

# On the Non-Existence of Unbiased Estimators in Constrained Estimation Problems

Anelia Somekh-Baruch, Amir Leshem and Venkatesh Saligrama

**Abstract**—We address the problem of existence of unbiased constrained parameter estimators. We show that if the constrained set of parameters is compact and the hypothesized distributions are absolutely continuous with respect to one another, then there exists no unbiased estimator. Weaker conditions for the absence of unbiased constrained estimators are also specified. We provide several examples which demonstrate the utility of these conditions.

**Index Terms**—Cramer-Rao bound, unbiased estimation, Estimation theory, constrained estimators

## I. INTRODUCTION

Unbiased estimation is a cornerstone of classical estimation theory and arises as an important concept in minimum variance unbiased estimation (MVUE) theory [1], [2], [3]. Information theoretic bounds are an important tool for the evaluation of estimators for identifying theoretical performance gaps. [4], [5], [6], [7].

In this context, the Cramer-Rao bound (CRB) has emerged as an important information inequality since it not only serves as a non-trivial lower bound on all unbiased estimators, but is also often readily computable particularly for unconstrained parameter estimation problems. While CRB can also be extended to bound the performance of biased estimators, these bounds are often not useful in many interesting cases. This is because they require the knowledge of the bias as a function of the parameters and are thus not agnostic to the estimation scheme employed.

In many real-world estimation problems we often encounter constraints on the parameter space in the form of side-information. For example in many communication systems we encounter positivity constraints, limited power constraints, bandwidth or delay constraints, circularity constraints, subspace constraints, and so on. Consequently, it makes sense to

incorporate these constraints for deriving constraint dependent lower bounds. Motivated by these scenarios a significant amount of research has been conducted on developing estimation techniques and performance bounds for constrained problems. In this context, many papers propose deriving constrained CRBs and in particular focus on bounds for unbiased estimators on account of its relative computational simplicity. This raises the question of as to the conditions under which unbiased estimators can exist in constrained scenarios. Before we describe our results we will discuss some of the related work in this context below.

CRBs for constrained parameters was introduced and studied by Gorman and Hero [8]. A constrained CRB on the error covariance of estimators of multi-dimensional parameters was derived. The derivation was based on a limiting form of a multi-parameter Chapman-Robbins [9] form of a Barankin-type bound [4]. It was shown that the constrained CRB is equivalent to the unconstrained CRB evaluated with a “constrained Fisher information matrix” for the case in which there is a general smooth functional inequality constraint of the form  $\mathcal{G}_\theta \leq 0$ . This constrained Fisher matrix was shown to be identical to the classical unconstrained Fisher matrix at all regular points of the constraint set, e.g., at interior points. However as noted by [8] at non-regular points, such as points governed by equality constraints, the constrained Fisher matrix was observed to be a rank-deficient matrix. It was also established that functional constraints necessarily decrease the CR bound for unbiased estimators.

Marzetta [10] provided a simpler proof for the case of constrained *unbiased* estimators. While Gorman and Hero’s derivation relied on an application of the Cauchy-Schwarz inequality to a pair of random vectors, one of which has a possibly singular covariance matrix, the derivation of Marzetta avoids using the Cauchy-Schwarz inequality and it avoids pseudoinverses. But, it focuses solely on unbiased estimators. Another result of [10] is a necessary condition for an unbiased estimator to achieve the constrained CRB with equality.

A subsequent work by Stoica and Ng [11] extends the constrained CRB (under differentiable, deterministic constraints on the parameters) to the case in which the Fisher information matrix for the unconstrained problem is not necessarily full rank. This case was not treated in previous works. It was assumed in [11] that the estimator was unbiased and it was shown that the expression for the constrained CRB in this case depends only on the unconstrained Fisher information matrix and a basis of the nullspace of the constraints gradient matrix. A necessary and sufficient condition for the existence of the constrained CRB was also derived.

A. Somekh-Baruch and Amir Leshem are with the Faculty of Engineering at Bar-Ilan University, Ramat-Gan, Israel. Emails: somekha@biu.ac.il, leshema@biu.ac.il. Amir Leshem’s work was funded by the visiting scholar program of Boston University. Venkatesh Saligrama is with the Department of Electrical and Computer Engineering, Boston University, Email: srv@bu.edu. This material is based upon work supported in part by NSF Grants CCF: 1320566, NSF Grant CNS: 1330008 NSF CCF: 1527618, the U.S. Department of Homeland Security, Science and Technology Directorate, Office of University Programs, under Grant Award 2013-ST-061-ED0001, by ONR Grant 50202168 and US AF contract FA8650-14-C-1728. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the social policies, either expressed or implied, of the NSF, U.S. DHS, ONR or AF. The paper was presented in part at the Information Theory and Application workshop ITA2017, in San-Diego, USA, Feb. 2017.

Copyright (c) 2017 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

There have been a number of follow-on works that have utilized constrained CRBs for a range of applications including sparse estimation problems [12] and MIMO Radars [13].

