Rethinking Biased Estimation: Improving Maximum Likelihood and the Cramér–Rao Bound

Rethinking Biased Estimation: Improving Maximum Likelihood and the Cramér–Rao Bound

Yonina C. Eldar

Department of Electrical Engineering
Technion — Israel Institute of Technology
Haifa 32000
Israel
yonina@ee.technion.ac.il

the essence of knowledge

Boston – Delft

Foundations and Trends® in Signal Processing

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is 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, 2007

ISBN: 978-1-60198-130-1

Foundations and Trends® in Signal Processing

Volume 1 Issue 4, 2007

Editorial Board

Editorial Scope

Foundations and Trends® in Signal Processing will publish survey and tutorial articles on the foundations, algorithms, methods, and applications of signal processing including the following topics:

- Adaptive signal processing
- Audio signal processing
- Biological and biomedical signal processing
- Complexity in signal processing
- Digital and multirate signal processing
- Distributed and network signal processing
- Image and video processing
- Linear and nonlinear filtering
- Multidimensional signal processing
- Multimodal signal processing
- Multiresolution signal processing
- Nonlinear signal processing
- Randomized algorithms in signal processing
- Sensor and multiple source signal processing, source separation
- Signal decompositions, subband and transform methods, sparse representations
- Signal processing for communications
- Signal processing for security and forensic analysis, biometric signal processing
- Signal quantization, sampling, analog-to-digital conversion, coding and compression
- Signal reconstruction, digital-to-analog conversion, enhancement, decoding and inverse problems
- Speech/audio/image/video compression
- Speech and spoken language processing
- Statistical/machine learning
- Statistical signal processing
- Classification and detection
- Estimation and regression
- Tree-structured methods

Information for Librarians

Foundations and Trends® in Signal Processing, 2007, Volume 1, 4 issues. ISSN paper version 1932-8346. ISSN online version 1932-8354. Also available as a combined paper and online subscription.

Foundations and Trends® in
Signal Processing
Vol. 1, No. 4 (2007) 305–449

DOI: 10.1561/2000000008

Rethinking Biased Estimation: Improving Maximum Likelihood and the Cramér–Rao Bound

Yonina C. Eldar[1]

[1] *Department of Electrical Engineering, Technion — Israel Institute of Technology, Haifa 32000, Israel, yonina@ee.technion.ac.il*

Abstract

One of the prime goals of statistical estimation theory is the development of performance bounds when estimating parameters of interest in a given model, as well as constructing estimators that achieve these limits. When the parameters to be estimated are deterministic, a popular approach is to bound the mean-squared error (MSE) achievable within the class of unbiased estimators. Although it is well-known that lower MSE can be obtained by allowing for a bias, in applications it is typically unclear how to choose an appropriate bias.

In this survey we introduce MSE bounds that are lower than the unbiased Cramér–Rao bound (CRB) for all values of the unknowns. We then present a general framework for constructing biased estimators with smaller MSE than the standard maximum-likelihood (ML) approach, regardless of the true unknown values. Specializing the results to the linear Gaussian model, we derive a class of estimators that dominate least-squares in terms of MSE. We also introduce methods for choosing regularization parameters in penalized ML estimators that outperform standard techniques such as cross validation.

Contents

1

Introduction

The problem of estimating a set of unknown deterministic parameters is ubiquitous in a vast variety of areas in science and engineering including, for example, communication, economics, signal processing, seismology, and control. Many engineering systems rely on estimation theory to extract required information by estimating values of unknown parameters. Statisticians use parameter estimation techniques to extract and infer scientific, medical, and social conclusions from numerical data which are subject to random uncertainties.

Parameter estimation has a rich history dating back to Gauss and Legendre who used the least-squares (LS) method to predict movements of planets [62, 63, 97]. Mathematically, in an estimation problem, we are given a set of observations $\mathbf{x}$ which we assume depend on an unknown parameter vector $\boldsymbol{\theta}_0$. In this survey, we treat the setting in which $\boldsymbol{\theta}_0$ is an unknown deterministic vector, i.e., the classical estimation setting as opposed to Bayesian inference. The problem then is to infer $\boldsymbol{\theta}_0$ from the data using an estimate $\hat{\boldsymbol{\theta}}$ which is a function of $\mathbf{x}$, and to gain insight into the theoretical effects of the parameters on the system output.

One of the prime goals of statistical estimation theory is the development of bounds on the best achievable performance in inferring parameters of interest in a given model, as well as determining estimators that achieve these limits. Such bounds provide benchmarks against which we can compare the performance of any proposed estimator, and insight into the fundamental limitations of the problem.

A classic performance benchmark is the Cramér–Rao bound (CRB) [27, 28, 30, 60, 119, 120], which characterizes the smallest achievable total variance of any *unbiased* estimator of $\boldsymbol{\theta}_0$. Although other variance bounds exist in the literature, the CRB is relatively easy to determine, and can often be achieved by the maximum likelihood (ML) method [100, 120]. Despite its popularity, the CRB limits only the variance of unbiased estimators. However, in some problems, restricting attention to unbiased approaches leads to unreasonable solutions, that may, for example, be independent of the problem parameters [71, 98]. More importantly, in many cases the variance can be made smaller at the expense of increasing the bias, while ensuring that the overall estimation error is reduced. Therefore, even though unbiasedness may be appealing intuitively, it does not necessarily lead to a small estimation error $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$ [34]. Consequently, the design of estimators is typically subject to a tradeoff between variance and bias [50, 58, 81, 104, 107, 136].

In this survey, we discuss methods to improve the accuracy of unbiased estimators used in many signal processing problems. At the heart of the proposed methodology is the use of the mean-squared error (MSE) as the performance criteria. The MSE is the average of the squared-norm error $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2$, and is equal to the sum of the variance and the squared-norm of the bias. In an estimation context, where our prime concern is inferring $\boldsymbol{\theta}_0$, the MSE (or weighted MSE) provides a direct measure of the relevant performance. Although herein we focus on the MSE, the essential ideas can be easily generalized to include weighted MSE criteria which measure the average weighted squared-norm error [48].

The approach we present is based on introducing a bias as a means of reducing the MSE. Biased estimation strategies are used extensively in a variety of different signal processing applications, such as image restoration [31, 108] where the bias corresponds to spatial resolution, smoothing techniques in time series analysis [115, 137], spectrum estimation [131], wavelet denoising [33], and diagonal loading in beamforming applications [21, 26, 56]. Despite the fact that biasing as a method for improving performance is a mainstream approach, very often the choice of bias is rather *ad-hoc*. In particular, although the biased algorithms mentioned above will improve the performance for certain choices of $\boldsymbol{\theta}_0$, they can in fact deteriorate the MSE for other parameter values. Thus, in general, conventional biasing methods are not guaranteed to dominate ML, i.e., do not necessarily have lower MSE for all choices of $\boldsymbol{\theta}_0$. Furthermore, many of these techniques include regularization parameters which are typically chosen by optimizing a data-error measure, i.e., an objective that depends on the estimated data $\hat{\mathbf{x}}$ obtained by replacing $\boldsymbol{\theta}_0$ by $\hat{\boldsymbol{\theta}}$ in the model equations. Here, we focus on biasing in a way that is guaranteed to improve the MSE for all parameter values. This is achieved by using objectives that are directly related to the estimation error and are not data-error driven.

