Recursive Algorithms for Computing the Cramer-Rao Bound[1]

Mohammad Usman, Alfred O. Hero, Anne Sauve, and Jeffrey A. Fessler

Abstract

Computation of the Cramer-Rao bound requires the inverse or the pseudo-inverse Fisher information matrix (FIM). Direct matrix inversion can be computationally intractable when the number of unknown parameters is large. In this paper we compare several recursive methods for extracting columns of the inverse Fisher information matrix including the algorithm of [1] and similar monotone matrix splitting algorithms, over-relaxation techniques, and preconditioned conjugate gradient algorithms. The latter algorithms can converge significantly faster than the monotone algorithms for equivalent computational load. We illustrate the methods developed in this paper for an important class of inverse problems.

Key Words: Performance bounds, multi-dimensional parameter estimation, monotone matrix splitting iterations, preconditioned conjugate gradient.

EDICS Number: SP 3.8.1

1 Introduction

The Cramer-Rao (CR) bound is a very widely used lower bound on the covariance of any unbiased estimator. For an n-dimensional parameter vector calculation of the CR bound involves finding elements of the inverse of the $n \times n$ Fisher information matrix (FIM) $\mathbf{F}_Y$. Direct matrix inversion, requiring $O(n^2)$ bytes of memory storage and $O(n^3)$ flops (floating point operations), is intractable if n is large. For example, in image reconstruction where the pixel intensities are the parameters to be estimated, a moderate size image of $128 \times 128 = 16384$ pixels has a FIM of dimension 16384×16384 which would require 268 Megabytes memory storage (single precision) and 4 Tera flops ($O(10^{-12})$ flops) for computation of the inverse FIM.

Often only a few components of the n-dimensional estimator are of interest in which case the entire inverse FIM is not needed. For example in image reconstruction one may be primarily interested in a small q-pixel region of interest (ROI) corresponding to a tumor or lesion. Methods of sequential partitioning [2] can be used in this case to calculate the $q \times q$ submatrix of the inverse Fisher matrix yielding the CR bound for the ROI. However, while requiring fewer flops than required for full matrix inversion, this algorithm still requires the same order $O(n^3)$ of flops.

In [1] and [3] a recursive method was presented for approximating columns of the CR bound. This method was derived by specifying a FIM-dominating matrix which allowed a geometric series decomposition of the inverse FIM. The geometric series algorithm requires only $O(n^2)$ flops per iteration per parameter so that if convergence speed is fast, a significant computational saving is achieved. The main advantage of this algorithm is its monotone convergence which guarantees a valid and improving lower bound on estimator covariance at each iteration. The disadvantage is slow speed of convergence. In [3, 1] special properties of the Fisher information matrix were exploited to specify the FIM-dominating matrix. In Section 3.2 we give an alternative dominating matrix based on purely algebraic considerations.

This paper places the methods of [1] in the context of the general class of iterative linear equation solvers known as stationary and non-stationary algorithms [4]. These methods include monotone and non-monotone matrix splitting algorithms, such as the method of [1], Jacobi, Gauss-Seidel, and preconditioned conjugate gradient algorithms. We also present an algorithm for computing the pseudo-inverse version of the CR bound for singular FIM based on matrix perturbation and differencing methods. These iterative algorithms are illustrated for a class of inverse problems arising in deconvolution, image restoration, and tomographic imaging.

The main conclusions of this paper are: 1) the non-monotone Gauss-Seidel and pre-conditioned conjugate gradient methods have an order of magnitude speed advantage relative to the monotone methods; 2) for sparse matrix inverse problems these methods can be implemented with significant savings in memory and computation; 3) for inverse problems having singular FIM the application of Gauss-Seidel and conjugate gradient algorithms yields stable and rapidly convergent methods for finding columns of the pseudoinverse of the FIM.

[1] J. Fessler, A. Hero, and A. Sauve are with the Dept. of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, MI 48109-2122. M. Usman is with the Punjab Institute of Computer Science, Pakistan. This work was supported in part by the National Science Foundation under grant BCS-9024370, a Government of Pakistan Postgraduate Fellowship, NIH grants CA-54362 and CA-60711, and DOE grant DE-FG02-87ER60561

2 The Cramer-Rao Bound

Let Y be an observed random variable with probability density $f_Y(y;\underline{\theta})$ dependent on an unknown parameter vector $\underline{\theta}=[\theta_1,...,\theta_n]^T$ lying in an open subset Θ of $\mathbb{R}^n$. The covariance $\mathrm{cov}_{\underline{\theta}}(\hat{\underline{\theta}}) = E_{\underline{\theta}}\left[(\hat{\underline{\theta}}-E_{\underline{\theta}}(\hat{\underline{\theta}}))(\hat{\underline{\theta}}-E_{\underline{\theta}}(\hat{\underline{\theta}}))^T\right]$ of any unbiased estimator $\hat{\underline{\theta}}=\hat{\underline{\theta}}(Y)$ of $\underline{\theta}$ obeys the Cramer-Rao lower bound [5, 6]

$$\mathrm{cov}_{\underline{\theta}}(\hat{\underline{\theta}}) \geq \mathbf{F}_Y^+(\underline{\theta}), \tag{1}$$

where $\mathbf{F}_Y$ is the FIM

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

the column $\nabla_{\underline{\theta}}$ denotes the gradient vector $[\frac{\partial}{\partial\theta_1},\frac{\partial}{\partial\theta_2},...,\frac{\partial}{\partial\theta_n}]^T$, and $\mathbf{F}_Y^+$ denotes the Moore-Penrose pseudo-inverse [7]. When $\mathbf{F}_Y$ is non-singular $\mathbf{F}_Y^+=\mathbf{F}_Y^{-1}$. Note that the pseudo-inverse form of the CR bound is generally only achievable for estimation of components of $\underline{\theta}$ in the rangespace of $\mathbf{F}_Y$ [?].

The matrix inequality (1) means that $\mathrm{cov}_{\underline{\theta}}(\hat{\underline{\theta}}) - \mathbf{F}_Y^+(\underline{\theta})$ is non-negative definite. In particular the lower bound on the variance of $\hat{\theta}_1$, corresponding to the top left (1,1) entry of $\mathrm{cov}_{\underline{\theta}}(\hat{\underline{\theta}})$, is given by the corresponding entry of $\mathbf{F}_Y^+$

$$\mathrm{var}_{\underline{\theta}}(\hat{\theta}_1) \geq \underline{e}_1^T\mathbf{F}_Y^+(\underline{\theta})\underline{e}_1, \tag{2}$$

where $\underline{e}_1=[1,0,...,0]^T$. Throughout this paper we will be interested in calculating the right hand side of (2), or more generally the first column of $\mathbf{F}_Y$. The results are easily extended to q -elements, $q << n$, by repeated application of the algorithms given here.

Direct computation of the right hand side of the CR inequality (2) can be accomplished in $O(n^3)$ flops by solving for $\underline{x}$ in the equation

$$\mathbf{F}_Y\,\underline{x} = \underline{e}_1. \tag{3}$$

For example, when $\mathbf{F}_Y$ is non-singular the solution $\underline{x}$ can be found by the method of Gaussian elimination ($2n^3/3$ flops) or by methods which exploit positive definiteness of $\mathbf{F}_Y$, such as methods based on the QR and Cholesky decomposition ($n^3/3$ flops) [8, Sec. 4.2].

