The Cramér–Rao Bound for Estimating a Sparse Parameter Vector

Zvika Ben-Haim, *Student Member, IEEE,*
and Yonina C. Eldar, *Senior Member, IEEE*

***Abstract*—The goal of this contribution is to characterize the best achievable mean-squared error (MSE) in estimating a sparse deterministic parameter from measurements corrupted by Gaussian noise. To this end, an appropriate definition of bias in the sparse setting is developed, and the constrained Cramér–Rao bound (CRB) is obtained. This bound is shown to equal the CRB of an estimator with knowledge of the support set, for almost all feasible parameter values. Consequently, in the unbiased case, our bound is identical to the MSE of the oracle estimator. Combined with the fact that the CRB is achieved at high signal-to-noise ratios by the maximum likelihood technique, our result provides a new interpretation for the common practice of using the oracle estimator as a gold standard against which practical approaches are compared.**

Index terms: Constrained estimation, Cramér–Rao bound, sparse estimation.

I. Introduction

The problem of estimating a sparse unknown parameter vector from noisy measurements has been analyzed intensively in the past few years, and has already given rise to numerous successful signal processing algorithms (see, e.g., [1], [2]). In this contribution, we consider the setting in which a deterministic sparse vector $\boldsymbol{x}_0$ is to be estimated from a small number of noisy measurements. Much of the interest in sparse estimation is a consequence of the fact that a variety of practical approaches are surprisingly successful in this task.

One way to determine the quality of an estimator is to compare its mean-squared error (MSE) with theoretical performance limits: if one is able to approach the bound, then efforts at further performance enhancements are futile. This motivates the development of lower bounds on the MSE of estimators in the sparse setting. While there are lower bounds on the worst-case achievable MSE among all possible parameter values [3, §7.4], the actual performance for a specific value, or even for most values, might be substantially lower. Our goal is to characterize the minimum MSE obtainable for each particular parameter vector. A standard method of achieving this objective is the Cramér–Rao bound (CRB) [4].

In this correspondence, we formulate the sparsity assumption as a constrained estimation problem, and our results are inspired by the well-studied theory of the constrained CRB [5]–[7]. However, deriving the CRB for sparse estimation requires the development of a bound suitable for non-smooth constraints [8], [9]. In obtaining this modified bound, we also provide new insight into the meaning of the constrained CRB. In particular, we will see that while the unconstrained CRB is relevant to estimators having a completely specified bias, the constrained bound applies to a wider class of techniques having a partially specified bias function. This distinction explains some curious properties of the constrained CRB.

In terms of the application to the sparse setting, our contribution is related to, but distinct from, the work of Babadi et al. [10], in which the CRB of the "oracle estimator" was derived. This estimator has knowledge of the locations of the nonzero components of $\boldsymbol{x}_0$. By contrast, our goal is to obtain a lower bound on the performance of estimators which are not endowed with oracular knowledge. It was also shown in [10] that when the measurements consist of Gaussian random mixtures of the parameter vector, there exists an estimator which asymptotically achieves the oracle CRB; this is shown to hold on average over realizations of the measurement mixtures. While a random measurement matrix is a common setup in the field of compressive sensing, there are many cases in which the matrix is fixed and given *a priori*. For example, images can be modeled as sparse combinations of empirically determined atoms [11], [12]. In this contribution, we will focus on the setting in which the measurement matrix is deterministic and pre-specified.

When examining the CRB, it is common to particularly emphasize the unbiased case. In the sparse estimation setting, the unbiased CRB can be summarized as follows. For parameters having maximal support, i.e., parameters whose representation requires the maximum allowed number s of atoms, the lower bound equals the MSE of the oracle estimator. On the other hand, for parameters which do not have maximal support (a set which has Lebesgue measure zero in $\mathcal{S}$), our lower bound is identical to the CRB for an unconstrained problem, which is substantially higher than the oracle MSE.

The correspondence between the unbiased CRB and the MSE of the oracle estimator (for all but a zero-measure subset of the feasible parameter set $\mathcal{S}$) is of practical interest since, unlike the oracle estimator, the CRB is achieved by the maximum likelihood (ML) estimator at high signal-to-noise ratio (SNR). Our bound can thus be viewed as an alternative justification for the common use of the oracle estimator as a baseline against which practical algorithms are compared. This gives further merit to recent results, which demonstrate that practical algorithms achieve near-oracle performance [2], [13]. However, the existence of parameters for which the bound is much higher indicates that oracular performance cannot be attained for *all* parameter values, at least using unbiased techniques. Indeed, as we will show, in many sparse estimation scenarios, one cannot construct *any* estimator which is unbiased for all sparsely representable parameters.

The rest of this paper is organized as follows. In Section II, we model the sparse setting as a constrained estimation problem. Section III defines a generalization of sparsity constraints, which we refer to as locally balanced constraint sets; the CRB is then derived in this general setting. In Section IV, our general results are applied back to the sparse estimation problem. The implications of this analysis are discussed in Section V.

Throughout the paper, boldface lowercase letters $\boldsymbol{v}$ denote vectors while boldface uppercase letters $\boldsymbol{M}$ denote matrices. Given a vector function $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^k$, we denote by $\partial \boldsymbol{f}/\partial \boldsymbol{x}$ the $k \times n$ matrix whose ijth element is $\partial f_i/\partial x_j$. The support of a vector, denoted $\operatorname{supp}(\boldsymbol{v})$, is the set of indices of the nonzero entries in $\boldsymbol{v}$. The Euclidean norm of a vector $\boldsymbol{v}$ is denoted $\|\boldsymbol{v}\|_2$, and the number of nonzero entries in $\boldsymbol{v}$ is $\|\boldsymbol{v}\|_0$. Finally, the symbols $\mathcal{R}(\boldsymbol{M})$, $\mathcal{N}(\boldsymbol{M})$, and $\boldsymbol{M}^\dagger$ refer, respectively, to the column space, null space, and Moore–Penrose pseudoinverse of the matrix $\boldsymbol{M}$.

II. Sparse Estimation Problems

Suppose we observe a measurement vector $\boldsymbol{y} \in \mathbb{R}^m$, given by

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x}_0 + \boldsymbol{w} \tag{1}$$

where $\boldsymbol{x}_0 \in \mathbb{R}^p$ is an unknown deterministic signal, $\boldsymbol{w}$ is independent, identically distributed (IID) Gaussian noise with zero mean and variance σ^2, and $\boldsymbol{H}$ is a known $m \times p$ matrix. It is known that

Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel. Phone: +972-4-8294700, fax: +972-4-8295757, E-mail: {zvikabh@tx, yonina@ee}.technion.ac.il. This work was supported in part by the Israel Science Foundation under Grant no. 1081/07 and by the European Commission in the framework of the FP7 Network of Excellence in Wireless COMmunications NEWCOM++ (contract no. 216715).

