21st International Symposium on
Mathematical Theory of Networks and Systems
July 7-11, 2014. Groningen, The Netherlands

Stochastic Methods for Hankel Structured Low Rank Approximation

Jonathan Gillard[1] and Anatoly Zhigljavsky[2]

Abstract— **In this paper, we investigate the complexity of the numerical construction of the Hankel structured low-rank approximation (HSLRA) problem, and develop a family of algorithms to solve this problem. Briefly, HSLRA is the problem of finding the closest (in some pre-defined norm) rank r approximation of a given Hankel matrix, which is also of Hankel structure. Unlike many other methods described in the literature the family of algorithms we propose has the property of guaranteed convergence.**

I. Introduction

A. Statement of the problem

Let L, K and r be given positive integers such that $1 \leq r < L \leq K$. Denote the set of all real-valued $L \times K$ matrices by $\mathbb{R}^{L\times K}$. Let $\mathcal{M}_r = \mathcal{M}_r^{L\times K} \subset \mathbb{R}^{L\times K}$ be the subset of $\mathbb{R}^{L\times K}$ containing all matrices with rank $\leq r$, and $\mathcal{H} = \mathcal{H}^{L\times K} \subset \mathbb{R}^{L\times K}$ be the subset of $\mathbb{R}^{L\times K}$ containing matrices of some known structure. The set of structured $L \times K$ matrices of rank $\leq r$ is $\mathcal{A} = \mathcal{M}_r \cap \mathcal{H}$.

Assume we are given a matrix $\mathbf{X}_* \in \mathcal{H}$. The problem of structured low rank approximation (SLRA) is:

$$f(\mathbf{X}) \to \min_{\mathbf{X}\in\mathcal{A}} \tag{1}$$

where $f(\mathbf{X}) = \rho^2(\mathbf{X}, \mathbf{X}_*)$ is a squared distance on $\mathbb{R}^{L\times K} \times \mathbb{R}^{L\times K}$.

In this paper we only consider the case where $\mathcal{H}$ is the set of Hankel matrices and thus refer to (1) as HSLRA. Recall that a matrix $\mathbf{X} = (x_{lk}) \in \mathbb{R}^{L\times K}$ is called Hankel if $x_{lk} =$ const for all pairs (l, k) such that $l + k =$ const; that is, all elements on the anti-diagonals of $\mathbf{X}$ are equal. There is a one-to-one correspondence between $L \times K$ Hankel matrices and vectors of size $N = L + K - 1$. For a vector $Y = (y_1, \ldots, y_N)^T$, the matrix $\mathbf{X} = \mathbb{H}(Y) = (x_{lk}) \in \mathbb{R}^{L\times K}$ with elements $x_{lk} = y_{l+k-1}$ is Hankel and vise-versa: for any matrix $\mathbf{X} \in \mathcal{H}$, we may define $Y = \mathbb{H}^{-1}(\mathbf{X})$ so that $\mathbf{X} = \mathbb{H}(Y)$.

B. Background

The aim of low-rank approximation methods is to approximate a matrix containing observed data, by a matrix of pre-specified lower rank r. The rank of the matrix containing the original data can be viewed as the order of complexity required to fit to the data exactly, and a matrix of lower complexity (lower rank) 'close' to the original matrix is often required. A further requirement is that if the original matrix of the observed data is of a particular structure, then the approximation should also have this structure. An example is the HSLRA problem as defined in the previous section.

HSLRA is a very important problem with applications in a number of different areas. In addition to the clear connection with time series analysis and signal processing, HSLRA has been extensively used in system identification (modeling dynamical systems) [1], in speech and audio processing [2], in modal and spectral analysis [3] and image processing [4]. Some discussion on the relationship of HSLRA with some well known subspace-based methods of time series analysis and signal processing is given in [5]. Similar structures used in (1) include Toeplitz, block Hankel and block Toeplitz structures. In image processing, there is much use of Hankel-block Hankel structures. Further details, references and specific applications of SLRA are provided in [6], [7], [8].

