Another important result is about the choice of the preassigned length m. As a rule of thumb, choose the value of m at least $J+1$ more than the given k value. This will give satisfactory results in most cases.

IV. CONCLUSION

It is clear that this method of discrete convolution can be employed more effectively and efficiently if the values of m, k, and J are chosen with utmost care. The best position to place the center point k is the middle point of the preassigned length of the incoming series. This choice of k will much improve the performance of the algorithm. This very important point is not emphasized in the paper by Porsani and Ulrych [1].

REFERENCES

[1] M. Porsani and T. J. Ulrych, "Discrete convolution by means of forward and backward modeling," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 1680–1686, Nov. 1989.

Selection of Observations in Signal Reconstruction

Stanley J. Reeves and Larry P. Heck

***Abstract*— In some signal reconstruction problems, the observation equations can be used as *a priori* information for selecting the best combination of observations before acquiring them. In this correspondence, we define a selection criterion and propose efficient methods for optimizing the criterion with respect to the combination of observations. Our examples illustrate the value of optimized sampling using the proposed methods.**

I. INTRODUCTION

Signal reconstruction uses measurements in one domain to estimate parameters or distributions in another domain. A common approach to the signal reconstruction problem is to model the observed signal y as a linear transformation of x observed in the presence of additive noise; that is

$$y = Ax + u \tag{1}$$

where u is additive noise and $A \in C^{m \times n}$ $(m \geq n)$ is full rank. For this problem, the goal is to reconstruct a good estimate of x given the observed signal y.

In many applications, the relationship between the observation y_i and the original signal x—the ith row of A—is known *a priori*. However, the number of observations of the elements of y that can be made is often limited since collecting the observations may be expensive, time-consuming, or even dangerous for some applications. This limitation exists in computed tomography, magnetic resonance imaging (MRI), and magnetic resonance spectroscopic imaging [1], [2]. We would like to select the limited set of observations that will yield the best possible reconstruction of x, using the known mapping from the original signal to the observations being considered to make the selection. This problem is equivalent to choosing rows for A that correspond to the best set of observations y. Once the best combination of observations is determined, the available resources can be devoted to acquiring only those observations.

Although aspects of this problem have been addressed for specific applications [3]–[7], we have not found the problem as a whole addressed for the general category of signal reconstruction. A related problem has been addressed in the statistics literature on optimal experiment design [8]. However, the goals and constraints of signal reconstruction differ somewhat from that of experiment design. In experiment design, one is typically concerned with obtaining good predictions rather than good regression parameters. Furthermore, optimal design usually assumes either total freedom in the choice of A or freedom within the confines of a specific regression model, such as a polynomial model.

We pose observation selection as a problem of candidate selection. We consider a candidate set of observations $\{y_{i_k} : k = 1, \ldots, M\}$, where $M > m$. The problem then is to find a combination of m out of M observations by minimizing an appropriate criterion based on the rows of A corresponding to the candidate observations. This general formulation of the problem allows us to address many applications of observation selection. To solve this problem, we must first define an appropriate criterion in terms of the candidate matrix A. Then we must identify an efficient means for optimizing the criterion with respect to the choice of observations.

II. OPTIMALITY CRITERION

If we let $A^\dagger = (A^H A)^{-1} A^H$, then the least-squares estimate of x is given by

$$\begin{aligned} x_{\mathrm{LS}} &= A^\dagger y \\ &= A^\dagger A x + A^\dagger u \\ &= x + A^\dagger u \ . \end{aligned} \tag{2}$$

Thus, we would like to choose A so that x_{LS} is in some sense a reliable estimate of x. That is, we want to reduce the uncertainty in the solution introduced by $A^\dagger u$. The sense in which the solution uncertainty is reduced is controlled by the criterion we adopt for the minimization problem. If we adopt a minimum 2-norm criterion on the reconstructed signal, we must minimize

$$\begin{aligned} \|x - x_{\mathrm{LS}}\|^2 &= \|x - (x + A^\dagger u)\|^2 \\ &= \|A^\dagger u\|^2 \end{aligned} \tag{3}$$