$\boldsymbol{x}_0$ is sparse, i.e.,

$$\boldsymbol{x}_0 \in \mathcal{S} \triangleq \{\boldsymbol{x} \in \mathbb{R}^p : \|\boldsymbol{x}\|_0 \leq s\}. \tag{2}$$

We will adopt the standard assumption that

$$\operatorname{spark}(\boldsymbol{H}) > 2s \tag{3}$$

where $\operatorname{spark}(\boldsymbol{H})$ is defined as the smallest integer k such that there exist k linearly dependent columns in $\boldsymbol{H}$ [14]. This condition is required to ensure identifiability of $\boldsymbol{x}_0$ from the measurements.

The goal of an estimator is to obtain an approximation $\widehat{\boldsymbol{x}}$ of $\boldsymbol{x}_0$ from the measurements $\boldsymbol{y}$. We will measure the quality of an estimator using the MSE, defined as $E\{\|\widehat{\boldsymbol{x}} - \boldsymbol{x}_0\|_2^2\}$. Since $\boldsymbol{x}_0$ is deterministic, the expectation is taken only over the noise $\boldsymbol{w}$, so that the MSE is in general a function of the unknown parameter.

A widely used approach for estimating parameters from noisy measurements is the ML technique. For the problem (1), the ML estimator is given by

$$\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|_2^2 \quad \text{s.t. } \|\boldsymbol{x}\|_0 \leq s. \tag{4}$$

Unfortunately, (4) is a nonconvex optimization problem and solving it is NP-hard, meaning that an efficient algorithm for calculating the ML estimator is unlikely to exist. Consequently, numerous practical alternatives have been proposed for estimating $\boldsymbol{x}_0$ [1], [2]. Previous research on the performance of these estimators has examined their worst-case MSE among all possible values of $\boldsymbol{x}_0 \in \mathcal{S}$. Specifically, it has been shown that, under suitable conditions on $\boldsymbol{H}$ and s, a variety of techniques obtain an MSE of $O(s\sigma^2 \log p)$ for all $\boldsymbol{x}_0 \in \mathcal{S}$ [2], [15]. Conversely, it is known that the worst-case error of *any* estimator is at least a constant times $s\sigma^2 \log p$ [3, §7.4]. Thus, there exist techniques which are optimal, up to a constant, in terms of worst-case error. Nevertheless, the MSE of these approaches for specific values of $\boldsymbol{x}_0$, even for a vast majority of such values, might be much lower. Our analysis differs from this line of work in that we characterize the *pointwise* performance of an estimator, i.e., the MSE for specific values of $\boldsymbol{x}_0$.

Another baseline with which practical techniques are often compared is the oracle estimator, given by

$$\widehat{\boldsymbol{x}}_{\text{oracle}} = \begin{cases} \boldsymbol{H}_{\boldsymbol{x}_0}^{\dagger}\boldsymbol{b} & \text{on the set } \operatorname{supp}(\boldsymbol{x}_0) \\ \boldsymbol{0} & \text{elsewhere} \end{cases} \tag{5}$$

where $\boldsymbol{H}_{\boldsymbol{x}_0}$ is the submatrix constructed from the columns of $\boldsymbol{H}$ corresponding to the nonzero entries of $\boldsymbol{x}_0$. In other words, $\widehat{\boldsymbol{x}}_{\text{oracle}}$ is the least-squares (LS) solution among vectors whose support coincides with $\operatorname{supp}(\boldsymbol{x}_0)$. Of course, in practice the support of $\boldsymbol{x}_0$ is unknown, so that $\widehat{\boldsymbol{x}}_{\text{oracle}}$ cannot actually be implemented. Nevertheless, one often compares the performance of true estimators with $\widehat{\boldsymbol{x}}_{\text{oracle}}$, whose MSE is given by [2]

$$\sigma^2 \operatorname{Tr}((\boldsymbol{H}_{\boldsymbol{x}_0}^T \boldsymbol{H}_{\boldsymbol{x}_0})^{-1}). \tag{6}$$

Is (6) a bound on estimation MSE? While $\widehat{\boldsymbol{x}}_{\text{oracle}}$ is a reasonable technique to adopt if $\operatorname{supp}(\boldsymbol{x}_0)$ is known, this does not imply that (6) is a lower bound on the performance of practical estimators. Indeed, when the SNR is low, it has been shown that some estimators outperform $\widehat{\boldsymbol{x}}_{\text{oracle}}$, thanks to the use of shrinkage [15]. Furthermore, if $\operatorname{supp}(\boldsymbol{x}_0)$ is known, then there exist (biased) techniques which are better than $\widehat{\boldsymbol{x}}_{\text{oracle}}$ for *all* values of $\boldsymbol{x}_0$ [16]. Thus, $\widehat{\boldsymbol{x}}_{\text{oracle}}$ is neither achievable in practice, nor optimal in terms of MSE. As we will see, one can indeed interpret (6) as a lower bound on the achievable MSE, but such a result requires a certain restriction of the class of estimators under consideration.

III. The Constrained Cramér–Rao Bound

Our goal in this paper is to calculate the CRB when it is known that the parameter $\boldsymbol{x}$ satisfies the sparsity constraint (2). While the constrained CRB has been studied extensively in the past [5]–[7], prior derivations assumed that the constraint set is given by

$$\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^p : \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{0},\ \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}\} \tag{7}$$

where $\boldsymbol{f}(\boldsymbol{x})$ and $\boldsymbol{g}(\boldsymbol{x})$ are continuously differentiable functions. We will refer to such $\mathcal{X}$ as continuously differentiable sets. Unfortunately, in some cases, including the sparse estimation scenario of Section II, the constraint set cannot be written in the form (7), and the aforementioned results are therefore inapplicable. Our goal in the current section is to close this gap by extending the constrained CRB to constraint sets $\mathcal{X}$ encompassing the sparse estimation scenario.

A. Bias Requirements in the Constrained CRB

In previous settings for which the constrained CRB was derived, it was noted that the resulting bound is typically lower than the unconstrained version [5, Remark 4]. At first glance, one would attribute the reduction in the value of the CRB to the fact that the constraints add information about the unknown parameter, which can then improve estimation performance. On the other hand, the CRB separately characterizes the achievable performance for each value of the unknown parameter $\boldsymbol{x}_0$. Thus, the CRB at $\boldsymbol{x}_0$ applies even to estimators designed specifically to perform well at $\boldsymbol{x}_0$. Such estimators surely cannot achieve further gain in performance if it is known that $\boldsymbol{x}_0 \in \mathcal{X}$. Why, then, is the constrained CRB lower than the unconstrained bound? The answer to this apparent paradox involves a careful definition of the class of estimators to which the bound applies.

