Canonical Filters and Rank-Reduction Algorithms

P. C. Hansen* P. S. K. Hansen* S. D. Hansen*† J. Aa. Sørensen*

August 18, 2000

ABSTRACT

Rank reduction is a common noise-reduction technique in signal processing. We analyze a class of rank-reduction algorithms based on orthogonal projection on certain subspaces, and show that the properties of these algorithms can be compared by means of FIR filters defined by the canonical vectors associated with the projections. We use our new analysis to demonstrate that ULV decompositions work well in connection with speech signals, also in the absence of a gap in the singular values (which is usually assumed to be present).

SP EDICS: SP 3.8.2 Algorithm-Specific Performance Analysis

*Department of Mathematical Modelling, Technical University of Denmark, Building 321, DK-2800 Lyngby, Denmark (pch,pskh,jaas@imm.dtu.dk).

†Steffen Duus Hansen (1936–1999).

I. Introduction

Rank reduction algorithms have proved themselves useful for noise-reduction and signal identification in a number of applications. The central idea is to approximate a matrix, derived from the data, with another matrix of lower rank from which the reconstructed signal is derived. As stated in [8],

> *Rank reduction is a general principle for finding the right trade-off between model bias and model variance when reconstructing signals from noisy data.*

The classical way to implement rank reduction is via the singular value decomposition (SVD) [4], which is the most reliable method for computing the numerical rank of a matrix. In the last decade, alternatives to the SVD have emerged, most notably the rank-revealing QR decomposition and the URV and ULV decompositions; see [5] for a survey and [1], [9], [10] for updating issues. The latter two decompositions are collectively known as UTV decompositions, and their main advantage to the SVD is that they can be computed and updated more efficiently.

In order to guarantee that the subspaces computed by the UTV decompositions are close to those defined in terms of the SVD it is necessary to assume a distinct gap in the singular values [2]. In practise such a gap rarely exists, and yet the noise reduction achieved by the UTV algorithms is comparable to the SVD-based noise reduction. Clearly, the noise-reduction performance of the algorithms is controlled by other quantities than the singular value spectrum.

The purpose of this paper is to provide a general scheme for analysis of rank-reduction algorithms, by means of which we can compare the methods and their noise-reducing capabilities. Our main tool is the notion of canonical angles and vectors, defined in terms of the signal subspaces computed by the rank-revealing algorithms. We show that the canonical vectors define FIR filter arrays similar to the filter-array representation introduced in [6], and we demonstrate that these canonical filters provide a natural framework for comparison of the algorithms. We also illustrate by numerical examples that the similarity of two rank-reduction algorithms is equivalent to their canonical FIR filters being similar in the frequency domain.

This work can be viewed as a continuation of [6], in the sense that we continue to explore the filter array interpretation of rank-reduction algorithms. The main contribution in this paper is to show how canonical angles and vectors along with the filter-array provide a natural framework for comparing the performance of rank-reduction algorithms.

Our paper is organized as follows. In Section II we introduce a wide class of rank-revealing decompositions now used in signal processing, and in Section III we discuss the filter array interpretation of these algorithms. The canonical filters are defined in Section IV, and we conclude with numerical examples in Section V.

II. Rank Reduction by Truncated Decompositions

Our point of departure is the $m \times n$ Hankel matrix A defined in terms of the real signal vector $x = (x_1, \ldots, x_{m+n-1})^T$ as

$$A = \mathcal{H}(x) = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ x_2 & x_3 & \cdots & x_{n+1} \\ x_3 & x_4 & \cdots & x_{n+2} \\ \vdots & \vdots & & \vdots \\ x_m & x_{m+1} & \cdots & x_{m+n-1} \end{pmatrix}, \tag{1}$$

where we use the notation $\mathcal{H}(x)$ from [6]. It is natural to consider x as the input signal to the noise-reduction algorithm. The next step is to compute a rank-revealing decomposition of A, and we assume that the resulting rank-k matrix A_k can be written as

$$A_k = A\, V_k V_k^T + \Delta \tag{2}$$

where k is the rank that we have chosen, V_k is a matrix with orthonormal columns (such that $V_k^T V_k = I_k$, the identity matrix of order k) coming from the rank-revealing decomposition, and Δ is a residual matrix with small norm (possibly zero) also determined by the decomposition. Note that A_k is not a Hankel matrix. The output from the algorithm is the vector s computed by averaging along all the $m - n + 1$ full-length antidiagonals of A_k, which we write in the short form

$$s = \mathcal{A}(A_k) \tag{3}$$

where s is a vector of length $m - n + 1$, and the symbol $\mathcal{A}$ denotes the averaging operation. The details behind this result are derived in [6], but the notation here is slightly different because our output vector s has length $m - n + 1$ to avoid dealing with end effects.