3 Recursive CR Bound Algorithms for Non-Singular FIM

Here we describe several methods for iteratively approximating the right hand side of the CR inequality (2) for non-singular FIM. We first describe the monotonically convergent algorithm of [1] in the context of standard splitting iterations [8, Sec. 10.1] , also known as stationary iterations [4], for approximating the solution to the linear equation (3).

3.1 Splitting Algorithms

Let $\mathbf{F}$ and $\mathbf{N}$ be $n\times n$ matrices which split $\mathbf{F}_Y$ in the sense that $\mathbf{F}-\mathbf{N}=\mathbf{F}_Y$ and assume that $\mathbf{F}$ is non-singular.

3.1.1 General Splitting Iterations

Define the *general splitting iterations for approximation of the CR bound*

INITIALIZE: $\eta^{(0)}=0$, $\underline{\beta}^{(0)}=\underline{0}$ an $n\times 1$ vector of zeroes.

RECURSION:

$$\begin{array}{ll} i). & \underline{t}=\mathbf{N}\underline{\beta}^{(k)}+\underline{e}_1 \\ ii). & \text{Solve}: \quad \mathbf{F}\underline{\beta}^{(k+1)}=\underline{t} \end{array} \tag{4}$$

CR BOUND APPROXIMATION:

$$\eta^{(k+1)} = \underline{e}_1^T\ \underline{\beta}^{(k+1)}.$$

Define the matrix $\mathbf{M}=\mathbf{F}^{-1}\mathbf{N}$. The root convergence factor $\rho(\mathbf{M})=|\lambda_i^M|_{\max}$ is defined as the maximum magnitude eigenvalue of $\mathbf{M}$, also known as the spectral radius of $\mathbf{M}$. If the $\rho(\mathbf{F}^{-1}\mathbf{N})<1$ then $\underline{\beta}^{(k)}$ converges to the first column of $\mathbf{F}_Y^{-1}$ [9, 8] and the approximating sequence $\eta^{(k)}$ converges to the CR bound $\underline{e}_1^T\mathbf{F}_Y^{-1}\underline{e}_1$. Furthermore, it can be shown in an analogous manner to [1, Eq. (9)] that if $\mathbf{F}$ is symmetric positive definite and $\mathbf{N}$ is symmetric non-negative definite then $\eta^{(k+1)}-\eta^{(k)} = \underline{e}_1^T\mathbf{F}^{-\frac{1}{2}}[\mathbf{F}^{-\frac{1}{2}}\mathbf{N}\mathbf{F}^{-\frac{1}{2}}]^k\mathbf{F}^{-\frac{1}{2}}\underline{e}_1$. Hence, under these conditions the approximating sequence $\eta^{(k)}$ is monotone non-decreasing in k and, if $\rho(\mathbf{F}^{-1}\mathbf{N})<1$, $\eta^{(k)}$ converges to the CR bound from below.

Many well known algorithms fall into the category of general splitting iterations such as the Jacobi and Gauss-Siedel iterations. Let the FIM have the additive decomposition $\mathbf{F}_Y=\mathbf{D}+\mathbf{U}+\mathbf{L}$ where $\mathbf{D}$ is diagonal and $\mathbf{U}$ and $\mathbf{L}$ are upper and lower diagonal matrices with zero diagonal entries. The Jacobi (J) iteration is obtained by making the identifications $\mathbf{F}=\mathbf{D}$, $\mathbf{N}=-(\mathbf{U}+\mathbf{L})$ in (4). For Jacobi iterations, the spectral radius of $\mathbf{F}^{-1}\mathbf{N}$ may exceed one and $\mathbf{N}$ is not generally non-negative definite. Therefore, the Jacobi iterations may not converge and are generally not monotone. The Gauss-Siedel (GS) iteration is obtained by identifying $\mathbf{F}=\mathbf{D}+\mathbf{L}$ and $\mathbf{N}=-\mathbf{U}$ in (4). Like Jacobi the Gauss-Seidel iterations yield non-monotonic approximations. On the other hand, the GS iterations always converge for positive definite FIM.

For large n step (i) in (4) requires $2n^2$ flops. If $\mathbf{F}$ is a diagonal matrix, as in Jacobi, step (ii) requires n flops. For $\mathbf{F}$ a banded diagonal matrix with upper and lower bandwidth of $p-1$, ($p=1$ gives a diagonal matrix) step (ii) can be computed with $2np^2$ flops ($n\gg p$) using the method of banded Gaussian elimination [8, Sec. 4.3]. Finally, when

$\mathbf{F}$ is lower triangular, as in Gauss-Siedel, step (ii) can be implemented by the method of back-substitution requiring n^2 flops.

3.1.2 Monotone Splitting (MS) Iterations

Assume that $\mathbf{F}_Y$ is positive definite and that one can find a positive definite $n \times n$ matrix $\mathbf{F}$ such that $\mathbf{F}$ dominates $\mathbf{F}_Y$ in the sense that $\mathbf{F} - \mathbf{F}_Y$ is non-negative definite. Now identify $\mathbf{N} = \mathbf{F} - \mathbf{F}_Y$ and note that $\mathbf{F}$ and $\mathbf{N}$ give a matrix splitting of $\mathbf{F}_Y$. Furthermore $\mathbf{N}$ is non-negative definite and all the eigenvalues of $\mathbf{F}^{-1}\mathbf{N} = \mathbf{I} - \mathbf{F}^{-1}\mathbf{F}_Y$ lie in the interval $[0, 1)$ [1]. Hence, using this choice of $\mathbf{F}$ and $\mathbf{N}$ in the general splitting iteration (4) we obtain an algorithm which gives an approximating sequence $\eta^{(k)}$ which converges monotonically to the CR bound from below. This monotonic splitting (MS) algorithm yields a sequence of increasingly tight bounds on $\mathrm{var}_{\underline{\theta}}(\hat{\theta}_1)$.

In [1] a diagonal splitting matrix, denoted $\mathbf{F}_{EM}$, was specified as the FIM associated with the complete data for a complete/incomplete data formulation. Since complete data is always at least as informatative as incomplete data $\mathbf{F}_{EM} \geq \mathbf{F}_Y$ and thus use of $\mathbf{F}_{EM}$ as the splitting matrix $\mathbf{F}$ in (4) yields a monotone MS algorithm.

If one is willing to sacrifice monotonic convergence, any monotonic MS algorithm can be accelerated by introducing a stationary (time invariant) over-relaxation parameter $\psi \in (1, 2)$ [10]. Specifically, assume that a dominating $\mathbf{F}$ has been found. An over-relaxed MS algorithm (MSOR) is specified by using $\mathbf{N} = \mathbf{F} - \psi\mathbf{F}_Y$ in the recursion (4). Note that the accelerated algorithm is generally non-monotonic since $\mathbf{N}$ can lose the property of non-negative definiteness when $\psi > 1$.

