

- [5] S. Van Huffel, H. Chen, C. Decanniere, and P. Van Hecke, "Algorithm for time-domain NMR data fitting based on total least squares," *J. Magn. Resonance A*, vol. 110, pp. 228–237, 1994.
- [6] J. Totz *et al.*, "The use of continuous regularization in the automated analysis of MRS time domain data," submitted to *J. Magn. Resonance A*, 1996.
- [7] H. Chen, S. Van Huffel, C. Decanniere, and P. Van Hecke, "A signal-enhancement algorithm for the quantification of NMR data in the time domain," *J. Magn. Resonance A*, vol. 109, pp. 46–55, 1994.
- [8] S. Y. Kung, K. S. Arun, and D. V. Bhaskar Rao, "State-space and SVD-based approximation methods for the harmonic retrieval problem," *J. Opt. Soc. Amer.*, vol. 73, no. 12, pp. 1799–1811, Dec. 1983.

## Recursive Algorithms for Computing the Cramer-Rao Bound

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 correspondence, 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 nonmonotone Gauss–Seidel and preconditioned conjugate gradient algorithms require significantly fewer flops for convergence than monotone "bound preserving" algorithms.

### 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  $O(n^3)$  floating-point operations (flops), 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 nonsingular 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

Manuscript received August 23, 1994; revised June 27, 1996. 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. The associate editor coordinating the review of this paper and approving it for publication was Prof. Michael P. Clark.

A. Hero, A. Sauve, and J. Fessler are with the Department of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, MI 48109-2122 USA (e-mail: hero@eecs.umich.edu.).

M. Usman is with Northwestern University, Evanston, IL 60201 USA.  
Publisher Item Identifier S 1053-587X(97)01878-3.

iteration. As will be shown in this paper, the price of monotonicity is slow convergence.

In this correspondence, we place the method of [1] in the setting of a general class of iterative algorithms known as stationary and nonstationary linear equation solvers [3]. In this setting, we develop rapidly convergent CR bound approximation methods that 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 nonmonotone 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 correspondence 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  $\underline{\theta} = [\theta_1, \dots, \theta_n]^T$  lying in an open subset  $\Theta$  of  $\mathbb{R}^n$ . Define the  $n \times n$  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]

$$\text{var}_{\underline{\theta}}(\hat{t}) \geq \underline{\mathbf{m}}^T \mathbf{F}_Y^+(\underline{\theta}) \underline{\mathbf{m}} \quad (1)$$

where  $\underline{\mathbf{m}} = \nabla_{\underline{\theta}} m$  is the column gradient vector  $[\frac{\partial m}{\partial \theta_1}, \dots, \frac{\partial m}{\partial \theta_n}]^T$ , and  $\mathbf{F}_Y^+$  denotes the Moore–Penrose pseudo-inverse [6]. When  $\mathbf{F}_Y$  is nonsingular  $\mathbf{F}_Y^+ = \mathbf{F}_Y^{-1}$ , which is the ordinary matrix inverse. Note that the pseudo-inverse form of the CR bound is generally not generally achievable unless the vector  $\underline{\mathbf{m}}$  lies in the range space of  $\mathbf{F}_Y$  [4].

Throughout this correspondence, 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 nonsingular 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

$$\mathbf{F}_Y \underline{x} = \underline{\mathbf{m}}. \quad (2)$$

### III. RECURSIVE CR BOUND ALGORITHMS FOR NONSINGULAR FIM

Here, we describe the monotonically convergent algorithm of [1] in the context of standard splitting iterations [7, Section 10.1], which are 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 that 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 nonsingular.

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

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

For any square matrix  $\mathbf{M}$  with eigenvalues  $\{\lambda_i^M\}$ , the root convergence factor is defined as  $\rho(\mathbf{M}) = |\lambda_i^M|_{\max}$  (which is 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  $\mathbf{F}_Y^{-1}\underline{\dot{m}}$ , and the approximating sequence  $\eta^{(k)}$  converges to the CR bound  $\underline{\dot{m}}^T \mathbf{F}_Y^{-1} \underline{\dot{m}}$  [7].

Many algorithms are splitting iterations, such as the Jacobi (J) and Gauss-Siedel (GS) iterations [7, Section 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 nonnegative 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 nonmonotonic 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

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

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  $\mathbf{F}_Y$ .

