Image Resolution-Variance Tradeoffs Using the Uniform Cramèr-Rao Bound

Thomas J. Kragh and Alfred O. Hero, III
University of Michigan
Department of Electrical Engineering and Computer Science
Ann Arbor, MI, 48109, U.S.A.
e-mail: tkragh@eecs.umich.edu, hero@eecs.umich.edu

ABSTRACT

In image reconstruction and restoration, there exists an inherent tradeoff between the recovered spatial resolution and statistical variance: lower variance can be bought at the price of decreased spatial resolution. This tradeoff can be captured for a particular regularized estimator by tracing out the resolution and variance as a curve indexed by the estimator's smoothing parameter. When the resolution of an estimator is well characterized by the norm of the estimator bias-gradient the uniform Cramèr-Rao (CR) lower bound can be applied. The bias-gradient norm fails, however, to constrain the width of the estimator point response function and the uniform CR bound with bias-gradient norm can give counter-intuitive results. In this paper we present a modified uniform CR bound on estimator variance which captures the width of the estimator point response. These results on the theoretically minimum attainable resolution-variance curve are useful both for exploring near optimality of practical image estimation algorithms and for optimizing the design of image acquisition systems.

1 INTRODUCTION

Image reconstruction and restoration are inherently ill-conditioned problems since physical imaging sensors are resolution limited. Consequently, the full resolution image is unrecoverable from the measurements, i.e. all finite variance estimators of the image are necessarily biased. For such problems there exists an inherent tradeoff between the recovered spatial resolution of an estimator, overall bias, and its statistical variance: lower variance can only be bought at the price of decreased spatial resolution and/or increased overall bias. The goal of this paper is to relate these three fundamental quantities in the analysis of imaging systems.

Let $\underline{\theta} = [\theta_1, \ldots, \theta_n]^T \in \Theta$ be a column vector of unknown, nonrandom parameters that parameterize the density $f_Y(y; \underline{\theta})$ of the observed random variable Y. The parameter space Θ is assumed to be an open subset of the n-dimensional Euclidean space $\mathbb{R}^n$. For a fixed $\underline{\theta}$, let $\hat{\theta}_j = \hat{\theta}_j(Y)$ be an estimator of the jth component of $\underline{\theta}$. Let this estimator have mean value $m_{\underline{\theta}} = E_{\underline{\theta}}[\hat{\theta}_j]$, bias $b_{\underline{\theta}} = m_{\underline{\theta}} - \theta_j$, and variance $\sigma^2_{\underline{\theta}} = E_{\underline{\theta}}[(\hat{\theta}_j - m_{\underline{\theta}})^2]$. In the context of image reconstruction and restoration, Y corresponds to a noise and blur degraded measurement of the true image $\underline{\theta}$, and $\hat{\theta}_j$ is an estimate of the jth pixel of the true image $\underline{\theta}$. Bias $b_{\underline{\theta}}$ is due to mismatch between the estimation algorithm and truth. Variance $\sigma^2_{\underline{\theta}}$ arises from statistical fluctuations due to uncertainty in the measured data Y. Resolution is defined as the effective width of the estimation algorithm point response which will be defined later.

For a particular choice of estimator, the tradeoff between bias and variance is often analyzed by sweeping out the measured bias $b_{\underline{\theta}}$ and variance $\sigma^2_{\underline{\theta}}$, indexed by the estimator's smoothing parameter. Although common in the analysis of imaging system performance, this method has its drawbacks. First, an estimator can always be found where the bias and variance are zero at some point $\underline{\theta}$. For example, setting the estimator value to an arbitrary constant results in a zero-variance (but highly biased) estimator. Second, the bias value $b_{\underline{\theta}}$ penalizes estimators that may have a large constant, and thus removable, bias. Third, these types of tradeoff curves only apply to the particular estimator in question, and do not say anything about the optimality of the particular estimator.

One method to determine the variance of a particular estimator is the Cramèr-Rao lower bound. Let $\mathbf{F}_Y$ be the $n \times n$ Fisher information matrix of the measurements Y, and let $\mathbf{F}_Y^{-1}$ be its inverse. When $\hat{\theta}_j$ is unbiased, its variance $\sigma^2_{\underline{\theta}}$ is bounded below by the jth diagonal element of $\mathbf{F}_Y^{-1}$. Since almost all estimation algorithms of interest used in image processing are biased, this bound is not very useful.