3.1.3 Convergence of Splitting Iterations

For the algorithm (4) to be computationally advantageous relative to direct Gaussian elimination methods we must ensure that the per iteration computational cost be low and that the speed of convergence be fast. Define the approximation error $\Delta\beta^{(k)} = \mathbf{F}_Y^{-1}\underline{e}_1 - \beta^{(k)}$ at the k-th iteration of the recursion (4). As previously mentioned, it is well known that the the error will asymptotically decrease to zero as long as $\rho(\mathbf{M}) < 1$ [9, 8] where $\mathbf{M} = \mathbf{F}^{-1}\mathbf{N}$. In this section we give a useful non-asymptotic convergence criterion which is used to derive an upper bound on the number of iterations required for convergence to within a user-specified tolerance.

From the general splitting iteration (4) it is easily shown that

$$\Delta\beta^{(k+1)} = \mathbf{M}^k \Delta\beta^{(0)},$$

where $\Delta\beta^{(0)} = \mathbf{F}_Y^{-1}\underline{e}_1$. Using properties of vector and matrix norms [8, Sec. 2.3]

$$\begin{aligned} \|\Delta\beta^{(k+1)}\|_2 &\leq |||\mathbf{M}^k|||_2 \cdot \|\Delta\beta^{(0)}\|_2 \\ &\leq |||\mathbf{M}|||_2^k \cdot \|\Delta\beta^{(0)}\|_2, \end{aligned} \tag{5}$$

where $\|\underline{u}\|_2 = \sqrt{\underline{u}^T\underline{u}}$ and $|||\mathbf{M}|||_2 = \max_{\underline{x}} \frac{\|\mathbf{M}\underline{x}\|_2}{\|\underline{x}\|_2}$ denote the euclidean norm and the matrix-2 norm, respectively. The matrix-2 norm of $\mathbf{M}$ is equal to the maximum magnitude singular value of $\mathbf{M}$.

Note that if the initial error $\Delta\beta^{(0)}$ is equal to zero subsequent errors remains zero. For non-zero $\Delta\beta^{(0)}$ we specify a convergence criterion in terms of a user specified threshold $\gamma \in [0, 1]$.

Definition *An iterative algorithm is said to have converged after N iterations (to within a factor $\gamma \in [0, 1]$) if*

$$\frac{\|\Delta\beta^{(k)}\|_2}{\|\Delta\beta^{(0)}\|_2} < \gamma, \quad \forall k \geq N. \tag{6}$$

The minimum integer $N = N(\gamma)$ such that this inequality is satisfied is called the required number of iterations for convergence.

Stopping the recursions at $k = N$ will ensure convergence to the CR bound within the factor γ. Now assume that the integer N satisfies $|||\mathbf{M}|||_2^{N-1} < \gamma < 1$. Then from inequality (5) $\frac{\|\Delta\beta^{(k)}\|_2}{\|\Delta\beta^{(0)}\|_2} < \gamma$ for all $k \geq N$. Thus we have the following upper bound on $N(\gamma)$

$$N(\gamma) \leq N^u(\gamma) = \frac{\log\gamma}{\log|||\mathbf{M}|||_2} + 1. \tag{7}$$

As γ approaches zero the required number of iterations $N(\gamma)$ goes to infinity and the root convergence factor can be directly related to the ratio $\|\Delta\beta^{(k)}\|_2/\|\Delta\beta^{(0)}\|_2$ through (5) and the identity $\rho(\mathbf{M}) = \lim_{N\to\infty}\left(|||\mathbf{M}^N|||_2\right)^{\frac{1}{N}}$ [8]. However, as pointed out in [10], while the root convergence factor is a good predictor of the asymptotic behavior of this ratio, the matrix-2 norm $|||\mathbf{M}|||_2$ is more reliable as a predictor of the actual number of iterations required for convergence. This will be quantitatively illustrated in Section 6.

If an accurate and simply computed estimate of $|||\mathbf{M}|||_2$ is available, e.g. from application of power iteration methods [8], then the function $N^u(\gamma)$ can be computed. In this case the relation (7) can be used to determine if the standard splitting algorithm (4) is likely to require fewer flops than Gaussian elimination. To illustrate, recall that for $\mathbf{F}$ a p-diagonal banded matrix each iteration of the splitting algorithm (4) requires $2n^2 + 2np^2$. Hence, for specified $\gamma \in [0, 1]$, the splitting algorithm (4) is guaranteed to require fewer flops than Gaussian elimination ($n^3/3$ flops or more) when

$$N^u(\gamma) < \frac{n^2}{n + p^2}\frac{3}{2}. \tag{8}$$

3.2 A Simple Class of Monotone Splitting Matrices

If monotone convergence of recursion (4) is to be ensured we need select dominating splitting matrix $\mathbf{F}$ such that

$\mathbf{N} = \mathbf{F} - \mathbf{F}_Y$ is non-negative definite. Furthermore, to promote fast convergence, $|||\mathbf{F}^{-1}\mathbf{N}|||_2$ should be small. This suggests posing the constrained optimization problem: find $\mathbf{F}$ to minimize $|||\mathbf{I} - \mathbf{F}^{-1}\mathbf{F}_Y|||_2$ subject to the constraint $\mathbf{F} - \mathbf{F}_Y \geq 0$. It can be simply verified that even for the case of diagonal $\mathbf{F}$ the optimum $\mathbf{F}$ is impractical to implement as it is a function of the eigenvalues of $\mathbf{F}_Y$.

As an alternative, consider the following p-diagonal splitting matrix $\mathbf{F} = \mathbf{D}_p$ defined in terms of the elements f_{ij} of the matrix $\mathbf{F}_Y$

$$\mathbf{D}_p = \mathbf{Q} + \mathrm{diag}\left(|\mathbf{F}_Y - \mathbf{Q}|\underline{1}\right). \tag{9}$$

where $\mathbf{Q} = ((f_{ij}))_{|i-j|<p}$ is a p-diagonal matrix, $|\mathbf{A}|$ denotes a matrix whose elements are the absolute values of those of the matrix $\mathbf{A}$, $\underline{1} = [1, \ldots, 1]^T$, and $\mathrm{diag}(\underline{x})$ is a diagonal matrix with the elements of the vector $\underline{x}$ along the diagonal. In particular, $\mathbf{D}_1$ is a diagonal matrix with i-th diagonal element $\sum_{j=1}^n |f_{ij}|$ and $\mathbf{D}_2$ is the tridiagonal matrix

$$D_2 = \begin{bmatrix} \sum_{j\neq 2}|f_{1j}| & f_{12} & & & \cdots & 0 \\ f_{21} & \sum_{j\neq 1,3}|f_{2j}| & \ddots & & & \vdots \\ & \ddots & \ddots & \ddots & & \\ & & \ddots & \sum_{j\neq k-1,k+1}|f_{kj}| & \ddots & \\ \vdots & & & \ddots & \ddots & f_{n-1,n} \\ 0 & \cdots & & & f_{n,n-1} & \sum_{j\neq n-1}|f_{nj}| \end{bmatrix}.$$

The following lemma follows directly from the diagonal dominance of the matrix $\mathbf{D}_p - \mathbf{F}_Y = \mathrm{diag}\left(|\mathbf{F}_Y - \mathbf{Q}|\underline{1}\right) - (\mathbf{F}_Y - \mathbf{Q}_p)$ [9, Corollary 7.2.1] and the easily verifiable fact that when $\mathbf{F}_Y$ has non-negative entries: $(\mathbf{D}_p - \mathbf{F}_Y)\underline{1} = 0$.