To obtain a meaningful bound, one must exclude some estimators from consideration. Unless this is done, the bound will be tarnished by estimators of the type $\widehat{\boldsymbol{x}} = \boldsymbol{x}_{\mathrm{u}}$, for some constant $\boldsymbol{x}_{\mathrm{u}}$, which achieve an MSE of 0 at the specific point $\boldsymbol{x} = \boldsymbol{x}_{\mathrm{u}}$. It is standard practice to circumvent this difficulty by restricting attention to estimators having a particular bias $\boldsymbol{b}(\boldsymbol{x}) \triangleq E\{\widehat{\boldsymbol{x}}\} - \boldsymbol{x}$. In particular, it is common to examine unbiased estimators, for which $\boldsymbol{b}(\boldsymbol{x}) = \boldsymbol{0}$. If $\boldsymbol{x}$ is known to belong to a constrained set $\mathcal{X}$, then it seems reasonable to seek estimators which are unbiased for all $\boldsymbol{x} \in \mathcal{X}$. However, as we will see, this requirement can be too strict: in some cases it is impossible to construct estimators which are unbiased for all $\boldsymbol{x} \in \mathcal{X}$. Moreover, the CRB is a *local* bound, meaning that it determines the achievable performance at a particular value of $\boldsymbol{x}$ based on the statistics at $\boldsymbol{x}$ and at nearby values. Thus, it is irrelevant to introduce requirements on estimation performance for parameters which are distant from the value $\boldsymbol{x}$ of interest.

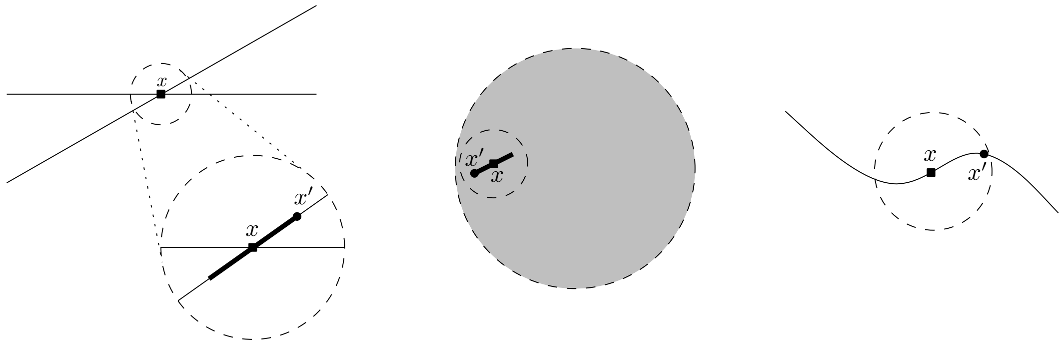
Since we seek a locally unbiased estimator, one possibility is to require unbiasedness at a single point, say $\boldsymbol{x}_{\mathrm{u}}$. As it turns out, it is always possible to construct such a technique: this is again $\widehat{\boldsymbol{x}} = \boldsymbol{x}_{\mathrm{u}}$, which is unbiased at $\boldsymbol{x}_{\mathrm{u}}$ but nowhere else. To avoid this loophole, one can require an estimator to be unbiased in the neighborhood

$$\mathcal{B}_\varepsilon(\boldsymbol{x}_0) = \{\boldsymbol{x} \in \mathcal{X} : \|\boldsymbol{x} - \boldsymbol{x}_0\|_2 < \varepsilon\} \tag{8}$$

of $\boldsymbol{x}_0$, for some small ε. Note that $\mathcal{B}_\varepsilon(\boldsymbol{x}_0)$ includes only those points $\boldsymbol{x}$ which are close to $\boldsymbol{x}_0$ and satisfy the constraints $\mathcal{X}$. In the unconstrained case for which $\mathcal{X} = \mathbb{R}^p$, this implies that both the bias $\boldsymbol{b}(\boldsymbol{x})$ and the bias gradient

$$\boldsymbol{B}(\boldsymbol{x}) \triangleq \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{x}} \tag{9}$$

vanish at $\boldsymbol{x} = \boldsymbol{x}_0$, and leads to the standard derivation of the unconstrained unbiased CRB. When constraints are imposed, such



(a) (b) (c)

Fig. 1. In a locally balanced set such as a union of subspaces (a) and an open ball (b), each point is locally defined by a set of feasible directions along which an infinitesimal movement does not violate the constraints. The curve (c) is not characterized in this way and thus is not locally balanced.

that $\mathcal{X} \neq \mathbb{R}^p$, one obtains a weaker requirement on the bias gradient $\boldsymbol{B}$ at $\boldsymbol{x}_0$. Specifically, the derivatives of the bias need only be specified in directions which do not violate the constraints.

It is worth emphasizing that the dependence of the CRB on the constraints is manifested through the class of estimators being considered, or more specifically, through the allowed estimators' bias gradient matrices. By contrast, the unconstrained CRB applies to estimators having a fully specified bias gradient matrix. Consequently, the constrained bound applies to a wider class of estimators, and is thus usually lower than the unconstrained version of the CRB. In other words, estimators which are unbiased in the constrained setting, and thus applicable to the unbiased constrained CRB, are likely to be biased in the unconstrained context. Since a wider class of estimators is considered by the constrained CRB, the resulting bound is lower, thus explaining the puzzling phenomenon described in the beginning of this subsection.

B. Locally Balanced Constraints

We now consider a class of constraint sets, called locally balanced sets, which encompass the sparsity constraints of Section II. Roughly speaking, a locally balanced set is one whose neighborhood at each point is defined by the directions along which one can move without leaving the set. Formally, a metric space $\mathcal{X}$ is said to be locally balanced if, for all $\boldsymbol{x} \in \mathcal{X}$, there exists an open set $\mathcal{C} \subset \mathcal{X}$ such that $\boldsymbol{x} \in \mathcal{C}$ and such that, for all $\boldsymbol{x}' \in \mathcal{C}$ and for all $|\lambda| \leq 1$, we have

$$\boldsymbol{x} + \lambda(\boldsymbol{x}' - \boldsymbol{x}) \in \mathcal{C}. \tag{10}$$

An example of a locally balanced set is given in Fig. 1(a), which represents a union of two subspaces. In this example, for any point $\boldsymbol{x} \in \mathcal{X}$, and for any point $\boldsymbol{x}' \in \mathcal{X}$ sufficiently close to $\boldsymbol{x}$, the entire line segment between $\boldsymbol{x}$ and $\boldsymbol{x}'$, as well as the line segment in the opposite direction, are also in $\mathcal{X}$. This illustrates the fact that any union of subspaces is locally balanced, and, in particular, so is the sparse estimation setting of Section II [17]–[19]. As another example, consider any open set, such as the open ball in Fig. 1(b). For such a set, any point $\boldsymbol{x}$ has a sufficiently small neighborhood $\mathcal{C}$ such that, for any $\boldsymbol{x}' \in \mathcal{C}$, the line segment connecting $\boldsymbol{x}$ to $\boldsymbol{x}'$ is contained in $\mathcal{X}$. On the other hand, the curve in Fig. 1(c) is not locally balanced, since the line connecting $\boldsymbol{x}$ to any other point on the set does not lie within the set.[1]

[1]We note in passing that since the curve in Fig. 1(c) is continuously differentiable, it can be locally approximated by a locally balanced set. Our derivation of the CRB can be extended to such approximately locally balanced sets in a manner similar to that of [5], but such an extension is not necessary for the purposes of this paper.

