RECURSIVE CR BOUNDS: ALGEBRAIC AND STATISTICAL ACCELERATION

Mohammad Usman[1] *Alfred O. Hero*[1]

[1] Dept. of Electrical Engineering and Computer Science
The University of Michigan, Ann Arbor MI 48109-2122

ABSTRACT

Computation of the Cramer-Rao bound involves inversion of the Fisher information matrix (FIM). The inversion can become computationally intractable when the number of unknown parameters is large. Hero has presented a recursive, monotonically convergent and computationally efficient algorithm to invert sub-matrices of the FIM corresponding to a small region of interest in image reconstruction [1]. The convergence rate of this algorithm depends on a splitting matrix which can be interpreted as a complete-data FIM. In this paper we investigate the acceleration of the algorithm using several different choices of the complete-data FIM. We also present a conjugate gradient based algorithm which achieves a much faster convergence rate at the expense of monotone convergence. We apply the methods developed in this paper to emission tomography.

1. INTRODUCTION

The Cramer-Rao (CR) bound determines a lower bound on the variance of any estimator. Calculation of the CR bound involves inversion of a non-singular Fisher Information Matrix (FIM) F_Y. Direct matrix inversion, requiring $O(n^3)$ flops (FLOating Point operationS), could be computationally intractable if the number n of parameters to be estimated is large. For example, in emission tomography, where the pixel intensities are the parameters to be estimated, a moderate size image of 128×128 pixels has a FIM of dimension $(128)^2 \times (128)^2$ and will require 4.4×10^{12} flops to compute its inverse.

Often we require a bound on few estimator components of interest, called the Region Of Interest (ROI). Hero et al. [1] presents a recursive and computationally efficient method for calculating a small portion of matrix ${F_Y}^{-1}$ which corresponds to a $q \times q$ ROI. As presented in [4], this algorithm requires that we find the FIM F_X of an imaginary "complete-data" set, called the splitting matrix, such that F_X dominates F_Y in the sense that $F_X - F_Y \geq 0$. In this paper we obtain an optimal F_X by purely algebraic, non-statistical approach, as a solution to a constrained optimization problem. The main advantage of the algorithm of [1, 2] is its monotone convergence which generates a valid improving lower bound on estimator covariance at each iteration of the algorithm. However, the price paid for monotone convergence is slow linear convergence rate. For applications where a strict lower bound is not required at each iteration, a non-monotone conjugate gradient algorithm is given which has a significantly faster convergence rate.

We apply the methods developed in this paper to emission tomography using several different images for a Single Photon Emission Computed Tomography (SPECT) system. Finally some conclusions are presented.

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

2. CR BOUND

Given a measurement Y that is a random variable and has probability distribution $f_Y(Y;\underline{\theta})$ dependent on an unknown parameter vector $\underline{\theta}$, we want to estimate $\underline{\theta} = [\theta_1, \theta_2, ..., \theta_n]^T$. A vector function $\hat{\underline{\theta}}(y)$ is a parameter estimator based on the observation $Y = y$.

The Cramer-Rao lower bound on the covariance of an unbiased parameter estimator is given by the inverse of FIM:

$$cov(\hat{\underline{\theta}}) \geq F_Y^{-1}(\underline{\theta}). \quad (1)$$

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 $E_{\underline{\theta}}$ denotes statistical expectation with respect to $f_Y(Y;\underline{\theta})$.

Under broad conditions the unbiased CR bound (1) is known to be asymptotically achievable for increasing observation times [3].

3. RECURSIVE ALGORITHMS

3.1. Monotone Convergent Algorithm

Assume that we want to calculate the CR-Bound for only the first component of $\underline{\theta}$, corresponding to the top left entry of F_Y^{-1}. The algorithm can be easily extended to q parameters, $q \leq n$ [5]. Let $\underline{e}_1 = [1, 0, 0, ..., 0]^T$. Then a recursive algorithm for computing the first column of F_Y^{-1} is given as follows:

Algorithm

INITIALIZATIONS:

$\underline{\beta}^{(0)} = \underline{0}$ an $n \times 1$ vector of zeroes

RECURSION:

FOR k := 0,1,...,l DO

$$\underline{\beta}^{(k+1)} = \underline{\beta}^{(k)} - F_X^{-1}(\ F_Y\,\underline{\beta}^{(k)} - \underline{e}_1\ ) \quad (2)$$

END DO

BOUND:

$$B^{(l)} = \underline{e}_1^T \ \underline{\beta}^{(l)}. \tag{3}$$

In [1] it is shown that if F_X dominates F_Y in the sense of positive semi-definitiveness of $F_X - F_Y$ then the algorithm generates a sequence of approximations $B^{(l)}$ which monotonically converges to 1×1 sub-matrix B of ${F_Y}^{-1}$ as $l \to \infty$. The monotone convergence property of this algorithm guarantees that the matrix $B^{(l)}$ calculated at each recursion by this algorithm is a valid lower bound.

Since (2) requires only $O(n^2)$ flops, if l is less than n, we obtain a bound with fewer flops than required for direct inversion of F_Y ($O(n^3)$), e.g. using the method of sequential partitioning [6]. For this algorithm to be computationally efficient we require F_X^{-1} to be easy to compute. The rate of convergence of this algorithm is directly proportional to the spectral radius $\rho(M) = \rho([I - \ {F_X}^{-1} F_Y]) = \lambda_{max}^M$, where λ_{max}^M is the largest eigenvalue of $I - \ {F_X}^{-1} F_Y$.

Thus we can obtain a faster speed of convergence by reducing $\rho(M)$. This suggests that we find a sparse F_X, that is easy to invert, satisfies $F_X - F_Y \geq 0$ and, gives the minimum spectral radius $\rho(M)$. One can show that even for a simple case of a diagonal F_X the solution to this constrained min-max optimization problem involves the calculation of eigenvalues and eigenvectors of an $n \times n$ matrix [5]. However, instead of directly minimizing the spectral norm we can minimize the matrix Frobenius norm $\| F_X - F_Y \|_F$. The inequality $\lambda_{max}^M \leq \| F_Y^{-1} \|_F \ \| F_X - F_Y \|_F$ relates λ_{max}^M to $\| F_X - F_Y \|_F$, guaranteeing that minimization of $\| F_X - F_Y \|_F$ at-least makes the spectral norm λ_{max}^M small.

Define the p-diagonal matrix $D_p =$

$$\begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1p} & & \cdots & 0 \\ d_{21} & d_{22} & & & \ddots & & \vdots \\ \vdots & & \ddots & & & \ddots & \\ d_{p1} & & & \ddots & & & d_{n-p,n} \\ & \ddots & & & \ddots & & \vdots \\ \vdots & & \ddots & & & \ddots & d_{n-1,n} \\ 0 & \cdots & & d_{n,n-p} & \cdots & d_{n,n-1} & d_{nn} \end{bmatrix}. \tag{4}$$

The following theorem is proven in [5]:

Theorem 1 *Let an $n \times n$ matrix $A = ((a))$ be a symmetric Positive Semi-Definite (PSD) matrix. Then the p-diagonal matrix D_p, as defined in (4), that minimizes the Frobenius norm $\| D_p - A \|_F$ s.t. $D_p - A$ is PSD is given by $D_p =$*

$$\begin{bmatrix} \sum\limits_{j \neq 2,\ldots,n} |a_{1j}| & a_{12} & \cdots & a_{1p} & & \cdots & 0 \\ a_{21} & \sum\limits_{j \neq 1,3,\ldots,n+1} |a_{2j}| & & & \ddots & & \vdots \\ \vdots & & \ddots & & & \ddots & \\ a_{p1} & & & \ddots & & & a_{n-p,n} \\ & \ddots & & & \ddots & & \vdots \\ \vdots & & \ddots & & & \ddots & a_{n-1,n} \\ 0 & \cdots & & a_{n,n-p} & \cdots & a_{n,n-1} & \sum\limits_{j \neq n-p,\ldots,n-1} |a_{nj}| \end{bmatrix}. \tag{5}$$

Thus selecting $F_X = D_p$ minimizes the Frobenius norm between F_X and F_Y over all p-diagonal matrices. We need F_X^{-1} to be easy to compute ($O(n^2)$ or less). Efficient inversion algorithms exist for diagonal (1-diagonal) and tri-diagonal (2-diagonal) matrices [9]. Using theorem 1 we can directly write the optimal 1-diagonal and optimal 2-diagonal F_X.

