

Unbiased Estimation of a Sparse Vector in White Gaussian Noise

Alexander Jung^a, Zvika Ben-Haim^b, Franz Hlawatsch^a, and Yonina C. Eldar^b

^aInstitute of Communications and Radio-Frequency Engineering, Vienna University of Technology
Gusshausstrasse 25/389, 1040 Vienna, Austria

Phone: +43 1 58801 38963, Fax: +43 1 58801 38999, E-mail: {ajung, fhlawats}@nt.tuwien.ac.at

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

Abstract

We consider unbiased estimation of a sparse nonrandom vector corrupted by additive white Gaussian noise. We show that while there are infinitely many unbiased estimators for this problem, none of them has uniformly minimum variance. Therefore, we focus on locally minimum variance unbiased (LMVU) estimators. We derive simple closed-form lower and upper bounds on the variance of LMVU estimators or, equivalently, on the Barankin bound (BB). Our bounds allow an estimation of the threshold region separating the low-SNR and high-SNR regimes, and they indicate the asymptotic behavior of the BB at high SNR. We also develop numerical lower and upper bounds which are tighter than the closed-form bounds and thus characterize the BB more accurately. Numerical studies compare our characterization of the BB with established biased estimation schemes, and demonstrate that while unbiased estimators perform poorly at low SNR, they may perform better than biased estimators at high SNR. An interesting conclusion of our analysis is that the high-SNR behavior of the BB depends solely on the value of the smallest nonzero component of the sparse vector, and that this type of dependence is also exhibited by the performance of certain practical estimators.

Index Terms

Sparsity, unbiased estimation, denoising, Cramér–Rao bound, Barankin bound, Hammersley–Chapman–Robbins bound, locally minimum variance unbiased estimator.

This work was supported by the FWF under Grant S10603-N13 (Statistical Inference) within the National Research Network SISE, by the WWTF under Grant MA 07-004 (SPORTS), by the Israel Science Foundation under Grant 1081/07, and by the European Commission under the FP7 Network of Excellence in Wireless Communications NEWCOM++ (contract no. 216715). Parts of this work were previously presented at ICASSP 2010, Dallas, TX, March 2010.

I. INTRODUCTION

Research in the past few years has led to a recognition that the performance of signal processing algorithms can be boosted by exploiting the tendency of many signals to have sparse representations. Applications of this principle include signal reconstruction (e.g. in the context of compressed sensing [1, 2]) and signal enhancement (e.g. in the context of image denoising and deblurring [3–5]).

In this work, we consider the estimation of an S -sparse, finite-dimensional vector $\mathbf{x} \in \mathbb{R}^N$. By “ S -sparse” we mean that the vector \mathbf{x} has at most S nonzero entries, which is denoted by $\|\mathbf{x}\|_0 \triangleq |\text{supp}(\mathbf{x})| \leq S$, where $\text{supp}(\mathbf{x})$ denotes the set of indices of the nonzero entries of \mathbf{x} . The “sparsity” S is assumed to be known, and typically $S \ll N$. However, the positions of the nonzero entries (i.e., $\text{supp}(\mathbf{x})$) as well as the values of the nonzero entries are unknown. We investigate how much we can gain in estimation accuracy by knowing *a priori* that the vector \mathbf{x} is S -sparse. We will use the frequentist setting [6] of estimation theory, i.e., we will model \mathbf{x} as unknown but deterministic. This is in contrast to Bayesian estimation theory, where one treats \mathbf{x} as a random vector whose probability density function (pdf) or certain moments thereof are assumed to be known. In the Bayesian setting, the sparsity can be modeled by using a pdf that favors sparse vectors, see e.g. [7–9].

A fundamental concept in the frequentist setting is that of unbiasedness [6, 10, 11]. An unbiased estimator is one whose expectation always equals the true underlying vector \mathbf{x} . The restriction to unbiased estimators is important as it excludes trivial and practically useless estimators, and it allows us to study the difficulty of the estimation problem using established techniques such as the Cramér–Rao bound (CRB) [10–12]. Another justification of unbiasedness is that for typical estimation problems, when the variance of the noise is low, it is necessary for an estimator to be unbiased in order to achieve a small mean-squared estimation error (MSE) [6].

These reasons notwithstanding, there is no guarantee that unbiased estimators are necessarily optimal. In fact, in many settings, including the scenario described in this paper, there exist biased estimators which are strictly better than any unbiased technique in terms of MSE [13–15]. Nevertheless, for simplicity and because of the reasons stated above, we focus on bounds for unbiased estimation in this work. As we will see, bounds on unbiased techniques give some indication of the general difficulty of the setting, and as such some of our conclusions will be shown empirically to characterize biased techniques as well.

Our main contribution is a characterization of the optimal performance of unbiased estimators $\hat{\mathbf{x}}(\mathbf{y})$ that are based on observing

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

where $\mathbf{A} \in \mathbb{R}^{M \times N}$ ($M \geq N$) is a known matrix with orthonormal columns, i.e., $\mathbf{A}^T \mathbf{A} = \mathbf{I}_N$, and $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_M)$ denotes zero-mean white Gaussian noise with known variance σ^2 (here, \mathbf{I}_N denotes the identity matrix of size $N \times N$). Note that without loss of generality we can then assume that $\mathbf{A} = \mathbf{I}_N$ and $M = N$, i.e., $\mathbf{y} = \mathbf{x} + \mathbf{n}$, since premultiplication of the model (1) by \mathbf{A}^T will reduce the estimation problem to an equivalent problem $\mathbf{y}' = \mathbf{A}' \mathbf{x} + \mathbf{n}'$ in which $\mathbf{A}' = \mathbf{A}^T \mathbf{A} = \mathbf{I}_N$ and the noise $\mathbf{n}' = \mathbf{A}^T \mathbf{n}$ is again zero-mean white Gaussian with variance σ^2 . Such a sparse signal model can be used, e.g., for channel estimation [16] when the channel consists only of few significant taps and an orthogonal training signal is used [17]. Another application that fits our scope is image denoising using an orthonormal wavelet basis [3]. We note that parts of this work were previously presented in [18].

The estimation problem (1) with $\mathbf{A} = \mathbf{I}_N$ was studied by Donoho and Johnstone [19, 20]. Their work was aimed at demonstrating asymptotic minimax optimality, i.e., they considered estimators having optimal worst-case behavior when the problem dimensions N, S tend to infinity. By contrast, we consider the finite-dimensional setting, and attempt to characterize the performance at each value of \mathbf{x} , rather than analyzing worst-case behavior. Such a ‘‘pointwise’’ approach was also advocated by the authors of [21, 22], who studied the CRB for the sparse linear model (1) with arbitrary \mathbf{A} . However, the CRB is a local bound, in the sense that the performance characterization it provides is only based on the statistical properties in the neighborhood of the specific value of \mathbf{x} being examined. In particular, the CRB for a given \mathbf{x} is only based on a *local* unbiasedness assumption, meaning that the estimator is only required to be unbiased at \mathbf{x} and in its infinitesimal neighborhood. Our goal in this paper is to obtain performance bounds for the more restrictive case of globally unbiased estimators, i.e., estimators whose expectation equals the true \mathbf{x} for each S -sparse vector \mathbf{x} . Since any globally unbiased estimator is also locally unbiased, our lower bounds will be tighter than those of [21, 22].

Our contributions and the organization of this paper can be summarized as follows. In Section II, we show that whereas only one unbiased estimator exists for the ordinary (nonsparse) signal in noise model, there are infinitely many unbiased estimators for the sparse signal in noise model; on the other hand, none of them has uniformly minimum variance. In Sections III and IV, we characterize the performance of *locally* minimum variance unbiased estimators by providing, respectively, lower and upper bounds on their mean-squared error (MSE). These bounds can equivalently be viewed as lower and upper bounds on the *Barankin bound* [23, 24]. Finally, numerical studies exploring and extending our performance bounds and comparing them with established estimator designs are presented in Section V.

Notation: Throughout the paper, boldface lowercase letters (e.g., \mathbf{x}) denote column vectors while boldface uppercase letters (e.g., \mathbf{M}) denote matrices. We denote by $\text{tr}(\mathbf{M})$, \mathbf{M}^T , and \mathbf{M}^\dagger the trace, the trace, and the conjugate transpose, respectively.

transpose, and Moore-Penrose pseudoinverse of \mathbf{M} , respectively. The identity matrix of size $N \times N$ is denoted by \mathbf{I}_N . The notation $\mathbf{M} \succeq \mathbf{N}$ indicates that $\mathbf{M} - \mathbf{N}$ is a positive semidefinite matrix. The set of indices of the nonzero entries of a vector \mathbf{x} is denoted by $\text{supp}(\mathbf{x})$, and $\|\mathbf{x}\|_0$ is defined as the size of this set. The k th entry of \mathbf{x} is written x_k . We also use the signum function of a real number y , $\text{sgn}(y) \triangleq y/|y|$. The sets of nonnegative, nonpositive, and positive real numbers will be denoted by \mathbb{R}_+ , \mathbb{R}_- , and \mathbb{R}_{++} , respectively.

II. THE SPARSE SIGNAL IN NOISE MODEL

A. Problem Setting

Let $\mathbf{x} \in \mathbb{R}^N$ be an unknown deterministic vector which is known to be S -sparse, i.e.,

$$\mathbf{x} \in \mathcal{X}_S, \quad \text{with } \mathcal{X}_S \triangleq \{\mathbf{x} \in \mathbb{R}^N : \|\mathbf{x}\|_0 \leq S\}.$$

The vector \mathbf{x} is to be estimated based on the observation of a vector \mathbf{y} which is the sum of \mathbf{x} and zero-mean white Gaussian noise. Thus

$$\mathbf{y} = \mathbf{x} + \mathbf{n}, \quad \text{with } \mathbf{x} \in \mathcal{X}_S, \quad \mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_N) \quad (2)$$

where the noise variance σ^2 is assumed to be nonzero and known. It follows that the pdf of \mathbf{y} , parameterized by the deterministic but unknown parameter $\mathbf{x} \in \mathcal{X}_S$, is

$$f(\mathbf{y}; \mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{x}\|_2^2\right). \quad (3)$$

We refer to (2) as the *sparse signal in noise model* (SSNM). As explained previously, settings of the form (1) with an orthonormal matrix \mathbf{A} can be converted to the SSNM (2). The case $S=N$ corresponds to the situation in which no sparsity assumption is made. As we will see, this case is fundamentally different from the sparse setting $S < N$, which is our focus in this paper.

An *estimator* $\hat{\mathbf{x}}(\mathbf{y})$ of the parameter \mathbf{x} is a function that maps (a realization of) the observation \mathbf{y} to (a realization of) the estimated vector $\hat{\mathbf{x}}$, i.e.,

$$\hat{\mathbf{x}}(\cdot) : \mathbb{R}^N \rightarrow \mathbb{R}^N : \mathbf{y} \mapsto \hat{\mathbf{x}}.$$

With an abuse of notation, we will use the symbol $\hat{\mathbf{x}}$ for both the estimator (which is a function) and the estimate (a specific function value). The meaning should be clear from the context. The question now is how we can exploit the information that \mathbf{x} is S -sparse in order to construct “good” estimators. Our

measure of the quality of an estimator $\hat{\mathbf{x}}(\cdot)$ for a given parameter value $\mathbf{x} \in \mathcal{X}_S$ will be the estimator's MSE, which is defined as

$$\varepsilon(\mathbf{x}; \hat{\mathbf{x}}) \triangleq \mathbb{E}_{\mathbf{x}} \{ \|\hat{\mathbf{x}}(\mathbf{y}) - \mathbf{x}\|_2^2 \}.$$

Here, the notation $\mathbb{E}_{\mathbf{x}}\{\cdot\}$ means that the expectation is taken with respect to the pdf $f(\mathbf{y}; \mathbf{x})$ of the observation \mathbf{y} parameterized by \mathbf{x} . Note that even though \mathbf{x} is known to be S -sparse, the estimates $\hat{\mathbf{x}}$ are not constrained to be S -sparse.

The MSE can be written as the sum of a bias term and a variance term, i.e.,

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

where the bias $\mathbf{b}(\mathbf{x}; \hat{\mathbf{x}}) \triangleq \mathbb{E}_{\mathbf{x}}\{\hat{\mathbf{x}}(\mathbf{y})\} - \mathbf{x}$ accounts for systematic estimation errors and the variance $V(\mathbf{x}; \hat{\mathbf{x}}) \triangleq \mathbb{E}_{\mathbf{x}}\{\|\hat{\mathbf{x}}(\mathbf{y}) - \mathbb{E}_{\mathbf{x}}\{\hat{\mathbf{x}}(\mathbf{y})\}\|_2^2\}$ accounts for errors due to random fluctuations of the estimate. Thus, for unbiased estimators ($\mathbf{b}(\mathbf{x}; \hat{\mathbf{x}}) = \mathbf{0}$ for all $\mathbf{x} \in \mathcal{X}_S$), the MSE is equal to the variance, i.e., $\varepsilon(\mathbf{x}; \hat{\mathbf{x}}) = V(\mathbf{x}; \hat{\mathbf{x}})$.

We will also consider the mean power (second moment) of an estimator,

$$P(\mathbf{x}; \hat{\mathbf{x}}) \triangleq \mathbb{E}_{\mathbf{x}}\{\|\hat{\mathbf{x}}(\mathbf{y})\|_2^2\} = V(\mathbf{x}; \hat{\mathbf{x}}) + \|\mathbb{E}_{\mathbf{x}}\{\hat{\mathbf{x}}(\mathbf{y})\}\|_2^2. \quad (4)$$

For unbiased estimators, $\|\mathbb{E}_{\mathbf{x}}\{\hat{\mathbf{x}}(\mathbf{y})\}\|_2^2 = \|\mathbf{x}\|_2^2$; thus, minimizing the variance $V(\mathbf{x}; \hat{\mathbf{x}})$ at a fixed $\mathbf{x} \in \mathcal{X}_S$ among all unbiased estimators is equivalent to minimizing $P(\mathbf{x}; \hat{\mathbf{x}})$.

B. Estimator Design

Two well-established estimator designs are the least squares (LS) estimator defined by

$$\hat{\mathbf{x}}_{\text{LS}}(\mathbf{y}) \triangleq \arg \min_{\mathbf{x}' \in \mathcal{X}_S} \|\mathbf{y} - \mathbf{x}'\|_2^2 \quad (5)$$

and the maximum likelihood (ML) estimator defined by

$$\hat{\mathbf{x}}_{\text{ML}}(\mathbf{y}) \triangleq \arg \max_{\mathbf{x}' \in \mathcal{X}_S} f(\mathbf{y}; \mathbf{x}'). \quad (6)$$

For the SSNM, due to (3), the LS and ML estimators coincide; they are easily seen to be given by

$$\hat{\mathbf{x}}_{\text{LS}}(\mathbf{y}) = \hat{\mathbf{x}}_{\text{ML}}(\mathbf{y}) = \mathbf{P}_S(\mathbf{y}) \quad (7)$$

where \mathbf{P}_S is an operator that retains the S largest (in magnitude) components and zeros out all others. The LS/ML estimator is biased unless $S = N$. Note that this estimator is not based on a direct minimization

of the MSE. Indeed, if the sparsity constraint is removed ($S=N$) and $N \geq 3$, it has been shown [13–15] that there exist estimators which yield a better MSE performance than that of the LS/ML estimator.

The MSE $\varepsilon(\mathbf{x}; \hat{\mathbf{x}})$ of a specific estimator $\hat{\mathbf{x}}(\cdot)$ depends on the value of the parameter \mathbf{x} . This makes it difficult to define optimality in terms of minimum MSE. For example, if an estimator $\hat{\mathbf{x}}(\cdot)$ performs well (i.e., has a small MSE) for a specific parameter value \mathbf{x}_1 , it may still exhibit poor performance (i.e., a large MSE) for a different parameter value \mathbf{x}_2 . Ideally, an optimal estimator should have minimum MSE for all parameter values *simultaneously*. However, such an optimality criterion is unobtainable since the minimum MSE achievable at a specific parameter value \mathbf{x}_1 is zero; it is achieved by the trivial estimator $\hat{\mathbf{x}}(\mathbf{y}) \equiv \mathbf{x}_1$ which is constant and completely ignores the observation \mathbf{y} . Therefore, if there were a *uniformly minimum MSE* estimator, it would have to achieve zero MSE for all parameter values, which is obviously impossible. Thus, requiring the estimator to minimize the MSE at all parameter values simultaneously makes no sense.

One useful optimality criterion is the minimax approach, which considers the worst-case MSE

$$\sup_{\mathbf{x} \in \mathcal{X}_S} \varepsilon(\mathbf{x}; \hat{\mathbf{x}})$$

of an estimator $\hat{\mathbf{x}}(\cdot)$. An optimal estimator in the minimax sense minimizes the worst-case MSE, i.e., is a solution of the optimization problem

$$\inf_{\hat{\mathbf{x}}(\cdot)} \sup_{\mathbf{x} \in \mathcal{X}_S} \varepsilon(\mathbf{x}; \hat{\mathbf{x}}).$$

Considerable effort has been spent to identify minimax estimators for sparse models such as the SSNM in (2); see, e.g., [19, 20, 25]. However, these results only apply in the asymptotic regime, i.e., when $N, S \rightarrow \infty$. By contrast, our goal is to analyze estimator performance for finite problem dimensions. There are no known closed-form expressions of the minimax risk or of minimax-optimal estimators for the SSNM in this case.

In this work, rather than pursuing the minimax criterion, we consider *unbiased* estimators $\hat{\mathbf{x}}(\cdot)$ for the SSNM. An unbiased estimator is one for which the bias $\mathbf{b}(\mathbf{x}; \hat{\mathbf{x}})$ is zero for all S -sparse parameter vectors i.e.,

$$\mathbb{E}_{\mathbf{x}}\{\hat{\mathbf{x}}(\mathbf{y})\} = \mathbf{x} \quad \text{for all } \mathbf{x} \in \mathcal{X}_S. \quad (8)$$

Let \mathcal{U} denote the set of all unbiased estimators $\hat{\mathbf{x}}(\cdot)$ for the SSNM. Constraining an estimator to be unbiased excludes such trivial estimators as $\hat{\mathbf{x}}(\mathbf{y}) \equiv \mathbf{x}_1$ where $\mathbf{x}_1 \in \mathcal{X}_S$ is some fixed S -sparse parameter vector.

C. Unbiased Estimation for the SSNM

We now study the set \mathcal{U} of unbiased estimators for the SSNM in more detail. In particular, we will show that with the exception of the case $S = N$, this set is uncountably large, i.e., there are infinitely many unbiased estimators. We will also show that there exists no uniformly minimum variance unbiased estimator unless $S = N$. In what follows, we will say that an estimator $\hat{\mathbf{x}}$ has a bounded MSE if $\varepsilon(\mathbf{x}; \hat{\mathbf{x}}) \leq C$ for all $\mathbf{x} \in \mathbb{R}^N$, where C is a constant which may depend on N , S , and σ^2 .

Theorem 1. *Consider the SSNM (2) with $S = N$, i.e., without a sparsity constraint, in which case $\mathcal{X}_S = \mathbb{R}^N$. Then, there exists exactly one unbiased estimator having bounded MSE (up to deviations having zero measure). This estimator is given by $\hat{\mathbf{x}}(\mathbf{y}) = \mathbf{y}$, which equals the LS/ML estimator in (5)–(7).*

The proof of this result can be found in Appendix A. By contrast with Theorem 1, when sparsity constraints are imposed there exists a large family of unbiased estimators, as we now show.

Theorem 2. *For $1 \leq S < N$, there are uncountably infinitely many unbiased estimators for the SSNM.*

Proof. Consider the class of estimators defined by

$$\hat{x}(\mathbf{y}) = \mathbf{y} + ay_1 \left[\prod_{k=2}^{S+1} h^{(c,d)}(y_k) \right] (1 \ 0 \ \dots \ 0)^T, \quad a \in \mathbb{R}, \ c, d \in \mathbb{R}_{++} \quad (9)$$

where

$$h^{(c,d)}(y) \triangleq \begin{cases} \text{sgn}(y), & |y| \in [c, c+d] \\ 0, & \text{else.} \end{cases} \quad (10)$$

A straightforward calculation shows that each estimator of this uncountably infinite class is an unbiased estimator for the SSNM. \square

This (constructive) proof points at a noteworthy fact. Consider a particular parameter value \mathbf{x} . By an appropriate choice of the parameters a, c, d in (9), one can reduce the magnitude of the estimate $\hat{\mathbf{x}}(\mathbf{y})$ for sets of realizations \mathbf{y} with high probability, i.e., for which $f(\mathbf{y}; \mathbf{x})$ is large. This results in a reduced mean power and (since the estimator is unbiased) in a reduced variance and MSE at the specific parameter value \mathbf{x} . One can thus construct an unbiased estimator that performs better than the (biased) LS/ML estimator at the given \mathbf{x} .

In view of Theorems 1 and 2, we will only consider the case $S < N$ in the following. Since in this case there are infinitely many unbiased estimators, we would like to find an unbiased estimator

having minimum variance (and, thus, minimum MSE) among all unbiased estimators. If there exists an unbiased estimator $\hat{\mathbf{x}}(\cdot) \in \mathcal{U}$ which minimizes the variance *simultaneously* for all S -sparse parameter vectors $\mathbf{x} \in \mathcal{X}_S$, then this estimator is called a *uniformly minimum variance unbiased* (UMVU) estimator [6]. In other words, a UMVU estimator for the SSNM solves the optimization problem

$$\arg \min_{\hat{\mathbf{x}}(\cdot) \in \mathcal{U}} V(\mathbf{x}; \hat{\mathbf{x}}) \quad (11)$$

simultaneously for all $\mathbf{x} \in \mathcal{X}_S$. In the nonsparse case $S = N$, it is well known that the LS estimator is the UMVU estimator [10]; however, in light of Theorem 1, this is not a very strong result, since $\hat{\mathbf{x}}_{\text{LS}}$ is the *only* unbiased estimator in that case. On the other hand, for the sparse case $S < N$, the following negative result is shown in Appendix B.