II. HSLRA as an optimization problem

A. Matrix optimal solution and its approximation

Consider the HSLRA problem (1). Since $0 \leq f(\mathbf{X}) < \infty$ for any $\mathbf{X} \in \mathcal{A}$, $f(\mathbf{X})$ is a continuous function on $\mathcal{A}$ and $f(\mathbf{X}) \to \infty$ as $||\mathbf{X}|| \to \infty$, a solution to (1) always exists. However, the solution is not necessarily unique. Set

$$\mathfrak{X}^* = \{\mathbf{X}^* = \arg\min_{\mathbf{X}\in\mathcal{A}} f(\mathbf{X})\},\ f^* = f(\mathbf{X}^*) = \min_{\mathbf{X}\in\mathcal{A}} f(\mathbf{X})\,.$$

Any algorithm designed to solve the optimization problem (1) should return a matrix $\mathbf{X}_{appr}$ which can be considered as an approximation to one of the solutions $\mathbf{X}^* \in \mathfrak{X}^*$; approximations to the value f^* alone (without approximations to $\mathfrak{X}^*$) are not sufficient. Ideally, the matrix $\mathbf{X}_{appr}$ should belong to the set of matrices $\mathcal{A}$.

A typical optimization algorithm designed for solving the problem (1) could be represented as a procedure which generates a sequence of matrices $\mathbf{X}_0, \mathbf{X}_1, \ldots$ such that some of the matrices $\mathbf{X}_n$ for large n can be considered as approximations to $\mathbf{X}^*$, a solution of (1). This matrix sequence must have at least one limiting point (matrix) in the set $\mathcal{A}$. Denote by $\mathbf{X}_\infty$ the limiting point of the algorithm which has the smallest value of f among all its limiting points belonging to $\mathcal{A}$. In general, $f(\mathbf{X}_\infty) \geq f^*$. The algorithm (theoretically) converges to the optimal solution if $f(\mathbf{X}_\infty) = f^*$; that is, if $\mathbf{X}_\infty \in \mathfrak{X}^*$.

B. Distances defining the objective function

There are two natural distance functions ρ which define the objective function f in (1). The most natural squared distance ρ^2 is determined by the squared Frobenius norm:

$$||\mathbf{X}||_F^2 = \sum_{l=1}^{L}\sum_{k=1}^{K} x_{lk}^2 \quad \text{for } \mathbf{X} \in \mathbb{R}^{L\times K}\,. \tag{2}$$

[1]Jonathan Gillard is with Cardiff School of Mathematics, Cardiff University GillardJW@Cardiff.ac.uk

[2]Anatoly Zhigljavsky is with Cardiff School of Mathematics, Cardiff University GillardJW@Cardiff.ac.uk

ISBN: 978-90-367-6321-9

Every $L \times K$ Hankel matrix $\mathbf{X} \in \mathcal{H}$ is in a one-to-one correspondence with some vector $Y = (y_1, \ldots, y_N)^T$, with $N = L + K - 1$. Let the function $\mathbb{H} : \mathbb{R}^N \to \mathcal{H}^{L\times K}$ be defined such that $\mathbb{H}(Y) = ||y_{l+k-1}||_{l,k=1}^{L,K}$ for $Y = (y_1, \ldots, y_N)^T$; that is, $\mathbb{H}(Y)$ maps a vector $Y \in \mathbb{R}^N$ to an $L \times K$ Hankel matrix. Each element of the vector Y is repeated in $\mathbf{X} = \mathbb{H}(Y)$ several times. Let $\mathbf{E} = (e_{lk}) \in \mathbb{R}^{L\times K}$ be the matrix consisting entirely of ones. We can compute the sum of each anti-diagonal of $\mathbf{E}$, denoted v_n, as

$$v_n = \sum_{l+k=n+1} e_{lk} = \begin{cases} n & \text{for } n = 1, \ldots, L-1\,, \\ L & \text{for } n = L, \ldots, K-1, \\ N-n+1 & \text{for } n = K, \ldots, N. \end{cases} \tag{3}$$

The value v_n is the number of times the element y_n of the vector Y is repeated in the Hankel matrix $\mathbb{H}(Y)$. Denote by $\mathbf{V} = \operatorname{diag}(v_1, \ldots, v_N)$ the diagonal matrix with diagonal elements $v_1, \ldots, v_N$.

