mathbf{z} \in \mathbb{R}^D,\end{aligned}$$

for all $\widetilde{f}(\cdot) \in \mathcal{H}_{\mathrm{LGM},\mathbf{x}_0}$.

Proof: Consider the two sets

$$\mathcal{A} \triangleq \left\{\widetilde{f}_{\mathbf{x}}(\cdot) \triangleq \exp\left(\frac{1}{2\sigma^2}\|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma}\mathbf{x}^T\mathbf{H}^T\mathbf{H}\mathbf{x}_0\right) \times R_{\mathrm{G}}\big(\cdot, \widetilde{\mathbf{H}}^{\dagger}\mathbf{x}\big)\right\}_{\mathbf{x}\in\mathbb{R}^N}$$

and

$$\mathcal{B} \triangleq \big\{f_{\mathbf{x}}(\cdot) \triangleq R_{\mathrm{LGM},\mathbf{x}_0}(\cdot, \sigma\mathbf{x})\big\}_{\mathbf{x}\in\mathbb{R}^N}.$$

As shown in [22, p. 86], the sets $\mathcal{A}$ and $\mathcal{B}$ span the RKHS $\mathcal{H}_{\mathrm{LGM},\mathbf{x}_0}$ and $\mathcal{H}(R_{\mathrm{G}})$, respectively. Furthermore, K_{G} is a continuous linear mapping that maps $f_{\mathbf{x}}(\cdot) \in \mathcal{H}(R_{\mathrm{G}})$ to $\widetilde{f}_{\mathbf{x}}(\cdot) \in \mathcal{H}_{\mathrm{LGM},\mathbf{x}_0}$ for any $\mathbf{x} \in \mathbb{R}^N$, and there is $\langle \widetilde{f}_{\mathbf{x}_1}(\cdot), \widetilde{f}_{\mathbf{x}_2}(\cdot)\rangle_{\mathcal{H}_{\mathrm{LGM},\mathbf{x}_0}} = \langle f_{\mathbf{x}_1}(\cdot), f_{\mathbf{x}_2}(\cdot)\rangle_{\mathcal{H}(R_{\mathrm{G}})}$ for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^N$. Hence, according to [16, p. 263], K_{G} is a congruence from $\mathcal{H}(R_{\mathrm{G}})$ to $\mathcal{H}_{\mathrm{LGM},\mathbf{x}_0}$. For a detailed proof we refer to [22, Th. 5.2.2]. □

The congruence K_{G} reduces the characterization of the RKHS $\mathcal{H}_{\mathrm{LGM},\mathbf{x}_0}$ to that of the RKHS $\mathcal{H}(R_{\mathrm{G}})$. A simple characterization of $\mathcal{H}(R_{\mathrm{G}})$ in terms of an orthonormal basis can be obtained by noting that the kernel $R_{\mathrm{G}}(\cdot,\cdot)$ is infinitely often differentiable and applying the results for RKHSs with differentiable kernels presented in [58]. In particular, one can show [58], [22, Th. 5.2.4] that for any $\mathbf{p} \in \mathbb{Z}_+^D$, the RKHS $\mathcal{H}(R_{\mathrm{G}})$ contains the function $r^{(\mathbf{p})}(\cdot)\colon \mathbb{R}^D \to \mathbb{R}$ given by

$$r^{(\mathbf{p})}(\mathbf{z}) \triangleq \frac{1}{\sqrt{\mathbf{p}!}}\frac{\partial^{\mathbf{p}} R_{\mathrm{G}}(\mathbf{z}, \mathbf{z}_2)}{\partial \mathbf{z}_2^{\mathbf{p}}}\bigg|_{\mathbf{z}_2=\mathbf{0}} = \frac{1}{\sqrt{\mathbf{p}!}}\mathbf{z}^{\mathbf{p}}.$$

Furthermore, the set of functions $\big\{r^{(\mathbf{p})}(\cdot)\big\}_{\mathbf{p}\in\mathbb{Z}_+^D}$ is an orthonormal basis for $\mathcal{H}(R_{\mathrm{G}})$, and the inner product of an arbitrary function $f(\cdot) \in \mathcal{H}(R_{\mathrm{G}})$ with $r^{(\mathbf{p})}(\cdot)$ is given by

$$\big\langle f(\cdot), r^{(\mathbf{p})}(\cdot)\big\rangle_{\mathcal{H}(R_{\mathrm{G}})} = \frac{1}{\sqrt{\mathbf{p}!}}\frac{\partial^{\mathbf{p}} f(\mathbf{z})}{\partial \mathbf{z}^{\mathbf{p}}}\bigg|_{\mathbf{z}=\mathbf{0}}. \tag{57}$$

It follows that a function $f(\cdot)\colon \mathbb{R}^D \to \mathbb{R}$ belongs to $\mathcal{H}(R_{\mathrm{G}})$ if and only if it can be written pointwise as

$$f(\mathbf{z}) = f_a(\mathbf{z}) \triangleq \sum_{\mathbf{p}\in\mathbb{Z}_+^D} a[\mathbf{p}]\, r^{(\mathbf{p})}(\mathbf{z}) = \sum_{\mathbf{p}\in\mathbb{Z}_+^D} \frac{a[\mathbf{p}]}{\sqrt{\mathbf{p}!}}\mathbf{z}^{\mathbf{p}}, \tag{58}$$

with a unique coefficient sequence $a[\mathbf{p}] \in \ell^2(\mathbb{Z}_+^D)$ that is given by (57), i.e.,

$$a[\mathbf{p}] = \frac{1}{\sqrt{\mathbf{p}!}}\frac{\partial^{\mathbf{p}} f(\mathbf{z})}{\partial \mathbf{z}^{\mathbf{p}}}\bigg|_{\mathbf{z}=\mathbf{0}}.$$

The characterization (20) of the RKHS $\mathcal{H}_{\mathrm{LGM},\mathbf{x}_0}$ then follows in a straightforward manner since any function $\widetilde{f}(\cdot) \in \mathcal{H}_{\mathrm{LGM},\mathbf{x}_0}$ is the image of some $f(\cdot) = f_a(\cdot) \in \mathcal{H}(R_{\mathrm{G}})$ of the form (58) under the congruence $\mathsf{K}_{\mathrm{G}}[\cdot]$ given by (56). Furthermore, since $\mathsf{K}_{\mathrm{G}}[\cdot]$ is a congruence and the functions $\big\{r^{(\mathbf{p})}(\mathbf{z}) = \frac{1}{\sqrt{\mathbf{p}!}}\mathbf{z}^{\mathbf{p}}\big\}_{\mathbf{p}\in\mathbb{Z}_+^D}$ form an orthonormal basis for $\mathcal{H}(R_{\mathrm{G}})$, we have

$$\|\widetilde{f}_a(\cdot)\|_{\mathcal{H}_{\mathrm{LGM},\mathbf{x}_0}} = \|f_a(\cdot)\|_{\mathcal{H}(R_{\mathrm{G}})} = \|a[\cdot]\|_{\ell^2(\mathbb{Z}_+^D)},$$

which is (22).

Appendix B
Proof of Corollary III.3

1) Consider a bias function $c(\cdot) : \mathcal{X}_S \to \mathbb{R}$ that is valid for $\mathcal{E}_{\mathrm{SLGM}} = \big(\mathcal{X}_S, f_{\mathbf{H}}(\mathbf{y};\mathbf{x}), g(\cdot)\big)$ at $\mathbf{x}_0 \in \mathcal{X}_S$, and let $\gamma(\cdot) = c(\cdot) + g(\cdot)$ be the corresponding mean function. By (13) and (21), the minimum achievable variance $M_{\mathrm{SLGM}}(c(\cdot), \mathbf{x}_0)$ is given by

$$\begin{aligned}&M_{\mathrm{SLGM}}(c(\cdot), \mathbf{x}_0) \\ &\quad\stackrel{(13)}{=} \|\gamma(\cdot)\|^2_{\mathcal{H}_{\mathrm{SLGM},\mathbf{x}_0}} - \gamma^2(\mathbf{x}_0) \\ &\quad\stackrel{(21)}{=} \min_{\substack{\widetilde{f}(\cdot)\in\mathcal{H}_{\mathrm{LGM},\mathbf{x}_0} \\ \widetilde{f}(\cdot)|_{\mathcal{X}_S}=\gamma(\cdot)}} \|\widetilde{f}(\cdot)\|^2_{\mathcal{H}_{\mathrm{LGM},\mathbf{x}_0}} - \gamma^2(\mathbf{x}_0).\end{aligned} \tag{59}$$