In their seminal work, Stein and James showed that for the independent, identically-distributed (iid) linear Gaussian model, it is possible to construct a nonlinear estimate of $\boldsymbol{\theta}_0$ with lower MSE than that of ML for all values of the unknowns [88, 128]. Such a strategy is said to dominate ML. In general an estimator $\hat{\boldsymbol{\theta}}_1$ dominates a different estimator $\hat{\boldsymbol{\theta}}_2$ if its MSE is no larger than that of $\hat{\boldsymbol{\theta}}_2$ for all feasible $\boldsymbol{\theta}_0$, and is strictly smaller for at least one choice of $\boldsymbol{\theta}_0$; an estimator is admissible if it is not dominated by any other approach. Stein's landmark idea has since been extended in many different directions and has inspired the work on ML-dominating methods which is the focus of this survey.

Here we go beyond the iid Gaussian model, and address a broad variety of estimation problems within an unified, systematic framework. To characterize the best possible bias-variance tradeoff in a general setting we would like to obtain a bound on the smallest achievable MSE in a given estimation problem. However, since $\boldsymbol{\theta}_0$ is deterministic, the MSE will in general depend on $\boldsymbol{\theta}_0$ itself. Therefore, the MSE cannot be used as a design criterion for choosing an optimal bias. Indeed, the point-wise minimum of the MSE is given by the trivial zero bound, which can be achieved with $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$.

To overcome this obstacle, instead of attempting to minimize the MSE over all possible estimators, which includes the trivial solution

$\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$, we restrict attention to methods that lie in a suitable class; the CRB is an example where we consider only methods with zero bias. Allowing for a broader set of bias vectors will result in MSE bounds that are lower than the CRB for all values of $\boldsymbol{\theta}_0$. Furthermore, as part of the proposed framework we introduce explicit methods that achieve these lower bounds resulting in estimators with performance superior to unbiased approaches. In cases where the ML is efficient, namely it achieves the CRB, this methodology guarantees the existence of estimators that have lower MSE than ML for all values of $\boldsymbol{\theta}_0$.

The strategy we outlined is based on first developing MSE performance bounds, and then designing estimators that achieve these limits, thus ensuring MSE improvement over existing unbiased solutions. An alternative technique to improve traditional estimates which is prevalent in the literature is the use of regularization, first systematically studied by Tikhonov [135, 136] and later extended to general estimation problems via the penalized ML (PML) approach [65, 66]. In general, regularization methods measure both the fit to the observed data and the physical plausibility of the estimate. Traditional applications of PML and regularization techniques have relied on data-error measures for selecting the regularization parameters [17, 61, 64, 72, 73, 89, 110].

As part of the proposed framework in this survey, we introduce methods for choosing the required regularization parameters based on measures of estimation error rather than data error. A popular design strategy in this spirit is to minimize Stein's unbiased risk estimate (SURE) [32, 122, 129, 130], which is an unbiased estimate of the MSE. This method is appealing as it allows to directly approximate the MSE of an estimate from the data, without requiring knowledge of $\boldsymbol{\theta}_0$. Besides leading to significant performance improvement over standard data-driven approaches in many practical problems, this technique can often be shown to dominate ML. In fact, the celebrated James–Stein estimate [88, 128], although originally derived based on different considerations, can be obtained from the SURE principle, as can many other ML-dominating approaches.

In most of the survey, we focus on problems in which the relationship between the data $\mathbf{x}$ and the unknown parameters $\boldsymbol{\theta}_0$ is given by a statistical model. In the last section, we depart from this framework and discuss methods for bounded error estimation in which the statistical model is replaced by the assumption that $\boldsymbol{\theta}_0$ is restricted to some deterministic set, defined by prior constraints. The link to the rest of the survey is that in this context as well, we can replace traditional data-error strategies by methods that are inherently based on the error between the estimate $\hat{\boldsymbol{\theta}}$ and the true parameter $\boldsymbol{\theta}_0$. Although this approach is deterministic in nature, it can also be used in a statistical setting where the constraints are dictated by the underlying statistical properties. For example, given measurements $\mathbf{x} = \boldsymbol{\theta}_0 + \mathbf{w}$, where $\mathbf{w} \in \mathbb{R}^n$ is a zero-mean random vector with covariance $\sigma^2\mathbf{I}$, we can assume that $\boldsymbol{\theta}_0$ lies in the constraint set $\|\mathbf{x} - \boldsymbol{\theta}_0\|^2 \leq n\sigma^2$. Despite the fact that this restriction is not always satisfied, using it in conjunction with the proposed estimation strategy leads to an estimate that dominates the constrained ML solution. Therefore, this approach can also be used to develop MSE-dominating techniques when a statistical model exists.

Our focus here is on static models. In recent years, there has been increasing interest in inference techniques and performance bounds for dynamical systems [134]. We believe that the essential ideas introduced can be extended to the dynamical setting as well.

1.1 Estimation Model

Throughout the survey, our goal is to estimate a *deterministic* parameter vector $\boldsymbol{\theta}_0$ from measurements $\mathbf{x}$. For concreteness, we assume that $\boldsymbol{\theta}_0$ is a real length-m vector, and $\mathbf{x}$ is a real length-n vector. However, all the results are valid for the complex case as well with obvious modifications. The relationship between $\mathbf{x}$ and $\boldsymbol{\theta}_0$ is described by the probability density function (pdf) $p(\mathbf{x};\boldsymbol{\theta}_0)$ of $\mathbf{x}$ characterized by $\boldsymbol{\theta}_0$. We emphasize that $\boldsymbol{\theta}_0$ is a deterministic unknown vector, so that no Bayesian prior is assumed on $\boldsymbol{\theta}_0$. Consequently, $p(\mathbf{x};\boldsymbol{\theta}_0)$ is not a joint pdf, but rather a pdf of $\mathbf{x}$ in which $\boldsymbol{\theta}_0$ figures as an unknown parameter. As we will see throughout the survey, this renders the problem considerably more challenging, but at the same time more intriguing than its Bayesian counterpart.

As an example, suppose we have a Bernoulli random variable x_i which takes on the value 1 with probability (w.p.) θ_0 and 0 w.p. $1-\theta_0$. Our goal is to estimate θ_0 from n iid measurements. Denoting by $\mathbf{x} = (x_1,\ldots,x_n)^T$ the vector whose components are the measurements x_i, the pdf of $\mathbf{x}$ can be written as

$$p(\mathbf{x};\boldsymbol{\theta}_0) = \theta_0^{\sum_{i=1}^n x_i}(1-\theta_0)^{n-\sum_{i=1}^n x_i}. \tag{1.1}$$

Another important class of examples, which we will study in detail in Section 4, is the linear Gaussian model. In this case the unknown vector $\boldsymbol{\theta}_0 \in \mathbb{R}^m$ is related to $\mathbf{x} \in \mathbb{R}^n$ through the linear model:

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta}_0 + \mathbf{w}. \tag{1.2}$$

Here $\mathbf{H}$ is a known $n \times m$ model matrix with full column-rank, and $\mathbf{w}$ is a zero-mean Gaussian random vector with covariance matrix $\mathbf{C}$, which for simplicity is assumed to be positive definite. For the model (1.2), the pdf of $\mathbf{x}$ is

$$p(\mathbf{x};\boldsymbol{\theta}_0) = \frac{1}{\sqrt{(2\pi)^n|\mathbf{C}|}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{H}\boldsymbol{\theta}_0)^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{H}\boldsymbol{\theta}_0)\right\}. \tag{1.3}$$

Although we assume that $\mathbf{H}$ is known in the model (1.2), similar ideas to those developed here can be used when $\mathbf{H}$ is subject to deterministic or random uncertainty [8, 44, 51, 56, 144, 145].

