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

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

Technical Report No. 305, Nov. 1996

Communication and Signal Processing Laboratory (CSPL)
Department of EECS
The University of Michigan
Ann Arbor 48109-2122

[1]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

Abstract

Computation of the Cramer-Rao bound (CRB) on estimator variance 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 iterative methods for approximating the CRB using matrix splitting and preconditioned conjugate gradient algorithms. We specify matrix splitting algorithms which have a monotonic property: the approximation sequence increases at each iteration and converges to the CRB from below. Non-monotone Gauss-Seidel and preconditioned conjugate gradient algorithms converge significantly faster than the monotone algorithms for comparable computational load. We illustrate the methods developed in this paper for an important class of singular and non-singular inverse problems.
Key Words: Performance bounds, multi-dimensional parameter estimation, monotone matrix splitting iterations, Gauss-Seidel, preconditioned conjugate gradient.

1 Introduction

The Cramer-Rao (CR) bound is a very widely used lower bound on estimator covariance. When there are n unknown parameters the calculation of the CR bound involves inversion 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 analysis the pixel intensities are the unknown parameters and a moderate size image of $128 \times 128 = 16384$ pixels has a FIM of dimension 16384×16384, requiring 268 Megabytes memory storage (single precision) and 4 Tera flops (4×10^{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 analysis one may be primarily interested in a small q-pixel region of interest (ROI) corresponding to a tumor or lesion. In this case methods of sequential partitioning [1] can be used 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 direct full matrix inversion, this algorithm still requires the same order $O(n^3)$ of flops.

In [2] and [3] a recursive method was presented for approximating columns of the CR bound for unbiased estimation of an element of the parameter vector and for non-singular FIM. This method was derived by specifying a FIM-dominating matrix which allowed a geometric series decomposition of the inverse FIM. This FIM-dominating matrix was specified by exploiting special properties of the Fisher information in a complete-incomplete data reformulation of the estimation problem. The geometric series algorithm requires only $O(n^2)$ flops per iteration per parameter so that if convergence 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. As will be shown in this paper, the price of monotonicity is slow convergence.

In this paper we place the method of [2] in the setting of a general class of iterative algorithms, known as stationary and non-stationary linear equation solvers [4]. Using this setting we develop rapidly convergent CR bound approximation methods which can be applied to the cases of biased parameter estimation, estimation of a function of the parameters, and singular FIM. The following iterative equation solvers are considered: monotone and non-monotone matrix splitting algorithms, such as the method of [2], Jacobi, Gauss-Seidel, and preconditioned conjugate gradient algorithms. The extension of these algorithms to singular FIM is achieved by using matrix perturbation. We illustrate these algorithms for an important class of inverse problems arising in deconvolution, image restoration, and tomographic reconstruction.

The main conclusions of this paper are: 1) iterative equation solving methods are effective for approximating the CRB on estimators of any scalar function of the parameters; 2) for sparse-matrix inverse problems these methods can be implemented with significant savings in memory and computation load; 3) if monotonicity can be sacrificed for the user's application, the non-monotone Gauss-Seidel and pre-conditioned conjugate gradient methods should be implemented due to their advantage of very rapid convergence.

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 $\mathbb{R}^n$. Define the $n \times n$ Fisher information matrix (FIM)

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

where $\nabla_{\underline{\theta}}$ denotes the gradient operator (a column vector). Let $t = t(\underline{\theta})$ be a known scalar function of the unknown parameter vector and let $\hat{t} = \hat{t}(Y)$ be an arbitrary estimator of $t(\underline{\theta})$ having known mean function $m(\underline{\theta}) = E_{\underline{\theta}}[\hat{t}]$.

The CR lower bound on the variance of the estimator $\hat{t}(Y)$ is [5, 6, 7]

$$\mathrm{var}_{\underline{\theta}}(\hat{t}) \geq \underline{\dot{\mathrm{m}}}^T \mathbf{F}_Y^+(\underline{\theta}) \underline{\dot{\mathrm{m}}}, \tag{1}$$

where $\underline{\dot{\mathrm{m}}} = \nabla_{\underline{\theta}} m$ is the column gradient vector $[\frac{\partial m}{\partial \theta_1}, ..., \frac{\partial m}{\partial \theta_n}]^T$, and $\mathbf{F}_Y^+$ denotes the Moore-Penrose pseudo-inverse [8]. When $\mathbf{F}_Y$ is non-singular $\mathbf{F}_Y^+ = \mathbf{F}_Y^{-1}$ the ordinary matrix inverse. Note that the pseudo-inverse form of the CR bound is generally not generally achievable unless the $\underline{\dot{\mathrm{m}}}$ lies in the rangespace of $\mathbf{F}_Y$ [5].

Throughout this paper we will be interested in calculating the right hand side of (1). The results are easily extended to calculation of the uniform lower bound presented in [5] which applies to estimators with unknown mean function $m(\underline{\theta})$. The results are easily extended to estimation of vector-valued functions $[t_1(\underline{\theta}), \ldots, t_q(\underline{\theta})]^T$, $(q \ll n)$ by repeated application of the algorithms given here.

Direct computation of the right hand side of the CR inequality (1) can be accomplished in $O(n^3)$ flops by solving for $\underline{x}$ in the equation

$$\mathbf{F}_Y \ \underline{x} = \underline{\dot{\mathrm{m}}}. \tag{2}$$

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 the QR and Cholesky decompositions ($n^3/3$ flops) [9, Sec. 4.2].

3 Recursive CR Bound Algorithms for Non-Singular FIM

Here we describe the monotonically convergent algorithm of [2] in the context of standard splitting iterations [9, Sec. 10.1], also known as stationary iterations [4], for approximating the solution to the linear equation (2).

3.1 Splitting Algorithms

Let $\mathbf{F}$ and $\mathbf{N}$ be $n \times n$ matrices which split $\mathbf{F}_Y$ in the sense that $\mathbf{F} - \mathbf{N} = \mathbf{F}_Y$. The matrix $\mathbf{F}$ is called a preconditioning matrix and is assumed to be non-singular.

3.1.1 General Splitting Iterations

The following *general splitting iteration for approximation of the CR bound* requires an initial vector $\underline{\beta}^{(0)}$

$$\begin{array}{lll} i). & & \underline{u} = \mathbf{N}\underline{\beta}^{(k)} + \underline{\dot{\mathrm{m}}} \\ ii). & \texttt{Solve :} & \mathbf{F}\underline{\beta}^{(k+1)} = \underline{u} \\ & & \\ & & \eta^{(k+1)} = \underline{\dot{\mathrm{m}}}^T \ \underline{\beta}^{(k+1)} \quad \text{(CRB APPROX)} \end{array} \tag{3}$$

For any square matrix $\mathbf{M}$ define the root convergence factor $\rho(\mathbf{M})$as the maximum magnitude eigenvalue of $\mathbf{M}$, also known as the spectral radius of $\mathbf{M}$. It is well known [10, 9] that if $\rho(\mathbf{F}^{-1}\mathbf{N}) < 1$ then $\underline{\beta}^{(k)}$ converges to the vector $\mathbf{F}_Y^{-1}\underline{\dot{\mathrm{m}}}$ and the approximating sequence $\eta^{(k)}$ converges to the CR bound $\underline{\dot{\mathrm{m}}}^T\mathbf{F}_Y^{-1}\underline{\dot{\mathrm{m}}}$.

