Bias-Variance Tradeoffs Analysis Using Uniform CR Bound For a SPECT System

Mohammad Usman, Alfred O. Hero, Jeffrey A. Fessler and W. L. Rogers
University of Michigan

Abstract

We quantify fundamental bias-variance tradeoffs for the image reconstruction problem in radio-pharmaceutical tomography using Cramer-Rao (CR) bound analysis. The image reconstruction problem is very often biased and the classical or the unbiased CR bound on the mean square error performance of the estimator can not be used. We use a recently developed 'uniform' CR bound which applies to biased estimators whose bias gradient satisfies a user specified length constraint. We demonstrate the use of the 'uniform' CR bound for a simple SPECT system using several different examples.

I. Introduction

As in any parametric estimation problem, in image reconstruction there always exists a tradeoff between bias and variance of the reconstructed image. For example in penalized maximum likelihood reconstruction a smoothness penalty reduces the variance at the expense of introducing bias. This paper provides an estimator independent fundamental tradeoff inherent in emission tomography. In [6, 9] classical CR bound was applied to unbiased image reconstruction in emission tomography. Very often in tomographic imaging problems an unbiased image reconstruction with a smoothness constraint may results in large variance. For biased image reconstruction classical biased CR bound can be used [3]. However the biased CR bound is not very useful in that it is only applicable to the estimators with a given bias gradient $\nabla_{\underline{\theta}} b(\underline{\theta})$. Therefore it is unable to give a meaningful comparison of different biased estimators that may have acceptable bias but different bias gradients. Hero [2] gives a 'uniform' CR bound that is applicable to all the estimators whose bias gradient length is less than a prespecified threshold, permitting a meaningful comparison for different biased estimators to be performed.

For images, the calculation of the CR bound involves inversion of a huge $n \times n$ Fisher Information Matrix (FIM), where n is the total number of image pixels. Direct inversion algorithms require $O(n^3)$ flops and can become computationally intractable. The system matrix in image reconstruction problem is in general sparse. Here, we give an inversion algorithm based on conjugate gradient approach that exploits the sparsity of the system matrix and can compute only a few columns of the inverse of FIM corresponding to pixels of interest.

We apply the methods of [2] to a simple SPECT system for several imaging examples. The examples include determining the optimal collimator opening as a function of the tradeoff between resolution and sensitivity; deriving bias-variance tradeoffs curves for different angular sampling; and performance comparison of different unbiased and biased image reconstruction algorithms using the uniform CR bound.

A. System Description

The system used in this study is shown in Figure 1, called the SPRINT II system [1]. SPRINT II was designed specifically for brain imaging and consists of a ring of detectors and a ring of collimators. During the experiment the ring of collimators are rotated through small steps for adequate spatial sampling. The system parameters are given in the appendix and unless otherwise specified are those used in the simulations.

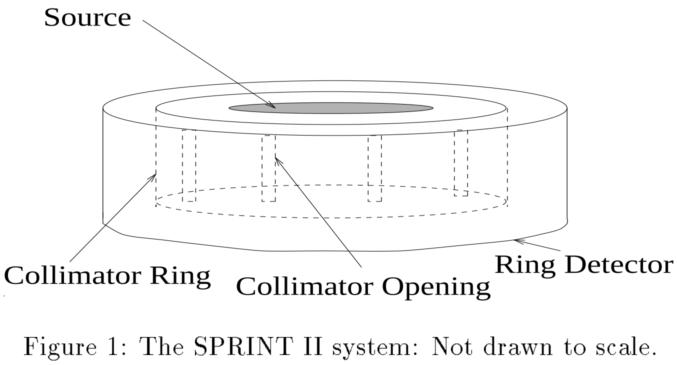

Figure 1: The SPRINT II system: Not drawn to scale.

II. CR bound

Given the projection data, an n-pixel image and a probability density $f_Y(y; \underline{\theta})$, the objective is to estimate the intensity in a small Region Of Interest (ROI) $\underline{\theta}_{ROI} = [\theta_1, \theta_2, ..., \theta_q]^T$. In the sequel we specialize to the case when the ROI consists of a single pixel θ_1. A function $\hat{\theta}_1(y)$ is a parameter estimator based on the observation $Y = y$. $\hat{\theta}_1$ is unbiased if the bias $b_1(\underline{\theta}) = E[\hat{\theta}_1 - \theta_1]$ is equal to zero.