To substantiate Eq. (2) we consider some specific rank-revealing decompositions. The SVD takes the form

$$A = \begin{pmatrix} U_k & U_0 \end{pmatrix} \begin{pmatrix} \Sigma_k & 0 \\ 0 & \Sigma_0 \end{pmatrix} \begin{pmatrix} V_k^T \\ V_0^T \end{pmatrix} \tag{4}$$

and the truncated SVD matrix is given by $A_k = U_k \Sigma_k V_k^T = A\, V_k V_k^T$ which is clearly of the form (2) with $\Delta = 0$. The URV decomposition is given by

$$A = \begin{pmatrix} \widetilde{U}_k & \widetilde{U}_0 \end{pmatrix} \begin{pmatrix} R_k & E \\ 0 & F \end{pmatrix} \begin{pmatrix} \widetilde{V}_k^T \\ \widetilde{V}_0^T \end{pmatrix} \tag{5}$$

and the truncated URV matrix is by definition $\widetilde{A}_k = \widetilde{U}_k R_k \widetilde{V}_k^T = A\, \widetilde{V}_k \widetilde{V}_k^T$, again of the form (2) with $\Delta = 0$. Similarly, the ULV decomposition is

$$A = \begin{pmatrix} \overline{U}_k & \overline{U}_0 \end{pmatrix} \begin{pmatrix} L_k & 0 \\ G & H \end{pmatrix} \begin{pmatrix} \overline{V}_k^T \\ \overline{V}_0^T \end{pmatrix} \tag{6}$$

and the truncated ULV matrix is by definition $\overline{A}_k = \overline{U}_k L_k \overline{V}_k^T = A\, \overline{V}_k \overline{V}_k^T - \overline{U}_0\, G\, \overline{V}_k^T$. Hence, $\overline{A}_k$ is of the general form (2) with $\Delta = -\overline{U}_0\, G\, \overline{V}_k^T$. Finally, the VSV decomposition (recently introduced in [7]) of a symmetric matrix A takes the form

$$A = \begin{pmatrix} \breve{V}_k & \breve{V}_0 \end{pmatrix} \begin{pmatrix} S_k & K^T \\ K & M \end{pmatrix} \begin{pmatrix} \breve{V}_k^T \\ \breve{V}_0^T \end{pmatrix} . \tag{7}$$

The corresponding truncated VSV matrix is $\breve{A}_k = \breve{V}_k S_k \breve{V}_k^T = A\, \breve{V}_k \breve{V}_k^T - \breve{V}_0\, K\, \breve{V}_k^T$, which also fits into the form (2) with $\Delta = -\breve{V}_0\, K\, \breve{V}_k^T$.

Due to the way that the ULV and VSV decompositions are defined, the submatrices G and K have small norm. The actual size of the norm depends on the singular vector estimation used in the particular implementation, and the norm can be made as small as desired by refining the singular vector estimates, at the cost of more work. Hence the norm of the residual matrix Δ can be assumed to be small compared to the norm of $\overline{A}_k$ and $\breve{A}_k$.

III. FIR Filter Representations

We shall now consider a general FIR filter interpretation of the expression for the output signal

$$s = \mathcal{A}(A\,V_k V_k^T) + \mathcal{A}(\Delta)$$

which is obtained by combining (2) and (3). The FIR filter representation of the truncated SVD approach was derived in [6], and using the same approach it is straightforward to show that if

$$V_k = (v_1, \ldots, v_k)$$

then

$$\mathcal{A}(A\,V_k V_k^T) = n^{-1} \sum_{i=1}^{k} \mathcal{H}(Av_i) J v_i$$

where $\mathcal{H}(Av_i)$ is the Hankel matrix defined from the vector Av_i, cf. (1), and J is the $n \times n$ exchange matrix consisting of the columns of the identity matrix in reverse order. Thus the output signal s consists of a sum of k signals $n^{-1}s_i$ plus a residual signal $r = \mathcal{A}(\Delta)$,

$$s = n^{-1}(s_1 + \cdots + s_k) + r,$$

where the signal vectors s_i are given by

$$s_i = \mathcal{H}(Av_i)\, J v_i, \quad i = 1, \ldots, k. \tag{8}$$

This shows that s_i is obtained by passing the input signal x through a pair of FIR filters with filter coefficients v_i and Jv_i.

