Correspondence

A Recursive Algorithm for Computing Cramer–Rao-Type Bounds on Estimator Covariance

Alfred Hero, *Member, IEEE,*
and Jeffrey A. Fessler, *Member, IEEE*

Abstract—**We give a recursive algorithm to calculate submatrices of the Cramer–Rao (CR) matrix bound on the covariance of any unbiased estimator of a vector parameter $\underline{\theta}$. Our algorithm computes a sequence of lower bounds that converges monotonically to the CR bound with exponential speed of convergence. The recursive algorithm uses an invertible "splitting matrix" to successively approximate the inverse Fisher information matrix. We present a statistical approach to selecting the splitting matrix based on a "complete-data–incomplete-data" formulation similar to that of the well-known EM parameter estimation algorithm. As a concrete illustration we consider image reconstruction from projections for emission computed tomography.**

Index Terms—**Multidimensional parameter estimation, estimator covariance bounds, complete–incomplete-data problem, image reconstruction.**

I. Introduction

The Cramer–Rao (CR) bound on estimator covariance is an important tool for predicting fundamental limits on best achievable parameter estimation performance. For a vector parameter $\underline{\theta} \in \Theta \subset \mathbb{R}^n$, an observation Y, and probability density function (pdf) $f_Y(y; \underline{\theta})$, one seeks a lower bound on the minimum achievable variance of an unbiased estimator $\hat{\Theta}_1 = \hat{\theta}_1(Y)$ of a scalar parameter θ_1 of interest. More generally, if, without loss of generality, the p parameters $\theta_1, \cdots, \theta_p$ are of interest, $p \leq n$, one may want to specify a $p \times p$ matrix which lower bounds the error covariance matrix for unbiased estimators $\hat{\theta}_1, \cdots, \hat{\theta}_p$. The upper left $p \times p$ submatrix of the $n \times n$ inverse Fisher information matrix F_Y^{-1} provides the CR lower bound for these parameter estimates. Equivalently, the first p columns of F_Y^{-1} provide this CR bound. The method of sequential partitioning [1] for computing the upper left $p \times p$ submatrix of F_Y^{-1} and Cholesky-based Gaussian elimination techniques [2] for computing the p first columns of F_Y^{-1} are efficient direct methods for obtaining the CR bound, but require $O(n^3)$ floating point operations. Unfornately, in many practical cases of interest, e.g., when there are a large number of nuisance parameters, high computation and memory requirements make direct implementation of the CR bound impractical.

Manuscript received April 18, 1992; revised April 9, 1993. This research was supported in part by the National Science Foundation under Grant BCS-9024370, a DOE Alexander Hollaender Postoctoral Fellowship, and DOE Grant DE-FG02-87ER60561. This paper was presented in part at the IEEE International Symposium Information Theory, San Antonio, TX, January 17–21, 1992.

A. O. Hero is with the Department of Electrical Engineering and Computer Science, the University of Michigan, Ann Arbor, MI 48109-2122.

J. A. Fessler, is with the Division of Nuclear Medicine, the University of Michigan, Ann Arbor, MI 48109-0552.

IEEE Log Number 9402455.

In this correspondence we give an iterative algorithm for computing columns of the CR bound which requires only $O(pn^2)$ floating point operations per iteration. This algorithm falls into the class of "splitting matrix iterations" [2] with the imposition of an additional requirement: the splitting matrix must be chosen to ensure that a valid lower bound results at each iteration of the algorithm. While a purely algebraic approach to specifying a suitable splitting matrix can also be adopted, here we exploit specific properties of Fisher information matrices arising from the statistical model. Specifically, we formulate the parameter estimation problem in a complete-data–incomplete-data setting and apply a version of the "data processing theorem" [3] for Fisher information matrices. This setting is similar to that which underlies the classical formulation of the maximum likelihood expectation maximization (ML-EM) parameter estimation algorithm. The ML-EM algorithm generates a sequence of estimates $\{\hat{\underline{\Theta}}^{(k)}\}_{k \geq 0}$ for $\underline{\theta}$ which successively increases the likelihood function and converges to the maximum likelihood estimator. In a similar manner, our algorithm generates a sequence of tighter and tighter lower bounds on estimator covariance which converges to the actual CR matrix bound.

The algorithms given here converge monotonically with exponential rate, where the asymptotic speed of convergence increases as the spectral radius $\rho(I - F_X^{-1}F_Y)$ decreases. Here I is the $n \times n$ identity matrix and F_X and F_Y are the complete- and incomplete-data Fisher information matrices, respectively. Thus when the complete data is only moderately more informative than the incomplete data, F_Y is close to F_X so that $\rho(I - F_X^{-1}F_Y)$ is close to 0 and the algorithm converges very quickly. To implement the algorithm, one must 1) precompute the first p columns of F_X^{-1}, and 2) provide a subroutine that can multiply $F_X^{-1}F_Y$ or $F_X^{-1}E_{\underline{\theta}}[\nabla^{11}Q(\underline{\theta}; \underline{\theta})]$ by a column vector (see (18)). By appropriately choosing the complete-data space, this precomputation can be quite simple, e.g., X can frequently be chosen to make F_X sparse or even diagonal. If the complete-data space is chosen intelligently, only a few iterations may be required to produce a bound which closely approximates the CR bound. In this case the proposed algorithm gives an order of magnitude computational savings as compared to conventional exact methods of computing the CR bound. This allows one to examine small submatrices of the CR bound for estimation problems that would have been intractable by exact methods due to the large dimension of F_Y.

The paper concludes with an implementation of the recursive algorithm for bounding the minimum achievable error of reconstruction for a small region of interest (ROI) in an image reconstruction problem arising in emission computed tomography. By using the complete data specified for the standard EM algorithm for PET reconstruction [4], [5], F_X is shown to be diagonal and the implementation of the recursive CR bound algorithm is very simple. As in the ML-EM PET reconstruction algorithm, the rate of convergence of the iterative CR bound algorithm depends on the image intensity and the tomographic system response matrix. Furthermore, due to the sparseness of the tomographic system response matrix, the computation of each column of the CR bound matrix recursion requires only

0018-9448/94$04.00

$O(n)$ memory storage as compared to $O(n^2)$ for the general algorithm.