Theorem 3. *For the SSNM with $S < N$, there exists no UMVU estimator, i.e., there is no unbiased estimator $\hat{\mathbf{x}} \in \mathcal{U}$ that minimizes $V(\mathbf{x}; \hat{\mathbf{x}})$ simultaneously for all parameter vectors $\mathbf{x} \in \mathcal{X}_S$.*

Despite the fact that a UMVU estimator does not exist for the SSNM, one can still attempt to solve the optimization problem (11) separately for each value of $\mathbf{x} \in \mathcal{X}_S$. An unbiased estimator which solves (11) for a specific value of \mathbf{x} is said to be *locally minimum variance unbiased* (LMVU) [6]. The MSE of this estimator at \mathbf{x} cannot be improved upon by any unbiased estimator. When viewed as a function of \mathbf{x} , this minimum MSE is known as the *Barankin bound* (BB) [23, 24]. Thus, the BB characterizes the minimum MSE achievable by any unbiased estimator for each value of $\mathbf{x} \in \mathcal{X}_S$; it is the highest and tightest lower bound on the MSE of unbiased estimators. As such, the BB serves as a measure of the difficulty of estimating \mathbf{x} .

Computing the BB is equivalent to calculating $\min_{\hat{\mathbf{x}}(\cdot) \in \mathcal{U}} V(\mathbf{x}; \hat{\mathbf{x}})$ for each parameter vector $\mathbf{x} \in \mathcal{X}_S$ separately. Unfortunately, there does not appear to be a simple closed-form expression of the BB, and the numerical computation of the BB seems to be difficult as well. Therefore, in the remainder of this paper, we will provide lower and upper bounds on the BB. When these bounds are close to one another, they provide an accurate characterization of the BB.

III. LOWER BOUNDS ON THE MINIMUM MSE

In this section, we will develop a lower bound on the BB (which is thus a lower bound on the MSE of any unbiased estimator) by calculating a limiting case of the Hammersley–Chapman–Robbins bound [23] for the SSNM.

A. Review of the CRB

A variety of techniques exist for developing lower bounds on the MSE of unbiased estimators. The simplest is the CRB [11, 12, 26], which was previously derived for a more general sparse estimation setting in [21, 22]. In the current setting, i.e., for the SSNM (1), the CRB is given by

$$\varepsilon(\mathbf{x}; \hat{\mathbf{x}}) \geq \begin{cases} S\sigma^2, & \|\mathbf{x}\|_0 = S \\ N\sigma^2, & \|\mathbf{x}\|_0 < S \end{cases} \quad (12)$$

where $\hat{\mathbf{x}} \in \mathcal{U}$, i.e., $\hat{\mathbf{x}}(\cdot)$ is any unbiased estimator for the SSNM.

In the case of parameter values $\mathbf{x} \in \mathcal{X}_S$ with non-maximal support, i.e., $\|\mathbf{x}\|_0 < S$, the CRB is $N\sigma^2$. This is the MSE of the trivial unbiased estimator $\hat{\mathbf{x}}(\mathbf{y}) = \mathbf{y}$. Since the CRB is thus achieved by an unbiased estimator, we conclude that the CRB is a *maximally tight* lower bound for $\|\mathbf{x}\|_0 < S$; no other lower bound can be tighter (higher). We also conclude that for $\|\mathbf{x}\|_0 < S$, the trivial estimator $\hat{\mathbf{x}}(\mathbf{y}) = \mathbf{y}$ is the LMVU estimator; no other unbiased estimator can have a smaller MSE.

For parameter values $\mathbf{x} \in \mathcal{X}_S$ with maximal support, i.e., $\|\mathbf{x}\|_0 = S$, we will see that the CRB is not maximally tight, and the trivial estimator $\hat{\mathbf{x}}(\mathbf{y}) = \mathbf{y}$ is not the LMVU estimator. Indeed, one problem with the CRB in (12) is that it is discontinuous in the transition between $\|\mathbf{x}\|_0 = S$ and $\|\mathbf{x}\|_0 < S$. Since the MSE of any estimator is continuous [6], this discontinuity implies that the CRB is not the tightest lower bound obtainable for unbiased estimators. In order to obtain tighter bounds for $\|\mathbf{x}\|_0 = S$, it is important to realize that the CRB is a local bound, which assumes unbiasedness only in a neighborhood of \mathbf{x} . Since we are interested in estimators that are unbiased for all $\mathbf{x} \in \mathcal{X}_S$, which is a more restrictive constraint than local unbiasedness, tighter (i.e., higher) lower bounds can be expected for unbiased estimators in the case $\|\mathbf{x}\|_0 = S$.

B. Hammersley–Chapman–Robbins Bound

An alternative lower bound for unbiased estimators is the Hammersley–Chapman–Robbins bound (HCRB) [23, 27, 28], which can be stated, in our context, as follows.

Proposition 4. *Given a parameter value $\mathbf{x} \in \mathcal{X}_S$, consider a set of p “test points” $\{\mathbf{v}_i\}_{i=1}^p$ such that $\mathbf{x} + \mathbf{v}_i \in \mathcal{X}_S$ for all $i = 1, \dots, p$. Then, the covariance of any unbiased estimator $\hat{\mathbf{x}}(\cdot)$, $C(\mathbf{x}; \hat{\mathbf{x}}) \triangleq \mathbb{E}_{\mathbf{x}}\{[\hat{\mathbf{x}}(\mathbf{y}) - \mathbb{E}_{\mathbf{x}}\{\hat{\mathbf{x}}(\mathbf{y})\}][\hat{\mathbf{x}}(\mathbf{y}) - \mathbb{E}_{\mathbf{x}}\{\hat{\mathbf{x}}(\mathbf{y})\}]^T\}$, satisfies*

$$C(\mathbf{x}; \hat{\mathbf{x}}) \succeq \mathbf{V}\mathbf{J}^\dagger\mathbf{V}^T \quad (13)$$

where

$$\mathbf{V} \triangleq (\mathbf{v}_1 \cdots \mathbf{v}_p) \in \mathbb{R}^{N \times p} \quad (14)$$

and the (i, j) th entry of the matrix $\mathbf{J} \in \mathbb{R}^{p \times p}$ is given by

$$(\mathbf{J})_{i,j} \triangleq \exp\left(\frac{\mathbf{v}_i^T \mathbf{v}_j}{\sigma^2}\right) - 1. \quad (15)$$

In particular, the MSE of $\hat{\mathbf{x}}(\cdot)$ satisfies

$$\varepsilon(\mathbf{x}; \hat{\mathbf{x}}) \geq \text{tr}(\mathbf{V}\mathbf{J}^\dagger\mathbf{V}^T). \quad (16)$$

The proof of Proposition 4, which can be found in Appendix C, involves an application of the multivariate HCRB of Gorman and Hero [23] to the SSNM setting. Note that both the number of test points p and their values \mathbf{v}_i are arbitrary and can depend on \mathbf{x} . In general, including additional test points \mathbf{v}_i will result in a tighter HCRB [23]. Our goal in this section is to choose test points \mathbf{v}_i which result in a tight but analytically tractable bound.

Before attempting to derive a bound which is tighter than the CRB, we first observe that the CRB itself can be obtained as the limit of a sequence of HCRBs with appropriately chosen test points. Indeed, consider the specific test points given by¹

$$\{t\mathbf{e}_i\}_{i \in \text{supp}(\mathbf{x})}, \quad \|\mathbf{x}\|_0 = S \quad (17a)$$

$$\{t\mathbf{e}_i\}_{i \in \{1, \dots, N\}}, \quad \|\mathbf{x}\|_0 < S \quad (17b)$$

where $t > 0$ is a constant and \mathbf{e}_i represents the i th column of the $N \times N$ identity matrix. Note that $p = S$ in (17a) and $p = N$ in (17b). Each value of t yields a different set of test points and, via Proposition 4, a different lower bound on the MSE of unbiased estimators. We show in Appendix D that the CRB in (12) is the limit of a sequence of such bounds as $t \rightarrow 0$, and that it is tighter than any bound that can be obtained via Proposition 4 using the test points (17) for a fixed $t > 0$.

Can a set of test points different from (17) yield a lower bound that is tighter than the CRB? As discussed above, this is only possible for parameter values \mathbf{x} having maximal support, i.e., $\|\mathbf{x}\|_0 = S$, because for $\|\mathbf{x}\|_0 < S$ the CRB is already maximally tight. Therefore, let us consider a parameter \mathbf{x} with $\|\mathbf{x}\|_0 = S$. Suppose one of the entries within the support, x_j for some $j \in \text{supp}(\mathbf{x})$, has a small magnitude. Such a parameter \mathbf{x} just barely qualifies as having maximal support, so it makes sense to

¹Note that, with a slight abuse of notation, the index i of the test points is now allowed to take on non-sequential values from the set $\{1, \dots, N\}$.

adapt the optimal test points (17b) from the non-maximal support case. However, including a test point $t\mathbf{e}_i$ with $i \notin \text{supp}(\mathbf{x})$ is not allowed, since in this case $\mathbf{x} + t\mathbf{e}_i$ is not in \mathcal{X}_S . Instead, one could include the test point $\mathbf{v}_i = t\mathbf{e}_i - x_j\mathbf{e}_j$, which satisfies the requirement $\mathbf{x} + \mathbf{v}_i \in \mathcal{X}_S$ and is still close to $t\mathbf{e}_i$ if x_j is small. More generally, for any maximal-support parameter \mathbf{x} , we propose the set of N test points given by

$$\mathbf{v}_i = \begin{cases} t\mathbf{e}_i, & i \in \text{supp}(\mathbf{x}) \\ t\mathbf{e}_i - \xi\mathbf{e}_{(S)}, & i \notin \text{supp}(\mathbf{x}) \end{cases} \quad (18)$$

for $i = 1, \dots, N$. Here, ξ denotes the smallest (in magnitude) of the S nonzero components of \mathbf{x} and $\mathbf{e}_{(S)}$ denotes the corresponding unit vector. These test points \mathbf{v}_i satisfy the condition $\mathbf{x} + \mathbf{v}_i \in \mathcal{X}_S$. Note that the test points in (17a), which yield the CRB, are a subset of the test points in (18). It can be shown [23] that this implies that the bound induced by (18) will always be at least as tight as that obtained from (17a). It is important to note that (18) uses N test points for parameter values with maximal support, just as (17b) does for parameter values with non-maximal support. In fact, there is a smooth transition between the optimal test points (17b) for non-maximal support and the proposed test points (18) for maximal support.

While an expression of the HCRB can be obtained by simply plugging (18) into (16), the resulting bound is extremely cumbersome and not very insightful. Instead, in analogy to the derivation of the CRB above, one can obtain a simple result by taking the limit for $t \rightarrow 0$. This leads to the following theorem, which combines the cases of maximal support ((16) using (18) for $t \rightarrow 0$) and non-maximal support ((16) using (17b) for $t \rightarrow 0$), and whose proof can be found in Appendix E.

Theorem 5. *The MSE of any unbiased estimator $\hat{\mathbf{x}} \in \mathcal{U}$ for the SSNM satisfies*

$$\varepsilon(\mathbf{x}; \hat{\mathbf{x}}) \geq \text{HCRB}(\mathbf{x}) \triangleq \begin{cases} S\sigma^2 + (N-S-1)e^{-\xi^2/\sigma^2}\sigma^2, & \|\mathbf{x}\|_0 = S \\ N\sigma^2, & \|\mathbf{x}\|_0 < S, \end{cases} \quad (19)$$

where, in the case $\|\mathbf{x}\|_0 = S$, ξ is the smallest (in magnitude) of the S nonzero entries of \mathbf{x} .

For simplicity, we will continue to refer to (19) as an HCRB, even though it was obtained as a limit of HCRBs. Note that when $\|\mathbf{x}\|_0 < S$, the HCRB in (19) is identical to the CRB in (12), since in that case the CRB is maximally tight and cannot be improved. The HCRB also approaches the CRB when $\|\mathbf{x}\|_0 = S$ and all components of \mathbf{x} are much larger than σ : here $e^{-\xi^2/\sigma^2}$ is negligible and the respective bound in (19) converges to $S\sigma^2$, which is equal to the CRB in (12). This is due to the fact that the CRB

is achieved by the ML estimator asymptotically² as $\xi^2/\sigma^2 \rightarrow \infty$, and is therefore also maximally tight when $\xi \gg \sigma$. Furthermore, if we define the “worst-case component SNR” (briefly denoted as SNR) as ξ^2/σ^2 , then Theorem 5 hints that the convergence to the high-SNR limit is exponential in the SNR.

One of the motivations for improving the CRB (12) was that (12) is discontinuous in the transition between $\|\mathbf{x}\|_0 = S$ and $\|\mathbf{x}\|_0 < S$. While the HCRB (19) is still discontinuous in this transition, the discontinuity is much smaller than that of the CRB. Indeed, the transition from $\|\mathbf{x}\|_0 = S$ to $\|\mathbf{x}\|_0 < S$ corresponds to $\xi \rightarrow 0$, in which case the first bound in (19) tends to $(N-1)\sigma^2$, whereas the second bound, valid for $\|\mathbf{x}\|_0 < S$, is $N\sigma^2$; thus, the difference between the two bounds in (19) is σ^2 . By contrast, the difference between the two bounds in (12) is $(N-S)\sigma^2$, which is typically much larger. Again, the discontinuity of (19) implies that (19) is not the tightest lower bound obtainable for unbiased estimators. In Section V, we will demonstrate experimentally that this discontinuity can be eliminated altogether by using a much larger number of test points. However, in that case the resulting bound no longer has a simple closed-form expression and can only be evaluated numerically.

IV. UPPER BOUND ON THE MINIMUM MSE

As pointed out in the previous section, the lower bound $\text{HCRB}(\mathbf{x})$ on the BB is not maximally tight since it is discontinuous in the transition between parameter vectors with $\|\mathbf{x}\|_0 = S$ and those with $\|\mathbf{x}\|_0 < S$. In other words, there is a gap between the HCRB and the BB. How large is this gap? We will address this issue by deriving an *upper* bound on the BB. This will be done by finding a constrained solution of (11). If this upper bound is close to the lower bound $\text{HCRB}(\mathbf{x})$, we can conclude that both bounds are fairly tight and thus provide a fairly accurate characterization of the BB. As before, we consider the nontrivial case $\|\mathbf{x}\|_0 = S$.

We first note (cf. (4)) that (11) is equivalent to the optimization problem $\arg \min_{\hat{\mathbf{x}}(\cdot) \in \mathcal{U}} \mathbb{E}_{\mathbf{x}} \{ \|\hat{\mathbf{x}}(\mathbf{y})\|_2^2 \}$ = $\arg \min_{\hat{\mathbf{x}}(\cdot) \in \mathcal{U}} \sum_{k=1}^N \mathbb{E}_{\mathbf{x}} \{ (\hat{x}_k(\mathbf{y}))^2 \}$, where \hat{x}_k denotes the k th entry of $\hat{\mathbf{x}}$. This, in turn, is equivalent to the N individual scalar optimization problems

$$\arg \min_{\hat{x}_k(\cdot) \in \mathcal{U}^k} \mathbb{E}_{\mathbf{x}} \{ (\hat{x}_k(\mathbf{y}))^2 \}, \quad k = 1, \dots, N \quad (20)$$

²This can be explained by the fact that according to (7), the ML estimator for the SSNM retains the S largest components in \mathbf{y} and zeros out all other components. For noise variances σ^2 that are extremely small compared to the nonzero entries, i.e., for $\xi^2/\sigma^2 \rightarrow \infty$, the probability that the ML estimator selects the true components becomes very close to one. Therefore, for high ξ^2/σ^2 , the ML estimator behaves like an oracle estimator which knows the support of \mathbf{x} and whose MSE is equal to $S\sigma^2$.

where \mathcal{U}^k denotes the set of unbiased estimators of the k th entry of \mathbf{x} , i.e.,

$$\mathcal{U}^k \triangleq \{\hat{x}_k(\cdot) \mid \mathbb{E}_{\mathbf{x}}\{\hat{x}_k(\mathbf{y})\} = x_k \text{ for all } \mathbf{x} \in \mathcal{X}_S\}.$$

By combining the unbiased estimators $\hat{x}_k(\cdot)$ for $k = 1, \dots, N$ into a vector, we obtain an unbiased estimator of the parameter \mathbf{x} .

It will be convenient to write the k th scalar estimator as

$$\hat{x}_k(\mathbf{y}) = y_k + \hat{x}'_k(\mathbf{y}) \tag{21}$$

with $\hat{x}'_k(\mathbf{y}) \triangleq \hat{x}_k(\mathbf{y}) - y_k$. Since for any $\hat{x}_k(\cdot) \in \mathcal{U}^k$ we have $\mathbb{E}_{\mathbf{x}}\{\hat{x}_k(\mathbf{y})\} = \mathbb{E}_{\mathbf{x}}\{y_k\} + \mathbb{E}_{\mathbf{x}}\{\hat{x}'_k(\mathbf{y})\} = x_k + \mathbb{E}_{\mathbf{x}}\{\hat{x}'_k(\mathbf{y})\}$, the unbiasedness condition $\hat{x}_k(\cdot) \in \mathcal{U}^k$ is equivalent to

$$\mathbb{E}_{\mathbf{x}}\{\hat{x}'_k(\mathbf{y})\} = 0 \quad \text{for all } \mathbf{x} \in \mathcal{X}_S.$$

For $k \in \text{supp}(\mathbf{x})$, the solution of the optimization problem (20) is stated in the following lemma, which is proved in Appendix F. In what follows, it will be convenient to denote by $\hat{\mathbf{x}}^{(\mathbf{x})}(\mathbf{y})$ a solution of the optimization problem (11) for a given parameter vector $\mathbf{x} \in \mathcal{X}_S$. We recall that the estimator $\hat{\mathbf{x}}^{(\mathbf{x})}(\mathbf{y})$ is an LMVU at the parameter value \mathbf{x} , and its MSE, $\varepsilon(\mathbf{x}; \hat{\mathbf{x}}^{(\mathbf{x})}) = \min_{\hat{\mathbf{x}}(\cdot) \in \mathcal{U}} V(\mathbf{x}; \hat{\mathbf{x}})$, equals the BB at \mathbf{x} .

Lemma 6. *Consider a parameter vector $\mathbf{x} \in \mathcal{X}_S$ with maximal support, i.e., $\|\mathbf{x}\|_0 = S$. Then, for any $k \in \text{supp}(\mathbf{x})$, the solution of the optimization problem (20) is given by*

$$\hat{x}_k^{(\mathbf{x})}(\mathbf{y}) = y_k, \quad k \in \text{supp}(\mathbf{x}).$$

Moreover, this is the LMVU for $k \in \text{supp}(\mathbf{x})$. The MSE of this estimator is σ^2 .

Because Lemma 6 describes the scalar LMVU estimators for all indices $k \in \text{supp}(\mathbf{x})$, it remains to consider the scalar problem (20) for $k \notin \text{supp}(\mathbf{x})$. Since $\varepsilon(\mathbf{x}; \hat{\mathbf{x}}^{(\mathbf{x})})$ is the minimum of $\varepsilon(\mathbf{x}; \hat{\mathbf{x}})$ as defined by the optimization problem (11), we can obtain an upper bound on $\varepsilon(\mathbf{x}; \hat{\mathbf{x}}^{(\mathbf{x})})$ by placing further constraints on the estimator $\hat{\mathbf{x}}(\cdot)$ to be optimized. We will thus consider the modified optimization problem

$$\arg \min_{\hat{\mathbf{x}}(\cdot) \in \mathcal{U} \cap \mathcal{A}_{\mathbf{x}}} V(\mathbf{x}; \hat{\mathbf{x}}) \tag{22}$$

where the set $\mathcal{A}_{\mathbf{x}}$ is chosen such that a simpler problem is obtained. We will define $\mathcal{A}_{\mathbf{x}}$ in a componentwise fashion. More specifically, the k th component $\hat{x}_k(\mathbf{y})$ of $\hat{\mathbf{x}}(\mathbf{y})$, where $k \notin \text{supp}(\mathbf{x})$, is said to belong to the set $\mathcal{A}_{\mathbf{x}}^k$ if the correction term $\hat{x}'_k(\mathbf{y}) = \hat{x}_k(\mathbf{y}) - y_k$ (see (21)) satisfies the following two properties.

- *Odd symmetry* with respect to k and all indices in $\text{supp}(\mathbf{x})$:

$$\hat{x}'_k(\dots, -y_l, \dots) = -\hat{x}'_k(\dots, y_l, \dots), \quad \text{for all } l \in \{k\} \cup \text{supp}(\mathbf{x}). \quad (23)$$

- *Independence* with respect to all other indices:

$$\hat{x}'_k(\dots, y_l, \dots) = \hat{x}'_k(\dots, 0, \dots), \quad \text{for all } l \notin \{k\} \cup \text{supp}(\mathbf{x}). \quad (24)$$

We then define $\mathcal{A}_{\mathbf{x}}$ as the set of estimators $\hat{\mathbf{x}}(\mathbf{y})$ such that $\hat{x}_k(\mathbf{y}) \in \mathcal{A}_{\mathbf{x}}^k$ for all $k \notin \text{supp}(\mathbf{x})$. Note that any function $\hat{x}(\mathbf{y}) \in \mathcal{A}_{\mathbf{x}}^k$ is fully specified by its values for all arguments \mathbf{y} such that $\text{supp}(\mathbf{y}) = \{k\} \cup \text{supp}(\mathbf{x})$ and all entries of \mathbf{y} are nonnegative. The values of $\hat{x}(\mathbf{y})$ for all other \mathbf{y} follow by the decomposition (21) and the properties (23) and (24).

To solve the modified optimization problem (22), we consider the equivalent scalar form

$$\arg \min_{\hat{x}_k(\cdot) \in \mathcal{U}^k \cap \mathcal{A}_{\mathbf{x}}^k} \mathbb{E}_{\mathbf{x}} \{ (\hat{x}_k(\mathbf{y}))^2 \}, \quad k \notin \text{supp}(\mathbf{x}). \quad (25)$$

The resulting minimum MSE is stated by the following lemma, whose proof can be found in Appendix G.