Unfortunately, as we shall see in this paper, in many interesting cases of constrained parameter sets, an unbiased estimator does not exist. This limits the applicability of the results, mainly to cases of biased estimators, and to cases of unbiased estimators in which the constrained parameter set is a manifold without boundaries [14], i.e., there are no non-trivial inequality constraints. Moreover, often, the use of the CRBs for biased estimators is quite complicated since it requires explicit expressions for both the bias and the gradient of the bias. *One of the most striking consequences of the results of this paper is that if the constraint set is the set of solutions (zeros) of a set of continuous non-linear functions which is bounded, then there exists no unbiased estimator.*

This result in turn provides an explanation for the observation of [8] that inequality constraints do not reduce the bound at internal points of the constrained parameter set. The variance is not reduced simply because there are no constrained unbiased estimators in this case. This implies that in order to use the Gorman-Hero bound, one needs to know the bias of the estimator with the constraint. In fact, in another related work [15], a method for experimentally determining the bias gradient norm is presented in order to explore bias-variance tradeoffs.

In this paper we focus on the case of an estimation problem of a parameter vector  $\theta$  that lies in  $\mathbb{R}^d$ . Our first result concerns the case where  $\theta$  is constrained to a set, which has at least one extreme point,  $\theta_e$ . In addition we assume that the distribution of the observed random variable  $X$  given  $\theta_e$  is absolutely continuous with respect to (w.r.t.) the distribution of  $X$  given another parameter value in the constrained set. Under these assumptions we show that there exists no unbiased estimator whose range is the constrained set. We then specialize the result to the important case of a constrained parameter set which is compact (and not a singleton) and show that the result continues to hold in this special case. Furthermore, we extend the result to a case of a bounded constraint set, which can be open, and whose closure contains an extreme point and where the distribution of the observed random variable has a bounded Radon Nikodym derivative. We further provide examples to demonstrate that the conditions we specify are necessary and cannot be further relaxed.

Our results imply that almost in every constrained problem that one can think of, there exists no unbiased estimator. This result is surprising in light of the scarcity of examples which appear in the literature for the non existence of unbiased constrained estimators (e.g. [16]). In fact, the non-existence of unbiased estimators is the more common/natural case, a fact that will be defined more explicitly. Moreover, this fact has an important implication on CR bounds in the constrained case; since the applicable case is the biased one, the CR bound which requires explicit expressions for both the bias and the gradient of the bias might be useless in many setups.

## II. MAIN RESULTS

We begin by presenting a theorem which specifies conditions under which an unbiased estimator of a vector of parameters does not exist.

**Lemma 1.** *Let  $\Theta \subseteq \mathbb{R}^d$  be a set of parameters which is not a singleton, and let  $P_\theta$ ,  $\theta \in \Theta$  be a corresponding collection of distributions. Let  $\mathbf{X} \sim P_\theta$  be a random variable over some measurable space. Suppose there exists an extreme point  $\theta_e \in \Theta$  and an additional point  $\theta' \in \Theta$ ,  $\theta' \neq \theta_e$ , such that  $P_{\theta'}$  and  $P_{\theta_e}$  are absolutely continuous w.r.t. one another. Then there exists no unbiased estimator  $g(\mathbf{X})$  for  $\theta$ , which is measurable w.r.t. the Lebesgue  $\sigma$ -algebra whose range is  $\Theta$ .*

Note that Lemma 1 as well as the proceeding Theorem 1 do not pose restrictive assumptions on the measurable space on which the random variable  $\mathbf{X}$  is defined. Hence, as a special case, it holds for  $\mathbf{X}$  which is a random  $N$ -Vector over  $\mathbb{R}^N$ .

Note also that absolute continuity and contiguity can be formulated in terms of Rényi divergences as pointed out in [17].

*Proof.* Let  $\theta_e$  be an extreme point of  $\Theta$ , the existence of which is guaranteed by assumption.

Assume that  $g(\mathbf{X})$  is an unbiased estimator of  $\theta$  whose range is  $\Theta$ . For notational convenience we denote by  $P(g(\mathbf{X})|\theta)$  the distribution of  $g(\mathbf{X})$ . By unbiasedness we have

$$\mathbb{E}_{P_{\theta_e}}(g(\mathbf{X})|\theta_e) = \theta_e, \quad (1)$$

where  $\mathbb{E}_P(\cdot)$  denote expectation w.r.t. the distribution  $P$ . Therefore, by definition of the expectation we obtain

$$\begin{aligned} \theta_e &= \mathbb{E}_{P_{\theta_e}}(g(\mathbf{X})|\theta_e) \\ &= \int_{\mathcal{X}} g(\mathbf{X}) dP_{\theta_e} \\ &= \int_{\Theta} g(\mathbf{X}) dP(g(\mathbf{X})|\theta_e), \end{aligned} \quad (2)$$

where  $\mathcal{X}$  is the alphabet of  $\mathbf{X}$  and the last step follows from the assumption that the range of  $g$  is  $\Theta$ .