II. CR Bound and Iterative Algorithm

A. Background and General Assumptions

Let Θ_i be an open subset of the real line $\mathbb{R}$. Define $\underline{\theta} = [\theta_1, \cdots, \theta_n]^T$ a real, nonrandom parameter vector residing in $\Theta = \Theta_1 \times \cdots \times \Theta_n$. Let $\{P_{\underline{\theta}}\}_{\underline{\theta} \in \Theta}$ be a family of probability measures for a certain random variable Y taking values in a set $\mathcal{Y}$. Assume that for each $\underline{\theta} \in \Theta$, $P_{\underline{\theta}}$ is absolutely continuous with respect to a dominating measure μ, so that for each $\underline{\theta}$ there exists a density function $f(y; \underline{\theta}) = dP_{\underline{\theta}}(y)/d\mu$ for Y When $\int_{\mathcal{Y}} |y| dP_{\underline{\theta}}$ is finite, we define the expectation $E_{\underline{\theta}}[Y] = \int_{\mathcal{Y}} y\, dP_{\underline{\theta}}$.

The family of densities $\{f_Y(y; \underline{\theta})_{\underline{\theta} \in \Theta}$ is said to be a *regular family* [6] if Θ is an open subset of $\mathbb{R}^n$ and 1) $f_Y(y; \underline{\theta})$ is a continuous function on Θ for μ-almost all y; 2) $\ln f(Y; \underline{\theta})$ is mean-square differentiable in $\underline{\theta}$; and 3) $\nabla_{\underline{\theta}} \ln f(Y; \underline{\theta})$ is mean-square continuous in $\underline{\theta}$. These three conditions guarantee that the nonnegative definite $n \times n$ Fisher information matrix $F_Y(\underline{\theta})$ exists and is finite:

$$F_Y(\underline{\theta}) \stackrel{\text{def}}{=} E_{\underline{\theta}}[\nabla_{\underline{u}} \ln f_Y(Y; \underline{u})|_{\underline{u}=\underline{\theta}}]^T [\nabla_{\underline{u}} \ln f_Y(Y; \underline{u})|_{\underline{u}=\underline{\theta}}], \quad (1)$$

where $\nabla_{\underline{\theta}} = [\partial/\partial\theta_1, \cdots, \partial/\partial\theta_n]$ is the (row) gradient operator.

Under the additional assumption that the mixed partials $(\partial^2/\theta_i\theta_j) f_Y(Y; \underline{\theta})$, $i, j = 1, \cdots, n$, exist, are continuous in $\underline{\theta}$, and are absolutely integrable in Y, the Fisher information matrix is equivalent to the Hessian, or "curvature matrix," of the mean of $\ln f_Y(Y; \underline{\theta})$:

$$\begin{aligned} F_Y(\underline{\theta}) &= -E_{\underline{\theta}}[\nabla_{\underline{u}}^T \partial_{\underline{u}} \ln f_Y(Y; \underline{u})|_{\underline{u}=\underline{\theta}}] \\ &= -\nabla_{\underline{u}}^T \nabla_{\underline{u}} E_{\underline{\theta}}[\ln f_Y(Y; \underline{u})]|_{\underline{u}=\underline{\theta}}. \end{aligned} \quad (2)$$

Finally we recall convergence results for linear recursions of the form

$$\underline{v}^{i+1} = A\underline{v}^i, \qquad i = 1, 2, \cdots,$$

where $\underline{v}^i$ is a vector and A is a matrix. Let $\rho(A)$ denote the spectral radius, i.e., the maximum magnitude eigenvalue, of A. If $\rho(A) < 1$, then $\underline{v}^i$ converges exponentially to zero and the asymptotic rate of convergence increases as the *root convergence factor* $\rho(A)$ decreases [7].

B. The CR Lower Bound

Let $\hat{\underline{\theta}} = \hat{\underline{\theta}}(Y)$ be an unbiased estimator of $\underline{\theta} \in \Theta$, and assume that the densities $\{f_Y(y; \underline{\theta})\}_{\underline{\theta} \in \Theta}$ are a regular family. Additionally assume that the Fisher information F_Y is positive definite. Then the covariance matrix of $\hat{\underline{\theta}}$ satisfies the matrix CR lower bound [6]:

$$\text{cov}_{\underline{\theta}}(\hat{\underline{\theta}}) \geq B(\underline{\theta}) = F_Y^{-1}(\underline{\theta}). \quad (3)$$

We refer to the above as the unbiased CR bound.

Assume that among the n unknown quantities $\underline{\theta} = [\theta_1, \cdots, \theta_n]^T$, only a small number $p \ll n$ of parameters $\underline{\theta}^I = [\theta_1, \cdots, \theta_p]^T$ are directly of interest, the remaining $n - p$ parameters being considered "nuisance parameters." Partition the Fisher information matrix F_Y as

$$F_Y = \begin{bmatrix} F_{11} & F_{12}^T \\ F_{12} & F_{22} \end{bmatrix}, \quad (4)$$

where F_{11} is the $p \times p$ Fisher information matrix for the parameters $\underline{\theta}^I$ of interest, F_{22} is the $(n-p) \times (n-p)$ Fisher information matrix for the nuisance parameters, and F_{12} is the $(n-p) \times p$ information coupling matrix. The CR bound on the covariance of any unbiased estimator $\hat{\underline{\theta}}^I = [\hat{\theta}_1, \cdots, \hat{\theta}_p]^T$ of the parameters of interest is simply the $p \times p$ submatrix in the upper left-hand corner of F_Y^{-1}:

$$\text{cov}_{\underline{\theta}}(\hat{\underline{\theta}}^I) \geq \mathcal{E}^T F_Y^{-1} \mathcal{E}, \quad (5)$$

where $\mathcal{E}$ is the $n \times p$ elementary matrix consisting of the first p columns of the $n \times n$ identity matrix, i.e., $\mathcal{E} = [\underline{e}_1, \cdots, \underline{e}_p]$, and $\underline{e}_j$ is the jth unit column vector in $\mathbb{R}^n$. Using a standard identity for the partitioned matrix inverse [2], the submatrix (5) can be expressed in terms of the partition elements (4) of F_Y, yielding the following equivalent form for the unbiased CR bound:

$$\text{cov}_{\underline{\theta}}(\hat{\underline{\theta}}^I) \geq [F_{11} - F_{12}^T F_{22}^{-1} F_{12}]^{-1}. \quad (6)$$

By using the method of sequential partitioning [1], the right-hand side of (6) could be computed with $O(n^3)$ floating point operations. Alternatively, the CR bound (5) is specified by the first p columns $F_Y^{-1}\mathcal{E}$ of F_Y^{-1}. These p columns are given by the columns of the $n \times p$ matrix solution U to $F_Y U = \mathcal{E}$. The topmost $p \times p$ block $\mathcal{E}^T U$ of U is equal to the right-hand side of the CR bound inequality (5). By using the Cholesky decomposition of F_Y and Gaussian elimination [2], the solution U to $F_Y U = \mathcal{E}$ could be computed with (n^3) floating point operations.