Lemma 7. *Consider a parameter vector $\mathbf{x} \in \mathcal{X}_S$ with maximal support, i.e., $\|\mathbf{x}\|_0 = S$. Then, for any $k \notin \text{supp}(\mathbf{x})$, the minimum MSE of any estimator $\hat{x}_k(\cdot) \in \mathcal{U}^k \cap \mathcal{A}_{\mathbf{x}}^k$, denoted by $\text{BB}_c^k(\mathbf{x})$, is given by*

$$\text{BB}_c^k(\mathbf{x}) = \left[1 - \prod_{l \in \text{supp}(\mathbf{x})} g(x_l; \sigma^2) \right] \sigma^2 \quad (26)$$

with

$$g(x; \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^\infty e^{-(x^2+y^2)/(2\sigma^2)} \sinh\left(\frac{xy}{\sigma^2}\right) \tanh\left(\frac{xy}{\sigma^2}\right) dy. \quad (27)$$

Lemma 7 identifies the minimum MSE of any unbiased estimator of the k th component of \mathbf{x} (where $k \notin \text{supp}(\mathbf{x})$) that is also constrained to be an element of $\mathcal{A}_{\mathbf{x}}^k$. Note that $\text{BB}_c^k(\mathbf{x})$ does not depend on k . It provides an upper bound on the minimum MSE of any unbiased estimator of the k th component of \mathbf{x} , for any $k \notin \text{supp}(\mathbf{x})$.

The total MSE of a vector estimator $\hat{\mathbf{x}}(\cdot)$ can be decomposed as $\varepsilon(\mathbf{x}; \hat{\mathbf{x}}) = \sum_{k \in \text{supp}(\mathbf{x})} \varepsilon(\mathbf{x}; \hat{x}_k) + \sum_{k \notin \text{supp}(\mathbf{x})} \varepsilon(\mathbf{x}; \hat{x}_k)$ with the component MSE $\varepsilon(\mathbf{x}; \hat{x}_k) \triangleq \mathbb{E}_{\mathbf{x}} \{ (\hat{x}_k(\mathbf{y}) - x_k)^2 \}$. Inserting the minimum component MSE for $k \in \text{supp}(\mathbf{x})$ (which is σ^2 according to Lemma 6) in the first sum and the upper bound $\text{BB}_c^k(\mathbf{x})$ on the minimum component MSE for $k \notin \text{supp}(\mathbf{x})$ in the second sum, we obtain the following upper bound on the minimum total MSE of any unbiased vector estimator.

Theorem 8. *The minimum MSE achievable by any unbiased estimator for the SSNM at a parameter vector $\mathbf{x} \in \mathcal{X}_S$ with $\|\mathbf{x}\|_0 = S$ satisfies*

$$\varepsilon(\mathbf{x}; \hat{\mathbf{x}}^{(\mathbf{x})}) \leq \text{BB}_c(\mathbf{x}) \triangleq S\sigma^2 + (N-S)\text{BB}_c^k(\mathbf{x}) \quad (28)$$

with $\text{BB}_c^k(\mathbf{x})$ given by (26).

Depending on the parameter vector \mathbf{x} , the upper bound $\text{BB}_c(\mathbf{x})$ varies between two extreme values. For decreasing SNR ξ^2/σ^2 , it converges to the low-SNR value $N\sigma^2$ (because the factor $g(\xi, \sigma^2)$ in (26) vanishes for $\xi^2/\sigma^2 \rightarrow 0$). On the other hand, we will show below that for increasing SNR, $\text{BB}_c(\mathbf{x})$ converges to its high-SNR value, which is given by $S\sigma^2$.

The lower bound $\text{HCRB}(\mathbf{x})$ in (19) for the case $\|\mathbf{x}\|_0 = S$, i.e., $S\sigma^2 + (N-S-1)e^{-\xi^2/\sigma^2}\sigma^2$, exhibits an exponential transition between the low-SNR and high-SNR regimes. More specifically, when considering a sequence of parameter vectors $\mathbf{x} \in \mathcal{X}_S$ with increasing SNR ξ^2/σ^2 , the bound transitions from the low-SNR value $(N-1)\sigma^2$ (obtained for $\xi^2/\sigma^2 = 0$) to the high-SNR value $S\sigma^2$ (obtained for $\xi^2/\sigma^2 \rightarrow \infty$); this transition is exponential in the SNR. The upper bound $\text{BB}_c(\mathbf{x})$ in (28) also exhibits a transition that is exponential in ξ^2/σ^2 . In fact, it is shown in Appendix H that

$$\text{BB}_c(\mathbf{x}) \leq S\sigma^2 + (N-S)3^S e^{-\xi^2/(2\sigma^2)}\sigma^2. \quad (29)$$

This shows that for increasing ξ^2/σ^2 , the upper bound $\text{BB}_c(\mathbf{x})$ —just like the lower bound $\text{HCRB}(\mathbf{x})$ —decays exponentially to its asymptotic value $S\sigma^2$, which is also the asymptotic value of $\text{HCRB}(\mathbf{x})$. It follows that the BB itself also converges exponentially to $S\sigma^2$ as ξ^2/σ^2 increases. This result will be further explored in Section V-C.

V. NUMERICAL RESULTS

In this section, we describe several numerical studies which explore and extend the theoretical bounds developed above. These include a numerical improvement of the bounds, a comparison with practical (biased) estimation techniques, an analysis of the performance at high SNR, and an examination of the ability to estimate the threshold region in which the transition from low to high SNR occurs.

We will first show that it is possible to obtain significantly tighter versions of the lower and upper bounds developed in Sections III and IV. These tightened versions can only be computed numerically and no longer have a simple form; consequently, they are less convenient for theoretical analyses. Nevertheless, they characterize the BB very accurately and therefore also provide an indication of the accuracy of the simpler, closed-form bounds.

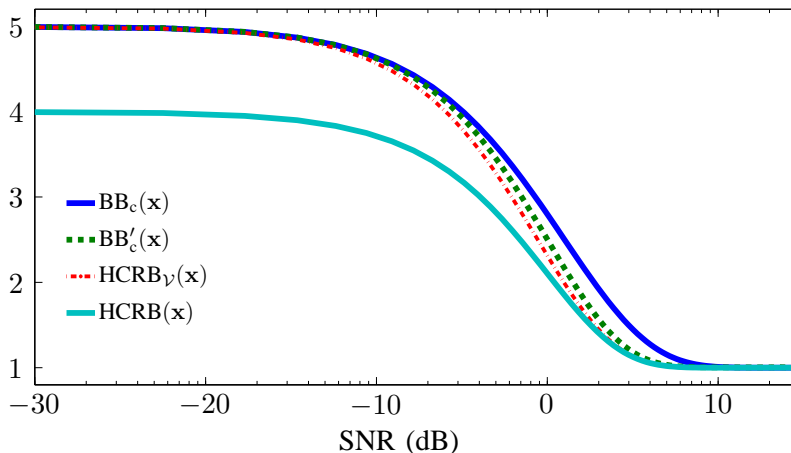


Fig. 1. Lower bounds $\text{HCRB}(\mathbf{x})$, $\text{HCRB}_{\mathcal{V}}(\mathbf{x})$ and upper bounds $\text{BB}_c(\mathbf{x})$, $\text{BB}'_c(\mathbf{x})$ on the MSE $\varepsilon(\mathbf{x}; \hat{\mathbf{x}}^{(\infty)})$ of the LMVU estimator at $\mathbf{x} = c(1 \ 0 \ 0 \ 0 \ 0)^T$, with c varied to obtain different values of $\text{SNR}(\mathbf{x}) = \xi^2/\sigma^2$. The SSNM parameters are $N=5$, $S=1$, and $\sigma^2=1$.

A. Numerical Lower Bound

For a parameter vector \mathbf{x} with $\|\mathbf{x}\|_0 = S$, let us reconsider the HCRB in (16). We will show that by using an increased number of appropriately chosen test points, we can obtain a lower bound that is higher (thus, tighter) than (19). Specifically, assume without loss of generality that $\text{supp}(\mathbf{x}) = \{1, \dots, S\}$, and consider the set of test points

$$\mathcal{V} \triangleq \mathcal{V}_0 \cup \bigcup_{k=1}^S (\mathcal{V}_k \cup \mathcal{W}_k)$$

with the component sets

$$\begin{aligned} \mathcal{V}_0 &\triangleq \bigcup_{l \in \text{supp}(\mathbf{x})} \{\alpha \mathbf{e}_l\} \\ \mathcal{V}_k &\triangleq \bigcup_{l \in \{S+1, \dots, N\}} \{\alpha \mathbf{e}_l - x_k \mathbf{e}_k\}, \quad k = 1, \dots, S \\ \mathcal{W}_k &\triangleq \bigcup_{l \in \{S+1, \dots, N\}} \{x_k \mathbf{e}_l - x_k \mathbf{e}_k\}, \quad k = 1, \dots, S \end{aligned}$$

where $\alpha = 0.02\sigma$. In Fig. 1, the HCRB (16) for the new set \mathcal{V} of test points—denoted $\text{HCRB}_{\mathcal{V}}(\mathbf{x})$ —is displayed versus the SNR and compared with $\text{HCRB}(\mathbf{x})$. For this figure, we chose $N=5$, $S=1$, $\sigma^2=1$, and $\mathbf{x} = c(1 \ 0 \ 0 \ 0 \ 0)^T$, where the parameter $c \in \mathbb{R}$ is varied to obtain different SNR values.³ As before,

³The use of a low-dimensional model is mandated by the complexity of the numerical approximation to the upper bound on the BB which will be described in Section V-B.

the SNR is defined as $\text{SNR}(\mathbf{x}) = \xi^2/\sigma^2$, where ξ is the S -largest (in magnitude) component of \mathbf{x} (in our example with $S = 1$, ξ is simply the single nonzero component). It can be seen from Fig. 1 that the numerical lower bound $\text{HCRB}_V(\mathbf{x})$ computed from the above test points is indeed tighter than the closed-form lower bound $\text{HCRB}(\mathbf{x})$ in (19).

B. Numerical Upper Bound

It is also possible to find upper bounds on the BB that are tighter (lower) than the upper bound $\text{BB}_c(\mathbf{x})$ in (28). Consider a parameter vector \mathbf{x} with $\|\mathbf{x}\|_0 = S$. We recall that $\text{BB}_c(\mathbf{x})$ was derived by constructing, for all $k \notin \text{supp}(\mathbf{x})$, unbiased estimators $\hat{x}_k(\mathbf{y}) = y_k + \hat{x}'_k(\mathbf{y})$ with $\hat{x}'_k(\mathbf{y})$ constrained by (23) and (24). We will now investigate how much we can improve on $\text{BB}_c(\mathbf{x})$ if we remove the constraint (23). Thus, in the optimization problem (22), the constraint set \mathcal{A}_x is hereafter considered to correspond only to the constraint (24).

In order to numerically solve this modified optimization problem (22), a discrete approximation for $\hat{x}'_k(\mathbf{y})$ was used. More specifically, we defined $\hat{x}'_k(\mathbf{y})$ to be piecewise constant in each of the components y_l with $l \in \{k\} \cup \text{supp}(\mathbf{x})$, and constant in the remaining components y_l (the latter being required by (24)). We used Q piecewise constant segments for each $l \in \{k\} \cup \text{supp}(\mathbf{x})$, with each segment of length $\Delta = 10\sigma/Q$. These arrays of constant segments were centered about $\mathbf{y} = \mathbf{x}$. The remaining values of $\hat{x}'_k(\mathbf{y})$ were set to 0. Thus, we obtained a function $\hat{x}'_k(\mathbf{y})$ with linear dependence on a finite number Q^{S+1} of parameters. For functions of this form, the optimization problem (22) becomes a finite-dimensional quadratic program with linear constraints, which can be solved efficiently [29]. The MSE of the resulting estimator, denoted by $\text{BB}'_c(\mathbf{x})$, is an upper bound on the BB. This bound is tighter than the closed-form upper bound $\text{BB}_c(\mathbf{x})$ in (28) if Q is large enough. In Fig. 1, we compare $\text{BB}'_c(\mathbf{x})$ for $Q = 20$ with $\text{BB}_c(\mathbf{x})$ as a function of the SNR. The improved accuracy of $\text{BB}'_c(\mathbf{x})$ relative to $\text{BB}_c(\mathbf{x})$ is evident, particularly at high SNR values. Moreover, the proximity of the numerical upper bound $\text{BB}'_c(\mathbf{x})$ to the numerical lower bound $\text{HCRB}_V(\mathbf{x})$ indicates that these two bounds achieve an accurate characterization of the BB, since the BB lies between them.

C. The Role of ξ

We have seen in Section IV that for $\|\mathbf{x}\|_0 = S$, the MSE of the LMVU estimator at high SNR is given by $S\sigma^2$, and furthermore, convergence to this value is exponential in the quantity ξ^2/σ^2 . A remarkable aspect of this conclusion is the fact that convergence to the high-SNR regime depends solely on ξ , the smallest nonzero component of \mathbf{x} , rather than having a more complex dependency on all the S nonzero

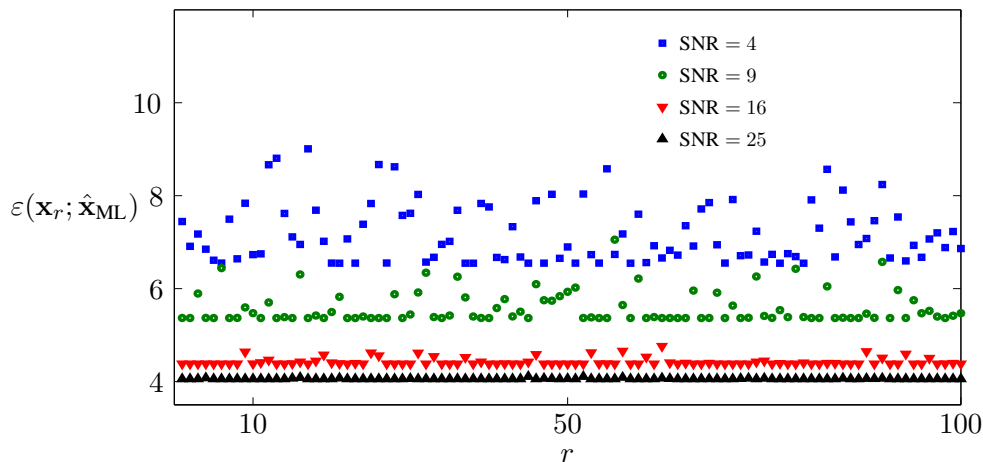


Fig. 2. MSE $\varepsilon(\mathbf{x}_r; \hat{\mathbf{x}}_{ML})$ of the ML estimator for randomly generated parameter vectors \mathbf{x}_r at four different SNRs ξ^2/σ^2 , for SSNM parameters $N=10$, $S=4$, and $\sigma^2=1$.

components of \mathbf{x} . For example, one might imagine the behavior of an estimator to be rather different when all nonzero components have the same value ξ , as opposed to the situation in which one component equals ξ and the others are much larger. However, our analysis shows that when $\xi \gg \sigma$, the remaining components of \mathbf{x} have no effect on the performance of the LMVU estimator. We will next investigate whether practical estimators also exhibit such an effect.

To answer this question, we examined the MSE of the ML estimator (7) for a wide range of parameter vectors \mathbf{x} having a predetermined smallest component ξ . More specifically, for a given value of ξ , we randomly generated 100 parameter vectors \mathbf{x}_r , $r = 1, \dots, 100$, with $\mathbf{x}_r \in \mathcal{X}_S$ and $\|\mathbf{x}_r\|_0 = S$, whose minimum nonzero component was equal to ξ . The other nonzero components were generated as independent, identically distributed realizations of the random variable $x = \xi(1 + 3\sigma|q|)$, where $q \sim \mathcal{N}(0, 1)$ is a standard Gaussian random variable and σ is the standard deviation of the noise. The MSE $\varepsilon(\mathbf{x}_r; \hat{\mathbf{x}}_{ML})$ of the ML estimator is shown in Fig. 2 for $N=10$, $S=4$, and four different SNRs ξ^2/σ^2 , with the horizontal axis representing the different choices of \mathbf{x}_r in arbitrary order. It is seen that for large ξ , the performance of the ML estimator, like that of the LMVU, depends almost exclusively on ξ . This suggests that the performance guarantees of Sections III and IV, while formally valid only for unbiased estimators, can still provide general conclusions which are relevant to biased techniques such as the ML estimator. Moreover, this result also justifies our definition of the SNR as the ratio ξ^2/σ^2 , since this is the most significant factor determining estimation performance for the SSNM.

D. Threshold Region Identification

In Sections III and IV, we characterized the performance of unbiased estimators as a means of quantifying the difficulty of estimation for the SSNM. A common use of this analysis is in the identification of the threshold region, a range of SNR values which constitutes a transition between low-SNR and high-SNR behavior [30–32]. Specifically, in many cases the performance of estimators can be calculated analytically when the SNR is either very low or very high. It is then important to identify the threshold region which separates these two regimes. Although the analysis is based on bounds for unbiased estimators, the result is often heuristically assumed to approximate the threshold region for biased techniques as well [30, 32].

For $\|\mathbf{x}\|_0 = S$, the lower and upper bounds on the BB (HCRB(\mathbf{x}) in (19), $\text{BB}_c(\mathbf{x})$ in (28)) exhibit a transition between a low-SNR region, where both bounds are on the order of $N\sigma^2$, and a high-SNR region, for which both bounds converge to $S\sigma^2$. The BB therefore also displays such a transition. One can define the threshold region of the SSNM (for unbiased estimation) as the range of values of ξ^2/σ^2 in which this transition takes place. Since the BB is itself a lower bound on the performance of unbiased estimators, one would expect the transition region of actual estimators to occur at slightly higher SNR values than that of the BB.

To test this hypothesis, we compared the bounds of Sections III and IV with the MSE of two well-known estimation schemes, namely, the ML estimator in (7) and the hard-thresholding (HT) estimator $\hat{\mathbf{x}}_{\text{HT}}(\mathbf{y})$, which is given componentwise as

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

for a given threshold $T > 0$. In our simulations, we chose the commonly used value $T = \sigma\sqrt{2\log N}$ [33]. Note that since the ML and HT estimators are biased, their MSE is not bounded by $\text{BB}_c(\mathbf{x})$, HCRB(\mathbf{x}), and the CRB. Assuming SSNM parameters $N = 10$ and $S = 4$, we generated a number of parameter vectors \mathbf{x} from the set $\mathcal{R} \triangleq \{c(1 \ 1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0)^T\}_{c \in \mathbb{R}}$, where c was varied to obtain a range of SNR values. For these \mathbf{x} , we calculated the MSE of the two estimators $\hat{\mathbf{x}}_{\text{ML}}$ and $\hat{\mathbf{x}}_{\text{HT}}$ by means of numerical integration (see Appendix I for a discussion of the computation of $\varepsilon(\mathbf{x}; \hat{\mathbf{x}}_{\text{ML}})$).

The results are displayed in Fig. 3 as a function of the SNR ξ^2/σ^2 . Although there is some gap between the lower bound (HCRB) and the upper bound (BB_c), a rough indication of the behavior of the BB is conveyed. As expected, the threshold region exhibited by the ML and HT estimators is somewhat higher than that predicted by the bounds. Specifically, the threshold region of the BB (as indicated by

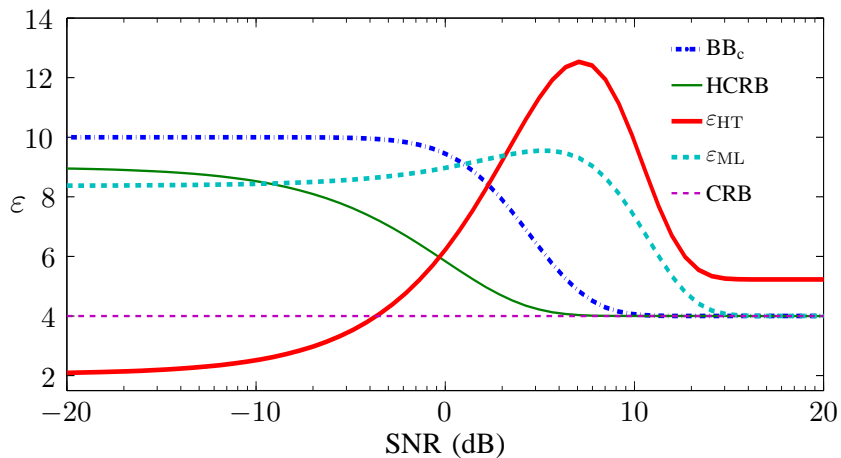


Fig. 3. MSE of the ML and HT estimators compared with the performance bounds $\text{BB}_c(\mathbf{x})$, $\text{HCRB}(\mathbf{x})$, and $\text{CRB} (\equiv S\sigma^2)$, as a function of the SNR ξ^2/σ^2 , for SSNM parameters $N=10$, $S=4$, and $\sigma^2=1$.

the bounds) can be seen to occur at SNR values between -5 and 5 dB, while the threshold region of the ML and HT estimators is at SNR values between 5 and 12 dB. Another effect which is visible in Fig. 3 is the convergence of the ML estimator to the BB at high SNR; this is a manifestation of the well-known fact that the ML estimator is asymptotically unbiased and asymptotically optimal. Finally, at low SNR, both the ML and HT estimators are better than the best unbiased approach. This is because unbiased methods generally perform poorly at low SNR, so that even the best unbiased technique is outperformed by the biased ML and HT estimators. On the other hand, for medium SNR, the MSE of the ML and HT estimators is significantly higher than the BB. Thus, there is a potential for unbiased estimators to perform better than biased estimators in the medium-SNR regime.

