DATA SELECTION FOR DETECTION OF KNOWN SIGNALS: THE RESTRICTED-LENGTH MATCHED FILTER

Charles K. Sestok

Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, MA 02139

ABSTRACT

Data selection algorithms in detection search for a small subset of the available data that is sufficient for making an accurate decision. This paper considers data selection for detection of a known signal in colored Gaussian noise. In our model, the performance of the matched filter detector for a specific subset is parameterized by a quadratic form. Selection of the best subset leads to a combinatorial optimization problem using the quadratic form as the objective function. Simulations show that heuristic search algorithms often find good solutions for the selected subset. Additionally, if the noise has a banded covariance matrix, a dynamic programming algorithm finds the optimal solution for any subset size.

1. INTRODUCTION

Data selection algorithms identify a subset of data for subsequent signal processing. A variety of hardware architectures amenable to data selection have been proposed. Specific examples include multiple-antenna wireless communication systems [4], and wireless sensor networks [2]. In these systems, collecting data can provide a large portion of the operating cost, so collecting only a subset of available measurements can yield significant resource savings. Additionally, if the density of sensors or antenna elements is sufficient, a small subset of the data can yield performance close to that possible with the full data set.

In practice, selection algorithms should be tuned to a particular underlying signal processing task. We focus on the matched filter algorithm for detection of a known signal in zero-mean Gaussian noise. In this model, the detector processes data denoted by the N-dimensional random vector $\mathbf{x}$. The target signal is given by $\mathbf{s}$ and the noise covariance is $\mathbf{\Lambda}$. The general form of the detector derived from the likelihood ratio test for this model [6] is

$$\mathbf{s}^T\mathbf{\Lambda}^{-1}\mathbf{x} \underset{\hat{H}=H_0}{\overset{\hat{H}=H_1}{\gtrless}} \eta. \tag{1}$$

The receiver operating characteristic (ROC) for this detector is parameterized by the quadratic form $\mathbf{s}^T\mathbf{\Lambda}^{-1}\mathbf{s}$. Increasing it improves the entire ROC. This quadratic form reduces to the traditional expression for signal-to-noise ratio when the noise is white, i.e. $\mathbf{\Lambda} = \sigma^2\mathbf{I}$. We refer to it as equivalent signal-to-noise ratio, and denote it by SNR.

2. THE RESTRICTED-LENGTH MATCHED FILTER

We formulate the data selection problem by restricting the size of the detector input to K measurements. The detector for any fixed subset is a matched filter, so the critical issue involves choice of the best such subset. We refer to a detector designed under the data selection constraint as a restricted-length matched filter (RLMF).

We can identify selected subsets using diagonal matrices with boolean entries with K non-zero entries on the main diagonal. One example is

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{2}$$

The notation for a selected data set with the identity of each data component preserved is $\mathbf{x}_g = \mathbf{G}\mathbf{x}$. Using the matrix displayed in equation (2), the selected data is $\mathbf{x}_g = [x_1\ 0\ x_3]^T$.

A compact notation for selection can be denoted by matrix multiplication between $\mathbf{x}$ and a non-square boolean matrix. Let $\tilde{\mathbf{G}}$ be a $K \times N$ matrix restricted such that each row has a single non-zero entry, and each column contains

This work was supported in part by participation in the Advanced Sensors Collaborative Technology Alliance (CTA) sponsored by the U.S. Army Research Laboratory under Cooperative Agreement DAAD19-01-2-008, in part by the Space and Naval Warfare Systems Center-San Diego and the Defense Advanced Research Projects Agency (DARPA) IXO under Contract N6601-00-C-8054, and in part by the Texas Instruments Leadership University Consortium Program. The views expressed are those of the authors and do not reflect the official policy or position of the U.S. Government.

at most one non-zero entry. An example corresponding to the same selection measurements shown in equation (2) is

$$\tilde{\mathbf{G}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{3}$$

Likewise, the selected data vector is denoted by $\tilde{\mathbf{x}}_g = \tilde{\mathbf{G}}\mathbf{x}$.