In [1] we presented a lower bound on estimator variance as a function of the norm of the estimator bias-gradient $\|\nabla b_{\underline{\theta}}\|$. When the resolution of an estimator is well characterized by the norm of the estimator bias-gradient the uniform Cramèr-Rao (CR) lower bound can be applied. However, the bias-gradient norm fails to constrain the width of the point spread function and the uniform CR bound with bias-gradient norm can give counter-intuitive results [2].

In this paper we present a modified uniform CR bound which captures the width of the estimator point spread function by placing an additional constraint on the second moment of the estimator mean-gradient $\nabla m_{\underline{\theta}}$. We characterize the estimator which attains this lower bound. For any fixed total bias and (2nd moment) resolution this estimator attains minimum variance at that particular resolution. This work generalizes the uniform CR bound of [1]. These results on the theoretically minimum attainable resolution-variance curve are useful both for exploring near optimality of practical image estimation algorithms and for optimizing the design of image acquisition systems.

2 THE BIASED CR BOUND

For a biased estimator $\hat{\theta}_j$ of θ_j with mean $m_{\underline{\theta}}$ the CR bound has the following form [3], referred to here as the *biased CR bound*.

$$\sigma^2_{\underline{\theta}} \quad \geq \quad (\nabla m_{\underline{\theta}})^T \mathbf{F}_Y^+ (\nabla m_{\underline{\theta}}) \tag{1}$$

where $\nabla m_{\underline{\theta}}$ is the gradient of the estimator mean value $m_{\underline{\theta}}$, $\mathbf{F}_Y = \mathbf{F}_Y(\underline{\theta})$ is the Fisher information matrix,

$$\mathbf{F}_Y = E_{\underline{\theta}} \left\{ [\nabla_{\underline{\theta}} \ln f_Y(Y;\underline{\theta})] \, [\nabla_{\underline{\theta}} \ln f_Y(Y;\underline{\theta})]^T \right\},$$

and $\mathbf{F}_Y^+$ denotes the Moore-Penrose pseudo-inverse of the possibly singular matrix $\mathbf{F}_Y$. Note that the scalar estimator $\hat{\theta}_j$ can be expressed in terms of the vector estimator $\hat{\underline{\theta}} = \hat{\underline{\theta}}(Y)$ by the inner-product $\hat{\theta}_j = \underline{e}_j{}^T \hat{\underline{\theta}}$, where $\underline{e}_j$ is the jth unit vector $(0, \ldots, 0, 1, 0, \ldots, 0)^T$. Thus the estimator mean- and bias-gradient vectors are related by $\nabla m_{\underline{\theta}} = \underline{e}_j + \nabla b_{\underline{\theta}}$, and the CR bound can be expressed in terms of the estimator bias-gradient,

$$\sigma^2_{\underline{\theta}} \quad \geq \quad (\underline{e}_j + \nabla b_{\underline{\theta}})^T \mathbf{F}_Y^+ (\underline{e}_j + \nabla b_{\underline{\theta}}) \tag{2}$$

However, the biased CR bound only applies to estimators with a given bias-gradient vector $\nabla b_{\underline{\theta}}$. Thus (2) can not be used to simultaneously bound the variance of several different estimators, each with comparable but non-equal bias-gradient vectors.

3 THE UNIFORM CR BOUND

The bias-gradient vector $\nabla b_{\underline{\theta}}$ can be interpreted as the *sensitivity* or *coupling* of the bias in the jth pixel estimate to perturbations in the remaining pixels of the image. Thus, its length or norm $\|\nabla b_{\underline{\theta}}\|$ is a measure of the overall bias in the estimate $\hat{\theta}_j$. When $\|\nabla b_{\underline{\theta}}\| \to 0$, the biased CR bound given in (2) reduces to $[\underline{e}_j]^T \mathbf{F}_Y^+ [\underline{e}_j] = [\mathbf{F}_Y^+]_{j,j}$, as one would expect from an unbiased estimator. More precisely, [1] showed that the norm $\delta = \|\nabla b_{\underline{\theta}}\|_{\mathbf{C}}$ of the bias-gradient with respect to a positive definite matrix $\mathbf{C}$ is an upper bound on the maximal bias variation over an ellipsoidal neighborhood $\mathcal{C}$ about $\underline{\theta}$.