One may argue that considering only parameter vectors \mathbf{x} in the set \mathcal{R} is not representative, since \mathcal{R} covers only a small part of the parameter space \mathcal{X}_S . However, the choice of \mathcal{R} is conservative in that the maximum deviation between $\text{HCRB}(\mathbf{x})$ and $\text{BB}_c(\mathbf{x})$ is largest when the nonzero entries of \mathbf{x} have approximately the same magnitude, which is the case for each element of \mathcal{R} . This is illustrated in Fig. 4, which shows the ratio between the two bounds versus the SNR ξ^2/σ^2 for three different configurations of the nonzero entries in the parameter vector. Specifically, we considered the two additional sets $\mathcal{R}_2 \triangleq \{c(10 \ 1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0)^T\}_{c \in \mathbb{R}}$ and $\mathcal{R}_3 \triangleq \{c(0.1 \ 1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0)^T\}_{c \in \mathbb{R}}$, in which the nonzero entries have different magnitudes. It can be seen from Fig. 4 that the ratio $\text{BB}_c(\mathbf{x})/\text{HCRB}(\mathbf{x})$ is indeed highest when \mathbf{x} is in \mathcal{R} .

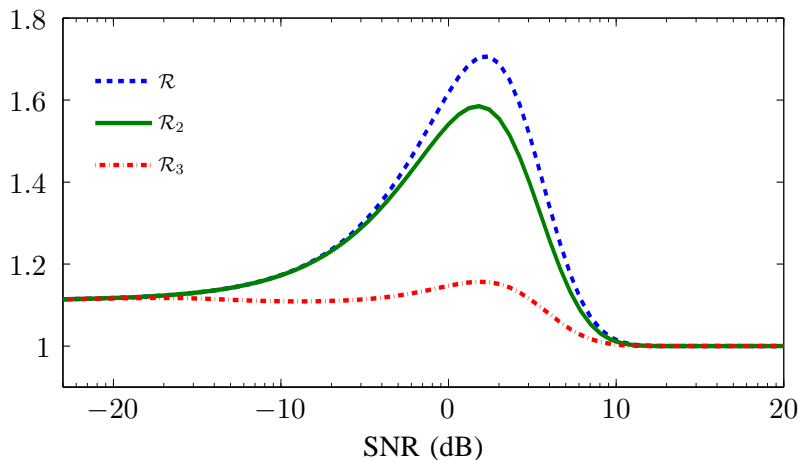


Fig. 4. Ratio $\text{BB}_c(\mathbf{x})/\text{HCRB}(\mathbf{x})$ versus the SNR ξ^2/σ^2 for different sets of parameter vectors \mathbf{x} .

VI. CONCLUSION

In this paper, we have studied unbiased estimation of a sparse vector in white Gaussian noise within a frequentist setting. As we have seen, without the assumption of sparsity, there exists only a single unbiased estimator. However, the addition of a sparsity assumption yields a rich family of unbiased estimators. The analysis of the performance of these estimators has been the primary goal of this paper. We first demonstrated that there exists no uniformly minimum variance unbiased estimator, i.e., no single unbiased estimator is optimal for all parameter values. Consequently, we focused on analyzing the Barankin bound (BB), i.e., the MSE of the locally minimum variance unbiased estimator, or equivalently, the smallest MSE achievable by an unbiased estimator for each value of the sparse vector.

For the sparse estimation problem considered, as for most estimation problems, the BB cannot be computed precisely. However, we demonstrated that it can be characterized quite accurately using numerical lower and upper bounds. Furthermore, we derived simple closed-form lower and upper bounds which are somewhat looser than the numerical bounds. These closed-form bounds allow an estimation of the threshold region separating the low-SNR and high-SNR regimes, and they indicate the asymptotic behavior of the BB at high SNR. In particular, a notable conclusion is that the high-SNR behavior of the BB depends solely on the value of the smallest nonzero component of the sparse vector.

While the unbiasedness property is intuitively appealing and related to several desirable asymptotic features of an estimator [6], one can often obtain biased estimators which outperform any unbiased estimator [13–15]. Thus, it is interesting to note that some of the conclusions obtained from our analysis of unbiased estimators appear to provide insight into the behavior of standard biased estimators. In

particular, we saw that the behavior of two commonly used biased estimators at high SNR corresponds to the predictions of our unbiased bounds, not only in terms of the asymptotically achievable MSE but also in certain finer details, such as the SNR range of the threshold region and the fact that the convergence to the high-SNR regime depends primarily on the value of the smallest nonzero component of the sparse vector, rather than on the entire vector. This gives additional merit to the analysis of achievable estimation performance within the unbiased setting.

APPENDIX A

PROOF OF THEOREM 1

We wish to show that for $S = N$, the only unbiased estimator with bounded MSE is the trivial estimator $\hat{\mathbf{x}}(\mathbf{y}) = \mathbf{y}$. We will first show that a bounded MSE implies that $\hat{\mathbf{x}}(\mathbf{y})$ is equivalent to a tempered distribution. This will allow us to reformulate the unbiasedness condition in the Fourier transform domain.

Using (3), the unbiasedness condition in (8) for $S = N$ reads

$$\frac{1}{(2\pi\sigma^2)^{N/2}} \int_{\mathbb{R}^N} \hat{\mathbf{x}}(\mathbf{y}) \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{x}\|_2^2\right) d\mathbf{y} = \mathbf{x} \quad \text{for all } \mathbf{x} \in \mathbb{R}^N. \quad (30)$$

The integral in (30) is the convolution of $\hat{\mathbf{x}}(\mathbf{y})$ with $\exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y}\|_2^2\right)$. The result of this convolution, viewed as a function of \mathbf{x} , must equal $(2\pi\sigma^2)^{N/2} \mathbf{x}$ for all parameter vectors \mathbf{x} . For absolutely integrable functions, the Fourier transform maps a convolution onto a pointwise product, and consequently it seems natural to consider the Fourier transform of condition (30) in order to simplify the analysis. However, typically, the estimator function $\hat{\mathbf{x}}(\mathbf{y})$ will be neither absolutely integrable nor square integrable, and thus its Fourier transform can only exist in the sense of a tempered distribution [34]. From a practical point of view, the class of tempered distributions is large enough so that it does not exclude reasonable estimators such as the LS estimator (7). The following lemma states that $\hat{\mathbf{x}}(\mathbf{y})$ can be viewed as a tempered distribution if it has a bounded MSE.

Lemma 9. *Consider an estimator $\hat{\mathbf{x}}$ for the SSNM (2) with $S = N$. If $\hat{\mathbf{x}}$ has a bounded MSE, i.e., $\varepsilon(\mathbf{x}; \hat{\mathbf{x}}) \leq C$ for all $\mathbf{x} \in \mathbb{R}^N$ (where C is a constant which may depend on N , S , and σ^2), then $\hat{\mathbf{x}}$ is equivalent to a tempered distribution.*

Proof. The proof of Lemma 9 is based on the following result which gives a sufficient condition for a function $\hat{\mathbf{x}}(\mathbf{y})$ to be (equivalent to) a tempered distribution.

Proposition 10 ([34]). *If there exist constants $B, n, R_0 \in \mathbb{R}_+$ such that*

$$\int_{\|\mathbf{y}\|_2 \leq R} \|\hat{\mathbf{x}}(\mathbf{y})\|_2^2 d\mathbf{y} \leq BR^n \quad \text{for all } R \geq R_0 \quad (31)$$

then $\hat{\mathbf{x}}(\mathbf{y})$ is equivalent to a tempered distribution.

Let $\hat{\mathbf{x}}(\mathbf{y})$ be an estimator function with bounded MSE, i.e., there exists a constant C such that

$$\mathbb{E}_{\mathbf{x}}\{\|\hat{\mathbf{x}}(\mathbf{y}) - \mathbf{x}\|_2^2\} \leq C \quad \text{for all } \mathbf{x} \in \mathcal{X}_S. \quad (32)$$

Defining the usual norm $\|\cdot\|_{\text{RV}}$ on the space of random vectors by $\|\mathbf{y}\|_{\text{RV}} \triangleq \sqrt{\mathbb{E}_{\mathbf{x}}\{\|\mathbf{y}\|_2^2\}}$, we can use the (reverse) triangle inequality $\|\hat{\mathbf{x}}(\mathbf{y}) - \mathbf{x}\|_{\text{RV}} \geq \|\hat{\mathbf{x}}(\mathbf{y})\|_{\text{RV}} - \|\mathbf{x}\|_{\text{RV}}$ to obtain

$$\sqrt{\mathbb{E}_{\mathbf{x}}\{\|\hat{\mathbf{x}}(\mathbf{y}) - \mathbf{x}\|_2^2\}} \geq \sqrt{\mathbb{E}_{\mathbf{x}}\{\|\hat{\mathbf{x}}(\mathbf{y})\|_2^2\}} - \sqrt{\mathbb{E}_{\mathbf{x}}\{\|\mathbf{x}\|_2^2\}} = \sqrt{\mathbb{E}_{\mathbf{x}}\{\|\hat{\mathbf{x}}(\mathbf{y})\|_2^2\}} - \|\mathbf{x}\|_2.$$

From this, it follows that

$$\sqrt{\mathbb{E}_{\mathbf{x}}\{\|\hat{\mathbf{x}}(\mathbf{y})\|_2^2\}} \leq \sqrt{\mathbb{E}_{\mathbf{x}}\{\|\hat{\mathbf{x}}(\mathbf{y}) - \mathbf{x}\|_2^2\}} + \|\mathbf{x}\|_2 \leq \sqrt{C} + \|\mathbf{x}\|_2 \quad \text{for all } \mathbf{x} \in \mathcal{X}_S,$$

where (32) has been used. Squaring both sides and using the inequality $(x+y)^2 \leq 2(x^2+y^2)$, we obtain

$$\mathbb{E}_{\mathbf{x}}\{\|\hat{\mathbf{x}}(\mathbf{y})\|_2^2\} \leq (\sqrt{C} + \|\mathbf{x}\|_2)^2 \leq 2(C + \|\mathbf{x}\|_2^2) \quad \text{for all } \mathbf{x} \in \mathcal{X}_S$$

or equivalently

$$\frac{1}{(2\pi\sigma^2)^{N/2}} \int_{\mathbb{R}^N} \|\hat{\mathbf{x}}(\mathbf{y})\|_2^2 e^{-\|\mathbf{y}-\mathbf{x}\|_2^2/(2\sigma^2)} d\mathbf{y} \leq 2(C + \|\mathbf{x}\|_2^2) \quad \text{for all } \mathbf{x} \in \mathcal{X}_S. \quad (33)$$

We will now show that (31) holds for $R_0 = 1$, i.e., $R \geq 1$. We define the N -dimensional grid

$$\mathcal{G} \triangleq \{-m\Delta, -(m-1)\Delta, \dots, -\Delta, 0, \Delta, \dots, m\Delta\}^N$$

where $0 < \Delta \leq R$ (hence, $R/\Delta \geq 1$) and $m = \lfloor R/\Delta \rfloor \leq R/\Delta$. The number of grid points in any single dimension satisfies

$$2m+1 \leq \frac{2R}{\Delta} + 1 \quad (34)$$

so that

$$|\mathcal{G}| = (2m+1)^N \leq \left(\frac{2R}{\Delta} + 1\right)^N. \quad (35)$$

We thus have

$$\begin{aligned} \sum_{\mathbf{x} \in \mathcal{G}} \|\mathbf{x}\|_2^2 &= \sum_{\mathbf{x} \in \mathcal{G}} \sum_{k=1}^N x_k^2 = \sum_{k=1}^N \sum_{\mathbf{x} \in \mathcal{G}} x_k^2 = \sum_{k=1}^N \left[(2m+1)^{N-1} \sum_{l=-m}^m (l\Delta)^2 \right] = N(2m+1)^{N-1} \sum_{l=-m}^m (l\Delta)^2 \\ &\leq N(2m+1)^{N-1} \Delta^2 \int_{x=-R/\Delta}^{R/\Delta} x^2 dx \leq N \left(\frac{2R}{\Delta} + 1 \right)^{N-1} \frac{2}{3} \frac{R^3}{\Delta} \end{aligned} \quad (36)$$

where (34) was used in the last step. Furthermore, for $c \triangleq \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-N\Delta^2/(2\sigma^2)}$, we have

$$\frac{1}{c} \frac{1}{(2\pi\sigma^2)^{N/2}} \sum_{\mathbf{x} \in \mathcal{G}} e^{-\|\mathbf{y}-\mathbf{x}\|_2^2/(2\sigma^2)} \geq 1, \quad \text{for all } \mathbf{y} \text{ with } \|\mathbf{y}\|_2 \leq R \quad (37)$$

In order to verify this inequality, consider an arbitrary $\mathbf{y} \in \mathbb{R}^N$ with $\|\mathbf{y}\|_2 \leq R$. Since $0 < \Delta \leq R$, and since $\|\mathbf{y}\|_2 \leq R$ implies that no component y_k of \mathbf{y} can be larger than R , there always exists a grid point $\tilde{\mathbf{x}} \in \mathcal{G}$ (dependent on \mathbf{y}) such that $|y_k - \tilde{x}_k| \leq \Delta$ for all $k \in \{1, \dots, N\}$. It follows that $\|\mathbf{y} - \tilde{\mathbf{x}}\|_2^2 \leq N\Delta^2$ and, in turn,

$$e^{-N\Delta^2/(2\sigma^2)} \leq e^{-\|\mathbf{y}-\tilde{\mathbf{x}}\|_2^2/(2\sigma^2)} \leq \sum_{\mathbf{x} \in \mathcal{G}} e^{-\|\mathbf{y}-\mathbf{x}\|_2^2/(2\sigma^2)}, \quad \|\mathbf{y}\|_2 \leq R$$

which is equivalent to (37).

Successively using (37), (33), (35), (36), and $1 \leq 2R/\Delta$, we obtain the following sequence of inequalities:

$$\begin{aligned} \int_{\|\mathbf{y}\|_2 \leq R} \|\hat{\mathbf{x}}(\mathbf{y})\|_2^2 d\mathbf{y} &\leq \int_{\|\mathbf{y}\|_2 \leq R} \|\hat{\mathbf{x}}(\mathbf{y})\|_2^2 \left[\frac{1}{c} \frac{1}{(2\pi\sigma^2)^{N/2}} \sum_{\mathbf{x} \in \mathcal{G}} e^{-\|\mathbf{y}-\mathbf{x}\|_2^2/(2\sigma^2)} \right] d\mathbf{y} \\ &\leq \frac{1}{c} \sum_{\mathbf{x} \in \mathcal{G}} \frac{1}{(2\pi\sigma^2)^{N/2}} \int_{\mathbb{R}^N} \|\hat{\mathbf{x}}(\mathbf{y})\|_2^2 e^{-\|\mathbf{y}-\mathbf{x}\|_2^2/(2\sigma^2)} d\mathbf{y} \\ &\leq \frac{1}{c} \sum_{\mathbf{x} \in \mathcal{G}} 2(C + \|\mathbf{x}\|_2^2) \\ &\leq \frac{2}{c} \left[\left(\frac{2R}{\Delta} + 1 \right)^N C + N \left(\frac{2R}{\Delta} + 1 \right)^{N-1} \frac{2}{3} \frac{R^3}{\Delta} \right] \\ &\leq \frac{2}{c} \left[\left(\frac{4R}{\Delta} \right)^N C + N \left(\frac{4R}{\Delta} \right)^{N-1} \frac{2}{3} \frac{R^3}{\Delta} \right]. \end{aligned} \quad (38)$$

It then follows from (38) that for $R \geq 1$

$$\begin{aligned} \int_{\|\mathbf{y}\|_2 \leq R} \|\hat{\mathbf{x}}(\mathbf{y})\|_2^2 d\mathbf{y} &\leq \frac{2}{c} \left[\left(\frac{4}{\Delta} \right)^N R^{N+2} C + N \left(\frac{4}{\Delta} \right)^{N-1} \frac{2}{3} \frac{R^{N+2}}{\Delta} \right] \\ &\leq \frac{2}{c} \frac{R^{N+2}}{\Delta^N} \left(4^N C + N 4^N \frac{2}{3} \right) \end{aligned}$$

$$= \frac{2^{2N+1}}{c\Delta^N} \left(C + \frac{2N}{3} \right) R^{N+2}.$$

Thus, we have established that under the conditions of Lemma 9 (bounded MSE), the bound (31) holds with $R_0 = 1$, $B = \frac{2^{2N+1}}{c\Delta^N} (C + 2N/3)$, and $n = N + 2$. Therefore, it follows from Proposition 10 that an estimator with bounded MSE is equivalent to a tempered distribution. This concludes the proof of Lemma 9. \square

We now continue our proof of Theorem 1. Any estimator $\hat{\mathbf{x}}(\mathbf{y})$ for the SSNM (2) can be written as

$$\hat{\mathbf{x}}(\mathbf{y}) = \mathbf{y} + \hat{\mathbf{x}}'(\mathbf{y}) \quad (39)$$

with the correction term $\hat{\mathbf{x}}'(\mathbf{y}) \triangleq \hat{\mathbf{x}}(\mathbf{y}) - \mathbf{y}$. Because $\mathbb{E}_{\mathbf{x}}\{\hat{\mathbf{x}}(\mathbf{y})\} = \mathbb{E}_{\mathbf{x}}\{\mathbf{y}\} + \mathbb{E}_{\mathbf{x}}\{\hat{\mathbf{x}}'(\mathbf{y})\} = \mathbf{x} + \mathbb{E}_{\mathbf{x}}\{\hat{\mathbf{x}}'(\mathbf{y})\}$, $\hat{\mathbf{x}}(\mathbf{y})$ is unbiased if and only if

$$\mathbf{b}(\mathbf{x}; \hat{\mathbf{x}}) = \mathbb{E}_{\mathbf{x}}\{\hat{\mathbf{x}}'(\mathbf{y})\} \equiv \frac{1}{(2\pi\sigma^2)^{N/2}} \int_{\mathbb{R}^N} \hat{\mathbf{x}}'(\mathbf{y}) e^{-\|\mathbf{y}-\mathbf{x}\|_2^2/(2\sigma^2)} d\mathbf{y} = \mathbf{0} \quad \text{for all } \mathbf{x} \in \mathcal{X}_S. \quad (40)$$

Remember that we assume that $\hat{\mathbf{x}}$ has a bounded MSE, so that according to our above proof of Lemma 9, the estimator function $\hat{\mathbf{x}}(\mathbf{y})$ satisfies condition (31) with $n = N + 2$, i.e.,

$$\int_{\|\mathbf{y}\|_2 \leq R} \|\hat{\mathbf{x}}(\mathbf{y})\|_2^2 d\mathbf{y} \leq BR^{N+2} \quad \text{for all } R \geq 1 \quad (41)$$

with B as given at the end of the proof of Lemma 9. We will also need the following bound, in which $\mathcal{R} \triangleq [-R, R]^N$:

$$\int_{\|\mathbf{y}\|_2 \leq R} \|\mathbf{y}\|_2^2 d\mathbf{y} \leq \int_{\mathcal{R}} \|\mathbf{y}\|_2^2 d\mathbf{y} = \sum_{k=1}^N \int_{\mathcal{R}} y_k^2 d\mathbf{y} = \sum_{k=1}^N (2R)^{N-1} \frac{2}{3} R^3 = \frac{N}{3} 2^N R^{N+2}. \quad (42)$$

We then have for the correction term $\hat{\mathbf{x}}'(\mathbf{y})$, for all $R \geq 1$,

$$\begin{aligned} \int_{\|\mathbf{y}\|_2 \leq R} \|\hat{\mathbf{x}}'(\mathbf{y})\|_2^2 d\mathbf{y} &= \int_{\|\mathbf{y}\|_2 \leq R} \|\hat{\mathbf{x}}(\mathbf{y}) - \mathbf{y}\|_2^2 d\mathbf{y} \\ &\leq \int_{\|\mathbf{y}\|_2 \leq R} 2(\|\hat{\mathbf{x}}(\mathbf{y})\|_2^2 + \|\mathbf{y}\|_2^2) d\mathbf{y} \\ &= 2 \left(\int_{\|\mathbf{y}\|_2 \leq R} \|\hat{\mathbf{x}}(\mathbf{y})\|_2^2 d\mathbf{y} + \int_{\|\mathbf{y}\|_2 \leq R} \|\mathbf{y}\|_2^2 d\mathbf{y} \right) \\ &\leq 2 \left(BR^{N+2} + \frac{N}{3} 2^N R^{N+2} \right) \\ &= \left(2B + \frac{N}{3} 2^{N+1} \right) R^{N+2} \end{aligned}$$

where (41) and (42) have been used. Therefore, the correction term $\hat{\mathbf{x}}'(\mathbf{y})$ also satisfies (31) and thus, according to Proposition 10, it is equivalent to a tempered distribution.

The bias function $\mathbf{b}(\mathbf{x}, \hat{\mathbf{x}})$ in (40) is the convolution of $\hat{\mathbf{x}}'(\mathbf{y})$ with the Gaussian function $(2\pi\sigma^2)^{-N/2} e^{-\|\mathbf{y}\|_2^2/(2\sigma^2)}$. Because $S = N$, we have $\mathcal{X}_S = \mathbb{R}^N$, and thus (40) holds for all $\mathbf{x} \in \mathbb{R}^N$. Since $\hat{\mathbf{x}}'(\mathbf{y})$ is a tempered distribution and the Gaussian function is in the Schwartz class, it follows that the Fourier transform of the convolution product (40) is a smooth function which can be calculated as the pointwise product $\bar{\mathbf{x}}'(\bar{\mathbf{y}}) e^{-\|\bar{\mathbf{y}}\|_2^2/(2\sigma^2)}$, where $\bar{\mathbf{x}}'(\bar{\mathbf{y}})$ denotes the Fourier transform of $\hat{\mathbf{x}}'(\mathbf{y})$ [34]. Therefore, (40) is equivalent to $\bar{\mathbf{x}}'(\bar{\mathbf{y}}) e^{-\|\bar{\mathbf{y}}\|_2^2/(2\sigma^2)} = \mathbf{0}$ for all $\bar{\mathbf{y}} \in \mathbb{R}^N$. This can only be satisfied if $\bar{\mathbf{x}}'(\bar{\mathbf{y}}) \equiv \mathbf{0}$, which in turn implies that $\hat{\mathbf{x}}'(\mathbf{y}) \equiv \mathbf{0}$ (up to deviations of zero measure) and further, by (39), that $\hat{\mathbf{x}}(\mathbf{y}) = \mathbf{y}$. Recalling that $\mathcal{X}_S = \mathbb{R}^N$, it is clear from (5) that $\hat{\mathbf{x}}(\mathbf{y}) = \mathbf{y}$ is the LS estimator. Thus, we have shown that $\hat{\mathbf{x}}_{\text{LS}}(\mathbf{y}) = \mathbf{y}$ is the unique unbiased estimator for the SSNM with $S = N$.