For a fixed subset indicated by $\mathbf{G}$, the RLMF is designed according to equation (1) using the covariance and the conditional mean of the random vector $\tilde{\mathbf{x}}_g$. The new mean and covariance are

$$\begin{array}{rcccl} E[\tilde{\mathbf{x}}_g; H_1] & = & \tilde{\mathbf{s}}_g & = & \tilde{\mathbf{G}}\mathbf{s} \\ E[\tilde{\mathbf{x}}_g\tilde{\mathbf{x}}_g^T; H_i] & = & \mathbf{\Lambda}_{\tilde{\mathbf{x}}_g} & = & \tilde{\mathbf{G}}\mathbf{\Lambda}\tilde{\mathbf{G}}^T, \quad i = 0, 1. \end{array} \tag{4}$$

The densities under H_0 and H_1 are the Gaussian densities determined by these parameters.

Since the probabilistic model for the hypothesis test after data selection has the same structure as the original model, the optimal detector is the matched filter determined by $\tilde{\mathbf{s}}_g$ and $\mathbf{\Lambda}_{\tilde{\mathbf{x}}_g}$. Thus, for a fixed subset of measurements, represented by a particular instance of $\mathbf{G}$, the performance of the associated whitened matched filter is given by the quadratic form

$$\mathrm{SNR}(\mathbf{G}) = \tilde{\mathbf{s}}_g^T \mathbf{\Lambda}_{\tilde{\mathbf{x}}_g}^{-1} \tilde{\mathbf{s}}_g. \tag{5}$$

The notation in equation (5) represents the SNR quadratic form for a subset of sensors indicated either symbolically as $\mathbf{G}$ or explicitly, such as $\{x_1, x_3, x_8\}$.

Selecting the data subset yielding the best RLMF is a combinatorial optimization problem. There are $\begin{pmatrix} N \\ K \end{pmatrix}$ selection matrices that satisfy the subset size constraint. Optimization of $\mathrm{SNR}(\mathbf{G})$ over this finite set produces the best detector. When the noise is white, selection of the subset that maximizes $\|\tilde{\mathbf{s}}_g\|^2$ maximizes $\mathrm{SNR}(\mathbf{G})$. In the general case, however, the maximum energy selection rule may work poorly.

To cope with non-white noise, a variety of heuristic solutions have been suggested [3, 5]. Common approaches involve tests for local optima, branch and bound search, and incremental searches that add measurements to maximize the change in SNR, referred to as greedy algorithms.

3. GREEDY SEARCH ALGORITHMS FOR THE RLMF

This section focuses on forward and backward greedy selection algorithms. The forward greedy algorithm builds a selected subset by adding measurements that maximize the incremental increase in SNR, terminating after selecting a K element subset. The backward greedy algorithm discards measurements one at a time, minimizing SNR loss in each stage, again terminating with K measurements. These algorithms initialize more complicated search procedures such as the local search and the branch-and-bound search.

An expression for the changes in SNR is derived using a formula for the inverse of a partitioned matrix [5]. To apply this result, we require notation to identify individual columns of $\mathbf{\Lambda}_{\tilde{\mathbf{x}}_g}$. Let $\mathbf{G}$ denote a subset, and consider the change in SNR from adding x_i, such that $\mathbf{G}_{ii} = 0$. Let $\mathbf{z}_i$ contain the elements of the ith column of $\mathbf{\Lambda}$ corresponding to the rows identified in $\mathbf{G}$. In terms of these quantities, the SNR after adding x_i to the subset is

$$\mathrm{SNR}(\{\mathbf{G} \cup x_i\}) = \tilde{\mathbf{s}}_g^T \mathbf{\Lambda}_{\tilde{\mathbf{x}}_g}^{-1} \tilde{\mathbf{s}}_g + \frac{\left(s_i - \mathbf{z}_i^T \mathbf{\Lambda}_{\tilde{\mathbf{x}}_g}^{-1} \tilde{\mathbf{s}}_g\right)^2}{[\mathbf{\Lambda}]_{ii} - \mathbf{z}_i^T \mathbf{\Lambda}_{\tilde{\mathbf{x}}_g}^{-1} \mathbf{z}_i}. \tag{6}$$