While the RHS of (2) is a convex combination of points in  $\Theta$ , the LHS of (2), that is,  $\theta_e$ , is an extreme point of  $\Theta$ . By definition of an extreme point in  $\Theta$  as a point which does not lie in any open line segment joining two points of  $\Theta$ , we conclude that

$$g(\mathbf{X}) \equiv \theta_e \text{ w.p. 1 w.r.t } P_{\theta_e}. \quad (3)$$

Since by assumption, the distributions  $P_{\theta'}$  and  $P_{\theta_e}$  are absolutely continuous w.r.t. one another it follows that

$$g(\mathbf{X}) \equiv \theta_e \text{ w.p. 1 w.r.t } P_{\theta'}. \quad (4)$$

To realize that (4) holds, assume in negation that it does not. Thus  $\exists \epsilon > 0$  such that  $P_{\theta'}(\{x : g(x) \neq \theta_e\}) > \epsilon$ , while from (3) we have  $P_{\theta_e}(\{x : g(x) \neq \theta_e\}) = 0$ , contradicting the absolute continuity of  $P_{\theta'}$  and  $P_{\theta_e}$  w.r.t. one another.

Equation (4) clearly contradicts the unbiasedness of the estimator for  $\theta'$ .  $\square$

The following theorem gives alternative (stricter) conditions for the non existence of an unbiased estimator.

**Theorem 1.** Let  $\Theta \subseteq \mathbb{R}^d$  be a compact set of parameters which is not a singleton. Assume further that  $P_\theta, \theta \in \Theta$  is a collection of distributions, which are absolutely continuous w.r.t. one another. Let  $\mathbf{X} \sim P_\theta$  be a random variable over some measurable space.

Then there exists no unbiased estimator  $g(\mathbf{X})$  for  $\theta$ , which is measurable w.r.t. Lebesgue  $\sigma$ -algebra whose range is  $\Theta$ .

*Proof.* The proof of Theorem 1 is a straightforward consequence of Lemma 1, which is established by proving that a compact set in  $\mathbb{R}^d$  must contain at least one extreme point. To this end, we use a known result, which can be found in [18, Page 185, Corollary 5.33]. This corollary states that a convex hull and the convex circled hull of a compact subset of a finite dimensional vector space are compact sets. Therefore, the convex hull of a compact set in  $\mathbb{R}^d$  is compact.

Further, we note that the set of extreme points of a compact set in  $\mathbb{R}^d$  is non-empty: By the Krein-Milman Theorem [19], the closure of  $\text{Conv}(\mathcal{A})$  is the convex hull of its extreme points. If  $\mathcal{A}$  is compact, then by the previous lemma,  $\text{Conv}(\mathcal{A})$  is closed. Therefore,  $\text{Conv}(\mathcal{A})$  is the convex hull of its extreme points. Hence, all the extreme points of  $\text{Conv}(\mathcal{A})$  must belong to  $\mathcal{A}$ . By definition, an extreme point is not a convex combination of other points in the set, hence all the extreme points must belong to  $\mathcal{A}$ .

Hence, we have established the existence of an extreme point in  $\Theta$ . Now, repeating steps (1)-(3) the condition that  $P_\theta, \theta \in \Theta$  are absolutely continuous w.r.t. one another, ensures that (4) holds for all  $\theta' \neq \theta_e$ , which contradicts the unbiasedness of the estimator for all  $\theta' \neq \theta_e$ .  $\square$

We next present a corollary of Theorem 1.

**Corollary 1.** Let  $f$  be a continuous non-linear function and let the set  $\Theta = \{\theta : f(\theta) = 0\}$  of solutions be bounded, then there is no unbiased estimator. The same result holds for a bounded set  $\Theta = \{\theta : f(\theta) \leq 0\}$ .

*Proof.* By the boundedness and the continuity of  $f$  we have that the set  $\Theta$  must also be closed and thus also compact. Therefore, the conditions of Theorem 1 are met and there exists no unbiased estimator.  $\square$

Next, we relax the condition that there must exist an extreme point within the constraint set, and replace it with a requirement that an extreme point exists on its closure. This allows to generalize the results to open constraint sets under an additional requirement on set of distributions  $P_\theta, \theta \in \Theta$ . We denote by  $\bar{\mathcal{A}}$  the closure of the set  $\mathcal{A}$ .

**Theorem 2.** Let  $\Theta \subseteq \mathbb{R}^d$  be a bounded set of parameters that is not a singleton. Let  $P_\theta, \theta \in \bar{\Theta}$  be a collection of distributions which are absolutely continuous w.r.t. the Lebesgue measure and w.r.t. one another, with Radon Nikodym derivatives (p.d.f.'s)  $f_\theta, \theta \in \bar{\Theta}$  which are uniformly bounded. Let  $\mathbf{X} \sim P_\theta$  be a random variable over some measurable space. Suppose there exists an extreme point  $\theta_e \in \bar{\Theta}$  and that there exists a sequence  $\theta_n$  which satisfies  $\lim_{n \rightarrow \infty} \theta_n = \theta_e$

and the sequence  $f_{\theta_n}$  converges globally in Lebesgue measure<sup>1</sup> to the p.d.f.  $f_{\theta_e}$ . Then there exists no unbiased estimator  $g(\mathbf{X})$  for  $\theta$ , which is measurable w.r.t. the Lebesgue  $\sigma$ -algebra whose range is  $\Theta$ .