This work was supported in part by National Science Foundation under grant BCS-9024370 and a Government of Pakistan Postgraduate Fellowship.

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

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

where,

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

and $\nabla_{\underline{\theta}}$ denotes the (row) gradient vector $[\frac{\partial}{\partial\theta_1}, \frac{\partial}{\partial\theta_2}, ..., \frac{\partial}{\partial\theta_n}]$ and $\underline{e}_1 = [1, 0, ..., 0]^T$.

While the unbiased CR bound (1) is known to be asymptotically achievable for large imaging times, in practice, most of the popular image reconstruction algorithms are biased and the unbiased CR bound is inapplicable.

A. Motivating Example

In Figure 2 the unbiased CR bound, for a high intensity pixel at the top of the image in Figure 4, is plotted as a function of the collimator rotations. Also plotted is the variance of the weighted least square algorithm for the same pixel. The variance of the weighted least square algorithm (section C.2) 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. The unbiased CR bound can not be used for performance analysis in this case.

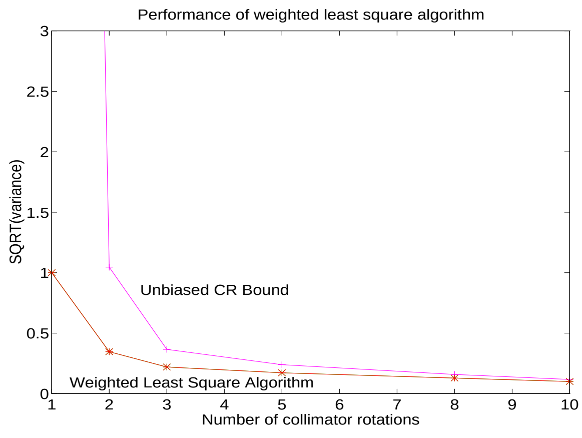

Figure 2: The regularized weighted least square estimator shows lower variance than unbiased CR bound at all the points.

III. Biased CR bound

For a biased estimator $\hat{\theta}_1$ the following form of the biased CR bound is well known [3]:

$$var(\hat{\theta}_1) \geq [\nabla_{\underline{\theta}} m_1(\underline{\theta})] \; F_Y^{-1}(\underline{\theta}) \; [\nabla_{\underline{\theta}} m_1(\underline{\theta})]^T, \qquad (2)$$

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

A. Uniform CR Bound

The application of the biased CR bound (2) is very restricted due to the fact that it is only applicable to estimators with a given bias gradient $\nabla_{\underline{\theta}} b_1(\underline{\theta})$. In [2] 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 length $\|\nabla_{\underline{\theta}} b_1(\underline{\theta})\|$ of the bias gradient $\nabla_{\underline{\theta}} b_1(\underline{\theta})$ is less than a small prespecified threshold:

$$\|\nabla b_1(\underline{\theta})\|^2 \leq \delta^2 < 1. \qquad (3\text{-a})$$

Then we have the following theorem [2].

Theorem 1 *Let θ_1 be an estimator with bias $b_1(\underline{\theta})$ whose n-element bias gradient vector $\nabla_{\underline{\theta}} b_1 = \underline{d}^T$ satisfies (3-a). Then the variance of θ_1 satisfies:*

$$var(\hat{\theta}_1) \geq B(\underline{\theta}, \delta), \qquad (3\text{-b})$$

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

$$B(\underline{\theta}, \delta) = [\underline{e}_1 + \underline{d}_{min}]^T F_Y^{-1} [\underline{e}_1 + \underline{d}_{min}]. \qquad (3\text{-c})$$

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

$$\underline{d}_{min} = -[I + \lambda F_Y]^{-1} \underline{e}_1, \qquad (3\text{-d})$$

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

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

In Theorem 1, $\underline{d}_{min}$ defined in (3-d) is an optimal bias gradient in the sense that it minimizes the biased CR bound (2) over all vectors $\nabla_{\underline{\theta}} b_1(\underline{\theta})$.