The forward SNR increment, the change in SNR due to adding measurement i to the subset identified in $\mathbf{G}$, is given by

$$\Delta^+(\mathbf{G}, i) = \frac{\left(s_i - \mathbf{z}_i^T \mathbf{\Lambda}_{\tilde{\mathbf{x}}_g}^{-1} \tilde{\mathbf{s}}_g\right)^2}{[\mathbf{\Lambda}]_{ii} - \mathbf{z}_i^T \mathbf{\Lambda}_{\tilde{\mathbf{x}}_g}^{-1} \mathbf{z}_i}. \tag{7}$$

The SNR increment when a measurement is discarded has a similar form. For a subset where $\mathbf{G}_{jj} = 1$, let $\mathbf{G}'(j)$ indicate the selection matrix formed when j is discarded. The reduction in SNR is $\Delta^+(\mathbf{G}'(j), j)$ using the form of equation (7). For convenience, the SNR reduction is also written as $\Delta^-(\mathbf{G}, j)$.

The greedy algorithms determine the measurement to select or discard by computing the SNR increments for each possible measurement. The forward greedy algorithm begins with no selected measurements and adds the unselected measurement that maximizes $\Delta^+(\mathbf{G}, i)$; the backward greedy algorithm begins with $\mathbf{G} = \mathbf{I}$, and removes the measurement that minimizes $\Delta^-(\mathbf{G}, i)$.

4. CONDITIONS FOR AN EXACT SOLUTION WITH MAXIMUM SIGNAL ENERGY SELECTION

In certain cases, selection algorithms less computationally complex than the greedy search can find the optimal RLMF subset. As mentioned in section 2, the maximum signal energy subset is optimal when $\mathbf{\Lambda} = \sigma^2\mathbf{I}$. This section generalizes this result, providing a sufficient condition for the optimality of the maximum signal energy selection rule in cases when $\mathbf{\Lambda}$ is not diagonal. In this section, $\mathbf{G}^*$ denotes the K element maximum signal energy subset, and $\tilde{\mathbf{s}}_{g^*}$ denotes the corresponding signal measurements.

The sufficient condition follows from bounds on $\mathrm{SNR}(\mathbf{G})$. Let $d_{\max}(\mathbf{\Lambda})$ and $d_{\min}(\mathbf{\Lambda})$ be maximum and minimum eigenvalues of $\mathbf{\Lambda}$. The relationship between $\mathbf{\Lambda}$ and $\mathbf{\Lambda}_{\tilde{\mathbf{x}}_g}$ [1, pg. 189] indicates that the inequalities

$$d_{\max}(\mathbf{\Lambda}) \geq d_{\max}(\mathbf{\Lambda}_{\tilde{\mathbf{x}}_g}) \geq d_{\min}(\mathbf{\Lambda}_{\tilde{\mathbf{x}}_g}) \geq d_{\min}(\mathbf{\Lambda}) \tag{8}$$

hold for any choice of $\mathbf{\Lambda}_{\tilde{\mathbf{x}}_g}$. Applying this result, we see

that $\mathrm{SNR}(\mathbf{G})$ is bounded by

$$\frac{\|\tilde{\mathbf{s}}_g\|^2}{d_{\min}(\mathbf{\Lambda})} \geq \mathrm{SNR}(\mathbf{G}) \geq \frac{\|\tilde{\mathbf{s}}_g\|^2}{d_{\max}(\mathbf{\Lambda})} \tag{9}$$

for any subset $\mathbf{G}$.