Corollary 1 *The optimal 1-diagonal matrix is:*

$$D_1 = \begin{bmatrix} \sum |a_{1j}| & & \cdots & 0 \\ & \sum |a_{2j}| & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & \sum |a_{nj}| \end{bmatrix}.$$

Corollary 2 *The optimal 2-diagonal matrix is:*

$$D_2 = \begin{bmatrix} \sum\limits_{j \neq 2} |a_{1j}| & a_{12} & & \cdots & 0 \\ a_{21} & \sum\limits_{j \neq 1,3} |a_{2j}| & \ddots & & \vdots \\ & \ddots & \ddots & \ddots & \\ \vdots & & \ddots & \ddots & a_{n-1,n} \\ 0 & \cdots & & a_{n,n-1} & \sum\limits_{j \neq n-1} |a_{nj}| \end{bmatrix}.$$

By identifying A with F_Y and F_X with D_1 or D_2 in the algorithm we can obtain an acceleration of the convergence rate.

3.2. Non-monotone Convergent Algorithm: Conjugate Gradient

Conjugate gradient is an algorithm to solve the linear system of equation $F_Y \underline{x} = \underline{b}$ [8, 7]. 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 F_Y^{-1}. In general we can get the inverse of the m-th column of F_Y if we choose $b_m = 1$ and $b_j = 0 : j \neq m$. The following conjugate gradient algorithm is applicable to symmetric, positive definite matrices F_Y [8].

Algorithm

INITIALIZATIONS:

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

RECURSION:

FOR i := 0 UNTIL |r(i)| < Tolerance DO

IF i = 0 DO

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

ELSE

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

END IF

$$\lambda_i = \frac{< \underline{r}^i, \underline{r}^i >}{< \underline{p}^i, F_Y \underline{p}^i >}$$
$$\underline{u}^{i+1} = \underline{u}^i + \lambda_i \ \underline{p}^i$$
$$\underline{r}^{i+1} = \underline{r}^i - \lambda_i \ F_Y \ \underline{p}^i$$

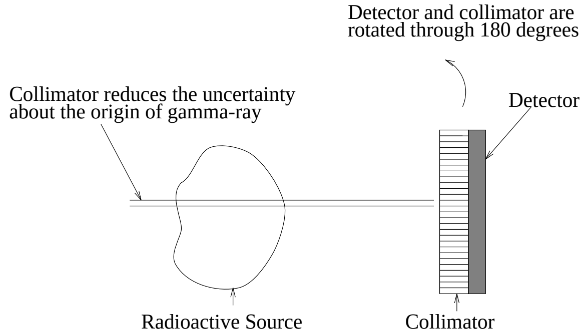

Figure 1. A SPECT system

```
END DO
BOUND:
```

$$B^{(i)} = \underline{b}^T \, \underline{u}^i.$$

The iterations are terminated when the residual error $|r^i|$ is less than a user specified tolerance. Although the above algorithm calculates only a single column of F_Y^{-1}, it can be easily extended to q columns: $q \leq n$ [5]. The convergence rate of the conjugate gradient depends on the condition number of F_Y and its eigenvalue distribution. One advantage in using the conjugate gradient algorithm is when computed with infinite precision, the algorithm is guaranteed to converge in maximum n iterations. Each iteration of the algorithm involves a matrix-vector multiplication $F_Y \, \underline{p}^i$, requiring $O(n^2)$ flops and therefore if the algorithm is stopped before n iterations, an $O(n^2)$ approximation to the CR bound is obtained.

In relation to the recursive algorithm of previous section, a disadvantage of this algorithm is its non-monotonic convergence. This means that if stopped at iteration $k < n$ then the quantity $B^{(i)}$ calculated by the conjugate gradient algorithm might not be a valid lower bound. This, however, may not be a problem if we let the algorithm run until a desired accuracy is achieved. Note that the conjugate gradient algorithm does not require the calculation of another complete-data FIM F_X.