Lemma 1 *Assume that $\mathbf{F}_Y$ is an $n \times n$ symmetric matrix. Then $\mathbf{D}_p - \mathbf{F}_Y$ is non-negative definite. Furthermore, if $\mathbf{F}_Y$ has only non-negative entries then $\mathbf{D}_p - \mathbf{F}_Y$ has rank at most $n-1$.*

It can be shown that a necessary condition for $|||\mathbf{I} - \mathbf{F}^{-1}\mathbf{F}|||_2$ to be minimized by an $\mathbf{F}_Y$-dominating p-diagonal matrix $\mathbf{F} = \mathbf{D}_p$ is that $\mathbf{D}_p - \mathbf{F}$ be rank deficient. Lemma 1 asserts that for non-negative Fisher matrices $\mathbf{D}_p$ satisfies this condition. Non-negative Fisher matrices arise in many applications including the Poisson inverse problem to be considered in Sec. 6. Extensive numerical experiments have indicated that $\mathbf{F} = \mathbf{D}_p$ comes very close to minimizing $|||\mathbf{I} - \mathbf{F}^{-1}\mathbf{F}_Y|||_2$ subject to $\mathbf{D}_p - \mathbf{F}_Y \geq 0$.

4 Preconditioned Conjugate Gradient Algorithm

When the matrix $\mathbf{F}_Y$ in the equation $\mathbf{F}_Y\underline{x} = \underline{e}_1$ is square, symmetric, and positive definite, the preconditioned conjugate gradient (CG) algorithm can be used to approximate the solution $\underline{x}$ [8, Sec. 10.3]. Thus we can apply CG algorithms to approximating a column of the matrix CR bound via solution of (3). The CG algorithm can be interpreted as resulting from non-monotone and non-stationary acceleration of the splitting algorithm (4) via the introduction of time-varying acceleration parameters. [8, sec 10.3.6]. The CG algorithm converges to the exact solution in n iterations when run with infinite precision arithmetic. However, when run to termination it is not computationally competitive with Gaussian elimination. We will show that with proper preconditioning matrix $\mathbf{F}$ the following prematurely stopped preconditioned CG algorithm [8, Algorithm 10.3.1] can have significantly faster convergence than the monotone splitting iterations.

CG RECURSION FOR CR BOUND

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

RECURSION:

$$\text{Solve:} \quad \mathbf{F}\underline{z}^{(k)} = \underline{r}^{(k)}$$

$$\begin{aligned} \alpha^{(k)} &= \begin{cases} 0, & k = 0 \\ \frac{<\underline{r}^{(k)}, \underline{z}^{(k)}>}{<\underline{r}^{(k-1)}, \underline{z}^{(k-1)}>}, & k > 0 \end{cases} \\ \underline{p}^{(k)} &= \begin{cases} \underline{z}^{(0)}, & k = 0 \\ \underline{z}^{(k)} + \alpha^{(k)}\,\underline{p}^{(k-1)}, & k > 0 \end{cases} \\ \lambda^{(k)} &= \frac{<\underline{r}^{(k)}, \underline{z}^{(k)}>}{<\underline{p}^{(k)}, \mathbf{F}_Y\,\underline{p}^{(k)}>} \\ \underline{r}^{(k+1)} &= \underline{r}^{(k)} - \lambda^{(k)}\,\mathbf{F}_Y\,\underline{p}^{(k)} \\ \underline{\beta}^{(k+1)} &= \underline{\beta}^{(k)} + \lambda^{(k)}\,\underline{p}^{(k)} \end{aligned}$$

CR BOUND APPROXIMATION:

$$\eta^{(k)} = \underline{e}_1^T\,\underline{\beta}^{(k)}. \tag{10}$$

When the preconditioner $\mathbf{F}$ is a banded p-diagonal matrix the CG algorithm requires the same number of flops per iteration $(2n^2 + 2np^2)$ as the splitting algorithms previously described. In the CG algorithm $\underline{r}^{(k)}$ is the forward residual $\underline{r}^{(k)} = \underline{e}_1 - \mathbf{F}_Y\underline{\beta}^{(k)} = \mathbf{F}_Y\Delta\underline{\beta}^{(k)}$ where $\Delta\underline{beta}^{(k)}$ is the approximation error $\Delta\underline{\beta}^{(k)}\mathbf{F}_Y^{-1}\underline{e}_1 - \underline{\beta}^{(k)}$. While it has long been oberved that the speed of convergence of preconditioned CG generally improves as the eigenvalue spread of $\mathbf{F}^{-1}\mathbf{F}_Y$ decreases, only a bound on convergence rate of the norm of the residuals $||\underline{r}^{(k)}||_2 = ||\underline{\beta}^{(k)} - \mathbf{F}_Y^{-1}\underline{e}_1||_{F_Y}$ is available, where $||\underline{u}||^2_{F_Y} = \underline{u}^T\mathbf{F}_Y\underline{u}$.

From [4, Sec. 2.3.1] we have $||\Delta\underline{\beta}^{(k)}||_{F_Y} \leq 2||\Delta\underline{\beta}^{(0)}||_{F_Y}\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k$ where κ is the spectral condition number of the matrix $\mathbf{F}^{-1}\mathbf{F}_Y$, defined as the ratio of the largest to the smallest magnitude eigenvalues. If one adopts $\|\Delta\underline{\beta}^{(k)}\|_{F_Y}$ as an appropriate measure of approximation error then, by analogous reasoning as used for convergence analysis of the splitting algorithms, we obtain the following bound on the number of iterations $N(\gamma)$ necessary to guarantee $\|\Delta\underline{\beta}^{(k)}\|_{F_Y}$ converges to within a factor γ: $N^u(\gamma) = \frac{\log\gamma - \log(2)}{\log(\sqrt{\kappa}-1)-\log(\sqrt{\kappa}+1)}$. Combining this N^u expression with the per iteration cost of the algorithm (10) yields a similar inequality as in (8) which must be satisfied if preconditioned CG is to be computationally advantageous as compared to Gaussian elimination methods. Offline estimates of κ can be implemented ($O(n^2)$ additional flops) using power iterations or iterative sup-norm approximations to the maximum and minimum eigenvalues. [8, Sec. 3.5.4]. Alternatively, on-line estimates of these extremal eigenvalues can be updated in $O(1)$ flops from quantities produced in the preconditioned CG recursion [4, Sec. 5.1].

5 CRB Approximation for Singular FIM

The splitting iterations and CG algorithm described in the previous section are only applicable to non-singular Fisher information matrices $\mathbf{F}_Y$. If it is known that $\underline{e}_1$ lies in the range space of $\mathbf{F}_Y$ the CR bound $\underline{e}_1^T\mathbf{F}_Y^+\underline{e}_1$ for singular $\mathbf{F}_Y$ can be found in $4n^3/3$ flops using the QR factorization to solve for the min-norm solution $\underline{x}$ to $\mathbf{F}_Y\underline{x} = \underline{e}_1$ [8, Alg. 5.7.2]. However, typically the range space of $\mathbf{F}_Y$ is unknown and much more computationally intensive algorithms are required, e.g. the singular value decomposition (SVD) ($20n^3$ flops). Here we present an iterative approximation to the singular CR bound $\underline{e}_1^T\mathbf{F}_Y^+\underline{e}_1$.

