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

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

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 note we compare several iterative methods for approximating the CRB using matrix splitting and preconditioned conjugate gradient algorithms. For a large class of inverse problems we show that non-monotone Gauss-Seidel and preconditioned conjugate gradient algorithms require significantly fewer flops for convergence than monotone "bound preserving" algorithms.

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

I. Introduction

The Cramer-Rao (CR) bound is a widely used lower bound on estimator covariance. When there are n unknown parameters the calculation of the CR bound involves calculation of the inverse or pseudo-inverse of the $n \times n$ Fisher information matrix (FIM). Direct methods of matrix inversion, requiring $O(n^2)$ bytes of memory storage and $O(n^3)$ flops (floating point operations), are intractable if n is large. 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 medical image analysis one may be primarily interested in a small q-pixel region of interest (ROI) corresponding to a tumor or lesion.

In [1] and [2] 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 requires only $O(n^2)$ flops per iteration per parameter so that, if convergence is fast, a computational saving is achieved. The important feature 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 [1] in the setting of a general class of iterative algorithms, known as stationary and non-stationary linear equation solvers [3]. In 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 [1], Jacobi, Gauss-Seidel, and preconditioned conjugate gradient algorithms. The extension of these algorithms to singular FIM is achieved by using matrix perturbation methods.

We illustrate these algorithms for an important class of inverse problems arising in tomographic reconstruction, deconvolution, and image restoration. We perform numerical studies for the special case of uptake estimation in radio-isotope imaging (PET). The uptake is the overall amount of radio-isotope delivered to a region of interest, and the uptake estimates are derived from a set of noisy tomographic projections. After treating uptake estimation with full rank FIM, we conclude the paper by treating the so-called "missing angle problem" where only a small range of projection data is available and the FIM is singular.

II. 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 column gradient operator. 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 bound on the variance of the estimator $\hat{t}(Y)$ is [4], [5]

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

where $\dot{\underline{m}} = \nabla_{\underline{\theta}} m$ 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 [6]. 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 vector $\dot{\underline{m}}$ lies in the range space of $\mathbf{F}_Y$ [4].

Throughout this paper we will be interested in calculating the right hand side of (1). The method easily extends to calculation of the uniform lower bound presented in [4] for biased estimators.

For non-singular FIM, the right hand side of the CR inequality (1) can be computed in $O(n^3)$ flops by solving for $\underline{x}$ in the equation

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

[1] A. Hero, A. Sauve and J. Fessler are with the Dept. of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, MI 48109-2122. M. Usman is with Northwestern University, Evanston, IL. 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

III. Recursive CR Bound Algorithms for Non-Singular FIM

Here we describe the monotonically convergent algorithm of [1] in the context of standard splitting iterations [7, Sec. 10.1], also known as stationary iterations [3], for approximating the solution to the linear equation (2).

A. 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 splitting matrix and is assumed to be non-singular.

A.1 General Splitting Iterations

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

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

For any square matrix $\mathbf{M}$ with eigenvalues $\{\lambda_i^M\}$ the root convergence factor is defined as $\rho(\mathbf{M}) = |\lambda_i^M|_{\max}$ (also known as the spectral radius of $\mathbf{M}$). If $\rho(\mathbf{F}^{-1}\mathbf{N}) < 1$ then $\underline{\beta}^{(k)}$ converges to the vector $\mathbf{F}_Y^{-1}\dot{\underline{m}}$ and the approximating sequence $\eta^{(k)}$ converges to the CR bound $\dot{\underline{m}}^T\mathbf{F}_Y^{-1}\dot{\underline{m}}$ [7].

Many algorithms are splitting iterations, such as the Jacobi (J) and Gauss-Siedel (GS) iterations [7, 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, J iterations may not converge and are generally not monotone. 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. 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. However, 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 iterations the matrix $\mathbf{F}$ is lower triangular and step (ii) of (3) could be accomplished using backsubstitution (n^2 flops). However, GS iterations are 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:

```
for j = 1 to M  (GS Iteration)
    r = (ṁ_j − f_{j*} β^(k)) / f_jj
    β_j^(k) = r
    β_{[j]_n+1}^(k) = 0
end
```

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 $\mathbf{F}_Y$.

