REDUCED-RANK NOISE REDUCTION: A FILTER-BANK INTERPRETATION

Søren Holdt Jensen
CPK, Aalborg University
Fredrik Bajers Vej 7
DK-9220 Aalborg Øst, Denmark
E-mail: shj@cpk.auc.dk

Per Christian Hansen
UNI•C, Building 304
Technical University of Denmark
DK-2800 Lyngby, Denmark
E-mail: Per.Christian.Hansen@uni-c.dk

ABSTRACT

The key step in reduced-rank noise reduction algorithms is to approximate a matrix by another one with lower rank, typically by truncating a singular value decomposition (SVD). We give an explicit and closed-form derivation of the filter properties of the rank reduction operation and interpret this operation in the frequency domain by showing that the reduced-rank output signal is identical to that from a filter-bank whose analysis and synthesis filters are determined by the SVD. Our analysis includes the important general case in which pre- and dewhitening is used.

1 INTRODUCTION

Reduced-rank noise reduction, in which a signal matrix is approximated by another one with lower rank, may be viewed as an "energy decomposition" filtering out parts of the corresponding spectrum with low energy [1]. This techniques is the underlying principle in the noise reduction algorithms proposed in [2, 3], where the key idea is to form a Hankel (or Toeplitz) matrix from the input signal, compute the *singular value decomposition* (*SVD*) of the matrix, discard small singular values to obtain a matrix with reduced rank, and finally construct the output signal from this generally unstructured matrix by arithmetic averaging along its antidiagonals (or diagonals). This algorithm, which we shall refer to as the *truncated SVD* (*TSVD*) algorithm, has been used successfully for reduction of white noise in speech signals [2].

Prewhitening of the signal—corresponding to multiplying the signal matrix with a certain matrix—is sometimes used if the noise cannot be considered as white. The prewhitening operation can be included as an integral part of the algorithm, which then requires the computation of the *quotient SVD* (*QSVD*) of the signal-prewhitener matrix pair. This is the *truncated QSVD* (*TQSVD*) algorithm [3], in which the reduced-rank matrix is obtained by discarding small quotient singular values. The *TQSVD* algorithm has been used successfully for reduction of non-white broad-band noise in speech signals [3].

In this paper we give an explicit and closed-form derivation of the filter properties of the rank reduction operation and show that the filtered signal can be obtained from the input signal by means of a filter-bank consisting of *finite-duration impulse response* (*FIR*) filters whose coefficients are derived from the *SVD*/*QSVD*.

Our paper is based on ideas presented in [4], where it is concluded that the *TSVD* algorithm corresponds to subtracting from the input signal information contained in eigen-residuals corresponding to the smallest singular values. Our analysis, we hope, is more intuitive in that it uses a filter-bank to describe the *TSVD*/*TQSVD* filtering. Moreover, our analysis is more general since we include in a natural way the case of prewhitening (i.e., the *TQSVD* algorithm).

2 NOTATION

Before going into details of our analysis, it is useful to introduce the notation that we use throughout the rest of this paper. Given a vector $\mathbf{x} = (x_1, x_2, \ldots, x_N)^T$ of length N, we define the following two Hankel matrices of dimension $m \times n$ and $(N+n-1) \times n$, respectively, where $m + n - 1 = N$:

$$\mathcal{H}(\mathbf{x}) \equiv \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ x_2 & x_3 & \cdots & x_{n+1} \\ \vdots & \vdots & \vdots & \vdots \\ x_m & x_{m+1} & \cdots & x_N \end{pmatrix},$$

$$\mathcal{H}_p(\mathbf{x}) \equiv \begin{pmatrix} 0 & 0 & \cdots & x_1 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & x_1 & \cdots & x_{n-1} \\ x_1 & x_2 & \cdots & x_n \\ x_2 & x_3 & \cdots & x_{n+1} \\ \vdots & \vdots & \vdots & \vdots \\ x_m & x_{m+1} & \cdots & x_N \\ x_{m+1} & x_{m+2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ x_N & 0 & \cdots & 0 \end{pmatrix}.$$

