King Fahd University of Petroleum & Minerals

DEPARTMENT OF MATHEMATICAL SCIENCES

Technical Report Series

TR 266

May 2001

Structured Methods for Solving Hankel Matrix Approximation Problems

Suliman Al-Homidan

DHAHRAN 31261 ● SAUDI ARABIA ● www.kfupm.edu.sa/math/ ● E-mail: mathdept@kfupm.edu.sa

Structured Methods for Solving Hankel Matrix Approximation Problems

Suliman Al-Homidan
Department of Mathematical Sciences, King Fahd University of Petroleum and Minerals, Dhahran 31261, PO Box 119, Saudi Arabia E-mail: homidan@kfupm.edu.sa

Abstract

The problem of finding the nearest positive semi-definite Hankel matrix of a given rank to an arbitrary matrix is considered. The problem is formulated as a nonlinear minimization problem with positive semi-definite Hankel matrix as constraints. Then an algorithm with rapid convergence is obtained by the l_1Sequential Quadratic Programming (SQP) method. A second approch is to formulate the problem as a smooth unconstrained minimization problem, for which rapid convergence can be obtained by, for example, the BFGS method. This paper studies both methods. Comparative numerical results are reported.

Key words : Non-smooth optimization, positive semi-definite matrix, Hankel matrix, SQP Method, BFGS Method.
AMS (MOS) subject classifications; 65F99, 99C25, 65F30.

1 Introduction

Hankel matrices appear naturally in a variety of problems of engineering interest: comunication, control engineering, filter design, identification, model reduction and broadband matching and in different fileds of mathematics, e.g., in systems theory, integral equations and operator theory.

Hankel matrices possess certain properties regarding their rank and positive semi-definite structures depending on the construction or arrangement of their elements. In practical applications, these matrices are constructed from noisy observations and hence some of their nice properties may be destroyed or changed. The signal processing problem is to estimate the matrices with desired properties so that the estimated matrix is close to the given observation in some reasonable sense.

We consider the following problem: Given an arbitrary data matrix $F \in \mathbb{R}^{n \times n}$, find the nearest positive semi-definite Hankel matrix H of rank m to F. Use of the Frobenius norm as a measure gives rise to

$$\begin{aligned} &\text{minimize} \quad \phi = \|F - H\| \\ &\text{subject to} \quad H \in K, \end{aligned} \tag{1.1}$$

where K is the set of all $n \times n$ symmetric positive semi-definite Hankel matrices

$$K = \{H : H \in \mathbb{R}^{n \times n},\ H \geq 0,\ \text{Rank}(H) = m \quad \text{and} \quad H \in \mathcal{H}\}, \tag{1.2}$$

where $\mathcal{H}$ is the set of all Hankel matrices.

The problem was studied by MacInnes [9]; he proposed a method for finding the best approximation of a matrix A by a full rank Hankel matrix. In [9], the initial problem of best approximation of one matrix by another is transformed into a problem involving best approximation of a given vector by a second vector whose elements are constrained so that its inverse image is a Hankel matrix. Related problems were also studied by [10, 11] and [12] in relation to signal processing problems.

A Hankel matrix H is denoted by

$$H = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ a_2 & a_3 & \cdots & a_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ a_n & a_{n+1} & \cdots & a_{2n-1} \end{bmatrix} = \text{Hankel}(a_1, a_2, a_3, \ldots, a_{2n-1}). \tag{1.3}$$

Section 2 contains a brief description of the SQP method for solving (1.1). The problem is formulated as a nonlinear minimization problem and then solved using techniques related to filterSQP . In Section 3 the problem is formulated as a smooth unconstrained minimization problem then solved using BFGS method. Finally, in Section 4 numerical comparisons of these methods are carried out.

2 The SQP Methods

In this section an iterative scheme is investigated in order to develop an algorithm for solving problem (1.1). The problem is formulated as a nonlinear minimization problem and then solved using techniques related to filterSQP [7] which provide global convergence at a second order rate.

