Cramér-Rao Lower Bounds for Biased Image Reconstruction

Jeffrey A. Fessler and Alfred O. Hero
University of Michigan
Ann Arbor, Michigan 48109-0552

Abstract

Since image reconstruction and restoration are ill-posed problems, unbiased estimators often have unacceptably high variance. To reduce the variance, one introduces constraints and smoothness penalties, which yields biased estimators. This bias precludes the use of the classical Cramér-Rao (CR) lower bound for the variance of an unbiased estimator. This paper presents a uniform bound for minimum variance subject to a bias gradient constraint. Since the bound is independent of any estimator, one can explore the fundamental tradeoff between bias and variance in ill-posed problems. We apply the bound to a linear Gaussian model, and demonstrate the optimality of a simple penalized least-squares estimator.

I. Biased CR bounds

Let $\mathbf{Y}$ be a random vector whose density $f(\boldsymbol{y};\boldsymbol{\theta})$ depends on a vector parameter $\boldsymbol{\theta}$ of length p. The Fisher information of $\mathbf{Y}$ for $\boldsymbol{\theta}$ is the $p \times p$ matrix

$$\mathbf{F}(\boldsymbol{\theta}) = E\{-\nabla^2 \log f(\mathbf{Y};\boldsymbol{\theta})\}.$$

Let $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\boldsymbol{y})$ be an estimator for $\boldsymbol{\theta}$. The bias of $\hat{\boldsymbol{\theta}}$ is

$$\boldsymbol{b}(\boldsymbol{\theta}) = E\{\hat{\boldsymbol{\theta}}\} - \boldsymbol{\theta} = \int \hat{\boldsymbol{\theta}}(\boldsymbol{y}) f(\boldsymbol{y};\boldsymbol{\theta})\, d\boldsymbol{y} - \boldsymbol{\theta}.$$

A. Ordinary CR Bound

The classical CR bound [1] for the minimum variance of an estimator with bias $\boldsymbol{b}(\boldsymbol{\theta})$ is given by:

$$\mathrm{Cov}\{\hat{\boldsymbol{\theta}}\} \geq (\mathbf{I} + \nabla \boldsymbol{b}(\boldsymbol{\theta}))' \mathbf{F}^{+}(\boldsymbol{\theta}) (\mathbf{I} + \nabla \boldsymbol{b}(\boldsymbol{\theta})),$$

where $'$ denotes matrix transpose, $\mathbf{I}$ is the $p \times p$ identity matrix, $\mathbf{F}^{+}$ is the Moore-Penrose inverse of $\mathbf{F}$, and ∇ denotes the row gradient operator. This bound is of very limited use if one hopes to compare a broad class of estimators, rather than just those that happen to have the bias gradient $\nabla \boldsymbol{b}(\boldsymbol{\theta})$.

This work was supported in part by a DOE Alexander Hollaender Postdoctoral Fellowship, DOE Grant DE-FG02-87ER65061, and NIH Grants CA54362, CA52880, and CA60711.

B. Uniform Bound

Since all regularized estimators entail a tradeoff between bias and variance, we would like to know how much one can reduce the variance by introducing some given amount of bias. As we describe later, it is advantageous to constrain the bias gradient rather than the bias itself. In [2], a "uniform" CR bound on the variance of θ_1 was derived for nonsingular Fisher information matrices. We generalize that derivation here by allowing singular Fisher information matrices. This is important since image reconstruction problems are often underdetermined (i.e., when the number of image pixels to be estimated is greated than the number of measurements). Note however that $\mathbf{F}$ will always be nonnegative definite [3].

For simplicity, we consider only the first component θ_1, and define the bias gradient column vector

$$\boldsymbol{g}(\boldsymbol{\theta}) = \nabla' \boldsymbol{b}(\boldsymbol{\theta}) \boldsymbol{e}_1,$$

where $\boldsymbol{e}_1$ is the first unit vector of length p. We constrain the magnitude of the bias gradient by:

$$\|\boldsymbol{g}(\boldsymbol{\theta})\|_{\mathbf{C}} \leq \delta, \qquad (1)$$

where $\|\boldsymbol{g}\|_{\mathbf{C}}^2 = \boldsymbol{g}' \mathbf{C} \boldsymbol{g}$. We assume $\mathbf{C}$ is positive definite. Consider the following sequence of inequalities:

$$\begin{aligned} \mathrm{Var}\{\hat{\theta}_1\} &\geq (\boldsymbol{e}_1 + \boldsymbol{g}(\boldsymbol{\theta}))' \mathbf{F}^{+} (\boldsymbol{e}_1 + \boldsymbol{g}(\boldsymbol{\theta})) \\ &\geq \min_{\boldsymbol{g}(\boldsymbol{\theta}) \,:\, \|\boldsymbol{g}(\boldsymbol{\theta})\|_{\mathbf{C}} \leq \delta} (\boldsymbol{e}_1 + \boldsymbol{g}(\boldsymbol{\theta}))' \mathbf{F}^{+} (\boldsymbol{e}_1 + \boldsymbol{g}(\boldsymbol{\theta})) \\ &\geq \min_{\boldsymbol{d} \,:\, \|\boldsymbol{d}\|_{\mathbf{C}} \leq \delta} (\boldsymbol{e}_1 + \boldsymbol{d})' \mathbf{F}^{+} (\boldsymbol{e}_1 + \boldsymbol{d}). \end{aligned}$$

The constrained minimization over $\boldsymbol{d}$ can be solved using a Lagrange multiplier, yielding the (unique) solution:

$$\boldsymbol{d}_\gamma = \begin{cases} -\boldsymbol{e}_1, & \gamma = 0 \\ \mathbf{0}, & \gamma = \infty \\ -(\gamma \mathbf{C} + \mathbf{F}^{+})^{-1} \mathbf{F}^{+} \boldsymbol{e}_1, & \gamma \in (0, \infty) \end{cases}, \qquad (2)$$

where $\gamma = \gamma(\delta)$ is the solution to

$$\boldsymbol{d}_\gamma' \mathbf{C} \boldsymbol{d}_\gamma = \delta^2.$$

Defining

$$B_1(\boldsymbol{\theta}, \delta) = (\boldsymbol{e}_1 + \boldsymbol{d}_{\gamma(\delta)})' \mathbf{F}^{+}(\boldsymbol{\theta}) (\boldsymbol{e}_1 + \boldsymbol{d}_{\gamma(\delta)}),$$

we see then that $B_1(\boldsymbol{\theta}, \delta)$ is a uniform lower bound on the variance of θ_1 for all estimators satisfying the constraint (1) on the magnitude of the bias gradient.

In general, computing the uniform bound for any particular δ is less informative then examining the graph $(\delta^2, B(\boldsymbol{\theta}, \delta))$ as δ varies over $[0, 1]$. Since there is a one-to-one relationship between δ and the Lagrange parameter γ, in the sequel we typically ignore δ and examine the graph $(\boldsymbol{d}_\gamma' \mathbf{C} \boldsymbol{d}_\gamma, B(\gamma))$ where

$$B(\gamma) = (\boldsymbol{e}_1 + \boldsymbol{d}_\gamma)' \mathbf{F}^+ (\boldsymbol{e}_1 + \boldsymbol{d}_\gamma).$$

As γ varies from 0 to ∞, the bound $B(\gamma)$ varies from 0 to $\mathbf{F}^+$, whereas the magnitude of the bias gradient varies between 1 and 0. One can show from (2) that

$$\boldsymbol{e}_1 + \boldsymbol{d}_\gamma = (\mathbf{I} + \gamma^{-1} \mathbf{C}^{-1} \mathbf{F}^+)^{-1} \boldsymbol{e}_1,$$

so

$$B(\gamma) = \boldsymbol{e}_1' (\mathbf{I} + \gamma^{-1} \mathbf{F}^+ \mathbf{C}^{-1})^{-1} \mathbf{F}^+ (\mathbf{I} + \gamma^{-1} \mathbf{C}^{-1} \mathbf{F}^+)^{-1} \boldsymbol{e}_1, \tag{3}$$

and

$$\boldsymbol{d}_\gamma' \mathbf{C} \boldsymbol{d}_\gamma = \|(\gamma \mathbf{C} + \mathbf{F}^+)^{-1} \boldsymbol{e}_1\|_{\mathbf{C}}^2. \tag{4}$$