Next, let $\mathbf{J}$ denote the symmetric matrix derived from the identity matrix by rearranging its columns in reverse order. We now define two Toeplitz matrices derived from $\mathcal{H}(\mathbf{x})$ and $\mathcal{H}_p(\mathbf{x})$ by reversing their columns, i.e.,

$$\mathcal{T}(\mathbf{x}) \equiv \mathcal{H}(\mathbf{x})\,\mathbf{J} \quad \text{and} \quad \mathcal{T}_p(\mathbf{x}) \equiv \mathcal{H}_p(\mathbf{x})\,\mathbf{J}.$$

Premultiplication of an n-vector $\mathbf{y}$ with either $\mathcal{H}(\mathbf{x})$ or $\mathcal{H}_p(\mathbf{x})$ corresponds to filtering $\mathbf{x}$ with a *FIR* filter whose coefficients are the elements of the vector $\mathbf{y}$. The only difference between the two operations lies in the way the two ends of the output signal vector are computed; the vector $\mathcal{H}(\mathbf{x})\mathbf{y}$ has length m, while the vector $\mathcal{H}_p(\mathbf{x})\mathbf{y}$ has length N. Similarly, premultiplication of $\mathbf{y}$ with either $\mathcal{T}(\mathbf{x})$ or $\mathcal{T}_p(\mathbf{x})$ corresponds to filtering $\mathbf{y}$ with a *FIR* filter whose coefficients are the elements of $\mathbf{x}$ in reverse order. To see this, note that $\mathbf{JJ}$ is the identity and therefore

$$\mathcal{T}_p(\mathbf{x})\mathbf{y} = (\mathcal{T}_p(\mathbf{x})\mathbf{J})\,(\mathbf{J}\mathbf{y}) = \mathcal{H}_p(\mathbf{x})\,(\mathbf{J}\mathbf{y})\,,$$

and similarly $\mathcal{T}(\mathbf{x})\mathbf{y} = \mathcal{H}(\mathbf{x})\,(\mathbf{J}\mathbf{y})$.

Finally, we define the averaging operator $\mathcal{A}$ that transforms a given $m \times n$ matrix $\mathbf{M}$ into a signal vector

$$\mathbf{s} \equiv \mathcal{A}(\mathbf{M})$$

of length N by arithmetic averaging along the $N = m + n - 1$ antidiagonals of $\mathbf{M}$:

$$s_i = \frac{1}{\beta - \alpha + 1} \sum_{k=\alpha}^{\beta} \mathbf{M}_{i-k+2,k}, \quad i = 1, \ldots, N,$$

where $\alpha = \max(1, i - m + 2)$ and $\beta = \min(n, i + 1)$.

For the special case where $\mathbf{M}$ is a rank-one matrix, i.e., where we can write $\mathbf{M} = \mathbf{u}\mathbf{v}^T$, the averaging operation can be expressed in the following simple form:

$$\mathcal{A}(\mathbf{M}) = \mathcal{A}(\mathbf{u}\mathbf{v}^T) = \mathbf{D}\mathcal{T}_p(\mathbf{u})\mathbf{v},$$

where $\mathbf{D}$ is an $N \times N$ diagonal matrix given by

$$\mathbf{D} = \operatorname{diag}(1, 2^{-1}, \ldots, \mu^{-1}, \mu^{-1} \ldots, \mu^{-1}, \ldots, 2^{-1}, 1),$$

with $\mu = \min(m, n)$. The first $\mu - 1$ and last $\mu - 1$ diagonal elements account for the fact that the corresponding antidiagonals of $\mathbf{M}$ do not have full length μ.

3 FILTER-BANK INTERPRETATION

In this section we give an explicit and closed-form derivation of the filter properties of the rank reduction operation.

3.1 *TSVD* ALGORITHM

Given an input signal vector $\mathbf{s}_{\text{in}}$ of length N, consisting of a pure signal plus additive white noise, we first form the associated $m \times n$ Hankel matrix

$$\mathbf{H} = \mathcal{H}(\mathbf{s}_{\text{in}}). \tag{1}$$

To simplify our exposition, we assume that $m \geq n$, which is usually the case in signal processing applications.