Consider the following matrix difference

$$\mathbf{G}(\epsilon) \stackrel{def}{=} \frac{a}{a-b}(\mathbf{F}_Y + a\,\epsilon\mathbf{I})^{-1} - \frac{b}{a-b}(\mathbf{F}_Y + b\,\epsilon\mathbf{I})^{-1}, \quad (11)$$

where $\epsilon > 0$ and $a, b \in (0,\infty)$, $a \neq b$, are free parameters. $\mathbf{G}(\epsilon)$ is a convergent approximation to the pseudo-inverse of $\mathbf{F}_Y$ in the sense

$$\mathbf{F}_Y^+ = \lim_{\epsilon\to 0^+}\mathbf{G}(\epsilon). \quad (12)$$

The representation (12) can be easily established by considering the eigendecomposition of $\mathbf{F}_Y$

$$\begin{aligned}\mathbf{F}_Y^+ - \mathbf{G}(\epsilon) &= \sum_{i=1}^r \frac{1}{\sigma_i}\underline{u}_i\underline{u}_i^T - \sum_{i=1}^r \frac{a}{(a-b)(\sigma_i + a\epsilon)}\underline{u}_i\underline{u}_i^T \\ &\quad + \sum_{i=1}^r \frac{b}{(a-b)(\sigma_i + b\epsilon)}\underline{u}_i\underline{u}_i^T \\ &= \sum_{i=1}^r \frac{\epsilon\,(a\,\sigma_i + b\,\sigma_i + a\,b\,\epsilon)}{\sigma_i\,(\sigma_i + a\,\epsilon)\,(\sigma_i + b\,\epsilon)}\underline{u}_i\underline{u}_i^T, \quad (13)\end{aligned}$$

where $\sigma_1 \geq, ..., \geq \sigma_r > 0$ are the r non-zero eigenvalues of $\mathbf{F}_Y$ arranged in decreasing order, and $\{\underline{u}_i\}_{i=1}^n$ is an orthonormal set of eigenvectors. Observe that the range space of $\mathbf{G}(\epsilon)$ corresponds to the range space of $\mathbf{F}_Y$ for all $\epsilon > 0$.

Note that for all $\epsilon > 0$, $\mathbf{F}_Y^+ - \mathbf{G}(\epsilon) \geq 0$ so that $\underline{e}_1^T\mathbf{G}(\epsilon)\underline{e}_1$ is itself a valid lower bound on $\mathrm{var}_{\underline{\theta}}(\hat{\theta}_1)$ which converges to the CR bound $\underline{e}_1^T\mathbf{F}_Y^+\underline{e}_1$ as $\epsilon \to 0$. An iterative approximation to $\underline{e}_1^T\mathbf{G}(\epsilon)\underline{e}_1$ is obtained from the weighted difference of iterative approximations to the solutions of the linear equations $[\mathbf{F}_Y + a\epsilon\mathbf{I}]\gamma_1 = \underline{e}_1$ and $[\mathbf{F}_Y + b\epsilon\mathbf{I}]\gamma_2 = \underline{e}_1$. In particular, using the following assignment for (4) we obtain a splitting algorithm for approximating the singular CR bound

$$\begin{aligned}\mathbf{F}_1\underline{\gamma}_1^{(k+1)} &= [\mathbf{F}_1 - \mathbf{F}_Y - a\epsilon\mathbf{I}]\,\underline{\gamma}_1^{(k)} - \underline{e}_1 \\ \mathbf{F}_2\underline{\gamma}_2^{(k+1)} &= [\mathbf{F}_2 - \mathbf{F}_Y - b\epsilon\mathbf{I}]\,\underline{\gamma}_2^{(k)} - \underline{e}_1 \\ \underline{\beta}^{(k)} &= \frac{a}{a-b}\underline{\gamma}_1^{(k)} - \frac{b}{a-b}\underline{\gamma}_2^{(k)}. \quad (14)\end{aligned}$$

For large n the computational burden of the iteration (14) is $4n^2+2np^2$ flops which, for $p \ll n$, which is approximately twice as expensive as the non-singular CR bound iteration (4).

Both the speed of convergence and the normalized asymptotic approximation error $\delta = \underline{e}_1^T(\mathbf{F}_Y^+ - \mathbf{G}(\epsilon))\underline{e}_1/\underline{e}_1^T\mathbf{F}_Y^+\underline{e}_1$ increase as ϵ increases. It is easily shown that for $\epsilon \ll \min_{i=1,...,r}\{\sigma_i\}$: $\underline{e}_1^T(\mathbf{F}_Y^+ - \mathbf{G}(\epsilon))\underline{e}_1 = \epsilon\sum_{i=1}^r[(a+b)\,\sigma_i + a\,b\,\epsilon]/[\sigma_i\,(\sigma_i + a\,\epsilon)\,(\sigma_i + b\,\epsilon)]\|\underline{u}_i^T\underline{e}_1\|^2 \approx \epsilon(a+b)\|\mathbf{F}_Y^+\underline{e}_1\|^2$. Hence the normalized asymptotic approximation error is $\delta \approx \epsilon(a+b)\|\mathbf{F}_Y^+\underline{e}_1\|^2/\underline{e}_1^T\mathbf{F}_Y^+\underline{e}_1$. The right hand side of this relation can be bounded from above and below (see Appendix) to yield a bound on asymptotic normalized error δ valid for $\epsilon \ll \min_{i=1,...,r}\{\sigma_i\}$

$$\epsilon(a+b)\frac{f_{11}}{\sum_{i=1}^n f_{1i}^2} \leq \delta \leq \epsilon(a+b)\sum_{i=1}^n\frac{1}{f_{ii}}. \quad (15)$$

Note that the bounds depend on the magnitudes of ϵ, a, and b only through the product $\epsilon(a+b)$. In the following Section relation (15) will be used to select appropriate values of these parameters given a desired magnitude of normalized error δ via the equivalent relation

$$\frac{\delta}{(a+b)\sum_{f_{ii}\neq 0}\frac{1}{f_{ii}}} \leq \epsilon \leq \frac{\delta\sum_{j=1}^n f_{ij}}{f_{11}(a+b)}. \quad (16)$$

6 Application to ECT Inverse Problem

We will briefly illustrate the CR bound algorithms for the following inverse problem arising in collimated emission computed tomography (ECT). One wishes to estimate the first component θ_1 of the vector $\underline{\theta}$ in the linear model

$$\underline{Y} = \mathbf{A}\underline{\theta} + \underline{w}, \quad (17)$$

where $\mathbf{A}$ is an $m \times n$ matrix, and $\underline{w}$ is an additive noise. Here $\mathbf{A}$ is a sparse spatially variant system matrix with ρmn nonzero entries, $\rho \ll 1$, and $\underline{w}$ is a Poisson distributed noise. The vector $\underline{\theta}$ is a vector of unknown image intensity values indexed over n pixels, $\underline{Y}$ is a vector of photon counts detected at each of m detector elements, and the ij-th element of the matrix $\mathbf{A}$ are the probabilities that a photon emitted at pixel index j is detected at detector element i. Note that m is typically several times larger than n in tomography applications. Under the standard assumption that for each detector the mean detected counts $\underline{\mu} = \mathbf{A}\underline{\theta}$ are strictly greater than zero the FIM is non-singular and has the form [1]

$$\mathbf{F}_Y(\underline{\theta}) = \mathbf{A}^T \operatorname{diag}\left(\underline{\mu}\right)^{-1} \mathbf{A}. \tag{18}$$