The maximum energy subset remains the source for the optimal RMF if the inequality

$$\mathrm{SNR}(\mathbf{G}^*) \geq \mathrm{SNR}(\mathbf{G}) \tag{10}$$

holds for all K sample subsets. An application of the lower bound in equation (9) to $\mathrm{SNR}(\mathbf{G}^*)$ and the upper bound to $\mathrm{SNR}(\mathbf{G})$ shows that the condition

$$\frac{\|\tilde{\mathbf{s}}_{g^*}\|^2}{d_{\max}(\mathbf{\Lambda})} \geq \frac{\|\tilde{\mathbf{s}}_g\|^2}{d_{\min}(\mathbf{\Lambda})} \tag{11}$$

guarantees that $\mathrm{SNR}(\mathbf{G}^*)$ exceeds $\mathrm{SNR}(\mathbf{G})$. If the inequality holds for every arrangement of $\mathbf{G}$ with K elements, the inequality gives a sufficient condition for the maximum energy subset to yield the best RMF. The condition can be checked by comparing $\mathbf{G}^*$ with the subset that has the second largest signal energy. If (11) holds for this subset, it holds for all other K element subsets. In cases where the condition does not hold, the maximum signal energy selection rule is not optimal, and the greedy search algorithm often performs better.

5. WORST-CASE PERFORMANCE OF THE GREEDY ALGORITHM

In general, the greedy algorithm is sub-optimal, but we can characterize its worst-case behavior. If the SNRs for the optimal subset and one selected by a greedy algorithm are $\mathrm{SNR}(\mathbf{G}_{\mathrm{opt}})$ and $\mathrm{SNR}(\mathbf{G}_{\mathrm{gr}})$, the approximation ratio for the greedy algorithm is $\alpha = \mathrm{SNR}(\mathbf{G}_{\mathrm{opt}})/\mathrm{SNR}(\mathbf{G}_{\mathrm{gr}})$. If $\mathbf{G}_{\mathrm{gr}} = \mathbf{G}_{\mathrm{opt}}$, the ratio is $\alpha = 1$. Large values of α correspond to poor RLMF performance.

The upper bound on α follows from the bounds in equation (9). Applying these bounds to the expression for α, we find that the approximation ratio is bounded by

$$\alpha \leq \frac{\|\mathbf{s}_{\mathrm{opt}}\|^2}{\|\mathbf{s}_{\mathrm{gr}}\|^2}\frac{d_{\max}(\mathbf{\Lambda})}{d_{\min}(\mathbf{\Lambda})} = \frac{\|\mathbf{s}_{\mathrm{opt}}\|^2}{\|\mathbf{s}_{\mathrm{gr}}\|^2}\kappa(\mathbf{\Lambda}), \tag{12}$$

where $\kappa(\mathbf{\Lambda})$ denotes the condition number of $\mathbf{\Lambda}$ and $\mathbf{s}_{\mathrm{opt}}$ and $\mathbf{s}_{\mathrm{gr}}$ are placeholders for the signal entries selected by $\mathbf{G}_{\mathrm{opt}}$ and $\mathbf{G}_{\mathrm{gr}}$. Note that equation (11) can be rearranged to resemble equation (12).

Thus, the greedy algorithm can perform poorly when $\kappa(\mathbf{\Lambda})$ is large and the ratio between $\|\mathbf{s}_{\mathrm{opt}}\|$ and the norm of any other selected subset exceeds some finite constant. Unfortunately, there are instances where the greedy algorithm selects a subset that meets the upper bound in equation (12) with equality [7].