Even if the number p of parameters of interest is small, for large n the feasibility of directly computing the CR bound (5) is limited by the high number $O(n^3)$ of floating point operations. For example, in the case of image reconstruction for a moderate-sized 256×256 pixelated image, F_Y is $256^2 \times 256^2$ so that direct computation of the CR bound on estimation errors in a small region of the image requires on the order of 256^6, or approximately 10^{19}, floating point operations!

C. A Recursive CR Bound Algorithm

The basic idea of the algorithm is to replace the difficult inversion of F_Y with an easily inverted matrix F. To simplify notation, we drop the dependence on $\underline{\theta}$. Let F be an $n \times n$ matrix. Assume that F_Y is positive definite and that $F \geq F_Y$, i.e., $F - F_Y$ is nonnegative definite. It follows that F is positive definite, so let $F^{1/2}$ be the positive definite matrix-square-root-factor of F. Then,

$$1 - F^{-1/2}(F - F_Y)F^{-1/2} = F^{-1/2} F_Y F^{-1/2} > 0,$$

and

$$F^{-1/2}(F - F_Y)F^{-1/2} \geq 0.$$

Hence $0 \leq I - F^{-1/2} F_Y F^{-1/2} < I$, so that all of the eigenvalues of $I - F^{-1/2} F_Y F^{-1/2}$ are nonnegative and strictly less than 1. Since $I - F^{-1/2} F_Y F^{-1/2}$ is similar to $I - F^{-1} F_Y$, it follows that the eigenvalues of $I - F^{-1} F_Y$ lie in $[0, 1)$ [8, Corollary 1.3.4]. Thus, applying the matrix form of the geometric series [8, Corollary 5.6.16]:

$$\begin{aligned} B = [F_Y]^{-1} &= [F - (F - F_Y)]^{-1} \\ &= [I - F^{-1}(F - F_Y)]^{-1} F^{-1} \\ &= \left(\sum_{k=0}^{\infty} [I - F^{-1} F_Y]^{(k)} \right) F^{-1}. \end{aligned} \quad (7)$$

This infinite series expression for the unbiased $n \times n$ CR bound B is the basis for the matrix recursion given in the following theorem.

Theorem 1: Assume that F_Y is positive definite and $F \geq F_Y$. When initialized with the $n \times n$ matrix of zeros $B^{(0)} = 0$, the

following recursion yields a sequence of matrix lower bounds $B^{(k)} = B^{(k)}(\underline{\theta})$ on the $n \times n$ covariance of unbiased estimators $\hat{\underline{\theta}}$ of $\underline{\theta}$. This sequence asymptotically converges to the $n \times n$ unbiased CR bound F_Y^{-1} with root convergence factor $\rho(A)$.

Recursive Algorithm: For $k = 0, 1, 2, \cdots$,

$$B^{(k+1)} = A \cdot B^{(k)} + F^{-1}, \tag{8}$$

where $A = I - F^{-1}F_Y$ has eigenvalues in $[0, 1)$. Furthermore, the convergence is monotone in the sense that $B^{(k)} \leq B^{(k+1)} \leq B = F_Y^{-1}$, for $k = 0, 1, 2, \cdots$.

Proof: Since all eigenvalues of $I - F^{-1}F_Y$ are in the range $[0, 1)$, we obviously have $\rho(I - F^{-1}F_Y) < 1$. Now consider

$$\begin{aligned} B^{(k+1)} - F_Y^{-1} &= (I - F^{-1}F_Y)B^{(k)} + F^{-1} - F_Y^{-1} \\ &= (I - F^{-1}F_Y)(B^{(k)} - F_Y^{-1}). \end{aligned} \tag{9}$$

Since the eigenvalues of $I - F^{-1}F_Y$ are in $[0, 1)$, this establishes that $B^{(k+1)} \to F_Y^{-1}$ as $k \to \infty$ with root convergence factor $\rho(I - F^{-1}F_Y)$. Similarly,

$$B^{(k+1)} - B^{(k)} = (I - F^{-1}F_Y)[B^{(k)} - B^{(k-1)}], \qquad k = 1, 2, \cdots,$$

with initial condition $B^{(1)} - B^{(0)} = F^{-1}$. By induction we have

$$B^{(k+1)} - B^{(k)} = F^{-1/2}[I - F^{-1/2}F_Y F^{-1/2}]^k F^{-1/2},$$

which is nonnegative definite for all $k \geq 0$. Hence the convergence is monotone. □

By right multiplying each side of the equality (8) by the matrix $\mathcal{E} = [\underline{e}, \cdots, \underline{e}_p]$, where $\underline{e}_j$ is the jth unit vector in $\mathbb{R}^n$, we obtain a recursion for the first p columns $B^{(k)}\mathcal{E} = [\underline{b}_1^{(k)}, \cdots, \underline{b}_p^{(k)}]$. Furthermore, the first p rows $\mathcal{E}^T B^{(k)}\mathcal{E}$ of $B^{(k)}\mathcal{E}$ correspond to the upper left-hand corner $p \times p$ submatrix of $B^{(k)}$ and, since $\mathcal{E}[B^{(k+1)} - B^{(k)}]\mathcal{E}$ is nonnegative definite, by Theorem 1, $\mathcal{E}^T B^{(k)}\mathcal{E}$ converges monotonically to $\mathcal{E}^T F_Y^{-1}\mathcal{E}$. Thus we have the following corollary to Theorem 1.