2) *Monotone Splitting (MS) Iterations:* Assume that  $\mathbf{F}$  is symmetric positive definite and  $\mathbf{N} = \mathbf{F} - \mathbf{F}_Y$  is symmetric nonnegative definite. Then, it follows from [1, (9)] that  $\eta^{(k+1)} - \eta^{(k)} = \underline{\dot{m}}^T \mathbf{F}^{-\frac{1}{2}} [\mathbf{F}^{-\frac{1}{2}} \mathbf{N} \mathbf{F}^{-\frac{1}{2}}]^k \mathbf{F}^{-\frac{1}{2}} \underline{\dot{m}} \geq 0$ . Assume also that the splitting algorithm (3) is initialized with  $\underline{\beta}^{(0)} = \underline{0}$ . Then, the approximating sequence  $\eta^{(k)}$  is monotone nondecreasing 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  $\text{var}_{\theta}(\hat{t})$ .

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 nonnegative definite [1]. This was accomplished in [1] 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  $\mathbf{D}_p$

$$\mathbf{D}_p = \mathbf{Q} + \text{diag}(|\mathbf{F}_Y - \mathbf{Q}| \underline{1}) \quad (4)$$

where  $\mathbf{Q} = ((f_{ij})_{|i-j| \leq p})$  is a  $(2p - 1)$ -diagonal banded matrix,  $\mathbf{F} = ((f_{ij}))$ ,  $|\mathbf{A}| = ((|a_{ij}|))$ ,  $\underline{1} = [1, \dots, 1]^T$ , and  $\text{diag}(\underline{x})$  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 the  $i$ 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  $\mathbf{D}_p - \mathbf{F}_Y = \text{diag}(|\mathbf{F}_Y - \mathbf{Q}| \underline{1}) - (\mathbf{F}_Y - \mathbf{Q}_p)$  [9, Corollary 7.2.1] and the easily verifiable fact that when  $\mathbf{F}_Y$  has nonnegative entries,  $(\mathbf{D}_p - \mathbf{F}_Y) \underline{1} = \underline{0}$ .

*Lemma 1:* If  $\mathbf{F}_Y$  is an  $n \times n$  symmetric matrix, then  $\mathbf{D}_p - \mathbf{F}_Y$  is nonnegative definite. Furthermore, if  $\mathbf{F}_Y$  has only nonnegative 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  $\mathbf{F} - \mathbf{F}_Y \geq \underline{0}$  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 nonnegative entries. Such Fisher matrices  $\mathbf{F}_Y$  arise in many applications, including the inverse problem considered in Section 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, Section 10.3], giving an approximation to the CR bound  $\underline{\dot{m}}^T \underline{x}$ . 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  $\underline{r}^{(0)} = \underline{\dot{m}} - \mathbf{F}_Y \underline{\beta}^{(0)}$

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

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  $\|\underline{r}^{(k)}\|_2 = \|\Delta \underline{\beta}^{(k)}\|_2^T \mathbf{F}_Y \Delta \underline{\beta}^{(k)}$  is upper bounded by  $2\|\Delta \underline{\beta}^{(0)}\|_{\mathbf{F}_Y} (\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1})^k$ , where  $\kappa$  is the spectral condition number of the matrix  $\mathbf{F}^{-1}\mathbf{F}_Y$ , which is defined as the ratio of its largest to the smallest magnitude eigenvalues [3, Section 2.3.1].

## V. CRB APPROXIMATION FOR SINGULAR FIM

The splitting iterations and CG algorithm described in the previous sections are only applicable to nonsingular FIM  $\mathbf{F}_Y$ . If it is known that  $\underline{\mathbf{m}}$  lies in the range space of  $\mathbf{F}_Y$ , the CR bound  $\underline{\mathbf{m}}^T \mathbf{F}_Y^+ \underline{\mathbf{m}}$  for singular  $\mathbf{F}_Y$  can be found in  $4n^3/3$  flops using the QR factorization to solve for the min-norm solution  $\underline{\mathbf{x}}$  to  $\mathbf{F}_Y \underline{\mathbf{x}} = \underline{\mathbf{m}}$  [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  $\mathbf{F}_Y \underline{\mathbf{m}} \neq 0$ .

