SIGNAL PROCESSING V: Theories and Applications
L. Torres, E. Masgrau, and M.A. Lagunas (eds.)
© Elsevier Science Publishers B.V., 1990

A SIGNAL SUBSPACE FRAMEWORK OF NONLINEARLY CONSTRAINED SOLUTIONS

S. Konyk, Jr.*, M. G. Amin*, M. A. Lagunas**

*Department of Electrical Engineering
Villanova University
Villanova, PA 19085, USA

**Procesado de Senal en Comunicaciones
E.T.S.I. Telecomunicacion
Apdo. 30.002 08080 Barcelona, Spain

A signal subspace based framework in presented for the nonlinearly constrained estimation problem in the contexts of both mean square error (MSE) and least squares error (LSE). The class of constraints emphasized were those with a null space spanned by a linear basis set and includes all linear constraints and nonlinear quadratic constraints. The approach taken employs a congruent transformation and maps the problem to one of distance minimization over an appropreate null space. Closed form constrained solutions are derived for the cases of linear constraints and nonlinear quadratic constraints.

1. INTRODUCTION

The problem of parameter estimation subject to constraints has been studied for numerous applications in signal ananysis and control theory [1]. Various methods have been employed to identify sytems with linear constraints [2]. Some of the identification methods are based on apriori statistical knowledge while other employ a finite sequence of data observations. This paper addresses the constrained mean squared error (MSE) and least squares error (LSE) estimation problems from the view point of signal space subspace, signal null space and constraint null space.

The null space of a constraint function is introduced as a vehicle with which to satisfy the constrained estimation problem with minimal or no increase in estimation error. The constraint function is invariant to updates in the weight vector which are restricted to the constraint function null space. Thus, the constraint function null space may be searched to obtain an optimal weight vector which satisfies the constraint with minimal penalty in estimation error.

The case where the constraint null space is spanned by a linear basis is emphasised. In this case analytical expressions are derived to calculate the elements of the optimal constrained weight vector. Included in this class of constraints are linear constraints and nonlinear quadratic constraints.

In the MSE estimation problem the error surface measure has a null space of dimension zero [3]. In this problem, the approach taken is to perform a congruent transformation on the weight vector resulting in a symmetric error surface. Under this transformation the optimal weight vector is obtained via distance minimization by searching the constraint null space for a weight vector with the shortest possible distance to the null space of the LSE surface.

The LSE estimation problem is considered in the context of the minimum norm constraint [4,5]. Here the result is a unique solution determined by the pseudo inverse of the data matrix and corresponds to selecting a particular member of the subspace of the weight vectors which intersect only the origin of the nullspace. The invariant performance imposed by the span of the null space is used as a reference in searching for and selecting an optimal LSE solution from a set of competitive suboptimal features.

Searching the nullspace is performed by first defining the LSE null space basis via singular value decomposition (SVD) of the data matrix [6]. These bases are then used to determine the additional null space weight components which when added to the minimum norm solution result in a total weight vector which has the shortest distance from the null space of the nonlinear constraints.

The paper presents a unified null space approach to solve the MSE and LSE estimation problems. Analytical solutions are presented for the case of linearly spanned null spaces of the constraint and error performance surfaces. Examples for the estimation problems are presented for the nonlinear quadratic smoothness constraint and may be extended to quadratic constraints in general.

The above cases of the least mean squared least squares estimation problems are considered in the context of a simple finite impulse response (FIR) linear predictor. The one step predictor output $\hat{u}(i)$ may be expressed as the convolution sum

$$\hat{d}(i) = \sum_{k=1}^{M} w_k u(i-k+1) \tag{1}$$

where M is the filter length, w_k are the filter weights and the $u(\cdot)$ are the tap inputs.

2. CONSTRAINT NULL SPACE INVARIANCE

In the case of real data, constraints on the weight vector, $\mathbf{w} \in \mathcal{R}^{\mathcal{M}}$, may involve equality and/or inequality relations. In the case of an equality constraint relation on the weight vector, $f(\mathbf{w}) = \mathbf{C}$, the image of the constraint function $f(\mathbf{w})$ may be vector valued with $\mathbf{C} \in \mathcal{R}^k$ for $1 \leq k \leq M$. In the case when an inequality constraint relation is imposed such as $f(\mathbf{w}) \leq C$, the constraint function is taken to be a functional with $C \in \mathcal{R}$.