for an appropriate choice of A. However, since we do not know u, we must settle for some other measure upon which to base our choice. Therefore, we assume that u is zero-mean unit variance white noise and take the expected value of (3). (Note that correlated noise with correlation matrix R_{uu} can be handled by premultiplying A by the whitening operator $R_{uu}^{-\frac{1}{2}}$ and then proceeding with $R_{uu}^{-\frac{1}{2}} A$ in place of A.) The minimization criterion then becomes

$$\begin{aligned} E\{\|A^\dagger u\|^2\} &= E\{u^H A (A^H A)^{-1} (A^H A)^{-1} A^H u\} \\ &= \text{trace}\, A (A^H A)^{-1} (A^H A)^{-1} A^H \\ &= \text{trace}\, (A^H A)^{-1} A^H A (A^H A)^{-1} \\ &= \text{trace}\, (A^H A)^{-1} \\ &= \|A^\dagger\|_F^2 \end{aligned} \tag{4}$$

Manuscript received December 23, 1993; revised August 24, 1994. The associate editor coordinating the review of this paper and approving it for publication was Dr. Barry Sullivan.

S. J. Reeves is with the Department of Electrical Engineering, Auburn University, Auburn, AL 36849 USA.

L. P. Heck is with the Acoustics and Radar Technology Laboratory, SRI International, Menlo Park, CA 94025 USA.

IEEE Log Number 9408245.

where $\|\cdot\|_F$ represents the Frobenius norm (F-norm). (Note that this criterion is proportional to the expected value of (3) regardless of the variance of u.)

III. OPTIMIZATION STRATEGY

Unfortunately, a simple definition of the minimization criterion is not sufficient to solve the problem. We know of no method for optimizing this criterion without considering either explicitly or implicitly all possible combinations of rows. Thus, we must also deal with a combinatoric problem if all possible combinations of m of M rows are evaluated. For example, suppose that we want to choose the best 10 of 30 rows. In this case, we must evaluate the criterion for over 30 000 000 combinations if exhaustive search is used! Clearly, exhaustive search is impractical even for problems of moderate size. Rather than using exhaustive search, we consider two methods used in pattern recognition for large-scale feature selection: sequential backward selection (SBS) [9] and branch-and-bound (B&B) [10]. The B&B algorithm is optimal and more efficient than exhaustive search, but it is significantly more complex than the SBS algorithm. Even though the SBS algorithm is suboptimal, in general it still performs quite well, as we will demonstrate in Section IV, and may be the preferred algorithm for applications involving a large number of rows.

A. Sequential Backward Selection

The idea behind SBS is to sequentially eliminate one row at a time from the candidate set until m rows remain. First, we form an A matrix from all M candidate rows. Then, we determine the row that when deleted from A has the least adverse affect on the signal reconstruction performance (i.e., best value of the criterion). Once the worst row has been eliminated, we follow the same procedure with a new A containing $M-1$ candidate rows. We continue the process until we are left with an A matrix containing m rows as desired. These rows represent the choice of observations to be made.

The SBS approach is suboptimal in the sense that it is not guaranteed to find the combination that minimizes the criterion, since the algorithm has no ability to backtrack; that is, once an observation (row) is removed from the set, it cannot be added at a later stage. This early commitment to delete an observation overlooks many combinations of observations that might perform better than the selected set. However, the fact that SBS does not backtrack significantly prunes the search space, and the algorithm generally provides near-optimal results.

In addition to the pruning of the search space, the SBS approach facilitates a sequential update of the criterion without having to compute it explicitly for each combination of rows [11]. This provides a significant computational savings. If A is a sparse matrix, which is often the case in signal reconstruction applications, it may be possible to reduce the computation and storage requirements even more.