A broader class of pdfs which includes (1.3) is the exponential family of distributions which can expressed in the form:

$$f(\mathbf{x};\boldsymbol{\theta}_0) = r(\mathbf{x})\exp\{\boldsymbol{\theta}_0^T\phi(\mathbf{x}) - g(\boldsymbol{\theta}_0)\}, \tag{1.4}$$

where $r(\mathbf{x})$ and $\phi(\mathbf{x})$ are functions of the data only, and $g(\boldsymbol{\theta}_0)$ depends on the unknown parameter $\boldsymbol{\theta}_0$. Exponential pdfs play an important role in statistics due to the Pitman–Koopman–Darmois theorem [29, 94, 117], which states that among distributions whose domain does not vary with the parameter being estimated, a sufficient statistic with bounded dimension as the sample size increases can be found only in exponential families [100]. Furthermore, efficient estimators achieving the CRB exist only when the underlying model is exponential. Many known distributions are of the exponential form, such as Gaussian, gamma, chi-square, beta, Dirichlet, Bernoulli, binomial, multinomial, Poisson, and geometric distributions. Exponential families will play an important role in Section 5 in the context of estimation based on the SURE criterion.

1.2 Minimum Variance Unbiased Estimation

Given data $\mathbf{x}$ and a model $p(\mathbf{x};\boldsymbol{\theta}_0)$ a pervasive inference strategy in signal processing applications is to seek a minimum variance unbiased (MVU) estimate of $\boldsymbol{\theta}_0$. This is typically accomplished by using the theory of sufficient statistics or the attainment of the CRB [93]. Although an MVU solution is not guaranteed to exist, in many problems of interest such an estimate can be found, at least asymptotically. The constraint of unbiasedness is often a practical one, since in many cases the variance, or the MSE, can be minimized over this class using functions of the data that are truly estimators, i.e., the statistic does not depend on the unknown parameter. However, there are several severe limitations of unbiased methods.

First, unbiased estimators are not always guaranteed to exist. An example is when inferring the odds ratio $p = \theta_0/(1-\theta_0)$ from n Bernoulli trials. It can be shown that there is no unbiased estimate for p [124, Sec. 7.12]. On the other hand, there exist many reasonable approximations such as $p = \hat{\theta}/(1-\hat{\theta})$, where $\hat{\theta} = (1/n)\sum_{i=1}^{n} x_i$.

Second, the unbiasedness requirement can sometimes produce nonsensical results. As an example, consider the problem of estimating the probability of success θ_0 in a set of Bernoulli trials, from the number of experiments x until success [25]. The pdf of x is given by

$$p(x;\theta_0) = \theta_0(1-\theta_0)^{x-1}, \quad x = 1,2,\ldots. \tag{1.5}$$

The only unbiased estimate for this problem, and hence the MVU solution, is

$$\hat{\theta}_0 = \begin{cases} 1, & x = 1; \\ 0, & \text{otherwise.} \end{cases} \tag{1.6}$$

Clearly this is an unreasonable estimate of θ_0. A more appealing choice is $\hat{\theta} = 1/x$.

As another example, suppose that x is a Poisson random variable with mean $\theta_0 > 0$, and we would like to estimate $p = \exp\{-2\theta_0\}$, which is the probability that no events occur in two units of time. Clearly the true value of p satisfies $p \in (0,1)$. However, the only unbiased estimate is given by

$$\hat{p} = \begin{cases} 1, & x \text{ even}; \\ -1, & x \text{ odd}, \end{cases} \tag{1.7}$$

which always falls outside the range $(0,1)$, and is extremely unreasonable [99] [132, Exercise 17.26] [124, Sec. 7.16]. A somewhat more complex example, in which the only unbiased estimator always ends up considerably outside the problem bounds, can be found in [77].

Finally, the most important objection to the constraint of unbiasedness is that it produces estimators $\hat{\boldsymbol{\theta}}$ whose optimality is based on the error between $\hat{\boldsymbol{\theta}}$ and the average value, not $\hat{\boldsymbol{\theta}}$ and the true value as measured by the MSE. It is the latter that is actually of prime importance in an estimation context as it is a direct measure of estimation error. Specifically, the MSE is defined by

$$E\left\{\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2\right\} = \int \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2 f(\mathbf{x};\boldsymbol{\theta}_0) d\mathbf{x} = \|\mathbf{b}(\boldsymbol{\theta}_0)\|^2 + v(\boldsymbol{\theta}_0), \tag{1.8}$$

where $\mathbf{b}(\boldsymbol{\theta}_0) = E\{\hat{\boldsymbol{\theta}}\} - \boldsymbol{\theta}_0$ is the bias of the estimate, and $v(\boldsymbol{\theta}_0) = E\{\|\hat{\boldsymbol{\theta}} - E\{\hat{\boldsymbol{\theta}}\}\|^2\}$ is its variance. Note that the MSE depends explicitly on $\boldsymbol{\theta}_0$. An MVU method minimizes the MSE only over a constrained class for which $\mathbf{b}(\boldsymbol{\theta}_0) = \mathbf{0}$ for all $\boldsymbol{\theta}_0$. Thus, even in problems in which the MVU approach leads to reasonable estimates, the MSE performance may still be improved using a biased technique.

The difficulty in using the MSE as a design objective is that in general it depends explicitly on $\boldsymbol{\theta}_0$. This parameter dependency also renders comparison between different estimators a difficult (and often impossible) task. Indeed, one method may be better than another for some values of $\boldsymbol{\theta}_0$, and worse for others. For instance, the trivial estimator $\hat{\boldsymbol{\theta}} = \mathbf{0}$ achieves optimal MSE when $\boldsymbol{\theta}_0 = \mathbf{0}$, but its performance is otherwise poor. Nonetheless, it is possible to impose a partial order among inference techniques [100] using the concepts of domination and admissibility. An estimator $\hat{\boldsymbol{\theta}}_1$ dominates an estimator $\hat{\boldsymbol{\theta}}_2$ on a given set

$\mathcal{U}$ if

$$\begin{aligned} E\{\|\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}\|^2\} &\leq E\{\|\hat{\boldsymbol{\theta}}_2-\boldsymbol{\theta}\|^2\}, \quad \text{for all } \boldsymbol{\theta}\in\mathcal{U}; \\ E\{\|\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}\|^2\} &< E\{\|\hat{\boldsymbol{\theta}}_2-\boldsymbol{\theta}\|^2\}, \quad \text{for some } \boldsymbol{\theta}\in\mathcal{U}. \end{aligned} \tag{1.9}$$

The estimator $\hat{\boldsymbol{\theta}}_1$ strictly dominates $\hat{\boldsymbol{\theta}}_2$ on $\mathcal{U}$ if

$$E\{\|\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}\|^2\} < E\{\|\hat{\boldsymbol{\theta}}_2-\boldsymbol{\theta}\|^2\}, \quad \text{for all } \boldsymbol{\theta}\in\mathcal{U}. \tag{1.10}$$

If $\hat{\boldsymbol{\theta}}_1$ dominates $\hat{\boldsymbol{\theta}}_2$ then clearly it is better in terms of MSE. An estimator $\hat{\boldsymbol{\theta}}$ is admissible if it is not dominated by any other method. If an estimator is inadmissible, then there exists another approach whose MSE is no larger than the given method for all $\boldsymbol{\theta}$ in $\mathcal{U}$, and is strictly smaller for some $\boldsymbol{\theta}$ in $\mathcal{U}$.