In either constraint relation, equality or inequality, the null space $N(f(\mathbf{w}))$ of the constraint relation $f(\mathbf{w})$ is a subset of the space of all possible weight vectors $\mathcal{R}^{\mathcal{M}}$. The null space of the constraint function is defined as the set

$$\begin{aligned} N(f(\mathbf{w})) = \{\mathbf{w} \in \mathcal{R}^M | f(\mathbf{w}) = \mathbf{0}, \\ \mathbf{0} \in \mathcal{R}^k, 1 \leq k \leq M\} \end{aligned} \tag{2}$$

The constraint function $f(\mathbf{w})$ is null space invariant if

$$f(\mathbf{w} + \mathbf{z}) = f(\mathbf{w}) \tag{3}$$

$\forall \mathbf{w} \in \mathcal{R}^M$ and $\forall \mathbf{z} \in N(f\mathbf{w})$. Furthermore, if the null sapce of the constraint function is spanned by a linear basis, the constraint is said to be null space linearly invariant or simply linearly invariant. The class of linearly invariant filter constraints includes linear and quadratic constraint functions.

2.1 Linear Constraint Case

The null space linear invariance of a linear constraint on the weight vector is readily proven. The linear constraint function takes the form $f(\mathbf{w}) = \mathbf{A}\mathbf{w}$ where $\mathbf{A}$ is an $M \times M$ matrix. The null sapce of this linear constraint may be expressed as

$$N(\mathbf{A}) = \{\mathbf{w} \in \mathcal{R}^M | \mathbf{A}\mathbf{w} = \mathbf{0}\} \tag{4}$$

The linear invariance of the constraint follows directly from the property

$$\mathbf{A}(\mathbf{w} + \mathbf{x}) = \mathbf{A}\mathbf{w} \; \forall \mathbf{x} \in N(\mathbf{A}) \tag{5}$$

The dimensionality of the constraint null space is directly related to the rank of the constraint matrix by $dim\{N(\mathbf{A})\} = rank(\mathbf{A}) - M$ The null space of a linear transformation is spanned by finite basis of $(rank(\mathbf{A}) - M)$ vectors and so is a subspace of $\mathcal{R}^M$. The linear constraint is thus linearly invariant.

2.2 Quadratic Constraint Case

Quadratic constraints on the weight vector possess a constraint function of the form $f(\mathbf{w}) = \mathbf{w}^T\mathbf{Q}\mathbf{w}$, where $\mathbf{Q}$ in an $M \times M$ symmetric matrix and the superscript $(\;)^T$ denotes the transpose operation. The null space of the quadratic constraint is characterized by

$$\begin{aligned} N(\mathbf{Q}) &= \{\mathbf{w} \in \mathcal{R}^M | \mathbf{w}^T\mathbf{Q}\mathbf{w} = \mathbf{0}, O \in \mathcal{R}\} \\ &\text{or equivalently} \\ &= \{\mathbf{w} \in \mathcal{R}^M | \mathbf{Q}\mathbf{w} = O, O \in \mathcal{R}\} \end{aligned} \tag{6}$$

Thus the null space of the quadratic constraint function has a finite dimensional set of basis vectors. The quadratic constraint function also possesses the null space linear invariance property since

$$\begin{aligned} f(\mathbf{w} + \mathbf{x}) &= (\mathbf{w} + \mathbf{x})^T\mathbf{Q}(\mathbf{w} + \mathbf{x}) \\ &= \mathbf{w}^T\mathbf{Q}\mathbf{w} + 2\mathbf{x}^T\mathbf{Q}\mathbf{w} + \mathbf{x}^T\mathbf{Q}\mathbf{x} \\ &= \mathbf{w}^T\mathbf{Q}\mathbf{w} \\ &= f(\mathbf{w}) \end{aligned} \tag{7}$$

for all $\mathbf{w} \in \mathcal{R}^M$ and $\mathbf{x} \in N(\mathbf{Q})$.