Observe that the neighborhood of a point $\boldsymbol{x}$ in a locally balanced set $\mathcal{X}$ is entirely determined by the set of feasible directions along which infinitesimal changes of $\boldsymbol{x}$ do not violate the constraints. These are the directions $\boldsymbol{v} = (\boldsymbol{x}' - \boldsymbol{x})/\|\boldsymbol{x}' - \boldsymbol{x}\|_2$ for all points $\boldsymbol{x}' \neq \boldsymbol{x}$ in the set $\mathcal{C}$ of (10). Recall that we seek a lower bound on the performance of estimators whose bias gradient is defined over the neighborhood of $\boldsymbol{x}_0$ restricted to the constraint set $\mathcal{X}$. Suppose for concreteness that we are interested in unbiased estimators. For a locally balanced constraint set $\mathcal{X}$, this implies that

$$\boldsymbol{B}\boldsymbol{v} = \boldsymbol{0} \tag{11}$$

for any feasible direction $\boldsymbol{v}$. In other words, all feasible directions must be in the nullspace of $\boldsymbol{B}$. This is a weaker condition than requiring the bias gradient to equal zero, and is thus more useful for constrained estimation problems. If an estimator $\widehat{\boldsymbol{x}}$ satisfies (11) for all feasible directions $\boldsymbol{v}$ at a certain point $\boldsymbol{x}_0$, we say that $\widehat{\boldsymbol{x}}$ is $\mathcal{X}$-unbiased at $\boldsymbol{x}_0$. This terminology emphasizes the fact that $\mathcal{X}$-unbiasedness depends both on the point $\boldsymbol{x}_0$ and on the constraint set $\mathcal{X}$.

Consider the subspace $\mathcal{F}$ spanned by the feasible directions at a certain point $\boldsymbol{x} \in \mathcal{X}$. We refer to $\mathcal{F}$ as the feasible subspace at $\boldsymbol{x}$. Note that $\mathcal{F}$ may include infeasible directions, if these are linear combinations of feasible directions. Nevertheless, because of the linearity of (11), any vector $\boldsymbol{u} \in \mathcal{F}$ satisfies $\boldsymbol{B}\boldsymbol{u} = \boldsymbol{0}$, even if $\boldsymbol{u}$ is infeasible. Thus, $\mathcal{X}$-unbiasedness is actually a property of the feasible subspace $\mathcal{F}$, rather than the set of feasible directions.

Since $\mathcal{X}$ is a subset of a finite-dimensional Euclidean space, $\mathcal{F}$ is also finite-dimensional, although different points in $\mathcal{X}$ may yield subspaces having differing dimensions. Let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_l$ denote an orthonormal basis for $\mathcal{F}$, and define the matrix

$$\boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_l]. \tag{12}$$

Note that $\boldsymbol{u}_i$ and $\boldsymbol{U}$ are functions of $\boldsymbol{x}$. For a given value of $\boldsymbol{x}$, different orthonormal bases can be chosen, but the choice of a basis is arbitrary and will not affect our results.

As we have seen, $\mathcal{X}$-unbiasedness at $\boldsymbol{x}_0$ can alternatively be written as $\boldsymbol{B}\boldsymbol{u} = \boldsymbol{0}$ for all $\boldsymbol{u} \in \mathcal{F}$, or, equivalently

$$\boldsymbol{B}\boldsymbol{U} = \boldsymbol{0}. \tag{13}$$

The constrained CRB can now be derived as a lower bound on all $\mathcal{X}$-unbiased estimators, which is a weaker requirement than "ordinary" unbiasedness.

Just as $\mathcal{X}$-unbiasedness was defined by requiring the bias gradient matrix to vanish when multiplied by any feasible direction vector, we can define $\mathcal{X}$-biased estimators by requiring a specific value (not necessarily zero) for the bias gradient matrix when multiplied by a feasible direction vector. In an analogy to (13), this implies that one must define a value for the matrix $\boldsymbol{BU}$. Our goal is thus to construct a lower bound on the covariance at a given $\boldsymbol{x}$ achievable by any estimator whose bias gradient $\boldsymbol{B}$ at $\boldsymbol{x}$ satisfies $\boldsymbol{BU} = \boldsymbol{P}$, for a given matrix $\boldsymbol{P}$. This is referred to as specifying the $\mathcal{X}$-bias of the estimator at $\boldsymbol{x}$. Specified in this manner, the $\mathcal{X}$-bias is dependent on the choice of the basis $\boldsymbol{U}$. This is, however, merely a notational effect: the bias gradient in any particular direction will not change if a different basis is chosen.

C. The CRB for Locally Balanced Constraints

It is helpful at this point to compare our derivation with prior work on the constrained CRB, which considered continuously differentiable constraint sets of the form (7). It has been previously shown [5] that inequality constraints of the type $\boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}$ have no effect on the CRB. Consequently, we will consider constraints of the form

$$\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^p : \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{0}\} \tag{14}$$

where $\boldsymbol{f} : \mathbb{R}^p \to \mathbb{R}^k$ is a continuously differentiable function defining the equality constraints. Define the $k \times p$ matrix $\boldsymbol{F}(\boldsymbol{x}) = \partial \boldsymbol{f}/\partial \boldsymbol{x}$. For simplicity of notation, we will omit the dependence of $\boldsymbol{F}$ on $\boldsymbol{x}$. Assuming that the constraints are non-redundant, $\boldsymbol{F}$ is a full-rank matrix, and thus one can define a $p \times (p-k)$ matrix $\boldsymbol{W}$ (also dependent on $\boldsymbol{x}$) such that

$$\boldsymbol{FW} = \boldsymbol{0}, \quad \boldsymbol{W}^T\boldsymbol{W} = \boldsymbol{I}. \tag{15}$$

The matrix $\boldsymbol{W}$ is closely related to the matrix $\boldsymbol{U}$ spanning the feasible direction subspace of locally balanced sets. Indeed, the column space $\mathcal{R}(\boldsymbol{W})$ of $\boldsymbol{W}$ is the tangent space of $\mathcal{X}$, i.e., the subspace of $\mathbb{R}^p$ containing all vectors which are tangent to $\mathcal{X}$ at the point $\boldsymbol{x}$. Thus, the vectors in $\mathcal{R}(\boldsymbol{W})$ are precisely those directions along which infinitesimal motion from $\boldsymbol{x}$ does not violate the constraints, up to a first-order approximation. It follows that if a particular set $\mathcal{X}$ is both locally balanced and continuously differentiable, its matrices $\boldsymbol{U}$ and $\boldsymbol{W}$ coincide. Note, however, that there exist sets which are locally balanced but not continuously differentiable (and vice versa).

