Katholieke Universiteit Leuven

Departement Elektrotechniek

ESAT-SISTA/TR 1994-95

Blind Source Separation by Higher-Order Singular Value Decomposition [1]

Lieven De Lathauwer[2] Bart De Moor[3] Joos Vandewalle

August 1994

Published in:
Proceedings of EUSIPCO-94
VIIth European Signal Processing Conference
Edinburgh, Scotland
September 13-16, 1994
vol. 1, pp. 175-178

[1]This report is available by anonymous ftp from *ftp.esat.kuleuven.ac.be* in the directory *pub/SISTA/delathau/reports/ldl-94-95.ps.Z*

[2]ESAT - Katholieke Universiteit Leuven, Kardinaal Mercierlaan 94, 3001 Leuven (Heverlee), Belgium, tel 32/16/22 09 31, fax 32/16/22 18 55, email: *Lieven.DeLathauwer@esat.kuleuven.ac.be*. Lieven De Lathauwer is a research assistant with the I.W.T. (Flemish Institute for Support of Scientific-Technological Research in Industry). The research reported in this paper was partially supported by the Belgian Programme on Interuniversity Poles of Attraction (IUAP-17, IUAP-50), initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture. It was also supported by the European Community Research program ESPRIT, Basic Research Working Group nr. 6620 (ATHOS). The scientific responsibility rests with its authors.

[3]Bart De Moor is a research associate with the N.F.W.O. (Belgian National Fund for Scientific Research).

Blind Source Separation by Higher-Order Singular Value Decomposition[1]

Lieven DE LATHAUWER, Bart DE MOOR and Joos VANDEWALLE

Group ESAT/SISTA, Dept. of Electronic and Electrical Engineering, K.U.Leuven, Kardinaal Mercierlaan 94, B-3001 Leuven (Heverlee), Belgium, Tel/Fax: +32 [(0)16] 220931 / 221855, E-Mail: Lieven.DeLathauwer@esat.kuleuven.ac.be

Abstract. In this paper, a new tool in multilinear algebra is highlighted: a higher-order generalization of the Singular Value Decomposition. We will show that this decomposition can be used to solve the blind source separation problem in Higher-Order Statistics. The derivation of the algorithm is established under noise-free conditions. It is indicated how to proceed for noisy environments. The approach offers considerable conceptual insight, e.g. it allows a further interpretation of Independent Component Analysis as the higher-order refinement of Principal Component Analysis.

1 Introduction

This paper deals with the problem of blind source separation. This problem can be stated as follows. Consider the linear transfer of a zero-mean stochastic "source vector" X to a zero-mean stochastic "output vector" Y when additive noise N is present:

$$Y = \mathbf{M}X + N \tag{1}$$

The components of X are statistically independent and the matrix $\mathbf{M}$ has linearly independent columns. The goal of blind source separation now consists of the determination of $\mathbf{M}$ and the corresponding realizations of X, given only realizations of Y. We will assume for convenience that the matrix $\mathbf{M}$ is square.

As illustrated in Section 4.1, only the column space of $\mathbf{M}$ can be identified when only second-order statistical information on Y is used and no extra constraints are added. In "Principal Component Analysis" (PCA) the solution is made essentially unique by selecting a matrix with orthonormal columns. To solve the initial problem however, one has to resort to the higher-order statistics of Y.

In recent years a lot of work has been done with respect to blind source separation. The approaches of Comon and Cardoso play a fundamental role here: starting from the observation that the higher-order statistics of a stochastic vector are higher-order tensors, they tackled the problem by the development of multilinear decomposition techniques. These decompositions are not only important for their application in higher-order statistics; they can also be considered as fundamental new tools in the general framework of tensor algebra. (We should motivate at this point the nomenclature adopted in this paper. We will use the term "higher-order array" to denote a higher-order table of numerical values. The term "higher-order tensor" is reserved for higher-order arrays that behave in a particular way under coordinate transformations of an underlying vector space, e.g. the space on which the stochastic vector is defined for the case of higher-order statistics [1].)