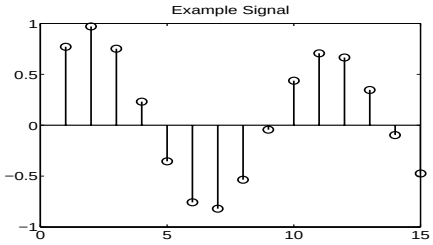

Fig. 1. Signal $\mathbf{s}$ for the performance comparison simulations

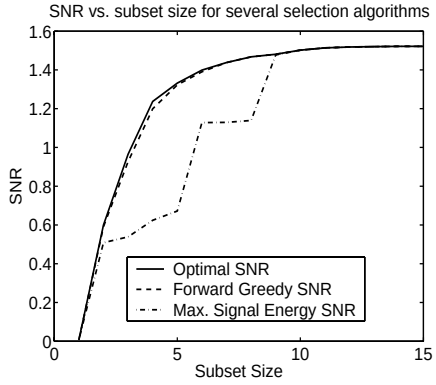

Fig. 2. SNR versus K for Forward Greedy selection algorithm, maximum signal energy selection, and exhaustive search selection.

6. PERFORMANCE COMPARISON FOR THE GREEDY ALGORITHM

In practice, the greedy algorithm often performs well even when $\kappa(\mathbf{\Lambda})$ is large. As an example, consider the signal in Figure 1. The signal is a 15-sample damped sinusoid. In the simulation, the covariance is a symmetric Toeplitz matrix generated by the sequence $12 * (.995)^i$, for $i = 0, \ldots, 14$. The condition for this covariance matrix is $\kappa(\mathbf{\Lambda}) = 5774.8$.

Despite the large condition number in this instance, the forward greedy algorithm performs well. Figure 2 shows the performance of the forward greedy algorithm, the maximum signal energy selection algorithm, and the optimal selection based on an exhaustive search. The forward greedy algorithm does almost as well as the exhaustive search, and significantly better than the maximum signal energy selection rule. Additionally, the RLMF with $K = 5$ performs almost as well as the full matched filter.

7. DYNAMIC PROGRAMMING SOLUTIONS FOR BANDED MATRICES

In some cases, we can use the special structure of $\mathbf{\Lambda}$ to derive search algorithms guaranteed to find the optimal RLMF subset. In this section, we consider instances of the RLMF

problem where $\mathbf{\Lambda}$ is a tri-band matrix. In these instances, all the non-zero entries of $\mathbf{\Lambda}$ lie along the main diagonal and two adjacent diagonals. Exploiting this structure, the best RLMF subset for any K can be determined by a dynamic programming algorithm. This section develops the ideas behind the algorithm and presents a brief example.

Regardless of the properties of $\mathbf{\Lambda}$, the vector $\mathbf{x}_g$ can be built from segments of consecutive data, which we refer to as fragments. For example, consider the selected subset

$$\mathbf{x}_g^T = \left[\begin{array}{cccc} x_1 & 0 & x_3 & x_4 \end{array} \right]. \tag{13}$$

In this example, the selected subset consists of $\{x_1, x_3, x_4\}$. For this subset, the two fragments are $\{x_1\}$ and $\{x_3, x_4\}$. For tri-band matrices, the fragment decomposition constrains the structure of $\mathbf{\Lambda}_{\tilde{\mathbf{x}}_g}$. In example (13), the selected covariance matrix is

$$\mathbf{\Lambda}_{\tilde{\mathbf{x}}_g} = \left[\begin{array}{cc} [\mathbf{\Lambda}]_{11} & \mathbf{0}^T \\ \mathbf{0} & \mathbf{\Lambda}_{\{3,4\}} \end{array} \right]. \tag{14}$$

Here, $\mathbf{\Lambda}_{\{3,4\}}$ is the covariance matrix for $\{x_3, x_4\}$ and $\mathbf{0}$ is a 2×1 vector of zeros. Note that measurements from distinct fragments are uncorrelated.