Consider the following matrix:

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

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

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

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

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

where  $\sigma_1 \geq \dots \geq \sigma_r > 0$  are the  $r$  nonzero eigenvalues of  $\mathbf{F}_Y$  arranged in decreasing order, and  $\{\underline{\mathbf{u}}_i\}_{i=1}^r$  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  $\underline{\mathbf{m}}^T \mathbf{G}(\epsilon) \underline{\mathbf{m}} \leq \underline{\mathbf{m}}^T \mathbf{F}_Y^+ \underline{\mathbf{m}}$ . Hence,  $\underline{\mathbf{m}}^T \mathbf{G}(\epsilon) \underline{\mathbf{m}}$  is a valid lower bound on  $\text{var}_{\underline{\theta}}(\hat{t})$ , which converges to the CR bound  $\underline{\mathbf{m}}^T \mathbf{F}_Y^+ \underline{\mathbf{m}}$  as  $\epsilon \rightarrow 0$ . In view of (5), we have the representation  $\underline{\mathbf{m}}^T \mathbf{G}(\epsilon) \underline{\mathbf{m}} = \underline{\gamma}^T \mathbf{F}_Y \underline{\gamma}$ , where  $\underline{\gamma}$  is the solution to the linear equation  $[\mathbf{F}_Y + \epsilon \mathbf{I}] \underline{\gamma} = \underline{\mathbf{m}}$ . Since the perturbed matrix  $[\mathbf{F}_Y + \epsilon \mathbf{I}]$  is nonsingular for  $\epsilon > 0$ , the CG, GS, and other previously discussed algorithms can be applied to the approximation of  $\underline{\gamma}$ .

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

$$2\epsilon \frac{\underline{\mathbf{m}}^T \mathbf{F}_Y \underline{\mathbf{m}}}{\|\mathbf{F}_Y \underline{\mathbf{m}}\|^2} \leq \delta \leq \frac{2\epsilon}{\sigma_{\min}} \quad (7)$$

which is valid for  $\epsilon \ll \sigma_{\min}$ . In Section VI, (7) will be used to select an appropriate value of  $\epsilon$  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

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

where

- $\mathbf{A}$   $m \times n$  matrix of coefficients,
- $\underline{\mathbf{w}}$  vector of independent random noises,
- $\underline{\theta}$  parameter of interest.

This model arises in computed tomography (CT), such as X-ray CT and emission CT, where

- $\underline{\mathbf{Y}}$  vector of  $m$  projections of a nonnegative object attenuation map (X ray CT) or object intensity (emission CT)  $\underline{\theta}$ ,
- $\mathbf{A}$  matrix of transition probabilities,

$\underline{\mathbf{w}}$  Gaussian distributed (X ray CT) or Poisson distributed (emission CT) random vector with diagonal covariance  $\mathbf{C}$ . For this model, the FIM is the nonsingular  $n \times n$  matrix

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

In many cases, the matrix  $\mathbf{A}$  is sparse with  $mn\nu$  nonzero entries, where  $\nu \ll 1$  is 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:  $\mathbf{F}_Y \underline{\beta}^{(k)} = \mathbf{A}^T [\mathbf{C}^{-1} \mathbf{A} \underline{\beta}^{(k)}]$  ( $4mn\nu$  flops). Furthermore,  $\mathbf{F}_Y$  need not be precomputed ( $2mn^2\nu$  flops) or stored ( $n^2$  bytes): Only storage of the  $mn\nu$  nonzero elements of  $\mathbf{A}$  and the  $m$  nonzero 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 correspondence, we consider a positron emission tomography PET application similar to that considered in [1]. A downsampled  $32 \times 32$  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; this is the so-called uptake estimation problem. Specifically, we defined  $t(\underline{\theta}) = \underline{\mathbf{1}}_{\text{region}}^T \underline{\theta}$ , where  $\underline{\mathbf{1}}_{\text{region}}$  is a vector indicator function of a square nine-pixel region of interest within the right ventricle of the brain phantom.