Many well known algorithms fall into the category of general splitting iterations, such as the Jacobi (J) and Gauss-Siedel (GS) iterations [9, Sec. 10.1]. 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 triangular matrices with zero diagonal entries. The J iteration is obtained by making the identifications $\mathbf{F} = \mathbf{D}$, $\mathbf{N} = -(\mathbf{U} + \mathbf{L})$ in (3). For J iterations, the spectral radius of $\mathbf{F}^{-1}\mathbf{N}$ may exceed one and $\mathbf{N}$ is not generally non-negative definite. Therefore, the J iterations may not converge and are generally not monotone. In general to ensure convergence the Jacobi algorithm must be relaxed, corresponding to using $\mathbf{N} = (1 - \psi)\mathbf{D} - \psi(\mathbf{U} + \mathbf{L})$ in place of $-(\mathbf{U} + \mathbf{L})$ where $\psi \in (1, 2)$ is an over-relaxation parameter [11]. The GS iteration is obtained from the general splitting iteration by identifying $\mathbf{F} = \mathbf{D} + \mathbf{L}$ and $\mathbf{N} = -\mathbf{U}$ in (3). Like J iterations the GS iterations yield non-monotonic approximations. On the other hand, the GS iterations always converge for positive definite FIM. Step (i) of (3) requires $2n^2$ flops while step (ii) requires a number of flops depending on the specific form of the matrix $\mathbf{F}$. When $\mathbf{F}$ is diagonal, as in J iterations, step (ii) requires n flops. For GS the matrix $\mathbf{F}$ is lower triangular and step (ii) of (3) could be accomplished using backsubstitution (n^2 flops). However, GS is never implemented in this way since, by rearranging the order of computation, steps (i) and (ii) of (3) can be accomplished in only $2n^2$ flops (large n) via the equivalent iteration:

$$
\begin{array}{l}
\texttt{for } j = 1 \texttt{ to } M \quad \text{(GS Iteration)} \\
\quad r = (\underline{\dot{\mathrm{m}}}_j - f_{j*}\underline{\beta}^{(k)})/f_{jj} \\
\quad \beta_j^{(k)} = r \\
\quad \beta_{[j]_n+1}^{(k)} = 0 \\
\texttt{end}
\end{array}
$$

where $[j]_n = j \bmod n$, f_{ij} denotes the ij-th element of $\mathbf{F}_Y$ and f_{j*} denotes the j-th row of $\mathbf{F}_Y$.

3.1.2 Monotone Splitting (MS) Iterations

Assume that $\mathbf{F}$ is symmetric positive definite and $\mathbf{N} = \mathbf{F} - \mathbf{F}_Y$ is symmetric non-negative definite. Then it can be shown in an analogous manner to [2, Eq. (9)] that $\eta^{(k+1)} - \eta^{(k)} = \underline{\dot{\mathrm{m}}}^T \mathbf{F}^{-\frac{1}{2}} [\mathbf{F}^{-\frac{1}{2}} \mathbf{N} \mathbf{F}^{-\frac{1}{2}}]^k \mathbf{F}^{-\frac{1}{2}} \underline{\dot{\mathrm{m}}} \geq 0$. Assume also that the splitting algorithm (3) is initialized with a vector $\underline{\beta}^{(0)}$ such that $\eta^{(0)} = \underline{\dot{\mathrm{m}}}^T \underline{\beta}^{(0)}$ is less than or equal to the CRB $\underline{\dot{\mathrm{m}}}^T \mathbf{F}_Y^{-1} \underline{\dot{\mathrm{m}}}$. This can be accomplished by setting $\underline{\beta}^{(0)} = \underline{0}$. 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. Such a monotonic splitting (MS) algorithm yields a sequence of increasingly tight bounds on $\mathrm{var}_{\underline{\theta}}(\hat{t})$.

We can ensure that $\rho(\mathbf{F}^{-1}\mathbf{N}) = \rho(\mathbf{I} - \mathbf{F}^{-1}\mathbf{F}_Y) < 1$ by selecting a preconditioning matrix $\mathbf{F}$ which dominates $\mathbf{F}_Y$ in the sense that $\mathbf{F} - \mathbf{F}_Y$ is non-negative definite [2]. In [2] a diagonal preconditioning matrix, denoted $\mathbf{F}_{EM}$, was specified as the FIM associated with the complete data for a complete/incomplete data formulation of the estimation problem. Since complete data is always at least as informative as incomplete data $\mathbf{F}_{EM} \geq \mathbf{F}_Y$ and thus use of $\mathbf{F}_{EM}$ as the preconditioning matrix $\mathbf{F}$ in (3) yields a monotone MS algorithm.

We next present a general class of FIM-dominating preconditioning matrices $\mathbf{F}$ which ensure monotone convergence under the initialization condition $\underline{\beta}^{(0)} = \underline{0}$. Define the $(2p-1)$-diagonal matrix $\mathbf{D}_p$ in terms of the elements f_{ij} of the matrix $\overline{\mathbf{F}}_Y$

$$\mathbf{D}_p = \mathbf{Q} + \operatorname{diag}\left(|\mathbf{F}_Y - \mathbf{Q}|\,\underline{1}\right). \tag{4}$$

where $\mathbf{Q} = ((f_{ij}))_{|i-j|<p}$ is a $(2p-1)$-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 $\operatorname{diag}(\underline{x})$ 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

$$\mathbf{D}_2 = \begin{bmatrix} \sum_{j\neq 2}|f_{1j}| & f_{12} & & & \cdots & 0 \\ f_{21} & \sum_{j\neq 1,3}|f_{2j}| & \ddots & & & \vdots \\ & \ddots & \ddots & \ddots & & \\ & & \ddots & \sum_{j\neq k-1,k+1}|f_{kj}| & \ddots & \\ \vdots & & & \ddots & \ddots & f_{n-1,n} \\ 0 & \cdots & & & f_{n,n-1} & \sum_{j\neq n-1}|f_{nj}| \end{bmatrix}.$$

The following lemma follows directly from the diagonal dominance of the matrix $\mathbf{D}_p - \mathbf{F}_Y = \operatorname{diag}\left(|\mathbf{F}_Y - \mathbf{Q}|\,\underline{1}\right) - (\mathbf{F}_Y - \mathbf{Q}_p)$ [10, Corollary 7.2.1] and the easily verifiable fact that when $\mathbf{F}_Y$ has non-negative entries: $(\mathbf{D}_p - \mathbf{F}_Y)\underline{1} = 0$. We give a direct proof in Appendix B.