If we compute the norm (2) for the Hankel matrix $\mathbf{X} = \mathbb{H}(Y)$ and express this formula in terms of the associated vector Y, then we obtain

$$||\mathbf{X}||_F^2 = \sum_{n=1}^{N} v_n y_n^2 = Y^T \mathbf{V} Y \text{ for } \mathbf{X} = \mathbb{H}(Y) \text{ with } x_{lk} = y_{l+k-1}\,. \tag{4}$$

The squared Euclidian norm of the vector Y (associated with the matrix $\mathbf{X} = \mathbb{H}(Y)$) defines another common distance ρ:

$$||\mathbf{X}||^2 = \sum_{n=1}^{N} y_n^2 = Y^T Y \quad \text{for } \mathbf{X} = \mathbb{H}(Y)\,. \tag{5}$$

The general weighted squared distance is defined as

$$||\mathbf{X}||_W^2 = Y^T \mathbf{W} Y \tag{6}$$

where $\mathbf{W}$ is an arbitrary non-negative definite matrix which can sometimes be interpreted as a covariance matrix of the observations Y. For the cases $\mathbf{W} = \mathbf{V}$ and $\mathbf{W} = \mathbf{I}_N$, the squared distance (6) reduces to (4) and (5), respectively.

C. Projection to $\mathcal{M}_r$ for $\mathbf{W} = \mathbf{V}$ (Frobenius norm)

Unlike $\mathcal{H}$, the set $\mathcal{M}_r$ is clearly not convex. However, the projections from $\mathbb{R}^{L\times K}$ to $\mathcal{M}_r$ for the standard Frobenius norm can be easily computed using the singular value decomposition (SVD) of $\mathbf{X}$, see [9]. Let $\sigma_i = \sigma_i(\mathbf{X})$, the singular values of $\mathbf{X}$, be ordered such that $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_L$. Denote $\Sigma_0 = \operatorname{diag}(\sigma_1, \sigma_2, \ldots, \sigma_L)$ and $\Sigma = \operatorname{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r, 0, \ldots, 0)$. Then the SVD of $\mathbf{X}$ can be written as $\mathbf{X} = U\Sigma_0 V^T$ and the matrix $\pi^{(r)}(\mathbf{X}) = U\Sigma V^T$ belongs to $\mathcal{M}_r$ and minimizes the value $||\mathbf{X} - \mathbf{X}'||_F^2$ over $\mathbf{X}' \in \mathcal{M}_r$. The projection $\pi^{(r)}(\mathbf{X})$ of $\mathbf{X}$ onto $\mathcal{M}_r$ is uniquely defined if and only if $\sigma_r > \sigma_{r+1}$. The squared (Frobenius) distance between matrix $\mathbf{X}$ and $\mathcal{M}_r$ is

$$\begin{aligned} \rho^2(\mathbf{X}, \mathcal{M}_r) &= \min_{\mathbf{X}' \in \mathcal{M}_r} \rho^2(\mathbf{X}, \mathbf{X}') = \\ = \rho^2(\mathbf{X}, \pi^{(r)}(\mathbf{X})) &= ||\mathbf{X} - \pi^{(r)}(\mathbf{X})||_F^2 = \sum_{i=r+1}^{L} \sigma_i^2(\mathbf{X})\,. \end{aligned}$$

D. Projection to $\mathcal{H}$

The space $\mathcal{H} = \mathcal{H}^{L\times K}$ of $L \times K$ Hankel matrices is a linear subspace of $\mathbb{R}^{L\times K}$. The closest Hankel matrix (in the norm defined by (6) with any diagonal matrix $\mathbf{W}$ including the Frobenius norm and the norm defined by (5)) to any given matrix is obtained by using the simple diagonal averaging procedure.