In [2] the blind source separation problem is solved in a two-step procedure. In the first step the estimated sources are made uncorrelated by a congruence transformation of the output covariance matrix. In the second step the remaining statistical dependence is minimized by diagonalizing, as far as possible, a higher-order output cumulant. This leads to a higher-order extension of the Jacobi-method for computation of the Eigenvalue Decomposition (EVD) of a Hermitean matrix. Due to the minimization of statistical dependence, the way of analysis is called "Independent Component Analysis" (ICA). [3] follows a two-step procedure too. The second step consists here of a generalized EVD of the fourth-order output cumulant. In [4] it is proved that, although the congruence transformation of a Hermitean matrix is underdetermined and unicity usually obtained by putting orthogonality conditions, the appropriate fourth-order generalization is essentially unique under less stringent conditions of symmetry and linear independence. The resulting "Super-Symmetric Decomposition" offers the possibility to identify more sources than there are sensors available, making use of higher-order statistics only. [5, 6] describe a trade-off for the same idea: the assumption that the number of sources does not exceed the number of sensors leads to an implementation that is more robust towards perturbation of the data.

In this paper another tool in multilinear algebra is highlighted. For the the third-order case, the basics have been developed in the field of Psychometrics [7]. We have investigated the proposed way of data analysis from an algebraic point of view and proved that it yields a generalization of the Singular Value Decomposition (SVD) to the case of higher-order arrays. Many properties, like e.g. the link with the EVD, have already been generalized. They all show a strong analogy between the matrix and the higher-order case [8, 9].

The paper is organized as follows. Section 3 introduces the concept of Higher-Order Singular Value Decomposition (HO SVD). The model is first proposed for the general case of an Nth-order array with complex elements. For reasons of clarity, the furher discussion will be restricted to third-

[1]The research reported in this paper was partially supported by the Belgian Program on Interuniversity Attraction poles (IUAP-17, IUAP-50), the European Community Research program ESPRIT, Basic Research Working Group nr. 6620 (ATHOS), the Belgian Institute for Support of Scientific Research in Industry and Agriculture (I.W.O.N.L.) and the Belgian National Fund for Scientific Research (N.F.W.O.) Lieven De Lathauwer is a Research Assistant supported by the I.W.O.N.L. Bart De Moor is a Research Associate with the N.F.W.O.

order arrays with real elements. We will show the analogy between the third-order decomposition and the classical matrix decomposition, and demonstrate how the HO SVD can be computed. In Section 4 we derive a new algorithm to perform the blind source separation, based on HO SVD. It will bring up an interesting relationship between PCA and ICA.

2 Definitions and notations

2.1 Scalar product, orthogonality, norm of higher-order arrays

Geometrical conditions to be discussed in Section 3, that replace the morphological constraint of diagonality of the matrix of singular values in the second-order case, make it necessary to generalize the well-known definitions of scalar product, orthogonality and Frobenius-norm to arrays of arbitrary order. Consider two Nth-order $(I_1 \times I_2 \times \ldots I_N)$-arrays Φ and Ψ with complex elements, then we have:

Definition 1 *The scalar product $\langle \Phi, \Psi \rangle$ of two arrays $\Phi, \Psi \in \mathbb{C}^{I_1 \times I_2 \times \ldots I_N}$ is defined as*

$$\langle \Phi, \Psi \rangle \stackrel{\mathrm{def}}{=} \sum_{i_1} \sum_{i_2} \ldots \sum_{i_N} \Psi^*_{i_1 i_2 \ldots i_N} \Phi_{i_1 i_2 \ldots i_N} \tag{2}$$

in which $*$ denotes the complex conjugation.

Due to the generalization of the scalar product it becomes possible to treat arrays in a geometrical way. We have e.g.:

Definition 2 *Arrays of which the scalar product equals 0, are mutually orthogonal.*

Definition 3 *The Frobenius-norm of an array Φ is given by*

$$\|\Phi\| \stackrel{\mathrm{def}}{=} \sqrt{\langle \Phi, \Phi \rangle} \tag{3}$$

The Frobenius-norm can be interpreted as the "size" of the array. The square of this norm can be seen as the "energy" in the array.