The study of admissibility is sometimes restricted to linear methods. A linear admissible estimator is one which is not dominated by any other linear strategy. The class of linear admissible techniques can be characterized by a simple rule [24, 43, 83, 121], and given any linear inadmissible estimator, it is possible to construct a linear admissible alternative which dominates it by using convex analysis tools [43]. However, the problem of admissibility is considerably more intricate when the linearity restriction is removed; generally, admissible estimators are either trivial (e.g., $\hat{\boldsymbol{\theta}}=\mathbf{0}$) or exceedingly complex [105]. As a result, much research has focused on finding simple nonlinear techniques that dominate ML.

1.3 Maximum Likelihood Estimation

One of the most popular estimation strategies is the ML method in which the estimate $\hat{\boldsymbol{\theta}}$ is chosen to maximize the likelihood of the observations:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} p(\mathbf{x};\boldsymbol{\theta}). \tag{1.11}$$

This approach was pioneered by Fisher between 1912 and 1922 [1, 59] and has widespread applications in various fields. The ML estimator enjoys several appealing properties, including asymptotic efficiency under suitable regularity conditions. Thus, asymptotically, and in many non-asymptotic cases, the ML approach is MVU optimal. Nonetheless, its MSE can be improved upon in the non-asymptotic regime in many different settings.

As is evident from (1.11) the ML technique is data driven, meaning the quality of the estimator is determined by how well it describes the observations. However, the ML objective is not related to the MSE which is a direct measure of estimation error. This distinction is clearly seen when considering the linear Gaussian model (1.2). In this case the ML criterion coincides with the weighted LS objective:

$$\arg\max_{\boldsymbol{\theta}} p(\mathbf{x};\boldsymbol{\theta}) = \arg\min_{\boldsymbol{\theta}} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta}). \tag{1.12}$$

Evidently, the ML solution is designed to minimize the error between the given data and the estimated data $\hat{\mathbf{x}} = \mathbf{H}\hat{\boldsymbol{\theta}}$. Assuming $\mathbf{H}$ has full column-rank, the resulting LS estimate is given by

$$\hat{\boldsymbol{\theta}}_{\mathrm{LS}} = (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}^{-1} \mathbf{x}. \tag{1.13}$$

It is well known that $\hat{\boldsymbol{\theta}}_{\mathrm{LS}}$ is also MVU optimal for Gaussian noise [93].

To illustrate the fact that minimizing data error does not necessarily imply a small estimation error, in Figure 1.1 we consider an example of the model (1.2) in which $\boldsymbol{\theta}_0$ represents the 2D signal in Figure 1.1(a). Our goal is to recover this image from the observation $\mathbf{x}$ of Figure 1.1(b) which is obtained after shifting and blurring with a Gaussian kernel, and corruption by additive Gaussian noise. We assume that the distortion and noise variance are known. Using the LS estimate results in the image in Figure 1.1(c) in which the original signal

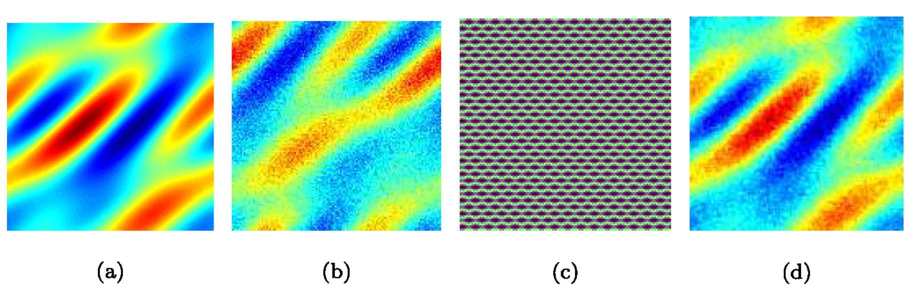

Fig. 1.1 Image recovery using least-squares (LS) and a biased minimax estimate. (a) original 2D signal. (b) Corrupted image. (c) Recovery using LS. (d) Recovery based on a minimax strategy.

is completely destroyed. On the other hand, using a minimax estimate, which we will discuss in Section 4, we obtain a pretty good recovery of the signal, as can be seen in Figure 1.1(d). Clearly the fact that the data error is smaller in Figure 1.1(c) is not sufficient to guarantee good signal recovery.

As another example, consider estimating a signal $\theta_0(t)$ that is observed through the heat integral equation and corrupted by additive noise. The true and observed signals are shown in Figures 1.2(a) and 1.2(b), respectively. In Figure 1.2(c) we compare the estimated signal using LS and a bounded-error approach (RCC) based on controlling the minimax estimation error, which we present in Section 6. Evidently, the latter strategy, referred to as the Chebyshev center, leads to substantial performance improvement.

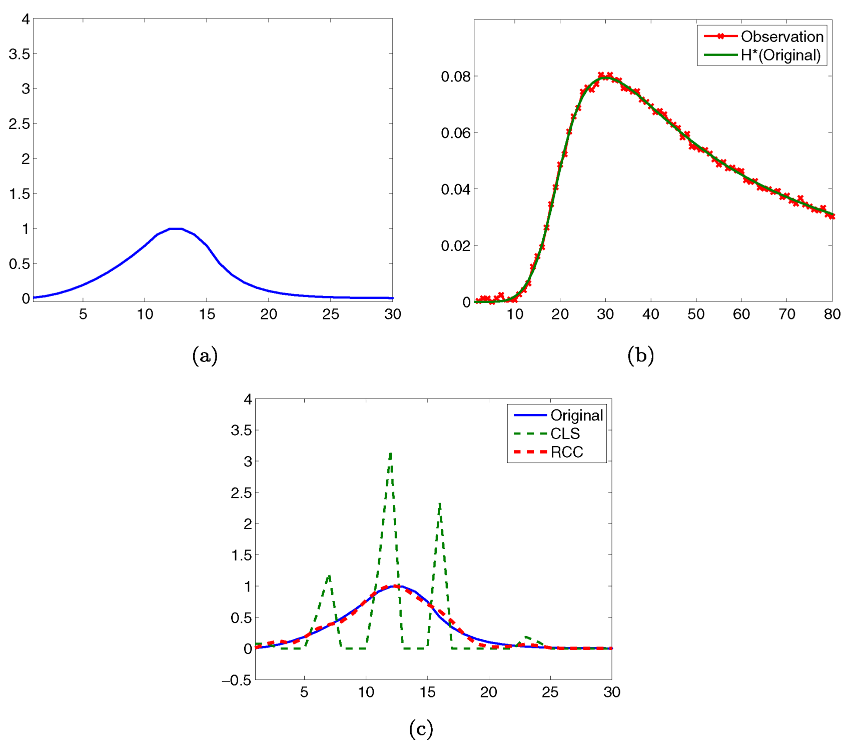

Fig. 1.2 Signal recovery using least-squares (CLS) and the Chebyshev estimate (RCC). (a) True signal. (b) Observed signal. (c) Recovery using CLS and RCC.