APPENDIX B

PROOF OF THEOREM 3

We must show that there exists no UMVU estimator for the SSNM with $S < N$. The outline of our proof is as follows. We first demonstrate that the unique solution of the optimization problem (11) at the parameter value $\mathbf{x} = \mathbf{0}$, i.e., $\arg \min_{\hat{\mathbf{x}}(\cdot) \in \mathcal{U}} V(\mathbf{0}; \hat{\mathbf{x}})$, is the estimator $\hat{\mathbf{x}}^{(0)}(\mathbf{y}) = \mathbf{y}$. We then show that there exist unbiased estimators which have lower variance than $\hat{\mathbf{x}}^{(0)}$ at other points \mathbf{x} . This implies that neither $\hat{\mathbf{x}}^{(0)}$ nor any other estimator uniformly minimizes the variance for all \mathbf{x} among all unbiased estimators.

The estimator $\hat{\mathbf{x}}^{(0)}(\mathbf{y}) = \mathbf{y}$ is a solution of (11) when $\mathbf{x} = \mathbf{0}$ because the minimum variance at $\mathbf{x} = \mathbf{0}$ of any unbiased estimator is bounded below by $N\sigma^2$ and $\hat{\mathbf{x}}^{(0)}(\mathbf{y}) = \mathbf{y}$ achieves this lower bound [21]. To show that $\hat{\mathbf{x}}^{(0)}$ is the unique solution of (11) for $\mathbf{x} = \mathbf{0}$, suppose by contradiction that there exists a second unbiased estimator $\hat{\mathbf{x}}_a$ different from $\hat{\mathbf{x}}^{(0)}$, also having variance $N\sigma^2$ at $\mathbf{x} = \mathbf{0}$. Consider the estimator $\hat{\mathbf{x}}_{\text{new}} \triangleq (\hat{\mathbf{x}}^{(0)} + \hat{\mathbf{x}}_a)/2$. Since $\hat{\mathbf{x}}^{(0)}$ and $\hat{\mathbf{x}}_a$ are unbiased, $\hat{\mathbf{x}}_{\text{new}}$ is unbiased as well. Thus, its variance is (see (4)) $V(\mathbf{x}; \hat{\mathbf{x}}_{\text{new}}) = P(\mathbf{x}; \hat{\mathbf{x}}_{\text{new}}) - \|\mathbf{x}\|_2^2$. In particular, we obtain for $\mathbf{x} = \mathbf{0}$

$$\begin{aligned} V(\mathbf{0}; \hat{\mathbf{x}}_{\text{new}}) &= P(\mathbf{0}; \hat{\mathbf{x}}_{\text{new}}) = \mathbb{E}_{\mathbf{x}=\mathbf{0}} \left\{ \left\| \frac{1}{2}(\hat{\mathbf{x}}^{(0)} + \hat{\mathbf{x}}_a) \right\|_2^2 \right\} \\ &= \frac{1}{4} \left[\mathbb{E}_{\mathbf{x}=\mathbf{0}} \{ \|\hat{\mathbf{x}}^{(0)}\|_2^2 \} + \mathbb{E}_{\mathbf{x}=\mathbf{0}} \{ \|\hat{\mathbf{x}}_a\|_2^2 \} + 2 \mathbb{E}_{\mathbf{x}=\mathbf{0}} \{ (\hat{\mathbf{x}}^{(0)})^T \hat{\mathbf{x}}_a \} \right] \\ &\stackrel{(*)}{<} \frac{1}{4} \left[\mathbb{E}_{\mathbf{x}=\mathbf{0}} \{ \|\hat{\mathbf{x}}^{(0)}\|_2^2 \} + \mathbb{E}_{\mathbf{x}=\mathbf{0}} \{ \|\hat{\mathbf{x}}_a\|_2^2 \} + 2\sqrt{\mathbb{E}_{\mathbf{x}=\mathbf{0}} \{ \|\hat{\mathbf{x}}^{(0)}\|_2^2 \} \mathbb{E}_{\mathbf{x}=\mathbf{0}} \{ \|\hat{\mathbf{x}}_a\|_2^2 \}} \right] \\ &= \frac{1}{4} \cdot 4N\sigma^2 = N\sigma^2 \end{aligned}$$

where the strict inequality (*) follows from the Cauchy-Schwarz inequality applied to the inner product $E_{\mathbf{x}=\mathbf{0}}\{(\hat{\mathbf{x}}^{(0)})^T \hat{\mathbf{x}}_a\}$, combined with the fact that $\hat{\mathbf{x}}^{(0)}$ and $\hat{\mathbf{x}}_a$ are not linearly dependent (indeed, $\hat{\mathbf{x}}_a \neq c\hat{\mathbf{x}}^{(0)}$ since $\hat{\mathbf{x}}^{(0)}$ and $\hat{\mathbf{x}}_a$ were assumed to be different unbiased estimators). This inequality means that the variance of $\hat{\mathbf{x}}_{\text{new}}$ at $\mathbf{x}=\mathbf{0}$ is lower than $N\sigma^2$. But this is impossible, as $N\sigma^2$ is the minimum variance at $\mathbf{x}=\mathbf{0}$ achieved by any unbiased estimator. Thus, we have shown that $\hat{\mathbf{x}}^{(0)}$ is the unique solution of (11) for $\mathbf{x}=\mathbf{0}$.

Next, still for $S < N$, we consider the specific parameter value $\mathbf{x}' \in \mathcal{X}_S$ whose components are given by

$$x'_k = \begin{cases} 1, & k = 2, \dots, S+1, \\ 0, & \text{else.} \end{cases}$$

The estimator $\hat{\mathbf{x}}^{(0)}$ has variance $V(\mathbf{x}'; \hat{\mathbf{x}}^{(0)}) = N\sigma^2$ at \mathbf{x}' (and at all other $\mathbf{x} \in \mathcal{X}_S$). We will now construct an unbiased estimator $\hat{\mathbf{x}}_b(\mathbf{y})$ whose variance at \mathbf{x}' is smaller than $N\sigma^2$. The components of this estimator are defined as

$$\hat{x}_{b,k}(\mathbf{y}) \triangleq \begin{cases} y_1 + Ay_1 \prod_{l=2}^{S+1} h(y_l), & k = 1 \\ y_k, & k = 2, \dots, N \end{cases} \quad (43)$$

where

$$h(y) \triangleq \begin{cases} \text{sgn}(y), & |y| \in [0.4, 0.6] \\ 0, & \text{else} \end{cases}$$

and $A \in \mathbb{R}$ is a parameter to be determined shortly.⁴ A direct calculation shows that $\hat{\mathbf{x}}_b(\mathbf{y})$ is unbiased for all $\mathbf{x} \in \mathcal{X}_S$. Note that $\hat{\mathbf{x}}_b(\mathbf{y})$ is identical to $\hat{\mathbf{x}}^{(0)}(\mathbf{y}) = \mathbf{y}$ except for the first component, $\hat{x}_{b,1}(\mathbf{y})$.

We recall that for unbiased estimators, minimizing the variance $V(\mathbf{x}; \hat{\mathbf{x}})$ is equivalent to minimizing the mean power $P(\mathbf{x}; \hat{\mathbf{x}}) = E_{\mathbf{x}}\{\|\hat{\mathbf{x}}(\mathbf{y})\|_2^2\}$ (see (4)); furthermore, $P(\mathbf{x}; \hat{\mathbf{x}}) = \sum_{k=1}^N P(\mathbf{x}; \hat{x}_k)$ with $P(\mathbf{x}; \hat{x}_k) \triangleq E_{\mathbf{x}}\{(\hat{x}_k(\mathbf{y}))^2\}$. For the proposed estimator $\hat{\mathbf{x}}_b$, $P(\mathbf{x}'; \hat{x}_{b,k}) = P(\mathbf{x}'; \hat{x}_k^{(0)})$ except for $k=1$. Therefore, our goal is to choose A such that $P(\mathbf{x}'; \hat{x}_{b,1})$ is smaller than $P(\mathbf{x}'; \hat{x}_1^{(0)}) = \sigma^2 + (x'_1)^2 = \sigma^2$. We have

$$P(\mathbf{x}'; \hat{x}_{b,1}) = E_{\mathbf{x}'}\left\{\left(y_1 + Ay_1 \prod_{l=2}^{S+1} h(y_l)\right)^2\right\} = \alpha A^2 + \beta A + \gamma \quad (44)$$

⁴The interval $[0.4, 0.6]$ in the definition of $h(y)$ is chosen rather arbitrarily. Any interval which ensures that β in (44) is nonzero can be used.

with

$$\alpha = \mathbb{E}_{\mathbf{x}'} \left\{ y_1^2 \prod_{l=2}^{S+1} h^2(y_l) \right\}, \quad \beta = \mathbb{E}_{\mathbf{x}'} \left\{ 2y_1^2 \prod_{l=2}^{S+1} h(y_l) \right\}, \quad \gamma = \mathbb{E}_{\mathbf{x}'} \{ y_1^2 \} = \sigma^2.$$

Note that $\gamma = P(\mathbf{x}'; \hat{\mathbf{x}}_1^{(0)})$. From (44), the A minimizing $P(\mathbf{x}'; \hat{x}_{b,1})$ is obtained as $-\beta/(2\alpha)$; the associated minimum $P(\mathbf{x}'; \hat{x}_{b,1})$ is given by $\gamma - \beta^2/(4\alpha^2)$. It can be shown that β is nonzero due to the construction of $h(y)$. It follows that β is positive, and therefore $P(\mathbf{x}'; \hat{x}_{b,1})$ is smaller than $\gamma = P(\mathbf{x}'; \hat{\mathbf{x}}_1^{(0)})$. Thus, using $A = -\beta/(2\alpha)$ in (43), we obtain an estimator $\hat{\mathbf{x}}_b$ which has a smaller component power $P(\mathbf{x}'; \hat{x}_{b,1})$ than $\hat{\mathbf{x}}^{(0)}$. Since $P(\mathbf{x}'; \hat{x}_{b,k}) = P(\mathbf{x}'; \hat{x}_k^{(0)})$ for $k = 2, \dots, N$, it follows that the overall mean power of $\hat{\mathbf{x}}_b$ at \mathbf{x}' is smaller than that of $\hat{\mathbf{x}}^{(0)}$, i.e., $P(\mathbf{x}'; \hat{\mathbf{x}}_b) < P(\mathbf{x}'; \hat{\mathbf{x}}^{(0)})$. Since both estimators are unbiased, this moreover implies that at \mathbf{x}' , the variance of $\hat{\mathbf{x}}_b$ is smaller than that of $\hat{\mathbf{x}}^{(0)}$. Thus, $\hat{\mathbf{x}}^{(0)}$ cannot be the LMVU estimator at $\mathbf{x} = \mathbf{x}'$. On the other hand, as we have seen, $\hat{\mathbf{x}}^{(0)}$ is the unique LMVU estimator at $\mathbf{x} = \mathbf{0}$. We conclude that there does not exist a single unbiased estimator which simultaneously minimizes the variance for all parameters $\mathbf{x} \in \mathcal{X}_S$.

APPENDIX C

PROOF OF PROPOSITION 4

We begin by stating the multivariate HCRB.

Proposition 11 (Gorman and Hero [23]). *Let $f(\mathbf{y}; \mathbf{x})$ be a family of pdf's of \mathbf{y} indexed by $\mathbf{x} \in \mathcal{X}_S$, and let $\mathbf{x} + \mathbf{v}_1, \dots, \mathbf{x} + \mathbf{v}_p$ be a set of points in \mathcal{X}_S . Given an estimator $\hat{\mathbf{x}}$, define*

$$\begin{aligned} \mathbf{m}_{\mathbf{x}} &\triangleq \mathbb{E}_{\mathbf{x}} \{ \hat{\mathbf{x}} \} \\ \delta_i \mathbf{m}_{\mathbf{x}} &\triangleq \mathbf{m}_{\mathbf{x} + \mathbf{v}_i} - \mathbf{m}_{\mathbf{x}} \\ \boldsymbol{\delta} \mathbf{m}_{\mathbf{x}} &\triangleq (\delta_1 \mathbf{m}_{\mathbf{x}} \cdots \delta_p \mathbf{m}_{\mathbf{x}})^T \end{aligned}$$

and

$$\begin{aligned} \delta_i f &\triangleq f(\mathbf{y}; \mathbf{x} + \mathbf{v}_i) - f(\mathbf{y}; \mathbf{x}) \\ \boldsymbol{\delta} f &\triangleq (\delta_1 f \cdots \delta_p f)^T \\ \mathbf{Q} &\triangleq \mathbb{E}_{\mathbf{x}} \left\{ \frac{\boldsymbol{\delta} f}{f} \frac{\boldsymbol{\delta} f^T}{f} \right\}. \end{aligned} \tag{45}$$

Then, the covariance matrix of $\hat{\mathbf{x}}$ satisfies

$$C(\mathbf{x}; \hat{\mathbf{x}}) \succeq \boldsymbol{\delta} \mathbf{m}_{\mathbf{x}}^T \mathbf{Q}^\dagger \boldsymbol{\delta} \mathbf{m}_{\mathbf{x}}. \tag{46}$$

We will now prove Proposition 4 by applying the multivariate HCRB (46) to the case of unbiased estimation under Gaussian noise. For an unbiased estimator $\hat{\mathbf{x}}$, we have $\mathbf{m}_{\mathbf{x}} = \mathbf{x}$, so $\delta_i \mathbf{m}_{\mathbf{x}} = \mathbf{v}_i$ and further

$$\delta \mathbf{m}_{\mathbf{x}} = \mathbf{V} \triangleq (\mathbf{v}_1 \cdots \mathbf{v}_p) \quad (47)$$

(see (14)). We next show that the matrix \mathbf{Q} in (45) coincides with \mathbf{J} in (15). Because of the Gaussian noise, $f(\mathbf{y}; \mathbf{x}) = (2\pi\sigma^2)^{-N/2} \exp(-\|\mathbf{y} - \mathbf{x}\|_2^2/(2\sigma^2))$, and thus we obtain by direct calculation

$$\frac{\delta_i f}{f} = \exp\left(\frac{2\mathbf{v}_i^T(\mathbf{y} - \mathbf{x}) - \|\mathbf{v}_i\|_2^2}{2\sigma^2}\right) - 1$$

and consequently

$$\begin{aligned} (\mathbf{Q})_{i,j} &= \mathbb{E}_{\mathbf{x}} \left\{ \frac{\delta_i f}{f} \frac{\delta_j f}{f} \right\} \\ &= 1 - \exp\left(-\frac{\|\mathbf{v}_i\|_2^2}{2\sigma^2}\right) \mathbb{E}_{\mathbf{x}} \left\{ \exp\left(\frac{\mathbf{v}_i^T(\mathbf{y} - \mathbf{x})}{\sigma^2}\right) \right\} - \exp\left(-\frac{\|\mathbf{v}_j\|_2^2}{2\sigma^2}\right) \mathbb{E}_{\mathbf{x}} \left\{ \exp\left(\frac{\mathbf{v}_j^T(\mathbf{y} - \mathbf{x})}{\sigma^2}\right) \right\} \\ &\quad + \exp\left(-\frac{\|\mathbf{v}_i\|_2^2 + \|\mathbf{v}_j\|_2^2}{2\sigma^2}\right) \mathbb{E}_{\mathbf{x}} \left\{ \exp\left(\frac{(\mathbf{v}_i + \mathbf{v}_j)^T(\mathbf{y} - \mathbf{x})}{\sigma^2}\right) \right\}. \end{aligned}$$

Now $\mathbb{E}_{\mathbf{x}} \left\{ \exp(\mathbf{a}^T(\mathbf{y} - \mathbf{x})) \right\}$ is the moment-generating function of the zero-mean Gaussian random vector $\mathbf{y} - \mathbf{x}$, which equals $\exp(\|\mathbf{a}\|_2^2 \sigma^2/2)$. We thus have

$$\begin{aligned} (\mathbf{Q})_{i,j} &= 1 - \exp\left(-\frac{\|\mathbf{v}_i\|_2^2}{2\sigma^2}\right) \exp\left(\frac{\|\mathbf{v}_i\|_2^2}{2\sigma^2}\right) - \exp\left(-\frac{\|\mathbf{v}_j\|_2^2}{2\sigma^2}\right) \exp\left(\frac{\|\mathbf{v}_j\|_2^2}{2\sigma^2}\right) \\ &\quad + \exp\left(-\frac{\|\mathbf{v}_i\|_2^2 + \|\mathbf{v}_j\|_2^2}{2\sigma^2}\right) \exp\left(\frac{\|\mathbf{v}_i + \mathbf{v}_j\|_2^2}{2\sigma^2}\right) \\ &= -1 + \exp\left(\frac{\mathbf{v}_i^T \mathbf{v}_j}{\sigma^2}\right) \end{aligned} \quad (48)$$

which equals $(\mathbf{J})_{i,j}$ in (15). Inserting (47) and (48) into (46), we obtain (13). Finally, taking the trace of both sides of (13) yields (16).

APPENDIX D

OBTAINING THE CRB FROM THE HCRB

We will demonstrate that the CRB (12) can be obtained as a limit of HCRBs (16) by choosing the test points \mathbf{v}_i according to (17) and letting $t \rightarrow 0$. Since the test points (17) are orthogonal vectors, it

follows from (15) that the matrix \mathbf{J} is diagonal. More specifically, we have

$$\mathbf{J} = \begin{cases} [\exp(t^2/\sigma^2) - 1] \mathbf{I}_S, & \|\mathbf{x}\|_0 = S \\ [\exp(t^2/\sigma^2) - 1] \mathbf{I}_N, & \|\mathbf{x}\|_0 < S. \end{cases}$$

Thus, both for $\|\mathbf{x}\|_0 = S$ and for $\|\mathbf{x}\|_0 < S$, the pseudoinverse of \mathbf{J} is obtained simply by inverting the diagonal entries of \mathbf{J} . From (16), we then obtain

$$\varepsilon(\mathbf{x}; \hat{\mathbf{x}}) \geq \begin{cases} \frac{St^2}{\exp(t^2/\sigma^2) - 1}, & \|\mathbf{x}\|_0 = S \\ \frac{Nt^2}{\exp(t^2/\sigma^2) - 1}, & \|\mathbf{x}\|_0 < S. \end{cases} \quad (49)$$

We now use the third-order Taylor series expansion

$$\exp\left(\frac{t^2}{\sigma^2}\right) = 1 + \frac{t^2}{\sigma^2} + \frac{\tau^4}{2\sigma^4}, \quad \text{where } \tau \in [0, t]. \quad (50)$$

Substituting (50) into (49) yields

$$\varepsilon(\mathbf{x}; \hat{\mathbf{x}}) \geq \begin{cases} \frac{St^2}{t^2/\sigma^2 + \tau^4/(2\sigma^4)}, & \|\mathbf{x}\|_0 = S \\ \frac{Nt^2}{t^2/\sigma^2 + \tau^4/(2\sigma^4)}, & \|\mathbf{x}\|_0 < S. \end{cases} \quad (51)$$

In the limit as $t \rightarrow 0$, $\tau^4 \in [0, t^4]$ decays faster than t^2 , and thus the bound (51) converges to the CRB (12).

The CRB can also be obtained by formally replacing $\exp(t^2/\sigma^2)$ with $1 + t^2/\sigma^2$ in (49). From (50), we have $\exp(t^2/\sigma^2) \geq 1 + t^2/\sigma^2$ for all $t > 0$. This shows that for any $t > 0$, the bound (49) is lower than the CRB (12). Thus, the CRB (which, as shown above, is obtained using the test points (17) in the limit $t \rightarrow 0$) is tighter than any bound that is obtained using the test points (17) for any fixed $t > 0$.

APPENDIX E

PROOF OF THEOREM 5

We will prove the HCRB-type bound in (19). For $\|\mathbf{x}\|_0 < S$, (19) was already demonstrated by the CRB (12), and thus it remains to show (19) for $\|\mathbf{x}\|_0 = S$. This will be done by plugging the test points (18) into the HCRB (16), calculating the resulting bound for an arbitrary constant $t > 0$, and then taking the limit as $t \rightarrow 0$. We will use the following lemma, whose proof is provided at the end of this appendix.