A.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 follows from [1, Eq. (9)] that $\eta^{(k+1)} - \eta^{(k)} = \dot{\underline{m}}^T\mathbf{F}^{-\frac{1}{2}}[\mathbf{F}^{-\frac{1}{2}}\mathbf{N}\mathbf{F}^{-\frac{1}{2}}]^k\mathbf{F}^{-\frac{1}{2}}\dot{\underline{m}} \geq 0$. Assume also that the splitting algorithm (3) is initialized with $\underline{\beta}^{(0)} = \underline{0}$. Then 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 matrix $\mathbf{F}$ for which $\mathbf{F} - \mathbf{F}_Y$ is non-negative definite [1]. This was accomplished by selecting a diagonal matrix, denoted $\mathbf{F}_{EM}$, which was the FIM associated with a complete data space.

We next present a general class of FIM-dominating splitting matrices $\mathbf{F}$ which ensure monotone convergence when $\underline{\beta}^{(0)} = \underline{0}$. Define the $(2p-1)$-diagonal banded matrix $\mathbf{D}_p$

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

where $\mathbf{Q} = ((f_{ij}))_{|i-j|\leq p}$ is a $(2p-1)$-diagonal banded matrix, $\mathbf{F} = ((f_{ij}))$, $|\mathbf{A}| = ((|a_{ij}|))$, $\underline{1} = [1, \ldots, 1]^T$, 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 a tridiagonal matrix [8].

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: If $\mathbf{F}_Y$ 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 $n-1$.

One can show that a necessary condition for a $(2p-1)$-diagonal banded matrix $\mathbf{F}$ to minimize the root convergence factor $\rho(\mathbf{I} - \mathbf{F}^{-1}\mathbf{F}_Y)$ subject to $\mathbf{F} - \mathbf{F}_Y \geq 0$, is that $\mathbf{F} - \mathbf{F}_Y$ be rank deficient. Lemma 1 asserts that $\mathbf{F} = \mathbf{D}_p$ satisfies this condition when $\mathbf{F}_Y$ has non-negative entries. Such Fisher matrices $\mathbf{F}_Y$ arise in many applications including the inverse problem considered in Sec. VI.

IV. Preconditioned Conjugate Gradient Algorithm

When the FIM $\mathbf{F}_Y$ is positive definite, the preconditioned conjugate gradient (CG) algorithm can be used to approximate the solution $\underline{x}$ [7, Sec. 10.3] giving an approximation to the CR bound $\dot{\underline{m}}^T\underline{x}$. 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 [7, Algorithm 10.3.1] is quite competitive with direct methods and has significantly faster convergence than MS iterations.

A. Preconditioned CG Recursion for CRB

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

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

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 $\underline{r}^{(k)} = \underline{\dot{m}} - \mathbf{F}_Y \underline{\beta}^{(k)} = \mathbf{F}_Y \Delta\underline{\beta}^{(k)}$ where $\Delta\underline{\beta}^{(k)}$ is the approximation error $\Delta\underline{\beta}^{(k)} = \mathbf{F}_Y^{-1}\underline{\dot{m}} - \underline{\beta}^{(k)}$. The speed of convergence of preconditioned CG generally improves as the eigenvalue spread of $\mathbf{F}^{-1}\mathbf{F}_Y$ decreases. The asymptotic rate of decrease of $\|\underline{r}^{(k)}\|_2 = [\Delta\underline{\beta}^{(k)}]^T \mathbf{F}_Y [\Delta\underline{\beta}^{(k)}]$ is upper bounded by $2\|\Delta\underline{\beta}^{(0)}\|_{F_Y} \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k$, where κ is the spectral condition number of the matrix $\mathbf{F}^{-1}\mathbf{F}_Y$, defined as the ratio of its largest to the smallest magnitude eigenvalues [3, Sec. 2.3.1].

V. CRB Approximation for Singular FIM

The splitting iterations and CG algorithm described in the previous sections are only applicable to non-singular FIM $\mathbf{F}_Y$. 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 $\mathbf{F}_Y \underline{x} = \underline{\dot{m}}$ [7, 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 the case of singular FIM and $\mathbf{F}_Y \underline{\dot{m}} \neq 0$.