Due to the simple product form (18) of $\mathbf{F}_Y$ the implementation of the generic CR bound algorithms presented previous sections can be simplified. In each of these algorithms the most intensive computation reduces to a vector matrix multiply of the form $\mathbf{A}^T \underline{x}^{(k)}$ (see algorithm descriptions in Fig. ??) requiring only on the order of $2mn\rho$ flops per iteration. As a result of this reduction the Fisher matrix $\mathbf{F}_Y$ need not be precomputed or stored. This is significant since the computation of $\mathbf{F}_Y$ requires $O(\rho mn^2)$ flops and $O(n^2)$ memory storage. Implementation of the iterative CR bound algorithms presented in previous sections only require precomputation of $\underline{\mu} = \mathbf{A}\underline{theta}$ ($2\rho mn$ flops) and storage of ρmn non-zero elements of $\mathbf{A}$ and m elements of $\underline{\mu}$. For matrix sparsity factors $\rho < 0.01$ commonly encountered in collimated ECT this corresponds to a substantial savings in storage and precomputation requirements.

For these studies the 32×32 Hoffman brain phantom shown in Fig. ??) was used as the true image. It was assumed that the overall size, location, and orientation of the brain within the image was known. Thus for computation of the bound the unknown parameter vector $\underline{\theta}$ consisted of a lexicographical ordering of the $n = 640$ pixel intensities within the ellipsoidal brain boundary. A system matrix $\mathbf{A}$ corresponding to axially collimated PET was constructed which acquires projections of the planar phantom over 40 detector angles and 80 radial detector bins. This gave a $\mathbf{A}$ with 3600 rows and 640 columns and sparsity factor $\rho = 0.1$.

While the CR bound approximation algorithms were investigated for many different pixels, we only present results for a representative case: the pixel located at coordinates $(16, 21)$ indicated on Fig. ??. It was found that the convergence trajectories of the CG and GS algorithms were relatively insensitive to the pixel location. Figs. 1-3 show the trajectories of various iterative CR bound approximation algorithms for this pixels.

Fig. 1 shows the monotone splitting algorithms which are mathematically guaranteed to improve the bound approximation at each successive iteration. MD1 denotes the

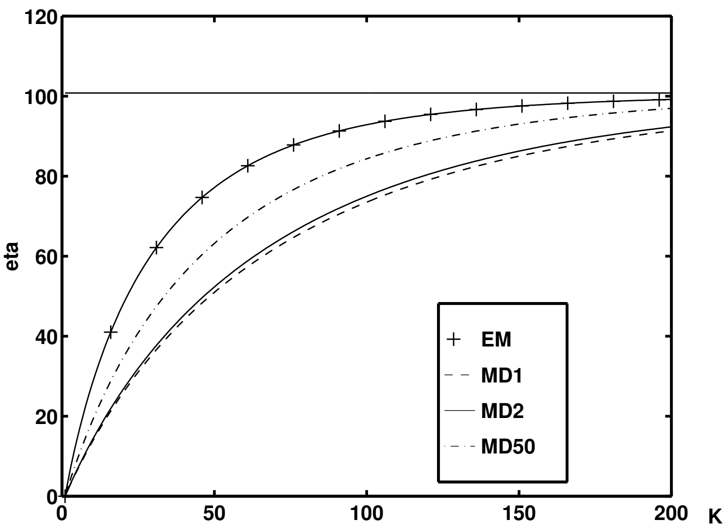

Figure 1: The monotone splitting algorithms for approximating the non-singular CR bound at pixel $(16, 21)$. Straight line denotes the true value of the CR bound. MD1 uses the diagonal splitting matrix $\mathbf{D}_1$ defined in Sec. ??; MD2 uses the tridiagonal splitting matrix $\mathbf{D}_2$; MD50 uses the 99-diagonal matrix $\mathbf{D}_{50}$; and EM uses the diagonal splitting matrix derived in [?]

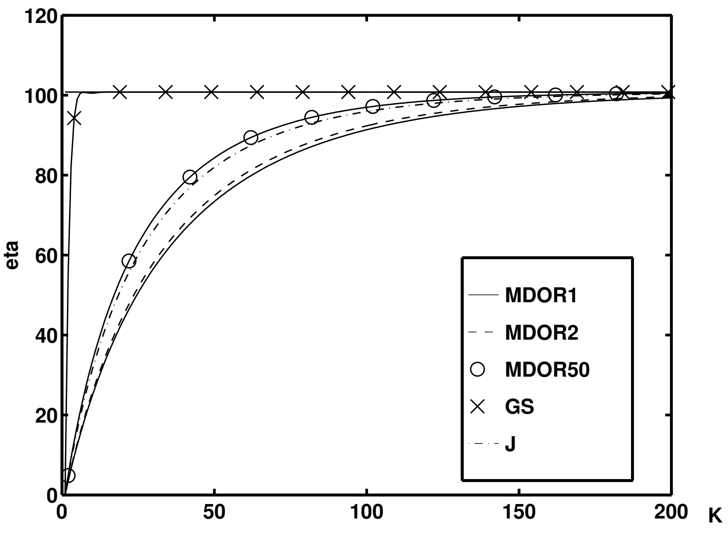

Figure 2: The non-monotone splitting algorithms for approximating the non-singular CR bound at pixel $(16, 21)$. MDOR1, MDOR2 and MDOR50 are over-relaxed versions of MD1, MD2 and MD50, where the over-relaxation parameter was numerically optimized to maximize asymptotic convergence. Also shown are the Jacobi (J) and Gauss-Seidel (GS) algorithms.

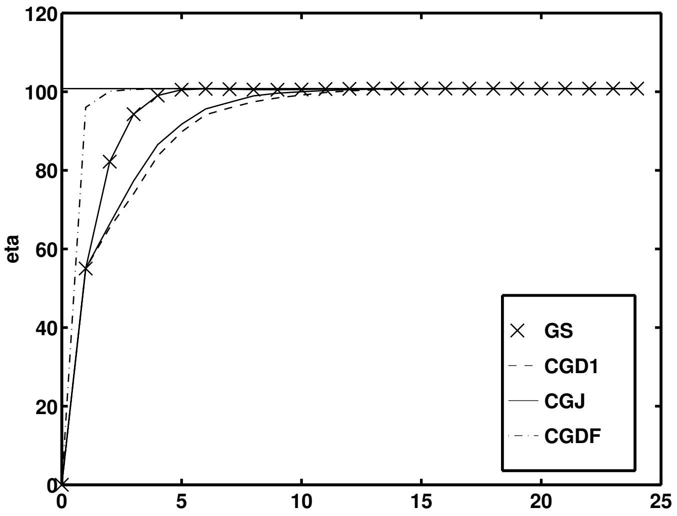

Figure 3: The Gauss-Seidel (GS) and preconditioned conjugate gradient algorithms for approximating the non-singular CR bound at pixel $(16, 21)$. CGJ uses the standard diagonal Jacobi preconditioner, CGD1 uses the diagonal matrix $\mathbf{D}_1$ preconditioner given in Sec. ??, and CGDF uses a preconditioner tailored for the $\mathbf{A}$ matrix arising in the spatially invariant ECT geometry considered here.