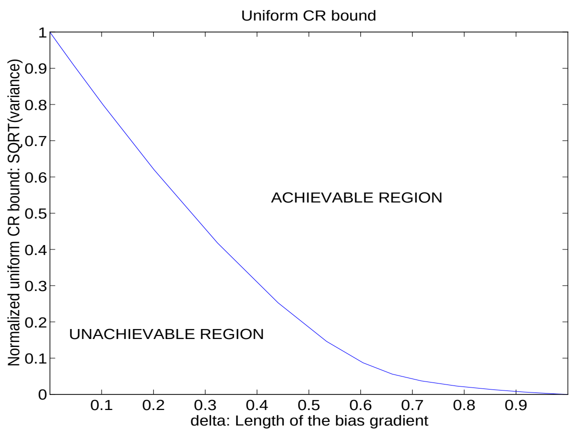

Figure 3: The Uniform CR bound

Figure 3 shows a typical bias-variance tradeoff curve. The curve is drawn for the same pixel as in Figure 2. 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 $\hat{\theta}_1$=Constant.

B. Inversion Algorithm

Equation (3-d) in the calculation of uniform CR bound requires computation of the first column of $[I + \lambda\, F_Y]^{-1}$ for each λ. 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 (3-d).

Conjugate gradient is a recursive method to solve the linear system of equation $C\,\underline{x} = \underline{b}$. If we substitute $\underline{b} = [1, 0, 0, ..., 0]^T$ then it is easy to recognize that the solution to such a system of equations will be the first column of C^{-1}. In general we can get the inverse of the m-th column of C if we choose $b_m = 1$ and $b_j = 0 : j \neq m$. The following algorithm is applicable to only symmetric, positive definite matrices C [4].

Algorithm

`INITIALIZATIONS:`

$\underline{u}^0 = \underline{0};\ \underline{b} = [1, 0, 0, ..., 0]^T;\ \underline{r}^0 = \underline{b}$

`FOR i := 0 UNTIL | C u - b | < tolerance DO`

`IF i = 0 DO`

$$\underline{p}^i = \underline{r}^i$$

`ELSE`

$$\begin{aligned} \alpha_i &= \frac{< \underline{r}^i, \underline{r}^i >}{< \underline{r}^{i-1}, \underline{r}^{i-1} >} \\ \underline{p}^i &= \underline{r}^i + \alpha_i\, \underline{p}^{i-1} \end{aligned}$$

`END IF`

$$\begin{aligned} \lambda_i &= \frac{< \underline{r}^i, \underline{r}^i >}{< \underline{p}^i, A\,\underline{p}^i >} \\ \underline{u}^{i+1} &= \underline{u}^i + \lambda_i\, \underline{p}^i \\ \underline{r}^{i+1} &= \underline{b} - C\,\underline{u}^{i+1} \end{aligned}$$

`END DO`

The iterations are terminated when the error $|\,C\,\underline{u} - \underline{b}\,|$ is less than a user specified tolerance. The convergence rate of the conjugate gradient depends on the condition number of C and its eigenvalue distribution. One advantage in using the conjugate gradient algorithm is that the algorithm is guaranteed to converge in maximum n iterations. Each iteration of the algorithm involves a matrix-vector multiplication $A\,\underline{p}^i$, requiring $O(n^2)$ flops and therefore the algorithm can take $O(n^3)$ flops to converge. For SPECT the system matrix in general is sparse (typically less than 5%) and, if the sparsity of A is taken into account, then the total number of flops can be reduced significantly.

In relation to the recursive algorithm for the unbiased CR bound given in [6], a disadvantage of this algorithm is its non-monotonic convergence. This means that if stopped at iteration $k < n$ then the quantity calculated by the algorithm might not be a valid lower bound on $var(\hat{\theta}_1)$. This, however, may not be a problem if we let the algorithm run until a desired accuracy is achieved.

IV. Applications

The object used in all the following simulations has two point sources in a uniform intensity background (figure 4). In all cases the pixel of interest was the high intensity pixel at the top of the image. Noise due to scatter was neglected, however the effect of attenuation was considered.