It is difficult to deal with the matrix set constraint in (1.2) since it is not easy to specify if the elements are feasible. Using partial LDL^T factorization of H, this difficulty can be overcome. Since m, the rank of H^*, is known, therefore for H sufficiently close to H^*, the partial factors $H = LDL^T$ can be calculated such that

$$L = \begin{bmatrix} L_{11} & \\ L_{21} & I \end{bmatrix}, D = \begin{bmatrix} D_1 & \\ & D_2 \end{bmatrix}, H = \begin{bmatrix} H_{11} & H_{21}^T \\ H_{21} & H_{22} \end{bmatrix}, \tag{2.1}$$

where L_{11}, D_1 and H_{11} are $m \times m$ matrices; I, D_2 and H_{22} are $n - m \times n - m$ matrices; L_{21} and H_{21} are $n - m \times m$ matrices; D_1 is diagonal and $D_1 > 0$ and D_2 have no particular structure other than symmetry. At the solution, $D_2 = 0$ and H is symmetric positive semi-definite Hankel matrix. In general,

$$D_2(H) = H_{22} - H_{21} H_{11}^{-1} H_{21}^T. \tag{2.2}$$

Now if the structure of the matrix H is in a Hankel form, i.e.,

$$H = \begin{bmatrix} x_1 & \cdots & x_n \\ \vdots & \ddots & \vdots \\ x_n & \cdots & x_{2n-1} \end{bmatrix} = \text{Hankel}(x_1, \cdots, x_{2n-1}), \tag{2.3}$$

then (2.2) enables the constraint $H \in K$ to be written in the form

$$D_2(H(\mathbf{x})) = 0. \tag{2.4}$$

Hence, (1.1) can now be expressed as

$$\begin{aligned} &\text{minimize} \quad \phi \\ &\text{subject to} \quad D_2(H(\mathbf{x})) = 0 = Z^T H Z, \end{aligned} \tag{2.5}$$

where $Z = \begin{bmatrix} -H_{11}^{-1} H_{21}^T \\ I \end{bmatrix}$ is the basis matrix for the null space of H when $D_2 = 0$. The Lagrange multipliers for the constraint (2.4) are Λ relative to the basis Z and the Lagrangian for porblem (2.5) is

$$\mathcal{L}(\mathbf{x}^{(k)}, \Lambda^{(k)}) = \phi - \Lambda : Z^T H Z. \tag{2.6}$$

The above approach has been studied in a similar way by [5].

The structure of the Hankel matrix D has been given in (2.3), then

$$\phi = \sum_{i,j=1}^{n} (f_{ij} - h_{ij})^2 = \sum_{i,j=1}^{n} (f_{ij} - x_{i+j-1})^2, \tag{2.7}$$

and $\nabla\phi = [\frac{\partial\phi}{\partial x_1} \cdots \frac{\partial\phi}{\partial x_{2n-1}}]^T$ where ∇ denotes the gradient operator $(\partial/\partial x_1, \ \ldots, \ \partial/\partial x_{2n-1})^T$. Therefore

$$\begin{aligned} \frac{\partial\phi}{\partial x_s} &= 2\sum_{i=1}^{s}(x_s - f_{i\ s-i+1}) \qquad s = 1, \ldots, n \\ \frac{\partial\phi}{\partial x_s} &= 2\sum_{i=1}^{2n-s}(x_s - f_{n-i+1\ s+i-n}) \quad s = n+1, \ldots, 2n-1. \end{aligned} \tag{2.8}$$

Differentiating again gives

$$\frac{\partial^2\phi}{\partial x_r \partial x_s} = 0 \qquad \textit{if} \quad r \neq s,$$

where $s, r = 1, \cdots, 2n-1$, and

$$\begin{aligned} \frac{\partial^2\phi}{\partial x_s^2} &= 2s \qquad\qquad s = 1, \ldots, n \\ \frac{\partial^2\phi}{\partial x_s^2} &= 2(2n-s). \quad s = n+1, \ldots, 2n-1. \end{aligned} \tag{2.9}$$

The advantage of formula (2.4) is that expressions for both first and second derivatives of the constraints with respect to the elements of H can be obtained. The simple form of (2.2) is utilized by writing the constraints $D_2(H) = 0$ in the following form:

$$d_{ij}(\mathbf{x}) \;=\; x_{i+j-1} - \sum_{k,l=1}^{m} x_{i+k-1}[H_{11}^{-1}]_{kl}\; x_{j+l-1} = 0 \tag{2.10}$$

where $i, j = m+1, \cdots, n$ and $[H_{11}^{-1}]_{kl}$ denotes the element of H_{11}^{-1} in kl–position. Thus (2.5) can be expressed as

$$\begin{aligned} \text{minimize} \quad \phi &= \sum_{i,j=1}^{n} (f_{ij} - x_{i+j-1})^2. \\ \text{subject to } d_{ij}(\mathbf{x}) &= 0 \end{aligned} \tag{2.11}$$

In this problem, since the equivalent constraints $d_{ij}(\mathbf{x}) = 0$ and $d_{ji}(\mathbf{x}) = 0$ are both present, they would be stated only for $i \geq j$.