Figs. 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  $\mathbf{D}_{50}$ , which are defined in Section III-A2. These algorithms have convergence rates that improve with the number  $2p - 1$  of nonzero off-diagonal bands in  $\mathbf{D}_p$ . The monotone algorithm labeled EM uses the diagonal splitting matrix  $\mathbf{F} = \text{diag}_i(\mathbf{A}_{*i}^T \underline{\mathbf{1}} / \theta_i)$  given in [1], where  $\mathbf{A}_{*i}$  is the  $i$ th column of  $\mathbf{A}$  and  $\text{diag}_i(a_i)$  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  $[\text{diag}(\mathbf{F}_Y)]^{-1} \mathbf{F}_Y$ . 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 nonmonotone behavior after the first 60 iterations.

In Fig. 2, we zoom into the trajectories of the nonmonotone algorithms graphed in Fig. 1. The conjugate gradient algorithm labeled CGD uses the standard diagonal Jacobi preconditioning matrix  $\mathbf{F} = \text{diag}(\mathbf{F}_Y)$ . The conjugate gradient algorithm labeled CGDF uses a special preconditioning matrix  $\mathbf{F}$  consisting of a diagonal matrix, which is 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

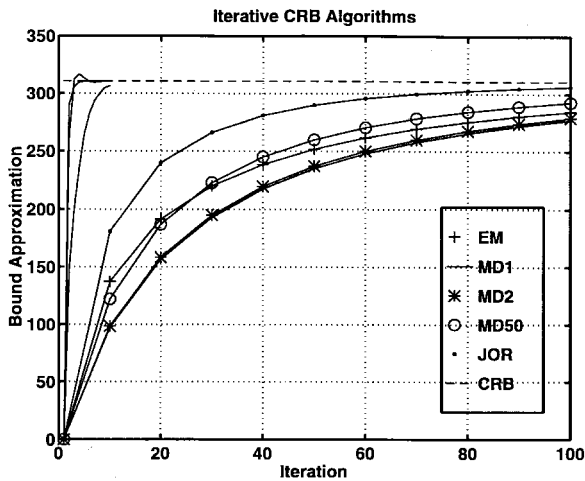


Fig. 1. Trajectories of iterative algorithms for approximating the nonsingular CR bound for estimates of uptake in a nine-pixel neighborhood. Dotted line labeled CRB denotes the true value of the CR bound. Rapidly convergent nonmonotone Gauss-Seidel and preconditioned conjugate gradient algorithms are unlabeled curves at far left of graph (see Fig. 2).

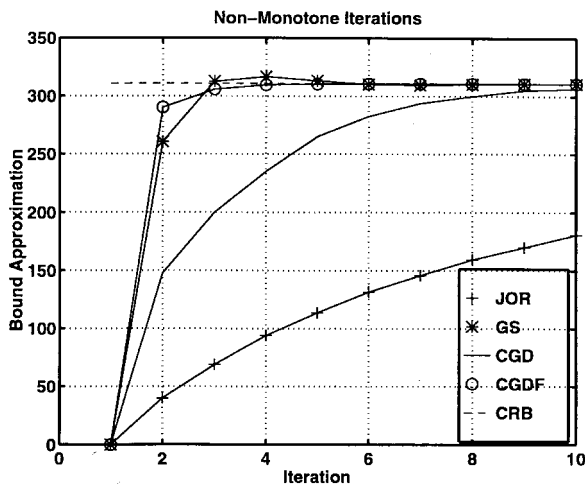


Fig. 2. Magnified view of the nonmonotone algorithms including Gauss-Seidel (GS) and preconditioned conjugate gradient algorithms shown in Fig. 1. CGD is a conjugate gradient algorithm using the standard diagonal Jacobi preconditioner, and CGDF uses a preconditioner tailored to the  $\mathbf{A}$  matrix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

slightly outdone by CGDF. However, the GS displays a prominent (2%) overshoot that does not occur in any of the other algorithms.

The convergence properties of these algorithms are quantified in Table I. The asymptotic convergence factors (second column of Table I) 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  $\kappa$  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 ( $2mnv$  flops to compute  $\mathbf{F}_Y$  plus  $n^3/3$  flops to compute  $\mathbf{F}_Y^{-1}\underline{\hat{m}}$  via Cholesky decomposition). For all algorithms except MD50, the number of flops required per iteration is approximately  $4mnv = 0.4$  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

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 WITH DIRECT COMPUTATION OF THE CRB

Alg.	Asy. Conv. Factor	5%	0.5%	Break Even
EM	$\rho = 0.9999998$	143	521	540
MD1	$\rho = 0.9983984$	160	383	540
MD2	$\rho = 0.9983889$	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