Consider the following matrix

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

where $\epsilon > 0$ is a free parameter. $\mathbf{G}(\epsilon)$ is a convergent approximation to the pseudo-inverse of $\mathbf{F}_Y$ in the sense

$$\mathbf{F}_Y^+ = \lim_{\epsilon \to 0+} \mathbf{G}(\epsilon). \tag{5}$$

The representation (5) can be easily established by considering the eigendecomposition of $\mathbf{F}_Y$

$$\mathbf{F}_Y^+ - \mathbf{G}(\epsilon) = \sum_{i=1}^{r} \frac{\epsilon\,(2\sigma_i + \epsilon)}{\sigma_i\,(\sigma_i + \epsilon)^2} \underline{u}_i \underline{u}_i^T, \tag{6}$$

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 $\mathbf{F}_Y^+ - \mathbf{G}(\epsilon) \geq 0$ 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 (5) we have the representation $\underline{\dot{m}}^T \mathbf{G}(\epsilon) \underline{\dot{m}} = \underline{\gamma}^T \mathbf{F}_Y \underline{\gamma}$ where $\underline{\gamma}$ is the solution to the linear equation $[\mathbf{F}_Y + \epsilon \mathbf{I}]\underline{\gamma} = \underline{\dot{m}}$. Since the perturbed matrix $[\mathbf{F}_Y + \epsilon \mathbf{I}]$ is non-singular for $\epsilon > 0$ the CG, GS and other previously discussed algorithms can be applied to approximating $\underline{\gamma}$.

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 via Eq. (6) that for $\epsilon \ll \sigma_{min} = \min_{i=1,...,r}\{\sigma_i\}$: $\underline{\dot{m}}^T (\mathbf{F}_Y^+ - \mathbf{G}(\epsilon)) \underline{\dot{m}} \approx 2\epsilon \|\mathbf{F}_Y^+ \underline{\dot{m}}\|^2$. Hence the normalized asymptotic approximation error is $\delta \approx 2\epsilon \|\mathbf{F}_Y^+ \underline{\dot{m}}\|^2 / \underline{\dot{m}}^T \mathbf{F}_Y^+ \underline{\dot{m}}$. The right hand side of this relation can be bounded (see Appendix) to yield

$$2\epsilon \frac{\underline{\dot{m}}^T \mathbf{F}_Y \underline{\dot{m}}}{\|\mathbf{F}_Y \underline{\dot{m}}\|^2} \leq \delta \leq \frac{2\epsilon}{\sigma_{min}}, \tag{7}$$

which is valid for $\epsilon \ll \sigma_{min}$. In the following Section, relation (7) will be used to select an appropriate value of ϵ to attain a desired magnitude of asymptotic normalized error.

VI. Application to an Inverse Problem

We illustrate the iterative CR bound approximations for the inverse problem consisting of estimating the vector $\underline{\theta}$ from the model

$$\underline{Y} = \mathbf{A}\underline{\theta} + \underline{w},$$

where $\mathbf{A}$ is an $m \times n$ matrix of coefficients, $\underline{w}$ is a vector of independent random noises, and $\underline{\theta}$ is the parameter of interest. This model arises in computed tomography (CT), such as X-ray CT and emission CT, where $\underline{Y}$ is a vector of m projections of a non-negative object attenuation map (X ray CT) or object intensity (emission CT) $\underline{\theta}$, $\mathbf{A}$ is a matrix of transition probabilities, and $\underline{w}$ is Gaussian distributed (X ray CT) or Poisson distributed (emission CT) random vector with diagonal covariance $\mathbf{C}$. For this model the FIM is the non-singular $n \times n$ matrix

$$\mathbf{F}_Y(\underline{\theta}) = \mathbf{A}^T\, \mathbf{C}^{-1}\, \mathbf{A}. \tag{8}$$

In many cases the matrix $\mathbf{A}$ is sparse with $mn\nu$ nonzero entries, $\nu \ll 1$ being the matrix sparsity factor. Due to the simple form (8) of the FIM the matrix-vector product $\mathbf{F}_Y \underline{\beta}^{(k)}$ ($2n^2$ flops) can be performed in two nested vector-matrix multiplications: $\mathbf{F}_Y \underline{\beta}^{(k)} = \mathbf{A}^T \left[\mathbf{C}^{-1} \mathbf{A} \underline{\beta}^{(k)} \right]$ ($4mn\nu$ flops). Furthermore, $\mathbf{F}_Y$ need not be precomputed ($2mn^2\nu$ flops) or stored (n^2 bytes): only storage of the $mn\nu$ non-zero elements of $\mathbf{A}$ and the m non-zero elements of $\mathbf{C}$ is necessary. Therefore, for matrix sparsity factors $\nu < 0.05$ (more than 95% of all elements of $\mathbf{A}$ are zero) commonly encountered in tomography, use of iterative methods for computing the CR bound can give a substantial reduction in storage and computation.