monotone algorithm which uses the diagonal splitting matrix $\mathbf{D}_1$ defined in Sec. ??; MD2 uses the tridiagonal splitting matrix $\mathbf{D}_2$; MD50 uses the 99-diagonal matrix $\mathbf{D}_{50}$; and EM uses the diagonal splitting matrix derived in [1]. As expected, increasing the number $2k-1$ of diagonals in the splitting matrix $\mathbf{D}_k$ improves the convergence rate. Figure 2 shows several splitting algorithms which use splitting matrices which do not guarantee monotonicity. In the figure MDOR1, MDOR2 and MDOR50 denote over-relaxed versions of MD1, MD2 and MD50, where the over-relaxation parameter was numerically optimized to maximize asymptotic convergence rate. Also shown are Gauss-Seidel (GS) and relaxed Jacobi (J), with relaxation parameter ψ numerically determined to ensure convergence and maximize asymptotic convergence rate. The standard unrelaxed Jacobi algorithm diverged for all cases studied and is not shown. Note that GS exhibits much faster convergence than any of the other splitting algorithms. Finally several preconditioned CG algorithms are shown in Figs. 3. In the figure the CGJ algorithm uses the standard diagonal Jacobi preconditioner, i.e. $\mathbf{F} = \text{diag}(\mathbf{F}_Y)$ the diagonal elements of the FIM, CGD1 uses the diagonal matrix $\mathbf{D}_1$ given in Sec. ??) as preconditioner. The algorithm denoted CGDF uses a special preconditioner $\mathbf{F}$ consisting of a diagonal preconditioner, which makes $\mathbf{F}_Y$ almost circulant, followed by a Fourier-type preconditioner. The preconditioner used in CGDF is tailored to the spatially invariant PET application and is described in [?] in the context of fast least squares PET reconstruction algorithms. Also plotted for comparison is the GS algorithm. All of the algorithms shown in In Figs. 3 and ?? have much faster convergence rate than the monotonic splitting algorithms; compare CGDF which converges to within 5% of the true CR bound in only 2 iterations while MD50 takes more than 200 iterations to converge to within the same accuracy. It is interesting to note that the GS and CG algorithms are almost monotonic, exhibiting only a very small amount of overshoot.

Convergence properties of the splitting algorithms are quantified in Table ?? and those of the conjugate gradient algorithms are illustrated in ??. The first 3 columns of Table ?? contain the matrix-2 norm $|||\Delta|||_2$, the simpler to compute Frobenius norm difference $|||\mathbf{F}-\mathbf{F}_Y|||_{\mathcal{F}}$, and the root convergence factor λ_{max}^{Δ} for the matrix $\mathbf{I}-\mathbf{F}^{-1}\mathbf{F}_Y$. The 4th column of Table ?? is the actual number of iterations and total number of required Mflops to convergence to within $\delta = 0.05$, or 5% of the limit. Note, as discussed in Section ?? the matrix-2 norm and Frobenius norms are better predictors of the actual time to converge than is the root convergence factor. In particular, observe that the Frobenius norm correctly predicts that MD1 converges in fewer iterations than the algorithm using the splitting matrix $\mathbf{F} = D_\lambda$ which has minimum root convergence factor.

The first column of Table ?? contains the ratio $\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}$ where κ is the spectral condition number of the preconditioned FIM $\mathbf{F}^{-1}\mathbf{F}_Y$. The second and third columns of Table ?? are the number of iterations and the number of flops required to converge within $\delta = 0.05$ of the limit.

Next we turn to the case of singular FIM arising in the so-called "missing angle problem." In this problem the system must estimate image parameters with a greatly reduced number and range of projections. For this study only 10 angles from 0 to $\pi/4$ and 40 radial bins per angle are available; leading to decimating $\mathbf{A}$ (3) by a factor of eight. This resulted in a $\mathbf{A}$ matrix of dimension 400×640 with rank 400 and condition number on the order of 1000. We implemented the perturbed differencing algorithm (14) to approximate the first column of the pseudo-inverse of $\mathbf{F}_Y$. It was found that over the range $0.001 \le \frac{a-b}{a+b} \le 1000$ the choice of a and b has only a minor effect on the convergence rate; here we only show results for $a-b = 0.0001$ which appeared to maximize the convergence rate of all algorithms. The maximum allowable asymptotic normalized error was set to $\delta = 0.05$ or 5% and epsilon was selected according to (16) as the average of the lower and upper bounds: $\epsilon = \frac{1}{2}\frac{\delta}{a+b}\left(\frac{1}{\sum_{f_{ii}\neq 0}\frac{1}{f_{ii}}} + \frac{\sum_{j=1}^{n} f_{ij}}{f_{11}}\right)$. Figure ?? illustrates the closeness of this selection rule to the true epsilon which exactly meets the constraint δ on asymptotic normalized error. Figure ?? illustrates the convergence trajectories of the GS and CG1 and CGD algorithms applied to the pair of iterations (14). It is interesting to note that while the CG algorithms converged most rapidly on their respective difference terms $\underline{\beta}_1^{(k)}$ and $\underline{\beta}_2^{(k)}$, after performing the differencing $\frac{a}{a-b}\underline{\beta}_1^{(k)} - \frac{b}{a-b}\underline{\beta}_2^{(k)}$, the GS algorithm converged the most rapidly. Finally, while space limitations prevent showing any numerical results, it was observed that a significant tradeoff exists between the convergence rate and the asymptotic normalized error δ.

7 Conclusion

We have studied the convergence properties of several algorithms for recursive approximation of the CR bound: including monotone convergent splitting matrix iterations; non-montonic over-relaxed splitting matrix iterations; and non-monotonic preconditioned conjugate gradient iterations. Not surprisingly, to obtain similar convergence rates the computational cost of a monotonically convergent algorithm is quite high as compared with a non-monotonic algorithm in the context of the inverse problems considered. The general conclusion of this study is that unless monotonic convergence is an absolute necessity the preconditioned conjugate gradient and Gauss-Seidel algorithms should be seriously considered for iterative CR bound computation due to their rapid convergence rates.