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 that an $\mathbf{F}_Y$-dominating $(2p-1)$-diagonal matrix $\mathbf{F} = \mathbf{D}_p$ minimize the root convergence factor $\rho(\mathbf{I} - \mathbf{F}^{-1}\mathbf{F})$, and thus maximize rate of convergence, is that $\mathbf{D}_p - \mathbf{F}_Y$ be rank deficient. Lemma 1 asserts that for Fisher matrices with non-negative entries $\mathbf{D}_p$ satisfies this condition. Such Fisher matrices arise in many applications including the inverse problems to be considered in Sec. 6. A large number of numerical experiments indicate that $\mathbf{F} = \mathbf{D}_p$ comes very close to minimizing the root convergence factor over all $(2p-1)$-diagonal preconditioning matrices $\mathbf{F}$ satisfying $\mathbf{F} - \mathbf{F}_Y \geq 0$.

4 Preconditioned Conjugate Gradient Algorithm

When the FIM $\mathbf{F}_Y$ in the linear equation (2) is positive definite, the preconditioned conjugate gradient (CG) algorithm can be used to approximate the solution $\underline{x}$ [9, Sec. 10.3] and we obtain an approximation to the CR bound $\underline{\dot{\mathrm{m}}}^T \underline{x}$. The CG algorithm can be interpreted as resulting from non-monotone and non-stationary acceleration of the splitting algorithm (3) via the introduction of time-varying acceleration parameters [9, sec 10.3.6]. The CG algorithm converges to the exact solution $\underline{x}$ 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 [9, Algorithm 10.3.1] is quite competitive with direct methods and has significantly faster convergence than MS iterations.

4.1 Preconditioned CG Recursion for CRB

The following preconditioned CG iteration requires initialization of $\underline{\beta}^{(0)}$ and $\underline{r}^{(0)} = \underline{\dot{\mathrm{m}}} - \mathbf{F}_Y \underline{\beta}^{(0)}$:

$$
\begin{aligned}
\texttt{Solve:} \quad & \mathbf{F}\underline{z}^{(k)} = \underline{r}^{(k)} \\
\alpha^{(k)} &= \begin{cases} 0, & k = 0 \\ \frac{<\underline{r}^{(k)}, \underline{z}^{(k)}>}{<\underline{r}^{(k-1)}, \underline{z}^{(k-1)}>}, & k > 0 \end{cases} \\
\underline{p}^{(k)} &= \begin{cases} \underline{z}^{(0)}, & k = 0 \\ \underline{z}^{(k)} + \alpha^{(k)}\, \underline{p}^{(k-1)}, & k > 0 \end{cases} \\
\lambda^{(k)} &= \frac{<\underline{r}^{(k)}, \underline{z}^{(k)}>}{<\underline{p}^{(k)}, \mathbf{F}_Y\, \underline{p}^{(k)}>} \\
\underline{r}^{(k+1)} &= \underline{r}^{(k)} - \lambda^{(k)}\, \mathbf{F}_Y\, \underline{p}^{(k)} \\
\underline{\beta}^{(k+1)} &= \underline{\beta}^{(k)} + \lambda^{(k)}\, \underline{p}^{(k)} \\
\eta^{(k+1)} &= \underline{\dot{\mathrm{m}}}^T\, \underline{\beta}^{(k+1)} \qquad \text{(CRB APPROXIMATION)}
\end{aligned}
$$

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 recursion $\underline{r}^{(k)}$ is the forward residual $\underline{r}^{(k)} = \underline{\dot{\mathrm{m}}} - \mathbf{F}_Y \underline{\beta}^{(k)} = \mathbf{F}_Y \Delta \underline{\beta}^{(k)}$ where $\Delta \underline{\beta}^{(k)}$ is the approximation error $\Delta \underline{\beta}^{(k)} = \mathbf{F}_Y^{-1} \underline{\dot{\mathrm{m}}} - \underline{\beta}^{(k)}$. While it has long been observed 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 the asymptotic rate of decrease of the norm of the residuals $\|\underline{r}^{(k)}\|_2 = \|\underline{\beta}^{(k)} - \mathbf{F}_Y^{-1}\underline{\dot{\mathrm{m}}}\|_{F_Y}$ is available, where $\|\underline{u}\|^2_{F_Y} = \underline{u}^T \mathbf{F}_Y \underline{u}$ [4, Sec. 2.3.1]. This relation is: $\|\Delta \underline{\beta}^{(k)}\|_{F_Y} \leq 2\|\Delta \underline{\beta}^{(0)}\|_{F_Y} \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k$, 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 of the matrix.

5 CRB Approximation for Singular FIM

The splitting iterations and CG algorithm described in the previous sections are only applicable to non-singular FIM $\mathbf{F}_Y$. One simple case is easily handled. If the result of the operation $\mathbf{F}_Y \underline{\dot{m}}$ (requires only $2n^2$ flops) returns zero then we know that $\underline{\dot{m}}$ lies in the null-space of $\mathbf{F}_Y$ in which case the CRB is identically zero. On the other hand, if it is known that $\underline{\dot{m}}$ lies in the range space of $\mathbf{F}_Y$ the CR bound $\underline{\dot{m}}^T \mathbf{F}_Y^{+} \underline{\dot{m}}$ 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 $\mathbf{F}_Y \underline{x} = \underline{\dot{m}}$ [9, 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 pseudo-inverse form of the CR bound for tha case of singular FIM and $\mathbf{F}_Y \underline{\dot{m}} \neq 0$.

Consider the following matrix

$$\begin{aligned} \mathbf{G}(\epsilon) &\stackrel{def}{=} \frac{a}{a-b}(\mathbf{F}_Y + a\,\epsilon \mathbf{I})^{-1} - \frac{b}{a-b}(\mathbf{F}_Y + b\,\epsilon \mathbf{I})^{-1} \\ &= (\mathbf{F}_Y + a\,\epsilon \mathbf{I})^{-1} \mathbf{F}_Y (\mathbf{F}_Y + b\,\epsilon \mathbf{I})^{-1}, \end{aligned} \tag{5}$$

where $\epsilon > 0$ and $a, b \in (0, \infty)$. Note that for the form (5) it is necessary to restrict a and b to satisfy $a \neq b$, while this is not necessary for the form (6).