2.2 Multiplication of a higher-order array by a matrix

Definition 4 *The n-mode product of $\Phi \in \mathbb{C}^{I_1 \times I_2 \times \ldots I_N}$ by a matrix $\mathbf{U} \in \mathbb{C}^{J_n \times I_n}$, denoted by $\Phi \times_n \mathbf{U}$, is an $(I_1 \times I_2 \times \ldots J_n \times \ldots I_N)$-array of which the entries are given by*

$$(\Phi \times_n U)_{i_1 i_2 \ldots j_n \ldots i_N} \stackrel{\mathrm{def}}{=} \sum_{i_n} \Phi_{i_1 i_2 \ldots i_n \ldots i_N} \mathbf{U}_{j_n i_n} \tag{4}$$

The n-mode product of a higher-order array and a matrix is a special case of the inner product in multilinear algebra [10] and, more generally, tensor analysis [11]. In literature it is mostly denoted using the Einstein summation convention, i.e. the summation sign is dropped for the index that is repeated. Especially in the field of tensor analysis this approach is advantageous, since an Einstein summation can be proved to have a basis-independent meaning. For our purpose however, the use of the $\times_n$-symbol will more clearly demonstrate the analogy between matrix and array SVD.

Corollary 1 *Given the array $\Phi \in \mathbb{C}^{I_1 \times I_2 \times \ldots I_N}$ and the matrices $\mathbf{F} \in \mathbb{C}^{J_n \times I_n}$, $\mathbf{G} \in \mathbb{C}^{J_m \times I_m}$ it follows that*

$$(\Phi \times_n \mathbf{F}) \times_m \mathbf{G} = (\Phi \times_m \mathbf{G}) \times_n \mathbf{F} = \Phi \times_n \mathbf{F} \times_m \mathbf{G} \tag{5}$$

Corollary 2 *Given the array $\Phi \in \mathbb{C}^{I_1 \times I_2 \times \ldots I_N}$ and the matrices $\mathbf{F} \in \mathbb{C}^{J_n \times I_n}$, $\mathbf{G} \in \mathbb{C}^{K_n \times J_n}$ it follows that*

$$(\Phi \times_n \mathbf{F}) \times_n \mathbf{G} = \Phi \times_n (\mathbf{G} \cdot \mathbf{F}) \tag{6}$$

Example 1 For the real matrices $\mathbf{F} \in \mathbb{R}^{I_1 \times I_2}$, $\mathbf{U} \in \mathbb{R}^{J_1 \times I_1}$, $\mathbf{V} \in \mathbb{R}^{J_2 \times I_2}$ the matrix product $\mathbf{U} \cdot \mathbf{F} \cdot \mathbf{V}^t$ can be written as $\mathbf{F} \times_1 \mathbf{U} \times_2 \mathbf{V}$.

3 The Higher-Order Singular Value Decomposition

3.1 The HO SVD-model

Theorem 1 *Every complex $(I_1 \times I_2 \times \ldots I_N)$-array Φ can be written as the product*

$$\Phi = \Sigma \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \ldots \times_N \mathbf{U}^{(N)} \tag{7}$$

in which:

- $\mathbf{U}^{(n)} = \left[\underline{u}_1^{(n)} \underline{u}_2^{(n)} \ldots \underline{u}_{I_1}^{(n)} \right]$ *is a complex orthogonal $(I_n \times I_n)$-matrix*
- *the core array Σ is a complex $(I_1 \times I_2 \times \ldots I_N)$-array of which the subarrays $\Sigma_{i_n=\alpha}$, obtained by fixing the nth index to α, have the properties of:*
  - *all-orthogonality: two subarrays $\Sigma_{i_n=\alpha}$ and $\Sigma_{i_n=\beta}$ are orthogonal for all possible values of n, α and β subject to $\alpha \neq \beta$:*

  $$\langle \Sigma_{i_n=\alpha}, \Sigma_{i_n=\beta} \rangle = 0 \quad \text{when} \quad \alpha \neq \beta \tag{8}$$

  - *ordering:*

  $$\|\Sigma_{i_n=1}\| \geq \|\Sigma_{i_n=2}\| \geq \ldots \geq \|\Sigma_{i_n=r_n}\| > 0 \tag{9}$$

  and

  $$\|\Sigma_{i_n=r_n+1}\| = \ldots = \|\Sigma_{i_n=I_N}\| = 0 \tag{10}$$

  for all possible values of n.

The Frobenius-norms $\|\Sigma_{i_n=i}\|$, symbolized by $\sigma_i^{(n)}$, are the n-mode singular values of Φ and the vector $\underline{u}_i^{(n)}$ is the ith n-mode singular vector.