With the above formulation, the CRB for continuously differentiable constraints can be stated as a function of the the matrix $\boldsymbol{W}$ and the bias gradient $\boldsymbol{B}$ [7]. In fact, the resulting bound depends on $\boldsymbol{B}$ only through $\boldsymbol{BW}$. This is to be expected in light of the discussion of Section III-A: The bias should be specified only for those directions which do not violate the constraint set. Furthermore, the proof of the CRB in [7, Theorem 1] depends not on the formulation (14) of the constraint set, but merely on the class of bias functions under consideration. Consequently, one can state the bound without any reference to the underlying constraint set. To do so, let $\boldsymbol{y}$ be a measurement vector with pdf $p(\boldsymbol{y};\boldsymbol{x})$, which is assumed to be differentiable with respect to $\boldsymbol{x}$. The Fisher information matrix (FIM) $\boldsymbol{J}(\boldsymbol{x})$ is defined as

$$\boldsymbol{J}(\boldsymbol{x}) = E\Big\{\boldsymbol{\Delta}\boldsymbol{\Delta}^T\Big\} \tag{16}$$

where

$$\boldsymbol{\Delta} = \frac{\partial \log p(\boldsymbol{y};\boldsymbol{x})}{\partial \boldsymbol{x}}. \tag{17}$$

We assume that the FIM is well-defined and finite. We further assume that integration with respect to $\boldsymbol{y}$ and differentiation with respect to $\boldsymbol{x}$ can be interchanged, a standard requirement for the CRB. We then have the following result.

Theorem 1. *Let $\widehat{\boldsymbol{x}}$ be an estimator and let $\boldsymbol{B} = \partial \boldsymbol{b}/\partial \boldsymbol{x}$ denote the bias gradient matrix of $\widehat{\boldsymbol{x}}$ at a given point $\boldsymbol{x}_0$. Let $\boldsymbol{U}$ be an orthonormal matrix, and suppose that $\boldsymbol{BU}$ is known, but that $\boldsymbol{B}$ is otherwise arbitrary. If*

$$\mathcal{R}(\boldsymbol{U}(\boldsymbol{U}+\boldsymbol{BU})^T)) \subseteq \mathcal{R}(\boldsymbol{UU}^T\boldsymbol{JUU}^T) \tag{18}$$

then the covariance of $\widehat{\boldsymbol{x}}$ at $\boldsymbol{x}_0$ satisfies

$$\mathrm{Cov}(\widehat{\boldsymbol{x}}) \succeq (\boldsymbol{U}+\boldsymbol{BU})\left(\boldsymbol{U}^T\boldsymbol{JU}\right)^{\dagger}(\boldsymbol{U}+\boldsymbol{BU})^T. \tag{19}$$

Equality is achieved in (19) *if and only if*

$$\widehat{\boldsymbol{x}} = \boldsymbol{x}_0 + \boldsymbol{b}(\boldsymbol{x}_0) + (\boldsymbol{U}+\boldsymbol{BU})\left(\boldsymbol{U}^T\boldsymbol{JU}\right)^{\dagger}\boldsymbol{U}^T\boldsymbol{\Delta} \tag{20}$$

in the mean square sense, where $\boldsymbol{\Delta}$ is defined by (17)*. Conversely, if* (18) *does not hold, then there exists no finite-variance estimator with the required bias gradient.*

As required, no mention of constrained estimation is made in Theorem 1; instead, partial information about the bias gradient is assumed. Apart from this restatement, the theorem is identical to [7, Theorem 1], and its proof is unchanged. However, the above formulation is more general in that it can be applied to any constrained setting, once the constraints have been translated to bias gradient requirements. In particular, as we will see below, Theorem 1 provides a CRB for locally balanced sets if the matrix $\boldsymbol{U}$ is chosen as a basis for the feasible direction subspace of Section III-B.

IV. CRB for Sparse Estimation

We are now ready to derive the CRB for the estimation problem (1) subject to the constraint set (2). We begin by identifying the feasible subspaces $\mathcal{F}$ corresponding to each of the elements in $\mathcal{S}$. To this end, consider first vectors $\boldsymbol{x} \in \mathcal{S}$ for which $\|\boldsymbol{x}\|_0 = s$, i.e., vectors having maximal support. Denote by $\{i_1, \ldots, i_s\}$ the support set of $\boldsymbol{x}$. Then, for all δ and for $k = 1, \ldots, s$, we have $\|\boldsymbol{x} + \delta\boldsymbol{e}_{i_k}\|_0 = \|\boldsymbol{x}\|_0 = s$, where $\boldsymbol{e}_j$ is the jth column of the identity matrix. Thus $\boldsymbol{x} + \delta\boldsymbol{e}_{i_k} \in \mathcal{S}$, and consequently, the vectors $\{\boldsymbol{e}_{i_1}, \ldots, \boldsymbol{e}_{i_s}\}$ are all feasible directions, as is any linear combination of these vectors. On the other hand, for any $j \notin \mathrm{supp}(\boldsymbol{x})$ and for any nonzero δ, we have $\|\boldsymbol{x}+\delta\boldsymbol{e}_j\|_0 = s+1$, and thus $\boldsymbol{e}_j$ is not a feasible direction; neither is any other vector which is not in $\mathrm{span}\{\boldsymbol{e}_{i_1}, \ldots, \boldsymbol{e}_{i_s}\}$. It follows that the feasible subspace $\mathcal{F}$ for points having maximal support is given by $\mathrm{span}\{\boldsymbol{e}_{i_1}, \ldots, \boldsymbol{e}_{i_s}\}$, and a possible choice for the matrix $\boldsymbol{U}$ of (12) is

$$\boldsymbol{U} = [\boldsymbol{e}_{i_1}, \ldots, \boldsymbol{e}_{i_s}] \quad \text{for } \|\boldsymbol{x}\|_0 = s. \tag{21}$$

The situation is different for points $\boldsymbol{x}$ having $\|\boldsymbol{x}\|_0 < s$. In this case, vectors $\boldsymbol{e}_i$ corresponding to *any* direction i are feasible directions, since $\|\boldsymbol{x}+\delta\boldsymbol{e}_i\|_0 \leq \|\boldsymbol{x}\|_0+1 \leq s$. It follows that $\mathcal{F} = \mathbb{R}^p$ and thus a convenient choice for the matrix $\boldsymbol{U}$ is

$$\boldsymbol{U} = \boldsymbol{I} \quad \text{for } \|\boldsymbol{x}\|_0 < s. \tag{22}$$

Consequently, whenever $\|\boldsymbol{x}\|_0 < s$, a specification of the $\mathcal{S}$-bias amounts to completely specifying the usual estimation bias $\boldsymbol{b}(\boldsymbol{x})$.