$\mathbf{G}(\epsilon)$ is a convergent approximation to the pseudo-inverse of $\mathbf{F}_Y$ in the sense

$$\mathbf{F}_Y^{+} = \lim_{\epsilon \to 0^+} \mathbf{G}(\epsilon). \tag{6}$$

The representation (6) can be easily established by considering the eigendecomposition of $\mathbf{F}_Y$

$$\begin{aligned} \mathbf{F}_Y^{+} - \mathbf{G}(\epsilon) &= \sum_{i=1}^{r} \frac{1}{\sigma_i} \underline{u}_i \underline{u}_i^T - \sum_{i=1}^{r} \frac{\sigma_i}{(\sigma_i + a\epsilon)(\sigma_i + b\epsilon)} \underline{u}_i \underline{u}_i^T \\ &= \sum_{i=1}^{r} \frac{\epsilon\,(\sigma_i(a+b) + a\,b\,\epsilon)}{\sigma_i\,(\sigma_i + a\,\epsilon)\,(\sigma_i + b\,\epsilon)} \underline{u}_i \underline{u}_i^T, \end{aligned} \tag{7}$$

where $\sigma_1 \geq, \ldots, \geq \sigma_r > 0$ 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 $\underline{\dot{m}}^T \mathbf{G}(\epsilon) \underline{\dot{m}} \geq \underline{\dot{m}}^T \mathbf{F}_Y^{+} \underline{\dot{m}}$. Hence $\underline{\dot{m}}^T \mathbf{G}(\epsilon) \underline{\dot{m}}$ is a valid lower bound on $\mathrm{var}_{\underline{\theta}}(\hat{t})$ which converges to the CR bound $\underline{\dot{m}}^T \mathbf{F}_Y^{+} \underline{\dot{m}}$ as $\epsilon \to 0$. In view of (6) an iterative approximation to $\underline{\dot{m}}^T \mathbf{G}(\epsilon) \underline{\dot{m}}$ can be obtained from applying iterative algorithms to approximate each of the solutions $\underline{\gamma}_1$ and $\underline{\gamma}_2$ to the linear equations $[\mathbf{F}_Y + a\epsilon \mathbf{I}]\underline{\gamma}_1 = \underline{\dot{m}}$ and $[\mathbf{F}_Y + b\epsilon \mathbf{I}]\underline{\gamma}_2 = \underline{\dot{m}}$, respectively. Since the perturbed matrices in these equations are non-singular the CG, GS and other previously discussed algorithms can be applied. For example, the following is a splitting algorithm for approximating the singular CR bound when $a \neq b$:

$$\begin{aligned} \mathbf{F}_1 \underline{\gamma}_1^{(k+1)} &= [\mathbf{F}_1 - \mathbf{F}_Y - a\epsilon \mathbf{I}]\, \underline{\gamma}_1^{(k)} - \underline{\dot{m}} \\ \mathbf{F}_2 \underline{\gamma}_2^{(k+1)} &= [\mathbf{F}_2 - \mathbf{F}_Y - b\epsilon \mathbf{I}]\, \underline{\gamma}_2^{(k)} - \underline{\dot{m}} \\ \eta^{(k)} &= [\underline{\gamma}_1^{(k)}]^T \mathbf{F}_y [\underline{\gamma}_2^{(k)}] \quad \text{(CRB APPROX)} \end{aligned} \tag{8}$$

where $\mathbf{F}_1$ and $\mathbf{F}_2$ are suitable preconditioning matrices. The singular splitting iteration (8) requires approximately three times as much computation as the non-singular CR bound splitting iteration (3). When $a = b$ one of the two iterative equation solvers in (8) can be eliminated, leading to $\underline{\gamma}_1 = \underline{\gamma}_2$ in (8) and the computational burden becomes only twice as expensive as the non-singular CR bound splitting iteration.

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

$$\epsilon(a+b)\frac{\underline{\dot{m}}^T \mathbf{F}_Y \underline{\dot{m}}}{||\mathbf{F}_Y \underline{\dot{m}}||^2} \leq \delta \leq \frac{\epsilon(a+b)}{\sigma_{min}}. \tag{9}$$

Note that the upper and lower bounds depend on the magnitudes of ϵ, a, and b only through the product $\epsilon(a+b)$. In the following Section, relation (9) will be used to select appropriate values of these free parameters to attain a desired magnitude of asymptotic normalized error.

6 Application to an Inverse Problem

We will briefly illustrate the iterative CR bound approximations for the following inverse problem arising in emission computed tomography (ECT). A subject is injected with a radiopharmaceutical which is taken up by a target object or organ where it emits photons with intensity proportional to the spatial distribution of the isotope. This spatial intensity is described by a vector $\underline{\theta}$ of non-negative values indexed over n pixels or voxels. One wishes to estimate the uptake $t(\underline{\theta})$ of the radiopharmaceutical over a specified region within the subject based on a set of noisy projections $\underline{Y}$ of $\underline{\theta}$ onto m detector surfaces. The uptake is mathematically defined as the sum of the object intensities $\underline{\theta}$ over this region: $t(\underline{\theta}) = \sum_{region} \theta_i$. The detector measurements $\underline{Y}$ are distributed as independent Poisson random variables with mean vector $E_{\underline{\theta}}[\underline{Y}] = \underline{\mu} = \mathbf{A}\underline{\theta} + \underline{w}$. The $m \times n$ matrix $\mathbf{A}$ has non-negative entries which correspond to probabilities that a given detector detects a photon emitted from a given pixel. This matrix is typically sparse with $mn\nu$ nonzero entries, $\nu \ll 1$ being the matrix sparsity factor. The vector $\underline{w} = \alpha \cdot \underline{1}$ is a column of scalars $\alpha > 0$ which correspond to a spatially homogeneous noise background level. Under the standard assumption that for each detector the mean number of detected photons are strictly greater than zero the FIM is non-singular and has the form [2]

$$\mathbf{F}_Y(\underline{\theta}) \quad = \quad \mathbf{A}^T\, \mathbf{D}^{-1}\, \mathbf{A}. \tag{10}$$

where $\mathbf{D} = \mathrm{diag}\left(\underline{\mu}\right)$ is a diagonal matrix containing the mean number of detected photons at each detector.