Proof: see [9]. ∎

3.2 Interpretation

For clarity the interpretation as well as the further discussion will be restricted to third-order arrays with real elements. So consider $\Phi \in \mathbb{R}^{I \times J \times K}$ and assume its HO SVD is given by:

$$\Phi = \Sigma \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W} \tag{11}$$

in which the real matrices $\mathbf{U}, \mathbf{V}, \mathbf{W}$ are orthogonal and the core array Σ is real, all-orthogonal and ordered. This decomposition is visualized in Figure 1.

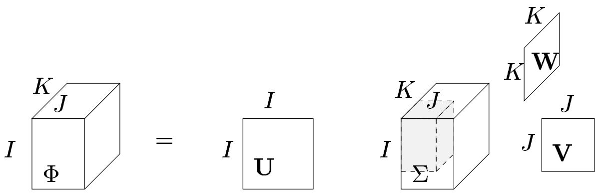

Figure 1. Visualization of the HO SVD for a third-order array.

Eq. (11) should be compared to the expression for the SVD of a real $(I \times J)$-matrix $\mathbf{F}$, which in our notation reads:

$$\mathbf{F} = \mathbf{S} \times_1 \mathbf{U} \times_2 \mathbf{V} \tag{12}$$

in which the matrices $\mathbf{U}, \mathbf{V}$ are orthogonal and the "core matrix" $\mathbf{S}$ is diagonal and contains r strict positive elements, put in non-increasing order (see also Figure 2).

Clearly eq. (11) is a formal generalization of eq. (12). Moreover, it can be proved that the HO SVD of a second-order array boils down to its matrix SVD [9].

3.3 Calculation

Reorganisation of eq. (11) in a matrix format shows that the matrices $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$ can be calculated as the left singular matrices of the $(I \times JK)$, $(K \times IJ)$ and $(J \times KI)$ matrix unfoldings of Φ, defined in accordance with Figure 3.

The core array is obtained by bringing the matrices in eq. (11) to the other side:

$$\Sigma = \Phi \times_1 \mathbf{U}^t \times_2 \mathbf{V}^t \times_3 \mathbf{W}^t \tag{13}$$

The way of calculation and the ordening constraint on the core array show that the HO SVD obeys analog unicity properties as its matrix equivalent: in the generic case, the singular vectors are determined up to the sign. When the sign of a singular vector is changed, the sign of the corresponding subarray in Σ alters too.

4 Application to blind source separation

We consider the noise-free version of eq. (1):

$$Y = \mathbf{M}X \tag{14}$$

The separation problem will be solved by factorisation of the transfer matrix:

$$\mathbf{M} = \mathbf{TQ} \tag{15}$$

in which $\mathbf{T}$ is regular and $\mathbf{Q}$ is orthogonal.

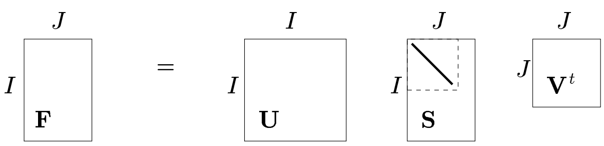

Figure 2. Visualization of the matrix SVD.

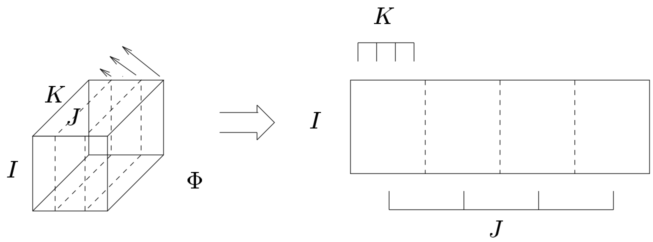

Figure 3. Unfolding of the $(I \times J \times K)$-array Φ to an $(I \times JK)$-matrix.

In a first step $\mathbf{T}$ will be determined from the second-order statistics of the output Y. The resulting degree of freedom, the orthogonal factor $\mathbf{Q}$, is recovered from the higher-order statistics of Y.