In this note we consider a positron emission tomography PET application similar to that considered in [1]. A down-sampled 32×32 Hoffman brain phantom was used as the true image intensity $\underline{\theta}$. Here the unknown parameter vector $\underline{\theta}$ consisted of a lexicographical ordering of the $n = 640$ pixel intensities within an 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 yields a matrix $\mathbf{A}$ with $m = 3600$ rows, $n = 640$ columns and sparsity factor $\nu = 0.0427$. The goal was to perform unbiased estimation of the integral of $\underline{\theta}$ over a specified region; the so called uptake estimation problem. Specifically, we defined $t(\underline{\theta}) = \underline{1}_{region}^T \underline{\theta}$ where $\underline{1}_{region}$ is a vector indicator function of a square 9-pixel region of interest within the right ventricle of the brain phantom.

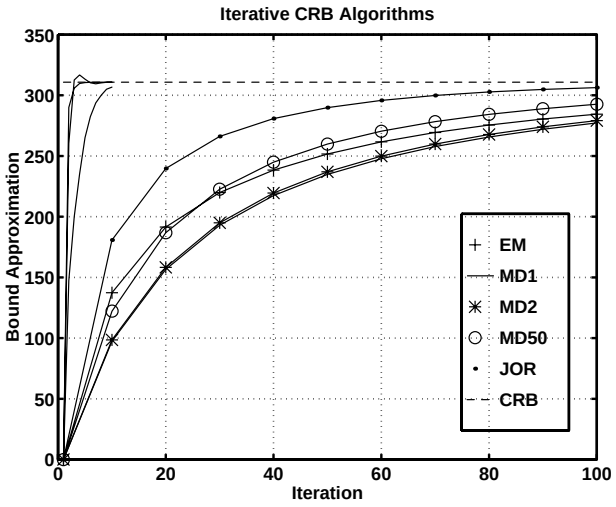

Fig. 1. Trajectories of iterative algorithms for approximating the non-singular CR bound for estimates of uptake in the 9-pixel neighborhood of pixel (21,16). Dotted line labeled CRB denotes the true value of the CR bound. Rapidly convergent non-monotone Gauss-Seidel and preconditioned conjugate gradient algorithms are unlabeled curves at far left of graph (see Figure 2).

Figures 1 and 2 show the convergence trajectories of eight algorithms. MD1, MD2 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. III-A.2. These algorithms have convergence rates which improve with the number $2p-1$ of non-zero off-diagonal bands in $\mathbf{D}_p$. The monotone algorithm labeled EM uses the diagonal splitting matrix $\mathbf{F} = \mathrm{diag}_i \left(\mathbf{A}_{*i}^T \underline{1} / \theta_i \right)$ given in [1], 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 MD1-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 $[\mathrm{diag}(\mathbf{F}_Y)]^{-1} \mathbf{F}_Y$. The standard unrelaxed Jacobi algorithm diverged for all cases studied and is not shown. The JOR algorithm converges faster than the monotone EM, MD1, MD2, and MD50 algorithms and appears to be monotonic. However, quantitative enumeration of the JOR trajectory reveals non-monotone behavior after the first 60 iterations.

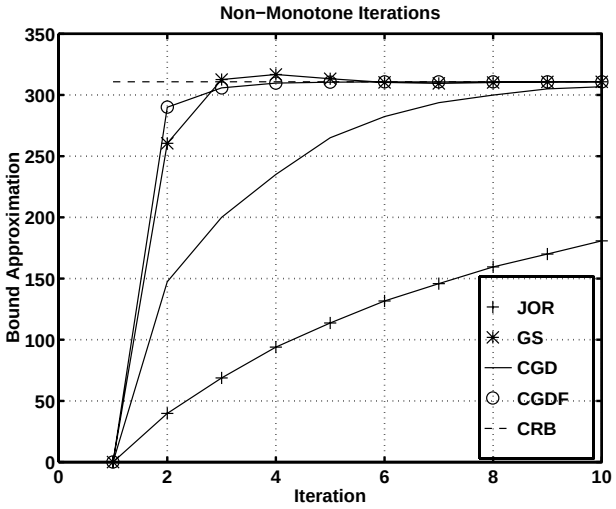

Fig. 2. A magnified view of the non-monotone algorithms including Gauss-Seidel (GS) and preconditioned conjugate gradient algorithms shown in Figure 1. CGD is a conjugate gradient algorithm using the standard diagonal Jacobi preconditioner and CGDF uses a preconditioner tailored to the $\mathbf{A}$ matrix considered here.