In all of the algorithms that we have discussed the computational bottlenecks occur in computing the vector-matrix product $\mathbf{F}_Y\underline{\beta}^{(k)}$ ($2n^2$ flops) and, for non diagonal $\mathbf{F}$, in computing the solution to the preconditioning equation $\mathbf{F}\underline{\beta}^{(k)} = \underline{u}$ ($np^2 + 8np$ flops for $\mathbf{F} = \mathbf{D}_p$ a p-diagonal preconditioner [9, Sec. 4.3.6]). However due to the simple sparse product form (10) of $\mathbf{F}_Y$ the vector-matrix multiplication can be accomplished in two nested vector-matrix multiplications: $\mathbf{F}_Y\underline{\beta}^{(k)} = \mathbf{A}^T\left[\mathbf{D}^{-1}\mathbf{A}\underline{\beta}^{(k)}\right]$ ($4mn\nu$ flops). Another result of these simplifications is that the Fisher matrix $\mathbf{F}_Y$ need not be precomputed or stored. This is significant since the computation of $\mathbf{F}_Y$ requires $2mn^2\nu$ flops and n^2 memory storage (typically $\mathbf{F}_Y$ does not inherit sparseness of $\mathbf{A}$). On the other hand, implementation of the iterative CR bound approximations only require precomputation of $\underline{\mu} = \mathbf{A}\underline{\theta}+\underline{w}$ ($2mn\nu$ flops) and storage of the $mn\nu$ non-zero elements of $\mathbf{A}$ and m elements of $\underline{\mu}$. For matrix sparsity factors $\nu < 0.05$ (more than 95% of all elements of $\mathbf{A}$ are zero), commonly encountered in conventional ECT, this corresponds to a substantial savings in storage and precomputation requirements.

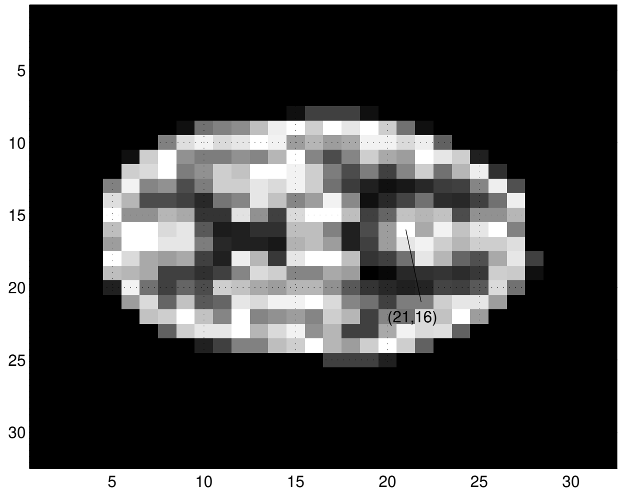

Figure 1: Undersampled Hoffman brain phantom used for numerical comparisons. The intensity ranges from 0 (black) to 2 (white). Region investigated is square 9-pixel neighborhood of pixel (21,16).

For these studies the 32×32 Hoffman brain phantom shown in Fig. 1 was used as the true image intensity $\underline{\theta}$. 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 matrix $\mathbf{A}$ with $m = 3600$ rows, $n = 640$ columns and sparsity factor $\nu = 0.0427$. For this application the memory storage advantages are substantial: direct methods require $n^2 \approx 400$ KBytes to store $\mathbf{F}_Y$ while the iterative methods require only $mn\nu + m \approx 100$ KBytes to store $\mathbf{A}$ and $\underline{\mu}$.

The CR bound approximation algorithms were investigated for unbiased uptake estimation over a specified region. For this case the gradient $\underline{\dot{m}}$ of the mean $m(\underline{\theta}) = t(\underline{\theta})$ is given by the indicator function $\underline{1}_{region}$ of the region, where the i-th element of $\underline{1}_{region}$ is 1 if the intensity θ_i corresponds to a pixel lying within the region while it is 0 otherwise. We present results for the region defined as the square 9-pixel neighborhood of the coordinate $(21, 16)$ (indicated on Fig. 1). We found that only for the preconditioned CG and the GS algorithms was the convergence behavior insensitive to the location of the selected uptake region.

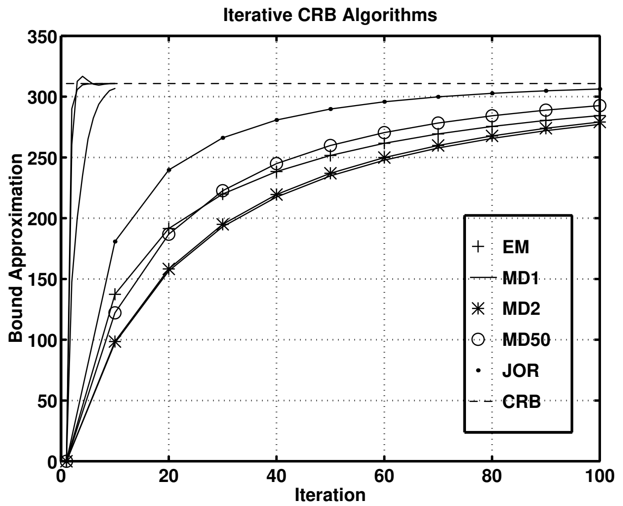

Figure 2: Trajectories of iterative algorithms for approximating the non-singular CR bound for estimates of uptake in the 9-pixel neighborhood of pixel $(21, 16)$. Dotted line labeled CRB denotes the true value of the CR bound. Rapidly convergent non-monotone Gauss-Seidel and preconditioned conjugate gradient algorithms are unlabeled curves at far left of graph (see Figure 3).

Figures 2 and 3 show the convergence trajectories of eight algorithms. Four of these algorithms, labeled EM, MD0, MD1, and MD50 are monotone, while the other four, labeled JOR, GS, CGD, CGDF are non-monotone. MD0, MD1 and MD50 denote the monotone splitting algorithms obtained by using the respective preconditioning matrices $\mathbf{D}_1$, $\mathbf{D}_2$ and $\mathbf{D}_{50}$ defined in Sec. 3.1.2. These algorithms have convergence rates which improve with

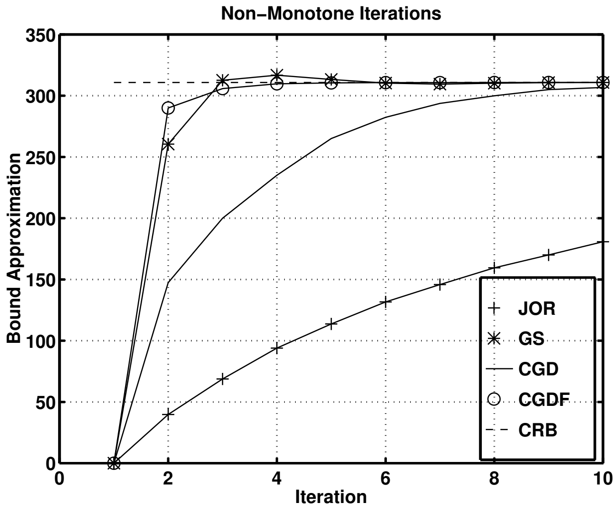

Figure 3: A magnified view of the non-monotone algorithms including Gauss-Seidel (GS) and preconditioned conjugate gradient algorithms shown in Figure 2. CGD is a conjugate gradient algorithm using the standard diagonal Jacobi preconditioner and CGDF uses a preconditioner tailored to the $\mathbf{A}$ matrix considered here.