4.1 Step 1: determination of $\mathbf{T}$ *from the second-order statistics of* Y

The covariance $\mathbf{C}_2^Y$ of Y is given by

$$\mathbf{C}_2^Y = \mathbf{M}\mathbf{C}_2^X\mathbf{M}^t \tag{16}$$

in which the covariance $\mathbf{C}_2^X$ of X is diagonal, since we claim that the source signals are uncorrelated. Assuming that the source signals have unity variance, we get:

$$\mathbf{C}_2^Y = \mathbf{M}\mathbf{M}^t \tag{17}$$

This assumption means just a scaling of the columns of $\mathbf{M}$ and is not detrimental to the method's generality: it is clear that $\mathbf{M}$ can at most be determined up to a scaling and a permutation of its columns.

We can conclude from eq. (17) that $\mathbf{M}$ can be determined, up to an orthogonal factor $\mathbf{Q}$, from a congruence transformation of $\mathbf{C}_2^Y$:

$$\mathbf{C}_2^Y = \mathbf{M}\mathbf{M}^t = (\mathbf{TQ})(\mathbf{TQ})^t = \mathbf{T}\mathbf{T}^t \tag{18}$$

One alternative is the computation, like in PCA, of the EVD of $\mathbf{C}_2^Y$:

$$\mathbf{C}_2^Y = \mathbf{E}\mathbf{D}^2\mathbf{E}^t = (\mathbf{ED})(\mathbf{ED})^t \tag{19}$$

When the output covariance is estimated following $\mathbf{C}_2^Y = \mathbf{A}_Y\mathbf{A}_Y^t$ in which $\mathbf{A}_Y$ is an $(I \times N)$-dimensional dataset containing N realizations of Y, then the factor $(\mathbf{ED})$ can be obtained in a numerically more reliable way from the SVD of $\mathbf{A}_Y$ [12].

4.2 Step 2: determination of $\mathbf{Q}$ *from the higher-order statistics of* Y

The third-order cumulant of Y is given by

$$\mathbf{C}_3^Y = \mathbf{C}_3^X \times_1 \mathbf{M} \times_2 \mathbf{M} \times_3 \mathbf{M} \tag{20}$$

in which the third-order cumulant $\mathbf{C}_3^X$ of X is diagonal, since we claim that the source signals are also higher-order independent. Substitution of eq. (15) in eq. (20) yields

$$\Phi = \mathbf{C}_3^X \times_1 \mathbf{Q} \times_2 \mathbf{Q} \times_3 \mathbf{Q} \tag{21}$$

in which the tensor Φ is defined as:

$$\Phi \stackrel{\text{def}}{=} \mathbf{C}_3^Y \times_1 \mathbf{T}^{-1} \times_2 \mathbf{T}^{-1} \times_3 \mathbf{T}^{-1} \tag{22}$$

Hence, due to the unicity property in Section 3.3, $\mathbf{Q}$ can be obtained from the HO SVD of Φ. (Eq. (21) is the third-order equivalent of the EVD of symmetric matrices.)

The transfer matrix $\mathbf{M}$ is now given by eq. (15). The $(I \times N)$-matrix $\mathbf{A}_X$, containing the corresponding N realizations of X, is then obtained from the set of linear equations:

$$\mathbf{M}\mathbf{A}_X = \mathbf{A}_Y \tag{23}$$

5 Discussion and conclusions

- We generalized the Singular Value Decomposition of matrices to the higher-order case. It was shown that this decomposition provides a way to solve the blind source separation problem.
- We want to stress the conceptual importance of the new approach. It reveals an important symmetry when considering the problems of PCA and ICA. In "classical" *second*-order statistics, the problem of interest is to remove the correlation from data measured after linear transfer of independent source signals. The key tool to realize this, comes from "classical" *linear* algebra: it is the matrix SVD. More recently, researchers also aimed at the removal of higher-order dependence, which is a problem of *Higher*-Order Statistics. We proved that one can resort to a tool from *multilinear* algebra, which is precisely the generalization of the SVD for higher-order tensors.
- One could think of incorporating the symmetry properties of Φ in existing algorithms, when computing the left singular matrix from the matrix unfolding of Φ. This would lead to a speed-up.
- For the application in the presence of noise, we make the distinction between the influence of Gaussian and non-Gaussian noise.