We now use the parametrization of $\widetilde{f}(\cdot)$ in terms of the coefficient sequence $a[\cdot] \in \ell^2(\mathbb{Z}_+^D)$ according to (20), i.e., $\widetilde{f}(\cdot) = \widetilde{f}_a(\cdot)$. The minimization constraint set in (59), $\big\{\widetilde{f}(\cdot) \in \mathcal{H}_{\mathrm{LGM},\mathbf{x}_0} \big| \widetilde{f}(\cdot)\big|_{\mathcal{X}_S} = \gamma(\cdot)\big\}$, equals the set $\big\{\widetilde{f}_a(\cdot) \in \mathcal{H}_{\mathrm{LGM},\mathbf{x}_0} \big| a[\cdot] \in \mathcal{C}(c)\big\}$. Moreover, $\big\|\widetilde{f}_a(\cdot)\big\|^2_{\mathcal{H}_{\mathrm{LGM},\mathbf{x}_0}} = \|a[\cdot]\|^2_{\ell^2(\mathbb{Z}_+^D)}$ (see (22)). Therefore, (59) can be reformulated as

$$M_{\mathrm{SLGM}}(c(\cdot), \mathbf{x}_0) = \min_{a[\cdot]\in\mathcal{C}(c)} \|a[\cdot]\|^2_{\ell^2(\mathbb{Z}_+^D)} - \gamma^2(\mathbf{x}_0),$$

which is (25).

2) Consider the congruence $\mathsf{J}[\cdot]$ from $\mathcal{H}_{\mathrm{LGM},\mathbf{x}_0}$ into the space $\mathcal{L}_{\mathcal{E}_{\mathrm{LGM}},\mathbf{x}_0}$ consisting of finite-variance estimators (cf. (11)). As verified in [22, Sec. 5.3], the estimator $\hat{g}(\mathbf{y})$ in (26) satisfies $\hat{g}(\cdot) = \mathsf{J}[\widetilde{f}_a(\cdot)]$, i.e., it is the image of the function $\widetilde{f}_a(\cdot)$ (defined in (20)) under the congruence $\mathsf{J}[\cdot]$. Comparing with (14), we conclude that $\hat{g}(\mathbf{y})$ is the

LMV estimator at $\mathbf{x}_0$ for the LGM and for the bias function $\widetilde{f}_a(\mathbf{x}) - g(\mathbf{x})$, which coincides with $c(\mathbf{x})$ for any $\mathbf{x} \in \mathcal{X}_S$ due to the constraint $a[\cdot] \in \mathcal{C}(c)$. Therefore, $\hat{g}(\mathbf{y})$ is trivially an allowed estimator (i.e., having finite variance at $\mathbf{x}_0$ and the prescribed bias) for the LGM. But an allowed estimator for the LGM is also an allowed estimator for the SLGM (since the SLGM is obtained by reducing the parameter set of the LGM). Therefore, the estimator (26) is an allowed estimator at $\mathbf{x}_0$ for the SLGM and bias function $c(\cdot)$.

3) Any estimator $\hat{g}(\mathbf{y})$ given by (26) for a coefficient sequence $a[\cdot] \in \mathcal{C}(c)$ yields the bias $c(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}_S$ and its variance at $\mathbf{x}_0$ is equal to $\|a[\cdot]\|^2_{\ell^2(\mathbb{Z}_+^D)} - \gamma^2(\mathbf{x}_0)$. According to (25), the minimum achievable variance at $\mathbf{x}_0$ is given by the minimum of $\|a[\cdot]\|^2_{\ell^2(\mathbb{Z}_+^D)} - \gamma^2(\mathbf{x}_0)$ over all $a[\cdot] \in \mathcal{C}(c)$. Therefore, the LMV estimator for the SLGM and bias function $c(\cdot)$ is given by (26) using the specific coefficient sequence $a[\cdot] \in \mathcal{C}(c)$ that minimizes $\|a[\cdot]\|^2_{\ell^2(\mathbb{Z}_+^D)}$, implying that the estimator variance achieves the minimum achievable variance (25). This minimizing sequence $a[\cdot]$ is unique, due to the uniqueness of the LMV estimator [19, p. 85] and the uniqueness of the coefficient sequence in the representation (58).

APPENDIX C
PROOF OF COROLLARY III.4

We first note that our assumption that the prescribed bias function $c(\cdot)$ is valid for $\mathcal{E}_{\text{SLGM}}$ at every $\mathbf{x} \in \mathcal{X}_S$ has two consequences. First, $M_{\text{SLGM}}(c(\cdot), \mathbf{x}) < \infty$ for every $\mathbf{x} \in \mathcal{X}_S$ (cf. our definition of the validity of a bias function in Section II-B) and therefore, $M_{\text{SLGM}}(c(\cdot), \mathbf{x})$ is a well-defined function of $\mathbf{x}$; second, due to (12), the prescribed mean function $\gamma(\cdot) = c(\cdot) + g(\cdot)$ belongs to $\mathcal{H}_{\text{SLGM},\mathbf{x}}$ for every $\mathbf{x} \in \mathcal{X}_S$.

A. Representation of $M_{\text{SLGM}}(c(\cdot), \mathbf{x})$

Let us denote by $\mathcal{L}_{\mathbf{x}}$ the set of all functions $f(\cdot) : \mathcal{X} \to \mathbb{R}$ that are finite linear combinations of the form

$$f(\cdot) = \sum_{l \in [L]} a_l R_{\text{SLGM},\mathbf{x}}(\cdot, \mathbf{x}_l), \quad \text{with } \mathbf{x}_l \in \mathcal{X}_S,\ a_l \in \mathbb{R},\ L \in \mathbb{N}. \tag{60}$$

As shown in [22, Th. 3.2.2], the set $\mathcal{L}_{\mathbf{x}}$ spans $\mathcal{H}_{\text{SLGM},\mathbf{x}}$. Hence [31, Ch. 4], the squared norm of any function $h(\cdot) \in \mathcal{H}_{\text{SLGM},\mathbf{x}}$ can be expressed as

$$\|h(\cdot)\|^2_{\mathcal{H}_{\text{SLGM},\mathbf{x}}} = \sup_{\substack{f(\cdot) \in \mathcal{L}_{\mathbf{x}} \\ \|f(\cdot)\|^2_{\mathcal{H}_{\text{SLGM},\mathbf{x}}} > 0}} \frac{\langle h(\cdot), f(\cdot)\rangle^2_{\mathcal{H}_{\text{SLGM},\mathbf{x}}}}{\|f(\cdot)\|^2_{\mathcal{H}_{\text{SLGM},\mathbf{x}}}}. \tag{61}$$

We can now develop the minimum achievable variance $M_{\text{SLGM}}(c(\cdot), \mathbf{x})$ as follows:

$$M_{\text{SLGM}}(c(\cdot), \mathbf{x}) \overset{(13)}{=} \|\gamma(\cdot)\|^2_{\mathcal{H}_{\text{SLGM},\mathbf{x}}} - \gamma^2(\mathbf{x}) \overset{(61)}{=} \sup_{\substack{f(\cdot) \in \mathcal{L}_{\mathbf{x}} \\ \|f(\cdot)\|^2_{\mathcal{H}_{\text{SLGM},\mathbf{x}}} > 0}} \frac{\langle \gamma(\cdot), f(\cdot)\rangle^2_{\mathcal{H}_{\text{SLGM},\mathbf{x}}}}{\|f(\cdot)\|^2_{\mathcal{H}_{\text{SLGM},\mathbf{x}}}} - \gamma^2(\mathbf{x}).$$