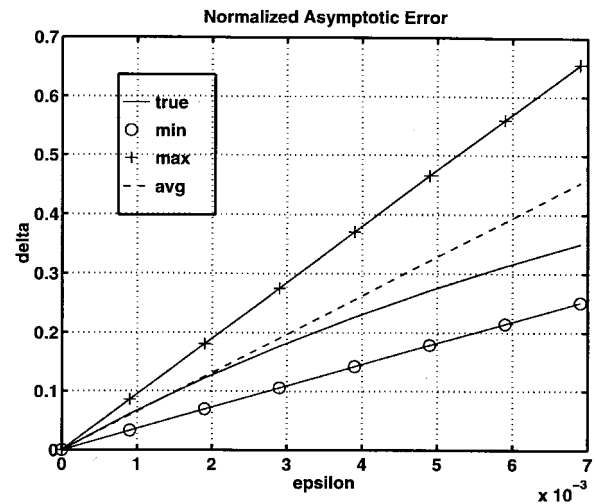


Fig. 3. Error bounds on the asymptotic normalized error  $\delta$  as a function of perturbation parameter  $\epsilon$  for the case of singular FIM. The exact calculated curve  $\delta = \delta(\epsilon)$  (labeled "true") and the average of the upper and lower bounds are also shown. Note that this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

entries in the fourth column monotonically decrease as  $\rho$  increases, the third column is not monotonically decreasing. This illustrates the fact that the asymptotic convergence factor can be a poor predictor of the nonasymptotic 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  $400 \times 640$  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  $\epsilon$ :  $\epsilon = [\frac{1}{2}\delta\sigma_{\min} + \frac{1}{2}\|\mathbf{F}_Y\underline{\hat{m}}\|^2 / (\underline{\hat{m}}^T \mathbf{F}_Y \underline{\hat{m}})] / 2$  (see Fig. 3). To implement this selection scheme, the minimum positive singular value  $\sigma_{\min}$  of  $\mathbf{F}_Y$  must be available. In practice,  $\sigma_{\min}$  can be estimated using successive power iterations [7] or using a slightly modified implementation of the preconditioned conjugate gradient algorithm [3].

Fig. 4 illustrates the trajectories of the GS and CGD iterations. The limiting value of both of these algorithms is 41.3, which, as

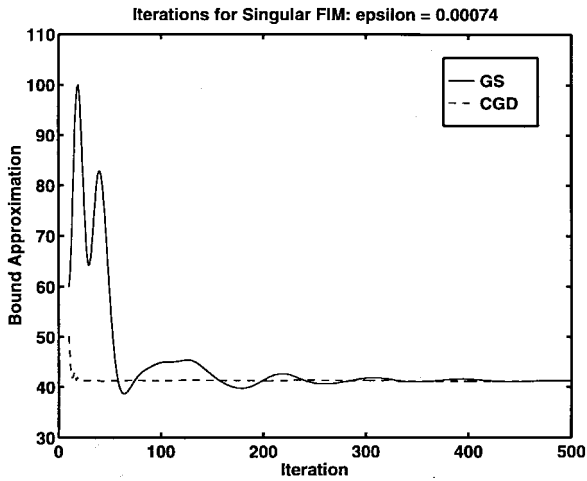


Fig. 4. Trajectories of the Gauss–Seidel (GS) and conjugate-gradient with diagonal Jacobi preconditioner (CGD) for the case of singular FIM and uptake estimation. CGD settles to within 5% of the true CR bound after fewer than 15 iterations.

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 trajectory that 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  $\delta$ . This is because decreasing  $\delta$  forces smaller  $\epsilon$ , and  $[\mathbf{F}_Y + \epsilon\mathbf{I}]$  becomes increasingly ill conditioned.

## VII. CONCLUSION

The main conclusions of this paper are as follows:

- 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 nonmonotone Gauss–Seidel and preconditioned conjugate gradient methods should be implemented due to their advantage of very rapid convergence.