the number $2p-1$ of non-zero diagonals in $\mathbf{D}_p$. The monotone algorithm labeled EM uses the diagonal preconditioning matrix $\mathbf{F} = \mathrm{diag}_i\left(\mathbf{A}_{*i}^T \underline{1}/\theta_i\right)$ given in [2], where $\mathbf{A}_{*i}$ is the i-th column of $\mathbf{A}$ and $\mathrm{diag}_i(a_i)$ denotes a diagonal matrix with the scalars a_i arranged along the diagonal. It is interesting that while this latter algorithm beats the monotone algorithms MD0-MD50 in the early iterations, it considerably undershoots the CRB in the later iterations and ends up converging to the CRB at a much slower asymptotic rate. The JOR algorithm is a Jacobi iteration implemented with relaxation parameter numerically selected to minimize the root convergence factor: $\psi = 2/[\min(|\lambda_i|) + \max(|\lambda_i|)]$, where $\{\lambda_i\}_i$ are the eigenvalues of $\mathrm{diag}(\mathbf{F}_Y)\mathbf{F}_Y$. The standard unrelaxed Jacobi algorithm diverged for all cases studied and is not shown. The JOR algorithm converges faster than the monotone EM, MD0, MD1, and MD50 algorithms and appears to be monotonic. However, close inspection of the JOR trajectory reveals non-monotone behavior after the first 60 iterations.

In Fig. 3 we zoom into the trajectories of the non-monotone algorithms graphed in Fig. 2. In Fig. 3 the conjugate gradient algorithm labeled CGD uses the standard diagonal Jacobi preconditioner, i.e. $\mathbf{F} = \mathrm{diag}(\mathbf{F}_Y)$ a diagonal matrix formed from the diagonal elements of the FIM (these can be precomputed from $\mathbf{F}_Y = \mathbf{A}\mathbf{D}^{-1}\mathbf{A}$ in $4nm\nu$ flops). The conjugate gradient algorithm labeled CGDF uses a special preconditioner $\mathbf{F}$ consisting of a diagonal matrix, chosen to make $\mathbf{F}_Y$ approximately circulant, followed by a Fourier-type preconditioner. The preconditioner used in CGDF is tailored to the spatially invariant PET application and is described in [12] in the context of fast least squares PET reconstruction algorithms. The GS algorithm shows very rapid convergence which is only slightly outdone by CGDF. On the

Alg.	Asy. Conv. Factor	5%	0.5%	Break Even
EM	$\rho = 0.9999998$	143	521	540
MD0	$\rho = 0.9983984$	160	383	540
MD1	$\rho = 0.9983889$	153	362	540
D50	$\rho = 0.9977878$	109	259	108
JOR	$\rho = 0.9975000$	60	152	540
GS	$\rho = 0.9376000$	3	6	540
CGD	$\rho^{'} = 0.9317000$	8	12	540
CGDF	$\rho^{'} = 0.7940000$	3	4	540

Table 1: Asymptotic and finite convergence properties of the iterative algorithms.

other hand, the GS displays a prominent (2%) overshoot which does not occur in any of the other algorithms.

The convergence properties of these algorithms are quantified in Table 1. The second column of Table 1 contains the asymptotic convergence factors for these algorithms. The asymptotic convergence factors are defined for the splitting algorithms as $\rho = \max\{|\lambda_i^{I-F^{-1}F_Y}|\}_{i=1}^n$, and for the conjugate gradient algorithms as the ratio $\rho^{'} = \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}$, where κ is the spectral condition number $\kappa = \lambda_{max}^{F^{-1}F_Y}/\lambda_{min}^{F^{-1}F_Y}$ of the preconditioned FIM $\mathbf{F}^{-1}\mathbf{F}_Y$. The third and fourth columns show the actual number of iterations to achieve convergence to within a 5% and a 0.5% tolerance of the CRB, respectively. The fifth column shows the number of iterations for which each algorithm would lose its advantage relative to direct computation of the CRB ($2mn\nu$ flops to compute $\mathbf{F}_Y$ plus $n^3/3$ flops to compute $\mathbf{F}_Y^{-1}\underline{\dot{m}}$ via Cholesky methods). For all algorithms except MD50 the number of flops required per iteration is approximately $4mn\nu = 0.4$ Mflops. MD50 requires an additional 1.6 Mflops per iteration to solve the preconditioning equation $\mathbf{D}_{50}\underline{\beta}^{(k)} = \underline{u}$. Comparing the second and fourth columns of the table we see that the asymptotic convergence factor accurately predicts the relative asymptotic rate of convergence of these algorithms: smaller ρ or ρ' implies more rapid attainment of 0.5% tolerance. On the other hand, by comparison to the third column it is seen that the asymptotic convergence factor is not a good predictor of finite convergence rates for less stringent (5%) tolerances.

Next we turn to the case of singular FIM arising in the so-called "missing angle problem" where image parameters must be estimated from 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 were used; corresponding to decimating the rows of $\mathbf{A}$ by a factor of eight. This resulted in a matrix $\mathbf{A}$ of dimension 400×640 with rank 400 and range-space condition number on the order of 1000. We implemented the matrix perturbation algorithm (8) to approximate the CR bound on variance for estimates of uptake in the same 9-pixel region as before.

It was found that over the range $0 \leq \frac{a-b}{a+b} \leq 1000$ the choice of a and b has some effect on the speed of convergence but that the dominating factor is the value of $\epsilon(a+b)/2$. It

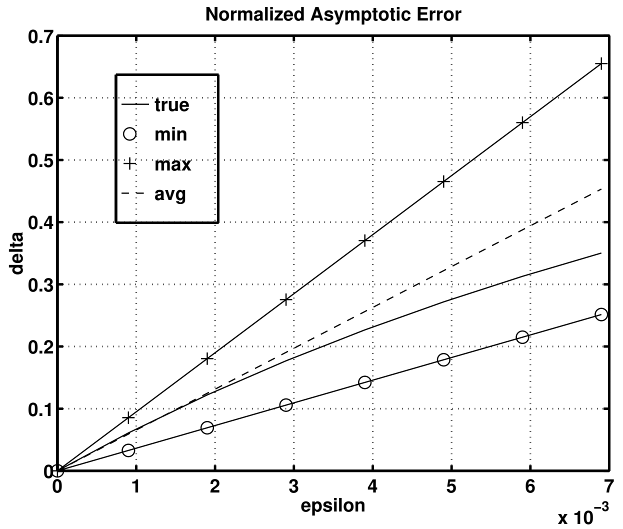

Figure 4: The error bounds on the asymptotic normalized error δ as a function of perturbation parameter ϵ for the case of singular FIM. Also shown is the exact calculated curve $\delta = \delta(\epsilon)$ (labeled "true") and the average of the upper and lower bounds. Note that this average is very close to the exact curve for all values of $\delta < 0.15$, i.e. 15% error or less.

