Katholieke Universiteit Leuven

Departement Elektrotechniek ESAT-SISTA/TR 03-105

Block-row Hankel Weighted Low Rank Approximation[1]

Mieke Schuermans, Philippe Lemmerling and Sabine Van Huffel[2]

July 2003
Submitted for publication to IEEE Transactions on Signal Processing

[1]This report is available by anonymous ftp from *ftp.esat.kuleuven.ac.be* in the directory *pub/sista/mschuerm/reports/03-105.ps.gz*

[2]K.U.Leuven, Dept. of Electrical Engineering (ESAT), Research group SISTA, Kasteelpark Arenberg 10, 3001 Leuven-Heverlee, Belgium, Tel. 32/16/32 17 10, Fax 32/16/32 19 70, WWW: *http://www.esat.kuleuven.ac.be/sista*. E-mail: *mieke.schuermans@esat.kuleuven.ac.be*. Dr. Sabine Van Huffel is a full professor, dr. Philippe Lemmerling is a postdoctoral researcher of the FWO (Fund for Scientific Research – Flanders) and Mieke Schuermans is a research assistant at the Katholieke Universiteit Leuven, Belgium. Our research is supported by Research Council KUL: GOA-Mefisto 666, IDO/99/003 (Predictive computer models for medical classification problems using patient data and expert knowledge), several PhD/postdoc & fellow grants; Flemish Government: FWO: PhD/postdoc grants, projects, G.0200.00 (damage detection in composites by optical fibers), G.0078.01 (structured matrices), G.0407.02 (support vector machines), G.0269.02 (magnetic resonance spectroscopic imaging), G.0270.02 (nonlinear Lp approximation), research communities (ICCoS, ANMMM); AWI: Bil. Int. Collaboration Hungary/Poland; IWT: PhD Grants, Belgian Federal Government: DWTC (IUAP IV-02 (1996-2001) and IUAP V-22 (2002-2006): Dynamical Systems and Control: Computation, Identification & Modelling)); EU: NICONET, INTERPRET, PDT-COIL, MRS/MRI signal processing (TMR); Contract Research/agreements: Data4s, IPCOS;

Block-row Hankel Weighted Low Rank Approximation

Mieke Schuermans*, Philippe Lemmerling, Sabine Van Huffel

***Abstract*— This paper extends the Weighted Low Rank Approximation (WLRA) approach towards linearly structured matrices. In the case of Hankel matrices with a special block structure an equivalent unconstrained optimization problem is derived and an algorithm for solving it is proposed.**

***Index Terms*— rank reduction, structured matrices, weighted norm.**

I. INTRODUCTION

The Structured Weighted Low Rank Approximation (SWLRA) problem can be formulated as follows :

$$\min_{\substack{R \\ rank(R)\leq r \\ R\in\Omega}} \parallel X - R \parallel_W^2, \qquad (1)$$

with $X, R \in \mathbb{R}^{n\times m}$, $n \geq m$ (if $n < m$, replace X by X^T), Ω the set of all matrices having the same structure as X and $\|M\|_W^2 \equiv vec(M)^T W vec(M)$ with W a positive definite symmetric weighting matrix. Problem (1) is an extension of the WLRA problem [10] which corresponds to (1) where $\Omega = \mathbb{R}^{n\times m}$. We define $\Delta r \equiv m - r$.

From this point on we will focus on a specific Ω, since as will become clear further on, it is not possible to derive a general algorithm that can deal with any type of linearly structured matrices. The block-Hankel structure with blocks equal to row vectors of length s will be investigated further on. This is the basic matrix structure in Multiple-Input Single-Output (MISO) system identification. Due to the applications in this field, we will work with block-row Hankel matrices of which the elements are Markov parameters. As a result also block-row Toeplitz matrices are dealt with since they can be converted into block-row Hankel matrices by a simple permutation of the rows. The solution of the corresponding SWLRA problem can be found by applying the same permutation to the solution of the Hankel case.

The aim of the paper is the extension of the concept and the algorithm presented in [10] to linearly structured matrices. In [13] the SWLRA for a particular structure, namely the scalar Hankel structure was presented. The main contribution of this paper is the extension to matrices with a specific block-row Hankel structure. Problem (1) can also be interpreted as an extension of the Structured Total Least Squares (STLS) problems described in [4], [5], [6], [7] since (1) is not limited to $\Delta r = 1$.