4. APPLICATIONS

In this section we apply the recursive algorithms described in the previous section to a simple Single Photon Emission Computed Tomography (SPECT) system to several different objects (Figure 2). We start with a brief description of the SPECT system.

4.1. Single Photon Emission Computed Tomography

A SPECT system consists of three basic components: 1) a source of γ-rays, 2) a γ-ray detector and, 3) a γ-ray collimator. The function of the collimator is to reduce the uncertainty associated with the origin of a γ-ray to a line or a path (Figure 1). During the imaging time, the γ-ray detector is rotated 180^o through small steps around the source. The size of the step is determined by the spatial sampling requirements. For reconstruction the source domain is divided into small regions, called pixels. Both the γ-ray emission and detection processes are governed by Poisson statistics. The objective is to reconstruct the intensity of each pixel $\underline{\theta} = [\theta_1, \theta_2, ..., \theta_n]^T$, defined as the average number of γ-ray photons emitted by a pixel during the imaging time, given the set of observations Y; which are the detected γ-rays over the m detector bins.

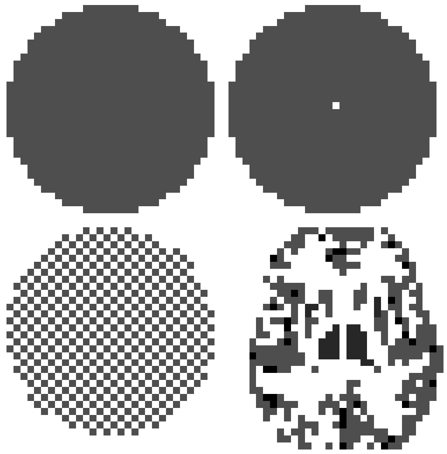

Figure 2. Objects used for numerical comparisons. The pixel of interest is the center pixel. Objects are: Top left: uniform, Top right: Point source, Bottom left: Checker board and, Bottom right: Brain image.

It can be shown that for emission computed tomography, the Fisher information matrix has the form [4]:

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

where A is a $m \times n$ weight matrix; m = total number of detector bins, $\underline{\mu} = A \, \underline{\theta}$, and $diag(\underline{\mu})$ is a diagonal matrix with diagonal elements $\mu_1, \mu_2, ..., \mu_m$. F_Y is an $n \times n$ symmetric positive semi-definite matrix ($F_Y \geq 0$), which is typically non-singular if $m \geq n$. The standard choice of complete-data in emission tomography is the set of observable γ-ray emissions over the n pixels. The FIM for the complete-data is [4]:

$$F_X = diag(\underline{1}^T \, A) \, [diag(\underline{\theta})]^{-1}.$$

4.2. Numerical Comparisons

Object: Several objects were used in the simulations (Figure 2). The high intensity pixels have a normalized intensity value of 2 and the low intensity pixels are set to 1. The object is within a disk of diameter 32 pixels: $n = 716$ pixels. In all the simulations the pixels of interest was the pixel at the center of the image.

The recursive algorithm was said to converge when the error between the actual and the calculated value was less than 10^{-4} and remained there for atleast 5 iterations.

Figure 3 shows the effect of increasing the number of diagonals of optimal F_X on the rate of convergence for the checker board image. As expected the rate of convergence improves with increasing number of diagonals since the Frobenius norm $\| F_X - F_Y \|_F$ decreases. For $p = n$ the algorithm converges in a single iteration since $F_X = F_Y$ in this case. Also plotted in the same figure is the convergence plot for the diagonal F_X called EM, which exhibits the slowest rate of convergence, and the convergence plot of the conjugate gradient algorithm, called CG, which has the

Table 1. Summary of convergence properties.