C. Symmetry

Since a Fisher information matrix $\mathbf{F}$ is symmetric non-negative definite, it is diagonalized by some orthonormal matrix $\mathbf{V}$. Let

$$\mathbf{F} = \mathbf{V} \boldsymbol{\Lambda} \mathbf{V}'$$

where $\boldsymbol{\Lambda}$ is diagonal with the eigenvalues of $\mathbf{F}$. Thus

$$\mathbf{F}^+ = \mathbf{V} \boldsymbol{\Lambda}^+ \mathbf{V}',$$

where $\boldsymbol{\Lambda}^+$ is diagonal with entries that are zero where $\boldsymbol{\Lambda}$ is zero and the reciprocal of $\boldsymbol{\Lambda}$ otherwise. Assume that $\mathbf{C}$ can also be diagonalized by $\mathbf{V}$:

$$\mathbf{C} = \mathbf{V} \boldsymbol{\Omega} \mathbf{V}'$$

(this is trivially true when $\mathbf{C} = \mathbf{I}$). Then from (2)

$$\boldsymbol{d}_\gamma = -\mathbf{V} (\gamma \boldsymbol{\Omega} + \boldsymbol{\Lambda}^+)^{-1} \boldsymbol{\Lambda}^+ \boldsymbol{q} \tag{5}$$

and

$$\boldsymbol{e}_1 + \boldsymbol{d}_\gamma = -\mathbf{V} (\mathbf{I} + \gamma^{-1} \boldsymbol{\Omega}^{-1} \boldsymbol{\Lambda}^+)^{-1} \boldsymbol{q}, \tag{6}$$

where $\boldsymbol{q} = \mathbf{V}' \boldsymbol{e}_1$ is the first column of $\mathbf{V}'$. Substituting (5) and (6) into (3) and (4) and commuting the diagonal matrices:

$$B(\gamma) = \boldsymbol{q}' (\gamma^{-1} \mathbf{I} + \boldsymbol{\Omega} \boldsymbol{\Lambda})^{-2} \boldsymbol{\Lambda} \boldsymbol{\Omega}^2 \boldsymbol{q}$$

and

$$\boldsymbol{d}_\gamma' \mathbf{C} \boldsymbol{d}_\gamma = \boldsymbol{q}' (\gamma \boldsymbol{\Omega} + \boldsymbol{\Lambda}^+)^{-2} \boldsymbol{\Omega} (\boldsymbol{\Lambda}^+)^2 \boldsymbol{q}.$$

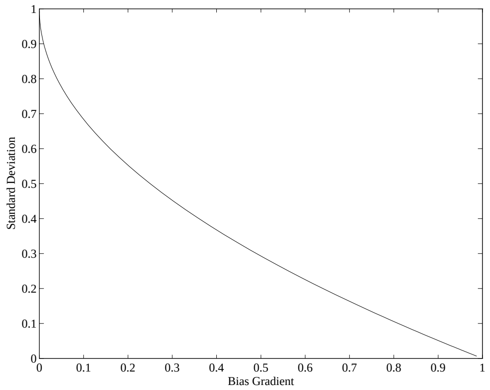

Figure 1: Canonical bias gradient versus standard deviation graph.

Let λ_k denote the diagonal entries of $\boldsymbol{\Lambda}$, and ω_k the diagonal entries of $\boldsymbol{\Omega}$, for $k = 1, \ldots, p$. Then

$$B(\gamma) = \sum_{k=1}^{p} q_k^2 \frac{\lambda_k \omega_k^2}{(\gamma^{-1} + \omega_k \lambda_k)^2}, \tag{7}$$

$$\boldsymbol{d}_\gamma' \mathbf{C} \boldsymbol{d}_\gamma = \sum_{k : \lambda_k \neq 0} q_k^2 \frac{\omega_k}{(1 + \gamma \omega_k \lambda_k)^2}. \tag{8}$$

From (7) and (8), we see that the bias-variance graph $(\boldsymbol{d}_\gamma' \mathbf{C} \boldsymbol{d}_\gamma, B(\gamma))$ is the weighted sum of p graphs of the form

$$\left(\frac{1}{(1 + \gamma \omega \lambda)^2}, \frac{\omega \lambda}{(\gamma^{-1} + \omega \lambda)^2}, \right)$$

for $\lambda \neq 0$, where the weights are $\omega_k q_k^2$. The basic form is shown in Figure 1. Note the steep slope at zero: a small bias can significantly reduce variance. We refer to this graph as the canonical bias-variance curve.

Note that if $\mathbf{F}$ is circulant, then $\mathbf{V}$ is the discrete Fourier basis, $q_k^2 = 1/p$, and each λ_k is the kth Fourier amplitude.