Here, it was possible to use (61) because $\gamma(\cdot) \in \mathcal{H}_{\text{SLGM},\mathbf{x}}$. Using (60) and letting $\mathcal{D} \triangleq \{\mathbf{x}_1, \ldots, \mathbf{x}_L\}$, $\mathbf{a} \triangleq (a_1 \cdots a_L)^T$, and $\mathcal{A}_{\mathcal{D}} \triangleq \big\{\mathbf{a} \in \mathbb{R}^L \big| \big\| \sum_{l \in [L]} a_l R_{\text{SLGM},\mathbf{x}}(\cdot, \mathbf{x}_l)\big\|_{\mathcal{H}_{\text{SLGM},\mathbf{x}}} > 0\big\}$, we obtain further

$$M_{\text{SLGM}}(c(\cdot), \mathbf{x}) = \sup_{\mathcal{D} \subseteq \mathcal{X},\, L \in \mathbb{N},\, \mathbf{a} \in \mathcal{A}_{\mathcal{D}}} h_{\mathcal{D},\mathbf{a}}(\mathbf{x}), \tag{62}$$

with

$$h_{\mathcal{D},\mathbf{a}}(\mathbf{x}) \triangleq \frac{\big\langle \gamma(\cdot), \sum_{l \in [L]} a_l R_{\text{SLGM},\mathbf{x}}(\cdot, \mathbf{x}_l)\big\rangle^2_{\mathcal{H}_{\text{SLGM},\mathbf{x}}}}{\big\| \sum_{l \in [L]} a_l R_{\text{SLGM},\mathbf{x}}(\cdot, \mathbf{x}_l)\big\|^2_{\mathcal{H}_{\text{SLGM},\mathbf{x}}}} - \gamma^2(\mathbf{x}).$$

Our notation $\sup_{\mathcal{D} \subseteq \mathcal{X},\, L \in \mathbb{N},\, \mathbf{a} \in \mathcal{A}_{\mathcal{D}}}$ in (62) indicates that the supremum is taken not only with respect to the elements $\mathbf{x}_l$ of $\mathcal{D}$ but also with respect to the size of $\mathcal{D}$, $L = |\mathcal{D}|$. For any finite set $\mathcal{D} = \{\mathbf{x}_1, \ldots, \mathbf{x}_L\} \subseteq \mathcal{X}$ and any $\mathbf{a} \in \mathcal{A}_{\mathcal{D}}$, it follows from the continuity of $R_{\text{SLGM},\mathbf{x}}(\cdot, \cdot)$ with respect to $\mathbf{x}$ on $\mathcal{X}_S$ (see (18)) and our assumption of continuity of $c(\cdot)$ (and, therefore, of $\gamma(\mathbf{x}) = c(\mathbf{x}) + x_k$) that the function $h_{\mathcal{D},\mathbf{a}}(\mathbf{x})$ is continuous in a neighborhood around any point $\mathbf{x}_0 \in \mathcal{X}_S$. Thus, for any $\mathbf{x}_0 \in \mathcal{X}_S$, there exists a radius $\delta_0 > 0$ such that $h_{\mathcal{D},\mathbf{a}}(\mathbf{x})$ is continuous on $\mathcal{B}(\mathbf{x}_0, \delta_0) \cap \mathcal{X}_S$, where $\mathcal{B}(\mathbf{x}_0, \delta_0)$ denotes the open ball of radius δ_0 around $\mathbf{x}_0$.

B. Lower Semi-Continuity of $M_{\text{SLGM}}(c(\cdot), \mathbf{x})$

We will now show that the function $M_{\text{SLGM}}(c(\cdot), \mathbf{x})$ given by (62) is lower semi-continuous [59] at every $\mathbf{x}_0 \in \mathcal{X}_S$, which means that for any $\mathbf{x}_0 \in \mathcal{X}_S$ and $\varepsilon > 0$ there is a radius $r > 0$ such that

$$M_{\text{SLGM}}(c(\cdot), \mathbf{x}) \geq M_{\text{SLGM}}(c(\cdot), \mathbf{x}_0) - \varepsilon, \quad \text{for all } \mathbf{x} \in \mathcal{B}(\mathbf{x}_0, r) \cap \mathcal{X}_S. \tag{63}$$

Due to (62), there must be a finite subset $\mathcal{D}_0 \subseteq \mathcal{X}_S$ and a vector $\mathbf{a}_0 \in \mathcal{A}_{\mathcal{D}_0}$ such that[11]

$$h_{\mathcal{D}_0,\mathbf{a}_0}(\mathbf{x}_0) \geq M_{\text{SLGM}}(c(\cdot), \mathbf{x}_0) - \frac{\varepsilon}{2}, \tag{64}$$

for any given $\varepsilon > 0$. Furthermore, since $h_{\mathcal{D}_0,\mathbf{a}_0}(\mathbf{x})$ is continuous on $\mathcal{B}(\mathbf{x}_0, \delta_0) \cap \mathcal{X}_S$ as shown above, there is a radius $r_0 > 0$ (with $r_0 < \delta_0$) such that

$$h_{\mathcal{D}_0,\mathbf{a}_0}(\mathbf{x}) \geq h_{\mathcal{D}_0,\mathbf{a}_0}(\mathbf{x}_0) - \frac{\varepsilon}{2}, \quad \text{for all } \mathbf{x} \in \mathcal{B}(\mathbf{x}_0, r_0) \cap \mathcal{X}_S. \tag{65}$$

By combining this inequality with (64), it follows that there is a radius $r > 0$ (with $r < \delta_0$) such that for any $\mathbf{x} \in \mathcal{B}(\mathbf{x}_0, r) \cap \mathcal{X}_S$ we have

$$h_{\mathcal{D}_0,\mathbf{a}_0}(\mathbf{x}) \overset{(65)}{\geq} h_{\mathcal{D}_0,\mathbf{a}_0}(\mathbf{x}_0) - \frac{\varepsilon}{2} \overset{(64)}{\geq} M_{\text{SLGM}}(c(\cdot), \mathbf{x}_0) - \varepsilon, \tag{66}$$

and further

$$M_{\text{SLGM}}(c(\cdot), \mathbf{x}) \overset{(62)}{=} \sup_{\mathcal{D} \subseteq \mathcal{X},\, L \in \mathbb{N},\, \mathbf{a} \in \mathcal{A}_{\mathcal{D}}} h_{\mathcal{D},\mathbf{a}}(\mathbf{x}) \geq h_{\mathcal{D}_0,\mathbf{a}_0}(\mathbf{x}) \overset{(66)}{\geq} M_{\text{SLGM}}(c(\cdot), \mathbf{x}_0) - \varepsilon.$$

[11] Indeed, if (64) were not true, we would have $h_{\mathcal{D},\mathbf{a}}(\mathbf{x}_0) < M_{\text{SLGM}}(c(\cdot), \mathbf{x}_0) - \varepsilon/2$ for every choice of $\mathcal{D}$ and $\mathbf{a}$. This, in turn, would imply that $\sup_{\mathcal{D} \subseteq \mathcal{X},\, L \in \mathbb{N},\, \mathbf{a} \in \mathcal{A}_{\mathcal{D}}} h_{\mathcal{D},\mathbf{a}}(\mathbf{x}_0) \leq M_{\text{SLGM}}(c(\cdot), \mathbf{x}_0) - \varepsilon/2 < M_{\text{SLGM}}(c(\cdot), \mathbf{x}_0)$, yielding the contradiction $M_{\text{SLGM}}(c(\cdot), \mathbf{x}_0) \overset{(62)}{=} \sup_{\mathcal{D} \subseteq \mathcal{X},\, L \in \mathbb{N},\, \mathbf{a} \in \mathcal{A}_{\mathcal{D}}} h_{\mathcal{D},\mathbf{a}}(\mathbf{x}_0) < M_{\text{SLGM}}(c(\cdot), \mathbf{x}_0)$.

Thus, for any given $\varepsilon > 0$, there is a radius $r > 0$ (with $r < \delta_0$) such that $M_{\text{SLGM}}(c(\cdot), \mathbf{x}) \geq M_{\text{SLGM}}(c(\cdot), \mathbf{x}_0) - \varepsilon$ for all $\mathbf{x} \in \mathcal{B}(\mathbf{x}_0, r) \cap \mathcal{X}_S$, i.e., (63) has been proved.