Lemma 12. Let \mathbf{P} be an $(r+1) \times (r+1)$ matrix with the following structure:

$$\mathbf{P} = \begin{pmatrix} a & b\mathbf{1}^T \\ b\mathbf{1} & \mathbf{M} \end{pmatrix} = \begin{pmatrix} a & b & b & b & \cdots & b \\ b & d & c & c & \cdots & c \\ b & c & d & c & \cdots & c \\ b & c & c & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & c \\ b & c & c & \cdots & c & d \end{pmatrix} \quad (52)$$

where $\mathbf{1}$ is the column vector of dimension r whose entries all equal 1, and

$$\mathbf{M} = (d-c)\mathbf{I}_r + c\mathbf{1}\mathbf{1}^T. \quad (53)$$

Let

$$q \triangleq rb^2 - ad - (r-1)ac \quad (54)$$

and assume that

$$d-c \neq 0, \quad d+(r-1)c \neq 0, \quad q \neq 0. \quad (55)$$

Then, \mathbf{P} is nonsingular and its inverse is given by

$$\mathbf{P}^{-1} = \begin{pmatrix} a' & b'\mathbf{1}^T \\ b'\mathbf{1} & \mathbf{M}' \end{pmatrix} = \begin{pmatrix} a' & b' & b' & b' & \cdots & b' \\ b' & d' & c' & c' & \cdots & c' \\ b' & c' & d' & c' & \cdots & c' \\ b' & c' & c' & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & c' \\ b' & c' & c' & \cdots & c' & d' \end{pmatrix} \quad (56)$$

where $\mathbf{M}' = (d'-c')\mathbf{I}_r + c'\mathbf{1}\mathbf{1}^T$ and

$$a' = -\frac{d+(r-1)c}{q}, \quad b' = \frac{b}{q}, \quad c' = \frac{ac-b^2}{(d-c)q}, \quad d' = \frac{(r-1)b^2 - (r-2)ac - ad}{(d-c)q}. \quad (57)$$

Let $\|\mathbf{x}\|_0 = S$, and assume for concreteness and without loss of generality that $\text{supp}(\mathbf{x}) = \{1, \dots, S\}$ and that ξ , the smallest (in magnitude) nonzero component of \mathbf{x} , is the S th entry. A direct calculation of the matrix \mathbf{J} in (15) based on the test points (18) then yields

$$\mathbf{J} = \begin{pmatrix} a\mathbf{I}_{S-1} & \mathbf{0}_{(S-1) \times (r+1)} \\ \mathbf{0}_{(r+1) \times (S-1)} & \mathbf{P} \end{pmatrix}.$$

Here, \mathbf{P} is an $(r+1) \times (r+1)$ matrix, where $r = N - S$, having the structure (52) with entries

$$a = e^{t^2/\sigma^2} - 1, \quad b = e^{-t\xi/\sigma^2} - 1, \quad c = e^{\xi^2/\sigma^2} - 1, \quad d = e^{(t^2+\xi^2)/\sigma^2} - 1. \quad (58)$$

We now apply Lemma 12 in order to show that \mathbf{J} is nonsingular and to calculate its inverse. More precisely, it suffices to calculate the inverse for all but a finite number of values of t , since any finite set of values can simply be excluded from consideration when t tends to 0. When applying Lemma 12, we first have to verify that the conditions (55) hold for all but a finite number of values of t . By substituting (58), it is seen that the left-hand sides of (55) are nonconstant entire functions of t , and thus have a finite number of roots on any compact set of values of t . By Lemma 12, this implies that \mathbf{J} is nonsingular for all but a finite number of values of values of t , and that the inverse (if it exists) is given by

$$\mathbf{J}^{-1} = \begin{pmatrix} \frac{1}{a}\mathbf{I}_{S-1} & \mathbf{0}_{(S-1) \times (r+1)} \\ \mathbf{0}_{(r+1) \times (S-1)} & \mathbf{P}^{-1} \end{pmatrix} \quad (59)$$

where \mathbf{P}^{-1} is given by (56) and (57), again with $r = N - S$. Next, we observe that for our choice of test points (18),

$$\mathbf{V}^T \mathbf{V} = \begin{pmatrix} t^2 \mathbf{I}_{S-1} & \mathbf{0}_{(S-1) \times (r+1)} \\ \mathbf{0}_{(r+1) \times (S-1)} & \tilde{\mathbf{P}} \end{pmatrix} \quad (60)$$

where $\tilde{\mathbf{P}}$ is an $(r+1) \times (r+1)$ matrix having the structure (52) with entries

$$\tilde{a} = t^2, \quad \tilde{b} = -t\xi, \quad \tilde{c} = \xi^2, \quad \tilde{d} = t^2 + \xi^2.$$

Using (16) together with (59) and (60), a direct calculation yields

$$\begin{aligned} \varepsilon(\mathbf{x}; \hat{\mathbf{x}}) &\geq \text{tr}(\mathbf{V}\mathbf{J}^\dagger\mathbf{V}^T) = \text{tr}(\mathbf{V}^T\mathbf{V}\mathbf{J}^{-1}) = \sum_{i=1}^N \sum_{j=1}^N (\mathbf{V}^T\mathbf{V})_{i,j} (\mathbf{J}^{-1})_{i,j} \\ &= (S-1) \frac{t^2}{a} + t^2 a' - 2rt\xi b' + r(r-1)\xi^2 c' + r(t^2 + \xi^2) d'. \end{aligned} \quad (61)$$

We now take the limit $t \rightarrow 0$ in (61). For the first term, we obtain

$$(S-1) \frac{t^2}{a} = (S-1) \frac{t^2}{e^{t^2/\sigma^2} - 1} = (S-1) \frac{t^2}{t^2/\sigma^2 + o(t^2)} \longrightarrow (S-1)\sigma^2 \quad (62)$$

where we have expanded e^{t^2/σ^2} into a second-order Taylor series. Here, $o(f(t))$ indicates terms which are negligible compared with $f(t)$ when $t \rightarrow 0$, i.e., $\lim_{t \rightarrow 0} o(f(t))/f(t) = 0$. To find the limit of the second

term in (61), $t^2 a' = -(t^2/q)[d + (r-1)c]$, we first consider the reciprocal of the first factor, t^2/q . We have

$$\frac{q}{t^2} = \frac{1}{t^2} [r(e^{-t\xi/\sigma^2} - 1)^2 - (e^{t^2/\sigma^2} - 1)(e^{(t^2+\xi^2)/\sigma^2} - 1) - (r-1)(e^{t^2/\sigma^2} - 1)(e^{\xi^2/\sigma^2} - 1)].$$

Expanding some of the t -dependent exponentials into Taylor series, dropping higher-order terms, and simplifying, we obtain

$$\begin{aligned} \frac{q}{t^2} &= \frac{1}{t^2} \left[r \left(\frac{-t\xi}{\sigma^2} + o(t) \right)^2 - \left(\frac{t^2}{\sigma^2} + o(t^2) \right) (e^{(t^2+\xi^2)/\sigma^2} - 1) - (r-1) \left(\frac{t^2}{\sigma^2} + o(t^2) \right) (e^{\xi^2/\sigma^2} - 1) \right] \\ &\rightarrow r \frac{\xi^2}{\sigma^4} - \frac{1}{\sigma^2} (e^{\xi^2/\sigma^2} - 1) - (r-1) \frac{1}{\sigma^2} (e^{\xi^2/\sigma^2} - 1) = \frac{r}{\sigma^4} [\xi^2 - \sigma^2 (e^{\xi^2/\sigma^2} - 1)]. \end{aligned} \quad (63)$$

For the second factor, we obtain

$$d + (r-1)c = e^{(t^2+\xi^2)/\sigma^2} - 1 + (r-1)(e^{\xi^2/\sigma^2} - 1) \rightarrow r(e^{\xi^2/\sigma^2} - 1). \quad (64)$$

Then, using (63) and (64), it is seen that the second term in (61) converges to

$$t^2 a' = -\frac{t^2}{q} [d + (r-1)c] \rightarrow -\frac{r(e^{\xi^2/\sigma^2} - 1)}{\frac{r}{\sigma^4} [\xi^2 - \sigma^2 (e^{\xi^2/\sigma^2} - 1)]} = \sigma^2 \left[1 + \frac{\xi^2}{\sigma^2 (e^{\xi^2/\sigma^2} - 1) - \xi^2} \right]. \quad (65)$$

Next, we consider the third term in (61), $-2rt\xi b'$, which can be written as $-2r\xi \frac{b/t}{q/t^2}$. We have

$$\frac{b}{t} = \frac{1}{t} (e^{-t\xi/\sigma^2} - 1) = \frac{1}{t} \left(\frac{-t\xi}{\sigma^2} + o(t) \right) \rightarrow -\frac{\xi}{\sigma^2}.$$

Combining with (63), we obtain

$$-2rt\xi b' \rightarrow 2r\xi \frac{\xi/\sigma^2}{\frac{r}{\sigma^4} [\xi^2 - \sigma^2 (e^{\xi^2/\sigma^2} - 1)]} = \frac{2\sigma^2 \xi^2}{\xi^2 - \sigma^2 (e^{\xi^2/\sigma^2} - 1)}. \quad (66)$$

The fourth and fifth terms in (61) have to be calculated together because each of them by itself diverges.

The sum of these terms is

$$\begin{aligned} r(r-1)\xi^2 c' + r(t^2 + \xi^2)d' &= \frac{r}{(d-c)q} [(r-1)\xi^2(ac - b^2) + (t^2 + \xi^2)[(r-1)b^2 - (r-2)ac - ad]] \\ &= \frac{r}{(d-c)q} [-\xi^2 a(d-c) + t^2[(r-1)b^2 - (r-2)ac - ad]] \\ &= -\frac{r\xi^2 a}{q} + \frac{rt^2}{(d-c)q} (q + ac - b^2) \\ &= \underbrace{-\frac{r\xi^2 a}{q}}_{z_1} + \underbrace{\frac{rt^2}{d-c}}_{z_2} + \underbrace{\frac{rt^2}{(d-c)q} (ac - b^2)}_{z_3}. \end{aligned} \quad (67)$$

Using (63), z_1 in (67) becomes

$$z_1 = -\frac{r\xi^2 a/t^2}{q/t^2} = -r\xi^2 \frac{(e^{t^2/\sigma^2} - 1)/t^2}{q/t^2} \rightarrow -r\xi^2 \frac{1/\sigma^2}{\frac{r}{\sigma^4} [\xi^2 - \sigma^2(e^{\xi^2/\sigma^2} - 1)]} = -\frac{\sigma^2 \xi^2}{\xi^2 - \sigma^2(e^{\xi^2/\sigma^2} - 1)}. \quad (68)$$

Furthermore, a direct calculation yields

$$z_2 = \frac{rt^2}{e^{(t^2+\xi^2)/\sigma^2} - e^{\xi^2/\sigma^2}} = r e^{-\xi^2/\sigma^2} \frac{t^2}{e^{t^2/\sigma^2} - 1} \rightarrow r\sigma^2 e^{-\xi^2/\sigma^2}. \quad (69)$$

To take the limit of z_3 , first note that

$$\begin{aligned} \frac{ac - b^2}{d - c} &= \frac{(e^{t^2/\sigma^2} - 1)(e^{\xi^2/\sigma^2} - 1) - (e^{-t\xi/\sigma^2} - 1)^2}{e^{(t^2+\xi^2)/\sigma^2} - e^{\xi^2/\sigma^2}} \\ &\rightarrow \frac{(t^2/\sigma^2)(e^{\xi^2/\sigma^2} - 1) - (-t\xi/\sigma^2)^2}{e^{\xi^2/\sigma^2} t^2/\sigma^2} = \frac{\sigma^2(e^{\xi^2/\sigma^2} - 1) - \xi^2}{\sigma^2 e^{\xi^2/\sigma^2}}. \end{aligned}$$

Together with (63), we thus have

$$z_3 = r \frac{t^2}{q} \frac{ac - b^2}{d - c} \rightarrow r \frac{1}{\frac{r}{\sigma^4} [\xi^2 - \sigma^2(e^{\xi^2/\sigma^2} - 1)]} \frac{\sigma^2(e^{\xi^2/\sigma^2} - 1) - \xi^2}{\sigma^2 e^{\xi^2/\sigma^2}} = -\sigma^2 e^{-\xi^2/\sigma^2}. \quad (70)$$

Adding the limits of z_1 , z_2 , and z_3 in (68)–(70), we find that the sum of the fourth and fifth terms in (61) converges to

$$z_1 + z_2 + z_3 \rightarrow \frac{-\sigma^2 \xi^2}{\xi^2 - \sigma^2(e^{\xi^2/\sigma^2} - 1)} + (r-1)\sigma^2 e^{-\xi^2/\sigma^2}. \quad (71)$$

Finally, adding the limits of all terms in (61) as given by (62), (65), (66), and (71) and simplifying, we obtain the following result for the limit of the bound (61) for $t \rightarrow 0$:

$$\varepsilon(\mathbf{x}; \hat{\mathbf{x}}) \geq S\sigma^2 + (r-1)\sigma^2 e^{-\xi^2/\sigma^2}.$$

This equals (19), as claimed.

Proof of Lemma 12: We first calculate the inverse of \mathbf{M} in (53). Applying the Sherman–Morrison–Woodbury formula [35, §2.8]

$$(\mathbf{A} + c\mathbf{u}\mathbf{v}^T)^{-1} = \mathbf{A}^{-1} - \frac{c}{1 + c\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}} \mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}$$

to (53) and simplifying yields

$$\mathbf{M}^{-1} = \frac{1}{d-c} \mathbf{I}_r - \frac{c}{(d-c)[d + (r-1)c]} \mathbf{1}\mathbf{1}^T. \quad (72)$$

Next, we invoke the block inversion lemma [35, §2.8]

$$\begin{pmatrix} \mathbf{A} & \mathbf{B}^T \\ \mathbf{B} & \mathbf{M} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{E}^{-1} & -\mathbf{E}^{-1}\mathbf{B}^T\mathbf{M}^{-1} \\ -\mathbf{M}^{-1}\mathbf{B}\mathbf{E}^{-1} & \mathbf{M}^{-1} + \mathbf{M}^{-1}\mathbf{B}\mathbf{E}^{-1}\mathbf{B}^T\mathbf{M}^{-1} \end{pmatrix}, \quad \text{with } \mathbf{E} \triangleq \mathbf{A} - \mathbf{B}^T\mathbf{M}^{-1}\mathbf{B}.$$

Specializing to $\mathbf{A} = a$ and $\mathbf{B} = b\mathbf{1}$ as is appropriate for \mathbf{P} in (52), we obtain for the inverse of \mathbf{P}

$$\mathbf{P}^{-1} = \begin{pmatrix} 1/e & -(b/e)\mathbf{1}^T\mathbf{M}^{-1} \\ -(b/e)\mathbf{M}^{-1}\mathbf{1} & \mathbf{M}^{-1} + (b^2/e)\mathbf{M}^{-1}\mathbf{1}\mathbf{1}^T\mathbf{M}^{-1} \end{pmatrix}, \quad \text{with } e \triangleq a - b^2\mathbf{1}^T\mathbf{M}^{-1}\mathbf{1}. \quad (73)$$

We now develop the various blocks of \mathbf{P}^{-1} by using the expression of \mathbf{M}^{-1} in (72). We first consider the upper-left block, $1/e$. We have

$$e = a - \frac{b^2}{d-c} \mathbf{1}^T \left[\mathbf{I}_r - \frac{c}{d+(r-1)c} \mathbf{1}\mathbf{1}^T \right] \mathbf{1} = a - \frac{b^2}{d-c} \left[r - \frac{cr^2}{d+(r-1)c} \right] = \frac{ad + (r-1)ac - rb^2}{d+(r-1)c}.$$

Thus, using the definitions in (54) and (57) yields

$$\frac{1}{e} = -\frac{d+(r-1)c}{q} = a' \quad (74)$$

which proves the validity of the upper-left entry of \mathbf{P}^{-1} in (56). Next, using (72) and (74) and simplifying, the upper-right block in (73) becomes

$$-\frac{b}{e} \mathbf{1}^T \mathbf{M}^{-1} = -ba' \left[\frac{1}{d-c} - \frac{rc}{(d-c)[d+(r-1)c]} \right] \mathbf{1}^T = -\frac{ba'}{d+(r-1)c} \mathbf{1}^T = \frac{b}{q} \mathbf{1}^T = b' \mathbf{1}^T.$$

Thus, we have shown the validity of the first row and first column of \mathbf{P}^{-1} in (56). Finally, to develop the remaining block $\mathbf{M}^{-1} + (b^2/e)\mathbf{M}^{-1}\mathbf{1}\mathbf{1}^T\mathbf{M}^{-1}$ in (73), we first calculate

$$\mathbf{u} \triangleq \mathbf{M}^{-1}\mathbf{1} = \frac{1}{d-c} \left[1 - \frac{rc}{d+(r-1)c} \right] \mathbf{1} = \frac{1}{d+(r-1)c} \mathbf{1}. \quad (75)$$

We then have

$$\mathbf{M}^{-1} + \frac{b^2}{e} \mathbf{M}^{-1}\mathbf{1}\mathbf{1}^T\mathbf{M}^{-1} = \mathbf{M}^{-1} + b^2 a' \mathbf{u}\mathbf{u}^T = \frac{1}{d-c} \mathbf{I}_r - \frac{1}{d+(r-1)c} \left[\frac{c}{d-c} + \frac{b^2}{q} \right] \mathbf{1}\mathbf{1}^T \quad (76)$$

where (72), (75), and the definition of a' in (57) were used. Using the definition of q in (54) and simplifying, the factor in brackets can be written as

$$\frac{c}{d-c} + \frac{b^2}{q} = \frac{cq + (d-c)b^2}{(d-c)q} = \frac{[d+(r-1)c](b^2-ac)}{(d-c)q}.$$

Substituting back into (76), we obtain

$$\mathbf{M}^{-1} + \frac{b^2}{e} \mathbf{M}^{-1}\mathbf{1}\mathbf{1}^T\mathbf{M}^{-1} = \frac{1}{d-c} \mathbf{I}_r - \frac{b^2-ac}{(d-c)q} \mathbf{1}\mathbf{1}^T = \frac{1}{d-c} \mathbf{I}_r + c' \mathbf{1}\mathbf{1}^T.$$

Thus, within the $r \times r$ lower-right block of \mathbf{P}^{-1} , the off-diagonal entries all equal c' , as required. Furthermore, the diagonal entries in this block are given by

$$\frac{1}{d-c} - \frac{b^2-ac}{(d-c)q} = \frac{(r-1)b^2 - ad - (r-2)ac}{(d-c)q} = d'$$

which completes the proof of the lemma. \square

APPENDIX F

PROOF OF LEMMA 6

Let $\mathbf{x} \in \mathcal{X}_S$ with $\|\mathbf{x}\|_0 = S$ and consider a fixed $k \in \text{supp}(\mathbf{x})$. We have to show that a solution of (20), i.e.,

$$\arg \min_{\hat{x}(\cdot) \in \mathcal{U}^k} \mathbf{E}_{\mathbf{x}} \{ (\hat{x}(\mathbf{y}))^2 \}, \quad \text{with } \mathcal{U}^k = \{ \hat{x}(\cdot) \mid \mathbf{E}_{\tilde{\mathbf{x}}} \{ \hat{x}(\mathbf{y}) \} = \tilde{x}_k \text{ for all } \tilde{\mathbf{x}} \in \mathcal{X}_S \} \quad (77)$$

is given by $\hat{x}_k^{(\mathbf{x})}(\mathbf{y}) = y_k$. Let $\varepsilon_0 \triangleq \min_{\hat{x}(\cdot) \in \mathcal{U}^k} \mathbf{E}_{\mathbf{x}} \{ (\hat{x}(\mathbf{y}))^2 \}$ denote the mean power of the LMVU estimator defined by (77). We will show that $\varepsilon_0 \geq \sigma^2 + x_k^2$ and, furthermore, that $\sigma^2 + x_k^2$ is achieved by the estimator $\hat{x}_k^{(\mathbf{x})}(\mathbf{y}) = y_k$.

Let $\mathcal{C}_{\mathbf{x}}^k$ denote the set of all S -sparse vectors $\tilde{\mathbf{x}}$ which equal \mathbf{x} except possibly for the k th component, i.e., $\mathcal{C}_{\mathbf{x}}^k \triangleq \{ \tilde{\mathbf{x}} \in \mathcal{X}_S \mid \tilde{x}_l = x_l \text{ for all } l \neq k \}$. Consider the modified optimization problem

$$\arg \min_{\hat{x}(\cdot) \in \mathcal{U}_{\mathbf{x}}^k} \mathbf{E}_{\mathbf{x}} \{ (\hat{x}(\mathbf{y}))^2 \}, \quad \text{with } \mathcal{U}_{\mathbf{x}}^k \triangleq \{ \hat{x}(\cdot) \mid \mathbf{E}_{\tilde{\mathbf{x}}} \{ \hat{x}(\mathbf{y}) \} = \tilde{x}_k \text{ for all } \tilde{\mathbf{x}} \in \mathcal{C}_{\mathbf{x}}^k \} \quad (78)$$

and let $\varepsilon'_0 \triangleq \min_{\hat{x}(\cdot) \in \mathcal{U}_{\mathbf{x}}^k} \mathbf{E}_{\mathbf{x}} \{ (\hat{x}(\mathbf{y}))^2 \}$ denote the mean power of the estimator defined by (78). Note the distinction between \mathcal{U}^k and $\mathcal{U}_{\mathbf{x}}^k$: \mathcal{U}^k is the set of estimators of x_k which are unbiased for all $\tilde{\mathbf{x}} \in \mathcal{X}_S$ whereas $\mathcal{U}_{\mathbf{x}}^k$ is the set of estimators of x_k which are unbiased for all $\tilde{\mathbf{x}} \in \mathcal{X}_S$ which equal a given, fixed \mathbf{x} except possibly for the k th component. Therefore, the unbiasedness requirement expressed by \mathcal{U}^k is more restrictive than that expressed by $\mathcal{U}_{\mathbf{x}}^k$, i.e., $\mathcal{U}^k \subseteq \mathcal{U}_{\mathbf{x}}^k$, which implies that

$$\varepsilon'_0 \leq \varepsilon_0. \quad (79)$$

We will use the following result, which is proved at the end of this appendix.

Lemma 13. *Given an arbitrary estimator $\hat{x}(\mathbf{y}) \in \mathcal{U}_{\mathbf{x}}^k$, the estimator*

$$\hat{x}_c(y_k) \triangleq \mathbf{E}_{\mathbf{x}} \{ \hat{x}(\mathbf{y}) \mid y_k \} \quad (80)$$

also satisfies the constraint $\hat{x}_c(y_k) \in \mathcal{U}_{\mathbf{x}}^k$, and its mean power does not exceed that obtained by \hat{x} , i.e., $\mathbf{E}_{\mathbf{x}} \{ (\hat{x}_c(y_k))^2 \} \leq \mathbf{E}_{\mathbf{x}} \{ (\hat{x}(\mathbf{y}))^2 \}$.

