Applications of the Uniform CR Bound for SPECT Performance Evaluation and System Design

Mohammad Usman*, Alfred O. Hero*, Leslie Rogers**,
and Jeffrey A. Fessler***
* Dept. of Electrical Engineering and Computer Science,
** Division of Nuclear Medicine, and *** Division of Internal Medicine,
The University of Michigan, Ann Arbor, MI 48109-2122

Abstract

We apply a recently developed uniform Cramer-Rao (CR) bound [7] to study fundamental trade-offs in SPECT imaging, namely, bias-variance and resolution-sensitivity. The uniform CR bound provides a means of specifying achievable and unachievable regions in the bias-variance trade-off plane. The applications considered in this paper are: i) the effect of angular sampling on reconstruction error, ii) optimal aperture design, and iii) performance comparison between penalized maximum likelihood and weighted least squares estimators.

Key Words: Single photon ring tomograph (SPRINT), space alternating generalized expectation-maximization, weighted least-squares estimator, aperture design.

Corresponding author:

Professor Alfred O. Hero III

Room 4229, Dept. of EE and CS

University of Michigan

1301 Beal Avenue

Ann Arbor, MI 48109-2122.

Tel: (313) 763-0564.

e-mail: hero@eecs.umich.edu.

---

[1] This work was supported in part by National Science Foundation under grant BCS-9024370, a Government of Pakistan Postgraduate Fellowship, NIH grant CA-60711, and DOE grant DE-FG02-87ER60561

1 Introduction

Independent of the imaging modality the goodness of the image is eventually decided by human operators and is therefore subjective. Although some measures such as Hotelling trace norm have proven useful [1], to make a designer's job easier, different measures are defined to predict image quality. In single photon emission computed tomography (SPECT) image reconstruction, the following design measures are commonly used.

1. **Spatial or Geometric Resolution:** is a measure of the closest distance at which a projections geometry can distinguish between two point sources. It is usually measured at Full Width at Half Maximum (FWHM) of the geometric point spread function (psf) of the system [3].
2. **Sensitivity:** is the average number of γ-ray photons striking the detectors, per unit area, over a given period of time, for a given source.
3. **Mean Square Error (MSE):** is the square of the difference between the actual intensity $\underline{\theta}$ of the image, and the estimated intensity $\widehat{\underline{\theta}}$. The MSE can be directly related to the estimator bias $\text{bias}_{\underline{\theta}}(\widehat{\underline{\theta}})$:

   $$\text{bias}_{\underline{\theta}}(\widehat{\underline{\theta}}) = E_{\underline{\theta}}(\widehat{\underline{\theta}}) - \underline{\theta},$$

   and the estimator variance $\text{var}_{\underline{\theta}}(\widehat{\underline{\theta}})$:

   $$\text{var}_{\underline{\theta}}(\widehat{\underline{\theta}}) = E_{\underline{\theta}}\left[\widehat{\underline{\theta}} - E_{\underline{\theta}}(\widehat{\underline{\theta}})\right]\left[\widehat{\underline{\theta}} - E_{\underline{\theta}}(\widehat{\underline{\theta}})\right]^T$$

   by:

   $$\text{MSE}_{\underline{\theta}}(\widehat{\underline{\theta}}) = \text{var}_{\underline{\theta}}(\widehat{\underline{\theta}}) + \text{bias}_{\underline{\theta}}(\widehat{\underline{\theta}})\text{bias}_{\underline{\theta}}(\widehat{\underline{\theta}})^T,$$

   where $E_{\underline{\theta}}(\widehat{\underline{\theta}})$ is the expected value of $\widehat{\underline{\theta}}$.

In any imaging system there are inherent resolution-sensitivity and bias-variance trade-offs. For example in SPECT, a wider collimator opening results in increased sensitivity at the price of decreased spatial resolution. Similarly, in image reconstruction, lower variance can only be bought at the price of increased bias via image smoothing [4]. In this paper we apply a recently developed uniform Cramer-Rao (CR) bound for biased estimators [7] to study the bias-variance and resolution-sensitivity trade-offs for several different examples.

The classical CR bound determines a lower bound on the variance of an unbiased parameter estimator. Although this CR bound is known to be asymptotically achievable for a large number of independent and identically distributed observations, in practice, many parameter estimation algorithms are biased. In [7] a 'uniform' CR bound was obtained which is applicable to all biased estimators whose bias gradient length is less than a pre-specified threshold 'δ'. The uniform CR bound is more useful than the standard extension of the CR bound [10] in that the latter extension only applies to estimators with the same bias gradient function $\nabla_{\underline{\theta}} \mathrm{bias}_{\underline{\theta}}(\widehat{\underline{\theta}}) = \nabla_{\underline{\theta}} \underline{b}$, where $\nabla_{\underline{\theta}} \underline{b}$ is a fixed function. The uniform CR bound provides a means for specifying achievable and unachievable regions in the bias gradient-variance trade-off plane, δ–σ, hence different estimators can be effectively compared by plotting their performance on the δ–σ trade-off plane. It is shown in [4] that given an accurate estimate of the bias gradient length of an estimator, the δ–σ trade-off plane can be translated into a true bias-variance trade-off plane b–σ for the estimator. Furthermore, it is shown in [11] that the bias gradient can be related to the spatial resolution.

The applications considered in this paper are: i) effect of angular sampling on image reconstruction, ii) optimal aperture design for intensity estimation, and iii) performance comparison of the biased weighted least-squares and penalized maximum likelihood estimators.

1.1 SPRINT System Description

The system investigated is shown in Figure 1 and is called single photon ring tomograph (SPRINT) system [2]. The system was designed specifically for brain imaging and consists of a ring of detectors and a rotating multi-slit collimator ring.

The system parameters are given in Appendix A and unless otherwise specified are those used in the simulations.

2 Statistical Model in SPECT

A radioactive source emits γ-rays in all directions with equal probability at spatial positions $\underline{V} = [V_1, ..., V_n]^T \in \mathcal{V}$. The source emissions are governed by a spatial intensity vector $\underline{\theta} = [\theta_1, ..., \theta_n]^T \in \Theta$, defined as the average number of γ-ray photons emitted by a pixel during the imaging time. The γ-rays photons are detected at positions $\underline{W} = [W_1, ..., W_d]^T \in \mathcal{W}$. The image reconstruction problem is stated as follows: given the projection data estimate the intensities $\underline{\theta} = [\theta_1, ..., \theta_n]^T \in \Theta$. The number of γ-ray emitted from each pixel, denoted by the vector $\underline{X} = [X_1, ..., X_n]^T$, and the number of detections in each detector bin, denoted by the vector $\underline{Y} = [Y_1, ..., Y_d]^T$ are Poisson distributed:

$$\begin{aligned} f_{\underline{X}}(\underline{x}; \underline{\theta}) &= \prod_{i=1}^{n} \frac{\theta_i^{x_i}}{x_i!} e^{\theta_i} \qquad (1) \\ f_{\underline{Y}}(\underline{y}; \underline{\theta}) &= \prod_{j=1}^{d} \frac{\mu_j^{y_j}}{y_j!} e^{\mu_j}, \qquad (2) \end{aligned}$$

where μ_j is the γ-ray intensity at detector j,

$$\underline{\mu} = A\,\underline{\theta} + \underline{s}, \qquad (3)$$

and A is the so-called weights or system matrix consisting of the transition pdfs $f_{\underline{W}/\underline{V}}(\underline{w}/\underline{v})$ that depends upon the system geometry and $\underline{s}$ respresents the scatter or the background events.