Appendix D
Proof of Theorem IV.1

Consider the subspace $\mathcal{U} = \operatorname{span}\{u_0(\cdot), \{\tilde{u}_i(\cdot)\}_{i \in [|\mathcal{K}|]}\} \subseteq \mathcal{H}_{\text{SLGM},\mathbf{x}_0}$ (see (30)). According to [22, Th. 3.1.9], the squared norm of the projection $\mathsf{P}_{\mathcal{U}}\gamma(\cdot)$ of the prescribed mean function $\gamma(\mathbf{x}) = c(\mathbf{x}) + x_k$ onto $\mathcal{U}$ is given by

$$\|\mathsf{P}_{\mathcal{U}}\gamma(\cdot)\|^2_{\mathcal{H}_{\text{SLGM},\mathbf{x}_0}} = \mathbf{c}^T \mathbf{G}^\dagger \mathbf{c} + \gamma^2(\mathbf{x}_0), \tag{67}$$

where the vector $\mathbf{c} \in \mathbb{R}^{|\mathcal{K}|}$ and the matrix $\mathbf{G} \in \mathbb{R}^{|\mathcal{K}| \times |\mathcal{K}|}$ are defined entrywise as $c_i \triangleq \langle \gamma(\cdot), \tilde{u}_i(\cdot) \rangle_{\mathcal{H}_{\text{SLGM},\mathbf{x}_0}}$, $i \in [|\mathcal{K}|]$ and $G_{m,n} \triangleq \langle \tilde{u}_m(\cdot), \tilde{u}_n(\cdot) \rangle_{\mathcal{H}_{\text{SLGM},\mathbf{x}_0}}$, $m, n \in [|\mathcal{K}|]$, respectively. We obtain

$$\begin{aligned} c_i &\overset{(30)}{=} \left\langle \gamma(\cdot), \frac{\partial R_{\text{SLGM},\mathbf{x}_0}(\cdot, \mathbf{x})}{\partial x_{l_i}} \bigg|_{\mathbf{x}=\tilde{\mathbf{x}}_0} \right\rangle_{\mathcal{H}_{\text{SLGM},\mathbf{x}_0}} \\ &\overset{(a)}{=} \frac{\gamma(\mathbf{x})}{\partial x_{l_i}} \bigg|_{\mathbf{x}=\tilde{\mathbf{x}}_0} \\ &= \frac{c(\mathbf{x})}{\partial x_{l_i}} \bigg|_{\mathbf{x}=\tilde{\mathbf{x}}_0} + \delta_{l_i,k}, \end{aligned}$$

for $i \in [|\mathcal{K}|]$. Here, step (a) is due to the derivative-reproducing property [22], [58] of an RKHS associated with a differentiable kernel function. Similarly,

$$\begin{aligned} &G_{m,n} \\ &\overset{(30)}{=} \left\langle \frac{\partial R_{\text{SLGM},\mathbf{x}_0}(\cdot, \mathbf{x}_1)}{\partial (\mathbf{x}_1)_{l_m}} \bigg|_{\mathbf{x}_1=\tilde{\mathbf{x}}_0}, \frac{\partial R_{\text{SLGM},\mathbf{x}_0}(\cdot, \mathbf{x}_2)}{\partial (\mathbf{x}_2)_{l_n}} \bigg|_{\mathbf{x}_2=\tilde{\mathbf{x}}_0} \right\rangle_{\mathcal{H}_{\text{SLGM},\mathbf{x}_0}} \\ &= \frac{\partial R_{\text{SLGM},\mathbf{x}_0}(\mathbf{x}_1, \mathbf{x}_2)}{\partial (\mathbf{x}_1)_{l_m} \partial (\mathbf{x}_2)_{l_n}} \bigg|_{\mathbf{x}_1=\mathbf{x}_2=\tilde{\mathbf{x}}_0}, \end{aligned} \tag{68}$$

for $m, n \in [|\mathcal{K}|]$. Inserting (18) into (68) and using some elementary linear algebra (see [22, p. 105]) yields

$$G_{m,n} = \frac{1}{\sigma^2} \mathbf{e}_{l_n}^T \mathbf{H}^T \mathbf{H} \mathbf{e}_{l_m} \exp\left(\frac{1}{\sigma^2} \|\mathbf{H}(\tilde{\mathbf{x}}_0 - \mathbf{x}_0)\|_2^2 \right),$$

where $\|\mathbf{H}(\tilde{\mathbf{x}}_0 - \mathbf{x}_0)\|_2^2 = \|\mathbf{P}\mathbf{H}\mathbf{x}_0 - \mathbf{H}\mathbf{x}_0\|_2^2 = \|(\mathbf{I}-\mathbf{P})\mathbf{H}\mathbf{x}_0\|_2^2$ (recall that $\mathbf{H}\tilde{\mathbf{x}}_0 = \mathbf{P}\mathbf{H}\mathbf{x}_0$). The bound (31) then follows upon inserting (67) into (27).

Appendix E
Proof of Corollary IV.2

Let $\mathbf{P} \in \mathbb{R}^{M \times M}$ denote the orthogonal projection matrix on the subspace $\operatorname{span}(\mathbf{H}_{\mathcal{K}}) \subseteq \mathbb{R}^M$. Because $\mathbf{H}$ satisfies the RIP of order S, $\mathbf{H}_{\mathcal{K}}$ has full column rank [21], and therefore $\mathbf{P} = \mathbf{H}_{\mathcal{K}}(\mathbf{H}_{\mathcal{K}}^T \mathbf{H}_{\mathcal{K}})^{-1} \mathbf{H}_{\mathcal{K}}^T$. We thus obtain $(\mathbf{I}-\mathbf{P})\mathbf{H}_{\mathcal{K}} = \mathbf{H}_{\mathcal{K}} - \mathbf{H}_{\mathcal{K}}(\mathbf{H}_{\mathcal{K}}^T \mathbf{H}_{\mathcal{K}})^{-1} \mathbf{H}_{\mathcal{K}}^T \mathbf{H}_{\mathcal{K}} = \mathbf{0}$, and hence $(\mathbf{I}-\mathbf{P})\mathbf{x}' = \mathbf{0}$ for every vector $\mathbf{x}' \in \operatorname{span}(\mathbf{H}_{\mathcal{K}})$. This implies that

$$\begin{aligned} (\mathbf{I}-\mathbf{P})\mathbf{H}\mathbf{x}_0 &= (\mathbf{I}-\mathbf{P})\mathbf{H}\big(\mathbf{x}_0^{\text{supp}(\mathbf{x}_0)\setminus\mathcal{K}} + \mathbf{x}_0^{\mathcal{K}}\big) \\ &= (\mathbf{I}-\mathbf{P})\mathbf{H}\mathbf{x}_0^{\text{supp}(\mathbf{x}_0)\setminus\mathcal{K}}, \end{aligned} \tag{69}$$

since $\mathbf{H}\mathbf{x}_0^{\mathcal{K}} \in \operatorname{span}(\mathbf{H}_{\mathcal{K}})$. Based on (69) and using the shorthand $\mathbf{x}_0' \triangleq \mathbf{x}_0^{\text{supp}(\mathbf{x}_0)\setminus\mathcal{K}}$, we obtain

$$\begin{aligned} \|(\mathbf{I}-\mathbf{P})\mathbf{H}\mathbf{x}_0\|_2^2 &= \|(\mathbf{I}-\mathbf{P})\mathbf{H}\mathbf{x}_0'\|_2^2 \\ &= \|\mathbf{H}\mathbf{x}_0'\|_2^2 - 2(\mathbf{x}_0')^T \mathbf{H}^T \mathbf{P} \mathbf{H} \mathbf{x}_0' \\ &\quad + (\mathbf{x}_0')^T \mathbf{H}^T \mathbf{P}^2 \mathbf{H} \mathbf{x}_0' \\ &\overset{(a)}{=} \|\mathbf{H}\mathbf{x}_0'\|_2^2 - (\mathbf{x}_0')^T \mathbf{H}^T \mathbf{P} \mathbf{H} \mathbf{x}_0' \\ &\overset{(b)}{\leq} \|\mathbf{H}\mathbf{x}_0'\|_2^2 \\ &\overset{(c)}{\leq} (1+\delta_S)\|\mathbf{x}_0'\|_2^2, \end{aligned} \tag{70}$$

where step (a) follows from $\mathbf{P}^2 = \mathbf{P}$ [21], step (b) follows from the fact that $\mathbf{P}$ is positive semidefinite, and step (c) is due to $\|\mathbf{x}_0'\|_0 \leq S$ and the assumption that $\mathbf{H}$ satisfies the RIP (33) of order S with RIP constant δ_S. The bound (34) then follows by reformulating (31) using (70):

$$\begin{aligned} &M_{\text{SLGM}}(c(\cdot), \mathbf{x}_0) \\ &\overset{(31)}{\geq} \exp\left(-\frac{1}{\sigma^2} \|(\mathbf{I}-\mathbf{P})\mathbf{H}\mathbf{x}_0\|_2^2 \right) \sigma^2 \mathbf{b}_{\mathbf{x}_0}^T \big(\mathbf{H}_{\mathcal{K}}^T \mathbf{H}_{\mathcal{K}}\big)^{-1} \mathbf{b}_{\mathbf{x}_0} \\ &\overset{(70)}{\geq} \exp\left(-\frac{1+\delta_S}{\sigma^2} \big\|\mathbf{x}_0^{\text{supp}(\mathbf{x}_0)\setminus\mathcal{K}}\big\|_2^2 \right) \sigma^2 \mathbf{b}_{\mathbf{x}_0}^T \big(\mathbf{H}_{\mathcal{K}}^T \mathbf{H}_{\mathcal{K}}\big)^{-1} \mathbf{b}_{\mathbf{x}_0}. \end{aligned}$$

Since (33) implies $\|\mathbf{H}_{\mathcal{K}}^T \mathbf{H}_{\mathcal{K}}\|_2 \leq 1 + \delta_S$ and, in turn, $\|(\mathbf{H}_{\mathcal{K}}^T \mathbf{H}_{\mathcal{K}})^{-1}\|_2 \geq 1/(1+\delta_S)$ (here, $\|\cdot\|_2$ denotes the spectral matrix norm), we finally obtain (34).