Thus, to each estimator $\hat{x}(\mathbf{y}) \in \mathcal{U}_{\mathbf{x}}^k$ which depends on the entire observation \mathbf{y} , we can always find at least one estimator $\hat{x}_c(y_k) \in \mathcal{U}_{\mathbf{x}}^k$ which depends only on the observation component y_k and is at least as good. Therefore, with no loss in optimality, we can restrict the optimization problem (78) to estimators $\hat{x}(y_k) \in \mathcal{U}_{\mathbf{x}}^k$ which depend on \mathbf{y} only via its k th component y_k . This means that (78) can be replaced by

$$\arg \min_{\hat{x}(\cdot) \in \tilde{\mathcal{U}}^k} \mathbb{E}_{\mathbf{x}} \{ (\hat{x}(y_k))^2 \}, \quad \text{with } \tilde{\mathcal{U}}^k \triangleq \{ \hat{x}(\cdot) \mid \mathbb{E}_{\tilde{\mathbf{x}}} \{ \hat{x}(y_k) \} = \tilde{x}_k \text{ for all } \tilde{\mathbf{x}} \in \mathbb{R}^N \}. \quad (81)$$

Note that in the definition of $\tilde{\mathcal{U}}^k$, we can use the requirement $\tilde{\mathbf{x}} \in \mathbb{R}^N$ instead of $\tilde{\mathbf{x}} \in \mathcal{C}_{\mathbf{x}}^k$ since the expectation $\mathbb{E}_{\tilde{\mathbf{x}}} \{ \hat{x}(y_k) \}$ does not depend on the components \tilde{x}_l with $l \neq k$. The corresponding minimum mean power $\min_{\hat{x}(\cdot) \in \tilde{\mathcal{U}}^k} \mathbb{E}_{\mathbf{x}} \{ (\hat{x}(y_k))^2 \}$ is still equal to ε'_0 . However, the new problem (81) is equivalent to the classical problem of finding the LMVU estimator of a scalar x_k based on the observation $y_k = x_k + n_k$, with $n_k \sim \mathcal{N}(0, \sigma^2)$. A solution of this latter problem is the estimator $\hat{x}(y_k) = y_k$, whose variance and mean power are σ^2 and $\sigma^2 + x_k^2$, respectively [10]. Thus, a solution of (81) or, equivalently, of (78) is the trivial estimator $\hat{x}(y_k) = y_k$, and

$$\varepsilon'_0 = \sigma^2 + x_k^2. \quad (82)$$

Combining (79) and (82), we see that the minimum mean power for our original optimization problem (77) satisfies

$$\varepsilon_0 \geq \sigma^2 + x_k^2.$$

As we have shown, this lower bound is achieved by the estimator $\hat{x}(y_k) = y_k$. In addition, $\hat{x}(y_k) = y_k$ is an element of \mathcal{U}^k , the constraint set of (77). Therefore, it is a solution of (77).

Proof of Lemma 13: Consider a fixed $\mathbf{x} \in \mathcal{X}_S$ and an estimator $\hat{x}(\mathbf{y}) \in \mathcal{U}_{\mathbf{x}}^k$. In order to show the first statement of the lemma, $\hat{x}_c(y_k) \in \mathcal{U}_{\mathbf{x}}^k$, we first note that

$$\mathbb{E}_{\mathbf{x}} \{ \hat{x}(\mathbf{y}) | y_k \} = \mathbb{E}_{\tilde{\mathbf{x}}} \{ \hat{x}(\mathbf{y}) | y_k \}, \quad \text{for any } \tilde{\mathbf{x}} \in \mathcal{C}_{\mathbf{x}}^k. \quad (83)$$

We now have for $\tilde{\mathbf{x}} \in \mathcal{C}_{\mathbf{x}}^k$

$$\mathbb{E}_{\tilde{\mathbf{x}}} \{ \hat{x}_c(y_k) \} \stackrel{(a)}{=} \mathbb{E}_{\tilde{\mathbf{x}}} \{ \mathbb{E}_{\mathbf{x}} \{ \hat{x}(\mathbf{y}) | y_k \} \} \stackrel{(b)}{=} \mathbb{E}_{\tilde{\mathbf{x}}} \{ \mathbb{E}_{\tilde{\mathbf{x}}} \{ \hat{x}(\mathbf{y}) | y_k \} \} \stackrel{(c)}{=} \mathbb{E}_{\tilde{\mathbf{x}}} \{ \hat{x}(\mathbf{y}) \} \stackrel{(d)}{=} \tilde{x}_k$$

where we used the definition (80) in (a), the identity (83) in (b), the law of total probability [36] in (c), and our assumption $\hat{x}(\mathbf{y}) \in \mathcal{U}_{\mathbf{x}}^k$ in (d). Thus, $\hat{x}_c(y_k) \in \mathcal{U}_{\mathbf{x}}^k$.

Next, the inequality $\mathbb{E}_{\mathbf{x}} \{ (\hat{x}_c(y_k))^2 \} \leq \mathbb{E}_{\mathbf{x}} \{ (\hat{x}(\mathbf{y}))^2 \}$ is proved as follows:

$$\mathbb{E}_{\mathbf{x}} \{ (\hat{x}(\mathbf{y}))^2 \} \stackrel{(a)}{=} \mathbb{E}_{\mathbf{x}} \{ \mathbb{E}_{\mathbf{x}} \{ (\hat{x}(\mathbf{y}))^2 | y_k \} \} \stackrel{(b)}{\geq} \mathbb{E}_{\mathbf{x}} \{ (\mathbb{E}_{\mathbf{x}} \{ \hat{x}(\mathbf{y}) | y_k \})^2 \} \stackrel{(c)}{=} \mathbb{E}_{\mathbf{x}} \{ (\hat{x}_c(y_k))^2 \}$$

where we used the law of total probability in (a), Jensen's inequality for convex functions [29] in (b), and the definition (80) in (c). \square

APPENDIX G

PROOF OF LEMMA 7

We wish to solve the componentwise optimization problem (25), i.e., $\arg \min_{\hat{x}(\cdot) \in \mathcal{U}^k \cap \mathcal{A}_x^k} \mathbb{E}_x \{ (\hat{x}(\mathbf{y}))^2 \}$, for $k \notin \text{supp}(\mathbf{x})$. Note that $x_k = 0$ and, thus, the variance equals the mean power $\mathbb{E}_x \{ (\hat{x}(\mathbf{y}))^2 \}$.

We first observe that the constraint $\hat{x} \in \mathcal{A}_x^k$ implies that the estimator \hat{x} is unbiased, and thus $\mathcal{U}^k \cap \mathcal{A}_x^k = \mathcal{A}_x^k$. Indeed, using (21) and $x_k = 0$, we have

$$\begin{aligned}
\mathbb{E}_x \{ \hat{x}(\mathbf{y}) \} &= \underbrace{\mathbb{E}_x \{ y_k \}}_{x_k (=0)} + \mathbb{E}_x \{ \hat{x}'(\mathbf{y}) \} \\
&= x_k + \frac{1}{(2\pi\sigma^2)^{N/2}} \int_{\mathbb{R}^N} \hat{x}'(\mathbf{y}) e^{-\|\mathbf{y}-\mathbf{x}\|_2^2/(2\sigma^2)} d\mathbf{y} \\
&= x_k + \frac{1}{(2\pi\sigma^2)^{N/2}} \int_{\mathbb{R}^{N-1}} e^{-\|\mathbf{y}_{\sim k}-\mathbf{x}_{\sim k}\|_2^2/(2\sigma^2)} \underbrace{\left[\int_{-\infty}^{\infty} \hat{x}'(\mathbf{y}) e^{-(y_k-0)^2/(2\sigma^2)} dy_k \right]}_0 d\mathbf{y}_{\sim k} \\
&= x_k
\end{aligned} \tag{84}$$

where $\mathbf{x}_{\sim k}$ and $\mathbf{y}_{\sim k}$ denote the $(N-1)$ -dimensional vectors obtained from \mathbf{x} and \mathbf{y} by removing the k th component x_k and y_k , respectively, and the result in (84) follows because $\int_{-\infty}^{\infty} \hat{x}'(\mathbf{y}) e^{-y_k^2/(2\sigma^2)} dy_k = 0$ due to the odd symmetry assumption (23). Thus, we can replace the constraint $\hat{x}(\cdot) \in \mathcal{U}^k \cap \mathcal{A}_x^k$ in (25) by $\hat{x}(\cdot) \in \mathcal{A}_x^k$.

A solution of (25) can now be found by noting that for any $\hat{x}(\cdot) \in \mathcal{A}_x^k$, we have

$$\begin{aligned}
\mathbb{E}_x \{ (\hat{x}(\mathbf{y}))^2 \} &= \frac{1}{(2\pi\sigma^2)^{N/2}} \int_{\mathbb{R}^N} (y_k + \hat{x}'(\mathbf{y}))^2 e^{-\|\mathbf{y}-\mathbf{x}\|_2^2/(2\sigma^2)} d\mathbf{y} \\
&= \frac{1}{(2\pi\sigma^2)^{N/2}} \int_{\mathbb{R}^N} y_k^2 e^{-\|\mathbf{y}-\mathbf{x}\|_2^2/(2\sigma^2)} d\mathbf{y} \\
&\quad + \frac{1}{(2\pi\sigma^2)^{N/2}} \int_{\mathbb{R}^N} [2y_k \hat{x}'(\mathbf{y}) + (\hat{x}'(\mathbf{y}))^2] e^{-\|\mathbf{y}-\mathbf{x}\|_2^2/(2\sigma^2)} d\mathbf{y}.
\end{aligned}$$

The first term is equal to $\sigma^2 + x_k^2 = \sigma^2$. Regarding the second term, let \mathbf{y}_k be the length- $(S+1)$ subvector of \mathbf{y} that comprises all y_l with $l \in \{k\} \cup \text{supp}(\mathbf{x})$. Due to (24), $\hat{x}'(\mathbf{y})$ depends only on \mathbf{y}_k and can thus be written (with some abuse of notation) as $\hat{x}'(\mathbf{y}_k)$. Let $\bar{\mathbf{y}}_k$ denote the complementary subvector of \mathbf{y} , i.e., the length- $(N-S-1)$ subvector comprising all y_l with $l \notin \{k\} \cup \text{supp}(\mathbf{x})$. Furthermore, let \mathbf{x}_k and

$\bar{\mathbf{x}}_k$ denote the analogous subvectors of \mathbf{x} . The second integral can then be written as the product

$$\begin{aligned} & \frac{1}{(2\pi\sigma^2)^{(S+1)/2}} \int_{\mathbb{R}^{S+1}} [2y_k \hat{x}'(\mathbf{y}_k) + (\hat{x}'(\mathbf{y}_k))^2] e^{-\|\mathbf{y}_k - \mathbf{x}_k\|_2^2 / (2\sigma^2)} d\mathbf{y}_k \\ & \quad \times \frac{1}{(2\pi\sigma^2)^{(N-S-1)/2}} \int_{\mathbb{R}^{N-S-1}} e^{-\|\bar{\mathbf{y}}_k - \bar{\mathbf{x}}_k\|_2^2 / (2\sigma^2)} d\bar{\mathbf{y}}_k. \end{aligned}$$

The second factor is 1, and thus we have

$$\mathbb{E}_{\mathbf{x}}\{(\hat{x}(\mathbf{y}))^2\} = \sigma^2 + \frac{1}{(2\pi\sigma^2)^{(S+1)/2}} \int_{\mathbb{R}^{S+1}} [2y_k \hat{x}'(\mathbf{y}_k) + (\hat{x}'(\mathbf{y}_k))^2] e^{-\|\mathbf{y}_k - \mathbf{x}_k\|_2^2 / (2\sigma^2)} d\mathbf{y}_k. \quad (85)$$

Using the symmetry property (23), this can be written as

$$\mathbb{E}_{\mathbf{x}}\{(\hat{x}(\mathbf{y}))^2\} = \sigma^2 + \frac{2}{(2\pi\sigma^2)^{(S+1)/2}} \int_{\mathbb{R}_+^{S+1}} [2\hat{x}'(\mathbf{y}_k) b(\mathbf{y}_k) + (\hat{x}'(\mathbf{y}_k))^2 c(\mathbf{y}_k)] d\mathbf{y}_k, \quad (86)$$

with

$$b(\mathbf{y}_k) \triangleq y_k e^{-y_k^2 / (2\sigma^2)} \prod_{l \in \text{supp}(\mathbf{x})} [e^{-(y_l - x_l)^2 / (2\sigma^2)} - e^{-(y_l + x_l)^2 / (2\sigma^2)}] \quad (87)$$

$$c(\mathbf{y}_k) \triangleq e^{-y_k^2 / (2\sigma^2)} \prod_{l \in \text{supp}(\mathbf{x})} [e^{-(y_l - x_l)^2 / (2\sigma^2)} + e^{-(y_l + x_l)^2 / (2\sigma^2)}]. \quad (88)$$

We sketch the derivation of expressions (87) and (88) by showing the first of $S + 1$ similar sequential calculations. For simplicity of notation and without loss of generality, we assume for this derivation that $k = 1$ and $\text{supp}(\mathbf{x}) = \{2, \dots, S + 1\}$. The integral in (85) then becomes

$$\begin{aligned} & \int_{\mathbb{R}^{S+1}} [2y_k \hat{x}'(\mathbf{y}_k) + (\hat{x}'(\mathbf{y}_k))^2] e^{-\|\mathbf{y}_k - \mathbf{x}_k\|_2^2 / (2\sigma^2)} d\mathbf{y}_k \\ & \quad = \int_{\mathbb{R}^{S+1}} [2y_1 \hat{x}'(\mathbf{y}_1) + (\hat{x}'(\mathbf{y}_1))^2] \left[\prod_{l=1}^{S+1} e^{-(y_l - x_l)^2 / (2\sigma^2)} \right] d\mathbf{y}_1. \end{aligned} \quad (89)$$

The $\int_{\mathbb{R}^{S+1}}$ integration can now be represented as $\int_{\mathbb{R}^S \times (\mathbb{R}_+ \cup \mathbb{R}_-)}$, where the component $\int_{\mathbb{R}^S}$ refers to y_1, \dots, y_S and the component $\int_{\mathbb{R}_+ \cup \mathbb{R}_-}$ refers to y_{S+1} . Then (89) can be further processed as

$$\begin{aligned} & \int_{\mathbb{R}^S \times \mathbb{R}_+} [2y_1 \hat{x}'(\mathbf{y}_1) + (\hat{x}'(\mathbf{y}_1))^2] \left[\prod_{l=1}^{S+1} e^{-(y_l - x_l)^2 / (2\sigma^2)} \right] d\mathbf{y}_1 \\ & \quad + \int_{\mathbb{R}^S \times \mathbb{R}_-} [2y_1 \hat{x}'(\mathbf{y}_1) + (\hat{x}'(\mathbf{y}_1))^2] \left[\prod_{l=1}^{S+1} e^{-(y_l - x_l)^2 / (2\sigma^2)} \right] d\mathbf{y}_1 \\ & \stackrel{(*)}{=} \int_{\mathbb{R}^S \times \mathbb{R}_+} [2y_1 \hat{x}'(\mathbf{y}_1) (e^{-(y_{S+1} - x_{S+1})^2 / (2\sigma^2)} - e^{-(y_{S+1} + x_{S+1})^2 / (2\sigma^2)}) \\ & \quad + (\hat{x}'(\mathbf{y}_1))^2 (e^{-(y_{S+1} - x_{S+1})^2 / (2\sigma^2)} + e^{-(y_{S+1} + x_{S+1})^2 / (2\sigma^2)})] \left[\prod_{l=1}^S e^{-(y_l - x_l)^2 / (2\sigma^2)} \right] d\mathbf{y}_1 \end{aligned}$$

where the odd symmetry property (23) was used in (*). After performing this type of manipulation S times, the integral is obtained in the form

$$\int_{\mathbb{R} \times \mathbb{R}_+^S} \left[2y_1 \hat{x}'(\mathbf{y}_1) \prod_{l=2}^{S+1} (e^{-(y_l-x_l)^2/(2\sigma^2)} - e^{-(y_l+x_l)^2/(2\sigma^2)}) \right. \\ \left. + (\hat{x}'(\mathbf{y}_1))^2 \prod_{l=2}^{S+1} (e^{-(y_l-x_l)^2/(2\sigma^2)} + e^{-(y_l+x_l)^2/(2\sigma^2)}) \right] e^{-y_1^2/(2\sigma^2)} d\mathbf{y}_1$$

where $x_1 = 0$ was used. With $y_1 \hat{x}'(y_1, \dots) = (-y_1) \hat{x}'(-y_1, \dots)$, this becomes further

$$\int_{\mathbb{R}_+ \times \mathbb{R}_+^S} \left[2y_1 \hat{x}'(\mathbf{y}_1) 2e^{-y_1^2/(2\sigma^2)} \prod_{l=2}^{S+1} (e^{-(y_l-x_l)^2/(2\sigma^2)} - e^{-(y_l+x_l)^2/(2\sigma^2)}) \right. \\ \left. + (\hat{x}'(\mathbf{y}_1))^2 2e^{-y_1^2/(2\sigma^2)} \prod_{l=2}^{S+1} (e^{-(y_l-x_l)^2/(2\sigma^2)} + e^{-(y_l+x_l)^2/(2\sigma^2)}) \right] d\mathbf{y}_1.$$

Finally, removing our “notational simplicity” assumptions $k=1$ and $\text{supp}(\mathbf{x}) = \{2, \dots, S+1\}$, this can be written for general k and $\text{supp}(\mathbf{x})$ as

$$2e^{-y_k^2/(2\sigma^2)} \int_{\mathbb{R}_+^{S+1}} \left[2y_k \hat{x}'(\mathbf{y}_k) \prod_{l \in \text{supp}(\mathbf{x})} (e^{-(y_l-x_l)^2/(2\sigma^2)} - e^{-(y_l+x_l)^2/(2\sigma^2)}) \right. \\ \left. + (\hat{x}'(\mathbf{y}_k))^2 \prod_{l \in \text{supp}(\mathbf{x})} (e^{-(y_l-x_l)^2/(2\sigma^2)} + e^{-(y_l+x_l)^2/(2\sigma^2)}) \right] d\mathbf{y}_k. \quad (90)$$

Inserting (90) into (85) yields (86).

The integral $\int_{\mathbb{R}_+^{S+1}} [2\hat{x}'(\mathbf{y}_k)b(\mathbf{y}_k) + (\hat{x}'(\mathbf{y}_k))^2 c(\mathbf{y}_k)] d\mathbf{y}_k$ is minimized with respect to $\hat{x}'(\mathbf{y}_k)$ by minimizing the integrand $2\hat{x}'(\mathbf{y}_k)b(\mathbf{y}_k) + (\hat{x}'(\mathbf{y}_k))^2 c(\mathbf{y}_k)$ pointwise for each value of $\mathbf{y}_k \in \mathbb{R}_+^{S+1}$. This is easily done by completing the square in $\hat{x}'(\mathbf{y}_k)$, yielding the optimization problem $\min_{\hat{x}'(\mathbf{y}_k)} [\hat{x}'(\mathbf{y}_k) + b(\mathbf{y}_k)/c(\mathbf{y}_k)]^2$. Thus, the optimal $\hat{x}'(\mathbf{y}_k)$ is obtained as

$$\hat{x}'_{k,\mathbf{x}}(\mathbf{y}_k) \triangleq -\frac{b(\mathbf{y}_k)}{c(\mathbf{y}_k)} = -y_k \prod_{l \in \text{supp}(\mathbf{x})} \tanh\left(\frac{x_l y_l}{\sigma^2}\right) \quad \text{for all } \mathbf{y}_k \in \mathbb{R}_+^{S+1}$$

and the corresponding pointwise minimum of the integrand is given by $-(b(\mathbf{y}_k))^2/c(\mathbf{y}_k)$. The extension $\hat{x}'_{k,\mathbf{x}}(\mathbf{y})$ to all $\mathbf{y} \in \mathbb{R}^N$ is then obtained using the properties (23) and (24), and the optimal component estimator solving (25) follows as $\hat{x}_{k,\mathbf{x}}(\mathbf{y}) = y_k + \hat{x}'_{k,\mathbf{x}}(\mathbf{y})$. The corresponding minimum variance, denoted by $\text{BB}_c^k(\mathbf{x})$, is obtained by substituting the minimum value of the integrand, $-(b(\mathbf{y}_k))^2/c(\mathbf{y}_k)$, in (86). This yields

$$\text{BB}_c^k(\mathbf{x}) \triangleq \mathbb{E}_{\mathbf{x}}\{(\hat{x}_{k,\mathbf{x}}(\mathbf{y}))^2\} = \sigma^2 - \frac{2}{(2\pi\sigma^2)^{(S+1)/2}} \int_{\mathbb{R}_+^{S+1}} \frac{(b(\mathbf{y}_k))^2}{c(\mathbf{y}_k)} d\mathbf{y}_k. \quad (91)$$

Inserting (87) and (88) into (91) and simplifying gives (26).