The paper is structured as follows. The exact problem formulation is given in section II. In section III a solution is derived for the SWLRA problem involving block-row Hankel matrices. Section IV introduces an algorithm for solving problem (1) and in the last section some numerical results are presented to compare the SWLRA with the TLS-ESPRIT like algorithm HTLSstack described in [15], [16]. By fitting the Markov parameters (elements of a Hankel/Toeplitz matrix) to an exponential data model of order r, HTLSstack finds a suboptimal Hankel/Toeplitz matrix approximation of rank r by filling in the fitted elements instead of the original ones. Whereas HTLS [2] only applies to scalar Hankel/Toeplitz matrices, HTLSstack [15], [16] extends the HTLS algorithm to block-row Hankel/Toeplitz matrices.

Manuscript received July 02, 2003.
Dr. Sabine Van Huffel is a full professor at the Katholieke Universiteit Leuven, Belgium. Dr. P. Lemmerling is a postdoctoral researcher with the FWO. Research supported by
Research Council KUL: GOA-Mefisto 666, IDO /99/003 and /02/009 (Predictive computer models for medical classification problems using patient data and expert knowledge), several PhD/postdoc & fellow grants;
Flemish Government:
- FWO: PhD/postdoc grants, projects, G.0200.00 (damage detection in composites by optical fibres), G.0078.01 (structured matrices), G.0407.02 (support vector machines), G.0269.02 (magnetic resonance spectroscopic imaging), G.0270.02 (nonlinear Lp approximation), research communities (ICCoS, ANMMM);
- AWI: Bil. Int. Collaboration Hungary/ Poland;
- IWT: PhD Grants,
Belgian Federal Government: DWTC (IUAP IV-02 (1996-2001) and IUAP V-22 (2002-2006): Dynamical Systems and Control: Computation, Identification & Modelling)); EU: NICONET, INTERPRET, PDT-COIL, MRS/MRI signal processing (TMR);
M. Schuermans, P. Lemmerling and S. Van Huffel are with the Department of Electrical Engineering, ESAT-SCD, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, 3001 Leuven, Belgium (phone:+32-16-321710; fax:+32-16-321970; e-mail: mieke.schuermans@esat.kuleuven.ac.be; philippe.lemmerling@esat.kuleuven.ac.be; sabine.vanhuffel@esat.kuleuven.ac.be).

The difference between HTLS(stack) and TLS-ESPRIT [14] is that HTLS(stack) only applies to (block-row) Hankel/Toeplitz structured matrices and makes use of the Singular Value Decomposition (SVD) of the data matrix X instead of the eigendecomposition of the sample covariance matrix. In [15] it is shown that HTLS(stack) in fact correspond to TLS-ESPRIT, because the sample covariance matrix can be written as cXX^*, where c is some multiplication factor and X^* represents the conjugate transpose of the matrix X.

II. Problem formulation

From a statistical point of view it only makes sense to treat identical elements in an identical way. Therefore (1) is reformulated as:

$$\min_{\substack{vec_2(R)\\ rank(R)\leq r}} \|X-R\|_W^2, \tag{2}$$

with $\|A\|_W^2 \equiv vec_2(A)^T W vec_2(A)$, where $vec_2(A)$ is a minimal vector representation of the linearly structured matrix A. E.g., when Ω represents the set of block-row Hankel matrices $A \in \mathbb{R}^{n\times st}$ of the form

$$A = \begin{bmatrix} a_1 & a_2 & a_3 & \dots & a_t \\ a_2 & a_3 & \dots & & \\ a_3 & \dots & & & \\ \vdots & & & & \vdots \\ a_n & \dots & & & a_{n+t-1} \end{bmatrix}$$

with a_i a row vector of length s for $i = 1, \dots, n+t-1$, $vec_2(A) = [a_1 \dots a_{n+t-1}]^T$. Note that due to the one-to-one relation between A and $vec_2(A)$ the condition $R \in \Omega$ no longer appears in (2).
As in [10] problem (2) can be reformulated as the following equivalent double minimization problem:

$$\min_{\substack{N\in\mathbb{R}^{st\times(st-r)}\\ N^TN=I}} (\min_{\substack{R\in\mathbb{R}^{n\times st}\\ RN=0}} \|X-R\|_W^2). \tag{3}$$