Appendix F
Proof of Theorem V.1

This proof is, in its first part, essentially analogous to that in Appendix A; however, the domain is different ($\mathcal{X}_S$ rather than $\mathbb{R}^N$ or $\mathbb{R}^D$). We will need the following lemma, which is analogous to Lemma A.1.

Lemma F.1: The RKHS $\mathcal{H}_{\text{SSNM},\mathbf{x}_0}$ is isometric to the RKHS $\mathcal{H}(R_{\text{e}})$ whose kernel $R_{\text{e}}(\cdot,\cdot)\colon \mathcal{X}_S \times \mathcal{X}_S \to \mathbb{R}$ is given by

$$R_{\text{e}}(\mathbf{x}_1, \mathbf{x}_2) = \exp\big(\mathbf{x}_1^T \mathbf{x}_2\big), \quad \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}_S. \tag{71}$$

A congruence from $\mathcal{H}(R_{\text{e}})$ to $\mathcal{H}_{\text{SSNM},\mathbf{x}_0}$ is constituted by the mapping $\mathsf{K}_{\text{e}}[\cdot]\colon \mathcal{H}(R_{\text{e}}) \to \mathcal{H}_{\text{SSNM},\mathbf{x}_0}$ given by

$$\mathsf{K}_{\text{e}}[f(\cdot)] = \widetilde{f}(\mathbf{x}) \triangleq f\left(\frac{\mathbf{x}}{\sigma}\right) \nu_{\mathbf{x}_0}(\mathbf{x}), \quad \mathbf{x} \in \mathcal{X}_S, \tag{72}$$

for all $f(\cdot) \in \mathcal{H}(R_{\text{e}})$, with $\nu_{\mathbf{x}_0}(\mathbf{x}) \triangleq \exp\big(\frac{1}{2\sigma^2}\|\mathbf{x}_0\|_2^2 - \frac{1}{\sigma^2}\mathbf{x}^T\mathbf{x}_0\big)$. A congruence from $\mathcal{H}_{\text{SSNM},\mathbf{x}_0}$ to $\mathcal{H}(R_{\text{e}})$ is constituted by the inverse mapping $\mathsf{K}_{\text{e}}^{-1}[\cdot]\colon \mathcal{H}_{\text{SSNM},\mathbf{x}_0} \to \mathcal{H}(R_{\text{e}})$ given by

$$\mathsf{K}_{\text{e}}^{-1}[\widetilde{f}(\cdot)] = f(\mathbf{x}) \triangleq \frac{\widetilde{f}(\sigma\mathbf{x})}{\nu_{\mathbf{x}_0}(\sigma\mathbf{x})}, \quad \mathbf{x} \in \mathcal{X}_S, \tag{73}$$

for all $\widetilde{f}(\cdot) \in \mathcal{H}_{\text{SSNM},\mathbf{x}_0}$.

Proof: The sets

$$\mathcal{A} \triangleq \big\{\widetilde{f}_{\mathbf{x}}(\cdot) \triangleq R_{\text{SSNM},\mathbf{x}_0}(\cdot, \sigma\mathbf{x})\big\}_{\mathbf{x}\in\mathcal{X}_S}$$

and

$$\mathcal{B} \triangleq \left\{ f_{\mathbf{x}}(\cdot) \triangleq \exp\left(\frac{1}{2\sigma^2}\|\mathbf{x}_0\|_2^2 - \frac{1}{\sigma}\mathbf{x}^T\mathbf{x}_0 \right) R_{\text{e}}(\cdot, \mathbf{x}) \right\}_{\mathbf{x}\in\mathcal{X}_S}$$

span the RKHSs $\mathcal{H}_{\mathrm{SSNM},\mathbf{x}_0}$ and $\mathcal{H}(R_e)$, respectively [22, Sec. 5.5]. A continuous linear mapping that maps $f_{\mathbf{x}}(\cdot) \in \mathcal{H}(R_e)$ to $\widetilde{f}_{\mathbf{x}}(\cdot) \in \mathcal{H}_{\mathrm{SSNM},\mathbf{x}_0}$ for any $\mathbf{x} \in \mathcal{X}_S$ is given by K_e, and moreover $\langle \widetilde{f}_{\mathbf{x}_1}(\cdot), \widetilde{f}_{\mathbf{x}_2}(\cdot)\rangle_{\mathcal{H}_{\mathrm{SSNM},\mathbf{x}_0}} = \langle f_{\mathbf{x}_1}(\cdot), f_{\mathbf{x}_2}(\cdot)\rangle_{\mathcal{H}(R_e)}$ for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}_S$. Hence, according to [16, p. 263], K_e is a congruence from $\mathcal{H}(R_e)$ to $\mathcal{H}_{\mathrm{SSNM},\mathbf{x}_0}$. A detailed proof is provided in [22, Th. 5.2.2]. □

Note that the kernel R_e in (71) is the restriction of the kernel R_G in (55) to the subdomain $\mathcal{X}_S \times \mathcal{X}_S \subseteq \mathbb{R}^N \times \mathbb{R}^N$.

Due to Lemma F.1, the characterization of the RKHS $\mathcal{H}_{\mathrm{SSNM},\mathbf{x}_0}$ is reduced to a characterization of the RKHS $\mathcal{H}(R_e)$. A simple characterization of $\mathcal{H}(R_e)$ in terms of an orthonormal basis can be obtained by exploiting the differentiability of the kernel $R_e(\cdot,\cdot)$. In particular, one can show [58], [22, Th. 5.5.1] that for any $\mathbf{p} \in \mathbb{Z}_+^N \cap \mathcal{X}_S$, $\mathcal{H}(R_e)$ contains the function $r^{(\mathbf{p})}(\cdot)\colon \mathcal{X}_S \to \mathbb{R}$ given by

$$r^{(\mathbf{p})}(\mathbf{x}) \triangleq \frac{1}{\sqrt{\mathbf{p}!}} \frac{\partial^{\mathbf{p}} R_e(\mathbf{x},\mathbf{x}_2)}{\partial \mathbf{x}_2^{\mathbf{p}}}\bigg|_{\mathbf{x}_2=\mathbf{0}} = \frac{1}{\sqrt{\mathbf{p}!}}\mathbf{x}^{\mathbf{p}}. \tag{74}$$

Furthermore, $\{r^{(\mathbf{p})}(\cdot)\}_{\mathbf{p}\in\mathbb{Z}_+^N\cap\mathcal{X}_S}$ is an orthonormal basis for $\mathcal{H}(R_e)$, and the inner product of an arbitrary function $f(\cdot) \in \mathcal{H}(R_e)$ with $r^{(\mathbf{p})}(\cdot)$ is given by

$$\big\langle f(\cdot), r^{(\mathbf{p})}(\cdot)\big\rangle_{\mathcal{H}(R_e)} = \frac{1}{\sqrt{\mathbf{p}!}} \frac{\partial^{\mathbf{p}} f(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}}}\bigg|_{\mathbf{x}=\mathbf{0}}. \tag{75}$$