The filter pair v_i and Jv_i in (8) corresponds to a single FIR filter with $2n-1$ coefficients given by the vector

$$c_i = v_i * Jv_i = \mathcal{H}(v_i)\, v_i, \tag{9}$$

where $*$ denotes convolution. Since v_i is real, the frequency response of this filter is given by

$$\widehat{c}_i(f) = \widehat{v}_i(f)\, \mathrm{conj}\,(\widehat{v}_i(f)) = |\widehat{v}_i(f)|^2 \tag{10}$$

where $\widehat{\ }$ denotes Fourier transform, showing that c_i defines a zero-phase filter.

The individual contributions to the output signal s can now be judged by means of the following result.

Theorem 1 *The 2-norm of each vector s_i in (8) is bounded above by*

$$\|s_i\|_2 \le n^{1/2}\, \|Av_i\|_2 \tag{11}$$

while the 2-norm of r is bounded above by

$$\|r\|_2 \le m^{1/2}\, \|\Delta\|_2. \tag{12}$$

Proof. Since $\mathcal{H}(Av_i) = \|Av_i\|_2 \mathcal{H}(z_i)$ with $z_i = Av_i \|Av_i\|_2^{-1}$ it follows from (8) that

$$\|s_i\|_2 \le \|Av_i\|_2\, \|\mathcal{H}(z_i)\|_2\, \|Jv_i\|_2$$

where $\|Jv_i\|_2 = 1$. Moreover,

$$\|\mathcal{H}(z_i)\|_2 \le \|\mathcal{H}(z_i)\|_{\mathrm{F}} = n^{1/2}\, \|z_i\|_2 = n^{1/2},$$

and we obtain the upper bound in (11). Now let $\Delta = \{\delta_{ij}\}$. Due to the averaging, each element r_i of $r = \mathcal{A}(\Delta)$ satisfies $|r_i| \le \max_{ij} |\delta_{ij}|$ and therefore

$$\|r\|_2 \le m^{1/2} \max_{ij} |\delta_{ij}| \le m^{1/2}\, \|\Delta\|_2$$

which is (12). $\square$

Unfortunately we have not been able to derive a rigorous lower bound for $\|s_i\|_2$, but we can obtain some insight via the power of the two signals Av_i and s_i. If $\Gamma_x(f)$ denotes the power density spectrum of x then it follows from (8) and (10) that

$$\|A\,v_i\|_2^2 = \int_{-\infty}^{\infty} |\widehat{v}_i(f)|^2 \; \Gamma_x(f)\, df$$

and

$$\|s_i\|_2^2 = \int_{-\infty}^{\infty} |\widehat{v}_i(f)|^4 \; \Gamma_x(f)\, df.$$

This shows that the combined filter c_i in (9) represents a filter similar to v_i, only sharper, and it is therefore reasonable to assume that $\|s_i\|_2$ is of the same order of magnitude as $\|Av_i\|_2$. In particular, for the truncated SVD algorithm we have $\|Av_i\|_2 = \sigma_i$, and the UTV and VSV algorithms are designed such that $\|Av_i\|_2$ approximates σ_i while the norm of the nonzero off-diagonal block is small.

When Δ is nonzero, its 2-norm is equal to that of the off-diagonal block G in (6) or K in (7), and therefore the residual signal r makes only a minor contribution to the output signal s. In the following we will therefore neglect the residual vector r in our analysis.

IV. Canonical Vectors and Filters

Although the rank-k matrices in the above algorithms are defined in terms of the submatrices of the particular decompositions, the matrix $A\,V_k V_k^T$ in (2) is independent of the choice of the columns $v_1, \ldots, v_k$ of the matrix V_k, as long as they are orthonormal and span the same subspace. To see this, let the columns of the matrix

$$W_k = V_k\, Q \qquad (13)$$

be a second set of vectors, where Q is $k \times k$ and orthogonal, and we see that $W_k W_k^T = V_k Q\, Q^T V_k^T = V_k V_k^T$. Another way to state this is to observe that $V_k V_k^T$ is an orthogonal projection matrix.

This fact allows us — for each rank-reduction algorithm — to choose a new set of vectors $w_1, \ldots, w_k$ that may better describe the output signal s than the vectors $v_1, \ldots, v_k$, knowing that s stays the same. And since these vectors define FIR filter coefficients in a filter array interpretation, this means that we are free to choose the filters as long as (13) is satisfied.

In particular, if we want to compare the output of two rank-reduction algorithms, then we can try to choose the vectors $w_1, \ldots, w_k$ for the two algorithms such that they are as similar as possible.

The solution to this problem of choosing the vectors comes in the form of the *canonical vectors* associated with the subspaces spanned by the columns of the V_k-matrices for the two algorithms. To illustrate this, let us compare the truncated SVD and ULV algorithms, in which we work with the two matrices V_k and $\overline{V}_k$ (and ignore the ULV residual vector r), and we let $\mathcal{V}_k$ and $\overline{\mathcal{V}}_k$ denote the subspaces spanned by the columns of these two matrices.