Corollary 1: Assume that F_Y is positive definite and $F \geq F_Y$, and let $\mathcal{E} = [\underline{e}_1, \cdots, \underline{e}_p]$ be the $n \times p$ elementary matrix whose columns are the first p unit vectors in $\mathbb{R}^n$. When initialized with the $n \times p$ matrix of zeros $\beta^{(0)} = 0$, the top $p \times p$ block $\mathcal{E}^T\beta^{(k)}$ of $\beta^{(k)}$ in the following recursive algorithm yields a sequence of lower bounds on the covariance of any unbiased estimator of $\underline{\theta}^I = [\theta_1, \cdots, \theta_p]^T$ which asymptotically converges to the $p \times p$ CR bound $\mathcal{E}^T F_Y^{-1}\mathcal{E}$ with root convergence factor $\rho(A)$:

Recursive Algorithm: For $k = 0, 1, 2, \cdots,$,

$$\beta^{(k+1)} = A \cdot \beta^{(k)} + \mathcal{F}^{-1}, \tag{10}$$

where $A = I - F^{-1}F_Y$ has eigenvalues in $[0, 1)$ and $\mathcal{F}^{-1} = F^{-1}\mathcal{E}$ is the $n \times p$ matrix consisting of the first p columns of F^{-1}. Furthermore, the convergence is monotone in the sense that $\mathcal{E}^T\beta^{(k)} \leq \mathcal{E}^T\beta^{(k+1)} \leq \mathcal{E}^T F_Y^{-1}\mathcal{E}$, for $k = 0, 1, 2, \cdots$.

Given F^{-1} and A the $n \times n$ times $n \times p$ matrix multiplication $A \cdot \beta^{(k)}$ requires only $O(pn^2)$ floating point operations.

D. Discussion

We make the following comments on the recursive algorithms of Theorem 1 and Corollary 1.

1) In order that the algorithm (10) for computing columns of F_Y^{-1} have significant computational advantages relative to the direct sequential partitioning and Cholesky-based methods discussed in Section II.B, the precomputation of the matrix inverse F^{-1} must be simple, and the iterations must converge reasonably quickly. By choosing an F that is sparse or diagonal, the computation of F^{-1} requires only $O(n^2)$ floating point operations. If in addition F can be chosen such that $\rho(I - F^{-1}F_Y)$ is small, then the algorithm (10) will converge to within a small fraction of the corresponding column of F_Y^{-1} with only a few iterations and thus will be an order of magnitude less costly than direct methods requiring $O(n^3)$ operations.
2) From the relation $I - F_Y B^{(k+1)} = A^T \cdot [I - F_Y B^{(k)}]$, obtained in a similar manner as (9) of the proof of Theorem 1, we obtain the following recursion for the normalized difference $\Delta\beta^{(k)} = F_Y[F_Y^{-1} - B^{(k)}]\mathcal{E}$ between $\beta^{(k)} = B^{(k)}\mathcal{E}$ and its asymptotic limit $F_Y^{-1}\mathcal{E}$:
$$\Delta\beta^{(k+1)} = A^T\Delta\beta^{(k)}, \qquad k = 1, 2, \cdots,$$
with $\Delta\beta^{(0)} = \mathcal{E}$. This recursion can be implemented in parallel with (10) to monitor the progress of the iterative CR bound algorithm towards its limit.
3) For $p = 1$, the iteration of Corollary 1 is related to the "matrix splitting" method [2] for iteratively approximating the solution $\underline{u}$ to a linear equation $C\underline{u} = \underline{c}$. In this method, a decomposition $C = F - N$ is found for the nonsingular matrix C such that F is nonsingular and $\rho(F^{-1}N) < 1$. Once this decomposition is found, the algorithm below produces a sequence of vectors $\underline{u}^{(k)}$ which converges to the solution $\underline{u} = C^{-1}\underline{c}$ as $k \to \infty$:
$$\underline{u}^{(k+1)} = F^{-1}N\underline{u}^{(k)} + F^{-1}\underline{c}. \tag{11}$$
Identifying C as the incomplete-data Fisher information F_Y, N as the difference $F - F_Y$, $\underline{u}$ as the jth column of F_Y^{-1}, and $\underline{c}$ as the jth unit vector $\underline{e}_j$ in $\mathbb{R}^n$, the splitting algorithm (11) is equivalent to the column recursion of Corollary 1. The novelty of the recursion of Corollary 1 is that we can identify splitting matrices F that guarantee *monotone convergence* of jth component of $\underline{b}_j^{(k)}$ to the scalar CR bound on $\mathrm{var}_{\underline{\theta}}(\hat{\theta}_j)$ based on *purely statistical considerations* (see next section). Moreover, for general $p \geq 1$, the recursion of Corollary 1 implies that when p parallel versions of (11) are implemented with $\underline{c} = \underline{e}_j$ and $\underline{u}^{(k)} = \underline{u}_j^{(k)}$, $j = 1, \cdots, p$, respectively, the first p rows of the concatenated sequence $[\underline{u}_1^{(k)}, \cdots, \underline{u}_p^{(k)}]$ converge monotonically to the $p \times p$ CR bound on $\mathrm{cov}_{\underline{\theta}}(\hat{\underline{\theta}}^I)$, $\hat{\underline{\theta}}^I = [\hat{\theta}, \cdots, \hat{\theta}_p]^T$. Monotone convergence is important in the statistical estimation context since it ensures that no matter when the iterative algorithm is stopped, a valid lower bound is obtained.
4) The basis for the matrix recursion of Theorem 1 is the geometric series (7). A geometric series approach was also employed in [9, Section 5] to develop a method to speed up the asymptotic convergence of the EM parameter estimation algorithm. This method is a special case of Aitken's acceleration which requires computation of the inverse of the *observed Fisher information* $\hat{F}_Y(\underline{\theta}) = -\nabla_{\underline{\theta}}\nabla_{\underline{\theta}}^T \ln f_Y(Y; \underline{\theta})$ evaluated at successive EM iterates, $\underline{\theta} = \underline{\theta}^{(k)}$, $k = m, m+1, m+2, \cdots$, where m is a large positive integer. If $\hat{F}(\underline{\theta}^{(k)})$ is positive definite, then Theorem 1 of this paper can be applied to iteratively compute this inverse. Unfornately, $\hat{F}_Y(\underline{\theta})$ is not guaranteed to be positive definite except within a small neighborhood $\{\underline{\theta}: \|\underline{\theta} - \hat{\underline{\theta}}\| \leq \delta\}$ of the MLE, so that in practice such an approach may fail to produce a convergent algorithm.