Let $\pi_{\mathcal{H}}(\mathbf{X})$ denote the projection of a matrix $\mathbf{X} \in \mathbb{R}^{L\times K}$ onto the space $\mathcal{H}$. Then the element $\tilde{x}_{ij}$ of $\pi_{\mathcal{H}}(\mathbf{X})$ is given by

$$\tilde{x}_{ij} = v_{i+j}^{-1} \sum_{l+k=i+j} x_{lk}\,.$$

The squared (Frobenius) distance between matrix $\mathbf{X}$ and the space $\mathcal{H}$ is

$$\rho^2(\mathbf{X}, \mathcal{H}) = \min_{\mathbf{X}' \in \mathcal{H}} \rho^2(\mathbf{X}, \mathbf{X}') = ||\mathbf{X} - \pi_{\mathcal{H}}(\mathbf{X})||_F^2\,.$$

Since projecting to $\mathcal{H}$ is very easy, using this subspace as the feasible domain for the HSLRA problem (1) is more natural than using the original space $\mathbb{R}^{L\times K}$.

III. Algorithms based on the use of alternating projections

In this section we consider algorithms for solving the HSLRA problem represented as optimization problems using alternating projections between the spaces $\mathcal{H}$ and $\mathcal{M}_r$. We restrict our attention to the distance function associated with the matrix Frobenuis norm (4), that is, we take $\mathbf{W} = \mathbf{V}$ in (6).

A. Classical algorithms and their modifications

1) Alternating projections (AP): The algorithm (7) below is the direct implementation of the alternating projections. For brevity we will refer to this algorithm as AP.

$$\mathbf{X}_0 = \mathbf{X}_*, \quad \mathbf{X}_{n+1} = \pi_{\mathcal{H}}\left[\pi^{(r)}(\mathbf{X}_n)\right] \quad \text{for } n = 0, 1, \ldots \tag{7}$$

These projections have also been studied in [10] and are sometimes known as Cadzow iterations [11]. Here we simply alternate projections to the space $\mathcal{M}_r$ with projections to the space $\mathcal{H}$. In this form of alternating projections, we have $\mathbf{X}_n \in \mathcal{H}$ for all $n = 0, 1, \ldots$.

Note also that one iteration of AP for HSLRA corresponds to the basic version of the technique of time series analysis known as singular spectrum analysis (SSA), see [12]; for further details regarding the link between AP and SSA; see, for example, Gillard [11].

2) Discussion on the convergence of AP: Despite AP often appearing to be myopic and too greedy by only aiming at minimizing the distance $\rho^2(\mathbf{X}, \mathcal{M}_r)$, it is very popular in practice. The popularity of AP is explained by the simplicity of the algorithm and by the fact that convergence to the space $\mathcal{A}$ is guaranteed, see [13]. AP often converges to a matrix which is far away from the set of optimal solutions $\mathfrak{X}^*$.

As shown in [14, Th. 6.1], AP converges linearly; that is, there exist constants $c < 1$ and $A > 0$ such that

$\rho^2(\mathbf{X}_\infty, \mathbf{X}_n) < Ac^n$, $\forall n$, where $\mathbf{X}_\infty$ is some matrix in $\mathcal{A}$. Moreover, it is easy to prove monotonicity of AP iterations. As derived by Chu et al. [13], we have

$$||\mathbf{X}_{n+1} - \pi^{(r)}(\mathbf{X}_{n+1})||_F^2 \leq ||\mathbf{X}_{n+1} - \pi^{(r)}(\mathbf{X}_n)||_F^2 \leq$$
$$\leq ||\mathbf{X}_n - \pi^{(r)}(\mathbf{X}_n)||_F^2 .$$

Numerical results show that the resulting approximation is never worse than the approximation obtained by (7) (usually it is slightly better than the approximation obtained by AP); it also converges faster to the set $\mathcal{A}$. Examples show that typically AP and do not converge to the set of optimal solutions $\mathfrak{X}^*$.

B. Alternating Projections with Backtracking and Randomization

In this section, we describe a family of algorithms which can be run as a random multistart-type algorithm, as a multistage algorithm and also as an evolutionary method. The main steps of this algorithm are summarized by its title 'Alternating Projections with Backtracking and Randomization' and we abbreviate this algorithm APBR. Here we describe two versions of this algorithm, Multistart APBR and APBR with selection. APBR with selection significantly reduces the number of computations by terminating non-prospective trajectories at early stages.