II. Linear gaussian model

In the remainder we focus on the linear Gaussian model:

$$\mathbf{Y} \sim \mathcal{N}(\mathbf{A} \boldsymbol{\theta}, \boldsymbol{\Pi}),$$

where $\boldsymbol{\theta} \in \Re^p$ is the parameter to be estimated, and $\boldsymbol{\Pi}$ is a positive-definite noise covariance matrix. The Fisher information of $\mathbf{Y}$ for $\boldsymbol{\theta}$ is given by

$$\mathbf{F} = \mathbf{A}' \boldsymbol{\Pi}^{-1} \mathbf{A}.$$

A. Linear Estimators

Define the following SVD:

$$\boldsymbol{\Pi}^{-1/2} \mathbf{A} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}',$$

where $\mathbf{U} \in \Re^{n \times n}$, $\mathbf{V} \in \Re^{p \times p}$, $\mathbf{\Sigma} \in \Re^{n \times p}$. Then the Fisher information is

$$\mathbf{F} = \mathbf{A}'\mathbf{\Pi}^{-1}\mathbf{A} = \mathbf{V}\mathbf{\Sigma}'\mathbf{\Sigma}\mathbf{V},$$

so $\mathbf{\Lambda} = \mathbf{\Sigma}'\mathbf{\Sigma}$, i.e. $\lambda_k = \sigma_k^2$. Any linear estimator for $\boldsymbol{\theta}$ can be written:

$$\hat{\boldsymbol{\theta}} = \mathbf{L}\mathbf{Y} = \mathbf{V}\mathbf{W}\mathbf{U}'\mathbf{\Pi}^{-1/2}\mathbf{Y}$$

where $\mathbf{W} \in \Re^{p \times n}$. The bias of this estimator is

$$\boldsymbol{b} = \mathbf{V}\mathbf{W}\mathbf{U}'\mathbf{\Pi}^{-1/2}\mathbf{A}\boldsymbol{\theta} - \boldsymbol{\theta}$$

$$= \mathbf{V}(\mathbf{W}\mathbf{\Sigma} - \mathbf{I})\mathbf{V}'\boldsymbol{\theta}.$$

Thus the bias gradient is

$$\boldsymbol{g} = \mathbf{V}(\mathbf{W}\mathbf{\Sigma} - \mathbf{I})'\mathbf{V}'\boldsymbol{e}_1,$$

so

$$\boldsymbol{g}'\mathbf{C}\boldsymbol{g} = \|\mathbf{\Omega}^{1/2}(\mathbf{W}\mathbf{\Sigma} - \mathbf{I})'\boldsymbol{q}\|^2.$$

The covariance of this estimator is

$$\mathbf{L}\mathbf{\Pi}\mathbf{L}' = \mathbf{V}\mathbf{W}\mathbf{W}'\mathbf{V}',$$

so the variance of $\hat{\theta}_1$ is

$$\boldsymbol{e}_1'\mathbf{V}\mathbf{W}\mathbf{W}'\mathbf{V}'\boldsymbol{e}_1 = \|\mathbf{W}'\boldsymbol{q}\|^2.$$

Note to minimize variance we would like $\mathbf{W}$ to be small, but to minimize bias we would like $\mathbf{W}\mathbf{\Sigma} - \mathbf{I}$ to be small. These conflicting objectives epitomize the bias-variance tradeoff.

If the elements of $\mathbf{W}$ are zero, except for the first $p \times p$ block being diagonal with entries $\{w_k\}_{k=1}^p$, then for this linear estimator the graph of (bias gradient, variance) is:

$$\left(\sum_{k=1}^p q_k^2 \omega_k (\sigma_k w_k - 1)^2, \sum_{k=1}^p q_k^2 w_k^2\right). \tag{9}$$

In particular, if

$$w_k = \frac{\omega_k \sigma_k}{\gamma^{-1} + \omega_k \sigma_k^2}, \tag{10}$$

then the graph (9) has the same form as the bound (7)-(8), except that the bias term (8) only sums over nonzero λ_k. Thus if the Fisher information matrix is nonsingular, then the linear estimator with weights given by (10) achieves the uniform bound (7). We show below that this corresponds to a penalized least-squares estimator. Apparently the choice (10) is the *only* choice that achieves the bound, thus the penalized least-squares estimator is in some sense optimal. In particular, the penalty method appears to be superior to the "truncated SVD" estimator that has been popular in imaging problems.