Hence, a function $f(\cdot) : \mathcal{X}_S \to \mathbb{R}$ belongs to $\mathcal{H}(R_e)$ if and only if it can be written pointwise as

$$f(\mathbf{x}) = \sum_{\mathbf{p}\in\mathbb{Z}_+^N\cap\mathcal{X}_S} a[\mathbf{p}]\, r^{(\mathbf{p})}(\mathbf{x}) = \sum_{\mathbf{p}\in\mathbb{Z}_+^N\cap\mathcal{X}_S} \frac{a[\mathbf{p}]}{\sqrt{\mathbf{p}!}}\mathbf{x}^{\mathbf{p}}, \tag{76}$$

with a unique coefficient sequence $a[\mathbf{p}] \in \ell^2(\mathbb{Z}_+^N \cap \mathcal{X}_S)$ that is given by (75), i.e.,

$$a[\mathbf{p}] = \frac{1}{\sqrt{\mathbf{p}!}} \frac{\partial^{\mathbf{p}} f(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}}}\bigg|_{\mathbf{x}=\mathbf{0}}. \tag{77}$$

We are now ready to prove the claims made in the theorem. To show (37), we note that according to (12), a prescribed bias function $c(\cdot) : \mathcal{X}_S \to \mathbb{R}$ is valid for $\mathcal{E}_{\mathrm{SSNM}}$ at $\mathbf{x}_0$ if and only if the corresponding mean function $\gamma(\cdot) = c(\cdot) + g(\cdot)$ belongs to $\mathcal{H}_{\mathrm{SSNM},\mathbf{x}_0}$. Furthermore, due to Lemma F.1, any function $\gamma(\cdot) \in \mathcal{H}_{\mathrm{SSNM},\mathbf{x}_0}$ is the image $\mathsf{K}_e[f(\cdot)]$ of some function $f(\cdot) \in \mathcal{H}(R_e)$ under the congruence K_e in (72), i.e.,

$$\begin{aligned}\gamma(\mathbf{x}) &= f\left(\frac{\mathbf{x}}{\sigma}\right)\nu_{\mathbf{x}_0}(\mathbf{x}) \\ &\overset{(76)}{=} \nu_{\mathbf{x}_0}(\mathbf{x}) \sum_{\mathbf{p}\in\mathbb{Z}_+^N\cap\mathcal{X}_S} a[\mathbf{p}]\, r^{(\mathbf{p})}\left(\frac{\mathbf{x}}{\sigma}\right) \\ &= \nu_{\mathbf{x}_0}(\mathbf{x}) \sum_{\mathbf{p}\in\mathbb{Z}_+^N\cap\mathcal{X}_S} \frac{a[\mathbf{p}]}{\sqrt{\mathbf{p}!}}\left(\frac{\mathbf{x}}{\sigma}\right)^{\mathbf{p}},\end{aligned} \tag{78}$$

with a unique coefficient sequence $a[\mathbf{p}] \in \ell^2(\mathbb{Z}_+^N \cap \mathcal{X}_S)$.

Next, we will show (39) and (40). Consider a prescribed bias function $c(\cdot)$ that is valid for $\mathcal{E}_{\mathrm{SSNM}}$ at $\mathbf{x}_0$, i.e., the corresponding mean function $\gamma(\cdot) = c(\cdot) + g(\cdot)$ is given by (78). Then, since $\mathsf{K}_e[\cdot]$ is a congruence and the coefficients $a[\mathbf{p}]$ are the expansion coefficients of the function $f(\cdot)$ with respect to the orthonormal basis $\{r^{(\mathbf{p})}(\cdot)\}_{\mathbf{p}\in\mathbb{Z}_+^N\cap\mathcal{X}_S}$, we have

$$\begin{aligned}\|\gamma(\cdot)\|^2_{\mathcal{H}_{\mathrm{SSNM},\mathbf{x}_0}} = \|\mathsf{K}_e[f(\cdot)]\|^2_{\mathcal{H}_{\mathrm{SSNM},\mathbf{x}_0}} &= \|f(\cdot)\|^2_{\mathcal{H}(R_e)} \\ &= \|a[\cdot]\|^2_{\ell^2(\mathbb{Z}_+^N\cap\mathcal{X}_S)}.\end{aligned} \tag{79}$$

The expression (39) for the minimum achievable variance $M_{\mathrm{SSNM}}(c(\cdot),\mathbf{x}_0)$ then follows by

$$\begin{aligned}M_{\mathrm{SSNM}}(c(\cdot),\mathbf{x}_0) &\overset{(13)}{=} \|\gamma(\cdot)\|^2_{\mathcal{H}_{\mathrm{SSNM},\mathbf{x}_0}} - \gamma^2(\mathbf{x}_0) \\ &\overset{(79)}{=} \|a[\cdot]\|^2_{\ell^2(\mathbb{Z}_+^N\cap\mathcal{X}_S)} - \gamma^2(\mathbf{x}_0).\end{aligned}$$

Furthermore, the expression (40) for the coefficients $a[\mathbf{p}]$ follows from (77) and (73), i.e.,

$$a[\mathbf{p}] = \frac{1}{\sqrt{\mathbf{p}!}} \frac{\partial^{\mathbf{p}}\big(\gamma(\sigma\mathbf{x})/\nu_{\mathbf{x}_0}(\sigma\mathbf{x})\big)}{\partial \mathbf{x}^{\mathbf{p}}}\bigg|_{\mathbf{x}=\mathbf{0}}.$$

Finally, to derive (41), we insert (78) into (14):

$$\begin{aligned}\hat{g}^{(c(\cdot),\mathbf{x}_0)}(\mathbf{y}) &\overset{(14)}{=} \mathsf{J}[\gamma(\cdot)] \\ &\overset{(78)}{=} \mathsf{J}\bigg[\nu_{\mathbf{x}_0}(\mathbf{x}) \sum_{\mathbf{p}\in\mathbb{Z}_+^N\cap\mathcal{X}_S} a[\mathbf{p}]\, r^{(\mathbf{p})}\left(\frac{\mathbf{x}}{\sigma}\right)\bigg] \\ &= \sum_{\mathbf{p}\in\mathbb{Z}_+^N\cap\mathcal{X}_S} a[\mathbf{p}]\, \mathsf{J}\bigg[\nu_{\mathbf{x}_0}(\mathbf{x})\, r^{(\mathbf{p})}\left(\frac{\mathbf{x}}{\sigma}\right)\bigg].\end{aligned} \tag{80}$$

We have

$$\begin{aligned}&\nu_{\mathbf{x}_0}(\mathbf{x})\, r^{(\mathbf{p})}\left(\frac{\mathbf{x}}{\sigma}\right) \\ &\overset{(38),(74)}{=} \exp\left(\frac{1}{2\sigma^2}\|\mathbf{x}_0\|_2^2 - \frac{1}{\sigma^2}\mathbf{x}^T\mathbf{x}_0\right)\frac{1}{\sqrt{\mathbf{p}!}}\left(\frac{\mathbf{x}}{\sigma}\right)^{\mathbf{p}} \\ &= \frac{1}{\sqrt{\mathbf{p}!}}\exp\left(-\frac{1}{2\sigma^2}\|\mathbf{x}_0\|_2^2\right)\exp\left(\frac{1}{\sigma^2}\|\mathbf{x}_0\|_2^2 - \frac{1}{\sigma^2}\mathbf{x}^T\mathbf{x}_0\right) \\ &\quad \times \frac{\partial^{\mathbf{p}}\exp\left(\frac{1}{\sigma}\tilde{\mathbf{x}}^T\mathbf{x}\right)}{\partial\tilde{\mathbf{x}}^{\mathbf{p}}}\bigg|_{\tilde{\mathbf{x}}=\mathbf{0}} \\ &= \frac{1}{\sqrt{\mathbf{p}!}}\exp\left(-\frac{1}{2\sigma^2}\|\mathbf{x}_0\|_2^2\right) \\ &\quad \times \frac{\partial^{\mathbf{p}}\exp\left(\frac{1}{\sigma^2}(\mathbf{x}-\mathbf{x}_0)^T(\sigma\tilde{\mathbf{x}}-\mathbf{x}_0) + \frac{1}{\sigma}\mathbf{x}_0^T\tilde{\mathbf{x}}\right)}{\partial\tilde{\mathbf{x}}^{\mathbf{p}}}\bigg|_{\tilde{\mathbf{x}}=\mathbf{0}} \\ &\overset{(36)}{=} \frac{1}{\sqrt{\mathbf{p}!}}\exp\left(-\frac{1}{2\sigma^2}\|\mathbf{x}_0\|_2^2\right) \\ &\quad \times \frac{\partial^{\mathbf{p}}\big[R_{\mathrm{SSNM},\mathbf{x}_0}(\mathbf{x},\sigma\tilde{\mathbf{x}})\exp\left(\frac{1}{\sigma}\mathbf{x}_0^T\tilde{\mathbf{x}}\right)\big]}{\partial\tilde{\mathbf{x}}^{\mathbf{p}}}\bigg|_{\tilde{\mathbf{x}}=\mathbf{0}}.\end{aligned}$$