It can be shown that for emission computed tomography, the Fisher information matrix has the form

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

where A is a $d \times n$ weight matrix; d = total number of detectors, and $\underline{\mu} = A\,\underline{\theta}$. Therefore F_Y is an $n \times n$ symmetric positive semi-definite matrix ($F_Y \geq 0$). Here we will only consider non-singular $F_Y (> 0)$. The system used in this

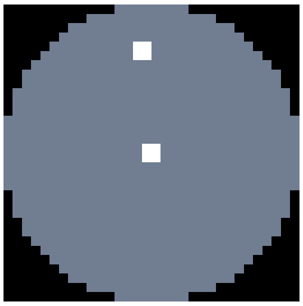

Figure 4: The phantom used in the simulations. The pixel of interest is the pixel at the top of the phantom. The image dimensions are 32x32 pixels.

study is the SPRINT II system as described before.

A. Spatial Sampling Study

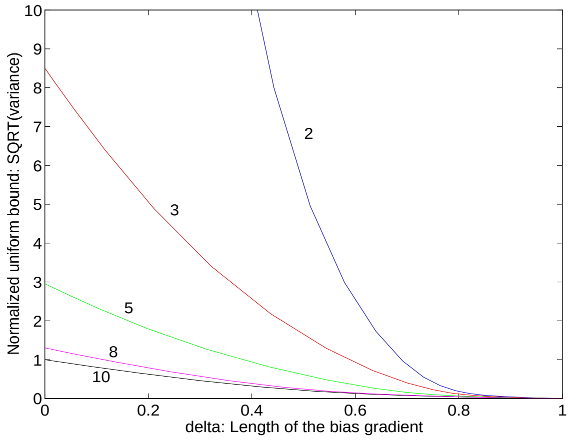

Figure 5: The effect of angular sampling on the unbiased CR bound. Curves denote lower bound for 2,3,5,8 and 10 rotation increments of the collimator.

In Figure 5 the bias-variance curves are displayed for varying degrees of spatial sampling. Spatial sampling was varied by rotating the collimator ring through different number of steps. The time for each step was kept constant. Note the monotonic nature of the curves: they do not intersect each other due to the fact that the projections from the lower sampled image is a subset of those of the higher sampled image. Also it can be seen that beyond a sampling rate of 8 increasing sampling rate does not reduce the bound significantly.

An interesting point to note is that the unbiased CR bound (zero bias gradient) for 10 rotations is greater than the bound at a bias gradient length of 0.5 for only 5 rotations. This means that a biased algorithm with lower sampling can perform better than an unbiased algorithm with higher sampling. For a time normalized case, when the total time for each curve is kept constant, the curves may intersect each other (Figure 6). They, however, display the same trend as above.

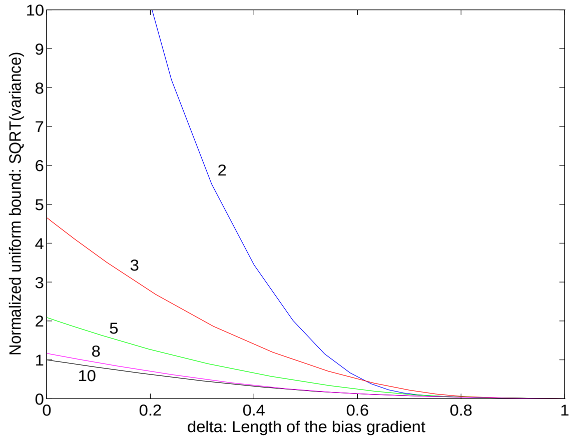

Figure 6: The effect of angular sampling on the unbiased CR bound. Time normalized case.

B. Optimal System Design

Spatial resolution and sensitivity are the two most important criteria for tomographic system design. These criteria are coupled, higher resolution can only be bought at the price of lower sensitivity. We demonstrate the use of the uniform CR bound (3-b) by determining an aperture opening that optimizes the tradeoff between resolution and sensitivity. For these simulations 10 collimator rotations were used and the aperture opening was varied from very narrow (ray width = 0.25 pixels) to very wide (ray width = 10 pixels). The total 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. Figure 7 shows a sharp minimum over all δ when the width of the ray is approximately one pixel. Therefore for the object and the ROI studied, this one pixel aperture width is 'universally' optimal for estimation, irrespective of the bias of the estimator.