B. Penalized Weighted Least Squares

A natural regularized estimator for the Gaussian image reconstruction problem is the following penalized weighted least-squares estimator:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} (\boldsymbol{y} - \mathbf{A}\boldsymbol{\theta})'\mathbf{\Pi}^{-1}(\boldsymbol{y} - \mathbf{A}\boldsymbol{\theta}) + \beta\boldsymbol{\theta}'\mathbf{R}\boldsymbol{\theta},$$

where $\mathbf{R}$ is a symmetric nonnegative definite regularization matrix. We assume $\mathbf{F} + \beta\mathbf{R}$ is positive definite, which is reasonable if $\mathbf{R}$ is to be very useful. In the absence of constraints, the solution to this is given by:

$$\hat{\boldsymbol{\theta}} = (\mathbf{F} + \beta\mathbf{R})^{-1}\mathbf{A}'\mathbf{\Pi}^{-1}\boldsymbol{y} \triangleq \mathbf{L}\boldsymbol{y},$$

where $\mathbf{F} = \mathbf{A}'\mathbf{\Pi}^{-1}\mathbf{A}$. Recall that $\mathbf{F} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}'$, and let

$$\mathbf{D}_R = \mathbf{V}'\mathbf{R}\mathbf{V}.$$

Then

$$\hat{\boldsymbol{\theta}} = \mathbf{V}(\mathbf{\Lambda} + \beta\mathbf{D}_R)^{-1}\mathbf{V}'\mathbf{A}'\mathbf{\Pi}^{-1}\boldsymbol{y}.$$

The covariance of this estimator is:

$$\mathrm{Cov}\{\hat{\boldsymbol{\theta}}\} = (\mathbf{F} + \beta\mathbf{R}')^{-1}\mathbf{F}(\mathbf{F} + \beta\mathbf{R})^{-1}$$

$$= \mathbf{V}(\mathbf{\Lambda} + \beta\mathbf{D}_R')^{-1}\mathbf{\Lambda}(\mathbf{\Lambda} + \beta\mathbf{D}_R)^{-1}\mathbf{V}',$$

so

$$\mathrm{Var}\{\hat{\boldsymbol{\theta}}_1\} = \boldsymbol{q}'(\mathbf{\Lambda} + \beta\mathbf{D}_R')^{-1}\mathbf{\Lambda}(\mathbf{\Lambda} + \beta\mathbf{D}_R)^{-1}\boldsymbol{q},$$

The bias is

$$\boldsymbol{b} = (\mathbf{L}\mathbf{A} - \mathbf{I})\boldsymbol{\theta},$$

thus the bias gradient vector for θ_1 is

$$\boldsymbol{g} = (\mathbf{L}\mathbf{A} - \mathbf{I})'\boldsymbol{e}_1 = (\mathbf{F}(\mathbf{F} + \beta\mathbf{R})^{-1} - \mathbf{I})\boldsymbol{e}_1.$$

Therefore

$$\begin{aligned} \boldsymbol{g} &= -\beta\mathbf{R}(\mathbf{F} + \beta\mathbf{R})^{-1}\boldsymbol{e}_1 \\ &= -\beta\mathbf{V}\mathbf{D}_R(\mathbf{\Lambda} + \beta\mathbf{D}_R)^{-1}\boldsymbol{q}, \end{aligned}$$

so

$$\boldsymbol{g}'\mathbf{C}\boldsymbol{g} = \boldsymbol{q}'(\mathbf{D}_R' + \beta^{-1}\mathbf{\Lambda})^{-1}\mathbf{D}_R'\mathbf{\Omega}\mathbf{D}_R(\mathbf{D}_R + \beta^{-1}\mathbf{\Lambda})^{-1}\boldsymbol{q}.$$

In particular, if $\mathbf{D}_R$ is diagonal with entries r_k, then

$$\mathrm{Var}\{\hat{\boldsymbol{\theta}}_1\} = \sum_{k=1}^p q_k^2 \frac{\lambda_k}{(\beta r_k + \lambda_k)^2} \tag{11}$$

and

$$\boldsymbol{g}'\mathbf{C}\boldsymbol{g} = \sum_{k=1}^p q_k^2 \frac{\omega_k r_k^2}{(r_k + \beta^{-1}\lambda_k)^2}. \tag{12}$$

Thus, if $\mathbf{R} = \mathbf{C}^{-1}$ and $\beta = \gamma^{-1}$, then comparing (7) and (8) with (11) and (12) we see that this estimator achieves the uniform bound if $\mathbf{F}$ is nonsingular. If $\mathbf{F}$ is singular, it may be that the uniform bound is unachievable. Note that the results in [2] on achievability are only for the case where $\mathbf{F}$ is nonsingular. We conjecture that tighter bounds may exist for the case where $\mathbf{F}$ is singular.