These examples illustrate that minimizing data error does not necessarily imply a small estimation error. From a statistical perspective, MVU methods do not guarantee satisfactory estimation performance, even when they exist and lead to reasonable strategies.

1.4 Outline and Goals

Stein's discovery of ML-dominating techniques in the linear Gaussian model, half a century ago, shocked the statistics community. Since then many other examples of ML improvement have been discovered and analyzed. In this survey, we present a broad framework for constructing ML-dominating solutions in a broad variety of estimation problems. More specifically, we present general tools for reducing MSE by introducing a bias. An important aspect of the proposed approach is that the reduction in MSE is guaranteed for all choices of the unknown parameter vector. The methods we outline for constructing estimators are designed to explicitly optimize an objective based on estimation error rather than data error. The performance advantage of the algorithms we present is greatest in difficult problems, i.e., short data records or lower signal-to-noise ratios (SNRs). Applications include the design of estimation algorithms for sonar, radar, and communications, as well as a myriad of other disciplines that rely heavily on precise measurement of parameters.

It is our hope that this framework will provide additional support for ML dominating methods, both by supplying an intuitive understanding of this phenomenon, and by providing a wide class of powerful new estimators.

1.4.1 Outline

In Section 2, we begin by reviewing the standard unbiased CRB and then discuss extensions to biased estimation. In particular, we introduce the uniform CRB which provides a benchmark on the variance of any biased estimator with bias-gradient matrix whose norm is limited by a constant. This bound is asymptotically achieved by the PML method with a suitable regularization function. The uniform CRB is useful in problems in which the bias gradient norm has a physical interpretation;

this is the case in some imaging applications where the norm is related to image resolution [80, 108]. Furthermore, it requires the specification of only one parameter (the norm bound) rather than the entire bias gradient matrix, as in the standard biased CRB [140].

In Section 3, we study MSE bounds which directly limit the estimation error. These bounds depend on the unknown parameter vector $\boldsymbol{\theta}_0$, as well as on the bias of the estimate $\hat{\boldsymbol{\theta}}$. In order to optimize the bound we first consider the class of estimates with linear bias vectors, and seek the member from this set that minimizes the bound. A nice aspect of this approach is that once an optimal bias of this form is found, it can be used to construct a linear modification of the ML estimate that dominates ML whenever the latter is efficient. We demonstrate this methodology through several examples which illustrate how scaling can be used to reduce the MSE. As we show, it is often possible to improve the MSE for all $\boldsymbol{\theta}_0$ using a linear modification, without any prior knowledge on the true parameter values. This linear scaling is chosen as a solution to a minimax optimization problem.

Building on the linear results, in Section 4, we present the blind minimax technique which leads to nonlinear modifications of the ML solution. The approach is illustrated in the context of the linear Gaussian model and makes use of a two-stage process: first, a set is estimated from the measurements; next, a linear minimax method for this set is used to estimate the parameter itself. Surprisingly, the resulting estimate can be shown to dominate the ML solution even though no prior information is assumed. The blind minimax technique provides a framework whereby many different estimators can be generated, and provides insight into the mechanism by which these techniques outperform ML. In particular, we show how the celebrated James–Stein estimate can be derived within this framework.

An alternative approach for deriving ML-dominating methods is to use the SURE principle. In Section 5, we introduce the SURE objective and illustrate how it can be applied to construct methods that have lower MSE than ML. The essential idea is to choose a class of estimates, and then select the member that minimizes the MSE estimate. We demonstrate, in particular, the use of the SURE design method for selecting regularization parameters in PML estimation.

Finally, in Section 6, we extend the estimation-error methodology to a deterministic setting. We treat estimation problems in which there are prior constraints on $\boldsymbol{\theta}_0$, such as weighted norm restrictions or interval constraints on the individual components of $\boldsymbol{\theta}_0$. The standard approach in such settings is constrained ML in which the likelihood is maximized subject to the given restrictions. Instead, we introduce the Chebyshev center estimator which is based on minimizing the worst-case estimation error $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2$ over all feasible solutions. As we show, this strategy can reduce the estimation error dramatically with respect to the constrained ML method. This design technique can also be used in a statistical setting by replacing the statistical model with an appropriate constraint on $\boldsymbol{\theta}_0$. Even though this later restriction is not always satisfied in practice, the resulting estimate can be shown in some cases to dominate the constrained ML for the same problem setting.

The procedures we develop throughout the survey are based on convex optimization tools and minimax formulations. In the Appendix, we provide a brief overview of the basics of convex analysis, emphasizing the results needed in our presentation.

References

[1] J. Aldrich, "R. A. Fisher and the making of maximum likelihood 1912–1922," *Statistical Science*, vol. 12, no. 3, pp. 162–176, 1997.

[2] D. Amir, "Chebychev centers and uniform convexity," *Pacific Journal of Mathematics*, vol. 77, no. 1, pp. 1–6, 1978.

[3] R. K. Bahr and J. A. Bucklew, "Minimax estimation of unknown deterministic signals in colored noise," *IEEE Transactions on Information Theory*, vol. 34, pp. 632–641, 1988.

[4] A. J. Baranchik, "Multiple regression and estimation of the mean of a multivariate normal distribution," Technical Report, 51, Stanford University, 1964.

[5] A. Beck, *Convexity Properties Associated with Nonconvex Quadratic Matrix Functions and Applications to Quadratic Programming,* preprint.

[6] A. Beck, A. Ben-Tal, and Y. C. Eldar, "Robust mean-squared error estimation of multiple signals in linear systems affected by model and noise uncertainties," *Mathematical Programming*, vol. 107, no. 1, pp. 155–187, 2006.

[7] A. Beck and Y. C. Eldar, "Strong duality in nonconvex quadratic optimization with two quadratic constraints," *SIAM Journal of Optimization*, vol. 17, no. 3, pp. 844–860, 2006.

[8] A. Beck and Y. C. Eldar, "Regularization in regression with bounded noise: A Chebyshev center approach," *SIAM Journal of Matrix Analysis Applications*, vol. 29, no. 2, pp. 606–625, 2007.

[9] A. Beck, Y. C. Eldar, and A. Ben-Tal, "Mean-squared error estimation for linear systems with block circulant uncertainty," *SIAM Journal of Matrix Analysis Applications*, vol. 29, pp. 712–730, 2007.

