Reduction of Broad-Band Noise in Speech by Truncated QSVD

Søren Holdt Jensen, Per Christian Hansen, Steffen Duus Hansen, and John Aasted Sørensen

Abstract— We consider an algorithm for reduction of broad-band noise in speech based on signal subspaces. The algorithm is formulated by means of the quotient singular value decomposition (QSVD). With this formulation, a prewhitening operation becomes an integral part of the algorithm. We demonstrate that this is essential in connection with updating issues in real-time recursive applications. We also illustrate by examples that we are able to achieve a satisfactory quality of the reconstructed signal.

I. INTRODUCTION

The objective of noise reduction is to improve noisy signals. One important application is the enhancement of speech transmission degraded by broadband noise as, for instance, in hands-free mobile telephony, where speech communication is affected by the presence of acoustic noise. This effect is particularly serious when linear predictive coding (LPC) is used for the digital representation of speech signals at low bit rates, as in digital mobile telephony. Low-frequency acoustic noise severely affects the estimated LPC spectrum in both the low- and high-frequency regions. Consequently, the intelligibility of digitized speech using LPC often falls below the minimum acceptable level. Therefore, there is a great need for using efficient noise reduction algorithms before the analysis stage in speech coders in order to remove as much of the noise as possible and, hence, give a better set of LPC coefficients for the synthesis stage.

Several noise reduction algorithms have been proposed, and two categories of algorithms dominate in the current literature. One of them is based on spectral decomposition [1], [13], [15] and uses single microphone measurements. The other is based on least-mean squares (LMS) filtering [21] and uses noise cancelling methodologies that require as least two microphones. However, in a mobile environment, the first category of algorithms suffers from the fact that either the noise reduction is not sufficient or the speech distortion is too large [22]. The second class of algorithms suffers from the fact that it is difficult to place the microphones inside the car cabin such that the noise part in the speech signal can be canceled, cf. [3], [12]. Thus, there is a need to develop more general and robust algorithms for noise reduction in mobile telephony where small SNR’s are not uncommon. In the present paper, we investigate an approach for the reduction of broadband noise in speech based on the signal subspace paradigm. The key idea is to consider the signal as a vector in N-dimensional space and to separate the “pure signal” and the noise into two mutually orthogonal components lying in different subspaces. In this way, one is able to obtain improved noise reduction, compared with the more classical approaches.

Signal subspace methods have been used frequently in connection with, e.g., frequency estimation and direction-of-arrival problems, cf. [7], [18]. However, until recently, this approach has not been used much in speech processing. In the area of speech enhancement, signal subspace methods have been proposed by Dendrinos et al. in [4] and by Ephraim and Van Trees in [5]. The noise reduction algorithm proposed in [4] is based on singular value decomposition (SVD) and deals with the special case of white noise only. In essence, the algorithm first arranges the data in a Toeplitz-structured matrix and then computes a least-squares estimate of the signal-only data matrix and finally restores the Toeplitz structure of the computed least-squares estimate. The least-squares estimate of the signal-only data matrix is computed by neglecting small singular values, and the number of retained singular values is determined adaptively from the data in each speech segment. The algorithm in [5] is very similar, except that it uses the Karhunen-Loève transform instead of the SVD. There are two immediate drawbacks of these algorithms. The first is that the algorithms deal only with white noise, i.e., in case of general noise, a prewhitening is necessary. Prewhtening is difficult to treat in a recursive algorithm. Second, the least-squares estimate requires a good judicious accounting concerning the number of retained singular values, and much computational effort is spent in computing this number.

In this paper, we propose an extension of the algorithm by Dendrinos et al. Our algorithm is also inspired by work by Van Huffel [19], and it deals with general broadband noise and is suited for updating. Moreover, by using a minimum-variance estimate of the signal-only matrix (instead of the least-squares estimate), our algorithm is less sensitive to the choice of retained singular values.