III. Statistical Choice for Splitting Matrix

The matrix F must satisfy $F \geq F_Y$ and must also be easily invertible. For an arbitrary matrix F, verifying that $F \geq F_Y$ could be quite difficult. In this section we present a statistical

approach to choosing the matrix F; F is chosen to be the Fisher information matrix of the complete data that is intrinsic to a related EM parameter estimation algorithm. This approach guarantees that $F \geq F_Y$ due to the Fisher information version of the data processing inequality.

A. Incomplete-Data Formulation

Many estimation problems can be conveniently formulated as an incomplete–complete-data problem. The setup is the following. Imagine that there exists a different set of measurements X taking values in a set $\mathcal{X}$ whose probability density $f_X(x;\underline{\theta})$ is also a function of $\underline{\theta}$. Further assume that this hypothetical set of measurements X is larger and more informative as compared to Y in the sense that the conditional distribution of Y given X is functionally independent of $\underline{\theta}$. X and $\mathcal{X}$ are called the complete data and complete-data space, while Y and $\mathcal{Y}$ are called the incomplete data and incomplete-data space, respectively. This definition of incomplete–complete data is equivalent to defining Y as the output of a $\underline{\theta}$-independent possibly noisy channel having input X. Note that our definition contains as a special case the standard definition [2] whereby X and Y must be related via a deterministic functional transformation $Y = h(X)$, where h: $\mathcal{X} \rightarrow \mathcal{Y}$ is many-to-one.

1) The EM Algorithm

For an initial point $\underline{\theta}^{(0)}$, the EM algorithm produces a sequence of estimates $\{\underline{\hat{\theta}}^{(k)}\}_{k=1}^{\infty}$ by alternating between computing an estimate $Q(\underline{u};\underline{\hat{\theta}}^{(k)})$ of the complete-data log-likelihood function $f_X(x;\underline{u})$, called the expectation (E) step, and finding the maximum of $Q(\underline{u};\underline{\hat{\theta}}^{(k)})$ over $\underline{u}$, called the maximization (M) step [10]:

EM Algorithm: **For** $k = 0, 1, 2, \cdots$, **do:**

$$\begin{aligned} &\text{(E)} \quad \textbf{Compute:} \quad Q(\underline{u};\underline{\hat{\theta}}^{(k)}) = E_{\underline{\hat{\theta}}^{(k)}}\{\log f_X(X;\underline{u})|Y=y\} \\ &\text{(M)} \qquad\qquad \underline{\theta}^{(k+1)} = \textbf{argmax}_{\underline{u}}\, Q(\underline{u};\underline{\theta}^{(k)}) \end{aligned} \tag{12}$$

It can be shown [11] that the sequence $\underline{\hat{\theta}}^{(k)}$ monotonically increases likelihood in the sense that $f_Y(Y;\underline{\hat{\theta}}^{(k+1)}) \geq f_Y(Y;\underline{\hat{\theta}}^{(k)})$, $\forall k$. Furthermore, if the likelihood function is strictly concave over Θ, $\underline{\hat{\theta}}^{(k)}$ converges to the maximum likelihood estimate.

2) A Data Processing Theorem

Assume that a complete-data set X has been specified. For regular probability densities $f_X(x;\underline{\theta})$, $f_Y(y;\underline{\theta})$, $f_{X|Y}(x|y;\underline{\theta})$, we define the associated Fisher information matrices $F_X(\underline{\theta}) = -E_{\underline{\theta}}[\nabla_{\underline{\theta}}\nabla_{\underline{\theta}}^T \ln f_X(X;\underline{\theta})]$, $F_Y(\underline{\theta}) = -E_{\underline{\theta}}[\nabla_{\underline{\theta}}\nabla_{\underline{\theta}}^T \ln f_Y(Y;\underline{\theta})]$, $F_{X|Y}(\underline{\theta}) = -E_{\underline{\theta}}[\nabla_{\underline{\theta}}\nabla_{\underline{\theta}}^T \ln f_{X|Y}(X|Y;\underline{\theta})]$, respectively. The following gives a decomposition for $F_Y(\underline{\theta})$ in terms of $F_X(\underline{\theta})$ and $F_{X|Y}(\underline{\theta})$.

Lemma 1: Let X and Y be random variables which have a joint probability density $f_{X,Y}(x,y;\underline{\theta})$ relative to some product measure $\mu_X \times \mu_Y$. Assume that X is more informative than Y in the sense that the conditional distribution of Y given X is functionally independent of $\underline{\theta}$. Assume also that $\{f_X(x;\underline{\theta})\}_{\underline{\theta}\in\Theta}$ is a regular family of densities with mixed partials $(\partial^2/\theta_i\theta_j)f_X(x;\underline{\theta})$ which are continuous in $\underline{\theta}$ and absolutely integrable in x. Then $\{f_Y(x;\underline{\theta})\}_{\underline{\theta}\in\Theta}$ is a regular family of densities with continuous and absolutely integrable mixed partials, the above-defined Fisher information matrices $F_X(\underline{\theta})$, $F_Y(\underline{\theta})$, and $F_{X|Y}(\underline{\theta})$ exist, are finite, and

$$F_Y(\underline{\theta}) = F_X(\underline{\theta}) - F_{X|Y}(\underline{\theta}). \tag{13}$$

Proof of Lemma 1: Since X, Y has the density $f_{X,Y}(x,y;\underline{\theta})$ with respect to the measure $\mu_X \times \mu_Y$, there exist versions $f_{Y|X}(y|x;\underline{\theta})$ and $f_{X|Y}(x|y;\underline{\theta})$ of the conditional densities. Furthermore, by assumption, $f_{Y|X}(y|x;\underline{\theta}) = f_{Y|X}(y|x)$ does not depend on $\underline{\theta}$. Since $f_Y(y;\underline{\theta}) = \int_{\mathcal{X}} f_{Y|X}(y|x) f_X(x;\underline{\theta})\, d\mu_X$, it is straightforward to show that the family $\{f_Y(y;\underline{\theta})\}_{\underline{\theta}\in\Theta}$ inherits the regularity properties of the family $\{f_X(x;\underline{\theta})\}_{\underline{\theta}\in\Theta}$. Now for any y such that $f_Y(y;\underline{\theta}) > 0$, we have from Bayes' rule,

$$f_{X|Y}(x|y;\underline{\theta}) = \frac{f_{Y|X}(y|x) f_X(x;\underline{\theta})}{f_Y(y;\underline{\theta})}. \tag{14}$$