[10] Z. Ben-Haim and Y. C. Eldar, "Maximum set estimators with bounded estimation error," *IEEE Transactions on Signal Processing*, vol. 53, pp. 3172–3182, August 2005.
[11] Z. Ben-Haim and Y. C. Eldar, "Blind minimax estimation," *IEEE Transactions on Information Theory*, vol. 53, pp. 3145–3157, September 2007.
[12] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization*. MPS-SIAM Series on Optimization, 2001.
[13] A. Benazza-Benyahia and J.-C. Pesquet, "Building robust wavelet estimators for multicomponent images using Stein's principle," *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1814–1830, November 2005.
[14] J. Berger, "Improving on inadmissible estimators in continuous exponential families with applications to simultaneous estimation of gamma scale parameters," *Annals of Statistics*, vol. 8, no. 3, pp. 545–571, 1980.
[15] J. O. Berger, "Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss," *Annals of Statistics*, vol. 4, no. 1, pp. 223–226, January 1976.
[16] D. P. Bertsekas, *Nonlinear Programming*. Belmont MA: Athena Scientific, Second Edition, 1999.
[17] A. Björck, *Numerical Methods for Least-Squares Problems*. Philadelphia, PA: SIAM, 1996.
[18] M. E. Bock, "Minimax estimators of the mean of a multivariate normal distribution," *Annals of Statistics*, vol. 3, no. 1, pp. 209–218, January 1975.
[19] S. Boyd, L. E. Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*. Philadelphia, PA: SIAM, 1994.
[20] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
[21] B. D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 24, pp. 397–401, July 1988.
[22] A. Charras and C. van Eeden, "Bayes and admissibility properties of estimators in truncated parameter spaces," *The Canadian Journal of Statistics*, vol. 19, no. 2, pp. 121–114, June 1991.
[23] J. Chernoi and Y. C. Eldar, "Improving maximum likelihood for constrained denoising: An extended Chebyshev center approach," submitted to *IEEE Transactions on Signal Processing*.
[24] A. Cohen, "All admissible linear estimates of the normal mean," *Annals of Mathematical Statistics*, vol. 37, pp. 458–463, 1966.
[25] D. R. Cox and D. V. Hinkley, *Theoretical Statistics*. Chapman and Hall, 1974.
[26] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," *IEEE Transactions on Acoustics, Speech, Signal Processing*, vol. ASSP-35, pp. 1365–1376, October 1987.
[27] H. Cramér, "A contribution to the theory of statistical estimation," *Skandinavisk Aktuaries Tidskrif*, vol. 29, pp. 458–463, 1946.
[28] H. Cramér, *Mathematical Methods of Statistics*. Princeton, NJ: Princeton Univ. Press, 1946.

[29] G. Darmois, "Sur les lois de probabilites a estimation exhaustive," *C.R. Academy of Science Paris*, vol. 200, pp. 1265–1266, 1935.

[30] G. Darmois, "Sur les lois limites de la dispersion de certaines estimations," *Rev. Inst. Znt. Stat.*, vol. 13, pp. 9–15, 1945.

[31] G. Demoment, "Image reconstruction and restoration: Overview of common estimation structures and problems," *IEEE Transactions on Acoustics, Speech, Signal Processing*, vol. 37, pp. 2024–2036, 1989.

[32] L. Desbat and D. Girard, "The "minimum reconstruction error" choice of regularization parameters: Some effective methods and their application to deconvolution problems," *SIAM Journal of Science Computation*, vol. 16, no. 6, pp. 1387–1403, November 1995.

[33] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of American Statistical Association*, vol. 90, no. 432, pp. 1200–1224, December 1995.

[34] B. Efron, "Biased versus unbiased estimation," *Advances Mathematics*, vol. 16, pp. 259–277, 1975.

[35] B. Efron and C. Morris, "Combining possibly related estimation problems," *Journal of Royal Statistical Society B*, vol. 35, no. 3, pp. 379–421, 1973.

[36] B. Efron and C. Morris, "Steins estimation rule and its competitors: An empirical Bayes approach," *Journal of American Statistical Association*, vol. 68, pp. 117–130, 1973.

[37] B. Efron and C. Morris, "Stein's paradox in statistics," *Scientific American*, vol. 236, pp. 119–127, 1977.

[38] Y. C. Eldar, "The Chebyshev estimate: Statistical aspects," to appear in *Journal of Statistical Planning and Inference*.

[39] Y. C. Eldar, "Generalized SURE for exponential families: Applications to regularization," to appear in *IEEE Transactions on Signal Processing*; available at http://arxiv.org/abs/0804.3010.

[40] Y. C. Eldar, "MSE bounds with affine bias dominating the Cramér-Rao bound," to appear in *IEEE Transactions on Signal Processing*.

[41] Y. C. Eldar, "Minimum variance in biased estimation: Bounds and asymptotically optimal estimators," *IEEE Transactions on Signal Processing*, vol. 52, pp. 1915–1930, July 2004.

[42] Y. C. Eldar, "Robust deconvolution of deterministic and random signals," *IEEE Transactions on Information Theory*, vol. 51, pp. 2921–2929, August 2005.

[43] Y. C. Eldar, "Comparing between estimation approaches: Admissible and dominating linear estimators," *IEEE Transactions on Signal Processing*, vol. 54, pp. 1689–1702, May 2006.

[44] Y. C. Eldar, "Minimax estimation of deterministic parameters in linear models with a random model matrix," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 601–612, February 2006.

[45] Y. C. Eldar, "Minimax MSE estimation with noise covariance uncertainties," *IEEE Transactions on Signal Processing*, vol. 54, no. 1, pp. 138–145, January 2006.

[46] Y. C. Eldar, "Robust competitive estimation with signal and noise covariance uncertainties," *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4532–4547, October 2006.

[47] Y. C. Eldar, "Uniformly improving the Cramér-Rao bound and maximum-likelihood estimation," *IEEE Transactions on Signal Processing*, vol. 54, pp. 2943–2956, August 2006.

[48] Y. C. Eldar, "Universal weighted MSE improvement of the least-squares estimator," *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 1788–1800, 2008.

[49] Y. C. Eldar, A. Beck, and M. Teboulle, "A minimax Chebyshev estimator for bounded error estimation," *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1388–1397, April 2008.

[50] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Linear minimax regret estimation of deterministic parameters with bounded data uncertainties," *IEEE Transactions on Signal Processing*, vol. 52, pp. 2177–2188, August 2004.

[51] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Robust mean-squared error estimation in the presence of model uncertainties," *IEEE Transactions on Signal Processing*, vol. 53, pp. 168–181, January 2005.

[52] Y. C. Eldar and J. Chernoi, "A pre-test like estimator dominating the least-squares method," *Journal of Statistical Planning and Inference*, vol. 138, no. 10, pp. 3069–3085, 1 October 2008.

[53] Y. C. Eldar and N. Merhav, "A competitive minimax approach to robust estimation of random parameters," *IEEE Transactions on Signal Processing*, vol. 52, pp. 1931–1946, July 2004.

[54] Y. C. Eldar and N. Merhav, "Minimax MSE–ratio estimation with signal covariance uncertainties," *IEEE Transactions on Signal Processing*, vol. 53, pp. 1335–1347, April 2005.

[55] Y. C. Eldar and A. Nehorai, "Mean-squared error beamforming for signal estimation: A competitive approach," in *Robust Adaptive Beamforming*, (J. Li and P. Stoica, eds.), pp. 259–298, John Wiley & Sons, Inc., 2006.

[56] Y. C. Eldar, A. Nehorai, and P. S. La Rosa, "An expected least-squares beamforming approach to signal estimation with steering vector uncertainties," *IEEE Signal Processing Letters*, vol. 13, no. 5, pp. 288–291, May 2006.