For the expected value criterion in (4), we need to evaluate $\text{trace}\,(A^HA)^{-1}$. This requires that the inverse matrix or an eigen-decomposition be available. Since we are sequentially eliminating one row at a time, we can compute a single inverse for A with all M rows and then use the Sherman-Morrison matrix inversion formula to update the inverse of A^HA [11]. The update formula for the new inverse matrix $(\tilde{A}^H\tilde{A})^{-1}$ is as follows:

$$(\tilde{A}^H\tilde{A})^{-1} = (A^HA)^{-1} + \frac{1}{1-a_i(A^HA)^{-1}a_i^H}(A^HA)^{-1}a_i^Ha_i(A^HA)^{-1} \tag{5}$$

where a_i is the deleted row. We can simplify this expression once we take the trace

$$\begin{aligned}\text{trace}\,(\tilde{A}^H\tilde{A})^{-1} &= \text{trace}\,(A^HA)^{-1} \\ &\quad + \text{trace}\,\frac{1}{1-a_i(A^HA)^{-1}a_i^H}((A^HA)^{-1}a_i^H)((A^HA)^{-1}a_i^H)^H \\ &= \text{trace}\,(A^HA)^{-1} \\ &\quad + \frac{1}{1-a_i(A^HA)^{-1}a_i^H}\text{trace}\,a_i(A^HA)^{-1}(A^HA)^{-1}a_i^H \\ &= \text{trace}\,(A^HA)^{-1} \\ &\quad + \frac{1}{1-a_i(A^HA)^{-1}a_i^H}a_i(A^HA)^{-1}(A^HA)^{-1}a_i^H \ .\end{aligned} \tag{6}$$

At each row elimination step, the term $\text{trace}\,(A^HA)^{-1}$ on the right-hand side of (6) is a common term in the criterion for all i. Therefore, we can simply compare the term that is a function of the row that is left out

$$\frac{1}{1-a_i(A^HA)^{-1}a_i^H}a_i(A^HA)^{-1}(A^HA)^{-1}a_i^H \ . \tag{7}$$

The row yielding the lowest value for this term should be eliminated. That is, we eliminate the row for which the expected value increases the least in its absence.

The primary drawback to this approach as stated is that it requires computation and storage of the initial inverse matrix $(A^HA)^{-1}$. Subsequent inverses can be computed efficiently using the matrix update formula in (5); however, one must still store the inverse matrix at the current step. In many applications, A has a sparse matrix representation. Unfortunately, sparsity is generally destroyed by the inversion process. Therefore, if A is very large, $(A^HA)^{-1}$ may be too large to compute and represent in computer memory. This problem can be circumvented, since only the vector quantity $(A^HA)^{-1}a_i^H$ is actually needed to compare rows. This quantity can be computed by solving the linear system $A^HAv_i = a_i^H$ for v_i. If A is sparse, we can solve the linear system efficiently by an iterative method such as conjugate gradients. The criterion can then be rewritten as $\frac{1}{1-a_iv_i}v_i^Hv_i$.

B. Branch-and-Bound

The B&B algorithm is an optimal search method in which all possible subsets of k out of N rows of the matrix A are implicitly inspected without exhaustive search. The algorithm has its roots in integer programming problems [12] and was applied to the feature selection problem in pattern recognition by Narendra and Fukunaga [10] . For the observation selection problem, the B&B algorithm can be used to efficiently organize the search process so that the enumeration of many candidate rows of the A matrix can be avoided without undermining the optimality of the observation selection process. The algorithm is a top down search, starting with the complete set of possible rows as the top node. The successive levels below the top node are branched to by removing one of the rows from the complete set for each level in depth. Each level is represented by X_i where i is the number of remaining rows (see Fig. 1). The number of sets at any level in the tree is determined by a uniqueness rule, which ensures that every possible candidate subset is only enumerated once [13]. (The figure does not necessarily show all possible nodes for each level.)

For the B&B algorithm to be optimal, the evaluation criterion must be monotonic. The monotonicity property requires that for nested sets of rows, $X_1, X_2, \ldots, X_k$, the criterion functional J satisfy

$$X_1 \subset X_2 \subset \cdots \subset X_k \quad \Rightarrow \quad J(X_1) \geq J(X_2) \geq \cdots \geq J(X_k).$$

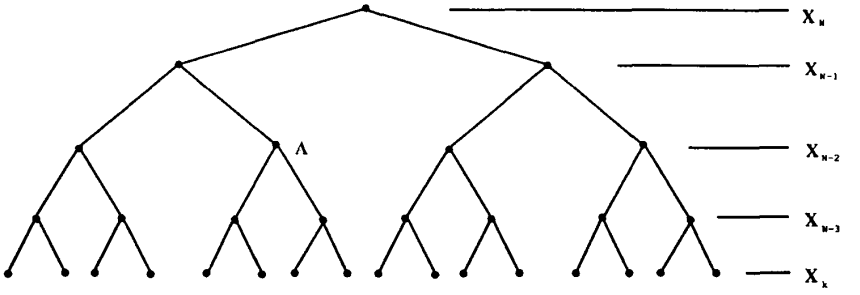


Fig. 1. Example search tree for branch and bound.