*Proof.* Let  $g(\mathbf{X})$  be a given estimator. Consider the sequence of measurable mappings  $A_n = g(\mathbf{X}) \cdot f_{\theta_n}(\mathbf{X}), n=1,2,\dots$ . By the assumptions and by definition,  $A_1, A_2, \dots$  is a uniformly integrable sequence which converges globally in Lebesgue measure to the limit  $A = g(\mathbf{X}) \cdot f_{\theta_e}(\mathbf{X})$ . Hence, there exists a subsequence  $A_{n_k}, k = 1, 2, \dots$  which converges a.s. to  $A$ .

Since  $f_{\theta_n}(\mathbf{X})$  is uniformly bounded by say  $M < \infty$ , and since  $g(\mathbf{X})$  must lie in  $\Theta$  which is also assumed to be bounded, one can invoke the Lebesgue Dominated Convergence Theorem (DCT) yielding

$$\lim_{k \rightarrow \infty} \mathbb{E}_{P_{\theta_{n_k}}}(g(\mathbf{X})) = \mathbb{E}_{P_{\theta_e}}(g(\mathbf{X})). \quad (5)$$

On the other hand, from the unbiasedness we have  $\mathbb{E}_{P_{\theta_{n_k}}}(g(\mathbf{X})) = \theta_{n_k}$  and therefore,

$$\lim_{k \rightarrow \infty} \mathbb{E}_{P_{\theta_{n_k}}}(g(\mathbf{X})) = \lim_{k \rightarrow \infty} \theta_{n_k} = \theta_e, \quad (6)$$

where the right inequality follows by assumption.

Equations (5) and (6) imply that

$$\mathbb{E}_{P_{\theta_e}}(g(\mathbf{X})) = \theta_e, \quad (7)$$

which means that the estimator is unbiased also at the point  $\theta_e$ . Therefore, the conditions of Lemma 1 are satisfied with  $\Theta$  substituted by  $\Theta \cup \theta_e$  and the theorem follows.  $\square$

We now generalize Theorem 2 to case where we the collection of measures are tight.

**Theorem 3.** Consider a bounded set  $\Theta \subset \mathbb{R}^d$  of parameters whose closure contains an extreme point  $\theta_e$ . Let  $P_\theta, \theta \in \Theta$  over observation space  $\mathcal{X}$  be a collection of tight probability measures on  $\mathbb{R}^n$ , with uniformly bounded Radon Nikodym derivatives w.r.t. the Lebesgue measure. Then every measurable estimator of the form  $g : \mathcal{X} \rightarrow \Theta$  is necessarily biased.

*Proof.* Let  $\theta_n$  be a sequence in the interior of  $\Theta$  converging to an extreme point  $\theta_e$ . Since the collection of probability measures is tight, it follows by Prokhorov's Theorem [21] that the corresponding sequence of measures,  $P_{\theta_n}$  is sequentially compact. Therefore, there is a subsequence,  $P_{\theta_{n_k}}$  that converges weakly to a limiting measure  $P_{\theta_e}$ .

Next we invoke Skorohod's Representation Theorem, which claims that weak convergence of measures implies corresponding existence of random variables on a common probability space with almost sure convergence. To this end, we construct random variables  $Z_{n_k}$  and  $Z$  on a common probability space  $([0, 1], \mathcal{B}, \lambda)$  with Lebesgue measure  $\lambda$  such that  $Z_{n_k} \sim P_{\theta_{n_k}}, Z \sim P_{\theta_e}$  with  $Z_{n_k} \xrightarrow{a.s.} Z$ . We now note that for an arbitrary unbiased estimator  $g : \mathcal{X} \rightarrow \Theta$  we have,

$$\theta_{n_k} = E_{P_{n_k}}[g(X)] = E_\lambda[g(X)f_{n_k}(X)] \xrightarrow{k \rightarrow \infty} \theta_e \quad (8)$$

<sup>1</sup>Global convergence in Lebesgue measure of the sequence  $f_{\theta_n}$  to  $f_{\theta_e}$  means that for every  $\epsilon > 0$ ,  $\lim_{n \rightarrow \infty} \mu(\{x : |f_{\theta_n}(x) - f_{\theta_e}(x)| > \epsilon\}) = 0$ ,  $\mu$  being the Lebesgue measure.

Note that for exponential families of measures other notions of convergence and closure have been discussed by Csiszár and Matúš [20].

where  $f_n(X)$ ,  $f(X)$  are the Radon-Nikodym derivative of  $P_{n_k}$ ,  $P_e$  w.r.t.  $\lambda$ . Note that the limiting statement above follows from our assumption that  $\theta_{n_k} \rightarrow \theta_e$ .