The next step is to compute the *SVD* of $\mathbf{H}$:

$$\mathbf{H} = \sum_{i=1}^{n} \sigma_i \mathbf{u}_i \mathbf{v}_i^T, \tag{2}$$

where the left and right singular vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ are orthonormal, and the singular values σ_i are nonnegative and appear in non-decreasing order, $\sigma_1 \geq \cdots \geq \sigma_n \geq 0$.

The third step is to approximate $\mathbf{H}$ by a rank-k matrix $\mathbf{H}_k$ with $k \leq n$. There are several possibilities here, and in a unified notation we can write $\mathbf{H}_k$ as

$$\mathbf{H}_k = \sum_{i=1}^{k} w_i \sigma_i \mathbf{u}_i \mathbf{v}_i^T, \qquad k \leq n. \tag{3}$$

The *least squares* approximation is obtained with $w_i = 1$, $i = 1, \ldots, k$, and the *minimum variance* approximation is obtained with $w_i = 1 - \sigma_{\text{noise}}^2/\sigma_i^2$, $i = 1, \ldots, k$, where σ_{noise}^2 is the variance of the white noise. For more details about w_i, see [5].

The final step is to compute the output signal vector $\mathbf{s}_{\text{out}}$ of length N from $\mathbf{H}_k$. This is done by arithmetic averaging along the antidiagonals of $\mathbf{H}_k$, i.e.,

$$\mathbf{s}_{\text{out}} = \mathcal{A}(\mathbf{H}_k). \tag{4}$$

To derive the filter-bank interpretation of the *TSVD* algorithm, we use the definitions and equations from the previous section together with the identity $\mathbf{H}\mathbf{v}_i = \sigma_i \mathbf{u}_i$, and we obtain

$$\mathbf{s}_{\text{out}} = \mathbf{D} \sum_{i=1}^{k} w_i\, \mathcal{H}_p(\mathcal{H}(\mathbf{s}_{\text{in}})\mathbf{v}_i)\,(\mathbf{J}\mathbf{v}_i). \tag{5}$$

This equation defines the precise relation between the input vector $\mathbf{s}_{\text{in}}$ and the output vector $\mathbf{s}_{\text{out}}$.

We see that the output signal essentially consists of a weighted sum of k intermediate signals $\mathbf{s}_i$ given by $\mathbf{s}_i = \mathcal{H}_p(\mathcal{H}(\mathbf{s}_{\text{in}})\mathbf{v}_i)\,(\mathbf{J}\mathbf{v}_i)$, $i = 1, \ldots, k$, of which $\mathcal{H}(\mathbf{s}_{\text{in}})\mathbf{v}_i$ is a signal obtained by passing $\mathbf{s}_{\text{in}}$ through a *FIR* filter with filter coefficients $\mathbf{v}_i$, and $\mathbf{s}_i$ is a signal obtained by passing $\mathcal{H}(\mathbf{s}_{\text{in}})\mathbf{v}_i$ through a *FIR* filter with filter coefficients $\mathbf{J}\mathbf{v}_i$, i.e., the coefficients of the first filter in reverse order. It is well known that this results in a zero-phase filtered version of $\mathbf{s}_{\text{in}}$. The weights simply represent k amplifiers with gain w_i and the matrix $\mathbf{D}$ represents an N-point window whose elements are the diagonal elements of $\mathbf{D}$.

From the above discussion it is evident that the *FIR* filters $\mathbf{v}_i$ and $\mathbf{J}\mathbf{v}_i$ constitute an analysis bank and a synthesis bank, respectively. Figure 1 summarizes this filter-bank interpretation of (5). For completeness we have included all n filters corresponding to the n *SVD* components of $\mathbf{H}$, plus a switch in each filter branch. The *TSVD* output signal $\mathbf{s}_{\text{out}}$ is then obtained by closing the first k switches corresponding to the largest k singular values used in (3). We note that exact reconstruction of $\mathbf{s}_{\text{in}}$ is obtained with $k = n$ and all $w_i = 1$.