Suppose that some branches of a solution tree have already been explored to the final depth of search and that the criterion function for the current best set of rows X_d equals J^*, that is

$$J(X_d) = J^* \ .$$

Now, consider a node at the l^{th} level ($l > d$) of an unexplored section of the tree where the value of the criterion functional is greater than the threshold J^*, i.e.

$$J(X_l) > J^* \ .$$

Due to the monotonicity property, it is not necessary to evaluate the sets of rows corresponding to the subtree branching from this node, since all these sets will be inferior to the current best set. In Fig. 1, if A is such a node at the lth level, then six candidate sets do not need to be evaluated. This pruning results in considerable savings because only a fraction of all sets needs to be evaluated.

Since the B&B algorithm is only applicable if the criterion that is used is monotonic, we explore this issue first for the expected value criterion.

Theorem 1: The criterion trace $(A^H A)^{-1}$ increases monotonically as rows are eliminated from A.

Proof: From (6), we only need to show that

$$\frac{1}{1 - a_i(A^H A)^{-1} a_i^H} a_i (A^H A)^{-1} (A^H A)^{-1} a_i^H > 0 \qquad (8)$$

for all i. The existence and symmetry of $(A^H A)^{-1}$ imply that it is positive definite. Thus, we know that $a_i(A^H A)^{-1}(A^H A)^{-1} a_i^H > 0$. It remains only to show that $a_i(A^H A)^{-1} a_i^H \leq 1$. We use a singular value decomposition of A. Let $A = U\Sigma V^H$. Then, $a_i = u_i \Sigma V^H$, where u_i is the ith row of U and u_{ij} is the ijth element of U. Note that $U^H U = V^H V = I$. It follows that

$$\begin{aligned} a_i(A^H A)^{-1} a_i^H &= u_i \Sigma V^H (U \Sigma^H V^H V \Sigma U^H)^{-1} V \Sigma^H u_i^H \\ &= u_i \Sigma (\Sigma^H \Sigma)^{-1} \Sigma^H u_i^H \\ &= \Sigma_{j=1}^{n} |u_{ij}|^2 \\ &\leq \Sigma_{j=1}^{m} |u_{ij}|^2 \\ &= 1. \end{aligned} \qquad (9)$$

(If the equality in (9) holds, then the inverse of the matrix does not exist with row i removed.)

This argument can be used as each row is eliminated to show that the criterion always increases. ■

IV. EXPERIMENTS

To demonstrate the potential of observation selection in an actual application, we generated a simple simulation example. This example is analogous to 2-D and 3-D reconstruction in magnetic resonance spectroscopic imaging (MRSI), where region of support information is available from a proton density scout image [1], [2]. We considered a 28-point 1-D signal. A region of support given by points (0, ..., 7, 8, 9, 20, ..., 24) was assumed to be known *a priori*. In MRI, such a region of support may be derived from the known characteristics of the object being imaged. MRI data acquisition involves acquiring samples in the spatial frequency domain. The original A matrix therefore represents a transformation of the signal from the spatial domain to the frequency domain; that is, each row of A corresponds to a Fourier transform of the signal evaluated at a particular frequency location. We assumed noise with unit variance at each frequency-domain sample.