The rest of the paper is organized as follows. In Section II, we define the signal and noise model and review the
conditions that allow us to derive the signal subspace from the SVD of a data matrix. We also present the least-squares and the minimum-variance estimates of the original data. In Section III, we outline the broadband noise reduction algorithm for speech enhancement. For convenience, we first describe the algorithm using prewhitening; then, we point out how the prewhitening step can be incorporated by using a generalization of the SVD to two matrices, namely, the quotient SVD (QSVD). After discussing the details of our algorithm, we present in Section IV several examples of the reconstruction of voiced and unvoiced speech sounds as well as the reconstruction of whole sentences contaminated by broadband noise. Finally, a summary and conclusions are given in Section V.

II. SIGNAL SUBSPACE ESTIMATION

A. Signal and Noise Model

We consider a noisy signal vector \( \mathbf{x} = [x_0, x_1, \ldots, x_{N-1}]^T \) of \( N \) samples, and we assume that the noise is additive and uncorrelated with the signal, i.e.,

\[
\mathbf{x} = \bar{\mathbf{x}} + \mathbf{n},
\]

where \( \bar{\mathbf{x}} \) contains the signal component, and \( \mathbf{n} \) represents the noise. From \( \mathbf{x} \), we can construct the following \( L \times M \) Hankel matrix \( \mathbf{H} \), where \( M + L = N + 1 \) and \( L \geq M \):

\[
\mathbf{H} = \begin{bmatrix}
x_0 & x_1 & \cdots & x_{M-1} \\
x_1 & x_2 & \cdots & x_M \\
\vdots & \vdots & \ddots & \vdots \\
x_{L-1} & x_L & \cdots & x_{N-1}
\end{bmatrix}
\]

(2)

We can always write \( \mathbf{H} \) as

\[
\mathbf{H} = \mathbf{\bar{H}} + \mathbf{N}
\]

(3)

where \( \mathbf{\bar{H}} \) and \( \mathbf{N} \) represent, respectively, the Hankel matrices derived from \( \bar{\mathbf{x}} \) and \( \mathbf{n} \) in (1). Moreover, we assume that \( \mathbf{\bar{H}} \) is rank deficient \( \text{rank}(\mathbf{\bar{H}}) = K < M \), and that \( \mathbf{H} \) and \( \mathbf{N} \) have full rank \( \text{rank}(\mathbf{H}) = \text{rank}(\mathbf{N}) = M \). This assumption is, e.g., satisfied when the samples \( \bar{x}_i \) of \( \bar{\mathbf{x}} \) consist of a sum of \( K \) sinusoids, and the samples \( n_i \) of \( \mathbf{n} \) consist of broadband noise. Such a model has often been attributed to speech, see, e.g., [16].

We remark that one can also choose to work with Toeplitz matrices instead of Hankel matrices. There are no fundamental differences between these two approaches.

B. The SVD and Signal Subspace Estimation for the White Noise Case

The SVD is a fundamental tool in subspace methods.

Theorem 1—(Singular Value Decomposition (SVD)): If \( \mathbf{H} \in \mathbb{R}^{L \times M} \) with \( L \geq M \), then there exist matrices \( \mathbf{U} \in \mathbb{R}^{L \times L} \) and \( \mathbf{V} \in \mathbb{R}^{M \times M} \) with orthonormal columns such that

\[
\mathbf{H} = \mathbf{U}\Sigma\mathbf{V}^T
\]

(4)

where \( \Sigma = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_M) \) with \( \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_M \geq 0 \).

Proof: For the proof, see [8, p. 71].