III. Derivation

The first step in the derivation consists of finding a closed form expression $f(N)$ for the solution of the inner minimization of (3). This is obtained as follows [10]: applying the technique of Lagrange multipliers to the inner minimization of (3) yields the Lagrangian

$$\begin{aligned}\psi(L,R) &= vec_2(X-R)^T W vec_2(X-R) \\ &\quad -tr(L^T(RN)),\end{aligned} \tag{4}$$

where $tr(A)$ stands for the trace of matrix A and L is the matrix of Lagrange multipliers. Using the equalities

$$vec(A)^T vec(B) = tr(A^TB),$$
$$vec(ABC) = (C^T \otimes A) vec(B),$$

(4) becomes

$$\begin{aligned}\psi(L,R) &= vec_2(X-R)^T W vec_2(X-R) \\ &\quad -vec(L)^T (N^T \otimes I) vec(R),\end{aligned} \tag{5}$$

where $\otimes$ denotes the Kronecker product and $vec(A)$ stands for the vectorized form of A, i.e. a vector constructed by stacking the consecutive columns of A in one vector. Note that when R is a linearly structured matrix it is straightforward to write down a relation between $vec(R)$ and its minimal vector representation $vec_2(R)$:

$$vec(R) = H vec_2(R), \tag{6}$$

where $H \in \mathbb{R}^{nst\times q}$ and q is the number of different elements in R. E.g., in the case of a block-row Hankel matrix with rows of length s, $q = s(n+t-1)$ and $vec_2(R)$ can be constructed from the elements of the different blocks of R. Substituting (6) in (5), the following expression is obtained for the Lagrangian:

$$\begin{aligned}\psi(L,R) &= vec_2(X-R)^T W vec_2(X-R) \\ &\quad -vec(L)^T (N^T \otimes I_n) H vec_2(R).\end{aligned} \tag{7}$$

Setting the derivatives of ψ w.r.t. $vec_2(R)$ and L equal to 0 yields the following set of equations:

$$\begin{bmatrix} 2W & -H^T(N\otimes I_n) \\ (N^T\otimes I_n)H & 0 \end{bmatrix} \begin{bmatrix} vec_2(R) \\ vec(L) \end{bmatrix} = \begin{bmatrix} 2W vec_2(X) \\ 0 \end{bmatrix}. \tag{8}$$

Using the fact that

$$\begin{bmatrix} A & -B \\ B^T & 0 \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} - A^{-1}B(B^TA^{-1}B)^{-1}B^TA^{-1} & * \\ -(B^TA^{-1}B)^{-1}B^TA^{-1} & * \end{bmatrix}, \tag{9}$$

it follows from (8) that

$$\begin{aligned} vec_2(R) &= (I_q - W^{-1}H^T(N\otimes I_n)[(N\otimes I_n)^T H \\ &\quad W^{-1}H^T(N\otimes I_n)]^{-1}(N\otimes I_n)^T H) vec_2(X) \Rightarrow \\ vec_2(X-R) &= W^{-1}H^T(N\otimes I_n)[(N\otimes I_n)^T H \\ &\quad W^{-1}H^T(N\otimes I_n)]^{-1}(N\otimes I_n)^T H vec_2(X) \end{aligned} \tag{10}$$

As a result, the double minimization problem (3) can be written as the following optimization problem:

$$\min_{\substack{N\in\mathbb{R}^{st\times(st-r)}\\ N^TN=I}} vec_2(X)^T H_2^T (H_2 W^{-1} H_2^T)^{-1} H_2 vec_2(X), \tag{11}$$

with $H_2 \equiv (N\otimes I_n)^T H$.