To invoke Theorem 1, we must also determine the FIM $\boldsymbol{J}(\boldsymbol{x})$. Under our assumption of white Gaussian noise, $\boldsymbol{J}(\boldsymbol{x})$ is given by [4, p. 85]

$$\boldsymbol{J}(\boldsymbol{x}) = \frac{1}{\sigma^2}\boldsymbol{H}^T\boldsymbol{H}. \tag{23}$$

Using (21), (22), and (23), it is readily shown that

$$\boldsymbol{U}^T\boldsymbol{JU} = \begin{cases} \frac{1}{\sigma^2}\boldsymbol{H}_{\boldsymbol{x}}^T\boldsymbol{H}_{\boldsymbol{x}} & \text{when } \|\boldsymbol{x}\|_0 = s \\ \frac{1}{\sigma^2}\boldsymbol{H}^T\boldsymbol{H} & \text{when } \|\boldsymbol{x}\|_0 < s \end{cases} \tag{24}$$

where $\boldsymbol{H}_{\boldsymbol{x}}$ is the $p \times s$ matrix consisting of the columns of $\boldsymbol{H}$ indexed by $\mathrm{supp}(\boldsymbol{x})$.

We now wish to determine the conditions under which (18) holds. Consider first points $\boldsymbol{x}_0$ for which $\|\boldsymbol{x}_0\|_0 = s$. Since, by (3), we have

$\text{spark}(\boldsymbol{H}) > s$, it follows that in this case $\boldsymbol{U}^T\boldsymbol{J}\boldsymbol{U}$ is invertible. Therefore $\mathcal{R}(\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{J}\boldsymbol{U}\boldsymbol{U}^T) = \mathcal{R}(\boldsymbol{U}\boldsymbol{U}^T)$. It thus follows from $\mathcal{R}(\boldsymbol{U}\boldsymbol{U}^T(\boldsymbol{I}+\boldsymbol{B}^T)) \subseteq \mathcal{R}(\boldsymbol{U}\boldsymbol{U}^T)$ that (18) holds when $\|\boldsymbol{x}_0\|_0 = s$.

However, (18) is no longer guaranteed when $\|\boldsymbol{x}_0\|_0 < s$. In this case, $\boldsymbol{U} = \boldsymbol{I}$, so that (18) is equivalent to $\mathcal{R}(\boldsymbol{I}+\boldsymbol{B}^T) \subseteq \mathcal{R}(\boldsymbol{H}^T\boldsymbol{H})$. Using the fact that $\mathcal{R}(\boldsymbol{H}^T\boldsymbol{H}) = \mathcal{R}(\boldsymbol{H}^T)$ and that, for any matrix $\boldsymbol{Q}$, $\mathcal{R}(\boldsymbol{Q}^T) = \mathcal{N}(\boldsymbol{Q})^\perp$, we find that (18) is equivalent to

$$\mathcal{N}(\boldsymbol{H}) \subseteq \mathcal{N}(\boldsymbol{I}+\boldsymbol{B}). \tag{25}$$

Combining these conclusions with Theorem 1 yields the following CRB for the problem of estimating a sparse vector.

Theorem 2. *Consider the estimation problem* (1) *with $\boldsymbol{x}_0$ given by* (2)*, and assume that* (3) *holds. For a finite-variance estimator $\widehat{\boldsymbol{x}}$ of $\boldsymbol{x}_0$ to exist, its bias gradient matrix $\boldsymbol{B}$ must satisfy* (25) *whenever $\|\boldsymbol{x}_0\|_0 < s$. Furthermore, the covariance of any estimator whose $\mathcal{S}$-bias gradient matrix is $\boldsymbol{B}\boldsymbol{U}$ satisfies*

$$\begin{aligned}\text{Cov}(\widehat{\boldsymbol{x}}) &\succeq \sigma^2(\boldsymbol{I}+\boldsymbol{B})(\boldsymbol{H}^T\boldsymbol{H})^\dagger(\boldsymbol{I}+\boldsymbol{B}^T) \\ &\qquad\qquad \textit{when } \|\boldsymbol{x}_0\|_0 < s, \\ \text{Cov}(\widehat{\boldsymbol{x}}) &\succeq \sigma^2(\boldsymbol{U}+\boldsymbol{B}\boldsymbol{U})(\boldsymbol{H}_{\boldsymbol{x}_0}^T\boldsymbol{H}_{\boldsymbol{x}_0})^{-1}(\boldsymbol{U}+\boldsymbol{B}\boldsymbol{U})^T \\ &\qquad\qquad \textit{when } \|\boldsymbol{x}_0\|_0 = s.\end{aligned} \tag{26}$$

Here, $\boldsymbol{H}_{\boldsymbol{x}_0}$ is the matrix containing the columns of $\boldsymbol{H}$ corresponding to $\text{supp}(\boldsymbol{x}_0)$.

Let us examine Theorem 2 separately in the well-determined and underdetermined cases. In the former setting, in which $\boldsymbol{H}$ has full row rank, the nullspace of $\boldsymbol{H}$ is trivial, so that (25) always holds. It follows that the CRB is always finite, in the sense that we cannot rule out the existence of an estimator having any given bias function. Some insight can be obtained in this case by examining the $\mathcal{S}$-unbiased case. Noting that in this setting $\boldsymbol{H}^T\boldsymbol{H}$ is invertible, the bound for $\mathcal{S}$-unbiased estimators is given by

$$\begin{aligned}\text{Cov}(\widehat{\boldsymbol{x}}) &\succeq \sigma^2(\boldsymbol{H}^T\boldsymbol{H})^{-1} && \text{when } \|\boldsymbol{x}_0\|_0 < s, \\ \text{Cov}(\widehat{\boldsymbol{x}}) &\succeq \sigma^2\boldsymbol{U}(\boldsymbol{H}_{\boldsymbol{x}_0}^T\boldsymbol{H}_{\boldsymbol{x}_0})^{-1}\boldsymbol{U}^T && \text{when } \|\boldsymbol{x}_0\|_0 = s.\end{aligned} \tag{27}$$