In Fig. 2 we zoom into the trajectories of the non-monotone algorithms graphed in Fig. 1. The conjugate gradient algorithm labeled CGD uses the standard diagonal Jacobi preconditioning matrix $\mathbf{F} = \mathrm{diag}(\mathbf{F}_Y)$. The conjugate gradient algorithm labeled CGDF uses a special preconditioning matrix $\mathbf{F}$ 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 [10] in the context of fast least squares PET reconstruction algorithms. The GS algorithm shows very rapid convergence which is only slightly outdone by CGDF. However, the GS displays a prominent (2%) overshoot which does not occur in any of the other

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
MD50	$\rho = 0.9977878$	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 I

ASYMPTOTIC AND FINITE CONVERGENCE PROPERTIES OF THE ITERATIVE ALGORITHMS. THE COLUMNS LABELED 5% AND 0.5% ARE ACTUAL NUMBER OF ITERATIONS REQUIRED FOR CONVERGENCE TO WITHIN A TOLERANCE OF 5% AND 0.5% OF THE CRB. COLUMN LABELED "BREAK EVEN" INDICATES THE NUMBER OF ITERATIONS FOR WHICH THE TOTAL NUMBER OF FLOPS OF EACH ALGORITHM WOULD BE COMPARABLE TO DIRECT COMPUTATION OF THE CRB.

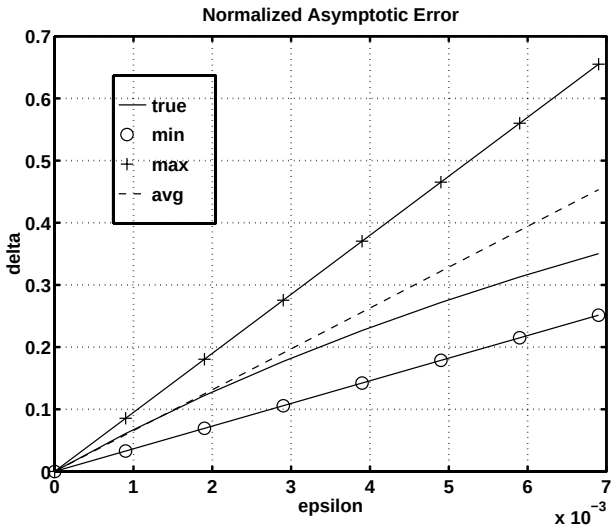

Fig. 3. The error bounds on the asymptotic normalized error δ as a function of perturbation parameter ϵ for the case of singular FIM. Also shown is the exact calculated curve $\delta = \delta(\epsilon)$ (labeled "true") and the average of the upper and lower bounds. Note that this average is close to the exact curve for all values of $\delta < 0.15$, i.e. 15% error or less.

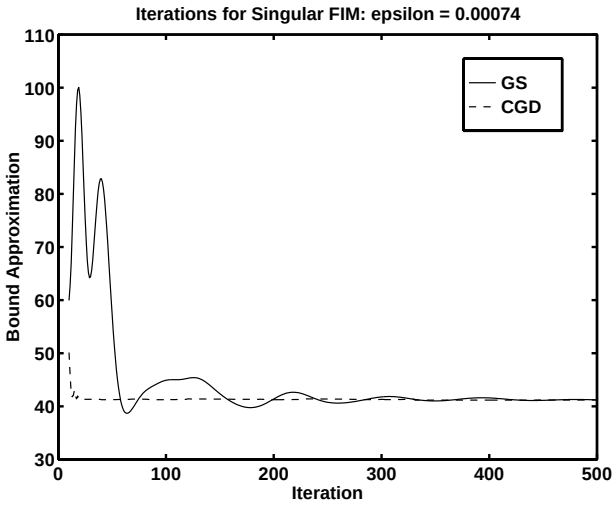

Fig. 4. The trajectories of the Gauss-Siedel (GS) and conjugate-gradient with diagonal Jacobi preconditioner (GCD) for the case of singular FIM and uptake estimation. CGD settles to within 5% of the true CR bound in fewer than 15 iterations.

algorithms.