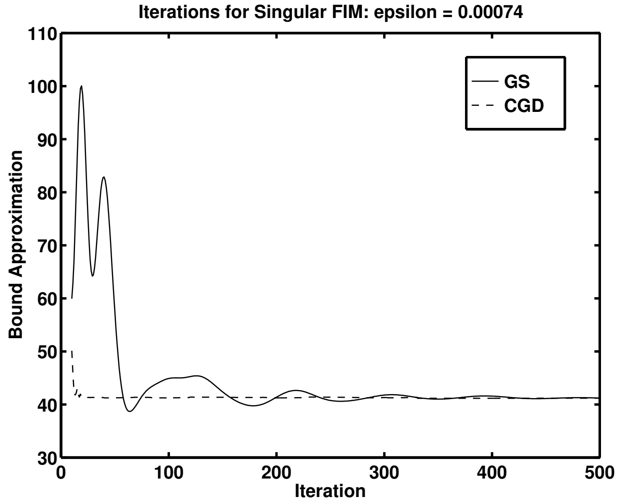

Figure 5: The trajectories of the Gauss-Siedel (GS) and conjugate-gradient with diagonal Jacobi preconditioner (GCD) for the case of singular FIM and uptake estimation. CGD settles to within 5% of the true CR bound after fewer then 15 iterations.

was also noticed that there is considerably more granular noise due to roundoff errors when implementing the differencing method (5) as contrasted with implementing the perturbed product method (6). The differencing method also produced large negatively valued bound approximations in the early iterations. Here we show results for the perturbed product method and $a = b = 1$. We selected a maximum allowable asymptotic normalized error criterion as $\delta = 0.05$, or 5%, and $\epsilon = 0.00074$ was selected according to (9) as the average of the induced lower and upper bounds on ϵ: $\epsilon = [\frac{1}{2}\delta\sigma_{min} + \frac{1}{2}\|\mathbf{F}_Y\underline{\dot{m}}\|^2/(\underline{\dot{m}}^T\mathbf{F}_Y\underline{\dot{m}})]/2$ (see Figure 4). We assumed that the minimum positive singular value σ_{min} of $\mathbf{F}_Y$ was available. In a larger problem this minimum positive eigenvalue would have to be estimated, e.g. using successive power iterations [9].

Figure 5 illustrates the trajectories of the GS and CGD algorithms applied to the iterations (8). The limiting value of both of these algorithms is 41.3 which, as expected, lies below the true CR bound, which we calculated to be 43.5, by approximately 5%. Note that the GS algorithm does not converge to within 5% of the limit before 250 iterations. On the other hand the CGD algorithm settles down to within 5% of the limit after fewer than 15 iterations. Finally it was observed that a significant tradeoff exists between the convergence rate and δ. This is because small δ forces small ϵ which renders perturbed matrices $[\mathbf{F}_Y + a\epsilon\mathbf{I}]$ and $[\mathbf{F}_Y + a\epsilon\mathbf{I}]$ closer to singular. As a result the asymptotic convergence factors of the pair $\underline{\gamma}_1$ and $\underline{\gamma}_2$ in (8) increase towards unity thereby lowering the convergence rate.

7 Conclusion

We have studied several iterative algorithms for approximation of the CR bound including monotone convergent splitting matrix iterations; non-monotonic splitting matrix iterations; and non-monotonic preconditioned conjugate gradient iterations. We have adapted these algorithms to the difficult case of singular Fisher information matrices for which the standard iterative algorithms are inapplicable. The computational cost of using a monotonically convergent algorithm is quite high as compared with a non-monotonic algorithm for the inverse problems considered. Our numerical studies indicate that unless monotonic convergence is an absolute necessity, the preconditioned conjugate gradient algorithm should be seriously considered for singular and non-singular CR bound approximation due to its very rapid convergence.

Acknowledegement

The authors wish to thank the anonymous reviewers for suggestions which significantly improved this paper. We also want to thank W. L. Rogers and N. H. Clinthorne for many helpful discussions of this work.

Appendix A: Bounds on Asymptotic Error

Lemma 2 *Assume that $\underline{\dot{m}}$ does not lie in the nullspace of $\mathbf{F}_Y$ and let δ be the normalized asymptotic error defined as $\delta = \underline{\dot{m}}^T(\mathbf{F}_Y^+ - G(\epsilon))\underline{\dot{m}}/(\underline{\dot{m}}^T\mathbf{F}_Y^+\underline{\dot{m}})$ defined in Sec. 5. Assume that ϵ is dominated by the smallest positive singular value σ_{min} of $\mathbf{F}$ so that to $o(\epsilon)$: $\delta = \epsilon(a+b)\frac{\underline{\dot{m}}^T[\mathbf{F}_Y^+]^2\underline{\dot{m}}}{\underline{\dot{m}}^T\mathbf{F}_Y^+\underline{\dot{m}}}$. Then*

$$\epsilon(a+b)\frac{\underline{\dot{m}}^T\mathbf{F}_Y\underline{\dot{m}}}{\|\mathbf{F}_Y\underline{\dot{m}}\|^2} \leq \delta \leq \epsilon(a+b)\frac{1}{\sigma_{min}}, \tag{11}$$

Proof: By assumption $\underline{\dot{m}}^T\mathbf{F}_Y^+\underline{\dot{m}} > 0$ so the inequalities are well defined. We first show the lower inequality. 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 $\underline{w} = \mathbf{F}_Y^+\underline{\dot{m}}$ and applying the above for $k = 2, 1, 0$ we obtain the sequence of inequalities

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

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 $\underline{\dot{m}}^T\mathbf{F}_Y^2\underline{\dot{m}} = \|\mathbf{F}_Y\underline{\dot{m}}\|^2$ we have established the lower inequality in (11). The upper inequality in (11) follows from the sequence of identities

$$\begin{aligned}
\frac{\underline{\dot{m}}^T[\mathbf{F}_Y^+]^2\underline{\dot{m}}}{\underline{\dot{m}}^T\mathbf{F}_Y^+\underline{\dot{m}}} &\leq \max_{\{\underline{\dot{m}}\,:\,\underline{\dot{m}}^T\mathbf{F}_Y\underline{\dot{m}}>0\}}\left\{\frac{\underline{\dot{m}}^T[\mathbf{F}_Y^+]^2\underline{\dot{m}}}{\underline{\dot{m}}^T\mathbf{F}_Y^+\underline{\dot{m}}}\right\} \\
&= \max_{\underline{\dot{m}}\neq 0}\left\{\frac{\underline{\dot{m}}^T\mathbf{F}_Y^+\underline{\dot{m}}}{\underline{\dot{m}}^T\underline{\dot{m}}}\right\} \\
&= \frac{1}{\sigma_{min}}.
\end{aligned}$$