In order to write down the SQP method applied to (2.11), it is necessay to derive first and second derivatives of d_{ij} which enable a second order rate of convergence to be achieved.

Let I_s be an $m \times m$ matrix given by

$$I_s \;=\; \text{Hankel}(0, \ldots, 0, 1, 0, \ldots, 0),$$

where the "1" appearing in the first row is in the sth column and the "1" appearing in the first column is in the sth row. Hence the matrix I_s is a matrix that contains "1"s in one across anti–diagonal and zeros elsewhere. Now differentiating $H_{11}H_{11}^{-1} = I$ gives

$$\begin{aligned} \frac{\partial H_{11}^{-1}}{\partial x_s} &= -\; H_{11}^{-1}\, I_s\, H_{11}^{-1} && s < 2m. \\ \frac{\partial H_{11}^{-1}}{\partial x_s} &= \mathbf{0} && s \geq 2m \end{aligned} \tag{2.12}$$

Hence from (2.2),

$$\frac{\partial D_2}{\partial x_s} \;=\; II_s + V^T I_s V + U^T + U, \tag{2.13}$$

where

$$V^T = -H_{21}H_{11}^{-1}, \quad U = III_s V, \quad II_s = \frac{\partial H_{22}}{\partial x_s} \text{ and } III_s = \frac{\partial H_{21}}{\partial x_s},$$

II_s and III_s are matrices similar to I_s with II_s being an $n-m \times n-m$ matrix which contains ones in one across anti–diagonal and zeros elsewhere, and III_s is an $n-m \times m$ matrix which contains ones in one across anti–diagonal and zeros elsewhere.

Furthermore, differentiating (2.12), we get

$$\frac{\partial^2 D_2}{\partial x_s \partial x_r} = Y + Y^T,$$

where

$$Y = -Z_r^T H_{11}^{-1} Z_s \quad \textit{and} \quad Z_t = I_t V - III_t^T.$$

Table 1 summarizes the state of the gradient and Hessian of D_2 with respect to x_s Now, let

$\frac{\partial D_2}{\partial x_s}$	Z_t	s
$V^T I_s V$	$I_t V$	$0 < s \le m$
$V^T I_s V + U^T + U$	$I_t V - III_t^T$	$m < s < 2m$
$U^T + U$	$-III_t^T$	$s = 2m$
$II_s + U^T + U$	$-III_t^T$	$2m < s < n + m$
II_s	$\mathbf{0}$	$n + m < s < 2n - 1$

Table 1: Gradient and Hessian formulas for D_2.

$$\begin{aligned} W &= \nabla^2 \mathcal{L}(\mathbf{x}, \Lambda) \\ &= \nabla^2 \phi - \sum_{i,j=m+1}^{n} \lambda_{ij} \nabla^2 d_{ij} \end{aligned} \tag{2.14}$$

where $\nabla^2 \phi$ is given by (2.9) and

$$\sum_{i,j=m+1}^{n} \lambda_{ij} \nabla^2 d_{ij} = \begin{bmatrix} \sum_{i,j} \lambda_{ij} \frac{\partial^2 d_{ij}}{\partial x_1 \partial x_1} & \cdots & \sum_{i,j} \lambda_{ij} \frac{\partial^2 d_{ij}}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \sum_{i,j} \lambda_{ij} \frac{\partial^2 d_{ij}}{\partial x_n \partial x_1} & \cdots & \sum_{i,j} \lambda_{ij} \frac{\partial^2 d_{ij}}{\partial x_n \partial x_n} \end{bmatrix}.$$

Usually, $\nabla^2 \mathcal{L}$ is positive definite, in which case, if $\mathbf{x}^{(k)}$ is sufficiently close to $\mathbf{x}^*$, the basic SQP method converges and the rate is second order (Fletcher [6]). Globally, however (2.11) may not converge. An algorithm with better convergence properties, when $\mathbf{x}^{(k)}$ is remote from $\mathbf{x}^*$, is suggested by Fletcher at. al. [7] in which the filterSQP can be used to slove (2.11). Now since the gradient and Hessian are both available, therefore filterSQP can be used to slove the problem.