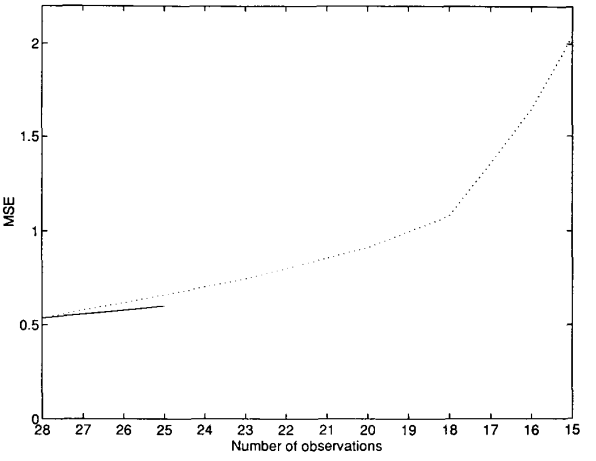


Fig. 2. MSE versus number of observations: equally spaced samples (solid line) and SBS selection of samples (dotted line).

The region of support given implies that there are 15 free variables, corresponding to the 15 unknown values in the region of support. Ideally, then, we should be able to solve for these values by taking only 15 frequency-domain samples. We investigated the effect of equally spaced sampling versus SBS and B&B sampling. Equally spaced sampling refers to sampling in equal intervals in the frequency domain over the continuous frequency range $(-\frac{\pi}{T}, \frac{\pi}{T}]$, where T is the sample spacing in the spatial domain. (Note that this describes the entire frequency range, assuming that the original signal is sampled without aliasing.) Sampling the frequency domain with equal spacing is a simple heuristic approach that has intuitive appeal. Fig. 2 shows mean squared error (MSE) as the number of equally spaced samples is decreased from a high of 28 samples to 15 samples. (The solid line shows MSE for equally spaced samples.) MSE grows slowly at first. However, at 24 equally spaced samples, the matrix $A^H A$ becomes singular and the criterion cannot be computed. (The abrupt end of the solid line reflects this fact.) To see why this happens,

note that the (k,l) entries of the $N \times 15$ matrix A in this case are given by $exp(-j2\pi kn(l)/N)$ for $k = [0,\ldots,N-1]$ and $n(l) = [0,\ldots,9,20,\ldots,24]$. It is easy to see that at least two of the 15 columns of the matrix will always be identical when $N \leq 24$, while the rest are orthogonal. Consider, for example, the 24×15 case. The first and the fifteenth columns will both be all 1's. Thus, $A^H A$ is singular.

To demonstrate the power of the SBS algorithm, we chose samples from a set of 28 equally spaced candidate samples. The plot (dotted line) shows MSE as the number of optimized samples is decreased. Note that for the proposed technique, MSE increases slowly and predictably as the number of optimally selected samples decreases.

We also optimized the selection criterion using B&B. B&B found a combination of samples yielding an MSE of 1.9073, as opposed to 2.0421 for SBS. This illustrates the suboptimality of the SBS algorithm. However, the price paid for the optimal solution was substantial. Whereas the SBS solution required about 1s on a Sun SPARC 10, the B&B solution required several days! Whether B&B is worth this computational effort in a particular application depends on the cost of acquiring each sample as well as the number of observations to be considered.

V. CONCLUSIONS

The experiments illustrate two points. 1) Choosing the observations using a matrix-based optimality criterion rather than by a simple heuristic method, such as equally spaced sampling in some domain, may improve the noise sensitivity of the reconstruction dramatically, and 2) the SBS algorithm works quite well. SBS yielded an MSE that was only 7% higher than optimal and with a computational cost orders of magnitude less than B&B. While the example presented is only a single anecdotal case, it illustrates that observation selection may in some circumstances improve the results considerably.

Criterion-based observation selection has the potential to improve signal reconstructions significantly beyond simple heuristic selection schemes. We have shown that the MSE criterion increases monotonically as samples are eliminated, thus allowing the use of B&B to find the optimal combination. However, the speed and performance of the suboptimal SBS scheme relative to B&B makes it an attractive alternative for large-scale observation selection problems.

REFERENCES

[1] X. Hu, D. N. Levin, P. C. Lauterbur, and T. A. Spraggins, "SLIM: Spectral localization by imaging," *Magnet. Reson. Med.*, vol. 8, pp. 314–322, 1988.

[2] E. M. Stokely and D. B. Twieg, "Functional image reconstruction enhancement (FIRE) for MR spectroscopic and nuclear medicine images," in *SPIE Medical Imaging VI*, 1992.