$\square$

Appendix B: Proof of Lemma 1

The matrix $D_p - A$ has the form:

$$\begin{bmatrix}
\sum\limits_{j\neq 2,\dots,p} |a_{1j}| & 0 & \cdots & 0 & -a_{1,p+1} & \cdots & -a_{1n} \\
0 & \sum\limits_{j\neq 1,3,\dots,p+1} |a_{2j}| & & & \ddots & & \vdots \\
\vdots & \ddots & & & & \ddots & -a_{n-p-1,n} \\
 & & \ddots & & & \ddots & \\
0 & & & \sum\limits_{\substack{j\neq k-p,\dots,k-1 \\ j\neq k+1,\dots,k+p-1}} |a_{kj}| & & & 0 \\
-a_{p+1,1} & \ddots & & & \ddots & & \\
 & & \ddots & & & \ddots & \vdots \\
\vdots & & \ddots & & & \ddots & 0 \\
-a_{n1} & \cdots & -a_{n,n-p-1} & 0 & \cdots & 0 & \sum\limits_{j\neq n-p,\dots,n-1} |a_{nj}|
\end{bmatrix}.$$

Since $D_p - A$ is a special case of $D_1 - A$ for $a_{ij} = 0, |i-j| \geq p$, it suffices to prove that the following matrix is Positive Semi-Definite (PSD):

$$D_1 - A = \begin{bmatrix}
\sum_{j\neq 1} a_1 j & -a_1 2 & \cdots & -a_{1n} \\
-a_{21} & \sum_{j\neq 2} -a_{2j} & \cdots & -a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
-a_{n1} & -a_{n2} & \cdots & \sum_{j\neq n} -a_{nj}
\end{bmatrix} \tag{12}$$

Define:

$$M^{(q)} = \begin{bmatrix} N & \underline{b} \\ \underline{b}^T & a \end{bmatrix},$$

where,

$$\begin{aligned}
\underline{b} &= [-a_{q1}, -a_{q2}, \dots, -a_{q,q-1}]^T, \\
a &= \sum_{i=1}^{q-1} |b_i|, \\
N &= M^{(q-1)} + \mathrm{diag}_i(|b_i|).
\end{aligned}$$

Since $D_1 - A$ is symmetric $M^{(q)}$ is the $q \times q$ upper left hand block of the $n \times n$ matrix $D_1 - A$. We proceed by induction. Assume that $M^{(q-1)}$ is positive semi-definite. Since $a = 0$ implies $\underline{b} = \underline{0}$, when $a = 0$, $M^{(q)}$ is obviously positive semi-definite. We therefore assume $a > 0$, in which case $\underline{b}$ is not identically zero. Consider the following factorization of $M^{(q)}$.

$$M^{(q)} = \begin{bmatrix} I & \frac{1}{a}\underline{b} \\ \underline{0}^T & 1 \end{bmatrix} \begin{bmatrix} N - \frac{1}{a}\underline{b}\underline{b}^T & \underline{0} \\ \underline{0}^T & a \end{bmatrix} \begin{bmatrix} I & \underline{0} \\ \frac{1}{a}\underline{b}^T & 1 \end{bmatrix}.$$

We only need to show that the $(q-1)\times(q-1)$ matrix,

$$N - \frac{1}{a}\underline{b}\underline{b}^T = M^{q-1} + \mathrm{diag}_i(|b_i|) - \frac{1}{a}\underline{b}\underline{b}^T$$

is PSD. Since the sum of PSD matrices is PSD, it is sufficient to show that

$$\mathrm{diag}_i(|b_i|) - \frac{1}{a}\underline{b}\underline{b}^T \geq 0.$$

Consider for any $\underline{x} = [x_1, x_2, ..., x_{q-1}]^T$:

$$\begin{aligned}
\underline{x}^T \left(\mathrm{diag}_i(|b_i|) - \frac{1}{a}\underline{b}\underline{b}^T\right)\underline{x} &= \sum_{i=1}^{q-1} x_i^2 |b_i| - \frac{1}{a}\left(\sum_{i=1}^{q-1} x_i b_i\right)^2 \\
&= \sum_{j=1}^{q-1} |b_j| \left(\sum_{i=1}^{q-1} x_i^2 \frac{|b_i|}{\sum_{j=1}^{q-1}|b_j|} - \left(\sum_{i=1}^{q-1} x_i \frac{b_i}{\sum_{j=1}^{q-1}|b_j|}\right)^2\right) \\
&\geq a\left(\sum_{i=1}^{q-1} x_i^2 \frac{|b_i|}{\sum_{j=1}^{q-1}|b_j|} - \left(\sum_{i=1}^{q-1} |x_i| \frac{|b_i|}{\sum_{j=1}^{q-1}|b_j|}\right)^2\right) \\
&= a\left(\sum_{i=1}^{q-1} p_i |x_i|^2 - \left(\sum_{i=1}^{q-1} p_i |x_i|\right)^2\right) \qquad (13) \\
&= a\sum_{i=1}^{q-1} p_i \left(|x_i| - \sum_{j=1}^{q-1} p_j |x_j|\right)^2 \\
&\geq 0.
\end{aligned}$$

Where in (13) we have defined $p_i = \frac{|b_i|}{\sum_{j=1}^{q-1}|b_j|}$, $p_i \in [0,1]$, $\sum_{i=1}^{q-1} p_i = 1$. Furthermore, for $\underline{x} = \underline{1}$, $(D_p - A)\underline{1} = \underline{0}$, where $\underline{1} = [1, ..., 1]^T$. This implies that $D_p - A$ has at least one zero eigenvalue, therefore $D_p - A$ is rank deficient with rank at most $n-1$. $\square$

References

[1] 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.

[2] 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.

[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] A. O. Hero, J. A. Fessler, and M. Usman, "Exploring bias-variance tradeoffs using the uniform CR bound," *IEEE Trans. on Signal Processing*, vol. in review, , 1994.

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

[7] B. Porat, *Digital processing of random signals*, Prentice-Hall, Englewood-Cliffs N.J., 1994.

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

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

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

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

[12] S. D. Booth and J. A. Fessler, "Combined diagonal/fourier preconditioning methods for image reconstruction in emission tomography," in *IEEE Int. Conf. on Image Processing*, vol. 3, Crystal City, VA, Nov. 1995.