This description of iterative schemes for solving (2.11) has so far ignored an important constraint, that, is $D_1 > 0$ in which the varibles $\mathbf{x}^{(k)}$ must permit the matrix $A^{(k)}$ to be factorized as in (2.1). However, since m is identified correctly and $\mathbf{x}^{(k)}$ is near the solution, this restriction will usually be inactive at the solution. If $\mathbf{x}^{(k)}$ is remote from the solution, additional constraints

$$d_{rr}^{(k)} > 0. \qquad r = 1, 2, \ldots, m$$

are introduced. However, strict inequalities are not permissible in an optimization problem and it is also advisable not to allow $d_{rr}(\mathbf{x}^{(k)})$ to come too close to zero, especially for small r, as this is likely to cause the factorization to fail. Hence the constraints

$$md_{rr}^{(k)}/r \geq 0 \qquad r = 1, 2, \ldots, m$$

are added to problem (2.11). Finally, it is possible that partial factors of the matrix $A^{(k)}$ in the form (2.1) do not exist for some iterates. In this case, the parameter in the filterSQP method $\rho^{(k+1)} = \rho^{(k)}/4$ is chosen for the next iteration in the trust region method.

3 Solution by Unconstrained Minimization

In this section, we consider a different approach to problem (1.1). The main idea is to replace (1.1) by a smooth unconstrained optimization problem in order to use superlinearly convergent quasi-Newton methods. Partial connection between the problem and signal processing is given in the following factorization.

Classical results about Hankel matrices that go back to [3] may be stated according to which a nonsingular positive semi-definite real Hankel matrix can be represented as the product of a Vandermonde matrix and its transpose and a diagonal matrix in between

$$H = VDV^T, \tag{3.1}$$

where D is an $m \times m$ diagonal matrix with positive diagonal entries and V is an $n \times m$ real Vandermonde matrix

$$V = [x_j^i], \quad i = 0, \ldots, n-1, \quad j = 1, \ldots, m \tag{3.2}$$

(see [1, 8]).

Since m, the rank of the matrix H^*, is known, it is possible to express (1.1) as a smooth unconstrained optimization problem in the following way. Since the unknown in (1.1) is the matrix H therefore the unknowns are chosen to be the elements of the matrices V; $x_1, \ldots, x_m$ and D; $d_{11}, \ldots, d_{mm}$ introduced in (3.1). This gives us an equivalent unconstrained optimization problem to (1.1) in $2m$ unknowns expressed as

$$\text{minimize}\, \phi(V, D) \;=\; \|F - VDV^T\|_F^2. \tag{3.3}$$

Then the objective function $\phi(V, D)$ is readily calculated by first forming H from V and D as indicated by (3.1) and (3.2), after which ϕ is given by $\phi(V, D) = \|F - H\|_F^2 = \|F - VDV^T\|_F^2$. The elements of the matrix H take the form

$$h_{ij} \;=\; \sum_{k=1}^{m} d_{kk} x_k^{i+j-2}. \tag{3.4}$$

Hence

$$\begin{aligned}\phi(V, D) &= \sum_{i,j=1}^{n} (h_{ij} - f_{ij})^2 \\ &= \sum_{i,j=1}^{n} (\{\sum_{k=1}^{m} d_{kk} x_k^{i+j-2}\} - f_{ij})^2. \end{aligned} \quad (3.5)$$

Our chosen method to minimize $\phi(X)$ is the BFGS quasi-Newton method (see, for example, [6]). This requires expressions for the first partial derivatives of ϕ, which are given from (3.5) by

$$\frac{\partial \phi}{\partial d_{ss}} = \sum_{i,j=1}^{n} 2(\{\sum_{k=1}^{m} d_{kk} x_k^{i+j-2}\} - f_{ij})(x_s^{i+j-2} \quad (3.6)$$

$$\frac{\partial \phi}{\partial x_s} = \sum_{\substack{i,j=1 \\ i=j\neq 1}}^{n} 2(\{\sum_{k=1}^{m} d_{kk} x_k^{i+j-2}\} - f_{ij})(i+j-2) d_{ss} x_s^{i+j-3} \quad (3.7)$$

The BFGS method also requires the Hessian approximation to be initialized. Where necessary, we do this using a unit matrix.