Definition 2 *Given two $n \times k$ matrices V_k and $\overline{V}_k$ with orthonormal columns. If*

$$V_k^T \overline{V}_k = Y\, \Theta\, Z^T \tag{14}$$

is the SVD of the cross-product matrix, then the canonical vectors are the columns of

$$W_k = V_k Y \quad \text{and} \quad \overline{W}_k = \overline{V}_k Z. \tag{15}$$

The singular values appearing in Θ are termed the canonical correlations, and they are equal *to the cosines of the canonical angles* $\theta_1, \dots, \theta_k$*. I.e.,*

$$\Theta = \text{diag}\,(\cos(\theta_1), \dots, \cos(\theta_k)) \tag{16}$$

with $0 \le \theta_1 \le \theta_2 \le \cdots \le \theta_k$.

We emphasize the following geometric interpretation of the canonical angles and vectors. The smallest canonical angle θ_1 is the smallest angle between any two vectors v and $\overline{v}$ in $\mathcal{V}_k$ and $\overline{\mathcal{V}}_k$, respectively, and it is attained for $v = w_1$ and $\overline{v} = \overline{w}_1$. Then the second canonical angle θ_2 is the smallest angle between any two vectors v and $\overline{v}$ orthogonal to w_1 and $\overline{w}_1$ in $\mathcal{V}_k$ and $\overline{\mathcal{V}}_k$, and it is attained for $v = w_2$ and $\overline{v} = \overline{w}_2$, etc.

Hence, canonical vectors associated with small canonical angles define subspaces of $\mathcal{V}_k$ and $\overline{\mathcal{V}}_k$ that are very similar, and zero canonical angles define canonical vectors in the intersection of the subspaces $\mathcal{V}_k$ and $\overline{\mathcal{V}}_k$. Zero canonical angles are always present when k is greater than $n/2$, for geometric reasons.

Theorem 3 *If* $2k > n$ *then*

$$\theta_1 = \cdots = \theta_{2k-n} = 0. \tag{17}$$

Proof. Both $\mathcal{V}_k$ and $\overline{\mathcal{V}}_k$ are k-dimensional subspaces of an n-dimensional space. Hence, if $2k > n$ then $\mathcal{V}_k$ and $\overline{\mathcal{V}}_k$ must have a nontrivial intersection of dimension $2k - n$. □

We can now compare the truncated SVD and ULV algorithms by comparing the *canonical FIR filters* determined by the canonical vectors $w_1, \dots, w_k$ and $\overline{w}_1, \dots, \overline{w}_k$. If $k > n/2$ then we are sure that $2k - n$ of these filters are identical, and if some of the nonzero canonical angles are small then the associated filters are also guaranteed to be similar.

Thus, small (and zero) canonical angles define FIR filters for the two algorithms that extract very similar (and identical) signal components.

Of course, there is more to this analysis than merely the canonical angles. Even if θ_i is quite large, meaning that the vectors w_i and $\overline{w}_i$ are quite different in the 2-norm, the associated filters may have similar properties in the frequency domain. For example, w_i and $\overline{w}_i$ may both represent band-pass filters with approximately the same center frequency and bandwidth.

Hence, it is the size of the canonical angles θ_i together with the frequency responses of the canonical FIR filters represented by w_i and $\overline{w}_i$ that provides a convenient tool for comparison of the similarities and differences in the output signals from the two algorithms characterized by V_k and $\overline{V}_k$.

The comparison technique can be brought even further. We can use the same approach to compare a noisy, filtered signal with a clean reference signal, such as

1. the output from applying the truncated ULV algorithm – or any other rank-reduction algorithm – to a noisy signal, and
2. a reference signal, obtained by applying truncated SVD to the noise-free signal, and thus providing approximations to the signal's eigenfilters.

We illustrate these issues in the next section.