Shao [3] has derived the weights matrix, A, for the SPRINT ring.

3 Uniform Cramer-Rao Bound

Given the projection data $\underline{Y}$, and a probability density associated with the detector intensity $f_{\underline{Y}}(\underline{y}; \underline{\theta})$, a common objective is to estimate the pixel intensities $\underline{\theta}$ in a small *region of interest* (ROI) $\underline{\theta}_{ROI} = [\theta_1, ..., \theta_q]^T$; $q \ll n$. In the sequel we specialize to the case when the ROI consists of a single pixel θ_1. A function $\widehat{\theta}_1(\underline{Y})$ is a parameter estimator based on the projections $\underline{Y}$. $\widehat{\theta}_1$ is unbiased if the bias function $b_1(\underline{\theta}) = \text{bias}_{\underline{\theta}}(\widehat{\theta}_1) = E_{\underline{\theta}}[\widehat{\theta}_1 - \theta_1]$ is equal to zero.

The Cramer-Rao lower bound on the variance of unbiased parameter estimator $\widehat{\theta}_1$ is given by the upper-left (1,1) element of the inverse of an $n \times n$, symmetric, positive definite Fisher information matrix (FIM) $F_Y = F_Y(\underline{\theta})$:

$$\mathrm{var}_{\underline{\theta}}(\hat{\theta}_1) \geq \underline{e}_1^T \; F_Y^{-1} \; \underline{e}_1, \tag{4}$$

where,

$$F_Y = E_{\underline{\theta}}[\nabla_{\underline{\theta}}^T \ln f_{\underline{Y}}(\underline{y};\underline{\theta}) \; \nabla_{\underline{\theta}} \ln f_{\underline{Y}}(\underline{y};\underline{\theta})],$$

$\nabla_{\underline{\theta}}$ denotes the (row) gradient vector $[\frac{\partial}{\partial\theta_1}, ..., \frac{\partial}{\partial\theta_n}]$, and $\underline{e}_1 = [1, 0, ..., 0]^T$ is a n-element unit vector.

While the unbiased CR bound (4) is known to be asymptotically achievable for large number of independent identically distributed measurements, in practice, most image reconstruction algorithms are biased and the unbiased CR bound is inapplicable.

3.1 Motivating Example

In Figure 2 the unbiased CR bound is plotted as a function of the collimator rotations. Also plotted is the variance of the weighted least-squares estimator (WLSE) [4, 15]. The variance of the WLSE is clearly less than the CR lower bound at all points. This is due to the fact that the weighted least-square estimator is biased. A biased version of the CR bound is needed to ensure that no bound violation occurs in examples such as this.

3.2 Biased CR Bound

For a biased estimator $\widehat{\theta}_1$ the following form of the biased CR bound is well known [10]:

$$\mathrm{var}_{\underline{\theta}}(\widehat{\theta}_1) \geq [\nabla_{\underline{\theta}} m_1] \; F_Y^{-1} \; [\nabla_{\underline{\theta}} m_1]^T, \tag{5}$$

where $\nabla_{\underline{\theta}} m_1 = \nabla_{\underline{\theta}} m_1(\underline{\theta}) = \nabla_{\underline{\theta}} b_1 + \underline{e}_1$ is an n element row vector of the gradient of the mean $E_{\underline{\theta}}(\widehat{\theta}_1) = m_1(\underline{\theta})$.

3.3 Uniform CR Bound

The application of the biased CR bound (5) is very restricted due to the fact that it is only applicable to estimators with a given bias gradient $\nabla_{\underline{\theta}} b_1$. In [7] Hero gives a 'uniform' CR bound on the variance of a single parameter θ_1 for non-singular F_Y. This bound is applicable to all biased estimators whose bias gradient length $||\nabla_{\underline{\theta}} b_1||$ satisfies:

$$||\nabla_{\underline{\theta}} b_1||^2 \leq \delta^2 < 1. \tag{6}$$

The following theorem is proven in [7].

Theorem 1 *Let $\widehat{\theta}_1$ be an estimator with bias $b_1(\underline{\theta})$ whose n-element bias gradient vector $\nabla_{\underline{\theta}} b_1$ satisfies (6). Assume that the FIM F_Y is non-singular. Then the variance of $\widehat{\theta}_1$ is given by:*

$$var_{\underline{\theta}}(\widehat{\theta}_1) \geq B(\underline{\theta}, \delta), \tag{7}$$

where $B(\underline{\theta}, \delta)$ is equal to:

$$\begin{aligned} B(\underline{\theta}, \delta) &= [\underline{e}_1 + \underline{d}_{min}]^T F_Y^{-1} [\underline{e}_1 + \underline{d}_{min}], & (8) \\ &= \lambda^2 \underline{e}_1^T \left[I + \lambda F_Y\right]^{-1} F_Y \left[I + \lambda F_Y\right]^{-1} \underline{e}_1 & (9) \end{aligned}$$

where $\underline{e}_1 = [1, 0, ..., 0]^T$ is an n-element unit vector and:

$$\underline{d}_{min} = -[I + \lambda\, F_Y]^{-1} \underline{e}_1, \tag{10}$$

and λ is given by the unique non-negative solution of the following equation involving the monotone decreasing, strictly convex function $g(\lambda) \in [0, 1]$:

$$g(\lambda) = \underline{d}_{min}^T \underline{d}_{min} = \delta^2 \qquad \lambda \geq 0. \tag{11}$$

A more general version of Theorem 1, which will not be required here, is given in [4] and applies to singular F_Y. Note that the use of the expression (9) does not suffer from any ill-conditioning of the FIM F_Y. In Theorem 1, $\underline{d}_{min}$ defined in (10) is an optimal bias gradient in the sense that it minimizes the biased CR bound (2) over all vectors $\nabla_{\underline{\theta}} b_1$.

Figure 3 shows a typical bias-variance trade-off curve. The region above and including the curve is the so called 'achievable' region where all the realizable estimators exist. Note that if an estimator lies on the curve then lower variance can only be bought at the price of increased bias and vice versa. At $\delta = 1$ the variance goes to zero. This corresponds to the trivial case $\widehat{\theta}_1$=Constant for which $\nabla_{\underline{\theta}} b_1 = \underline{e}_1$.

3.4 Recipe for Computing the Uniform Bound

Although the uniform bound is specified in terms of δ, it is simpler to sweep out the bound in terms of λ. Note that $g(\lambda) = \delta^2$ specifies a one-to-one map from $\lambda \in [0, \infty)$ to $\delta \in [0, 1]$ since $g(\lambda)$ is monotone over $\lambda \geq 0$. For each $\lambda \geq 0$ one can easily determine the corresponding δ by using the relation (11).