The underpinning idea for the family of the APBR algorithms has been suggested by the authors in [15] where the potential of the multistart APBR has been demonstrated on a number of examples.

The multistart version of APBR is described as follows. Let U denote a realization of a random number with uniform distribution in $[0, 1]$ and let $\tilde{\mathbf{X}}$ denote a random Hankel matrix which corresponds to a realization of a white noise Gaussian process $\tilde{Y} = (\xi_1, \ldots, \xi_N)$ with ξ_i, $i = 1, \ldots N$, independent Gaussian random variables with mean 0 and variance $s^2 \geq 0$.

In Multistart APBR, we run M independent trajectories in the space $\mathcal{H}$ starting at random Hankel matrices

$$\mathbf{X}_{0,j} = (1 - s_0)\mathbf{X}_* + s_0\tilde{\mathbf{X}}, \tag{8}$$

with some s_0 ($0 \leq s_0 \leq 1$), and use the updating formula

$$\mathbf{X}_{n+1,j} = \left(\mathrm{tr}\mathbf{Z}_{n,j}^T\mathbf{X}_*/\mathrm{tr}\mathbf{Z}_{n,j}^T\mathbf{Z}_{n,j}\right)\mathbf{Z}_{n,j} \tag{9}$$

where $j = 1, \ldots, M$,

$$\mathbf{Z}_{n,j} = (1 - \delta_n)\,\pi_{\mathcal{H}}\left[\pi^{(r)}(\mathbf{X}_{n,j})\right] + \delta_n\mathbf{X}_* + \sigma_n\tilde{\mathbf{X}} \tag{10}$$

and

$$\begin{cases} \delta_n = U/(n+1)^p, \quad \sigma_n = c/(n+1)^q, \\ \qquad\qquad \text{if } \rho^2(\mathbf{X}_{n,j}, \mathcal{M}_r) \geq \varepsilon, \\ \delta_n = 0, \;\; \sigma_n = 0, \qquad\qquad \text{otherwise}. \end{cases} \tag{11}$$

Each trajectory is either run until convergence or for a prespecified number of iterations. U could be either random or simply set to 1, $c \in \{0, 1\}$ and positive numbers p, q and ε can be chosen arbitrarily. A MATLAB implementation of this version of APBR, developed by the authors, is available at [16].

If $s_0 = \delta_n = \sigma_n = 0$ then the iterations in (9) coincide with iterations of AP (7) with some local improvement. If $s_0 > 0$ then the j-th trajectory of the algorithm starts at a random matrix in the neighbourhood of $\mathbf{X}_*$ (the width of this neighbourhood is controlled by the parameter s_0). If $\sigma_n > 0$ then there is a 'random mutation' at the n-th iteration (9). When $\delta_n > 0$, the current approximation 'backtracks' towards $\mathbf{X}_*$ conditionally that the backtracking does not worsen the distance $\rho^2(\mathbf{X}_{n,j}, \mathbf{X}_*)$. If $\rho^2(\mathbf{X}_{n,j}, \mathcal{M}_r) < \varepsilon$, we set $\delta_n = 0$ and $\sigma_n = 0$. That is, in the final stage for any trajectory of the APBR we perform AP iterations (7) to achieve faster convergence to $\mathcal{A}$.

IV. Numerical results for the example of De Moor

In this section we consider the data given by De Moor [17] to demonstrate the sub-optimality of AP (7). We have used Multistage APBR, with ρ^2 defined by the Frobenius norm (2) and the number of trajectories was chosen as $M = 1000$. Unless otherwise stated, U is a realization of a random number with uniform distribution in $[0, 1]$. For ease of exposition, (11) is defined as

$$\begin{cases} \delta_n = U/(n+1)^p, \sigma_n = c/(n+1)^q, & \text{for } n = 0, \ldots, P, \\ \delta_n = 0, \;\; \sigma_n = 0, & \text{for } n > P. \end{cases}$$