[57] Y. C. Eldar, A. Nehorai, and P. S. La Rosa, "A competitive mean-squared error approach to beamforming," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5143–5154, November 2007.

[58] Y. C. Eldar and A. V. Oppenheim, "Covariance shaping least-squares estimation," *IEEE Transactions on Signal Processing*, vol. 51, pp. 686–697, March 2003.

[59] R. A. Fisher, "On an absolute criterion for fitting frequency curves," *Messenger of Mathematics*, vol. 41, pp. 155–160, 1912.

[60] M. Fréchet, "Sur l'extension de certaines evaluations statistiques de petits echantillons," *Rev. Inst. Znt. Stat.*, vol. 11, pp. 128–205, 1943.

[61] N. P. Galatsanos and A. K. Katsaggelos, "Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation," *IEEE Transactions on Image Processing*, vol. 1, no. 3, pp. 322–336, 1992.

[62] K. F. Gauss, *Theoria Combination is Obsercationunt Erronbus Minimis Obnoxiae*. 1821.

[63] K. F. Gauss, *Theory of Motion of Heavenly Bodies Moving about the Sun in Conic Sections*. New York, NY: Dover, 1963.

[64] G. H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. 215–223, May 1979.

[65] I. J. Good and R. A. Gaskins, "Nonparametric roughness penalties for probability densities," *Biometrika*, vol. 58, pp. 255–277, 1971.

[66] I. J. Good and R. A. Gaskins, "Density estimation and bump hunting by the penalized likelihood method exemplified by scattering and meteorite data (with discussion and rejoinder)," *Journal of American Statistics*, vol. 10, pp. 811–824, 1980.

[67] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Transactions on Information Theory*, vol. 26, pp. 1285–1301, November 1990.

[68] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming (web page and software)," http://stanford.edu/˜boyd/cvx, March 2008.

[69] E. Greenberg and C. E. Webster Jr., *Advanced Econometrics*. New York: Wiley, Second Edition, 1983.

[70] M. H. J. Gruber, *Regression Estimators: A Comparative Study*. San Diego, CA: Academic Press Inc., 1990.

[71] P. R. Halmos, "The theory of unbiased estimation," *Annals of Mathematical Statistics*, vol. 17, pp. 34–43, 1946.

[72] M. Hanke and P. C. Hansen, "Regularization methods for large-scale problems," *Surveys Mathematical Industry*, vol. 3, no. 4, pp. 253–315, 1993.

[73] P. C. Hansen, "The use of the L-curve in the regularization of discrete ill-posed problems," *SIAM Journal of Science Statistical Computation*, vol. 14, pp. 1487–1503, 1993.

[74] P. C. Hansen, "Regularization tools, a matlab package for analysis of discrete regularization problems," *Numerical Algorithms*, vol. 6, pp. 1–35, 1994.

[75] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. Philadelphia, PA: SIAM, 1998.

[76] P. C. Hansen, J. G. Nagy, and D. P. O'Leary, *Deblurring Images: Matrices, Spectra, and Filtering*. Philadelphia, PA: SIAM, 2006.

[77] M. Hardy, "An illuminating counterexample," vol. 110, no. 3, pp. 234–238, 2003.

[78] A. O. Hero, "A Cramer-Rao type lower bound for essentially unbiased parameter estimation," Technical Report, 890, DTIC AD-A246666, MIT Lincoln Lab, Lexington, MA, January 1992.

[79] A. O. Hero and J. A. Fessler, "A recursive algorithm for computing CR-type bounds on estimator covariance," *IEEE Transactions on Information Theory*, vol. 40, pp. 1205–1210, July 1994.

[80] 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, August 1996.

[81] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, pp. 55–67, February 1970.

[82] K. Hoffmann, "Characterization of minimax linear estimators in linear regression," *Statistics*, vol. 10, no. 1, pp. 19–26, 1979.

[83] K. Hoffmann, "Admissible improvements of the least squares estimator," *Statistics*, vol. 11, pp. 373–388, 1980.

[84] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, UK: Cambridge Univ. Press, 1985.

[85] P. J. Huber, *Robust Statistics*. New York: NY, John Wiley & Sons, Inc., 1981.

[86] H. M. Hudson, "A natural identity for exponential families with applications in multiparameter estimation," *Annals of Statistics*, vol. 6, no. 3, pp. 473–484, 1978.

[87] J. T. Hwang, "Improving upon standard estimators in discrete exponential families with applications to poisson and negative binomial cases," *Annals of Statistics*, vol. 10, no. 3, pp. 857–867, 1982.

[88] W. James and C. Stein, "Estimation of quadratic loss," in *Proceedings of Fourth Berkeley Symposium on Mathematical Statistics Probability*, vol. 1, pp. 361–379, Berkeley: University of California Press, 1961.

[89] W. C. Karl, "Regularization in image restoration and reconstruction," in *Handbook of Image and Video Processing*, (A. Bovik, ed.), pp. 183–202, ELSEVIER, Second Edition, 2005.

[90] S. A. Kassam and H. V. Poor, "Robust signal processing for communication systems," *IEEE Communication Magazine*, vol. 21, pp. 20–28, 1983.

[91] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing: A survey," *IEEE Proceedings*, vol. 73, pp. 433–481, March 1985.

[92] S. Kay and Y. C. Eldar, "Rethinking biased estimation," *IEEE Signal Processing Magazine*, vol. 25, no. 3, pp. 133–136, May 2008.

[93] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice Hall Inc., 1993.

[94] B. Koopman, "On distribution admitting a sufficient statistic," *Transactions on American Mathematical Society*, vol. 39, pp. 399–409, 1936.

[95] T. Larsson, M. Patriksson, and A.-B. Strömberg, "On the convergence of conditional ϵ-subgradient methods for convex programs and convex-concave saddle-point problems," *European Journal of Operations Research*, vol. 151, no. 3, pp. 461–473, 2003.

[96] H. Later, "A minimax linear estimator for linear parameters under restrictions in form of inequalities," *Statistics*, vol. 6, pp. 689–695, 1975.

[97] A.-M. Legendre, *Nouvelles méthodes pour la détermination des orbites des cometes*. 1806.

[98] E. L. Lehmann, "A general concept of unbiasedness," *Annals of Mathematical Statistics*, vol. 22, pp. 587–592, December 1951.

[99] E. L. Lehmann, "Estimation with inadequate information," vol. 78, no. 383, pp. 624–627, September 1983.

[100] E. L. Lehmann and G. Casella, *Theory of Point Estimation*. New York, NY: Springer-Verlag, Inc., Second Edition, 1998.

[101] E. H. Lieb and M. Loss, *Analysis*. American Mathematical Society, Second Edition, 2001.

[102] F. Luisier, T. Blu, and M. Unser, "A new SURE approach to image denoising: Interscale orthonormal wavelet thresholding," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 593–606, 2007.

[103] J. H. Manton, V. Krishnamurthy, and H. V. Poor, "James-Stein state filtering algorithms," *IEEE Transactions on Signal Processing*, vol. 46, pp. 2431–2447, September 1998.

[104] D. W. Marquardt, "Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation," *Technometrics*, vol. 12, no. 3, pp. 592–612, August 1970.