  Additive Gaussian noise in eq. (1) doesn't affect the higher-order cumulant of Y. Hence its effect can be neutralized by replacing $\mathbf{C}_2^Y$ in Section 4.1 by the noise-free covariance $\mathbf{C}_2^Y - \sigma^2\mathbf{I}$, in which σ^2 is the noise variance on each data channel and $\mathbf{I}$ is the identity matrix. In a more-sensors-than-sources setup, σ^2 can be estimated as the mean of the "noise-eigenvalues" of $\mathbf{C}_2^Y$.

  The presence of non-Gaussian noise additionally makes that the third-order cumulant of Y can no longer be diagonalized, but the HO SVD will still yield three equal matrices of singular vectors, and the core tensor will be all-orthogonal and symmetric (invariant under permutation of its indices). The perturbation of the singular vectors and the core tensor will for sufficiently high signal-to-noise ratios be in the same order of magnitude as the perturbation of the third-order cumulant. This is due to the fact that the Unordered-Unsigned SVD (USVD) is analytic in the variation of parameters [13, 14]. In [9] explicit perturbation expressions have been derived for the HO SVD. One could possibly think of an extra optimization step (like [2]), using the perturbation results of the HO SVD as an analytic lower-bound of performance.

References

[1] **P. McCullagh (1987):** Tensor Methods in Statistics, *Chapman and Hall.*

[2] **P. Comon (1994):** Independent Component Analysis, A New Concept?, *Signal Processing* **36***(3), Special Issue on Higher Order Statistics*, 287-314.

[3] **J.-F. Cardoso and P. Comon (1990):** Tensor-Based Independent Component Analysis, *Signal Processing V: Theories and Applications*, 673-676.

[4] **J.-F. Cardoso (1991):** Super-Symmetric Decomposition of the Fourth-Order Cumulant Tensor. Blind Identification of More Sources than Sensors, *Proc. Int. Conf. ICASSP-91, vol V, 3109-3112.*

[5] **J.-F. Cardoso (1992):** Fourth-Order Cumulant Structure Forcing. Application to Blind Array Processing, *Proc. Int. IEEE Workshop on SSAP-92, 136-139.*

[6] **J.-F. Cardoso (1994):** A tetradic decomposition of 4th-order tensors. Application to the source separation problem. *Proc. Int. Workshop on SVD and Signal Processing*, Leuven, Belgium, August 1994.

[7] **L.R. Tucker (1964):** The extension of factor analysis to three-dimensional matrices, *H. Gulliksen and N. Frederiksen (Eds): Contributions to mathematical psychology, Holt, Rinehart & Winston*, 109-127.

[8] **L. De Lathauwer, B. De Moor and J. Vandewalle (1993):** A Singular Value Decomposition for Higher-Order Tensors, *Proc. Int. ATHOS Workshop on System Identification and High-Order Statistics*, Nice, France, September 1993.

[9] **L. De Lathauwer, B. De Moor and J. Vandewalle (1994):** The Higher-Order Singular Value Decomposition, *internal report 94-31, group ESAT/SISTA, K.U. Leuven.*

[10] **M. Marcus (1975):** Finite Dimensional Multilinear Algebra, *Dekker N.Y.*

[11] **D.C. Kay (1988):** Theory and Problems of Tensor Calculus, *McGraw-Hill.*

[12] **G.H. Golub and C.F. Van Loan (1991):** Matrix computations, *Johns Hopkins University Press.*

[13] **R.J. Vaccaro, D.W. Tufts and G.F. Boudreaux-Bartels (1988):** Advances in Principal Component Signal Processing, *E.F. Deprettere (Ed): SVD and Signal Processing, Algorithms, Applications and Architectures, Elsevier Science Publishers B.V.*, 115-146.

[14] **B. De Moor and S. Boyd (1989):** Analytic properties of singular values and vectors, *internal report 89-28, group ESAT/SISTA, K.U. Leuven.*

[15] **J. Vandewalle and D. Callaerts (1990):** Singular Value Decomposition: a Powerful Concept and Tool in Signal Processing, *Mathematics in Signal Processing* **2**, 539-559.

[16] **J. Vandewalle and B. De Moor (1991):** On the Use of the Singular Value Decomposition in Identification and Signal Processing, *NATO ASI Series: Numerical Linear Algebra, Digital Signal Processing and Parallel Algorithms, vol F70*, 321-360.