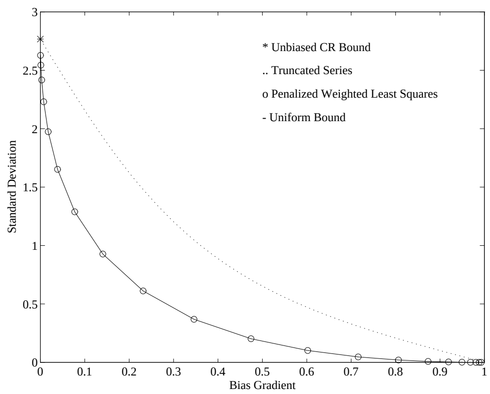

Figure 2: Comparison of truncated series estimation and penalized weighted least squares to uniform bound for a deconvolution problem.

III. Example

To compare the performance of a penalized least-squares estimator to the truncated series estimator and to the uniform bound, we consider a 1-D deconvolution problem. Assume $\mathbf{\Pi} = \mathbf{I}$, the system $\mathbf{A}$ is circulant with kernel (0.15, 0.7, 0.15), and $n = p = 64$. Figure 2 compares the calculated variances for specified bias gradients over a range of γ, β, and n. For a given bias gradient magnitude, the penalized estimator achieves the bound, whereas truncated series estimator does not. For more complicated nonlinear estimators, one has to perform Monte Carlo simulations to evaluate the performance for comparison with the bound. This type of analysis should be very useful for examining estimator performance.

IV. Bias gradient

One might think it would be preferable to have a bound on the minimum variance subject to a magnitude constraint on the bias vector, rather than on the bias gradient. To illustrate why such a bound does not exist, consider the model $Y \sim \mathcal{N}(\theta, 1)$. Applying the uniform bound, one finds that for a bias gradient magnitude constraint of δ, the minimum variance for an estimate of θ is $(1-\delta)^2$. Now consider the "shrinkage" estimator $\hat{\theta} = \beta\pi + (1-\beta)Y$. The variance of $\hat{\theta}$ is $(1-\beta)^2$, and the bias is $\beta(\pi - \theta)$, which could be as small as 0 if θ happened to equal π. Thus, the magnitude of the bias tells us nothing about the variance. On the other hand, the magnitude of the bias gradient for this estimator is $|\beta|$, which immediately tells us (from the uniform bound) that if $\beta = \delta$, the shrinkage estimator achieves the uniform bound.

Although the bias gradient may not be an intuitive object in general, in image reconstruction or restoration problems, the bias gradient is closely related to the bias of a point source. (This is expected, since imaging systems involve tradeoffs between resolution and noise.) To illustrate, recall that for a linear estimator the bias gradient is:

$$\boldsymbol{g} = \mathbf{V}(\mathbf{W\Sigma} - \mathbf{I})'\mathbf{V}'\boldsymbol{e}_1,$$

and the bias is:

$$\boldsymbol{b} = \mathbf{V}(\mathbf{W\Sigma} - \mathbf{I})\mathbf{V}'\boldsymbol{\theta}.$$

If $(\mathbf{W\Sigma})' = (\mathbf{W\Sigma})$, which is usually the case, then one can rewrite the bias gradient as:

$$\boldsymbol{g} = \mathbf{V}(\mathbf{W\Sigma} - \mathbf{I})\mathbf{V}'\boldsymbol{e}_1.$$

Therefore, if the image is a point source, i.e. if $\boldsymbol{\theta} = \boldsymbol{e}_1$, then the bias vector equals the bias gradient vector. For many nonlinear estimators, we conjecture that a similar relationship can be established using perturbation analysis of a point source in a uniform image.

V. Discussion

We have analyzed a uniform bound on the variance for estimators whose bias gradient satisfies a magnitude constraint. For a linear Gaussian model with invertible Fisher information, a penalized least-squares estimator achieves the bound. Further study of the important underdetermined case is ongoing.

References

[1] J. D. Gorman and A. O. Hero. Lower bounds on parametric estimators with constraints. *IEEE Transactions on Information Theory*, 36(6):1285–1301, November 1990.

[2] A. O. Hero. A Cramer-Rao type lower bound for essentially unbiased parameter estimation. Technical Report 890, Lincoln Laboratory, MIT, January 1992.

[3] E. L. Lehmann. *Theory of Point Estimation.* Wiley, New York, 1983.

VI. Acknowledgement

The authors thank M. Usman for helpful discussions.

Comments or questions to `fessler@umich.edu` would be welcomed.