V. Numerical Example

We conclude with a numerical example where we compare the output from the truncated SVD and ULV algorithms. All computations were done in Matlab using the UTV Tools

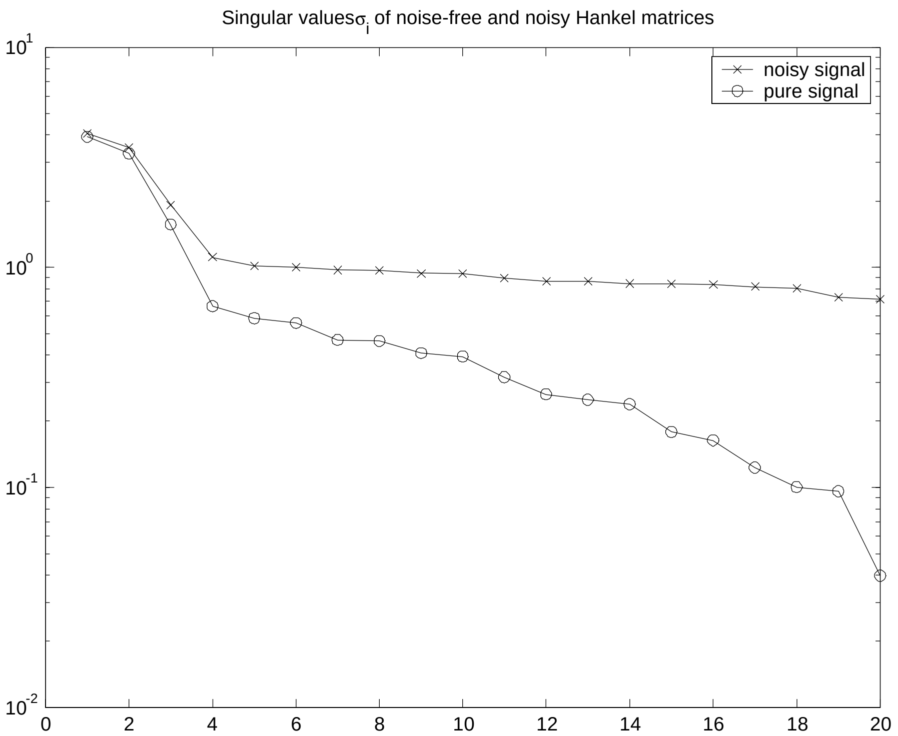

Figure 1: Singular values σ_i of the Hankel matrix $A = \mathcal{H}(x)$ corresponding to a pure and a noisy signal vector x.

package [3]. The signal x consists of 500 samples of a voiced speech signal sampled at 8 kHz, the number of columns in the Hankel matrix is $n = 20$, and the truncation parameter is $k = 9$.

The signal-to-noise ratio is 3 dB, and the norm of the input signal is $\|x\|_2 = 1.27$. The singular values of the two Hankel matrices corresponding to the pure and noisy signals are shown in Fig. 1, and we see that apart from the transition from σ_2 to σ_4 there is no particular gap. The influence of the noise is clearly visible as a plateau in the singular values for the noisy signal.

First we used the lulv implementation of the ULV decomposition, which seeks to yield

SVD and ULV filters

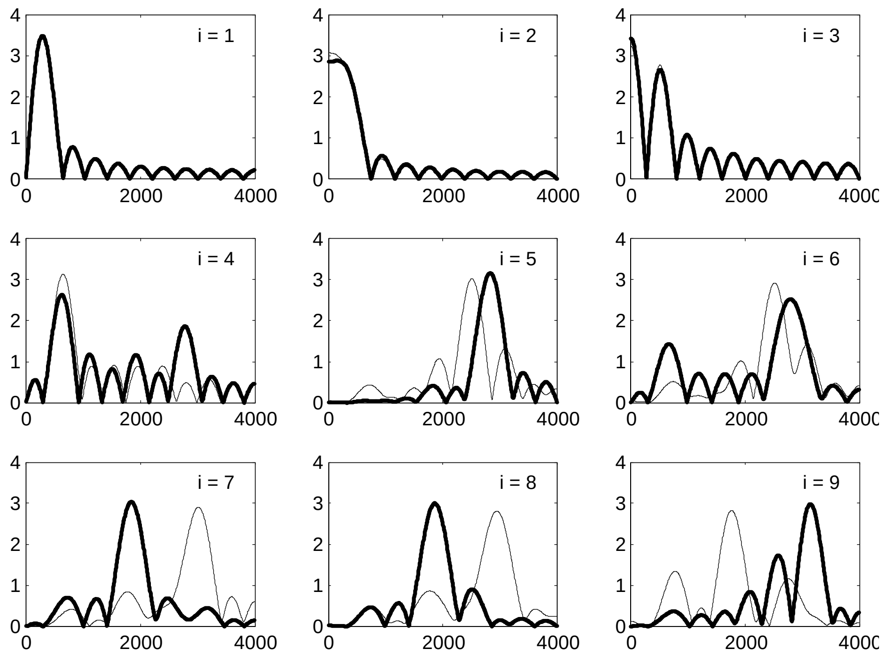

Figure 2: Frequency responses $|\widehat{v}_i(f)|$ and $|\widehat{\overline{v}}_i(f)|$ of the FIR filters for the truncated SVD algorithm (thick lines) and the truncated ULV algorithm (thin lines), defined via the vectors v_i and $\overline{v}_i$ for $i = 1, \ldots, 9$.