The concept behind the UCRB is that for a fixed value of bias-gradient norm $\delta > 0$, find an *optimal* bias-gradient vector $\underline{d}$ that minimizes (2) by performing a constrained minimization over the feasible set of bias gradient vectors $\nabla b_{\underline{\theta}} : \|\nabla b_{\underline{\theta}}\|_{\mathbf{C}} \leq \delta$,

$$\min_{\underline{d}:\|\underline{d}\|_{\mathbf{C}} \leq \delta} (\underline{e}_j + \underline{d})^T \mathbf{F}_Y^+ (\underline{e}_j + \underline{d}) \tag{3}$$

Derivation and proof of the optimal bias-gradient vector $\underline{d}$ in (3) is given in [1].

3.1 UCRB

The uniform CR bound for biased estimators with a given bias-gradient norm δ and non-singular Fisher information matrix $\mathbf{F}_Y$ is as follows. Let $\hat{\theta}_j$ be an estimator of the jth pixel of the true image $\underline{\theta}$. For a fixed $\delta \geq 0$, let the bias-gradient satisfy the norm constraint $\|\nabla b_{\underline{\theta}}\|_{\mathbf{C}} \leq \delta$ where $\mathbf{C}$ is a positive-definite symmetric matrix. Then the variance $\sigma^2_{\underline{\theta}}$ of $\hat{\theta}_j$ satisfies

$$\sigma^2_{\underline{\theta}} \geq B(\underline{\theta}, \delta) \tag{4}$$

where the variance lower bound $B(\underline{\theta}, \delta)$ is given by one of the following two cases:

1. If $\delta^2 \geq \nabla b_{\underline{\theta}}^T \mathbf{C} \nabla b_{\underline{\theta}}$, then $B(\underline{\theta}, \delta) = 0$

2. If $\delta^2 < \nabla b_{\underline{\theta}}^T \mathbf{C} \nabla b_{\underline{\theta}}$, then

$$B(\underline{\theta}, \delta) = (\underline{e}_j + \underline{d})^T \mathbf{F}_Y^{-1} (\underline{e}_j + \underline{d}) \tag{5}$$

where the optimal bias-gradient vector $\underline{d}$ is given by

$$\underline{d} = -[\mathbf{I} + \lambda_1 \mathbf{F}_Y \mathbf{C}]^{-1} \underline{e}_j \tag{6}$$

and λ_1 is the Lagrange multiplier given by the unique solution of $\delta^2 = \underline{d}^T \mathbf{C} \underline{d}$.

Note that the estimator variance lower bound $B(\underline{\theta}, \delta)$ is *independent* of the choice of estimator, and only depends on the Fisher information $\mathbf{F}_Y$ and choice of norm matrix $\mathbf{C}$.

3.2 Example: Limits of Image Restoration

Figure (1) shows a 64x64-pixel image of a Shepp-Logan head phantom, along with a noise- and blur-degraded simulated measurement. Image blur was simulated by convolving with a 5x5 pixel extent, shift-invariant, 1.5-pixel FWHM symmetric gaussian kernel, along with additive gaussian noise of variance $\sigma^2 = 1$.

Figure (2) shows the limiting square-root variance vs. bias-gradient norm of an estimate of pixel (32,32) in the presence of blur and additive gaussian noise. Two different cases are considered: a 1.5-pixel fwhm gaussian blur as in figure (1), along with a 1.75-pixel fwhm blur. In both cases the noise is additive gaussian with unity variance. The bias-gradient norm matrix $\mathbf{C}$ used in calculate the bias-gradient length δ was the identity matrix.

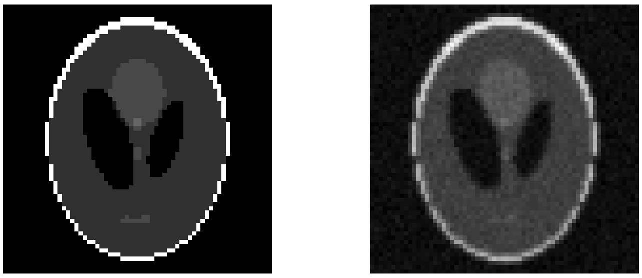

Figure 1: True Image (left), Noisy Image (right).