Hence,

$$\begin{aligned}&\mathsf{J}\bigg[\nu_{\mathbf{x}_0}(\mathbf{x})\, r^{(\mathbf{p})}\left(\frac{\mathbf{x}}{\sigma}\right)\bigg] \\ &\quad = \frac{1}{\sqrt{\mathbf{p}!}}\exp\left(-\frac{1}{2\sigma^2}\|\mathbf{x}_0\|_2^2\right) \\ &\qquad \times \mathsf{J}\bigg[\frac{\partial^{\mathbf{p}}\big[R_{\mathrm{SSNM},\mathbf{x}_0}(\mathbf{x},\sigma\tilde{\mathbf{x}})\exp\left(\frac{1}{\sigma}\mathbf{x}_0^T\tilde{\mathbf{x}}\right)\big]}{\partial\tilde{\mathbf{x}}^{\mathbf{p}}}\bigg|_{\tilde{\mathbf{x}}=\mathbf{0}}\bigg] \\ &\quad \overset{(*)}{=} \frac{1}{\sqrt{\mathbf{p}!}}\exp\left(-\frac{1}{2\sigma^2}\|\mathbf{x}_0\|_2^2\right)\end{aligned}$$

$$\times \frac{\partial^{\mathbf{p}} \mathsf{J}\big[R_{\mathrm{SSNM},\mathbf{x}_0}(\mathbf{x}, \sigma\tilde{\mathbf{x}})\big] \exp\big(\frac{1}{\sigma}\mathbf{x}_0^T\tilde{\mathbf{x}}\big)}{\partial \tilde{\mathbf{x}}^{\mathbf{p}}}\bigg|_{\tilde{\mathbf{x}}=\mathbf{0}}$$

$$\stackrel{(11)}{=} \frac{1}{\sqrt{\mathbf{p}!}} \exp\bigg(-\frac{1}{2\sigma^2}\|\mathbf{x}_0\|_2^2\bigg) \times \frac{\partial^{\mathbf{p}} \big[\rho_{\mathrm{SSNM},\mathbf{x}_0}(\mathbf{y}, \sigma\tilde{\mathbf{x}}) \exp\big(\frac{1}{\sigma}\mathbf{x}_0^T\tilde{\mathbf{x}}\big)\big]}{\partial \tilde{\mathbf{x}}^{\mathbf{p}}}\bigg|_{\tilde{\mathbf{x}}=\mathbf{0}},$$

where the derivative-reproducing property of the RKHS $\mathcal{H}_{\mathrm{SSNM},\mathbf{x}_0}$ [58] has been used in (*). Inserting $\rho_{\mathrm{SSNM},\mathbf{x}_0}(\mathbf{y},\mathbf{x}) = \exp\big(-\frac{1}{2\sigma^2}\big[2\mathbf{y}^T(\mathbf{x}_0-\mathbf{x}) + \|\mathbf{x}\|_2^2 - \|\mathbf{x}_0\|_2^2\big]\big)$ (note that this is $\rho_{\mathrm{LGM},\mathbf{x}_0}(\mathbf{y},\mathbf{x})$ in (16) for $\mathbf{H} = \mathbf{I}$) and using (42), we obtain further

$$\mathsf{J}\bigg[\nu_{\mathbf{x}_0}(\mathbf{x})\, r^{(\mathbf{p})}\bigg(\frac{\mathbf{x}}{\sigma}\bigg)\bigg] = \frac{1}{\sqrt{\mathbf{p}!}} \frac{\partial^{\mathbf{p}} \psi_{\mathbf{x}_0}(\tilde{\mathbf{x}}, \mathbf{y})}{\partial \tilde{\mathbf{x}}^{\mathbf{p}}}\bigg|_{\tilde{\mathbf{x}}=\mathbf{0}}. \qquad (81)$$

Finally, inserting (81) into (80) yields (41).

References

[1] C. Carbonelli, S. Vedantam, and U. Mitra, "Sparse channel estimation with zero tap detection," *IEEE Trans. Wireless Comm.*, vol. 6, no. 5, pp. 1743–1763, May 2007.

[2] S. Mallat, *A Wavelet Tour of Signal Processing—The Sparse Way*, 3rd ed. San Diego, CA, USA: Academic, 2009.

[3] M. Dong and L. Tong, "Optimal design and placement of pilot symbols for channel estimation," *IEEE Trans. Signal Process.*, vol. 50, no. 12, pp. 3055–3069, Dec. 2002.

[4] D. L. Donoho and J. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[5] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[6] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[7] E. J. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.

[8] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[9] G. Raskutti, M. J. Wainwright, and B. Yu, "Minimax rates of estimation for high-dimensional linear regression over ℓ_q-balls," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6976–6994, Oct. 2011.

[10] N. Verzelen, "Minimax risks for sparse regressions: Ultra-high dimensional phenomenons," *Electron. J. Statist.*, vol. 6, pp. 38–90, Jan. 2012.

[11] D. L. Donoho and I. M. Johnstone, "Minimax risk over ℓ_p-balls for ℓ_q-error," *Probab. Theory Rel. Fields*, vol. 99, no. 2, pp. 277–303, 1994.

[12] Z. Ben-Haim and Y. C. Eldar, "The Cramér-Rao bound for estimating a sparse parameter vector," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3384–3389, Jun. 2010.

[13] Z. Ben-Haim and Y. C. Eldar, "Performance bounds for sparse estimation with random noise," in *Proc. IEEE/SP 15th Workshop Statist. Signal Process.*, Cardiff, Wales, Aug./Sep. 2009, pp. 225–228.

[14] A. Jung, Z. Ben-Haim, F. Hlawatsch, and Y. C. Eldar, "Unbiased estimation of a sparse vector in white Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 57, no. 12, pp. 7856–7876, Dec. 2011.

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

[16] E. Parzen, "Statistical inference on time series by Hilbert space methods, I," Appl. Math. Statist. Lab., Stanford Univ., Stanford, CA, USA, Tech. Rep. 23, Jan. 1959.

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

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

[19] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, 2nd ed. New York, NY, USA: Springer-Verlag, 1998.

[20] S. G. Krantz and H. R. Parks, *A Primer of Real Analytic Functions*, 2nd ed. Boston, MA, USA: Birkhäuser, 2002.

[21] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD, USA: Johns Hopkins Univ. Press, 1996.

[22] A. Jung, "An RKHS approach to estimation with sparsity constraints," Ph.D. dissertation, Inst. Telecommun., Vienna Univ. Technol., Wien, Austria, 2011. [Online]. Available: http://arxiv.org/pdf/1311.5768.pdf

[23] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[24] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *Proc. Nat. Acad. Sci.*, vol. 100, no. 5, pp. 2197–2202, Mar. 2003.

[25] H. Leeb and B. M. Pötscher, "Sparse estimators and the oracle property, or the return of Hodges' estimator," *J. Econometrics*, vol. 142, no. 1, pp. 201–211, 2008.

[26] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York, NY, USA: Springer-Verlag, 1988.

[27] L. L. Scharf, *Statistical Signal Processing*. Reading, MA, USA: Addison-Wesley, 1991.

[28] A. O. Hero, J. A. Fessler, and M. Usman, "Exploring estimator bias-variance tradeoffs using the uniform CR bound," *IEEE Trans. Signal Process.*, vol. 44, no. 8, pp. 2026–2041, Aug. 1996.

[29] A. Jung, S. Schmutzhard, and F. Hlawatsch, "The RKHS approach to minimum variance estimation revisited: Variance bounds, sufficient statistics, and exponential families," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4050–4065, Jul. 2014.

[30] Y. C. Eldar, *Rethinking Biased Estimation: Improving Maximum Likelihood and the Cramér–Rao Bound* (Foundations and Trends in Signal Processing), vol. 1. Hanover, MA, USA: now, 2007.

[31] W. Rudin, *Real and Complex Analysis*, 3rd ed. New York, NY, USA: McGraw-Hill, 1987.

[32] L. D. Brown, *Fundamentals of Statistical Exponential Families*. Hayward, CA, USA: Institute of Mathematical Statistics, 1986.