From this formulation, the behavior of the CRB can be described as follows. When $\boldsymbol{x}_0$ has non-maximal support ($\|\boldsymbol{x}_0\|_0 < s$), the CRB is identical to the bound which would have been obtained had there been no constraints in the problem. This is because $\boldsymbol{U} = \boldsymbol{I}$ in this case, so that $\mathcal{S}$-unbiasedness and ordinary unbiasedness are equivalent. The bound $\sigma^2(\boldsymbol{H}^T\boldsymbol{H})^{-1}$ is achieved by the unconstrained LS estimator

$$\widehat{\boldsymbol{x}} = (\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{y} \tag{28}$$

which is the minimum variance unbiased estimator in the unconstrained case. On the other hand, when $\boldsymbol{x}_0$ has maximal support, Theorem 2 states that $\mathcal{S}$-unbiased estimators can perform, at best, as well as the oracle estimator, which is equivalent to the LS approach when the support of $\boldsymbol{x}_0$ is known.

The situation is similar in the underdetermined setting, but the condition (25) for the existence of an estimator having a given bias gradient matrix no longer automatically holds in this case. To interpret this condition, it is helpful to introduce the mean gradient matrix $\boldsymbol{M}(\boldsymbol{x})$, defined as

$$\boldsymbol{M}(\boldsymbol{x}) = \frac{\partial E\{\widehat{\boldsymbol{x}}\}}{\partial \boldsymbol{x}} = \boldsymbol{I}+\boldsymbol{B}. \tag{29}$$

The matrix $\boldsymbol{M}(\boldsymbol{x})$ is a measure of the sensitivity of an estimator to changes in the parameter vector. Thus, $\mathcal{N}(\boldsymbol{M})$ denotes the subspace of directions to which $\widehat{\boldsymbol{x}}$ is insensitive. Likewise, $\mathcal{N}(\boldsymbol{H})$ is the subspace of directions for which a change in $\boldsymbol{x}$ does not modify $\boldsymbol{H}\boldsymbol{x}$. The condition (25) therefore states that for an estimator to exist, it must be insensitive to changes in $\boldsymbol{x}$ which are unobservable through $\boldsymbol{H}\boldsymbol{x}$, at least when $\|\boldsymbol{x}\|_0 < s$. No such requirement is imposed in the case $\|\boldsymbol{x}\|_0 = s$, since in this case there are far fewer feasible directions.

It follows from this analysis that when $\boldsymbol{H}$ is underdetermined, an estimator cannot be $\mathcal{S}$-unbiased for all $\boldsymbol{x}$. To see this, recall from (13) that $\mathcal{S}$-unbiased estimators are defined by the fact that $\boldsymbol{B}\boldsymbol{U} = \boldsymbol{0}$. When $\|\boldsymbol{x}\|_0 < s$, we have $\boldsymbol{U} = \boldsymbol{I}$ and thus $\mathcal{S}$-unbiasedness implies $\boldsymbol{B} = \boldsymbol{0}$, so that $\mathcal{N}(\boldsymbol{I}+\boldsymbol{B}) = \{\boldsymbol{0}\}$. But since $\boldsymbol{H}$ is underdetermined, $\mathcal{N}(\boldsymbol{H})$ is nontrivial, so that (25) cannot hold.

While an estimator cannot be unbiased for *all* $\boldsymbol{x} \in \mathcal{S}$, unbiasedness is possible at points $\boldsymbol{x}$ for which $\|\boldsymbol{x}\|_0 = s$. In this case, Theorem 2 produces a bound on the MSE of a $\mathcal{S}$-unbiased estimator, obtained by calculating the trace of (26) in the case $\boldsymbol{B}\boldsymbol{U} = \boldsymbol{0}$. This bound is given by

$$E\{\|\widehat{\boldsymbol{x}} - \boldsymbol{x}_0\|_2^2\} \geq \sigma^2\,\text{Tr}((\boldsymbol{H}_{\boldsymbol{x}_0}^T\boldsymbol{H}_{\boldsymbol{x}_0})^{-1}), \quad \|\boldsymbol{x}_0\|_0 = s. \tag{30}$$

The most striking feature of (30) is that it is identical to the oracle MSE (6). However, the CRB is of additional importance because of the fact that the ML estimator achieves the CRB in the limit when a large number of independent measurements are available, a situation which is equivalent in our setting to the limit $\sigma \to 0$. Thus, an MSE of (30) is achieved at high SNR by the ML approach (4). While the ML approach is computationally intractable in the sparse estimation setting, it is still implementable in principle, as opposed to $\widehat{\boldsymbol{x}}_{\text{oracle}}$, which relies on unavailable information (namely, the support set of $\boldsymbol{x}_0$). Thus, Theorem 2 gives an alternative interpretation to comparisons of estimator performance with the oracle.

Note that when examining performance averaged over realizations of a random measurement matrix $\boldsymbol{H}$, it has previously been shown that the oracle MSE can be achieved under much weaker conditions, namely when m, p and s simultaneously tend to infinity [10]. In the present setting, however, we consider a fixed, pre-specified matrix $\boldsymbol{H}$, which is only assumed to specify the spark requirement (3); this is a more difficult estimation setting and thus stronger conditions are required to guarantee oracle-quality performance.

Before concluding this section, we briefly consider the estimation setting in which the parameter $\boldsymbol{x}_0$ is not sparse itself, but has a sparse representation, i.e., the situation in which there exists a known dictionary $\boldsymbol{D}$ such that $\boldsymbol{x}_0 = \boldsymbol{D}\boldsymbol{\alpha}_0$, for some $\|\boldsymbol{\alpha}_0\|_0 \leq s$. We are interested in estimating $\boldsymbol{x}_0$ from measurements (1); thus, only the constraint set has changed. If $\boldsymbol{D}$ is a unitary matrix, then estimating $\boldsymbol{x}_0$ is equivalent to estimating $\boldsymbol{\alpha}_0$, which can be rewritten in the form of the sparse estimation setting (1)–(2). However, in many cases $\boldsymbol{D}$ is not unitary, and indeed the dictionary is often overcomplete (i.e., column rank deficient). In such cases, while the estimation problem is different, the constraint set is still locally balanced, so that one can again apply Theorem 1 to determine the CRB. The resulting bound shares many of the characteristics of Theorem 2 described above. In particular, when $\boldsymbol{x}_0 = \boldsymbol{D}\boldsymbol{\alpha}_0$ with $\|\boldsymbol{\alpha}_0\|_0 < s$, the CRB equals the unconstrained bound, whereas when $\|\boldsymbol{\alpha}_0\|_0 = s$, the CRB coincides with the case in which the nonzero locations of $\boldsymbol{\alpha}_0$ are known. Due to space limitations, further details concerning this estimation setting are given in [9].