Some care has to be taken when choosing the initial value of the matrices V and D, in particular the rank m. If not, the minimization method may not be able to increase m. An extreme case occurs when the initial matrix $V = 0$ and $D = 0$ is chosen, and $F \neq 0$. It can be seen from (3.6) and (3.7) that the components of the gradient vector are all zero, so that $V = 0$ and $D = 0$ is a stationary point, but not a minimizer. A gradient method will usually terminate in this situation, and so fail to find the solution.

4 Numerical Results

In this section, we report our numerical results. Fortran codes have been written to program solver for (1.1) to both filterSQP and BFGS methods and executed on SUN workstation.

The results were obtained by applying the methods of Sections 2 and 3 as follows. A matrix H was formed from (3.1) by randomly choosing m weights d_i in matrix D, $0 \le d_j \le 1.0$ and m values $x_j, 0 \le x_j \le 1.0$ to determine the Vandermonde matrix V. The matrix thus obtained by (3.1) was perturbed by adding random noise matrix S to H, where elements of S vary between -0.10 and 0.10. The problem is to recover the m frequencies x_j and weights d_j that determine the matrix before the noise was added. The convergence criterion is that the maximum changes of the matrix $H^{(k)}$ should be less than 1×10^{-5}. Typically, n was chosen to be 20, 10, 4 with $m = 10,\ 4,\ 2$, respectively.

Table 2 illustrates an example of the approximation described in Sections 2 and 3. The first two columns give the weights d_j and frequencies x_j used to generate the

d_j	x_j
0.5916	0.7590
0.6690	0.4677
0.1158	0.2630
0.5040	0.1299
0.5890	0.7915
0.3539	0.5301
0.1753	0.6123
0.0388	0.7089
0.0647	0.5516
0.0822	0.7284

m	nq	ls	ϕ	d_j^*	x_j^*
10	113	10	0.32737	0.5823 0.7078 0.1126 0.5037 0.5823 0.3518 0.2236 0.0377 0.0419 0.0377	0.7771 0.4824 0.2148 0.1595 0.7771 0.5333 0.6231 0.7414 0.5118 0.7414
9	87	8	0.32731	0.6514 0.6848 0.0912 0.4733 0.6193 0.3613 0.1581 0.0797 0.0622	0.7402 0.4836 0.2772 0.1409 0.7940 0.4529 0.6479 0.7349 0.5377
8	72	27	0.32729	0.6878 0.7484 0.0563 0.4243 0.5840 0.4262 0.1386 0.1157	0.7274 0.4478 0.3155 0.1194 0.7979 0.4715 0.7275 0.7214
7	96	39	0.32729	0.7069 0.7380 0.0755 0.4493 0.6266 0.4144 0.1707	0.7265 0.4652 0.3005 0.1296 0.7956 0.4699 0.7276
6	116	21	0.32730	0.8067 0.7309 0.0819 0.4611 0.6954 0.4053	0.7223 0.4603 0.2936 0.1356 0.7931 0.4885
5	89	25	0.32730	0.7948 1.0051 0.1707 0.4875 0.7232	0.7191 0.4619 0.4655 0.1373 0.7921
4	120	30	0.32738	1.2740 0.9119 0.3758 0.6196	0.7743 0.5323 0.4785 0.1668
3	80	31	0.32738	1.2796 1.2719 0.6299	0.7741 0.5177 0.1685
2	79	12	0.33105	1.5741 1.6007	0.7593 0.3602
1	54	6	0.75111	2.8019	0.6735

Table 2: Comparing the methods with $n = 20$ and $m = 10$.

d_j	x_j	m	nq	ls	ϕ	d_j^*	x_j^*
0.5326	0.8249	5	70	55	0.1649725	0.5793 0.4719 0.6305 0.2394 0.0429	0.8233 0.2515 0.5186 0.5163 0.2326
0.7690	0.3051	4	65	41	0.1649723	0.5789 0.5183 0.6029 0.2639	0.8233 0.2509 0.5183 0.5194
0.4558	0.5136	3	77	28	0.1649723	0.5788 0.5192 0.8660	0.8234 0.2512 0.5188
0.2040	0.7090	2	63	12	0.166825	0.6925 1.2675	0.8095 0.3981
		1	89	8	0.573705	1.6696	0.6814

Table 3: Comparing the methods with $n = 10$ and $m = 4$.