A fundamental example of a quaratic constraint is the smoothness or flatness measure. According to this measure an equal coefficient filter corresponds to a maximally flat impulse response. This is equivalent to driving the filter towards a sinc function in the frequency domain and in this context minimizing filter bandwidth. Given a weight vector $\mathbf{w}$, a measure of its smoothness is expressed as

$$\begin{aligned} J_s(\mathbf{w}) &= \sum_{i=1}^{M} (w_i - \overline{w})^2 \\ &= \mathbf{w}^T\mathbf{Q}\mathbf{w} \end{aligned} \tag{8}$$

where $\overline{w}$ is the average value of the weights and $\mathbf{Q}$ is an $M \times M$ symmetric matrix with entries

$$q_{i,j} = \begin{cases} \frac{M-1}{M} & \text{for } i = j \\ \frac{-1}{M} & \text{for } i \neq j \end{cases} \tag{9}$$

The null space of $\mathbf{Q}$ is one dimensional and spanned by the $M \times 1$ unit vector

$$\hat{\mathbf{1}} = \begin{bmatrix} \frac{1}{\sqrt{M}} \\ \vdots \\ \frac{1}{\sqrt{M}} \end{bmatrix} \tag{10}$$

A maximally flat impulse response corresponds to zero variance in the weight vector which implies perfect smoothness and thus

$$\mathbf{w} = c\hat{\mathbf{I}} \quad \text{for some } c \in \mathcal{R} \tag{11}$$

3. LEAST MEAN SQUARED PERFORMANCE

Given a desired response $d(n)$, the least MSE estimation problem is one of determining the weight vector $\mathbf{w}$ which minimizes the mean squared value of the output error. This error, assuming real data, may be expressed as

$$\begin{aligned} J_{LMS} &= \mathcal{E}[(\hat{d}(n) - d(n))^2] \\ &= \sigma_d^2 - 2\mathbf{w}^T\mathbf{p} + 2\mathbf{w}^T\mathbf{R} \end{aligned} \tag{12}$$

where $\mathcal{E}$ denotes the expectation operator. The vector $\mathbf{p}$ is the cross-correlation vector between the input vector and the desired response, and $\mathbf{R}$ is the autocorrelation matrix of the input vector $\mathbf{u}(n)$.

In the case where no constraints are imposed upon the weight vector in the process of minimizing the MSE, the optimal weight vector in given by

$$\mathbf{w}_O = \mathbf{R}^{-1}\mathbf{p} \tag{13}$$

and the MSE has the form

$$J_{LMS} = J_{MIN} + (\mathbf{w} + \mathbf{w}_O)^T\mathbf{R}(\mathbf{w} - \mathbf{w} - \mathbf{w}_O) \tag{14}$$

where $J_{MIN} = J_{LMS}(\mathbf{w}_O)$.

The weight vector $\mathbf{w}_O$ is unique, the autocorrelation matrix $\mathbf{R}$ is of full rank and positive definite. Thus the constrained optimization problem may be divided into two cases.

Case I. The null space of the constraint has dimensionality zero. In this case, adherence to the constraint results in a direct sacrifice in MSE.

Case II. The null space of the constraint has dimensionality greater than zero. Thus, the null space of the constraint may be searched to minimize MSE while simultaneously satisfying the constraint. The approach taken is to perform a congruency transformation of the weight vector resulting in a symmetric MSE surface and then searching the constraint null space for the weight vector with the shortest distance away from the error surface minimum.

An example of *Case II.* is the smoothness constraint. The one dimensional null space of this constraint has $\hat{\mathbf{I}}$ as a basis. The congruency transformation resulting is a symmetric error surface if specified first. Let $\mathbf{V}$ be a unitary matrix which diagonalizes the autocorrelation matrix $\mathbf{R}$. Then,

$$\mathbf{R} = \mathbf{T}^{-1}\Lambda\mathbf{T} \tag{15}$$

where Λ is a diagonal matrix whose diagonal entries are the eigen values of $\mathbf{R}$. The desired congruency transformation may be expressed as

$$\mathbf{w} = \mathbf{T}_C\tilde{\mathbf{w}} \text{ where } \sqrt{\Lambda^{-1}}\mathbf{T} \tag{16}$$

which maps the weight vector $\mathbf{w}$ to $\tilde{\mathbf{w}}$.