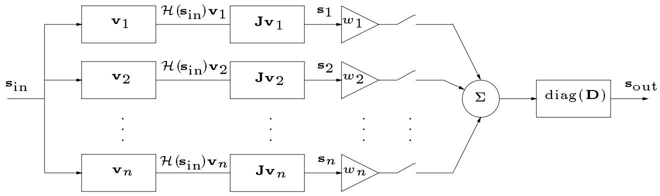

Figure 1: The filter-bank interpretation of the *TSVD* algorithm.

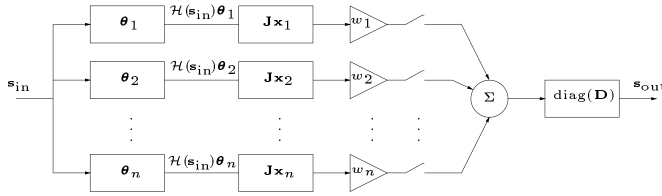

Figure 2: The filter-bank interpretation of the *TQSVD* algorithm.

3.2 *TQSVD* Algorithm.

The *TQSVD* algorithm [3], which is based on the *QSVD* of the matrix pair $(\mathcal{H}(\mathbf{s}_{\text{in}}), \mathcal{H}(\mathbf{e}))$, incorporates the pre- and dewhitening as an integral part of the algorithm. The *QSVD* is given by

$$\mathcal{H}(\mathbf{s}_{\text{in}}) = \sum_{i=1}^{n} \delta_i \hat{\mathbf{u}}_i \mathbf{x}_i^T, \qquad \mathcal{H}(\mathbf{e}) = \sum_{i=1}^{n} \mu_i \hat{\mathbf{v}}_i \mathbf{x}_i^T,$$

where $\hat{\mathbf{u}}_i$ and $\hat{\mathbf{v}}_i$ are orthonormal vectors, the vectors $\mathbf{x}_i$ are linearly independent, and $\boldsymbol{\Delta} = \text{diag}(\delta_1, \ldots, \delta_n)$ and $\mathbf{M} = \text{diag}(\mu_1, \ldots, \mu_n)$ with $\delta_i^2 + \mu_i^2 = 1$ for $i = 1, \ldots, n$. The quotient singular values are δ_i/μ_i, and they satisfy $\delta_1/\mu_1 \geq \cdots \geq \delta_n/\mu_n \geq 0$.

In the *TQSVD* algorithm, the rank-k matrix approximation that corresponds to $\mathbf{H}_k$ in (3) is given by

$$\mathbf{Z}_k = \sum_{i=1}^{k} w_i \delta_i \hat{\mathbf{u}}_i \mathbf{x}_i^T \ , \tag{6}$$

and the *TQSVD* output signal is $\mathbf{s}_{\text{out}} = \mathcal{A}(\mathbf{Z}_k)$. The weights w_i are computed by the same formulas as in the *TSVD* algorithm, but with the singular values σ_i replaced by the quotient singular values δ_i/μ_i and σ^2_{noise} being the noise variance of the prewhitened signal.

Also here, we obtain an expression for $\mathbf{s}_{\text{out}} = \mathcal{A}(\mathbf{Z}_k)$ that leads to a filter-bank interpretation. The main difference is that we need an additional set of vectors $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n$ defined such that $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n)^T = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^{-1}$. We note that the two sets of vectors are biorthonormal, i.e., $\boldsymbol{\theta}_i^T \mathbf{x}_j = 1$ for $i = j$ and zero otherwise. To derive the filter-bank interpretation of the *TQSVD* algorithm, we use the identity $\mathbf{H}\boldsymbol{\theta}_i = \delta_i \hat{\mathbf{u}}_i$ and obtain

$$\mathbf{s}_{\text{out}} = \mathbf{D} \sum_{i=1}^{k} w_i \, \mathcal{H}_p(\mathcal{H}(\mathbf{s}_{\text{in}})\boldsymbol{\theta}_i) \, (\mathbf{J}\mathbf{x}_i). \tag{7}$$

For details of the derivation, see [5]. Figure 2 summarizes our filter-bank interpretation of (7).