We are left to prove that the following equality is satisfied.

$$E_\lambda[g(X)f(X)] \triangleq E_{P_e}[g(X)] = \lim_{k \rightarrow \infty} E_{P_{n_k}}[g(X)] \triangleq \lim_{k \rightarrow \infty} E_\lambda[g(X)f_n(X)] \triangleq \theta_e. \quad (9)$$

We claim that this is an instance of the DCT. To show this note that  $\|g(X)\| \leq \max_{\theta \in \Theta} \|\theta\|$  is bounded and by the fact that the Radon-Nikodym derivatives,  $f_{n_k}(\cdot)$  are uniformly bounded. This implies that  $\|g(X)f_n(X)\| \leq C$  for some constant  $C > 0$ . By the Skorohod Representation Theorem we know that,  $f_{n_k}(\cdot)$  corresponding to the random variable  $Z_{n_k}$ , must converge pointwise almost everywhere to  $f(\cdot)$ . Consequently,  $\|g(X)f(X)\| \leq C$  almost everywhere. By the DCT we now have a sequence of functions  $g(X)f_n(X)$  that is bounded by an integrable function, namely,  $C\mathbb{I}_{[0,1]}$  under the Lebesgue measure and converges pointwise to  $g(X)f(X)$ . Therefore, our claim follows. We are now in the position of Theorem 2 and proof follows in an identical fashion.  $\square$

### III. EXAMPLES AND DISCUSSION

In this section we present some examples of unbiased constrained estimators, and characterize some important cases in which unbiased estimators do not exist.

#### A. Example: IID Bernoulli Random Variables

The classical example for unbiased (and even MVU) estimator over a compact parameter space is the estimation of the parameter of an IID sequence of Bernoulli random variables with parameter  $0 \leq p \leq 1$ . In this case, the sample mean is the best unbiased estimator. However, at the maximal and minimal values of the parameter, the distribution of the measurements is concentrated at a single value, i.e., when  $p \in \{0, 1\}$  all the measurements assume the value  $p$  exactly. This is not a coincidence as we saw above; had the distribution of the measurements with  $p = 0$  been absolutely continuous w.r.t. any other distribution with  $p \in (0, 1)$ , an unbiased estimator could not have existed.

We note that extended exponential families are relevant in this context; in certain cases, an unbiased estimate may not exist within an exponential family. Extending the exponential family by including distributions whose supports have positive probability on the face of the convex support (adding elements which are not mutually absolutely continuous) may solve the problem enabling the existence of an unbiased estimator, see [22]. However, for certain types of constraints this is not possible, e.g., if we constrain the parameter  $p$  to lie in the interval  $[1/4, 3/4]$ .

Example III-A can now be generalized to obtain non-trivial constraints: Assume that  $\theta \in [0, 2]$  and that given  $\theta$ ,  $X$  is a random vector with value  $\lfloor \theta \rfloor + z$  where  $z$  is a Bernoulli random IID vector with mean  $\theta - \lfloor \theta \rfloor$ . Similar to the example above we can find an unbiased estimator of  $\theta$  when we require  $1 \leq \theta \leq 2$ .

Moreover, in [16, Chapter 7.12], the problem of estimating the odds ratio  $\frac{p}{1-p}$  in  $N$  independent Bernoulli trials with

probability of success  $p$  is considered, as an example of non-existence of unbiased estimator. Any statistics  $T$  which maps the observable binary string of length  $N$  to a real number, would result in an expectation  $\mathbb{E}(T) = \sum_{j=1}^{2^N} t_j p^{n_j} (1-p)^{n-n_j}$ . Thus, an unbiased estimator should satisfy the condition that this polynomial would be equal to  $\frac{p}{1-p}$  for all  $p \in (0, 1)$ , which is clearly impossible.

#### B. Example: Estimation of the Variance of IID Gaussian Random Variables

A second interesting example is the case of estimating the variance of a Gaussian random vector from a sequence of  $N$  IID measurements  $X \sim \mathcal{N}(0, \theta \mathbf{I})$ . An unbiased estimator exists with the constraint  $\theta \geq 0$ . However, as a consequence of our results, one cannot find an unbiased estimator with any other constraint of the form  $\theta \geq c$ , where  $c$  is strictly positive. Since the point  $\theta_e = c$  is an extreme point of the interval  $[c, \infty)$ , and for any other  $\theta > c$ ,  $P_\theta$  is absolutely continuous w.r.t.  $P_{\theta_e}$  (as both are absolutely continuous w.r.t. the Lebesgue measure on  $\mathbb{R}^N$ ), there exists no unbiased estimator of  $\theta$ , which is confined to  $[c, \infty)$ . On the other hand, for  $\theta \geq 0$  we obtain that although  $\theta = 0$  is an extreme point, since a zero variance random variable is deterministic, its distribution is not absolutely continuous w.r.t. any other positive-variance Gaussian distribution, and therefore the conditions of Theorem 1 do not hold and indeed an unbiased estimator for  $\theta$  which lies in  $[0, \infty)$  exists.

Further, if one considers the open constraint set  $\Theta = (c_1, c_2)$ , the conditions of Theorem 2 are satisfied, and again there exists no unbiased estimator.