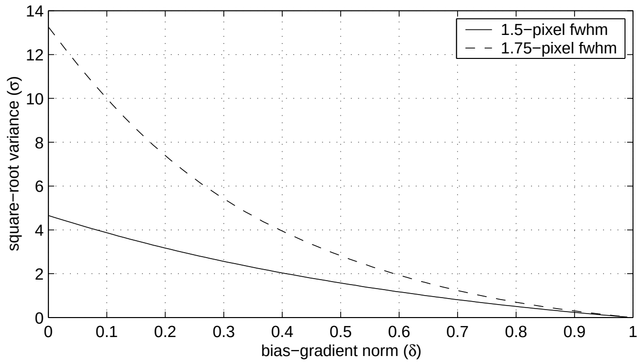

Figure 2: Pixel estimation performance in presence of blur and additive gaussian noise.

Note that the 1.75-pixel fwhm blur case has larger variance then the 1.5-pixel blur case. Estimating a pixel in the presence of larger blur is a more ill-posed problem and results in a noisier estimate for a given total bias.

Figure (3) shows a plot of mean-gradient images for the 1.75-pixel fwhm blur case, for bias-gradient norm $\delta = 0.1$ (left) and $\delta = 0.5$ (right). Note that with increasing bias-gradient norm, the mean-gradient is more spread out.

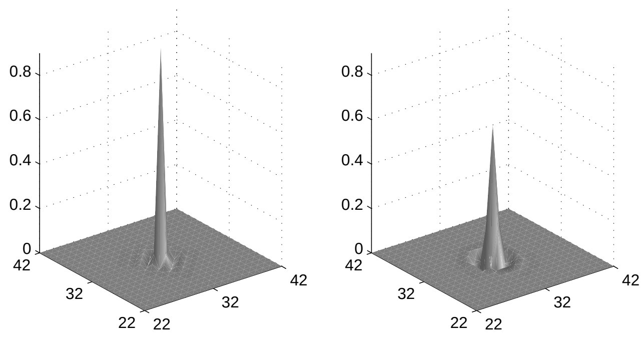

Figure 3: Mean gradients images, $\delta = 0.1, \delta = 0.5$.

3.3 Interpretation Difficulties of the UCRB

One problem with the bias-gradient norm as a measure of estimator resolution is that it is possible for different mean-gradients to have the exact same bias-gradient norm, but with dramatically different resolution properties. Figure (4) shows cross-sectional slices through two representative mean-gradients as a function of their pixel location. Their associated bias-gradients both have the same norm $\delta = 0.5$, however their spread or full-width-half-max are obviously different.

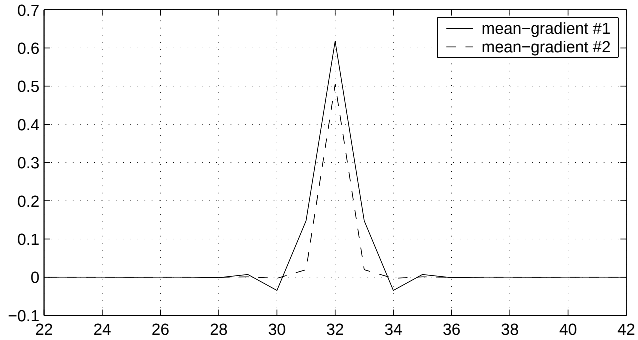

Figure 4: Mean gradient cross-sections, $\delta = 0.5$

4 MODIFICATIONS TO UCRB

As a modification to the UCRB we will add a constraint on the mean-gradient $\nabla m_{\underline{\theta}}$ and shows its connection to the estimator local impulse response. Let $\underline{\mu}(\underline{\theta}) = E_{\underline{\theta}}[\hat{\underline{\theta}}(Y)]$ be the expected value of the vector estimator $\hat{\underline{\theta}}$. Let ϵ be a small perturbation in the pth pixel of the source $\underline{\theta}$. For an estimator with mean $\underline{\mu}(\underline{\theta})$, define the local impulse response vector $\underline{h}$ of all reconstructed pixels due to a perturbation in the pth pixel of $\underline{\theta}$ as

$$\underline{h}^p(\underline{\theta}) = \lim_{\epsilon \to 0} \frac{\underline{\mu}(\underline{\theta} + \epsilon \underline{e}_p) - \underline{\mu}(\underline{\theta})}{\epsilon} = \frac{\partial}{\partial \theta_p} \underline{\mu}(\underline{\theta}) \qquad (7)$$