The filters that arise in the filter-bank interpretation of the *TQSVD* algorithm have filter coefficients $\boldsymbol{\theta}_i$ and $\mathbf{J}\mathbf{x}_i$, and they are not zero-phase filters as in the white-noise case. Perfect reconstruction is still obtained when $k = n$ and all $w_i = 1$.

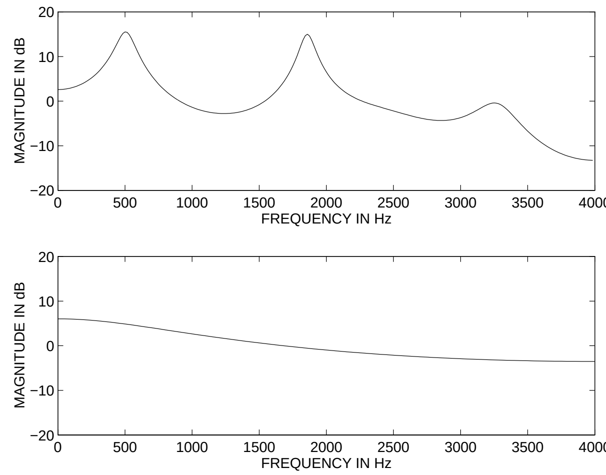

Figure 3: The *LPC* model spectrum of the pure preemphasized speech segment (top) and the *PSD* of the noise (bottom).

4 NUMERICAL EXAMPLE

In order to illustrate our filter-bank interpretation, we apply the *TQSVD* algorithm to a realistic example in speech enhancement.

We consider a segment $\mathbf{s}_{\text{in}}$ of voiced speech in additive colored noise. The speech signal was sampled at 8 kHz and filtered by a first order *FIR* preemphasis filter:

$$P(z) = 1 - \beta z^{-1}, \quad \beta = 0.95.$$

The noise was generated such that its *power spectral density* (*PSD*) was

$$S(\omega) = \frac{1}{(1+a^2) - 2a\cos(\omega T_s)}, \quad a = 0.5, \quad T_s = 125\mu\text{s}$$

and scaled such that the *SNR* of the signal $\mathbf{s}_{\text{in}}$ was 15 dB. Figure 3 shows the *linear predictive coding* (*LPC*) model spectrum of the pure speech segment (top) and the *PSD* of the noise (bottom). The segment length N was set to 160 samples corresponding to 20 ms, which is the typical block length in speech processing because this length is short enough for the segment to be nearly stationary. The size of both $\mathbf{H}$ and $\mathbf{N}$ was set to 141×20.

Table 1 lists the 20 quotient singular values δ_i/μ_i of $(\mathbf{H}, \mathbf{N})$, and Fig. 4 shows the frequency response

Table 1: The 20 quotient singular values of the matrix pair $(\mathbf{H}, \mathbf{N})$.

i	δ_i/μ_i	δ_{5+i}/μ_{5+i}	δ_{10+i}/μ_{10+i}	δ_{15+i}/μ_{15+i}
1	12.42	4.58	4.03	2.44
2	11.73	4.50	3.81	2.11
3	10.33	4.35	3.79	1.63
4	9.99	4.16	3.38	1.51
5	4.90	4.08	2.85	1.47