The decomposition can be extended to an arbitrary number of fragments. Any selection matrix $\mathbf{G}$ can be broken into f distinct fragments. The selection matrices identifying the fragments are expressed as $\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_f$. When $\mathbf{\Lambda}$ is tri-band, the selected covariance matrix $\mathbf{\Lambda}_{\tilde{\mathbf{x}}_g}$is block diagonal, as indicated in equation (14). Thus, the SNR can be decomposed as

$$\mathrm{SNR}(\mathbf{G}) = \sum_{i=1}^{f} \mathrm{SNR}(\mathbf{G}_i) \tag{15}$$

since the inverse of a block diagonal matrix is also block diagonal and the $\mathrm{SNR}(\mathbf{G})$ denotes a quadratic form.

A dynamic programming algorithm exploiting this structure can determine the optimal RLMF subset for any K. We give an example for the case of $K = 2$ that illustrates the general ideas in the algorithm. A brute-force search for the best $K = 2$ RLMF requires $O(N^2)$ computation. The dynamic programming algorithm reduces this complexity to $O(N \log N)$.

As a first step, consider a two-element subset with the first element fixed as x_i. If the subset elements are not consecutive, the subset SNR can be written

$$\mathrm{SNR}(\mathbf{G}) = \mathrm{SNR}(\{x_i\}) + \mathrm{SNR}(\{x_{i+j}\}), \tag{16}$$

where $j > 1$. Since the first element of the subset is x_i, optimization over all such subsets composed of two fragments is reduced to choosing the index of the second element according to

$$a(i) = \arg \max_{j \in [i+2, N]} \mathrm{SNR}(\{x_j\}). \tag{17}$$

As a second step, the best $K = 2$ subset with first element at index i can be determined by comparing $\mathrm{SNR}(\{x_i, x_{i+1}\})$ and $\mathrm{SNR}(\{x_i, x_{a(i)}\})$. Finally, the best $K = 2$ RLMF subset can be determined by calculating the best subsets with fixed first element for $i = 1$ to $i = N - 1$ and searching the list for the best subset.

Equation (17) indicates that the optimization procedure for $K = 2$ utilizes results for single element subsets. Specifically, a table of the maximum $\mathrm{SNR}(\{x_j\})$ element with index satisfying $j > i$, for $i = 1, \ldots, N - 1$ simplifies the optimization in the equation from a search through a list to a look-up. The computation necessary to construct this table is dominated by an $O(N \log N)$ sort of the single element SNRs. The computation of all the SNRs of the form $\mathrm{SNR}(\{x_i\})$ and $\mathrm{SNR}(\{x_i, x_{i+1}\})$ has $O(N)$ complexity, and calculating $a(i)$ for $i = 1, \ldots, N - 2$ from the sorted $K = 1$ results has $O(N)$ complexity, so the sort dominates the computational complexity of the algorithm for $K = 2$.

This dynamic programming procedure, where solutions for K are computed using intermediate solutions for subsets of up to $K - 1$ elements, can be expanded to computing optimal subsets for any value of K. The complete algorithm is developed and generalized to banded matrices with wider bands in [7].

8. REFERENCES

[1] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, UK, 1985.

[2] S. Kumar, F. Zhao, and D. Shepherd. Collaborative signal and information processing in microsensor networks. *IEEE Signal Processing Magazine*, vol. 19, no. 3, pp.13–14, March 2002.

[3] A. J. Miller. *Subset Selection in Regression*. Chapman and Hall, New York, NY, 1990.

[4] A. Molisch, M. Win, and J. Winters. Capacity of MIMO systems with antenna selection. In *Proc. IEEE International Conf. on Communication*, pp. 570–574, 2001.

[5] P. M. Narendra and K. Fukunaga. A branch and bound algorithm for feature subset selection. *IEEE Transactions on Computers*, vol. 26, pp.917–922, Sept. 1977.

[6] H. Vincent Poor. *An Introduction to Signal Detection and Estimation*. Springer-Verlag, New York, NY, Second edition, 1994.

[7] C. K. Sestok. *Data Selection in Binary Hypothesis Testing*. PHD dissertation, Mass. Inst. of Technology, Department of Electrical Engineering and Computer Science. Dec. 2003.