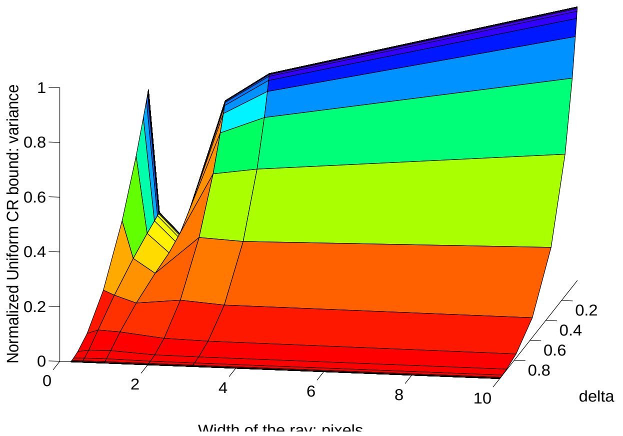

Figure 7: Optimal system design using the CR bound to study the effect of resolution and sensitivity tradeoffs

C. Performance Comparison

Here we will compare two least square reconstruction algorithms by placing them on the bias-variance plane and comparing them to the uniform CR lower bound. In the following simulations we used the standard SPRINT II system with 10 collimator rotations.

The least square estimator can be motivated by the following representation of $\underline{y}$:

$$\underline{y} = A\,\underline{\theta} + \underline{n}$$

Where $\underline{n}$ is the Poisson distributed noise and in SPECT it is generated by photon scatter.

C.1 Unweighted Linear Least Square Estimator

The unweighted Linear Least Square Estimator (LLSE) is given by:

$$\hat{\underline{\theta}}^{LLSE} = (A^T A + \alpha\, I)^{-1}\, A^T\, \underline{y} \tag{4}$$

where $\alpha \geq 0$ is a smoothing or regularization parameter to counteract any ill-conditioning of A. It can be easily shown that the bias vector $\underline{b}(\underline{\theta}) = E(\hat{\underline{\theta}}^{LLSE} - \underline{\theta})$ is :

$$\underline{b}(\underline{\theta}) = I - (A^T A + \alpha\, I)^{-1}\, A^T A\, \underline{\theta}$$

Therefore the bias gradient is the $n \times n$ matrix:

$$\nabla_{\underline{\theta}} b(\underline{\theta}) = I - (A^T A + \alpha\, I)^{-1}\, A^T A$$

The covariance of the LLSE $\hat{\underline{\theta}}^{LLSE}$ is:

$$cov(\hat{\underline{\theta}}^{LLSE}) = (A^T A + \alpha\, I)^{-1}\, A^T A\, (A^T A + \alpha\, I)^{-1} \tag{5}$$

Note that the bias gradient and bias are very simply related. In particular the bias gradient is equal to the bias for

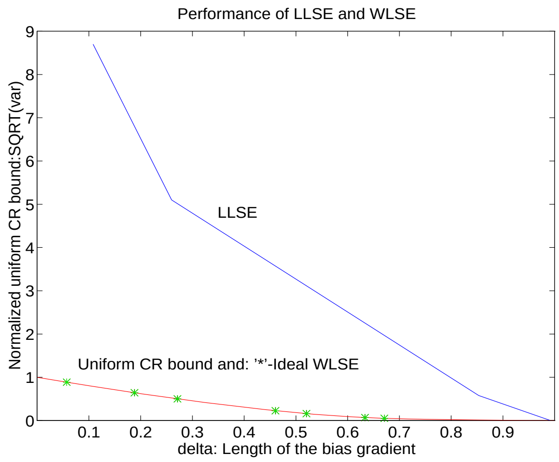

Figure 8: Linear Least Square Estimator

a point source intensity $\underline{\theta} = [1, 0, ..., 0]^T$. The bias gradient $\nabla_{\underline{\theta}} b_1$ of $\hat{\theta}_1^{LLSE}$ is simply the first row of $\nabla_{\underline{\theta}} b$. The length of the bias gradient $\nabla_{\underline{\theta}} b_1$ is a function of the smoothing parameter α, increasing from 0 as α increases from 0. In Figure 8 the uniform CR bound (3-b) is plotted as a function of the length of $\nabla_{\underline{\theta}} b_1$. Also plotted is the variance of $\hat{\theta}_1^{LLSE}$, obtained from the $(1, 1)$ element of the covariance matrix (5). Note that as the smoothing parameter α gets large the variance of LLSE approaches the CR bound but for small α there is a large difference.