Note that $f_{X,Y}(x,y;\underline{\theta}) > 0$ implies that $f_Y(y;\underline{\theta}) > 0$, $f_X(x;\underline{\theta}) > 0$, $f_{Y|X}(y|x) > 0$, and $f_{X|Y}(x|y;\underline{\theta}) > 0$. Hence, we can use (14) to express

$$\log f_{X|Y}(x|y;\underline{\theta}) = \log f_X(x;\underline{\theta}) - \log f_Y(y;\underline{\theta}) + \log f_{Y|X}(y|x), \tag{15}$$

whenever $f_{X,Y}(x,y;\underline{\theta}) > 0$. From this relation it is seen that $f_{X|Y}(x|y;\underline{\theta})$ inherits the regularity properties of the X and Y densities. Therefore, since the set $\{(x,y)\colon f_{X,Y}(x,y;\underline{\theta}) > 0\}$ has probability 1, we obtain from (15):

$$\begin{aligned} E_{\underline{\theta}}\left[-\nabla_{\underline{\theta}}^2 \log f_{X|Y}(X|Y;\underline{\theta})\right] &= E_{\underline{\theta}}[-\nabla_{\underline{\theta}}^2 \log f_X(Y;\underline{\theta})] \\ &\quad - E_{\underline{\theta}}[-\nabla_{\underline{\theta}}^2 \log f_Y(X;\underline{\theta})]. \end{aligned}$$

This establishes the lemma. □

Since the Fisher information matrix $F_{X|Y}$ is nonnegative definite, an important consequence of the decomposition of Lemma 1 is the matrix inequality

$$F_X(\underline{\theta}) \geq F_Y(\underline{\theta}). \tag{16}$$

The inequality (16) is a Fisher matrix version of the "data processing theorem" of information theory [3], which asserts that any irreversible processing of data X entails a loss in information in the resulting data Y.

B. Remarks

1) The inequality (16) on F_X is precisely the condition required of the splitting matrix F by the recursive CR bound algorithm (10). Furthermore, in many applications of the EM algorithm, the complete-data space is chosen such that the dependence of X on $\underline{\theta}$ is "uncoupled," so that F_X is diagonal or very sparse. Since many of the problems in which F_Y is difficult to invert are problems for which the EM algorithm has been applied, the Fisher information F_X of the corresponding complete-data space is thus a natural choice for F.
2) If the incomplete-data Fisher matrix F_Y is available, the matrix A in the recursion (8) can be precomputed as

$$A = I - F_X^{-1} F_Y. \tag{17}$$

On the other hand, if the Fisher matrix F_Y is not available, the matrix A in the recursion (8) can be computed directly from $Q(\underline{u},\underline{v}) = E\{\log f(X;\underline{\theta})|Y,\underline{\theta}\}$ arising from the E step of the EM parameter estimation algorithm (12). Note that, under the assumption that exchange of order of differentiation and expectation is justified:

$$\begin{aligned} F_{X|Y}(\underline{\theta}) &= E_{\underline{\theta}}\left[-\nabla_{\underline{u}}^2 E_{\underline{\theta}}[\ln f_{X|Y}(X|Y;\underline{u})|Y]\big|_{\underline{u}=\underline{\theta}}\right] \\ &= E_{\underline{\theta}}[-\nabla^{20} H(\underline{\theta};\underline{\theta})], \end{aligned}$$

where $H(\underline{u};\underline{v}) \stackrel{\text{def}}{=} E_{\underline{v}}\{\log f_{X|Y}(X|Y;\underline{u})|Y=y\}$. We can make use of an identity [10, Lemma 2]: $\nabla^{20} H(\underline{\theta};\underline{\theta}) = -\nabla^{11}(\underline{\theta};\underline{\theta})$. Furthermore, $\nabla^{11} H(\underline{\theta};\underline{\theta}) = \nabla^{11} Q(\underline{\theta};\underline{\theta})$. This

gives the identity $F_{X|Y}(\underline{\theta}) = E_{\underline{\theta}}[\nabla^{11}Q(\underline{\theta};\underline{\theta})]$; yielding an alternative expression to (17) for A:

$$A = F_X^{-1}E_{\underline{\theta}}[\nabla^{11}Q(\underline{\theta};\underline{\theta})]. \tag{18}$$

3) The form $\rho(I - F^{-1}F_Y)$ for the rate of convergence of the algorithms (8) and (10) implies that when $F = F_X$, for rapid convergence the complete-data space $\mathcal{X}$ should be chosen such that X is not significantly more informative than Y relative to the parameter $\underline{\theta}$.
4) The matrix recursion of Theorem 1 can be related to the following Frobenius normalization method for inverting a sparse matrix C:

$$B^{(k+1)} = [I - \alpha C]B^{(k)} + \alpha I, \tag{19}$$

where $\alpha = 1/\|C\|_2$ is the inverse of the Frobenius norm of C. When initialized with $B^{(0)} = I$, the above algorithm converges to C^{-1} as $k \to \infty$. For the case that C is the Fisher matrix F_Y, the matrix recursion (19) can be interpreted as a special case of the algorithm of Theorem 1 for a particular choice of complete data X. Specifically, let the complete data be defined as the concatenation $X = [Y^T, S^T]^T$ of the incomplete data Y and a hypothetical data set $S = [S_1,\cdots,S_m]^T$ defined by the following

$$S = \underline{c}(\underline{\theta}) + W, \tag{20}$$

where $W = [W_1,\cdots,W_m]^T$ are i.i.d. standard Gaussian random variables independent of Y, and $\underline{c} = [c_1,\cdots,c_m]^T$ is a vector function of $\underline{\theta}$. It is readily verified that the Fisher matrix F_S for $\underline{\theta}$ based on observing S is of the form $F_S = \sum_{j=1}^m \nabla^T c_j(\underline{\theta})\nabla c_j(\underline{\theta})$. Now since S and Y are independent, $F_X = F_S + F_Y$, so that if we could choose $c(\underline{\theta})$ such that $F_S = \|F_Y\| \cdot I - F_Y$, the recursion of Theorem 1 would be equivalent to (19) with $F_Y = C$, $F_X^{-1} = \alpha I$, $A = I - \alpha F_Y$. In particular, for the special case that F_Y is functionally independent of $\underline{\theta}$, we can take m equal to n and take the hypothetical data $S = [S_1,\cdots,S_n]^T$ as the n-dimensional linear Gaussian model:

$$S_j = \underline{c}_j\underline{\theta}^T + W_j, \qquad j = 1,\cdots,n,$$

where

$$\underline{c}_j = \left[\|F_Y\|_2 - \sqrt{\lambda_j}\right]\underline{v}_j, \qquad j = 1,\cdots,n,$$

and $\{\nu_1,\cdots,\nu_n\}$ are the eigenvectors and $\{\lambda_1,\cdots,\lambda_n\}$ are the eigenvalues of F_Y. With this definition of S, $F_Y = \sum_{j=1}^n \lambda_j \underline{v}_j\underline{v}_j^T$ is simply the eigendecomposition of the matrix $\|F_Y\| \cdot I - F_Y$, so that $F_X = \|F_Y\| \cdot I = \alpha I$, as required.

IV. Application to ECT Image Reconstruction

We consider the case of positron emission tomography (PET), where a set of m detectors is placed about an object to measure positions of emitted gamma rays. The mathematical formulation of PET is as follows [12]. Over a specified time interval, a number N_b of gamma rays are randomly emitted from pixels b, $b = 1,\cdots,n$, and a number Y_d of these gamma rays are detected at detectors d, $d = 1,\cdots,m$. The average number of emissions in pixels $1,\cdots,n$ is an unknown vector $\underline{\theta} = [\theta_1,\cdots,\theta_n]^T$, called the object intensity. It is assumed that the N_b's are independent Poisson random variables with rates θ_b, $b = 1,\cdots,n$, and the Y_d's are independent Poisson distributed with rates $\mu_d = \sum_{b=1}^n P_{d|b}\theta_b$, where $P_{d|b}$ is the transition probability corresponding to emitter location b and detector location d. For simplicity, we assume that $\mu_d > 0, \forall d$. The objective is to estimate a subset $[\theta_1,\cdots,\theta_p]^T$, $p \ll n$, of the object intensities within a p-pixel region of interest (ROI). In this section we develop the recursive CR bound for this estimation problem.

The log-likelihood function for $\underline{\theta}$ based on $Y = [Y_1,\cdots,Y_m]^T$ is simply

$$\ln f_Y(Y;\underline{\theta}) = \ln \prod_{k=1}^m \frac{[\mu_d]^{Y_d}}{Y_d!} e^{-\mu_d} \tag{21}$$

$$= -\sum_{d=1}^m \mu_d + \sum_{d1=}^m Y_d \ln \mu_D + \text{constant}. \tag{22}$$

From this, the Hessian matrix with respect to $\underline{\theta}$ is simply calculated and, using the fact that $E_{\underline{\theta}}[Y_d] = \mu_d$, the $n \times n$ Fisher information matrix F_Y is obtained:

$$F_Y = \sum_{d=1}^m \frac{1}{\mu_d}\underline{P}_{d|\cdot}^T\underline{P}_{d|\cdot} = \left(\left(\sum_{d=1}^m \frac{P_{d|i}P_{d|j}}{\mu_d}\right)\right)_{i,j=1,\cdots,n}, \tag{23}$$

where $\underline{P}_{d|\cdot} = [P_{d|1},\cdots,P_{d|n}]$ is the dth row of the $m \times n$ system matrix $((P_{j|i}))$. If $m \geq n$, and the linear span of $\{P_{d|\cdot}\}_{d=1}^n$ is $\mathbb{R}^n$, then F_Y is invertible and the CR bound exists. However, even for an imaging system of moderate resolution, e.g., a 256×256 pixel plane, direct computation of the $p \times p$ ROI submatrix $\mathcal{E}^T F_Y^{-1}\mathcal{E}$, $\mathcal{E} = [e_1,\cdots,e_p]$, of the $(256)^2 \times (256)^2$ Fisher matrix F_Y is impractical.

The standard choice of complete data for estimation of $\underline{\theta}$ via the EM algorithm is the set $\{N_{db}\}_{d=1,b=1}^{m,n}$, where N_{db} denotes the number of emissions in pixel b which are detected at detector d [4], [5]. $\{N_{db}\}$ are independent Poisson random variables with intensity $E_{\underline{\theta}}[N_{db}] = P_{d|b}\theta_b$, $d = 1,\cdots,m$, $b = 1,\cdots,n$. By Lemma 1 we know that, with F_X the Fisher information matrix associated with the complete data, $F_X - F_Y$ is nonnegative definite. Thus F_X can be used in Theorem 1 to obtain a monotonically convergent CR bound recursion.

The log-likelihood function associated with the complete-data set $X = [N_{db}]_{d=1,b=1}^{m,n}$ is of similar form to (22):

$$\ln f_X(X;\underline{\theta}) = -\sum_{d=1}^m \sum_{b=1}^n P_{d|b}\theta_b + \sum_{d=1}^m \sum_{b=1}^n N_{db} \ln \theta_b + \text{constants}.$$

The Hessian $\nabla_{\underline{\theta}}^2 \ln f_X(X;\underline{\theta})$ is easily calculated, and, assuming $\theta_b > 0, \forall b$, the Fisher information matrix F_X is obtained as

$$F_X = \text{diag}_b\left(\frac{\sum_{d=1}^m P_{d|b}}{\theta_b}\right), \tag{24}$$

where $\text{diag}_b(a_b)$ denotes a diagonal $n \times n$ matrix with a_b's indexed successively along the diagonal.