V. Discussion

In this paper, we extended the CRB to constraint sets satisfying the local balance condition (Theorem 1). This enabled us to derive a lower bound on the achievable performance in sparse estimation problems (Theorem 2). In simple terms, Theorem 2 can be summarized as follows. The behavior of the CRB differs depending on whether or not the parameter has maximal support (i.e., $\|\boldsymbol{x}_0\|_0 = s$). In the case of maximal support, the bound equals that which would be obtained if the sparsity pattern were known; this can be considered an "oracle bound". On the other hand, when $\|\boldsymbol{x}_0\|_0 < s$, performance is identical to the unconstrained case, and the bound is substantially higher. We now discuss some practical implications of this result. To simplify the discussion, we consider the case of unbiased estimators, though analogous conclusions can be drawn for any bias function.

When $\|\boldsymbol{x}_0\|_0 = s$ and all nonzero elements of $\boldsymbol{x}_0$ are considerably larger than the standard deviation of the noise, the support set can be recovered correctly with high probability (at least if computational considerations are ignored). Thus, in this case an estimator can mimic the behavior of the oracle, and the CRB is expected to be tight. Indeed, in the high SNR limit, the ML estimator achieves the unbiased CRB. The proposed bound can thus be viewed as an alternative justification for the common use of the oracle estimator as a baseline against which practical algorithms are compared, at least in the high SNR regime. This gives further merit to recent results, which demonstrate that a variety of algorithms achieve near-oracle performance [2], [15]. On the other hand, when the support of $\boldsymbol{x}_0$ is not maximal, the unbiasedness requirement demands sensitivity to changes in all components of $\boldsymbol{x}_0$, and consequently the bound coincides with the unconstrained CRB. Thus, in underdetermined cases no estimator is unbiased for all $\boldsymbol{x}_0 \in \mathcal{S}$. This emphasizes the need for bias in the construction of successful estimation algorithms. In particular, shrinkage techniques are designed to reduce MSE by lowering the variance at the expense of a small gain in bias.

An interesting observation can also be made concerning maximal-support points $\boldsymbol{x}_0$ for which some of the nonzero elements are close to zero. The CRB in this "low-SNR" case corresponds to the oracle MSE, but as we will see, the bound is loose for such values of $\boldsymbol{x}_0$. Intuitively, at low-SNR points, any attempt to recover the sparsity pattern will occasionally fail. Consequently, despite the optimistic CRB, it is unlikely that the oracle MSE can be achieved. Indeed, the covariance matrix of any finite-variance estimator is a continuous function of $\boldsymbol{x}_0$ [20], and the fact that performance is bounded by the (much higher) unconstrained bound when $\|\boldsymbol{x}_0\|_0 < s$ implies that performance must be similarly poor when the SNR is low.

This excessive optimism is a result of the local nature of the CRB: The bound depends on the estimation setting only in an ε-neighborhood of the parameter itself. Indeed, the CRB depends on the constraint set only through the feasible directions, which were defined in Section III-B as those directions which do not violate the constraints for *sufficiently small* deviations. Thus, for the CRB, it is entirely irrelevant if some of the components of $\boldsymbol{x}_0$ are close to zero, as long as $\operatorname{supp}(\boldsymbol{x}_0)$ is held constant. A tighter bound for sparse estimation problems can be obtained using a non-local technique such as the Hammersley–Chapman–Robbins approach, and will be described in a forthcoming publication [21].

Acknowledgements

The authors would like to thank Yaniv Plan for helpful discussions. The authors are also grateful to the anonymous reviewers and to the associate editor for their constructive comments.

References

[1] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030–1051, 2006.

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

[3] E. J. Candès, "Modern statistical estimation via oracle inequalities," *Acta Numerica*, pp. 1–69, 2006.

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

[5] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Trans. Inf. Theory*, vol. 26, no. 6, pp. 1285–1301, Nov. 1990.

[6] P. Stoica and B. C. Ng, "On the Cramér–Rao bound under parametric constraints," *IEEE Signal Process. Lett.*, vol. 5, no. 7, pp. 177–179, 1998.

[7] Z. Ben-Haim and Y. C. Eldar, "On the constrained Cramér-Rao bound with a singular Fisher information matrix," *IEEE Signal Process. Lett.*, vol. 16, no. 6, pp. 453–456, Jun. 2009.

[8] ——, "Performance bounds for sparse estimation with random noise," in *Proc. IEEE Workshop on Statistical Signal Processing*, Cardiff, Wales, UK, Sep. 2009.

[9] ——, "The Cramér–Rao bound for sparse estimation," Dept. of Electrical Engineering, Technion, Haifa, Israel, Tech. Rep. 1714, Feb. 2010. [Online]. Available: http://arxiv.org/abs/0905.4378

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

[11] K. Engan, S. O. Aase, and J. H. Husøy, "Multi-frame compression: Theory and design," *Signal Processing*, vol. 80, no. 10, pp. 2121–2140, 2000.

[12] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[13] Z. Ben-Haim, Y. C. Eldar, and M. Elad, "Near-oracle performance of basis pursuit under random noise," Mar. 2009. [Online]. Available: http://arxiv.org/abs/0903.4579

[14] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *Proc. Nat. Acad. Sci. USA*, vol. 100, no. 5, pp. 2197–2202, March 4, 2003.

[15] Z. Ben-Haim, Y. C. Eldar, and M. Elad, "Coherence-based performance guarantees for estimating a sparse vector under random noise," *IEEE Trans. Signal Process.*, 2009, submitted. [Online]. Available: http://arxiv.org/abs/0903.4579

[16] Z. Ben-Haim and Y. C. Eldar, "Blind minimax estimation," *IEEE Trans. Inf. Theory*, vol. 53, no. 9, pp. 3145–3157, Sep. 2007.

[17] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.

[18] Y. C. Eldar, "Compressed sensing of analog signals in shift-invariant spaces," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 2986–2997, Aug. 2009.

[19] K. Gedalyahu and Y. C. Eldar, "Time delay estimation from low rate samples: A union of subspaces approach," *IEEE Trans. Signal Process.*, 2010, to appear. [Online]. Available: http://arxiv.org/abs/0905.2429

[20] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, 2nd ed. New York: Springer, 1998.

[21] A. Jung, Z. Ben-Haim, F. Hlawatsch, and Y. C. Eldar, "On unbiased estimation of sparse vectors corrupted by Gaussian noise," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP 2010)*, Dallas, TX, Mar. 2010, to appear. [Online]. Available: http://arxiv.org/abs/1002.0110