Often we want to compute the bound at several different values of λ and therefore it is important to have a computationally efficient algorithm to compute (10). Furthermore, since we are mainly concerned with problems which have a large number of unknown parameters, it may not be practical to compute F_Y^{-1} or F_Y^{+} explicitly both due to computation time and as well as memory requirements. We use the method of conjugate gradient as presented in [14] for this purpose and give a recipe for efficiently calculating the uniform CR bound (8) with non-singular F_Y. The conjugate gradient (CG) can be implemented to solve the linear system of equations:

$$F_Y \, \underline{u} = \underline{w}. \tag{12}$$

The CG algorithm is an iterative method, but converges to the true solution in at most n iterations when run with infinite precision arithmetic. If we substitute $\underline{w} = \underline{e}_1 = [1, 0, 0, ..., 0]^T$ in (12), then it is easily seen that the solution to such a system of equations will be the first column of F_Y^{-1}. The CG algorithm given in Appendix B is only applicable to symmetric, positive definite matrices F_Y [9]. Additional computational savings can be obtained by exploiting any sparseness of F_Y.

3.4.1 Recipe

Suppose we want to compute the uniform CR bound on the k-th estimator component $\widehat{\theta}_1$ of an $n \times 1$ estimator vector $\underline{\widehat{\theta}} = [\widehat{\theta}_1, \ldots, \widehat{\theta}_n]^T$. Then the numerically stable form of the uniform CR bound (9) for non-singular F_Y can be computed as follows.

1. Interchange 1-st and k-th row and column of F_Y.
2. Choose $\lambda \in [0, \infty)$.
3. Compute $\underline{t} = [I + \lambda F_Y]^{-1}\, \underline{e}_1$ by applying the CG algorithm to solve: $[I + \lambda F_Y]\, \underline{t} = \underline{e}_1$.
4. Compute the bound (9): $B(\underline{\theta}, \delta) = \lambda^2 \underline{t}^T F_Y \underline{t}$, $\delta = ||\underline{t}||$.

For each λ the above algorithm requires only one application of the conjugate gradient algorithm. Applying this algorithm for several values of λ allows one to trace the curve $B(\underline{\theta}, \delta)$ in the δ–σ trade-off plane of Figure 3.

3.5 Estimation of the Bias Gradient

To compare a particular estimator to the uniform bound of Theorem 1 we require the length of the estimator bias gradient so that the estimator can be placed somewhere within the achievable region of Figure 3. In most cases the bias and the bias-gradient are analytically intractable. The method of moments is the standard method for experimentally determining bias and covariance of $\widehat{\underline{\theta}}$ which is based on forming the sample mean and sample covariance statistics for a sequence of L repeated experiments $\{\underline{Y}_i\}_{i=1}^{L}$ each generated from the density $f_{\underline{Y}_i}(\underline{y}_i; \underline{\theta})$. The method of moments for estimating the bias-gradient would require p additional sequences of L repeated experiments, each generated for a particular perturbation of a different component of the parameter vector $\underline{\theta}$. Such a direct approach is impractical. In [4] a method for experimentally determining the bias-gradient of an estimator $\widehat{\underline{\theta}}$ is presented that requires a single simulation of the same type as that commonly used to determine bias and covariance of $\widehat{\underline{\theta}}$. The unbiased estimate of the bias gradient for the estimate of $\widehat{\theta}_1$ is given by:

$$\widehat{\nabla_{\underline{\theta}} b_1} = \frac{1}{L} \sum_{i=1}^{L} (\hat{\theta}_1(\underline{Y}_i) - \zeta_i)\ \nabla_{\underline{\theta}} \ln f_{\underline{Y}_i}(\underline{y}_i; \underline{\theta}) - \underline{e}_1^T, \tag{13}$$

where $\zeta_i : i = 1, ..., n$ are the ancillary statistics i.e. unbiased estimators of zero [16]. It is shown in [4] that the following choice of ζ_i minimizes the asymptotic covariance of the bias gradient estimate:

$$\zeta_i = \frac{1}{L-1} \sum_{\substack{j=1 \\ j \neq i}}^{L} \hat{\theta}_1(\underline{Y}_j). \tag{14}$$

Substituting (14) in (13) yields a simplified form of the bias gradient estimate [4]:

$$\widehat{\nabla_{\underline{\theta}} b_1} = \frac{1}{L-1} \sum_{i=1}^{L} \left(\widehat{\theta}_1(\underline{Y}_i) - \sum_{j=1}^{L} \widehat{\underline{\theta}}(\underline{Y}_j) \right) \nabla_{\underline{\theta}} \ln f_{\underline{Y}}(\underline{Y}_i; \underline{\theta}) - \underline{e}_1^T .$$

In particular it is easy to show that for the Poisson model

$$\nabla_{\underline{\theta}} \ln f_{\underline{Y}}(\underline{y}; \underline{\theta}) = A^T \left[-\underline{1} + \underline{y} \oslash \underline{\mu} \right] ,$$

where $\oslash$ is a vector operation denoting element-by-element division, and $\underline{1} = [1, 1, ..., 1]^T$.

A few comments about the bias gradient are in order. The bias gradient $\nabla_{\underline{\theta}} b_1$ is a measure of the influence of each component parameter $\theta_1, ..., \theta_n$ on the mean $m_1(\underline{\theta})$ of the estimator $\widehat{\theta}_1$. Ideally, to be close to unbiased one would like $m_1(\underline{\theta})$ to be insensitive to the variations in the other parameters $\theta_2, ..., \theta_n$. Alternatively, since $b_1(\underline{\theta}) = m_1(\underline{\theta}) - \theta_1$, it is desirable that the components $\frac{\partial}{\partial \theta_k} b_1(\underline{\theta})$ be of small magnitude. The bias gradient therefore provides important information about the parameter coupling to the estimator mean. The bias gradient is in general only indirectly related to the estimator bias, with the exception that $\nabla_{\underline{\theta}} b_1 = 0$ implies $b_1(\underline{\theta}) =$ constant. An estimator that has a constant bias independent of any estimator parameter is removable, and therefore $\nabla_{\underline{\theta}} b_1 = 0$ implies that the estimation can be performed without bias. Conversely, a non-zero bias gradient implies non-removable estimator bias that is dependent on the estimator parameters. On the other hand, one can have a large bias gradient even though the bias is very small. Therefore the bias and the bias gradient together give a more complete picture of estimator behavior.

3.6 Bias-Variance Trade-Off Plane

When accurate estimates $\widehat{b}_1$, $\widehat{\nabla_{\underline{\theta}} b_1}$ and $\widehat{\sigma}^2$ of the estimator bias, bias gradient, and variance are available for a given estimator $\widehat{\theta}_1$ of θ_1, the uniform CR bound lying in the δ–σ plane can be easily mapped into the b–σ plane of variance and biases. This is accomplished by using the ordered triplet $(\widehat{b}_1, \widehat{\nabla_{\underline{\theta}} b_1}, \widehat{\sigma}^2)$ as a mapping between the δ–σ and the b–σ planes. The uniform CR bound on the variance as a function of bias is simply the ordered pair: $\left(\widehat{b}_1, \left[\underline{e}_1 + \widehat{\nabla^T_{\underline{\theta}} b_1} \right]^T F_Y^+ \left[\underline{e}_1 + \widehat{\nabla^T_{\underline{\theta}} b_1} \right] \right)$, denoted $B(\underline{\theta}; b)$ in the sequel.

4 Applications to Standard SPRINT

The object used in the first two applications consists of two point sources within a disk of uniform intensity of radius 16 pixels (Figure 4). The high intensity white pixels have a normalized intensity value of 2, while the low intensity black pixels are set to 1. For the uniform CR bound computation the pixel of interest was the high intensity pixel at the top of the image marked '1'. Noise due to scatter was neglected in these studies.