IV. Algorithm

The straightforward approach for solving (3) would be to apply a nonlinear least squares (NLLS) solver to (11). For $\Delta r = 1$ this works fine but when $\Delta r > 1$ this approach always breaks down by yielding the trivial solution $R = 0$. This can easily be understood by considering (3) with W equal to the identity matrix. In this case, as can be seen from (10), the inner minimization of (3) corresponds to an orthogonal projection of $vec_2(X)$ on the orthogonal complement of the column space of $H_2^T \in \mathbb{R}^{s(n+t-1)\times n\Delta r}$. Therefore it is clear that for $\Delta r = 1$ the orthogonal complement of the column space of H_2^T will never be empty but for $\Delta r > 1$ the latter orthogonal complement can be empty since $n \geq t$ (e.g. in the scalar case for $s = 1$). As a result the projection will yield $R = 0$. The problem obviously lies in an overparametrization of H_2^T (as explained in [13]).
To avoid this problem, a problem formulation equivalent to (3) is derived simply by replacing X with $\tilde{X} \in \mathbb{R}^{s(n+t-r-1)\times(r+1)}$ and by setting $\Delta r = 1$:

$$\tilde{X} = \begin{bmatrix} a_1 & a_2 & a_3 & \dots & a_t & \dots & a_{n+t-r-1} \\ a_2 & a_3 & \dots & & & & \\ a_3 & \dots & & & & & \\ \vdots & & & & & & \vdots \\ a_{r+1} & \dots & & & & \dots & a_{n+t-1} \end{bmatrix}^T \qquad (*)$$

with a_i a row vector of length s for $i = 1, \dots, n+t-1$.
The equivalency of this reformulated problem becomes clear as follows: because the elements of the block-row Hankel matrices that we work with in this paper are Markov parameters, every block-row a_i of the matrix $\tilde{X}$ can be parameterized by means of the state space representative matrices:
$a_i = CA^{i-1}B$ for $i = 1, \dots, n+t-1$.
The same parameterization holds for problem (3) because the same block-rows a_i appear in the matrix X. The only difference is that in this case, after rearranging, the system only needs a rank reduction by one to come up with matrices A^{i-1} of rank r, instead of a rank reduction by $st - r$.
For this equivalent problem formulations (11) and (10) thus become:

$$\min_{\substack{p\in\mathbb{R}^{r+1} \\ p^Tp=1}} vec_2(\tilde{X})^T \tilde{H}_2^T (\tilde{H}_2 W^{-1} \tilde{H}_2^T)^{-1} \tilde{H}_2 vec_2(\tilde{X}), \quad (12)$$

with $\tilde{H}_2 \equiv (p \otimes I)^T \tilde{H} \in \mathbb{R}^{s(n+t-r-1)\times s(n+t-1)}$ and I the identity matrix of size $s(n+t-r-1)$, where N is replaced by p, to indicate that it is a column vector i.o. a matrix.

A formulation of the vectorized form of $\tilde{R}$ is equal to:

$$vec_2(\tilde{R}) = (I_q - W^{-1}\tilde{H}_2^T[\tilde{H}_2 W^{-1}\tilde{H}_2^T]^{-1}\tilde{H}_2)\, vec_2(\tilde{X}) \qquad (13)$$

The algorithm for solving the SWLRA for block-row Hankel matrices can be summarized as follows:

Algorithm block-row SWLRA

Input: block-row Hankel matrix $X \in \mathbb{R}^{n\times st}$ with blocks=row vectors $a_1, a_2, \dots, a_{n+t-1}$ of length s, rank r and weighting matrix W.
Output: block-row Hankel matrix R of rank $\leq r$, such that R is as close as possible to X in $\|.\|_W$-sense.
Begin

- Step 1 Construct matrix $\tilde{X}$ by rearranging the elements of matrix X such that $\tilde{X} \in \mathbb{R}^{s(n+t-r-1)\times(r+1)}$ is a block Hankel matrix with blocks=column vectors $a_1^T, \dots, a_{n+t-1}^T$ (see $(*)$).
- Step 2 Compute SVD of $\tilde{X}$: $\tilde{X} = U\Sigma V^T$.
- Step 3 Initialize p_0 with the right singular vector of $\tilde{X}$ corresponding to its $(r+1)$th singular value.
- Step 4 Minimize the cost function in (12).
- Step 5 Compute $\tilde{R}$ using (13).
- Step 6 Rearrange the elements of $\tilde{R}$ into a matrix R such that $R \in \mathbb{R}^{n\times st}$ and R has the same structure as X (i.e. similar rearrangement as going from $\tilde{X}$ to X in $(*)$).