Appendix: Bounds on Asymptotic Error

Lemma 2 *Let δ be the normalized asymptotic error defined as $\delta = \underline{e}_1^T(\boldsymbol{F}_Y^+ - G(\epsilon))\underline{e}_1 / \underline{e}_1^T \boldsymbol{F}_Y^+ \underline{e}_1$ defined in Sec. 5. Assume that ϵ is dominated by the smallest positive eigenvalue of $\boldsymbol{F}$ so that to $o(\epsilon)$: $\delta = \epsilon(a+b)\frac{\underline{e}_1^T[\boldsymbol{F}_Y^+]^2\underline{e}_1}{\underline{e}_1^T \boldsymbol{F}_Y^+ \underline{e}_1}$. Then*

$$\epsilon(a+b)\frac{f_{11}}{\sum_{i=1}^n f_{1i}^2} \leq \delta \leq \epsilon(a+b)\sum_{f_{ii}\neq 0}\frac{1}{f_{ii}}, \tag{19}$$

Proof: We first show the lower inequality. If $f_{11} = 0$ then this inequality is trivial so hencforth we assume that $f_{11} > 0$. The Cauchy-Schwarz inequality states that $|\underline{u}^T\underline{v}|^2 \leq \underline{u}^T\underline{u}\cdot\underline{v}^T\underline{v}$ for any two vectors $\underline{u}$ and $\underline{v}$. Letting $\underline{u} = \mathbf{F}_Y^{k/2}\underline{w}$ and $\underline{v} = \mathbf{F}_Y^{k/2+1}\underline{w}$ we obtain

$$\frac{\underline{w}^T\mathbf{F}_Y^{k+1}\underline{w}}{\underline{w}^T\mathbf{F}_Y^{k+2}\underline{w}} \leq \frac{\underline{w}^T\mathbf{F}_Y^{k}\underline{w}}{\underline{w}^T\mathbf{F}_Y^{k+1}\underline{w}}.$$

Setting $\underline{w} = \mathbf{F}_Y^+\underline{e}_1$ and applying the above for $k = 0, 1, 2$ we obtain the sequence of inequalities

$$\frac{\underline{e}_1^T\mathbf{F}_Y\underline{e}_1}{\underline{e}_1^T\mathbf{F}_Y^2\underline{e}_1} \leq \frac{\underline{e}_1^T\mathcal{P}_{F_Y}\underline{e}_1}{\underline{e}_1^T\mathbf{F}_Y\underline{e}_1} \leq \frac{\underline{e}_1^T\mathbf{F}_Y^+\underline{e}_1}{\underline{e}_1^T\mathcal{P}_{F_Y}\underline{e}_1} \leq \frac{\underline{e}_1^T[\mathbf{F}_Y^+]^2\underline{e}_1}{\underline{e}_1^T\mathbf{F}_Y^+\underline{e}_1},$$

where $\mathcal{P}_{F_Y} = \mathbf{F}_Y^+\mathbf{F}_Y = \mathbf{F}_Y\mathbf{F}_Y^+$ is a symmetric idempotent matrix which projects vectors onto the column space of $\mathbf{F}_Y$. Since $\underline{e}_1^T\mathbf{F}_Y\underline{e}_1 = f_{11}$ and $\underline{e}_1^T\mathbf{F}_Y^2\underline{e}_1 = \|\mathbf{F}_Y\underline{e}_1\|^2 = \sum_{i=1}^n f_{1i}^2$ we have established the lower inequality in (19).

Now turn to the upper inequality in (19). Since $\sum_{k=1}^n \underline{e}_k\underline{e}_k^T = \mathbf{I}$ we have:

$$\begin{aligned}
\frac{\underline{e}_1^T[\mathbf{F}_Y^+]^2\underline{e}_1}{\underline{e}_1^T\mathbf{F}_Y^+\underline{e}_1} &= \frac{\sum_{k=1}^n|\underline{e}_1^T\mathbf{F}_Y^+\underline{e}_k|^2}{\underline{e}_1^T\mathbf{F}_Y^+\underline{e}_1} \\
&= \underline{e}_1^T\mathbf{F}_Y^+\underline{e}_1 + \sum_{k=2}^n\frac{|\underline{e}_1^T\mathbf{F}_Y^+\underline{e}_k|^2}{\underline{e}_1^T\mathbf{F}_Y^+\underline{e}_1} \\
&\leq \underline{e}_1^T\mathbf{F}_Y^+\underline{e}_1 + \sum_{k=2}^n\underline{e}_k^T\mathbf{F}_Y^+\underline{e}_k \\
&= \sum_{k=1}^n\underline{e}_k^T\mathbf{F}_Y^+\underline{e}_k,
\end{aligned} \tag{20}$$

where the Cauchy-Schwarz inequality was used in the third line. Let $\mathbf{P}_k$ be a permutation matrix such that $\mathbf{F}_k = \mathbf{P}_k^T\mathbf{F}_Y\mathbf{P}_k$ is the matrix resulting from exchanging the k-th and n-th rows and columns of $\mathbf{F}_Y$. Let $\mathbf{F}_k$ be partitioned as

$$\mathbf{F}_k = \begin{bmatrix} \mathbf{A} & \underline{c} \\ \underline{c}^T & \alpha \end{bmatrix},$$

where $\alpha = f_{kk}$ is the k-th diagonal element of $\mathbf{F}_Y$, $\mathbf{A}$ is an $(n-1)\times(n-1)$ matrix and $\underline{c}$ is an $(n-1)$-element column vector. Next we recall the formula for the pseudoinverse of partitioned non-negative definite matrices [11]

$$\mathbf{F}_k^+ = \begin{bmatrix} \mathbf{A}^+ + q^+\mathbf{A}^+\underline{c}\underline{c}^T\mathbf{A}^+ & -q^+\mathbf{A}^+\underline{c} \\ -q^+\underline{c}^T\mathbf{A}^+ & q^+ \end{bmatrix},$$

where $q = f_{kk} - \underline{c}^T\mathbf{A}^+\underline{c} \geq 0$. Hence $\underline{e}_k^T\mathbf{F}_Y^+\underline{e}_k = \underline{e}_n^T\mathbf{F}_k^+\underline{e}_n = q^+$. If $f_{kk} = 0$ then $q = q^+ = 0$. Otherwise, since $\mathbf{A}^+$ is non-negative definite, $q^+ = \frac{1}{f_{kk} - \underline{c}^T\mathbf{A}^+\underline{c}} \leq \frac{1}{f_{kk}}$. Plugging this into (20) we obtain the upper inequality of (19) which finishes the proof of the lemma. □

References

[1] A. O. Hero and J. A. Fessler, "A recursive algorithm for computing CR-type bounds on estimator covariance," *IEEE Trans. on Inform. Theory*, vol. 40, pp. 1205–1210, July 1994.

[2] A. R. Kuruc, "Lower bounds on multiple-source direction finding in the presence of direction-dependent antenna-array-calibration errors," Technical Report 799, M.I.T. Lincoln Laboratory, Oct., 1989.

[3] A. O. Hero and J. A. Fessler, "A fast recursive algorithm for computing CR-type bounds for image reconstruction problems," in *Proc. of IEEE Nuclear Science Symposium*, pp. 1188–1190, Orlando, FA, Oct. 1992.

[4] R. Barrett and etal, *Templates for the solution of linear systems: building blocks for iterative methods*, SIAM Press, Philadelphia, 1994.

[5] C. R. Rao, *Linear Statistical Inference and Its Applications*, Wiley, New York, 1973.

[6] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Trans. on Inform. Theory*, vol. IT-36, pp. 1285–1301, Nov. 1990.

[7] F. A. Graybill, *Matrices with Applications in Statistics*, Wadsworth Publishing Co., Belmont CA, 1983.

[8] G. H. Golub and C. F. Van Loan, *Matrix Computations (2nd Edition)*, The Johns Hopkins University Press, Baltimore, 1989.

[9] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge, 1985.

[10] D. M. Young, *Iterative solution of large linear systems*, Academic Press, New York, 1971.

[11] C. A. Rohde, "Generalized inverses of partitioned matrices," *J. Soc. Indust. Appl. Math. (SIAM)*, vol. 13, pp. 1033–1035, 1965.