[33] 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, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2010, pp. 1976–1980.

[34] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inf. Theory*, vol. 50, no. 3, pp. 1030–1051, Mar. 2004.

[35] Z. Ben-Haim, Y. C. Eldar, and M. Elad, "Coherence-based performance guarantees for estimating a sparse vector under random noise," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5030–5043, Oct. 2010.

[36] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.

[37] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[38] E. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *Ann. Statist.*, vol. 35, no. 6, pp. 2313–2351, 2007.

[39] D. L. Donoho and I. M. Johnstone, "Minimax estimation via wavelet shrinkage," *Ann. Statist.*, vol. 26, no. 3, pp. 879–921, 1998.

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

[41] G. Szegö, *Orthogonal Polynomials*. Providence, RI, USA: AMS, 1939.

[42] 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.

[43] M. J. Wainwright, "Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5728–5741, Dec. 2009.

[44] M. J. Wainwright, "Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso)," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2183–2202, May 2009.

[45] P. Maechler *et al.*, "VLSI design of approximate message passing for signal restoration and compressive sensing," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 2, no. 3, pp. 579–590, Sep. 2012.

[46] B. Babadi, N. Kalouptsidis, and V. Tarokh, "Asymptotic achievability of the Cramér–Rao bound for noisy compressive sampling," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1233–1236, Mar. 2009.

[47] A. W. van der Vaart, *Asymptotic Statistics*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[48] P. J. Huber, *Robust Statistics*. New York, NY, USA: Wiley, 1981.

[49] Y. C. Eldar, P. Kuppinger, and H. Bölcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.

[50] M. Mishali and Y. C. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4692–4702, Oct. 2008.
[51] Y. C. Eldar and H. Rauhut, "Average case analysis of multichannel sparse recovery using convex relaxation," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 505–519, Jan. 2009.
[52] D. Eiwen, G. Tauböck, F. Hlawatsch, and H. G. Feichtinger, "Group sparsity methods for compressive channel estimation in doubly dispersive multicarrier systems," in *Proc. IEEE SPAWC*, Marrakech, Morocco, Jun. 2010, pp. 1–5.
[53] A. Jung, G. Tauböck, and F. Hlawatsch, "Compressive spectral estimation for nonstationary random processes," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 3117–3138, May 2013.
[54] Z. Tian, "Compressed wideband sensing in cooperative cognitive radio networks," in *Proc. IEEE GLOBECOM*, New Orleans, LA, USA, Dec. 2008, pp. 1–5.
[55] Y. L. Polo, Y. Wang, A. Pandharipande, and G. Leus, "Compressive wide-band spectrum sensing," in *Proc. IEEE ICASSP*, Taipei, Taiwan, Apr. 2009, pp. 2337–2340.
[56] Z. Tian, Y. Tafesse, and B. M. Sadler, "Cyclic feature detection with sub-Nyquist sampling for wideband spectrum sensing," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 1, pp. 58–69, Feb. 2012.
[57] M. Mishali and Y. C. Eldar, "From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 375–390, Apr. 2010.
[58] D.-X. Zhou, "Derivative reproducing properties for kernel methods in learning theory," *J. Comput. Appl. Math.*, vol. 220, nos. 1–2, pp. 456–463, Oct. 2008.
[59] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ, USA: Princeton Univ. Press, 1970.

Alexander Jung (S'11–M'14) received the Diplom-Ingenieur and Dr. techn. degrees in electrical engineering/signal processing from Vienna University of Technology, Vienna, Austria, in 2008, and 2011, respectively. Since 2008, he has been a Research Assistant with the Institute of Telecommunications, Vienna University of Technology. His research interests are in statistical signal processing with emphasis on sparse estimation problems. He received several national awards and a Best Student Paper Award at IEEE ICASSP 2011.

Sebastian Schmutzhard received the Diplom-Ingenieur degree in technical mathematics from the University of Innsbruck, Austria in 2008, and the Dr. rer. nat. degree in mathematics from the University of Vienna, Austria in 2012. His research interests include numerical analysis, special functions of mathematical physics, and signal processing. He is coauthor of a paper that won a Best Student Paper Award at IEEE ICASSP 2011.

Franz Hlawatsch (S'85–M'88–SM'00–F'12) received the Diplom-Ingenieur, Dr. techn., and Univ.-Dozent (habilitation) degrees in electrical engineering/signal processing from Vienna University of Technology, Vienna, Austria in 1983, 1988, and 1996, respectively. Since 1983, he has been with the Institute of Telecommunications, Vienna University of Technology, where he is currently an Associate Professor. During 1991–1992, as a recipient of an Erwin Schrödinger Fellowship, he spent a sabbatical year with the Department of Electrical Engineering, University of Rhode Island, Kingston, RI, USA. In 1999, 2000, and 2001, he held one-month Visiting Professor positions with INP/ENSEEIHT, Toulouse, France and IRCCyN, Nantes, France. He (co)authored a book, three review papers that appeared in the IEEE SIGNAL PROCESSING MAGAZINE, about 200 refereed scientific papers and book chapters, and three patents. He coedited three books. His research interests include signal processing for sensor networks and wireless communications, statistical signal processing, and compressive signal processing.

Prof. Hlawatsch was Technical Program Co-Chair of EUSIPCO 2004 and served on the technical committees of numerous IEEE conferences. He was an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2003 to 2007 and for the IEEE TRANSACTIONS ON INFORMATION THEORY from 2008 to 2011. From 2004 to 2009, he was a member of the IEEE SPCOM Technical Committee. He is coauthor of papers that won an IEEE Signal Processing Society Young Author Best Paper Award and a Best Student Paper Award at IEEE ICASSP 2011.

Zvika Ben-Haim received the B.Sc. degree in electrical engineering and the B.A. degree in physics in 2000, the M.Sc. degree in electrical engineering in 2005, and the Ph.D. degree in electrical engineering in 2010, all from the Technion - Israel Institute of Technology, Haifa, Israel. He has been with the Google Israel R&D Center since 2011.

Yonina C. Eldar (S'98–M'02–SM'07–F'12) received the B.Sc. degree in physics and the B.Sc. degree in electrical engineering both from Tel-Aviv University (TAU), Tel-Aviv, Israel, in 1995 and 1996, respectively, and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 2002.

From January 2002 to July 2002, she was a Postdoctoral Fellow at the Digital Signal Processing Group at MIT. She is currently a Professor in the Department of Electrical Engineering at the Technion - Israel Institute of Technology, Haifa and holds the The Edwards Chair in Engineering. She is also a Research Affiliate with the Research Laboratory of Electronics at MIT and a Visiting Professor at Stanford University, Stanford, CA. Her research interests are in the broad areas of statistical signal processing, sampling theory and compressed sensing, optimization methods, and their applications to biology and optics.

Dr. Eldar was a Horev Fellow of the Leaders in Science and Technology program at the Technion from 2002-2005 and an Alon Fellow. In 2004, she was awarded the Wolf Foundation Krill Prize for Excellence in Scientific Research, in 2005 the Andre and Bella Meyer Lectureship, in 2007 the Henry Taub Prize for Excellence in Research, in 2008 the Hershel Rich Innovation Award, the Award for Women with Distinguished Contributions, the Muriel & David Jacknow Award for Excellence in Teaching, and the Technion Outstanding Lecture Award, in 2009 the Technion's Award for Excellence in Teaching, in 2010 the Michael Bruno Memorial Award from the Rothschild Foundation, and in 2011 the Weizmann Prize for Exact Sciences. In 2012 she was elected to the Young Israel Academy of Science and to the Israel Committee for Higher Education. In 2013 she received the Technion's Award for Excellence in Teaching, the Hershel Rich Innovation Award, and the IEEE Signal Processing Technical Achievement Award, and in 2014 the IEEE/AESS Fred Nathanson Memorial Radar Award. She received several best paper awards together with her research students and colleagues. She is the Editor in Chief of *Foundations and Trends in Signal Processing* and a member of the IEEE Sensor Array and Multichannel Technical Committee. In the past, she was a Signal Processing Society Distinguished Lecturer, member of the IEEE Signal Processing Theory and Methods and Bio Imaging Signal Processing technical committees, and served as an associate editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the *EURASIP Journal of Signal Processing*, the *SIAM Journal on Matrix Analysis and Applications*, and the *SIAM Journal on Imaging Sciences*.