It can be shown that for SPECT, the Fisher information matrix has the form [5]:

$$F_Y = A^T \, [\mathrm{diag}(\underline{\mu})]^{-1} \, A,$$

where A is a $d \times n$, $n \geq d$, weight matrix, $\underline{\mu} = A \, \underline{\theta} + \underline{s}$, and F_Y is a $n \times n$ symmetric positive definite matrix. We will assume that A is full rank so that $F_Y > 0$. The system used in this study is the SPRINT II system, without the vertex view, as described in Section 1.1.

4.1 Spatial Sampling Study

The effect of attenuation was neglected in this experiment. In Figure 5 the bias-variance curves are displayed for varying amounts of spatial sampling. Spatial sampling was varied by rotating the SPRINT collimator ring through different numbers of equal angle increments over $[0, \pi]$ radians. The time for each step was kept constant. Note the monotonic nature of the curves as a function of the number of rotation increments. The curves do not intersect each other due to the fact that the projections from the lower sampled image are a subset of those of the higher sampled image. Also it can be seen that beyond 8 rotation increments increasing sampling rate does not reduce the bound significantly. At this stage it is not clear how this reduction in the uniform CR bound relates to the image quality.

An interesting point to note is that the unbiased CR bound ($\nabla_{\underline{\theta}} b_1 = 0$) for 10 rotations is greater than the bound at a bias gradient length of 0.5 for only 5 rotations. This means that if the bias gradient length $\delta = 0.5$ is acceptable, then a biased algorithm with lower sampling can perform better than an unbiased algorithm with higher sampling.

For a time normalized case, when the total scan time for each curve is kept constant, the curves could intersect each other (Figure 6). They, however, display the same trend as above. Since in this case the total sensitivity of the system is constant, any reduction in the bound in Figure 6 is due to a higher resolution.

Note that in experimental setting, for a 64×64 image, 54 uniform steps did not produce a similar saturation behavior. Further increase in resolution was observed by using non-uniform angular rotations for the collimator ring.

4.2 Optimal System Design

In this experiment we apply the uniform CR bound (8) to standard SPRINT in order to determine an aperture opening that optimizes the trade-off between resolution and sensitivity. The effect of attenuation was not included in the FIM F_Y. For these simulations the aperture opening was varied from narrow (ray width = 0.25 pixels) to wide (ray width = 30 pixels). The imaging time was adjusted so that the total number of detected counts are the same for all cases; hence smaller exposure time for wider openings. The bias-variance trade-offs are displayed in Figure 7. Remarkably the variance shows a sharp minimum over all δ when the width of the ray is approximately one pixel at the pixel of interest. Therefore for the object and the ROI studied, this one pixel aperture width is 'universally' optimal for estimation, irrespective of the bias gradient length of the estimator.

For a time normalized case, where the total imaging time is kept constant, we expect the minimum to be shallower as compared to the count normalized case. This is due to normalization of the uniform CR bound with higher counts for wider collimator openings, which has an effect of reducing the uniform CR bound. Note that the variance is normalized by the reciprocal of the number of detected photons.

4.3 Performance Comparison Between Image Reconstruction Algorithms

We will apply the uniform CR bound to study the bias-variance trade-offs for a particular class of roughness penalized space alternating generalized expectation-maximization

(PML-SAGE) algorithm and penalized weighted least-squares (WLSE) image reconstruction algorithms. Since the penalty introduces bias in the estimator due to image smoothing, the standard unbiased CR bound is inapplicable in this study.

For the simulations in this section we used the standard SPRINT II ring geometry without the vertex view. The effect of attenuation was neglected. The total number of detected γ-ray counts were 10^9. Noise due to scatter were 5% of the total counts. In all the following cases, the algorithm was initialized by a uniform disk of intensity 1 and radius 16 pixels. Since both the algorithms considered in this section are non-linear, an analytic expression for the bias gradient is intractable, and therefore the bias gradient was estimated using (13). We used $L = 400$ realizations of the projection data $\underline{Y}$ to ensure the statistical accuracy of our estimator bias, bias gradient and variance.

The object is a disk of uniform intensity 1 with a high intensity region of 4 pixels in the center of uniform intensity 2, called the hot spot. The pixel of interest was the pixel at the upper edge of the hot spot, marked '1'. The diameter of the disk is 32 pixels.

4.3.1 Penalized Maximum Likelihood

The iterative space alternating generalized expectation-maximization (PML-SAGE) algorithm maximizes a penalized likelihood objective function of the form:

$$\hat{\underline{\theta}}(\underline{y}) = \underset{\underline{\theta} \in \Theta}{\text{argmax}} \left\{ \ln f_{\underline{Y}}(\underline{y}; \underline{\theta}) - \alpha P(\underline{\theta}) \right\},$$

where $P(\underline{\theta})$ is the penalty function and α is the smoothing parameter. Setting $\alpha = 0$ corresponds to no image smoothing while a large value of α corresponds to a significant amount of smoothing. PML-SAGE uses an intelligent choice of 'hidden data spaces' such that the E and M steps are analytically tractable. A detailed description of the PML-SAGE algorithm is given in [12].

For the first set of simulations the smoothing parameter α was varied (Figure 9 (a)). Points on the curves in Figures 9 (a) and (b) are labeled by the exponent of α. The bias,

bias gradient and variance were estimated and the uniform bound was plotted over the bias gradient length-variance trade-off plane, denoted δ–σ, and the bias-variance trade-off plane, denoted b–σ. The PML-SAGE algorithms were terminated after 100 iterations for each of the $L = 400$ trials. The 95% ellipsoidal confidence regions are not shown in the figure since they are smaller than the size of the plotting symbol '$*$'. Note that the bound, denoted by $B(\underline{\theta}; \delta)$ in Figure 9 (a), is achieved for large biases, i.e. large α. For α small, the curve 'B' tends to deviate more from the lower bound and saturate, i.e. lower α does not decrease the bias gradient. On the other hand the bias decreases to an asymptote near zero.

At points close to the unbiased point, i.e. the leftmost corner of the horizontal axis, in curve 'A', maximal reduction in bias is achieved at the price of significant increase in the variance. An interesting question to ask is: what is the best choice of α in this case? Unfortunately there is not a simple answer to this question since it is application dependent. However, using the two plots in Figure 9 (a), the bound on the variance as a function of bias can be easily translated into a bound on the estimator MSE, denoted $B(\underline{\theta}, \alpha) + b_1^2$, as a function of the smoothing parameter α. The bound, along with the estimator MSE is shown in Figure 9 (c). The point of least MSE is the point of maximum variance but least smoothing. This supports the widespread opinion that when taken in isolation the pixel intensity MSE is not suitable as an image quality metric. However, a suitable trade-off between the smoothing, which directly corresponds to the image bias, and the variance is to choose a point close to the knee of the MSE curve 'A' in Figure 9 (c), corresponding to $\alpha = 2^3$, or $\alpha = 2^4$. This choice of α ensures that the reconstructed image is not overly smoothed, at the same time not having an unacceptably large variance.