#### C. Example: Spectrum Estimation with Power Constraints:

We next revisit Example 3 discussed in [8]: Let  $(X_1, \dots, X_N)$  be a segment of a real wide sense stationary random process with power spectral density (PSD)  $S_X(f)$ ,  $f \in [-1/2, 1/2]$ . The objective is to estimate the PSD, at  $d$  distinct frequencies  $\theta_i = S_X(f_i)$ ,  $f_1, \dots, f_d$ . It is assumed that the average power of  $\{X_i\}$  is known over  $p$  non-overlapping frequency bands, i.e., the vector  $\theta^d = (\theta_1, \dots, \theta_d)$  is constrained to satisfy the equations

$$\begin{bmatrix} \chi_{1,1} & \chi_{1,2} & \cdots & \chi_{1,d} \\ \chi_{2,1} & \chi_{2,2} & \cdots & \chi_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ \chi_{p,1} & \chi_{p,2} & \cdots & \chi_{p,d} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_d \end{bmatrix} = \begin{bmatrix} \mathbb{E}_1 \\ \mathbb{E}_2 \\ \vdots \\ \mathbb{E}_p \end{bmatrix}, \quad (10)$$

where  $\{\chi_{i,j}\}$ ,  $i = 1, \dots, p$ ,  $j = 1, \dots, d$  is a specified matrix with binary  $(0, 1)$  entries, and  $\mathbb{E}_i$ ,  $i = 1, \dots, p$  are specified power levels. As mentioned in [8], these are in fact  $p$  linear constraints on the unknown PSD, known as the  $p$ -point constraint in robust Wiener filtering theory. In this example, the constraint set has no extreme points and therefore an unbiased estimator can exist.

#### D. Example: Intersection of Polyhedral Sets

Consider the case of inequality constraints

$$\sum_{i=1}^d \alpha_{i,k} \theta_i \leq B_k, \quad k = 1, \dots, d, \quad (11)$$

where  $B_k, \alpha_{i,k}, i \in \{1, \dots, d\}, k \in \{1, \dots, d\}$  are real constants. In the case in which the resulting set of allowable  $\theta$ 's in (11) defines a polyhedron, or even a set which possesses an extreme point, there exists no unbiased estimator for  $(\theta_1, \dots, \theta_d)$ .

#### E. Example: Continuous Non-linear Function with Equality Constraints

Revisiting the case discussed in Corollary 1, of a bounded set  $\Theta = \{\theta : f(\theta) = 0\}$  or  $\Theta = \{\theta : f(\theta) \leq 0\}$  of solutions; one such example is a  $p$ -norm equality constraint, i.e.,  $\|\theta\|_p = R$ . Other examples are when the parameter satisfies continuous non-linear inequality constraints, e.g.,  $\|\theta\| \leq R$  or when  $\theta$  belongs to an ellipsoid in  $\mathbb{R}^d$ . In these cases too the same argument holds since the boundary is the equality set.

#### F. Example: Constrained Sparse Estimation and Compressed Sensing Problem

Consider an observed vector

$$\mathbf{X} = A\theta + \mathbf{Z} \quad (12)$$

where  $\mathbf{Z}$  is a Gaussian  $n$ -vector,  $A$  is a known deterministic  $n \times d$  matrix, and  $\theta = (\theta_1, \dots, \theta_d)^T$  is a sparse  $d$ -vector of unknown parameters which satisfies

$$\sum_{i=1}^d |\theta_i| \leq 1. \quad (13)$$

Since the conditions of Theorem 1 are satisfied, there exists no unbiased estimator  $g(\mathbf{X}) = (g_1(\mathbf{X}), \dots, g_d(\mathbf{X}))$  for  $\theta$  which satisfies  $\sum_{i=1}^d |g_i(\mathbf{X})| \leq 1$ .

*Compressive Sensing:* Note that in the context of compressed sensing, many works typically consider the matrix  $A$  to be random (see [23], [24]). In these contexts one desires estimators  $\theta_A(X)$ , namely, estimators that attempt to estimate  $\theta$  based on realizing the random matrix  $A$ . Some of these works also impose constraints on coefficients [25]. Our results on lack of unbiased estimators extends to these constrained settings as well.

#### G. Extension to the case of periodic unbiasedness

The papers [26], [27], [28], [29] consider mean square periodic error criterion combined with periodic unbiasedness for which the conventional CR bound does not provide a valid bound. Lehmann-unbiasedness concept<sup>2</sup> [30] is used to introduce the concept of periodic unbiasedness<sup>3</sup>, and a CR

<sup>2</sup>An estimator  $g(X)$  is said to be Lehmann unbiased w.r.t. a cost function  $W : \Theta \times \Theta \rightarrow \mathbb{R}^+$ , if  $\mathbb{E}_\theta(W(\theta', g(X))) \geq \mathbb{E}_\theta(W(\theta, g(X))), \forall \theta', \theta \in \Theta$ . When the cost function  $W$  is equal to the MSE, the Lehmann unbiasedness degenerates to standard unbiasedness, i.e.,  $\mathbb{E}_\theta(g(X)) = \theta, \forall \theta \in \Theta$ .

<sup>3</sup>Periodic unbiasedness [26] is Lehmann unbiasedness with the cost function  $\mathbb{E}(\text{mod}_{2\pi}(g(X) - \theta)^2)$ , where the  $\text{mod}_{2\pi}$  maps the squared error to  $[-\pi, \pi]$ .

type bound on the mean square periodic error of any periodic unbiased estimator is derived. It is easy to realize that also in this case, if the set of parameters is constrained to a strictly smaller subset of  $[-\pi, \pi]$  which has an extreme point, there still is no Lehmann unbiased estimator w.r.t. the mean square periodic error.