De Moor's data and parameter settings are as follows: $Y_* = (3, 4, 2, 1, 5, 6, 7, 1, 2)^T$, $N = 9$, $L = 4$, and $r = 3$. Let $\mathbf{X}_* = \mathbb{H}(Y_*)$. Table I contains the Frobenius distances to $\mathbf{X}_*$ obtained using AP (7) for different values of the parameters L and r. Table I also contains the result obtained by De Moor. The approximation achieved by AP and by De Moor, Y_{AP} and Y_{DM} are provided in [17] (four decimal places only). One of the best Multistage APBR approximations for the De Moor's data is $Y_{APBR} = (3.451346, 3.533941, 2.002535, 1.487395, 4.039565, 7.078974, 5.995627, 1.720123, 1.613392)^T$. This Multistage APBR solution is slightly different from Y_{DM} while the values of the Frobenius distances to $\mathbf{X}_*$ coincide (with the precision provided by De Moor). We have no access to the software realizing the method of De Moor and therefore we cannot perform a comparative study involving this method.

In this example, the parameters of Multistage APBR are selected to be $M = 1000$, $P = 500$, $c = 1$, $s_0 = 0.25$, $s = 1$, $p = 0.5$ and $q = 1.5$. The total number of iterations was set at 600. Multistage APBR yields solutions similar to that obtained by the method of De Moor. For all values of L and r Multistage APBR provides better (or, in one case, similar) solutions than the other methods considered.

Table II contains the minimum Frobenius distance to $\mathbf{X}_*$ using Multistage APBR, varying the parameters p and q in (9). With $s_0 = 0.25$ and $s = 1$ it can be seen that there are many (p, q) parameter pairs which yield similar solutions, with Frobenius distances to $\mathbf{X}_*$ comparable to that achieved by De Moor's method. Summarizing the numerical results obtained in this example (and similar ones), we can give

	$L=3$			
	AP	Med (APBR)	Min (APBR)	DM
$r=1$	110.3142	110.0101	110.0095	-
$r=2$	73.6980	72.8550	72.8530	-
$r=3$	14.8251	14.1481	14.1478	14.1478

	$L=4$		
	AP	Med (APBR)	Min (APBR)
$r=1$	111.8552	111.5690	111.5625
$r=2$	73.3795	73.1790	73.1740
$r=3$	15.6168	14.9597	14.9519
$r=3$	3.4535	3.4535	3.4509

TABLE I

FROBENIUS DISTANCES OF THE APPROXIMATIONS TO $\mathbf{X}_*$ USING MULTISTAGE APBR, AP (7), AND DE MOOR (DM) RESULTS.

$p \backslash q$	0.25	0.5	0.75	1	1.5	2
0.25	14.2678	14.1528	14.1483	14.1478	14.1478	14.1478
0.5	14.3230	14.1540	14.1486	14.1478	14.1478	14.1478
0.75	16.6083	14.2294	14.1535	14.1479	14.1478	14.1478
1	17.1857	14.4427	14.1721	14.1498	14.1478	14.1478
1.5	36.1126	16.3141	14.3893	14.4292	14.1976	14.2249
2	38.4927	18.0746	14.9541	14.9674	14.4582	14.7119

TABLE II

MINIMUM FROBENIUS DISTANCES TO $\mathbf{X}_*$ USING MULTISTAGE APBR, VARYING THE PARAMETERS p AND q; $L=4$, $r=3$.

the following recommendations concerning the choice of parameters p and q of the Multistage APBR algorithm (9).

Backtracking (regulated by the parameter p), is extremely important. It is worth noting that numerical results show that in many examples the use of the random variable U in the formula for δ_n in (9) often works marginally better than a constant. For the data considered in this example, Multistage APBR was best performing with values of p in between 0.25 and 1, implying that the rate at which backtracking decreases should be slow.