Figures 10 and 11 show several graphs of reconstruction quantities for $\alpha = 2^4$, and $\alpha = 2^{10}$, respectively. For clarity in the figures, we down-sampled all the images by a factor of 2. For each image in Figures 10 and 11 the ordered pair at bottom indicates the minimum and maximum values for that image. In Figure 10, the mean reconstructed image is very close to the true image except around the edges. The correlation image, i.e. the column of F_Y^{-1} corresponding to the pixel of interest, θ_{ROI}, shows a strong correlation with the neighboring pixels. This implies that to estimate θ_{ROI} we must also estimate the strongly correlated neighboring pixels accurately, while the influence of the far pixels can be ignored. Ideally, one would like the correlation between the pixels to be zero so that the estimate of a certain pixel, θ_{ROI}, is independent of the estimates of all other pixels. The plot for the theoretically optimal bias gradient $\underline{d}_{min}$ shows a similar strong influence from the neighboring pixels.

The average bias gradient $\nabla_{\underline{\theta}} b_1$ for the reconstructed image is different from the theoretically optimal bias gradient $\underline{d}_{min}$. Thus the PML-SAGE image reconstruction algorithm does not take best advantage of its bias allocation since it is only by using the optimal bias gradient $\underline{d}_{min}$ given by (10) that the minimum bias length is achieved.

Figure 11 shows the same set of images as in Figure 10 but for $\alpha = 2^{10}$. Due to very high regularization, the hot spot is almost entirely smoothed out. Also, neither $\underline{d}_{min}$ nor the average bias gradient $\nabla_{\underline{\theta}} b_1$ for the reconstructed image show significant coupling between the pixel of interest and the neighboring pixels. This is to be expected since in the overly smoothed case the bias is principally determined by the smoothness penalty as opposed to the projection data.

4.3.2 Weighted Least-Squares Estimator

Similar to the PML-SAGE, the WLSE is penalized for roughness, but minimized over a quadratic objective function. The WLSE is given by [6]:

$$\underline{\widehat{\theta}}(\underline{y}) = \underset{\underline{\theta} \in \Theta,\ \underline{\theta} \geq 0}{\operatorname{argmin}} \left\{ \frac{1}{2} \left[(\underline{\widehat{y}} - A\underline{\theta})^T \Sigma^{-1} (\underline{\widehat{y}} - A\underline{\theta}) \right] + \alpha P(\underline{\theta}) \right\},$$

where Σ is a weight matrix, $P(\underline{\theta})$ is a regularization penalty, and A is the system matrix. We use a penalty function described in [6] which is imposed on the 8 neighboring pixels for each pixel of interest. The weight matrix Σ is diagonal, consisting of the covariance estimate of the observations. It is shown in [4, 15] that a WLSE with an identity penalty function and ideal weight matrix $\Sigma = \operatorname{diag}_i(\mu_i)$ exactly achieves the uniform CR bound for all biases.

Figure 9 (b) shows the δ–σ and b–σ plots for the WLSE. The WLSE follows the uniform CR bound closely for high bias and low variance, but tends to deviate away from the bound for low biases. An interesting point to note is that the PML-SAGE and the WLSE have similar bias-variance trade-off curves. However, the uniform bound on bias $B(\underline{\theta}, b)$ is different for PML-SAGE than that for WLSE since the bound on bias is indexed by algorithm bias gradient which is obviously algorithm dependent.

Figures 12 and 13 show several graphs of reconstruction quantities using the WLSE for $\alpha = 2^4$ and $\alpha = 2^{10}$. The comments for Figures 10 and 11 are valid here. The only exception being that the WLSE fails to accurately estimate the edges for small α. This is due to the fact that the estimates of covariance involving the projections that graze the image edges are less accurate.

The PML-SAGE and the WLSE have very similar bias gradient-variance and bias-variance trade-offs, and based on these simulations one can not prefer one algorithm over the other. However, the PML-SAGE takes longer to converge when compares to the WLSE.

A System Specifications

Ring Detector

The SPRINT parameters are given in Table 1. The SPRINT ring has a resolution of 13 mm at the center of the ring.

B Inversion Algorithm

We want to solve the linear system of equations:

$$(I + \lambda F_Y)\, \underline{u} = \underline{e}_1. \tag{15}$$

The conjugate gradient algorithm is given as follows [14].

CG Algorithm

INITIALIZE: $\underline{u}^{(0)} = \underline{0}$; $\underline{r}^{(0)} = \underline{e}_1$

RECURSION:

$$\begin{aligned}
\alpha^{(i)} &= \begin{cases} 0 & i = 0 \\ \frac{<\underline{r}^{(i)}, \underline{r}^{(i)}>}{<\underline{r}^{(i-1)}, \underline{r}^{(i-1)}>} & i > 0 \end{cases} \\
\underline{p}^{(i)} &= \underline{r}^{(i)} + \alpha^{(i)}\, \underline{p}^{(i-1)} \\
\lambda^{(i)} &= \frac{< \underline{r}^{(i)}, \underline{r}^{(i)} >}{< \underline{p}^{(i)}, (I + \lambda F_Y)\, \underline{p}^{(i)} >} \\
\underline{u}^{(i+1)} &= \underline{u}^{(i)} + \lambda^{(i)}\, \underline{p}^{(i)} \\
\underline{r}^{(i+1)} &= \underline{r}^{(i)} - \lambda^{(i)}\, (I + \lambda F_Y)\, \underline{p}^{(i)}
\end{aligned}$$

BOUND APPROXIMATION:

$$\eta^{(i)} = \underline{e}_1^T\, \underline{u}^{(i)}. \tag{16}$$

The iterations are terminated when the residual error $\|r^{(i)}\| = \|\underline{e}_1 - (I + \lambda F_Y)\underline{u}^{(i)}\|$ is less than a user specified tolerance.

References

[1] R.D. Fiete, H.H. Barrett, Fiete RD, W.E. Smith and K.J. Myers *Hotelling trace criterion and its correlation with human-observer performance.* Journal of Optical Society of America, 4:5, pp 945-953, May 1987.

[2] W.L. Rogers, N.H. Clinthorne, L.Shao, P.Chiao, J.Stamos, and K.F.Koral *SPRINT II, A Second Genration Single Photon Ring Tomograph,* IEEE Transactions on Medical Imaging, 7:4, pp. 291-297, 1988.

[3] L. Shao, A.O. Hero, W.L. Rogers and N.H. Clinthorne *The Mutual Information Criterion for SPECT Aperture Evaluation Design*, IEEE Transactions on Medical Imaging, 8:4, pp 322-336, Dec. 1989.

[4] M. Usman, A.O. Hero, and J.A. Fessler *Bias-Variance Trade-offs For Parametric Estimation Problems Using Uniform CR Bound*, Submitted to IEEE Transactiosn on Signal Processing.

[5] A.O. Hero and J.A. Fessler
A Recursive Algorithm for Computing CR-Type Bounds on Estimator Covariance, to appear in IEEE Transactions on Information Theory, 1994.

[6] J. A. Fessler
Penalized Weighted Least-Squares Image Reconstruction for Positron Emission Tomography, IEEE Transactions on Medical Imaging, 13:2, pp 290-300, June 1994.

[7] A. O. Hero
A Cramer-Rao Type Lower Bound for Essentially Unbiased parameter Estimation, Technical Report 890, Lincoln Laboratory, MIT 1992.

[8] G.H. Golub and C.F. van Loan
Matrix Computations (Second Edition), The Johns Hopkins University Press, 1989.