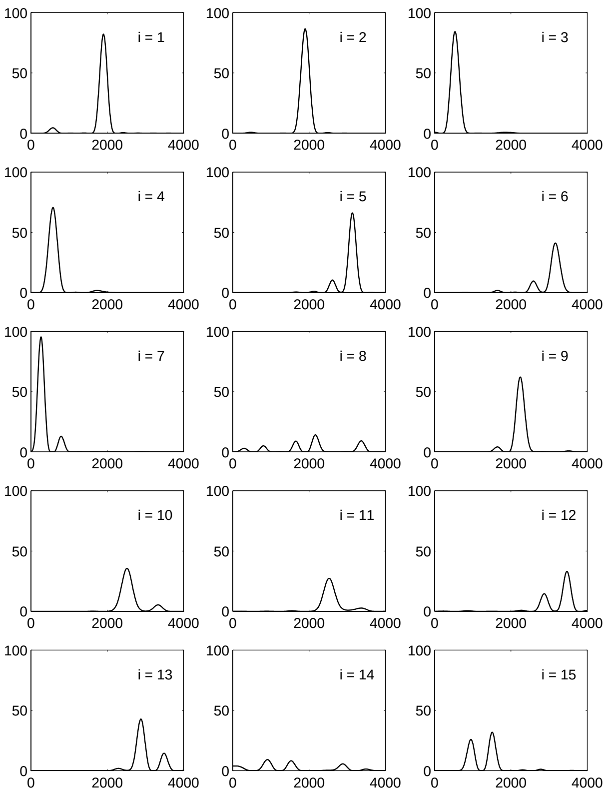

Figure 4: Frequency responses of the first 15 analysis/synthesis filters.

of the first 15 combined analysis/synthesis filters (cf. Fig. 2) associated with the *QSVD* of $(\mathbf{H}, \mathbf{N})$. It is clearly seen that the first 6 analysis/synthesis filters have band-pass characteristics and capture the 3 formants (i.e., the maxima in the *LPC* model spectrum) of the speech segment. These three formants, and in particular the two with lowest frequency, are so "peaked" that they can be modeled with good approximation by pure sinusoids—hence the shape of the corresponding *FIR* filters.

The remaining 14 filters (of which the first 9 are shown in Fig. 4) capture those spectral components of the signal that lie in between the formants (including the noise), and these components are much less important when reconstructing the signal. In this way, we achieve a good noise reduction of voiced speech sounds in the *TQSVD* method.

5 CONCLUSION

Reduced-rank noise reduction is interpreted as a filter bank, consisting of analysis/synthesis *FIR* filters whose filter coefficients are derived from the *SVD* of the signal matrix. The number of filter branches equals the rank of the approximating matrix.

In the white-noise case (no prewhitening), the analysis filter coefficients are the right singular vectors of the signal matrix. The synthesis filter coefficients are the same, except that they appear in reverse order.

When pre- and dewhitening is used, then the analysis and synthesis filters are determined by the right quotient singular vectors of the signal-prewhitener matrix pair. These coefficients, in turn, are related to the right singular vectors of the prewhitened signal.

Traditionally, reduced-rank techniques have been analyzed and interpreted in linear algebra terms, in particular signal and noise subspaces. Our analysis gives new insight: it interprets reduced-rank techniques as filter operations and therefore allows an analysis in the frequency domain. This insight can, for example, provide an aid to choosing an appropriate prewhitener.

ACKNOWLEDGEMENTS

Søren Holdt Jensen was supported in part by the European Community Research Program HCM, contract no. ERBCHBI-CT92-0182. Per Christian Hansen was supported in part by grant 9500764 from the Danish Natural Science Research Council.

REFERENCES

[1] L. L. Scharf and D. W. Tufts, "Rank Reduction for Modeling Stationary Signals," *IEEE Trans. Acoust., Speech, Signal Processing,* 35(3):350–355, 1987.

[2] M. Dendrinos, S. Bakamidis, and G. Carayannis, "Speech Enhancement from Noise: A Regenerative Approach", *Speech Communication*, 10(1):45–57, 1991.

[3] S. H. Jensen, P. C. Hansen, S. D. Hansen, and J. A. Sørensen, "Reduction of Broad-Band Noise in Speech by Truncated QSVD," *IEEE Trans. Speech, Audio Processing*, 3(6):439–448, 1995.

[4] I. Dologlou and G. Carayannis, "Physical Interpretation of Signal Reconstruction from Reduced Rank Matrices", *IEEE Trans. Signal Processing*, 39(7):1681–1682, 1991.

[5] P. C. Hansen and S. H. Jensen, "Filter-Bank Interpretation of Reduced-Rank Noise Reduction," *Report UNIC-95-12*, UNI•C, Lyngby, Denmark, Oct. 1995 (17 pages).