Randomization (regulated by the parameter q) could be useful too. Note also that the mechanism of this stochastic mutation in Multistage APBR resembles the mechanism of regularization of the Alternating Least Squares methods used in signal processing and in particular tensor decompositions, see [18]. Randomization appears to be beneficial both at the start of the iterations and throughout the running of the algorithm, but as illustrated in Table II, the rate at which randomization decreases should be slightly faster than for backtracking (that is, we recommend choosing $p < q$). For the data considered in this example, Multistage APBR was best performing with values of q in between 1 and 2.

V. CONCLUSION

This paper is devoted to the construction of numerical methods for solving the HSLRA problem. Finding optimal solutions to the HSLRA problem is very difficult. We construct algorithms for solving the HSLRA problem with trajectories in the space $\mathcal{H}$ of Hankel matrices which we believe lead to more tractable algorithms with higher chances of success. This is the approach which we undertook in the main body of the paper.

We have introduced our family of algorithms called APBR, which can be viewed as a global random search extension of AP. Examples show that Multistart APBR significantly outperforms AP and some other methods. Many other, possibly more efficient, techniques could be adapted for solving the HSLRA problem but this is a theme for future research.

REFERENCES

[1] I. Markovsky, J. C. Willems, S. Van Huffel, B. De Moor, and R. Pintelon, "Application of structured total least squares for system identification and model reduction," *IEEE Trans. Automat. Control*, vol. 50, no. 10, pp. 1490–1500, 2005.

[2] P. Lemmerling, N. Mastronardi, and S. Van Huffel, "Efficient implementation of a structured total least squares based speech compression method," *Linear Algebra Appl.*, vol. 366, pp. 295–315, 2003.

[3] A. Yeredor, "Multiple delays estimation for chirp signals using structured total least squares," *Linear Algebra Appl.*, vol. 391, pp. 261–286, 2004.

[4] A. Pruessner and D. P. O'Leary, "Blind deconvolution using a regularized structured total least norm algorithm," *SIAM J. Matrix Anal. Appl.*, vol. 24, no. 4, pp. 1018–1037, 2003.

[5] N. Golyandina, "On the choice of parameters in Singular Spectrum Analysis and related subspace-based methods," *Statistics and Its Interface*, vol. 3, pp. 259–279, 2010.

[6] I. Markovsky, "Structured low-rank approximation and its applications," *Automatica*, vol. 44, no. 4, pp. 891–909, 2008.

[7] ——, "Bibliography on total least squares and related methods," *Statistics and Its Interface*, vol. 3, no. 3, pp. 329–334, 2010.

[8] ——, *Low rank approximation: Algorithms, implementation, applications*. Springer, 2012.

[9] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, 1936.

[10] J. A. Cadzow, "Signal enhancement: A composite property mapping algorithm," *IEEE Trans. on Acoust., Speech, Signal Processing*, vol. 36, pp. 1070–1087, 1988.

[11] J. Gillard, "Cadzow's basic algorithm, alternating projections and singular spectrum analysis," *Statistics and Its Interface*, vol. 3, no. 3, pp. 335–343, 2010.

[12] N. Golyandina and A. A. Zhigljavsky, *Singular Spectrum Analysis for time series*, ser. Springer Briefs in Statistics. Springer, 2013.

[13] M. T. Chu, R. E. Funderlic, and R. J. Plemmons, "Structured low rank approximation," *Linear algebra and its applications*, vol. 366, pp. 157–172, 2003.

[14] F. Andersson and M. Carlsson, "Alternating projections on non-tangential manifolds," *arXiv:1107.4055*, 2011.

[15] J. Gillard and A. A. Zhigljavsky, "Optimization challenges in the structured low rank approximation problem," *Journal of Global Optimization (to appear)*.

[16] J. W. Gillard and A. A. Zhigljavsky, "Software for alternating projections with backtracking and randomization, *http://www.jonathangillard.co.uk*," 2012.

[17] B. De Moor, "Structured total least squares and L2 approximation problems," *Linear Algebra and its Applications*, vol. 188-189, no. 1036, pp. 163–205, 1993.

[18] P. Comon, X. Luciani, and A. L. F. de Almeida, "Tensor decompositions, alternating least squares and other tales," *Journal of Chemometrics*, vol. 23, pp. 393–405, 2009.