[9] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B.P. Flannery
Numerical Recipes in FORTRAN, The Art of Scientific Computing, (Second Edition), Cambridge University Press, 1992.

[10] H.L. van Trees
Detection, Estimation and Modulation Theory, (Part I), John Wiley and Sons. 1968.

[11] J. A. Fessler and A. O. Hero
Cramer-Rao Bounds for Biased Estimators in Image Reconstruction, Midwest Symposium on Circuits and Systems, pp 253-256, August 1993, Detroit, MI.

[12] J.A. Fessler and A.O. Hero
Space-Alternating Generalized Expectation-Maximization Algorithm, To appear in IEEE Transactions on Signal Processing, Oct., 1994.

[13] M. Usman, A. O. Hero and W. L. Rogers
Performance Gain Analysis for Adding Vertex View to a Standard SPECT, Midwest Symposium on Circuits and Systems, pp 418-421, August 1993, Detroit, MI.

[14] M. Usman and A.O. Hero
Recursive Algorithms for Computing CR Bounds, Submitted to IEEE Transactions on Signal Processing.

[15] M. Usman, A. O. Hero and J.A. Fessler
Bias-Variance Trade-offs Analysis Using Uniform CR Bound, IEEE Medical Imaging conference, San Francisco, Nov. 1993.

[16] E. L. Lehmann
Testing Statistical Hypothesis, 2nd Edition, pp 542-553, John Wiley and Sons, New York, 1986.

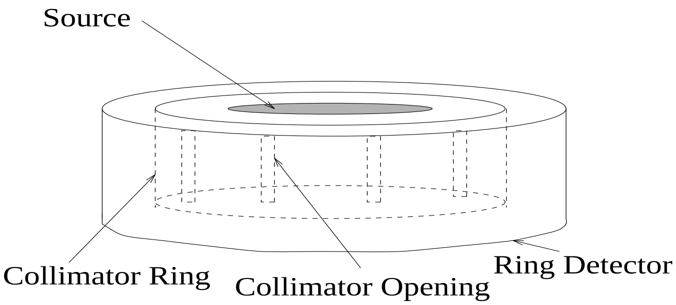


Figure 1: The SPRINT system. Not drawn to scale.

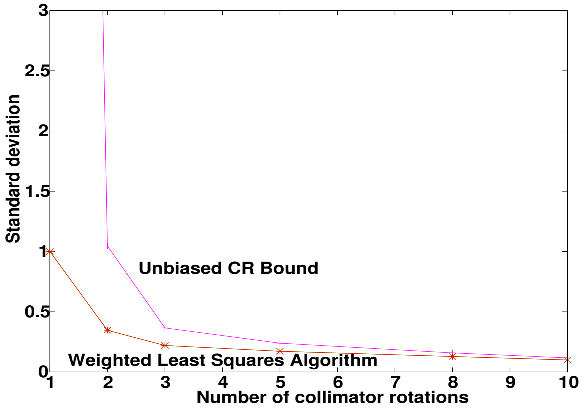


Figure 2: The regularized weighted least-squares estimator shows lower variance than the unbiased CR bound.

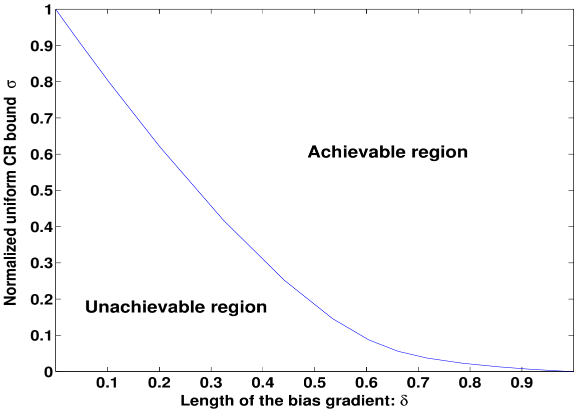


Figure 3: The Normalized Uniform CR bound.

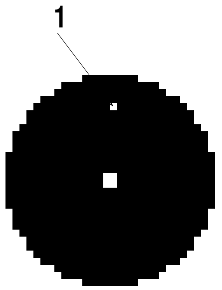


Figure 4: The object used in the simulations. The pixel of interest is the pixel at the top of the object. The image dimensions are 32x32 pixels.

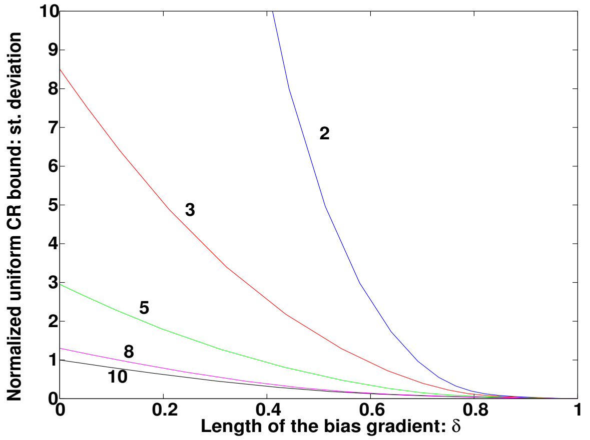


Figure 5: The effect of angular sampling on the unbiased CR bound. Curves denote lower bound for 2,3,5,8 and 10 rotation increments of the collimator.

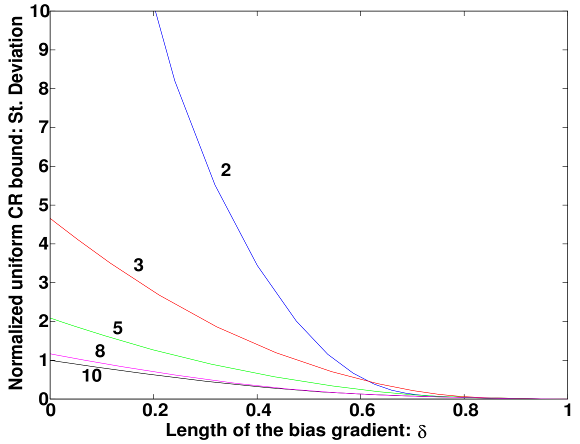


Figure 6: The effect of angular sampling on the unbiased CR bound. Time normalized case.

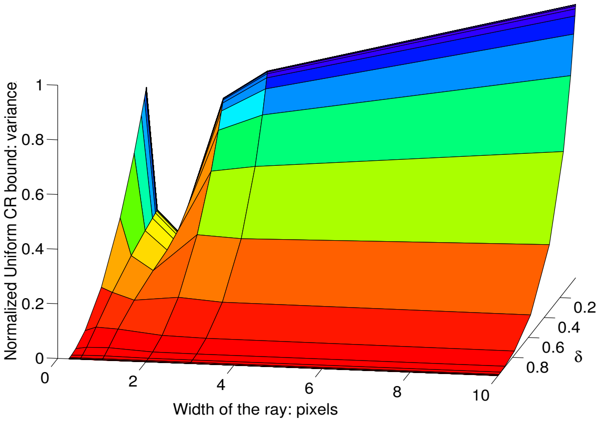


Figure 7: Optimal aperture design using the uniform CR bound. One pixel width shows minimum uniform bound.