The convergence properties of these algorithms are quantified in Table 1. The asymptotic convergence factors (second column of Table 1) 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 $\mathbf{F}^{-1}\mathbf{F}_Y$. 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 $\mathbf{F}_Y^{-1}\underline{\dot{\mathrm{m}}}$ via Cholesky decomposition). 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 $\mathbf{D}_{50}\underline{\beta}^{(k)} = \underline{u}$. Note that while the entries in the fourth column monotonically decrease as ρ increases, the third column is not monotone decreasing. This illustrates the fact that the asymptotic convergence factor can be a poor predictor of the non-asymptotic behavior of matrix iterations [11].

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. The matrix perturbation method discussed in Section V was used to approximate the CR bound for uptake estimation.

We selected a maximum allowable asymptotic normalized error criterion as $\delta = 0.05$, or 5%, and $\epsilon = 0.00074$ was selected according to (7) as the average of the induced lower and upper bounds on ϵ: $\epsilon = [\frac{1}{2}\delta\sigma_{min} + \frac{1}{2}\|\mathbf{F}_Y\underline{\dot{\mathrm{m}}}\|^2/(\underline{\dot{\mathrm{m}}}^T\mathbf{F}_Y\underline{\dot{\mathrm{m}}})]/2$ (see Figure 3). To implement this selection scheme the minimum positive singular value σ_{min} of $\mathbf{F}_Y$ must be available. In practice, σ_{min} can be estimated using successive power iterations [7] or using a slightly modified implementation of the preconditioned conjugate gradient algorithm [3].

Figure 4 illustrates the trajectories of the GS and CGD iterations. The limiting value of both of these algorithms is 41.3 which, as expected, lies below the true CR bound numerically calculated to be 43.5, by approximately 5%. Note that the GS algorithm has a highly oscillatory trajec-

tory which does not converge to within 5% of the limit until after 250 iterations. However, the CGD algorithm settles down to within 5% of the limit in fewer than 15 iterations. Finally, while space limitations prevent showing any supporting numerical results, it was observed that a significant tradeoff exists between the convergence rate and δ. This is because decreasing δ forces smaller ϵ and $[\mathbf{F}_Y+\epsilon\mathbf{I}]$ becomes increasingly ill-conditioned.

VII. Conclusion

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.

Appendix: 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 $\delta = \underline{\dot{m}}^T(\mathbf{F}_Y^+ - G(\epsilon))\underline{\dot{m}}/(\underline{\dot{m}}^T\mathbf{F}_Y^+\underline{\dot{m}})$. Assume that ϵ is dominated by the smallest positive singular value σ_{min} of $\mathbf{F}$ so that to $o(\epsilon)$: $\delta = 2\epsilon\frac{\underline{\dot{m}}^T[\mathbf{F}_Y^+]^2\underline{\dot{m}}}{\underline{\dot{m}}^T\mathbf{F}_Y^+\underline{\dot{m}}}$. Then

$$2\epsilon\frac{\underline{\dot{m}}^T\mathbf{F}_Y\underline{\dot{m}}}{\|\mathbf{F}_Y\underline{\dot{m}}\|^2} \leq \delta \leq 2\epsilon\frac{1}{\sigma_{min}}, \tag{9}$$

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

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

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 (9). The upper inequality in (9) follows from the sequence of identities

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

□

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. 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.

[3] R. Barrett, M. Berry, T. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. van der Vorst, *Templates for the solution of linear systems: building blocks for iterative methods*, SIAM Press, Philadelphia, 1994.

[4] A. O. Hero, J. A. Fessler, and M. Usman, "Exploring bias-variance tradeoffs using the uniform CR bound," *IEEE Trans. on Signal Processing*, vol. 44, pp. 2026–2042, Aug. 1996.

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

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

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

[8] A. O. Hero, M. Usman, A. Sauve, and J. A. Fessler, "Recursive algorithms for computing the Cramer-Rao bound," Technical report, Comm. and Sig. Proc. Lab. (CSPL), Dept. EECS, University of Michigan, Ann Arbor, Aug. 1994.

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

[10] 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*, volume 3, pp. 441–444, Crystal City, VA, Nov. 1995.

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