End

In step 4 a standard NLLS (Matlab's `lsqnonlin`) is used. In order to use a NLLS routine, the cost function has to be casted in the form f^Tf and in order to do so the Cholesky decomposition of $\tilde{H}_2^T(\tilde{H}_2 W^{-1}\tilde{H}_2^T)^{-1}\tilde{H}_2$ has to be computed. This can be done by means of a QR factorisation of $W^{-1/2}\tilde{H}_2^T$. The computationally most expensive step is step 4.

V. Numerical results

In this section we compare the statistical accuracy of algorithm block-row SWLRA and the TLS-ESPRIT like algorithm HTLSstack proposed in [15]. We expect that the block-row SWLRA algorithm will perform better than the TLS-ESPRIT like algorithm HTLSstack. Indeed, SWLRA seeks for the closest lower-rank Hankel/Toeplitz matrix by solving (12), while HTLSstack only computes a suboptimal Hankel/Toeplitz matrix estimate by exploiting the shift-invariant structure of the block-row Hankel matrix.

For comparison, we perform Monte-Carlo simulations. In table I, the results for a 25×12 block-row Hankel matrix X are presented with blocks of length 2, whereas table II contains the results for a 50×24 block-row Hankel matrix with blocks of length 4. The Monte-Carlo simulations are performed as follows. A block-row Hankel matrix of the required rank r is constructed. The latter matrix is perturbed by a block-row Hankel matrix that is constructed using a vector containing i.i.d. Gaussian noise of standard deviation σ. Finally, the SWLRA and the TLS-ESPRIT like algorithm HTLSstack are applied to this perturbed block-row Hankel matrix. Each Monte-Carlo simulation consists of 100 runs, i.e. considers 100 noisy realizations, for every considered σ and r. For both algorithms the relative error $\frac{\|X-R\|_F}{\|X\|_F}$ is presented. The presented results are averaged over the runs and obtained by implementing both algorithms in Matlab (version 6.1) on a PC i686 with 800 MHz and 256 MB memory.

From both tables we can conclude the following. As expected, the SWLRA algorithm performs better than the TLS-ESPRIT like algorithm HTLSstack for any choice of r.

VI. Conclusions

In this paper we have developed an extension of the Weighted Low Rank Approximation introduced by Manton et al. [10] towards linearly structured matrices. For a particular type of structure, namely the block-Hankel structure with blocks equal to rows, an algorithm was developed. Simulation experiments confirm the improved statistical accuracy of the latter algorithm compared to that of the TLS-ESPRIT like algorithm HTLSstack.

References

[1] S. Van Huffel and J. Vandewalle, ”The total least squares problem: computational aspects and analysis,” *Frontiers in Applied Mathematics series*, vol. 9, SIAM, Philadelphia, 1991.

[2] S. Van Huffel, H. Chen, C. Decanniere and P. Van Hecke, ”Algorithm for time-domain NMR data fitting based on total least squares,” *Journal of Magnetic Resonance A*, vol. 110, pp. 228-237, 1994.

[3] G.H. Golub and C.F. Van Loan, ”An analysis of the total least squares problem,” *SIAM Journal Numer. Anal.* vol. 17, pp.883-893, 1980.

[4] T. J. Abatzoglou and J. M. Mendel, ”Constrained total least squares,” *IEEE International Conf. on Acoustics, Speech & Signal Processing*, Dallas, pp. 1485-1488, 1987.

[5] T. J. Abatzoglou, J. M. Mendel and G. A. Harada, ”The constrained total least squares technique and its applications to harmonic superresolution,” *IEEE Transactions on Signal Processing*, vol. 39, pp. 1070-1086, 1991.

[6] B. De Moor, ”Total least squares for affinely structured matrices and the noisy realization problem,” *IEEE Transactions on Signal Processing*, vol. 42, pp. 3004-3113, 1994.

[7] J. B. Rosen, H. Park, and J. Glick, ”Total least norm formulation and solution for structured problems,” *SIAM Journal on Matrix Anal. Appl.*, vol. 17, no. 1, pp. 110-126, 1996.

[8] M. T. Chu, R. E. Funderlic and R. J. Plemmons, ”Structured lower rank approximation,” To appear in *Linear Algebra and Its Applications*. Preprint 2000.

[9] S. Van Huffel, H. Park and J. B. Rosen, ”Formulation and solution of structured total least norm problems for parameter estimation,” *IEEE Transactions on signal processing*, vol. 44, No 10, October 1996.