matrix H before the noise is added using (3.1). The matrix is 20×20 and of rank 10 before the perturbation. In the last six columns, the approximations are obtained, decreasing the rank of the approximation by 1 at each step. m is the rank of the approximation, nq is the number of quadratic programming problem solved by filter-SQP method to get convergence, ls is the number of line searches in the BFGS method to get convergence, ϕ give the norm of $F - H$ where H is the approximated matrix, d_j^* and x_j^* are the weights and frequencies in the approximating matrix. Note that the norm of $F - H$ decreases as the rank of the approximation decreases until rank seven (optimal rank) and then increases as the rank of the approximation decreases until rank one. It is clear that the rank changes from ten to seven and ϕ remains nonzero; this is because of the remaining noise.

d_j	x_j	m	nq	ls	ϕ	d_j^*	x_j^*
0.1763	0.9218	3	51	12	0.058136	0.1280 0.2530 0.1924	0.9386 0.6789 0.8297
0.4057	0.7382	2	47	6	0.059907	0.1730 0.3999	0.9382 0.7242
		1	72	8	0.066124	0.5584	0.8124

Table 4: Comparing the methods with $n = 4$ and $m = 2$.

In Table 2, we show the results with 10×10 matrix and of rank 4 before the perturbation. Comparing ϕ in all three tables, we find them proportional with the size of the matrix. For ϕ since the process of the methods is to obtain the nearest positive semi-definite Hankel matrix tends to minimize the effect of the noise. It is to be expected that the noise would be more significant in smaller matrices. The computations have shown that for matrices as large as 50×50, the results are quite good compared with 10×10. The results are not as good in the 4×4 case; see Tables 2, 3 and 4. It seems that the noises are quite big to some degree for the smaller matrices which makes ϕ almost equal in all cases in the four tables. Also, since ϕ is very small, this means that the approximated matrix is very close to the original one H.

5 Conclusions

In this paper, we have studied the Hankel matrix approximation problem involving the positive semi-definite matrix constraint, using both filterSQP and BFGS methods. Numerical comparisons are also given. The problem needs more study in terms of the hybrid methods involving both current method and projection method [2]. Also some numerical experiment comparisons with hybrid and projection methods need to be carried out.

Acknowledgements The author is grateful to King Fahd University of Petroleum and Minerals for providing research facilities.

References

[1] Akhiezer N. I, and Krein M. G., *Some Questions in the Theory of Moments,* Transl. Math. Monographs 2, Amer. Math. Soc., Providence, 1962.

[2] Al-Homidan, S. Alternating Projection Algorithm for Hankel Matrix Approximation, submitted.

[3] Fischer E., Uber das Caratheodorysche Problem, Potenzreihen mit positivem reellen Teil betreffend,*Rend Palermo 32* 1911.

[4] Fletcher R., Numerical experiments with an exact l_1 penalty function method, *Nonlinear Programming 4,* (Eds. O. L. Mangasarian, R. R. Meyer and S. M. Robinson), Academic Press, New York, 1981.

[5] Fletcher R., Semi–definite matrix constraints in optimization, *SIAM J. Control and Optimization, V. 23, pp. 493–513*, 1985.

[6] Fletcher R., *Practical methods of Optimization,* John Wiley and Sons, Chichester, 1987.

[7] Fletcher R. and Leyffer S., User manual for filterSQP, *Univerity of Dundee Numerical Analysis Report NA/181*, 1999.

[8] Iohvidov I. S., R., *Toeplitz and Hankel Matrices and Forms: Algebraic Theory,* transl. by G. Phillp A. Thijse, Birkhauser, Boston, 1987.

[9] MacInnes C. S., The Solution to a Structured Matrix Approximation Problem Using Grassman Coordinates *SIAM Journal on Matrix Analysis and Applications V. 21, pp. 446-453* 1999.

[10] Ponstein J., A matrix minimization problem involving ranks *Linear Alg. and Appl., V. 92, pp. 81–106,* 1987.

[11] Shaw A. K., Pokala S., and Kumaresan R., Toeplitz and Hankel matrix approximation using structured approach, *Proc. IEEE V. 86, pp.2349–2352,* 1998.

[12] Shaw A. K. and Kumaresan R., Some structured matrix approximation problems, in*Proc. IEEE ICASSP. V. 4, Internat Conf. Acoustics, Speech, and Signal Proccessing. Pp. 2324–2327.* 1988.