C.2 Weighted Least Squares Estimator

The weighted least squares estimator (WLSE) is:

$$\hat{\underline{\theta}}^{WLSE} = (A^T \, \Sigma \, A + \alpha \, I)^{-1} \; A^T \, \Sigma \; \underline{y}$$

where α is the regularization parameter Σ is an $n \times n$ weight matrix. Note that when $\Sigma = I$, $\hat{\underline{\theta}}$ is simply the unweighted least square estimator (4). It is easy to show that the bias gradient is:

$$\nabla_{\underline{\theta}} b(\underline{\theta}) = I - (A^T \, \Sigma \, A + \alpha \, I)^{-1} \; A^T \, \Sigma \, A$$

and the covariance is:

$$cov(\hat{\underline{\theta}}^{WLSE}) = (A^T \, \Sigma \, A + \alpha \, I)^{-1} \; A^T \, \Sigma A \, (A^T \, \Sigma \, A + \alpha \, I)^{-1}$$

For $\Sigma = diag(\frac{1}{\mu_i})$, where $\underline{\mu} = A \, \underline{\theta}$, the variance of $\hat{\theta}_1^{WLSE}$ is plotted in Figure 8. Observe that for this choice of weights the WLSE estimator achieves the bound exactly for all values of α as shown in [7]. The WLSE implemented with $\Sigma = diag(\frac{1}{\mu_i})$ is not physically realizable since μ_i depends on the unknown image $\underline{\theta}$. However, by replacing the mean $\underline{\mu}$ by the sample mean $\hat{\underline{\mu}}$, an estimated WLSE: $\hat{\underline{\theta}}^{WLSE} = (A^T \, \hat{\Sigma} \, A + \alpha \, I)^{-1} \; A^T \, \hat{\Sigma} \; \underline{y}$, $\hat{\Sigma} = diag(\frac{1}{\hat{\mu}_i})$ has been found to have similar variance and bias for large number of detected counts [8].

V. Conclusions

We have demonstrated, with the help of several examples, that the commonly used unbiased CR bound is not sufficient by itself for performance analysis and the uniform CR bound should be used to get a true overall picture.

The image reconstruction problem is very often ill-posed due to the insufficient radial sampling resulting in a singular FIM. Many of the results that are presented in this paper have been extended to the case of singular FIM but the space limitation does not permit us to include them here. For details please see [5].

A System Specifications

Ring Detector

Radius of the detector ring	25 cms
Number of detectors	512
Radius of the collimator ring	17 cms
Number of collimator slits	9 (uniformly spaced)
Slit Width	2.4 mm

References

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

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

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

[4] 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.

[5] M. Usman, A. O. Hero and J. A. Fessler
Uniform CR Bound for Ill-conditioned Systems
In preparation

[6] A. O. Hero, J. A. Fessler and W. L. Rogers
A Fast Recursive Algorithm for Computing CR-Type Bounds for Image Reconstruction Problems, IEEE Symposium on Nuclear Science and Medical Imaging Conference, October 1992, Orlando, FA.

[7] J. A. Fessler and A. O. Hero
Cramer-Rao Bounds for Biased Estimators in Image Reconstruction, MWSCS, August 1993, Detriot, MI.

[8] J. A. Fessler
Penalized Weighted Least-Squares Image Reconstruction for Positron Emission Tomography, Accepted by IEEE Transactions on Medical Imaging, 1993.

[9] M. Usman, A. O. Hero and W. L. Rogers
Performance Gain Analysis for Adding Vertex View to a Standard SPECT, MWSCS, August 1993, Detriot, MI.