Algorithm	Object											
	Uniform			Point Source			Checker Board			Brain Image		
	No. of iter.	$\sum_i \lambda_i$	λ_{max}	No. of iter.	$\sum_i \lambda_i$	λ_{max}	No. of iter.	$\sum_i \lambda_i$	λ_{max}	No. of iter.	$\sum_i \lambda_i$	λ_{max}
EM	803	337.17	.9997	1094	337.12	.9997	902	342.13	.9998	7892	349.11	.9999
1-diag	803	337.17	.9997	1094	337.12	.9997	686	337.12	.9997	879	335.67	.9997
2-diag	759	336.66	.9997	1036	336.65	.9997	622	336.62	.9997	842	335.14	.9997
Conjugate Gradient	29			42			31			42		

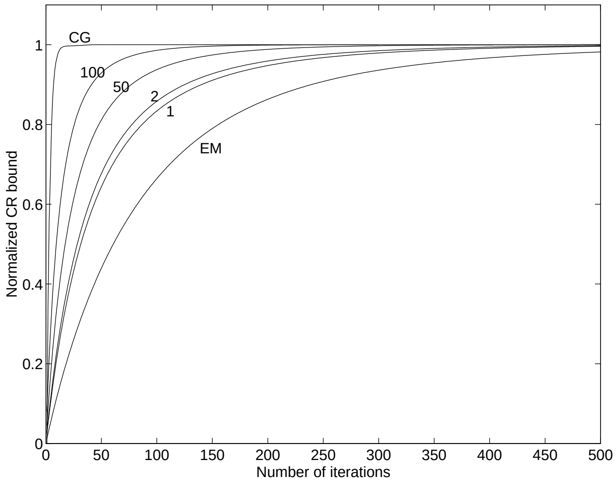

Figure 3. Effect of changing the number of diagonals for the optimal p-diagonal F_X and the checker board image. Numbers with each curve indicate the number of diagonals.

fastest rate of convergence. The CG plot oscillates around the point of convergence but it is not visible in the plot due to small magnitude of these oscillations.

Table 1 summarizes the performance of all algorithms for all the objects. Out of the optimal p-diagonal F_X we have only included optimal 1-diagonal and 2-diagonal F_X. There are several comments in relation to Table 1.

- For the uniform source both EM and the optimal 1-diagonal F_X are identical and hence have identical convergence. For a point source there is no difference either.
- For images with greater non-uniformity i.e. the checker board and the brain image, the optimal 1-diagonal and 2-diagonal show a significant improvement over the EM. In general for more irregular images we observe similar behavior.
- The rate of convergence of the conjugate gradient is far superior to any non-monotone convergent algorithm, at-least by an order of magnitude.
- Although conjugate gradient overshoots the true CR bound, and therefore it is not, strictly speaking, a valid bound, the percentage overshoot is very small (0.0011%) for this example. It was also observed that the percentage overshoot was sensitive to initial condition, here chosen as all zero image.

5. CONCLUSIONS

We have investigated several recursive algorithms for the inversion of a non-singular, positive-definite matrix and applied them to the computation of the CR bound in emission tomography. The algorithms presented are not restricted to this particular application. The results of the simulations should only be used as a guideline since the results may vary depending upon the application.

REFERENCES

[1] A.O. Hero, J.A. Fessler and W.L. Rogers
A Fast Recursive Algorithm for Computing CR-Type Bounds for Image Reconstruction Problems. *IEEE Nuclear Science Symposium and Medical Imaging Conference*, Orlando, 1983

[2] A.P. Dempster, N.M. Laird and D.B. Rubin
Maximum Likelihood From Incomplete Data via the EM Algorithm; *Journal of the Royal Statistical Society*, Ser. B, Vol. 39, pp 1-37, 1977

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

[4] 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.

[5] M. Usman and A.O. Hero
The Analysis of a Single Photon Emission Computed Tomography (SPECT) System Using Mutual Information and Cramer-Rao Bound as Performance Criteria, CSPL Technical Report, The University of Michigan, 1993.

[6] A. Kuruc
Lower bounds on Multiple-source Direction Finding in the Presence of Direction-dependent Antenna-array-calibration errors, Technical Report 799, MIT Lincoln Laboratory, Lexington MA, 1989.

[7] L.A. Hageman and D.M. Young
Applied Recursive Methods, Academic Press, 1981.

[8] 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.

[9] D. Rose and R.A. Willoughby
Sparse Matrices and their Applications, Plenum Press, 1972.