[105] Y. Maruyama, "A unified and broadened class of admissible minimax estimators of a multivariate normal mean," *Journal of Multivariate Analysis*, vol. 64, pp. 196–205, 1998.

[106] T. L. Marzetta, "A simple derivation of the constrained multiple parameter Cramér-Rao bound," *IEEE Transactions on Signal Processing*, vol. 41, pp. 2247–2249, June 1993.

[107] L. S. Mayer and T. A. Willke, "On biased estimation in linear models," *Technometrics*, vol. 15, pp. 497–508, August 1973.

[108] L. J. Meng and N. H. Clinthorne, "A modified uniform Cramer-Rao bound for multiple pinhole aperture design," *IEEE Transactions on Medical Imaging*, vol. 23, no. 7, pp. 896–902, July 2004.

[109] M. Milanese and G. Belforte, "Estimation theory and uncertainty intervals evaluation in presence of unknown but bounded errors: Linear families of models and estimators," *IEEE Transactions on Automatic Control*, vol. 27, no. 2, pp. 408–414, 1982.

[110] R. Molina, A. K. Katsaggelos, and J. Mateos, "Bayesian and regularization methods for hyperparameter estimation in image restoration," *IEEE Transactions on Image Processing*, vol. 8, no. 2, pp. 231–246, 1999.

[111] V. A. Morozov, *Methods for Solving Incorrectly Posed Problems*. New York, NY: Springer-Verlag, 1984.

[112] A. Nemirovski, "Prox-method with rate of convergence $o(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems," *SIAM Journal of Optimization*, vol. 15, pp. 229–251, 2004.

[113] Y. Nesterov and A. Nemirovski, *Interior-Point Polynomial Algorithms in Convex Programming*. Philadelphia, PE: SIAM, 1994.

[114] J. P. Norton, "Identification and application of bounded parameter models," *Automatica*, vol. 23, pp. 497–507, 1987.

[115] F. O'Sullivan, "A statistical perspective on ill-posed inverse problems," *Statistical Science*, vol. 1, no. 4, pp. 502–527, 1986.

[116] M. S. Pinsker, "Optimal filtering of square-integrable signals in Gaussian noise," *Problems Information on Transactions*, vol. 16, pp. 120–133, 1980.

[117] E. Pitman, "Sufficient statistics and intrinsic accuracy," *Proceedings of Cambridge Philosophical Society*, vol. 32, pp. 567–579, 1936.

[118] S. Ramani, T. Blu, and M. Unser, "Blind optimization of algorithm parameters for signal denoising by Monte-Carlo SURE," in *Proceedings of International Conference on Acoustics, Speech, Signal Processing (ICASSP-2008), (Las-Vegas, NV)*, April 2008.

[119] C. R. Rao, "Minimum variance and the estimation of several parameters," in *Proceedings of Cambridge Philosophical Society*, pp. 280–283, 1946.

[120] C. R. Rao, *Linear Statistical Inference and Its Applications*. New York, NY: John Wiley & Sons, Inc., Second Edition, 1973.

[121] C. R. Rao, "Estimation of a parameter in linear models," *Annals of Statistics*, vol. 4, pp. 1023–1037, 1976.

[122] J. Rice, "Choice of smoothing parameter in deconvolution problems," *Contemporary Mathematics*, vol. 59, pp. 137–151, 1986.

[123] R. T. Rockafellar, *Convex Analysis*. Princeton NJ: Princeton Univ. Press, 1970.

[124] J. P. Romano and A. F. Siegel, *Counterexamples in Probability and Statistics*. Monterey, CA: Wadsworth & Brooks, 1985.

[125] Y. Rong, Y. C. Eldar, and A. B. Gershman, "Performance tradeoffs among adaptive beamforming criteria," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 651–659, December 2007.

[126] D. L. Snyder and M. I. Miller, *Random Point Processes in Time and Space*. New York: Springer-Verlag, 1991.

[127] P. Speckman, "Spline smoothing and optimal rates of convergence in nonparametric regression models," *Annals of Statistics*, vol. 13, no. 3, pp. 970–983, 1985.

[128] C. Stein, "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution," in *Proceedings on Third Berkeley Symposium on Mathematical Statistical Probability*, vol. 1, pp. 197–206, Berkeley: University of California Press, 1956.

[129] C. M. Stein, "Estimation of the mean of a multivariate distribution," *Proceedings of Prague Symposium on Asymptotic Statistics*, pp. 345–381, 1973.

[130] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *Annals of Statistics*, vol. 9, no. 6, pp. 1135–1151, November 1981.

[131] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*. Upper Saddle River, NJ: Prentice Hall Inc., 1997.

[132] A. Stuart and J. K. Ord, *Kendall's Advanced Theory of Statistics*. Vol. 2, London: Edward Arnold, Fifth Edition, 1991.

[133] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11–12, pp. 625–653, 1999.

[134] P. Tichavsky, C. H. Muravchik, and A. Nehorai, "Posterior Cramer-Rao bounds for discrete-time nonlinear filtering," *IEEE Transactions on Signal Processing*, vol. 46, no. 5, pp. 1386–1396, May 1998.

[135] A. N. Tikhonov, "Solution of incorrectly formulated problems and the regularization method," *Soviet Mathematique Dokldy*, vol. 5, pp. 1035–1038, 1963.

[136] A. N. Tikhonov and V. Y. Arsenin, *Solution of Ill-Posed Problems*. Washington, DC: V.H. Winston, 1977.

[137] D. M. Titterington, "Common structure of smoothing techniques in statistics," *International Statistics Review*, vol. 53, pp. 141–170, 1985.

[138] K. C. Toh, M. J. Todd, and R. H. Tütüncü, "SDPT3 — a MATLAB software package for semidefinite programming, version 1.3," *Optimization Methods Software*, vol. 11/12, no. 1–4, pp. 545–581, 1999 (Interior point methods).

[139] J. F. Traub, G. Wasikowski, and H. Wozinakowski, *Information-Based Complexity*. New York: Academic, 1988.

[140] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*. John Willey and Sons Inc., 1968.

[141] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Review*, vol. 38, no. 1, pp. 40–95, March 1996.

[142] K. S. Vastola and H. V. Poor, "Robust Wiener-Kolmogorov theory," *IEEE Transactions on Information Theory*, vol. IT-30, pp. 316–327, March 1984.

[143] A. Wiesel, "Convex optimization methods in MIMO communication systems," Technical Report, Ph.D. thesis, Technion — Israel Institute of Technology, 2007.

[144] A. Wiesel, Y. C. Eldar, and A. Beck, "Maximum likelihood estimation in linear models with a Gaussian model matrix," *IEEE Signal Processing Letter*, vol. 13, no. 5, pp. 292–295, May 2006.

[145] A. Wiesel, Y. C. Eldar, and Y. Yeredor, "Linear regression with Gaussian model uncertainty: Algorithms and bounds," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2194–2205, June 2008.

[146] S. Zacks, *The Theory of Statistical Inference*. John Willey and Sons, Inc., 1971.

[147] X. P. Zhang and M. D. Desai, "Adapting denoising based on SURE risk," *IEEE Signal Processing Letters*, vol. 5, no. 10, pp. 265–267, 1998.