[3] Y. Cao and D. N. Levin, "Feature-guided acquisition and reconstruction of MR images," in *Information Processing in Medical Imaging* (H. H. Barrett and A. F. Gmitro, Eds.). New York: Springer-Verlag, 1993, pp. 278–292.

[4] W. H. Weedon, S. W. McKnight, and A. J. Devaney, "Selection of optimal angles for inversion of multiple-angle ellipsometry and reflectometry equations," *J. Opt. Soc. Amer. A*, vol. 8, pp. 1881–1891, Dec. 1991.

[5] C. H. Chen, A. Meyyappa, G. Wade, H. Shimizu, N. Chubachi, and J. Kushibiki, "Limited-angle planar diffraction tomography using optimal incident angles," in *Acoust. Imaging*, vol. 17, pp. 657–667, 1985.

[6] D. S. Chen and J. P. Allebach, "Analysis of error in reconstruction of two-dimensional signals from irregularly spaced samples," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 173–180, February 1987.

[7] M. Bertero, P. Boccacci, and E. R. Pike, "On the recovery and resolution of exponential relaxation rates from experimental data II. The optimum choice of experimental sampling points for laplace transform inversion," *Proc. Royal Soc. London A*, vol. 393, pp. 51–65, 1984.

[8] G. A. F. Seber, *Linear Regression Analysis*. New York: Wiley, 1977.

[9] T. Marill and D. M. Green, "On the effectiveness of receptors in recognition systems," *IEEE Trans. Inform. Theory*, vol. IT-9, pp. 11–17, 1963.

[10] P. M. Narendra and K. Fukunaga, "A branch and bound algorithm for feature subset selection," *IEEE Trans. Comput.*, vol. C-26, Sept. 1977.

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

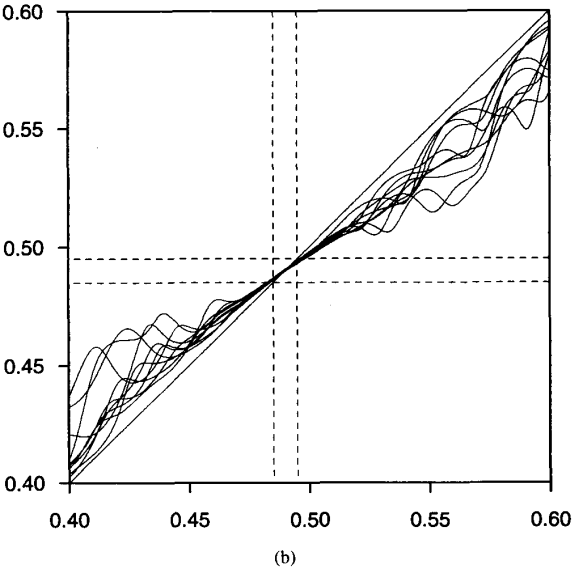
[12] A. H. Land and A. G. Doig, "An automatic method of solving discrete programming problems," *Econometrica*, vol. 28, no. 3, 1960.

[13] P. A. Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach*. Englewood Cliffs, NJ: Prentice Hall, 1982.

Correction to "Iterative Filtering for Multiple Frequency Estimation"

Ta-Hsin Li and Benjamin Kedem

In [1], due to an error in the graph generating code, the curves in Fig. 2(b) are incorrect. The correct Fig. 2(b) is shown.



(b)

Fig. 2. Least-squares mapping $\hat{\varphi}_{LS}(f)$ in the case of two sinusoids. The data length is $n = 100$ for each of the 10 realizations plotted, and the SNR is 0 dB per sinusoid. (b) $\eta = 1$ for closely spaced frequencies with $f_1 = 0.485$ and $f_2 = 0.495$.

Manuscript received August 22, 1994; revised September 1, 1994.

T. H. Li is with the Department of Statistics, Texas A&M University, College Station, TX 77843 USA.

B. Kedem is with the Department of Mathematics and the Institute for Systems Research, University of Maryland, College Park, MD 20742 USA.

IEEE Log Number 9408218.