The problem is now transformed to a distance minimization problem. The transformed constraint null space is searched to find the weight vector colsest to the image of $\mathbf{w}_O$. It can be readily shown that the null space component of the is characterized by $c_{opt}\hat{\mathbf{I}}$ where

$$c_{opt} = \frac{\hat{\mathbf{I}}^T\mathbf{P}}{\hat{\mathbf{I}}^T\mathbf{R}\hat{\mathbf{I}}} \tag{17}$$

4. LEAST SQUARES PERFORMANCE

The least squares estimation problem utilizes the sum of error squares performance measure. Employing the covariance method for windowing the input data $u(i), i = 1, \ldots, N$; the sum of error squares measure is given by

$$J_{\mathcal{E}}(w_1, \ldots, w_M) = \sum_{i=M}^{N} |e(i)|^2 \tag{18}$$

where

$$e(i) = u(i) - \hat{u}(i) \tag{19}$$

Equation (2) may be expressed as

$$\begin{aligned} J_{\mathcal{E}}(\mathbf{w}) &= \mathbf{r}^T\mathbf{r} \\ &= (\mathbf{b} - \mathbf{A}\mathbf{w})^T(\mathbf{b} - \mathbf{A}\mathbf{w}) \end{aligned} \tag{20}$$

where

$$\mathbf{w} = [w_1, w_2, \ldots, w_M]^T, \tag{21}$$

$$\mathbf{r} = [e(M), e(M+1), \ldots, e(N)]^T, \tag{22}$$

$$\mathbf{b} = [u(M+1), u(M+2), \ldots, u(N+1)]^T, \tag{23}$$

and

$$\mathbf{A} = \begin{bmatrix} u(M) & u(M-1) & \ldots & u(1) \\ u(M+1) & u(M) & \ldots & u(2) \\ \vdots & \vdots & \vdots & \vdots \\ u(N) & u(N-1) & \ldots & u(N-M+1) \end{bmatrix} \tag{24}$$

The superscript ^{T} denotes transposition, $\mathbf{w}$ is the $M \times 1$ tap weight vector, $\mathbf{u}(i)$ is the $M \times 1$ input vector, $\mathbf{r}$ is the $M \times 1$ residual vector, and $\mathbf{A}$ is the $(M - N + 1) \times N$ data matrix.

The least squares solution of the filter weights, which minimizes $J_{\mathcal{E}}$ over the data window, satisfies the normal equation

$$\mathbf{A}^T\mathbf{A}\hat{\mathbf{w}} = \mathbf{A}^T\mathbf{b} \tag{25}$$

If $rank(\mathbf{A}) = M$, $\mathbf{AA}^T$ is nonsingular and the tap weights ares uniquely determined as $\hat{\mathbf{w}} = (\mathbf{AA}^T)^{-1}\mathbf{A}^T\mathbf{b}$. On the other hand, for $rank(\mathbf{A}) = r < M$ the nullity of $\mathbf{A}$ is nonzero and the least squares solution ,$\hat{\mathbf{w}}$, is no longer unique. In this case, the pseudo inverse of the data matrix, denoted $\mathbf{A}^{\dagger}$, characterizes the minimum norm solution that is given by

$$\begin{aligned}\hat{\mathbf{w}} &= \mathbf{A}^{\dagger}\mathbf{b} \\ &= \mathbf{X}\Sigma^{-1}\mathbf{Y}^T\end{aligned} \tag{26}$$

The orthogonal transformation matrices $\mathbf{Y}$ and $\mathbf{X}$ of the singular value decomposition (SVD) of $\mathbf{A}$ characterize the null space, $N(\mathbf{A})$, and the range space, $R(\mathbf{A})$, of the data matrix, respectively. The first r diagonal entries of Σ satisfy $\sigma_1 \geq \sigma_2 \geq, \ldots, \sigma_M \geq 0$ while the rest of the matrix elements have zero values. The last $(M-r)+1$ columns of the matrix $\mathbf{X}$ form an orthonormal basis for $N(A)$.

In the constrained least squares problem four cases result based upon preformance and constraint nullity.