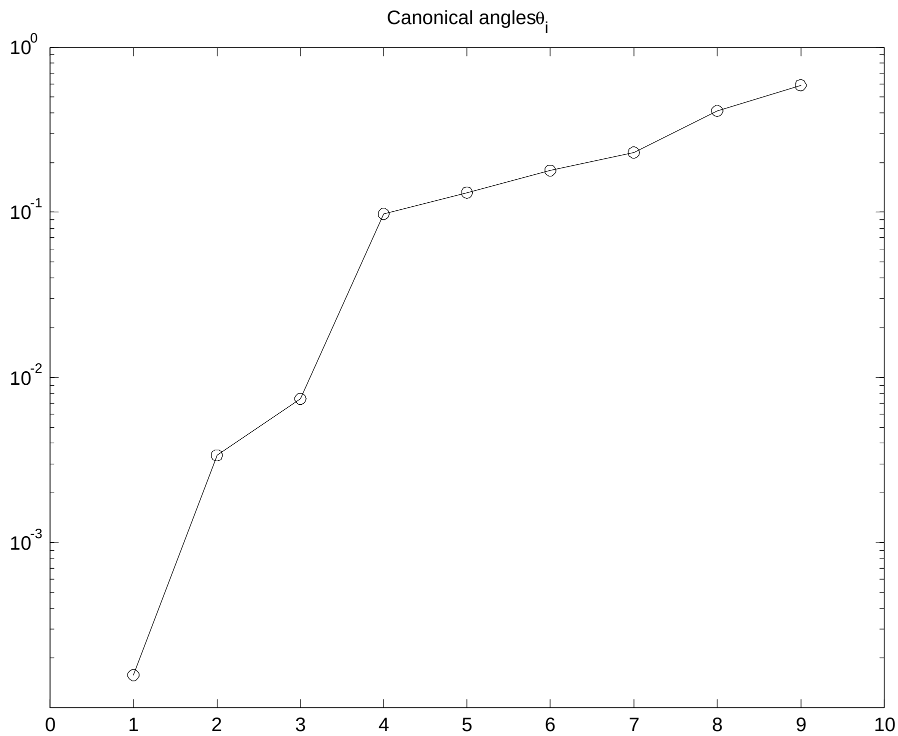

Figure 3: The canonical angles $\theta_1, \ldots \theta_9$ for the two 9-dimensional subspaces $\mathcal{V}_k$ and $\overline{\mathcal{V}}_k$ from the truncated SVD and ULV algorithms.

good approximations to the principal singular vectors. The FIR filters for the SVD and ULV algorithms, defined respectively by the vectors v_i and $\overline{v}_i$, $i = 1, \ldots, 9$, are shown in Fig. 2. The first three filters correspond to the three singular values lying above the noise level, and as expected these filters are very similar. The remaining six filters appear to be different. From Fig. 2 we might therefore immediately think that only a signal component lying in a three-dimensional subspace is recovered similarly by the two algorithms.

However, the analysis using canonical angles and vectors gives a different and more precise picture. Figure 3 shows the the 9 canonical angles, i.e., $\theta_1, \ldots, \theta_9$. Three canonical angles, corresponding to the above-mentioned subspace, are less than 10^{-2}. But there are