#### H. Conclusion

In this paper, we showed that under very general conditions, biasedness of the estimator is inevitable in constrained estimation problems. Sufficient conditions for the non-existence of an unbiased estimator whose range is the constrained set of parameters, is that the latter would have an extreme point, and that the distribution given the extreme value of  $\theta$  would be absolutely continuous w.r.t. another hypothesized distribution corresponding to another parameter in the constrained set. We extend the result to a case of open sets whose closure have extreme point. We also state more easily verifiable conditions which require that the constrained set be compact and the set of hypothesized distributions are continuous w.r.t. one another.

As mentioned in the introduction, in [10] the case in which there exists an unbiased estimator is considered for equality constraints. If the equality constraints on a non linear function  $f$  define a set of solutions which is bounded, then there is no unbiased estimator. It should be emphasized, however, that unbiased estimators can exist in the case in which the constrained parameter set is a manifold without boundaries, e.g., a intersection of hyperplanes in which there are no extreme points.

Our results have strong impact on the applicability of the results of [8], [10], [11].

#### REFERENCES

- [1] E. L. Lehmann and G. Casella, *Theory of point estimation*, 2nd ed. Springer-Verlag New-York, 1998.
- [2] H. L. Van Trees, *Detection, estimation, and modulation theory*. John Wiley & Sons, 2004.
- [3] L. L. Scharf, *Statistical signal processing*. Addison-Wesley Reading, MA, 1991, vol. 98.
- [4] E. Barankin, "Locally best unbiased estimates," *The Annals of Mathematical Statistics*, pp. 477–501, 1949.
- [5] K. L. Bell, Y. Steinberg, Y. Ephraim, and H. L. Van Trees, "Extended Ziv-Zakai lower bound for vector parameter estimation," *IEEE Transactions on Information Theory*, vol. 43, no. 2, pp. 624–637, 1997.
- [6] J. Ziv and M. Zakai, "Some lower bounds on signal parameter estimation," *IEEE transactions on Information Theory*, vol. 15, no. 3, pp. 386–391, 1969.
- [7] E. Weinstein and A. J. Weiss, "A general class of lower bounds in parameter estimation," *IEEE Transactions on Information Theory*, vol. 34, no. 2, pp. 338–342, 1988.
- [8] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Transactions on Information Theory*, vol. 36, no. 6, pp. 1285–1301, 1990.
- [9] D. G. Chapman and H. Robbins, "Minimum variance estimation without regularity assumptions," *The Annals of Mathematical Statistics*, pp. 581–586, 1951.
- [10] T. L. Marzetta, "A simple derivation of the constrained multiple parameter Cramer-Rao bound," *IEEE Transactions on Signal Processing*, vol. 41, no. 6, pp. 2247–2249, 1993.
- [11] P. Stoica and B. C. Ng, "On the Cramer-Rao bound under parametric constraints," *IEEE Signal Processing Letters*, vol. 5, no. 7, pp. 177–179, July 1998.
- [12] Z. Ben-Haim and Y. C. Eldar, "The Cramer-Rao bound for estimating a sparse parameter vector," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3384–3389, June 2010.

- [13] H. Wang, B. Pei, and J. Li, "Robust waveform design for MIMO radar with imperfect prior knowledge," *Circuits, Systems, and Signal Processing*, vol. 35, no. 4, pp. 1239–1255, 2016.
- [14] S. T. Smith, L. Scharf, and L. T. McWhorter, "Intrinsic quadratic performance bounds on manifolds," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, vol. 5, May 2006, pp. V–V.
- [15] A. O. Hero, J. A. Fessler, and M. Usman, "Exploring estimator bias-variance tradeoffs using the uniform cr bound," *IEEE Transactions on Signal Processing*, vol. 44, no. 8, pp. 2026–2041, Aug 1996.
- [16] J. P. Romano and A. F. Siegel, *Counter examples in probability and statistics*. Chapman and Hall/CRC, 1986.
- [17] T. van Erven and P. Harremoës, "Rényi divergence and Kullback-Leibler divergence," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 3797–3820, 2014.
- [18] D. C. Aliprantis and K. C. Border, *Infinite dimensional analysis: Ahichhiker's guide*, 3rd ed. Springer, 2006.
- [19] M. Krein and D. Milman, "On extreme points of regular convex sets," *Studia Mathematica*, vol. 9, no. 1, pp. 133–138, 1940. [Online]. Available: <http://eudml.org/doc/219061>
- [20] I. Csiszár and F. Matúš, "Closures of exponential families," *Annals of probability*, pp. 582–600, 2005.
- [21] Y. V. Prokhorov, "Convergence of random processes and limit theorems in probability theory," *Theory of Probability & Its Applications*, vol. 1, no. 2, pp. 157–214, 1956. [Online]. Available: <http://dx.doi.org/10.1137/1101016>
- [22] I. Csiszár and F. Matúš, "Information projections revisited," *IEEE Transactions on Information Theory*, vol. 49, no. 6, pp. 1474–1490, June 2003.
- [23] S. Aeron, V. Saligrama, and M. Zhao, "Information theoretic bounds for compressed sensing," *IEEE Transactions on Information Theory*, vol. 56, no. 10, pp. 5111–5130, 2010.
- [24] C. Aksoylar, G. K. Atia, and V. Saligrama, "Sparse signal processing with linear and nonlinear observations: A unified Shannon-theoretic approach," *IEEE Transactions on Information Theory*, vol. 63, no. 2, pp. 749–776, 2017.
- [25] M. H. Rohban, V. Saligrama, and D. M. Vaziri, "Minimax optimal sparse signal recovery with Poisson statistics," *IEEE Transactions on Signal Processing*, vol. 64, no. 13, pp. 3495–3508, 2016.
- [26] T. Rountenberg and J. Tabrikian, "Periodic CRB for non-Bayesian parameter estimation," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 2448–2451.
- [27] —, "Non-Bayesian periodic Cramér-Rao bound," *IEEE Transactions on Signal Processing*, vol. 61, no. 4, pp. 1019–1032, 2013.
- [28] —, "Performance bounds for constrained parameter estimation," in *Sensor Array and Multichannel Signal Processing Workshop (SAM), 2012 IEEE 7th*. IEEE, 2012, pp. 513–516.
- [29] T. Rountenberg and L. Tong, "The Cramér-Rao bound for estimation-after-selection," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 414–418.
- [30] E. L. Lehmann, "A general concept of unbiasedness," *Ann. Math. Statist.*, vol. 22, no. 4, pp. 587–592, 12 1951. [Online]. Available: <http://dx.doi.org/10.1214/aoms/1177729549>