APPENDIX H

PROOF OF EQUATION (29)

To show (29), we consider $g(x; \sigma^2)$ for $x \geq 0$ (this is sufficient since $g(-x; \sigma^2) = g(x; \sigma^2)$), and we use the simple bound $\tanh(x) \geq 1 - e^{-x}$, which can be verified using elementary calculus. We then obtain from (27), for $x \geq 0$,

$$\begin{aligned} g(x; \sigma^2) &\geq \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^\infty e^{-(x^2+y^2)/(2\sigma^2)} \sinh\left(\frac{xy}{\sigma^2}\right) (1 - e^{-xy/\sigma^2}) dy \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^\infty [e^{-(x-y)^2/(2\sigma^2)} - e^{-(x+y)^2/(2\sigma^2)}] (1 - e^{-xy/\sigma^2}) dy \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^\infty [e^{-(x-y)^2/(2\sigma^2)} - e^{-(x^2+y^2)/(2\sigma^2)} - e^{-(x+y)^2/(2\sigma^2)} + e^{-(x+y)^2/(2\sigma^2)} e^{-xy/\sigma^2}] dy \\ &\geq \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^\infty [e^{-(x-y)^2/(2\sigma^2)} - e^{-(x^2+y^2)/(2\sigma^2)} - e^{-(x+y)^2/(2\sigma^2)}] dy \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^\infty e^{-(x-y)^2/(2\sigma^2)} dy - \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^\infty [e^{-(x^2+y^2)/(2\sigma^2)} + e^{-(x+y)^2/(2\sigma^2)}] dy. \end{aligned}$$

The first integral can be written as $\frac{1}{\sqrt{2\pi\sigma^2}} \int_0^\infty e^{-(x-y)^2/(2\sigma^2)} dy = 1 - \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^0 e^{-(x-y)^2/(2\sigma^2)} dy = 1 - \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^\infty e^{-(x+y)^2/(2\sigma^2)} dy$. The bound thus becomes

$$\begin{aligned} g(x; \sigma^2) &\geq 1 - \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^\infty [2e^{-(x+y)^2/(2\sigma^2)} + e^{-(x^2+y^2)/(2\sigma^2)}] dy \\ &= 1 - \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^\infty [2e^{-2xy/(2\sigma^2)} + 1] e^{-(x^2+y^2)/(2\sigma^2)} dy \\ &\stackrel{(*)}{\geq} 1 - \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^\infty 3e^{-(x^2+y^2)/(2\sigma^2)} dy \\ &= 1 - \frac{3}{\sqrt{2\pi\sigma^2}} e^{-x^2/(2\sigma^2)} \int_0^\infty e^{-y^2/(2\sigma^2)} dy \\ &= 1 - \frac{3}{2} e^{-x^2/(2\sigma^2)} \end{aligned}$$

where $e^{-2xy/(2\sigma^2)} \leq 1$ was used in (*). This bound on $g(x; \sigma^2)$ is actually valid for all $x \in \mathbb{R}$ because $g(-x; \sigma^2) = g(x; \sigma^2)$. Inserting it in (26), we obtain

$$\text{BB}_c^k(\mathbf{x}) \leq \left[1 - \prod_{l \in \text{supp}(\mathbf{x})} \left(1 - \frac{3}{2} e^{-x_l^2/(2\sigma^2)} \right) \right] \sigma^2. \quad (92)$$

The statement in (29) follows since we have (note that $\sum_{\mathcal{I} \subseteq \text{supp}(\mathbf{x})}$ denotes the sum over all possible subsets \mathcal{I} of $\text{supp}(\mathbf{x})$, including $\text{supp}(\mathbf{x})$ and the empty set \emptyset)

$$\begin{aligned} 1 - \prod_{l \in \text{supp}(\mathbf{x})} \left(1 - \frac{3}{2} e^{-x_l^2/(2\sigma^2)} \right) &= 1 - \sum_{\mathcal{I} \subseteq \text{supp}(\mathbf{x})} \prod_{l \in \mathcal{I}} \left(-\frac{3}{2} e^{-x_l^2/(2\sigma^2)} \right) \\ &= - \sum_{\mathcal{I} \subseteq \text{supp}(\mathbf{x}), \mathcal{I} \neq \emptyset} \prod_{l \in \mathcal{I}} \left(-\frac{3}{2} e^{-x_l^2/(2\sigma^2)} \right) \\ &\leq \sum_{\mathcal{I} \subseteq \text{supp}(\mathbf{x}), \mathcal{I} \neq \emptyset} \prod_{l \in \mathcal{I}} \left(\frac{3}{2} e^{-x_l^2/(2\sigma^2)} \right) \\ &\leq \sum_{\mathcal{I} \subseteq \text{supp}(\mathbf{x}), \mathcal{I} \neq \emptyset} \prod_{l \in \mathcal{I}} \left(\frac{3}{2} e^{-\xi^2/(2\sigma^2)} \right) \\ &= \sum_{\mathcal{I} \subseteq \text{supp}(\mathbf{x}), \mathcal{I} \neq \emptyset} \left(\frac{3}{2} e^{-\xi^2/(2\sigma^2)} \right)^{|\mathcal{I}|} \\ &\leq \sum_{\mathcal{I} \subseteq \text{supp}(\mathbf{x}), \mathcal{I} \neq \emptyset} \left(\frac{3}{2} \right)^S e^{-\xi^2/(2\sigma^2)} \\ &\leq 2^S \left(\frac{3}{2} \right)^S e^{-\xi^2/(2\sigma^2)} \\ &= 3^S e^{-\xi^2/(2\sigma^2)} \end{aligned}$$

where we have used the fact that the number of different subsets $\mathcal{I} \subseteq \text{supp}(\mathbf{x})$ is $2^{|\text{supp}(\mathbf{x})|} = 2^S$. Inserting the last bound in (92) and, in turn, the resulting bound on $\text{BB}_c^k(\mathbf{x})$ in (28) yields (29).

APPENDIX I

MSE OF THE ML ESTIMATOR

We calculate the MSE $\varepsilon(\mathbf{x}; \hat{\mathbf{x}}_{\text{ML}})$ of the ML estimator $\hat{\mathbf{x}}_{\text{ML}}$ in (7). Let $\hat{x}_{\text{ML},k}$ denote the k th component of $\hat{\mathbf{x}}_{\text{ML}}$. We have

$$\varepsilon(\mathbf{x}; \hat{\mathbf{x}}_{\text{ML}}) = \sum_{k=1}^N \mathbb{E}_{\mathbf{x}} \{ (\hat{x}_{\text{ML},k} - x_k)^2 \}$$

$$\begin{aligned}
&= \sum_{k=1}^N [\mathbb{E}_{\mathbf{x}}\{\hat{x}_{\text{ML},k}^2\} - 2\mathbb{E}_{\mathbf{x}}\{\hat{x}_{\text{ML},k}\}x_k + x_k^2] \\
&= \sum_{k=1}^N [\mathbb{E}_{\mathbf{x}}\{\hat{x}_{\text{ML},k}^2\} + (\mathbb{E}_{\mathbf{x}}\{\hat{x}_{\text{ML},k}\} - x_k)^2 - (\mathbb{E}_{\mathbf{x}}\{\hat{x}_{\text{ML},k}\})^2]. \tag{93}
\end{aligned}$$

Thus, we have to calculate the quantities $\mathbb{E}_{\mathbf{x}}\{\hat{x}_{\text{ML},k}\}$ and $\mathbb{E}_{\mathbf{x}}\{\hat{x}_{\text{ML},k}^2\}$.

We recall that $\hat{x}_{\text{ML},k}(\mathbf{y}) = (\text{P}_S(\mathbf{y}))_k$, where P_S is an operator that retains the S largest (in magnitude) components and zeros out all others. Let \mathcal{L}_k denotes the set of vectors \mathbf{y} for which y_k is not among the S largest (in magnitude) components. We then have

$$\hat{x}_{\text{ML},k}(\mathbf{y}) = \begin{cases} y_k, & \mathbf{y} \notin \mathcal{L}_k \\ 0, & \mathbf{y} \in \mathcal{L}_k. \end{cases}$$

Equivalently, $\hat{x}_{\text{ML},k}(\mathbf{y}) = y_k[1 - \text{I}(\mathbf{y} \in \mathcal{L}_k)]$, where $\text{I}(\mathbf{y} \in \mathcal{L}_k)$ is the indicator function of the event $\{\mathbf{y} \in \mathcal{L}_k\}$ (i.e., $\text{I}(\mathbf{y} \in \mathcal{L}_k)$ is 1 if $\mathbf{y} \in \mathcal{L}_k$ and 0 else). Thus, we obtain $\mathbb{E}_{\mathbf{x}}\{\hat{x}_{\text{ML},k}\}$ as

$$\begin{aligned}
\mathbb{E}_{\mathbf{x}}\{\hat{x}_{\text{ML},k}\} &= \mathbb{E}_{\mathbf{x}}\{y_k[1 - \text{I}(\mathbf{y} \in \mathcal{L}_k)]\} \\
&= x_k - \mathbb{E}_{\mathbf{x}}\{y_k \text{I}(\mathbf{y} \in \mathcal{L}_k)\} \\
&\stackrel{(a)}{=} x_k - \mathbb{E}_{\mathbf{x}}^{(y_k)}\{\mathbb{E}_{\mathbf{x}}^{(\mathbf{y} \sim k)}\{y_k \text{I}(\mathbf{y} \in \mathcal{L}_k)|y_k\}\} \\
&\stackrel{(b)}{=} x_k - \mathbb{E}_{\mathbf{x}}^{(y_k)}\{y_k \mathbb{E}_{\mathbf{x}}^{(\mathbf{y} \sim k)}\{\text{I}(\mathbf{y} \in \mathcal{L}_k)|y_k\}\} \\
&= x_k - \mathbb{E}_{\mathbf{x}}^{(y_k)}\{y_k \text{P}_{\mathbf{x}}(\mathbf{y} \in \mathcal{L}_k|y_k)\} \tag{94}
\end{aligned}$$

where the notations $\mathbb{E}_{\mathbf{x}}^{(y_k)}$ and $\mathbb{E}_{\mathbf{x}}^{(\mathbf{y} \sim k)}$ indicate that the expectation is taken with respect to the random quantities y_k and $\mathbf{y} \sim k$, respectively (here, $\mathbf{y} \sim k$ denotes \mathbf{y} without the component y_k) and $\text{P}_{\mathbf{x}}(\mathbf{y} \in \mathcal{L}_k|y_k)$ is the conditional probability that $\mathbf{y} \in \mathcal{L}_k$, given y_k . Furthermore, we used the law of total probability in (a) and the fact that y_k is held constant in the conditional expectation $\mathbb{E}_{\mathbf{x}}\{y_k \text{I}(\mathbf{y} \in \mathcal{L}_k)|y_k\}$ in (b). Similarly,

$$\begin{aligned}
\mathbb{E}_{\mathbf{x}}\{\hat{x}_{\text{ML},k}^2\} &= \mathbb{E}_{\mathbf{x}}\{y_k^2[1 - \text{I}(\mathbf{y} \in \mathcal{L}_k)]^2\} \\
&= \mathbb{E}_{\mathbf{x}}\{y_k^2[1 - \text{I}(\mathbf{y} \in \mathcal{L}_k)]\} \\
&= \sigma^2 + x_k^2 - \mathbb{E}_{\mathbf{x}}\{y_k^2 \text{I}(\mathbf{y} \in \mathcal{L}_k)\} \\
&= \sigma^2 + x_k^2 - \mathbb{E}_{\mathbf{x}}^{(y_k)}\{y_k^2 \text{P}_{\mathbf{x}}(\mathbf{y} \in \mathcal{L}_k|y_k)\}. \tag{95}
\end{aligned}$$

Calculating $\mathbb{E}_{\mathbf{x}}\{\hat{x}_{\text{ML},k}\}$ and $\mathbb{E}_{\mathbf{x}}\{\hat{x}_{\text{ML},k}^2\}$ is thus reduced to calculating the conditional probability $\text{P}_{\mathbf{x}}(\mathbf{y} \in \mathcal{L}_k|y_k)$.

Let $\mathcal{M}_k \triangleq \{1, \dots, N\} \setminus \{k\}$, and let \mathcal{P} denote the set of all binary partitions $(\mathcal{A}, \mathcal{B})$ of the set \mathcal{M}_k , where \mathcal{A} is at least of cardinality S :

$$\mathcal{P} \triangleq \{(\mathcal{A}, \mathcal{B}) \mid \mathcal{A} \subseteq \mathcal{M}_k, \mathcal{B} \subseteq \mathcal{M}_k, \mathcal{A} \cap \mathcal{B} = \emptyset, \mathcal{A} \cup \mathcal{B} = \mathcal{M}_k, |\mathcal{A}| \geq S\}.$$

In order to evaluate the conditional probability $\mathbf{P}_{\mathbf{x}}(\mathbf{y} \in \mathcal{L}_k | y_k)$ of the event $\{\mathbf{y} \in \mathcal{L}_k\}$, i.e., of the event that a given y_k is not among the S largest (in magnitude) components of \mathbf{y} , we split the event $\{\mathbf{y} \in \mathcal{L}_k\}$ into several elementary events. More specifically, let $\mathcal{E}_{\mathcal{A}, \mathcal{B}}$ denote the event that every component y_l with $l \in \mathcal{A}$ satisfies $|y_l| > |y_k|$ and every component y_l with $l \in \mathcal{B}$ satisfies $|y_l| \leq |y_k|$. The events $\mathcal{E}_{\mathcal{A}, \mathcal{B}}$ for all $(\mathcal{A}, \mathcal{B}) \in \mathcal{P}$ are mutually exclusive, i.e., $(\mathcal{A}, \mathcal{B}) \neq (\mathcal{A}', \mathcal{B}') \Rightarrow \mathcal{E}_{\mathcal{A}, \mathcal{B}} \cap \mathcal{E}_{\mathcal{A}', \mathcal{B}'} = \emptyset$, and their union corresponds to the event $\{\mathbf{y} \in \mathcal{L}_k\}$, i.e., $\bigcup_{(\mathcal{A}, \mathcal{B}) \in \mathcal{P}} \mathcal{E}_{\mathcal{A}, \mathcal{B}} = \{\mathbf{y} \in \mathcal{L}_k\}$. Consequently,

$$\begin{aligned} \mathbf{P}_{\mathbf{x}}(\mathbf{y} \in \mathcal{L}_k | y_k = y) &= \sum_{(\mathcal{A}, \mathcal{B}) \in \mathcal{P}} \mathbf{P}_{\mathbf{x}}(\mathcal{E}_{\mathcal{A}, \mathcal{B}} | y_k = y) \\ &= \sum_{(\mathcal{A}, \mathcal{B}) \in \mathcal{P}} \prod_{l \in \mathcal{A}} \mathbf{P}_{\mathbf{x}}(|y_l| > |y_k| | y_k = y) \prod_{m \in \mathcal{B}} \mathbf{P}_{\mathbf{x}}(|y_m| \leq |y_k| | y_k = y) \\ &= \sum_{(\mathcal{A}, \mathcal{B}) \in \mathcal{P}} \prod_{l \in \mathcal{A}} \mathbf{P}_{\mathbf{x}}(|y_l| > |y|) \prod_{m \in \mathcal{B}} \mathbf{P}_{\mathbf{x}}(|y_m| \leq |y|) \\ &= \sum_{(\mathcal{A}, \mathcal{B}) \in \mathcal{P}} \prod_{l \in \mathcal{A} \cap \text{supp}(\mathbf{x})} \mathbf{P}_{\mathbf{x}}(|y_l| > |y|) \prod_{m \in \mathcal{B} \cap \text{supp}(\mathbf{x})} \mathbf{P}_{\mathbf{x}}(|y_m| \leq |y|) \\ &\quad \times \prod_{n \in \mathcal{A} \setminus \text{supp}(\mathbf{x})} \mathbf{P}_{\mathbf{x}}(|y_n| > |y|) \prod_{p \in \mathcal{B} \setminus \text{supp}(\mathbf{x})} \mathbf{P}_{\mathbf{x}}(|y_p| \leq |y|) \\ &= \sum_{(\mathcal{A}, \mathcal{B}) \in \mathcal{P}} \prod_{l \in \mathcal{A} \cap \text{supp}(\mathbf{x})} \left[Q\left(\frac{|y| - x_l}{\sigma}\right) + 1 - Q\left(\frac{-|y| - x_l}{\sigma}\right) \right] \\ &\quad \times \prod_{m \in \mathcal{B} \cap \text{supp}(\mathbf{x})} \left[-Q\left(\frac{|y| - x_m}{\sigma}\right) + Q\left(\frac{-|y| - x_m}{\sigma}\right) \right] \\ &\quad \times \prod_{n \in \mathcal{A} \setminus \text{supp}(\mathbf{x})} 2Q\left(\frac{|y|}{\sigma}\right) \prod_{p \in \mathcal{B} \setminus \text{supp}(\mathbf{x})} \left[1 - 2Q\left(\frac{|y|}{\sigma}\right) \right] \quad (96) \end{aligned}$$

where we have used the fact that the y_l are independent and $k \notin \mathcal{M}_k$; furthermore, $Q(y) \triangleq \frac{1}{\sqrt{2\pi}} \int_y^\infty e^{-x^2/2} dx$ is the right tail probability of a standard Gaussian random variable. Plugging (96) into (94) and (95) and, in turn, the resulting expressions into (93) yields a (very complicated) expression of $\varepsilon(\mathbf{x}; \hat{\mathbf{x}}_{\text{ML}})$. This expression is evaluated numerically in Section V.

REFERENCES

- [1] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.
- [2] E. J. Candès and T. Tao, “Near-optimal signal recovery from random projections: Universal encoding strategies?” *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.
- [3] D. L. Donoho and I. M. Johnstone, “Ideal spatial adaptation by wavelet shrinkage,” *Biometrika*, vol. 81, pp. 425–455, 1994.
- [4] K. Dabov, A. Foi, V. Katkovich, and K. Egiazarian, “Image restoration by sparse 3D transform-domain collaborative filtering,” in *Proc. SPIE Electronic Imaging '08*, no. 6812-07, San Jose, CA, Jan. 2008.
- [5] M. Protter and M. Elad, “Image sequence denoising via sparse and redundant representations,” *IEEE Trans. Image Processing*, vol. 18, no. 1, pp. 27–36, Jan. 2009.
- [6] E. L. Lehmann and G. Casella, *Theory of Point Estimation*. New York: Springer, 2003.
- [7] E. G. Larsson and Y. Selen, “Linear regression with a sparse parameter vector,” *IEEE Trans. Signal Processing*, vol. 55, no. 2, pp. 451–460, Feb. 2007.
- [8] M. E. Tipping, “Sparse Bayesian learning and the relevance vector machine,” *J. Machine Learn. Res. 1*, pp. 211–244, 2001.
- [9] P. Schniter, L. Potter, and J. Ziniel, “Fast Bayesian matching pursuit,” in *Inf. Theory Appl. Workshop*, San Diego, CA, Jan./Feb. 2008, pp. 326–333.
- [10] L. L. Scharf, *Statistical Signal Processing*. Reading (MA): Addison Wesley, 1991.
- [11] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.
- [12] H. Cramér, “A contribution to the theory of statistical estimation,” *Skand. Akt. Tidskr.*, vol. 29, pp. 85–94, 1945.
- [13] J. Stein, “Inadmissibility of the usual estimator for the mean of a multivariate normal distribution,” in *Proc. Third Berkeley Sympos. Math. Stat. Prob.* University of California Press, 1956, pp. 197–206.
- [14] Z. Ben-Haim and Y. C. Eldar, “Blind minimax estimation,” *IEEE Trans. Inf. Theory*, vol. 53, no. 9, pp. 3145–3157, Sep. 2007.
- [15] Y. C. Eldar, “Rethinking Biased Estimation: Improving Maximum Likelihood and the Cramér–Rao Bound,” *Foundations and Trends in Signal Processing*, vol. 1, no. 4, pp. 305–449, 2008.
- [16] 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.
- [17] M. Dong and L. Tong, “Optimal design and placement of pilot symbols for channel estimation,” *IEEE Trans. Signal Processing*, vol. 50, no. 12, pp. 3055–3069, Dec 2002.
- [18] A. Jung, Z. Ben-Haim, F. Hlawatsch, and Y. C. Eldar, “On unbiased estimation of sparse vectors corrupted by Gaussian noise,” in *Proc. IEEE ICASSP-2010*, Dallas, TX, March 2010, pp. 3990–3993.
- [19] D. Donoho and I. M. Johnstone, “Minimax risk over ℓ_p balls for ℓ_q -error,” *Probab. Theory Relat. Fields*, vol. 99, pp. 277–303, 1994.
- [20] —, “Minimax estimation via wavelet shrinkage,” *Ann. Statist.*, vol. 26, pp. 879–921, 1992.
- [21] Z. Ben-Haim and Y. C. Eldar, “The Cramér–Rao bound for estimating a sparse parameter vector,” *IEEE Trans. Signal Processing*, vol. 58, pp. 3384–3389, June 2010.
- [22] —, “Performance bounds for sparse estimation with random noise,” in *Proc. IEEE-SP Workshop Statist. Signal Process.*, Cardiff, Wales (UK), Aug. 2009, pp. 225–228.

- [23] 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.
- [24] E. W. Barankin, "Locally best unbiased estimates," *Ann. Math. Statist.*, vol. 20, no. 4, pp. 477–501, 1949.
- [25] I. M. Johnstone and B. W. Silverman, "Needles and straw in haystacks: Empirical Bayes estimates of possibly sparse sequences," *Ann. Statist.*, vol. 32, pp. 1594–1649, 2004.
- [26] C. R. Rao, "Information and accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc.*, vol. 37, pp. 81–91, 1945.
- [27] J. M. Hammersley, "On estimating restricted parameters," *J. Roy. Statist. Soc. B*, vol. 12, no. 2, pp. 192–240, 1950.
- [28] D. G. Chapman and H. Robbins, "Minimum variance estimation without regularity assumptions," *Ann. Math. Statist.*, vol. 22, no. 4, pp. 581–586, Dec. 1951.
- [29] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.
- [30] J. S. Abel, "A bound on mean-square-estimate error," *IEEE Trans. Inf. Theory*, vol. 39, no. 5, pp. 1675–1680, Sep. 1993.
- [31] P. Forster and P. Larzabal, "On lower bounds for deterministic parameter estimation," in *Proc. IEEE ICASSP-2002*, vol. 2, Orlando, FL, May 2002, pp. 1137–1140.
- [32] K. Todros and J. Tabrikian, "A new lower bound on the mean-square error of unbiased estimators," in *Proc. IEEE ICASSP-2008*, Las Vegas, NV, Mar./Apr. 2008, pp. 3913–3916.
- [33] S. G. Mallat, *A Wavelet Tour of Signal Processing – The Sparse Way*, 3rd ed. San Diego, CA: Academic Press, 2009.
- [34] R. S. Strichartz, *A Guide to Distribution Theory and Fourier Transforms*, 1st ed. Hackensack, NJ: World Scientific Publishing, 1994.
- [35] D. S. Bernstein, *Matrix Mathematics*. Princeton, NJ: Princeton Univ. Press, 2005.
- [36] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: Mc-Graw Hill, 1984.