Using the results (24) and (23) above, we obtain

$$A = I - F_X^{-1}F_Y = I - \left(\left(\sum_{l=1}^m \frac{\theta_i}{\sum_{d=1}^m P_{d|l}} \frac{P_{l|i}P_{l|j}}{\mu_l}\right)\right)_{i,j=1,\cdots,n}. \tag{25}$$

In many SPECT and PET tomographic geometries, the $m \times n$ ($m \geq n$) system response $((P_{j|i}))$ is a sparse matrix, i.e., its number of nonzero elements is only $O(n)$ as compared to $O(n^2)$ for the nonsparse case. Note, however, that even when the system response matrix is sparse, the matrix A (25) is not generally sparse, and it would appear that the recursive algorithm (10) of Corollary 1 requires $O(n^2)$ memory storage to store the $n \times n$ matrix A. In the present case, however, we only require $O(n)$ memory storage since it is seen that, using (25) in (10), the recursion collapses into a set of p vector recursions which only require storing the n parameters of the vector $\underline{\theta}$, the np entries of $\beta^{(k)}$, and the $O(n)$ nonzero entries of the sparse matrix $((P_{j|i}))$. Because of this feature, we have been able to implement this recursive CR bound on relatively large image reconstruction problems [13].

The rate of convergence of the recursive CR bound algorithm is determined by the maximum eigenvalue $\rho(A)$ of A specified by (25). For a fixed system matrix $((P_{j|i}))$, the magnitude of this eigenvalue will depend on the image intensity $\underline{\theta}$. Assume for simplicity that with probability 1 any emitted gamma ray is detected at some detector, i.e., $\sum_{d=1}^{m} P_{d|b} = 1$ for all b. Since $\text{trace}(A) = \sum_{i=1}^{n} \lambda_i$, where $\{\lambda_i\}_{i=1}^{n}$ are the eigenvalues of A, using (25) it is seen that the maximum eigenvalue $\rho(A)$ must satisfy

$$\frac{1}{n}\text{trace}(A) = 1 - \frac{1}{n}\sum_{l=1}^{m}\frac{\sum_{i=1}^{n} P_{l|i}^2\theta_i}{\sum_{j=1}^{n} P_{l|j}\theta_j} \leq \rho(A) < 1. \quad (26)$$

A consequence of the inequality $(\sum_i P_{l|i}\theta_i)^2 \leq \sum_i \theta_i \cdot \sum_i \theta_i P_{l|i}^2 \theta_i$ is

$$\frac{1}{n}\text{trace}(A) \leq 1 - \frac{1}{n}. \quad (27)$$

where equality occurs if $P_{l|i}$ is independent of i. On the other hand, as the intensity $\underline{\theta}$ concentrates an increasing proportion $1 - \epsilon$ of its mass on a single pixel k_o, e.g.,

$$\theta_i = \begin{cases} (1-\epsilon)\dfrac{n-1}{n}\displaystyle\sum_{b=1}^{n}\theta_b, & i = k_o, \\ \epsilon\dfrac{1}{n}\displaystyle\sum_{b=1}^{n}\theta_b, & i \neq k_o \end{cases},$$

we obtain $(1/n)\text{trace}(A) = 1 - 1/n + O(\epsilon)$. Thus for this case we have, from (26), $1 - 1/n + O(\epsilon) \leq \rho(A) < 1$. Since the number of pixels n is typically very large, this implies that the asymptotic convergence rate of the recursive algorithm will suffer for image intensities which approach that of an ideal point source, at least for this particular choice of splitting matrix F_X.

V. Conclusion and Future Work

We have given a recursive algorithm which can be used to compute submatrices of the CR lower bound F_Y^{-1} on unbiased multidimensional parameter estimation error covariance. The algorithm successively approximates the inverse Fisher information matrix F_Y^{-1} via a monotonically convergent splitting matrix iteration. We have also given a statistical methodology for selecting an appropriate splitting matrix F which involves application of a data processing theorem to a complete-data–incomplete-data formulation of the estimation problem. We are developing analogous recursive algorithms to compute matrix CR-type bounds for constrained and biased estimation, such as those developed in [14], [15].

References

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

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

[3] R. E. Blahut, *Principles and Practice of Information Theory*. Reading, MA: Addison-Wesley, 1987.

[4] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imag.*, vol. MI-1, no. 2, pp. 113–122, Oct. 1982.

[5] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comput. Assisted Tomogr.*, vol. 8, no. 2, pp. 306–316, Apr. 1984.

[6] I. A. Ibragimov and R. Z. Has'minskii, *Statistical Estimation: Asymptotic Theory*. New York: Springer-Verlag, 1981.

[7] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*. New York: Academic, 1970.

[8] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge: Cambrige Univ. Press, 1985.

[9] T. A. Louis, "Finding the observed information matrix when using the EM algorithm," *J. R. Stat. Soc., Ser. B*, vol. 44, no. 2, pp. 226–233, 1982.

[10] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc., Ser. B*, vol. 39, pp. 1–38, 1977.

[11] A. O. Hero and J. A. Fessler, "Asymptotic convergence properties of EM-type algorithms," Commun. Sig. Proc. Lab. (CSPL), Dept. EECS, University of Michigan, Ann Arbor, Technical Report insert 282, April 1993. Also to appear in *Statistica Sinica*, Jan. 1995.

[12] A. O. Hero and L. Shao, "Information analysis of single photon computed tomography with count losses," *IEEE Trans. Med. Imag.*, vol. 9, no. 2, pp. 117–127, June 1990.

[13] A. O. Hero and J. A. Fessler, "A fast recursive algorithm for computing CR-type bounds for image reconstruction problems," in *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf.*, (Orlando, FL), Oct. 1992, pp. 1188–1190.

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

[15] A. O. Hero, "A Cramer–Rao type lower bound for essentially unbiased parameter estimation," MIT Lincoln Laboratory, Lexington, MA, Tech. Rep. 890, Jan. 3, 1992.

Bounds on Achievable Convergence Rates of Parameter Estimators via Universal Coding

Neri Merhav

***Abstract*—Lower bounds on achievable convergence rates of parameter estimators towards the true parameter are derived via universal coding considerations. It is shown that for a parametric class of finite-alphabet information sources, if there exists a universal lossless code whose redundancy decays sufficiently rapidly, then it induces a limitation on the fastest achievable convergence rate of any parameter estimator, at any value of the true parameter, with a possible exception of a vanishingly small subset of parameter values. A specific choice of a universal**

Manuscript received December 1, 1992; revised December 2, 1993. This paper was presented in part at the 1994 IEEE International Symposium on Information Theory, January 1994.

The author is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel.

IEEE Log Number 9403843.