As noted in [4], this definition of impulse response reflects the space-varying nature of nonlinear estimators. It is space-varying through its dependence on the perturbing pixel index p, and object-dependent through $\underline{\theta}$. The mean-gradient and local impulse response are related by $\nabla m_{\underline{\theta}} = \left[h_j^1(\underline{\theta}), \ldots, h_j^n(\underline{\theta})\right]^T$, where h_j^p is the jth component of $\underline{h}^p$. In general $\nabla m_{\underline{\theta}}$ and $\underline{h}(\underline{\theta})$ are not equivilant except under certain conditions [1, 4]. Consider the case of an estimator whos mean is linear in $\underline{\theta}$: $\underline{\mu}(\underline{\theta}) = \mathbf{L}\underline{\theta}$ for some square matrix $\mathbf{L}$. The local impulse response $\underline{h}^j(\underline{\theta})$ due to a perturbation in the jth source pixel is the jth column of $\mathbf{L}$, while the mean-gradient $\nabla m_{\underline{\theta}} = \nabla E_{\underline{\theta}}[\hat{\theta}_j]$ is the jth row of $\mathbf{L}$. Thus when $\mathbf{L}$ is symmetric the mean-gradient is equivilant to the local impulse response.

The local impulse response of an estimator describes the coupling to all reconstructed pixels due to a perturbation in a single source pixel, while the mean-gradient describes the coupling into a single reconstructed pixel due to perturbations in all source pixels. A weighted norm of this coupling would be a natural measure of the estimator response about the jth pixel. In this case, as with 3.1, we want to find the *optimal* bias-gradient (and thus, mean-gradient) that results in a minimum variance estimate of $\hat{\theta}_j$ given constraints on $\nabla b_{\underline{\theta}}$ and $\nabla m_{\underline{\theta}}$.

4.1 Mean-gradient 2nd-moment

Define the 2nd-moment γ of the mean-gradient as

$$\gamma^2 = \frac{\sum_i d(j,i)^2 (\nabla m_{\underline{\theta}})_i^2}{\sum_i (\nabla m_{\underline{\theta}})_i^2} \tag{8}$$

where $d(j,i)$ is the distance between the jth and i-th pixel (nominally set equal to the Euclidean distance, although any other distance could be used). Since the mean-gradient is the sum of the unit vector $\underline{e}_j$ and the bias-gradient $\nabla b_{\underline{\theta}}$, (8) can be written as the ratio of two quadratic forms,

$$\gamma^2 = \frac{(\underline{e}_j + \nabla b_{\underline{\theta}})^T \mathbf{M}_j (\underline{e}_j + \nabla b_{\underline{\theta}})}{(\underline{e}_j + \nabla b_{\underline{\theta}})^T (\underline{e}_j + \nabla b_{\underline{\theta}})} \tag{9}$$

where $\mathbf{M}_j$ is a positive semi-definite diagonal matrix whose (i,i)-th diagonal entry is $d(j,i)^2$.

4.2 UCRB with Mean-Gradient Constraint

The uniform CR bound for biased estimators with a given bias-gradient norm δ, mean-gradient 2nd-moment γ and non-singular Fisher information $\mathbf{F}_Y$ is as follows. For a fixed $\delta, \gamma \geq 0$, let the bias-gradient satisfy the norm constraint

$$\nabla b_{\underline{\theta}}^T \mathbf{C} \nabla b_{\underline{\theta}} \leq \delta^2$$

and 2nd-moment constraint

$$\frac{(\underline{e}_j + \nabla b_{\underline{\theta}})^T \mathbf{M}_j (\underline{e}_j + \nabla b_{\underline{\theta}})}{(\underline{e}_j + \nabla b_{\underline{\theta}})^T (\underline{e}_j + \nabla b_{\underline{\theta}})} \leq \gamma^2$$

Then the variance $\sigma_{\underline{\theta}}^2$ of the estimator $\hat{\theta}_j$ satisfies

$$\sigma_{\underline{\theta}}^2 \geq B(\underline{\theta}, \delta, \gamma) \tag{10}$$

where the variance lower bound $B(\underline{\theta}, \delta, \gamma)$ is given by the following three cases:

1. If $\delta^2 \geq \nabla b_{\underline{\theta}}^T \mathbf{C} \nabla b_{\underline{\theta}}$, then $B(\underline{\theta}, \delta) = 0$
2. If $\delta^2 < \nabla b_{\underline{\theta}}^T \mathbf{C} \nabla b_{\underline{\theta}}$ and $\gamma \geq \gamma_*$, then

   $$B(\underline{\theta}, \delta, \gamma) = (\underline{e}_j + \underline{d})^T \mathbf{F}_Y^{-1} (\underline{e}_j + \underline{d}) \tag{11}$$

   where $\underline{d}$ is as given in (6) and

   $$\gamma_*^2 = \frac{(\underline{e}_j + \underline{d})^T \mathbf{M}_j (\underline{e}_j + \underline{d})}{(\underline{e}_j + \underline{d})^T (\underline{e}_j + \underline{d})} \tag{12}$$

3. If $\delta^2 < \nabla b_{\underline{\theta}}^T \mathbf{C} \nabla b_{\underline{\theta}}$ and $\gamma < \gamma_*$, then $B(\underline{\theta}, \delta, \gamma)$ is as given in (11), and $\underline{d} =$

   $$-[\mathbf{F}_Y^{-1} + \lambda_1 \mathbf{C} + \lambda_2 [\mathbf{M}_j - \gamma^2 \mathbf{I}]]^{-1} [\mathbf{F}_Y^{-1} - \lambda_2 \gamma^2 \mathbf{I}] \underline{e}_j \tag{13}$$

   where $\lambda_1, \lambda_2 \geq 0$ are Lagrange multipliers found implicitly through the two equality constraints

   $$(\underline{e}_j + \underline{d})^T [\mathbf{M}_j - \gamma^2 \mathbf{I}] (\underline{e}_j + \underline{d}) = 0 \tag{14}$$

   $$\underline{d}^T \mathbf{C} \underline{d} - \delta^2 = 0 \tag{15}$$

4.3 Interpretation

By the addition of a second constraint on the UCRB, we now define a minimum-variance *surface* that all estimators must lie above. For a given estimator, its variance follows a trajectory in (δ, γ), parameterized its regularization parameter. By analyzing the distance the particular estimator lies above the surface, one can determine how far from optimality the estimator is.

4.4 Example Calculation

In figure (5) we show the variance bound surface for the estimation task in 3.2. The image was degraded by a 1.5-pixel fwhm blur along with additive gaussian noise of unity variance. The variance-trajectory of a penalized weighted least-square estimator is superimposed on top. The estimator penalty $\mathbf{P}$ is a 1st-order neighbor

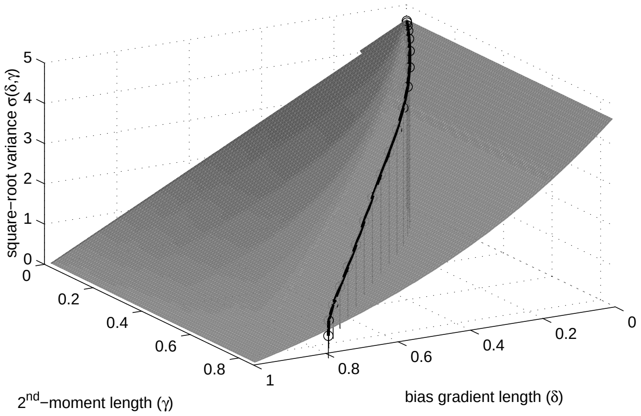

Figure 5: UCRB surface along with PWLS Estimator Trajectory

roughness penalty. The estimator was purposely mismatched from the true system model in order to show it lying above the bound surface (the estimator assumed a 1.75-pixel fwhm blur).

References

[1] A O Hero, J A Fessler, and M Usman, "Exploring estimator bias-variance tradeoffs using the uniform CR bound," *IEEE Tr. Sig. Proc.*, vol. 44, no. 8, pp. 2026–41, Aug. 1996.

[2] C-Y Ng, N H Clinthorne, J A Fessler, A O Hero, Y Zhang, and W L Rogers, "Preliminary studies on the feasibility of addition of vertex view to conventional brain SPECT," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, 1996, vol. 2, pp. 1057–61.

[3] Harry L Van Trees, *Detection, estimation, and modulation theory*, Wiley, New York, 1968.

[4] J A Fessler and W L Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction methods: Space-invariant tomographs," *IEEE Tr. Im. Proc.*, vol. 5, no. 9, pp. 1346–58, Sept. 1996.