**Anelia Somekh-Baruch**, (S'01–M'03) received the B.Sc. degree from Tel-Aviv University, Tel-Aviv, Israel, in 1996 and the M.Sc. and Ph.D. degrees from the Technion-Israel Institute of Technology, Haifa, Israel, in 1999 and 2003, respectively, all in electrical engineering. During 2003–2004, she was with the Technion Electrical Engineering Department. During 2005–2008, she was a Visiting Research Associate at the Electrical Engineering Department, Princeton University, Princeton, NJ. From 2008 to 2009 she was a researcher at the Electrical Engineering Department, Technion, and from 2009 she has been with the Bar-Ilan University School of Engineering, Ramat-Gan, Israel. Her research interests include topics in information theory and communication theory. Dr. Somekh-Baruch received the Tel-Aviv University program for outstanding B.Sc. students scholarship, the Viterbi scholarship, the Rothschild Foundation scholarship for postdoctoral studies, and the Marie Curie Outgoing International Fellowship.

**Amir Leshem** (leshema@biu.ac.il) received the B.Sc. (cum laude) in mathematics and physics, the M.Sc. (cum laude) in mathematics, and the Ph.D. degree in mathematics all from the Hebrew University, Jerusalem, Israel, in 1986, 1990, and 1998, respectively. From 1998 to 2000, he was with the Faculty of Information Technology and Systems, Delft University of Technology, The Netherlands, as a postdoctoral fellow working on algorithms for the reduction of terrestrial electromagnetic interference in radio-astronomical radio-telescope antenna arrays and signal processing for communication. From 2000 to 2003, he was Director of Advanced Technologies with Metalink Broadband where he was responsible for research and development of new DSL and wireless MIMO modem technologies and served as a member of several international standard setting groups such as ITU-T SG15, ETSI TM06, NIPP-NAI, IEEE 802.3 and 802.11. From 2000 to 2002, he was also a visiting researcher at Delft University of Technology. In 2002 he joined Bar-Ilan University where he is one of the founders of the faculty of engineering and a full professor and head of the communications research track. From 2003 to 2005, he was the technical manager of the U-BROAD consortium developing technologies to provide 100 Mbps and beyond over copper lines. In 2009 he spent his sabbatical at Delft University of Technology and Stanford University. His main research interests include multichannel wireless and wireline communication, applications of game theory to dynamic and adaptive spectrum management of communication networks, array and statistical signal processing with applications to multiple element sensor arrays and networks, wireless communications, radio-astronomical imaging, set theory, logic and foundations of mathematics. Prof. Leshem was an Associate Editor of the IEEE Transactions on Signal Processing from 2008 to 2011 and currently is AE for IEEE Traqns. on Signal and Information Processing over Networks. He was the Leading Guest Editor for special issues on signal processing for astronomy and cosmology in IEEE Signal Processing Magazine and the IEEE Journal of Selected Topics in Signal Processing.

**Venkatesh Saligrama** is a Professor at Boston University ECE and CS departments. He is also affiliated with the Institute for Data, Systems, and Society at MIT. His research interests include statistical machine learning, information and decision theory, and their applications in statistical and structured signal processing, resource constrained learning, anomaly detection and video analytics for surveillance. He has received a number of awards including the NSF CAREER Award, the PECASE award, ONR YIP, and a best paper at the NIPS14 Workshop on Analysis of Rank Data. He was a plenary speaker at the 2015 IEEE Machine Learning for Signal Processing conference. He has served as an Area Chair for NIPS, ICML, and AAAI conferences and has been an Associate Editor for IEEE Transactions on Information Theory and IEEE Transactions on Signal Processing.