## APPENDIX

### BOUNDS ON ASYMPTOTIC ERROR

*Lemmas 2:* Assume that  $\underline{\mathbf{m}}$  does not lie in the nullspace of  $\mathbf{F}_Y$ , and let  $\delta$  be the normalized asymptotic error  $\delta = \frac{\underline{\mathbf{m}}^T(\mathbf{F}_Y^+ - G(\epsilon))\underline{\mathbf{m}}}{(\underline{\mathbf{m}}^T\mathbf{F}_Y^+\underline{\mathbf{m}})}$ . Assume that  $\epsilon$  is dominated by the smallest positive singular value  $\sigma_{\min}$  of  $\mathbf{F}$  so that to  $o(\epsilon)$ :  $\delta = 2\epsilon \frac{\underline{\mathbf{m}}^T[\mathbf{F}_Y^+]^2\underline{\mathbf{m}}}{\underline{\mathbf{m}}^T\mathbf{F}_Y^+\underline{\mathbf{m}}}$ . Then

$$2\epsilon \frac{\underline{\mathbf{m}}^T\mathbf{F}_Y\underline{\mathbf{m}}}{\|\mathbf{F}_Y\underline{\mathbf{m}}\|^2} \leq \delta \leq 2\epsilon \frac{1}{\sigma_{\min}}. \quad (9)$$

*Proof:* By assumption  $\underline{\mathbf{m}}^T\mathbf{F}_Y^+\underline{\mathbf{m}} > 0$ ; therefore, the inequalities are well defined. We first show the lower inequality. The Cauchy–Schwarz inequality states that  $|\underline{\mathbf{u}}^T\underline{\mathbf{v}}|^2 \leq \underline{\mathbf{u}}^T\underline{\mathbf{u}} \cdot \underline{\mathbf{v}}^T\underline{\mathbf{v}}$  for any two vectors  $\underline{\mathbf{u}}$  and  $\underline{\mathbf{v}}$ . Letting  $\underline{\mathbf{u}} = \mathbf{F}_Y^{k/2}\underline{\mathbf{w}}$  and  $\underline{\mathbf{v}} = \mathbf{F}_Y^{k/2+1}\underline{\mathbf{w}}$ , we obtain

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

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

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

where  $\mathcal{P}_{F_Y} = \mathbf{F}_Y^+\mathbf{F}_Y = \mathbf{F}_Y\mathbf{F}_Y^+$  is a symmetric idempotent matrix that projects vectors onto the column space of  $\mathbf{F}_Y$ . Since  $\underline{\mathbf{m}}^T\mathbf{F}_Y^2\underline{\mathbf{m}} = \|\mathbf{F}_Y\underline{\mathbf{m}}\|^2$ , we have established the lower inequality in (9). The upper inequality in (9) follows from the sequence of identities

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

## REFERENCES

- [1] A. O. Hero and J. A. Fessler, "A recursive algorithm for computing CR-type bounds on estimator covariance," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1205–1210, July 1994.
- [2] —, "A fast recursive algorithm for computing CR-type bounds for image reconstruction problems," in *Proc. IEEE Nuclear Sci. Symp.*, Orlando, FL, Oct. 1992, pp. 1188–1190.
- [3] R. Barrett *et al.*, *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*. Philadelphia: SIAM, 1994.
- [4] A. O. Hero, J. A. Fessler, and M. Usman, "Exploring bias-variance tradeoffs using the uniform CR bound," *IEEE Trans. Signal Processing*, vol. 44, pp. 2026–2042, Aug. 1996.
- [5] C. R. Rao, *Linear Statistical Inference and Its Applications*. New York: Wiley, 1973.
- [6] F. A. Graybill, *Matrices with Applications in Statistics*. Belmont, CA: Wadsworth, 1983.
- [7] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1989.
- [8] A. O. Hero, M. Usman, A. Sauve, and J. A. Fessler, "Recursive algorithms for computing the Cramer-Rao bound," Tech. Rep., Commun. Signal Processing Lab. (CSPL), Dept. Elect. Eng. Comput. Sci., Univ. Michigan, Ann Arbor, Aug. 1994.
- [9] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, MA: MIT Press, 1985.
- [10] S. D. Booth and J. A. Fessler, "Combined diagonal/fourier preconditioning methods for image reconstruction in emission tomography," in *IEEE Int. Conf. Image Processing*, Crystal City, VA, Nov. 1995, vol. 3, pp. 441–444.
- [11] D. M. Young, *Iterative Solution of Large Linear Systems*. New York: Academic, 1971.