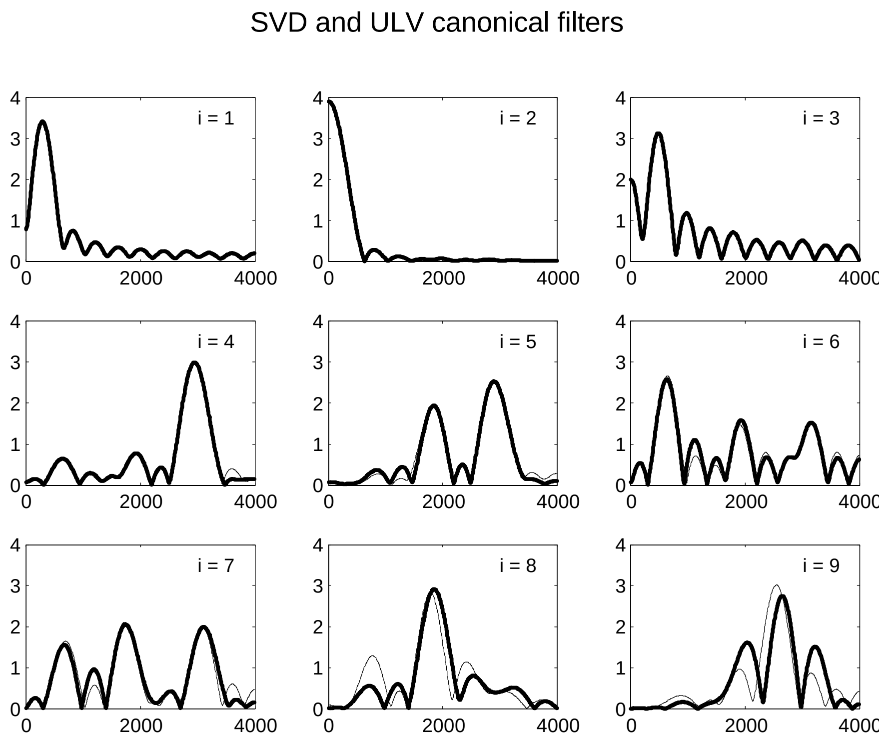

Figure 4: Frequency responses $|\widehat{w}_i(f)|$ and $|\widehat{\overline{w}}_i(f)|$ of the canonical FIR filters for the truncated SVD and ULV algorithms. Thick lines: SVD filters; thin lines: ULV filters.

also several canonical angles of the order 10^{-1}, and we would expect that the corresponding canonical filters are quite similar.

This is confirmed by the plots of the SVD-ULV canonical FIR filters shown in Fig. 4, where we see that actually the first seven canonical filters are very similar. The eights and ninth pair of filters have peaks abound 1.8 kHz and 2.6 kHz, respectively, but their sidelobes are different.

We conclude that for this particular noisy signal, the SVD and ULV algorithms produce filtered signals that have very similar signal components lying in a 7-dimensional subspace of the 9-dimensional output signal subspace. This is in spite of the fact that there is no gap

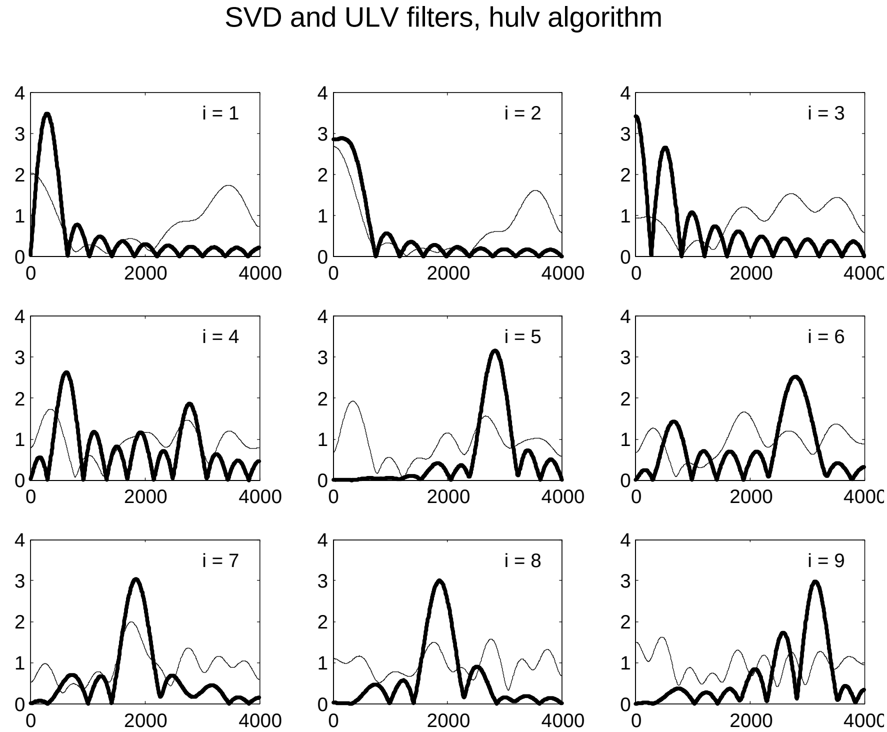

Figure 5: Frequency responses $|\widehat{v}_i(f)|$ and $|\widehat{\overline{v}}_i(f)|$ of the FIR filters for the SVD algorithm (thick lines) and the hulv-version of the ULV algorithm (thin lines).

whatsoever in the singular values around $i = 7$.

To elaborate on this point, we also used the hulv implementation of the ULV decomposition from [3]. This algorithm seeks to compute good approximations to the singular vectors corresponding to the smallest singular values, and we cannot expect that the principal singular vectors are approximated so well. This is confirmed by the SVD and ULV filters shown in Fig. 5: none of these filters are similar.