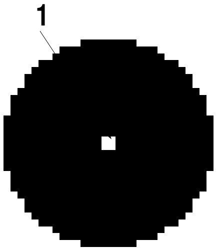
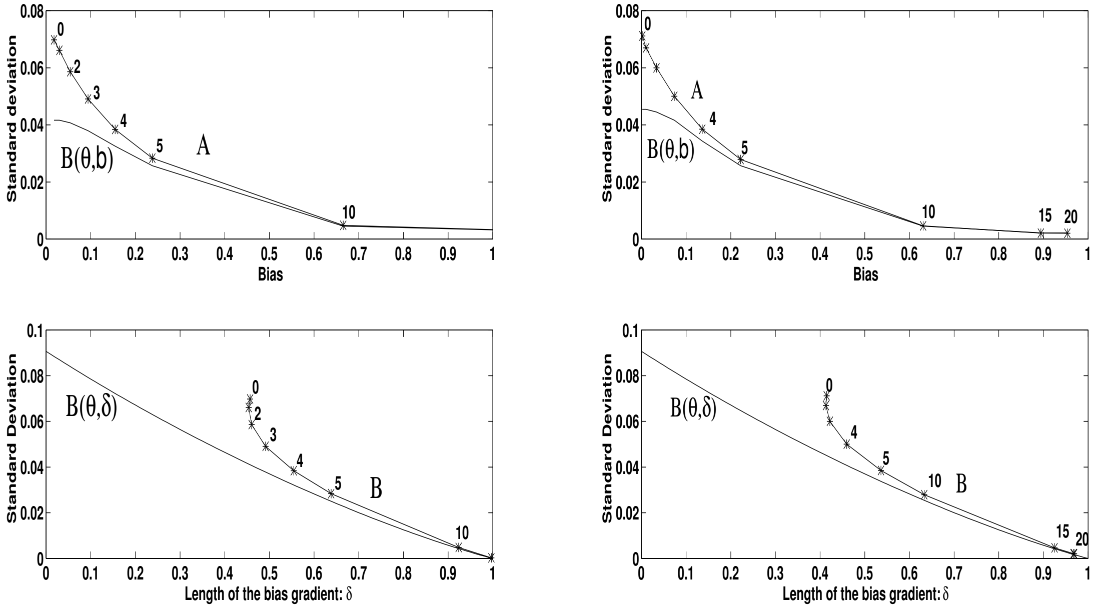
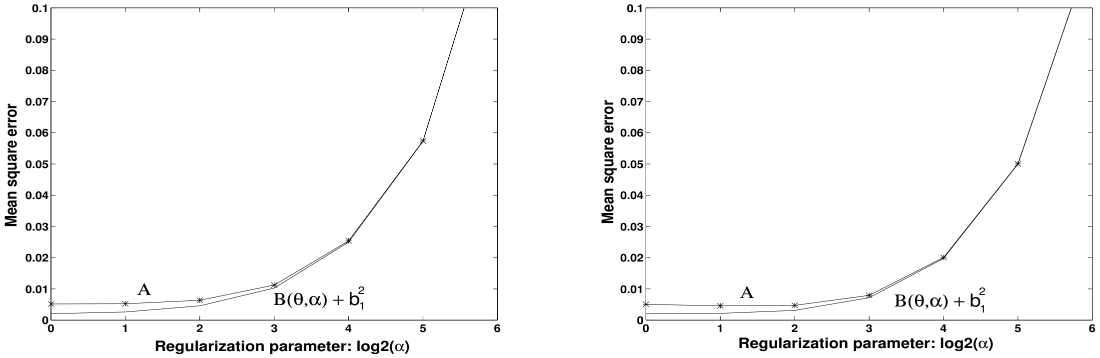


Figure 8: The object used in the simulations. The object dimensions are 32 × 32. The black pixels are of intensity 1 while the white pixels are of intensity 2.