[10] J. H. Manton, R. Mahony and Y. Hua, ”The geometry of weighted low rank approximations,” *IEEE Transactions on Signal Processing*, vol. 51, No.2, 2003.

[11] P. Lemmerling, N. Mastronardi and S. Van Huffel, ”Fast algorithm for solving the Hankel/Toeplitz structured total least squares problem,” *Numerical Algorithms*, vol. 23, pp. 371-392, 2000.

[12] T. Kailath and A. H. Sayed, editors, ”Fast reliable algorithms for matrices with structure,” *SIAM*, Philadelphia, 1999.

[13] M. Schuermans, P. Lemmerling and S. Van Huffel, ”Structured Weighted Low Rank Approximation,” ESAT-SCD Technical Report 03-04, Katholieke Univ. Leuven (Leuven, Belgium), January 2003.

[14] R. Roy and T. Kailath, ”Esprit-estimation of signal parameters via rotational invariance techniques,” *IEEE Trans. Acous., Speech, Signal Processing*, vol. 37, No.7, pp. 984-995, July 1989.

[15] G. Morren, P. Lemmerling and S. Van Huffel, ”Decimative subspace-based parameter estimation techniques,” *Signal Processing*, vol. 83, No. 5, pp. 1025-1033, May 2003.

[16] L. Vanhamme and S. Van Huffel, ”Multichannel quantification of biomedical Magnetic Resonance Spectroscopic signals,” invited paper for the session on ’Sensor array signal processing’, *Proceedings SPIE conference on Advanced Signal Processing: Algorithms, Architectures, and Implementations VIII*, ed. F.T. Luk, San Diego, pp. 237-248, 22-24 July, 1998.

TABLE I

APPROXIMATION OF A 25×12 BLOCK-HANKEL MATRIX X WITH BLOCKS 1×2 BY A MATRIX R OF RANK r. COMPARISON OF THE RELATIVE NORM $\frac{\|X-R\|_F}{\|X\|_F}$ COMPUTED VIA THE ALGORITHMS BLOCK-ROW SWLRA AND HTLSSTACK.

r	σ	SWLRA	HTLSstack
2	10^{-3}	0.00095	0.00096
	10^{-2}	0.00941	0.00950
	0.1	0.11534	0.11622
	1.0	0.63870	0.64713
4	10^{-3}	0.00105	0.00107
	10^{-2}	0.00932	0.00936
	0.1	0.10341	0.10452
	1.0	0.58715	0.59550
6	10^{-3}	0.00091	0.00093
	10^{-2}	0.00929	0.00965
	0.1	0.07790	0.08497
	1.0	0.55468	0.59148
8	10^{-3}	0.00088	0.00090
	10^{-2}	0.00652	0.00712
	0.1	0.07417	0.07476
	1.0	0.54678	0.67650
10	10^{-3}	0.00069	0.00075
	10^{-2}	0.00781	0.00873
	0.1	0.07522	0.08262
	1.0	0.52649	0.71059

TABLE II

APPROXIMATION OF A 50×24 BLOCK-HANKEL MATRIX X WITH BLOCKS 1×4 BY A MATRIX R OF RANK r. COMPARISON OF THE RELATIVE NORM $\frac{\|X-R\|_F}{\|X\|_F}$ COMPUTED VIA THE ALGORITHMS BLOCK-ROW SWLRA AND HTLSSTACK.

r	σ	SWLRA	HTLSstack
4	10^{-3}	0.00092	0.00093
	10^{-2}	0.00917	0.00929
	0.1	0.09080	0.09201
	1.0	0.63603	0.64525
8	10^{-3}	0.00087	0.00091
	10^{-2}	0.00872	0.00909
	0.1	0.08645	0.08935
	1.0	0.60860	0.69709
12	10^{-3}	0.00080	0.00085
	10^{-2}	0.00801	0.00852
	0.1	0.07879	0.08337
	1.0	0.57914	0.68251
16	10^{-3}	0.00082	0.00091
	10^{-2}	0.00820	0.00902
	0.1	0.08079	0.09395
	1.0	0.54452	0.77840
20	10^{-3}	0.00077	0.00120
	10^{-2}	0.00690	0.00810
	0.1	0.06773	0.07665
	1.0	0.48384	0.69717