But still the SVD and ULV algorithms produce signals that sound qualitatively the same, and the canonical FIR filters shown in Fig. 6 support this. We see that the first five canonical filters are very similar, the next two filters are qualitatively similar, and the last two filters

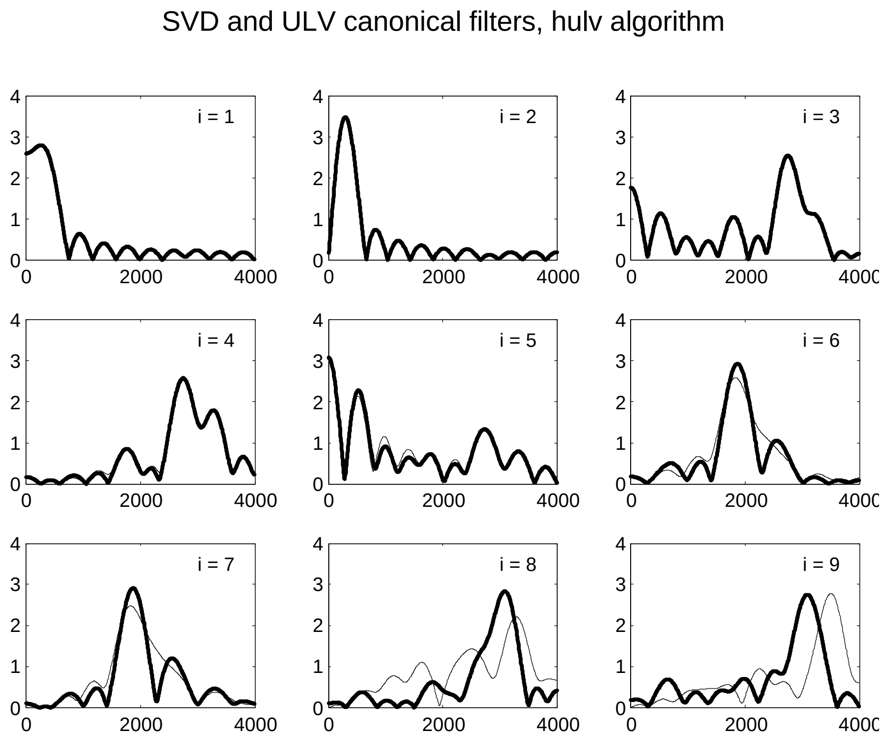

Figure 6: Frequency responses $|\widehat{w}_i(f)|$ and $|\widehat{\overline{w}}_i(f)|$ of the canonical filters for the SVD algorithm (thick lines) and the hulv-version of the ULV algorithm (thin lines).

are different. Thus, the two signals have similar signal components lying in a 5-dimensional (or perhaps a 7-dimensional) subspace.

VI. Conclusion

We have introduced the canonical angles and the canonical filters associated with the subspaces from two rank-reduction algorithms, and demonstrated their use in the analysis and comparison of the two algorithms. In particular we have demonstrated by a numerical example that these quantities provide additional insight beyond inspection of the FIR filter arrays that characterize the algorithms.

References

[1] C. H. Bischof and G. M. Shroff, *On updating signal subspaces*, IEEE Trans. Signal Proc., 40 (1992), pp. 96–105.

[2] R. D. Fierro and P. C. Hansen, *Accuracy of TSVD solutions computed from rank-revealing decompositions*, Numer. Math., 70, (1995), pp. 453–471.

[3] R. D. Fierro, P. C. Hansen and P. S. K. Hansen, *UTV Tools: Matlab templates for rank-revealing UTV decompositions*, Numer. Algo., 20 (1999), pp. 165–194.

[4] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3. Ed., Johns Hopkins University Press, 1996.

[5] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems*, SIAM, Philadelphia, 1998.

[6] P. C. Hansen and S. H. Jensen, *FIR filter representations of reduced-rank noise reduction*, IEEE Trans. Signal Proc., 46 (1998), pp. 1737–1741.

[7] P. C. Hansen and P. Y. Yalamov, *Computing symmetric rank-revealing decompositions via triangular factorization*, submitted to SIAM J. Matrix Anal. Appl.

[8] L. L. Scharf and D. W. Tufts, *Rank reduction for modeling stationary signals*, IEEE Trans. Acoust. Speech Proc., 35 (1987), pp. 350–355.

[9] G. W. Stewart, *An updating algorithm for subspace tracking*, IEEE Trans. Signal Proc., 40 (1992), pp. 1535–1541.

[10] G. W. Stewart, *Updating a rank-revealing ULV decomposition*, SIAM J. Matrix Anal. Appl., 14 (1993), pp. 494–499.