Case I. The data matrix $\mathbf{A}$ is full rank and the constraint null spcae has dimension zero. In this case a direct sacrifice of minimum LSE results in satisfying the constraint.

Case II. The data matrix $\mathbf{A}$ is full rank and the constraint null space dimension is greater than zero. Here the null space of the constraint may be searched so as to satisfy the constraint with mimimum increase in LSE which is analagous to the LSE problem.

Case III. The data matrix $\mathbf{A}$ is rank deficient and the constraint null space is dimension zero. If the constraint is relaxed, the null space of the data matrix may be searched to find the weight vector which best satisfies the constraint at minimum LSE again analagously to the MSE problem.

Case IV. The data matrix $\mathbf{A}$ is rank deficient and the dimension of the constraint null space is greater than zero. Now both the performance measure and the constraint null spaces may be search simultaneously to satisfy the constraint with minimum LSE penalty or to maintain the least possible MSE while satisfying the constraint as well as possible.

Consider *Case IV.* in which smoothness is to be maximized without sacrificing minimum LSE. The set of column vectors $\mathbf{x}_{r+1}, \ldots, \mathbf{x}_M$ of $\mathbf{X}$ form an orthonormal basis for $N(\mathbf{A})$. Moving a weight vector through $N(\mathbf{A})$ does not change its LSE error, i.e. $J_{\mathcal{E}}(\mathbf{w}) = J_{\mathcal{E}}(\mathbf{w} + \Delta\mathbf{w})$, $\forall \Delta\mathbf{w} \in N(\mathbf{A})$. Thus the least squares problem in terms of the bases for $N(\mathbf{A})$ and $N(\mathbf{Q})$ is: given the minimum norm least squares weight vector $\hat{\mathbf{w}}$, find the scalars $c, \alpha_{r+1}, \ldots, \alpha_M$ which minimize

$$J_{\mathcal{NE}}(\mathbf{w}) = \left\| c\hat{\mathbf{1}} - \hat{\mathbf{w}}_{TOT} \right\|^2$$

$$\hat{\mathbf{w}}_{TOT} = \sum_{i=r+1}^{M} \alpha_i \mathbf{x}_i \tag{27}$$

It can be shown that the scalars in the relation above which maximize smoothness without increasing the least squares error are

$$\hat{c} = \frac{\hat{\mathbf{1}}^T\hat{\mathbf{w}}}{1 - \sum_{i=r+1}^{M} \hat{\mathbf{1}}^T\mathbf{x}_i} \tag{16}$$

and

$$\hat{\alpha}_i = \hat{c}\hat{\mathbf{1}}\mathbf{x}_i \tag{17}$$

REFERENCES

[1] G. C. Goodwin and K. S. Sin, Adaptive Filtering, Prediction and Control, Prentice-Hall, Englewood Cliffs, NJ, 1984.

[2] T. Mu and L. Griffiths, "A solution space approach to achieving partially adaptive arrays," IEEE International Conference on Acoustics, Speech and Signal Process., New York, NY, April, 1988.

[3] S. Konyk and M. Amin,"A new class of nonlinearly constrained linear estimators," IEEE International Conference on Acoustics, Speech and Signal Process., New York, NY, April, 1988.

[4] S. Konyk, M. Amin and M. Lagunas,"Least squares null space variational characterization for nonminimum norm solutions," IEEE International Conference on Acoustics, Speech and Signal Process., Glasgow, Scotland, April, 1989.

[5] S. Haykin, Adaptive Signal Processing, Prentice-Hall, Englewood Cliffs, NJ, 1986.

[6] G. Golub and C. Van Loan, Matrix Computations, Johns Hopkins University Press, 1983.

In summary, the paper presented a null space approach to the least squares and mean square estimation problems. Optimal solutions are derived by searching the null spaces associacted with performance error surfaces and constraint functions for weight vectors with minimal distances to these spaces. In the MSE estimation problem only the null space of the constraint was searched, while in the LSE estimation problem both the signal null space and the constraint null space were searched simultaneously to obtain the optimal weight vector. Analytical results were derived for the case where each of the null spaces was spanned by a linear basis which includes the class of all nonlinear quadratic constraints.