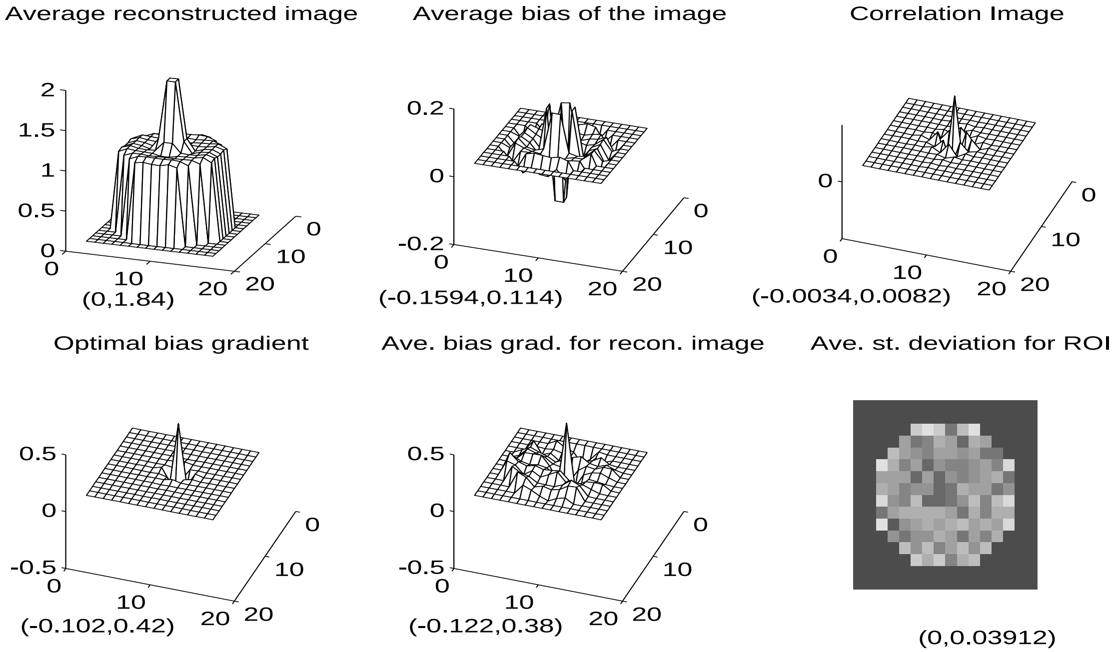


(a) PML-SAGE (b) WLSE



(c) MSE: PML-SAGE (d) MSE: WLSE

Figure 9: Comparison of the PML-SAGE and WLSE



Average reconstructed image

Average bias of the image

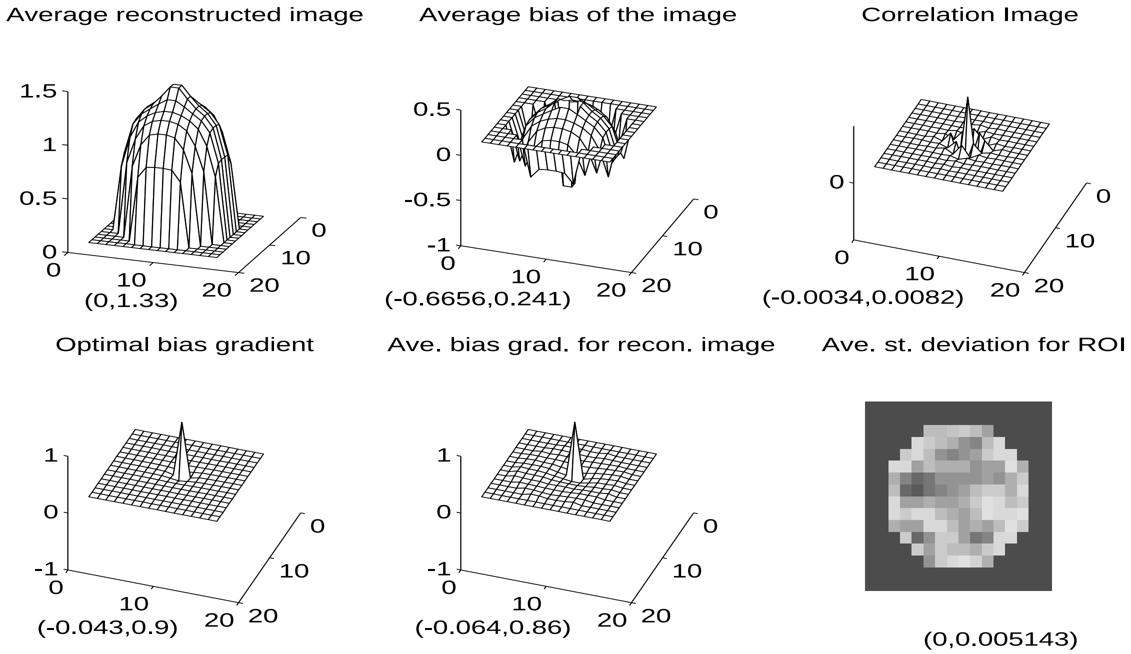
Correlation Image

Optimal bias gradient

Ave. bias grad. for recon. image

Ave. st. deviation for ROI

Figure 10: PML-SAGE: different graphs of reconstruction quantities for $\log_2(\alpha) = 4$. An ordered pair with each curve indicate the (minimum, maximum) value associated with that image.



Average reconstructed image

Average bias of the image

Correlation Image

Optimal bias gradient

Ave. bias grad. for recon. image

Ave. st. deviation for ROI

Figure 11: PML-SAGE: different graphs of reconstruction quantities for $\log_2(\alpha) = 10$.

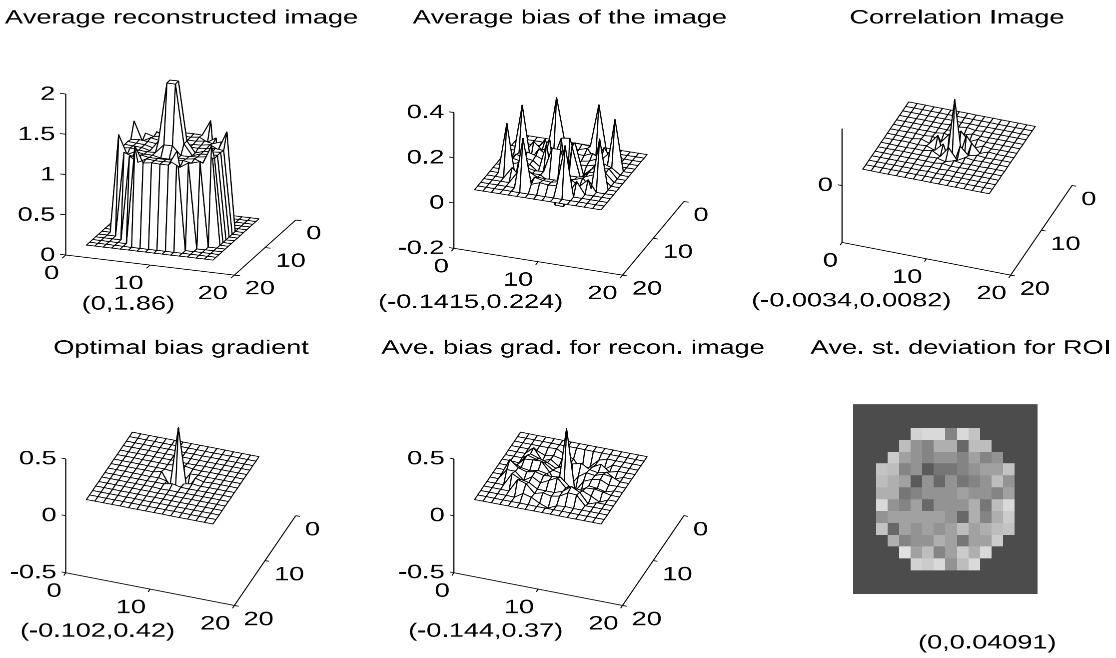


Figure 12: WLSE: different graphs of reconstruction quantities for $\log_2(\alpha) = 4$.

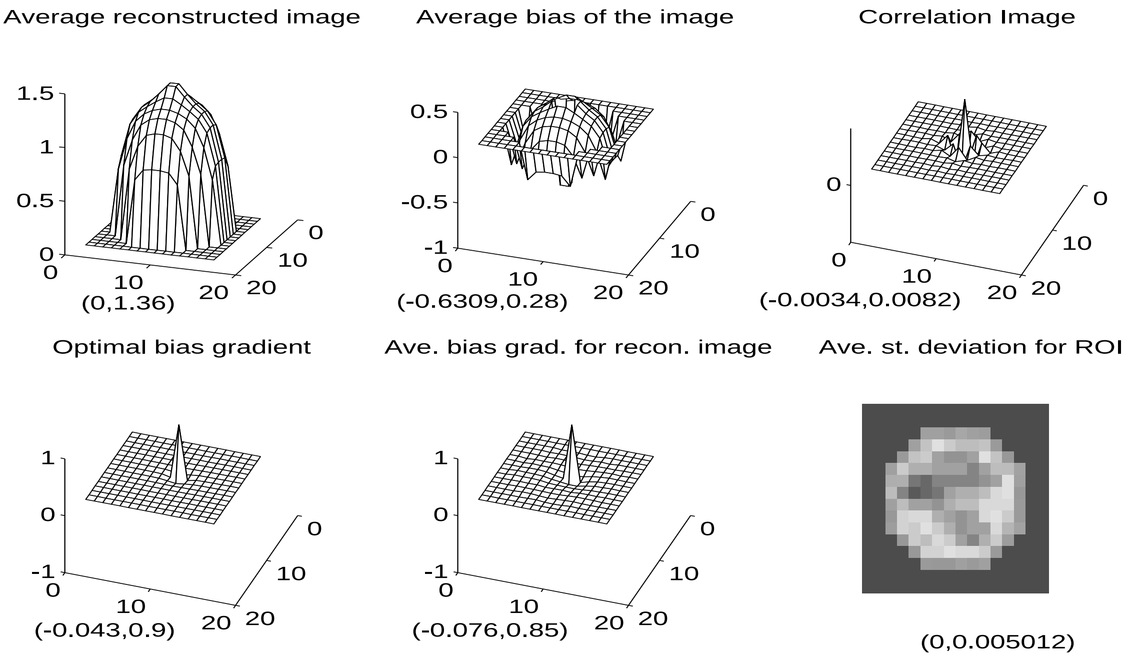


Figure 13: WLSE: different graphs of reconstruction quantities for $\log_2(\alpha) = 10$.

Radius of the detector ring	25 cms
Number of detectors	512
Radius of the collimator ring	17 cms
Number of collimator slits	10 (uniformly spaced)
Slit Width	2.4 mm

Table 1: SPRINT ring specifications

List of Figures

List of Tables