The diagonal elements of \( \Sigma \) are called singular values of \( \mathbf{H} \), and their set is called the singular-value spectrum. The columns of \( \mathbf{U} \) and \( \mathbf{V} \) are called left and right singular vectors. For the following discussion, it is convenient to partition the SVD of \( \mathbf{H} \) as follows:

\[
\mathbf{H} = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{bmatrix}
\]

(5)

where \( \mathbf{U}_1 \in \mathbb{R}^{L \times K} \), \( \Sigma_1 \in \mathbb{R}^{K \times K} \), and \( \mathbf{V}_1 \in \mathbb{R}^{M \times K} \). We also partition the SVD of \( \mathbf{\bar{H}} \) accordingly:

\[
\mathbf{\bar{H}} = \begin{bmatrix} \mathbf{\bar{U}}_1 & \mathbf{\bar{U}}_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} \mathbf{\bar{V}}_1^T \\ \mathbf{\bar{V}}_2^T \end{bmatrix}
\]

(6)

We now make the following three assumptions:

1) The signal is orthogonal to the noise in the sense: \( \mathbf{H}^T \mathbf{N} = 0 \).

2) The noise is white, i.e., the noise matrix \( \mathbf{N} \) has orthogonal columns and every column of \( \mathbf{N} \) has norm \( \sigma_{\text{noise}} \):

\[
\mathbf{N}^T \mathbf{N} = \sigma_{\text{noise}}^2 \mathbf{I}
\]

(7)

3) The smallest singular value of \( \Sigma_1 \) is strictly larger than the largest singular value of \( \Sigma_2 \) in the SVD of \( \mathbf{H} \), i.e.,

\[
\sigma_K > \sigma_{K+1}
\]

A detailed derivation of the algebraic and geometric conditions that allow us to derive the signal model from the SVD of the data matrix (5) is given by De Moor in [6]. Using a multivariate version of the classical Pythagorean lemma for triangles, he shows that the signal matrix \( \mathbf{\bar{H}} \) cannot be recovered consistently when the signal matrix is perturbed by additive noise.

Specifically, with the above three assumptions, the rank of \( \mathbf{\bar{H}} \), the row space of \( \mathbf{\bar{H}} \) (represented by \( \mathbf{\bar{V}}_1 \)), and the null space of \( \mathbf{\bar{H}} \) (represented by \( \mathbf{\bar{V}}_2 \)) can be estimated consistently, whereas the column space of \( \mathbf{\bar{H}} \) (represented by \( \mathbf{\bar{U}}_1 \)) cannot be recovered from the SVD of \( \mathbf{\bar{H}} \), not even asymptotically.

Despite the fact that \( \mathbf{\bar{U}}_1 \) cannot be consistently estimated from the SVD of \( \mathbf{\bar{H}} \), it is possible to find a least-square or minimum-variance estimate of \( \mathbf{\bar{H}} \) from the SVD of \( \mathbf{\bar{H}} \).

1) Least-Square Estimate of \( \mathbf{\bar{H}} \): The simplest estimate of \( \mathbf{\bar{H}} \), given \( \mathbf{\bar{H}} \), is obtained when we approximate \( \mathbf{\bar{H}} \) by a matrix of rank \( K \) in the least-squares (LS) sense:

\[
\min_{\text{rank}(\mathbf{H}_{LS} = K)} \| \mathbf{H} - \mathbf{\bar{H}}_{LS} \|^2_F.
\]

(8)

The solution is a classical result, see, e.g., [8, p. 73], and follows directly from the SVD of \( \mathbf{\bar{H}} \):

\[
\mathbf{\bar{H}}_{LS} = \mathbf{U}_1 \Sigma_1 \mathbf{V}_1^T.
\]

(9)

This estimate is, e.g., used in the SVD-based method for estimating the signal components of a noisy data vector proposed by Tufts et al. in [17] and in the SVD-based method for speech enhancement proposed by Dendrinos et al., in [4].

2) Minimum-Variance Estimate of \( \mathbf{\bar{H}} \): Another estimate is the minimum-variance (MV) estimate, as used in the MV estimation method described in [19]. The MV estimation problem can be formulated as follows [6], [19]: Given the
matrix $H$ as in (3), with $\text{rank}(H) = \text{rank}(N) = M$ and $\text{rank}(H) = K$. Find the matrix $T$ that minimizes

$$
\min_{T \in \mathbb{R}^{M \times M}} \|HT - \tilde{H}\|_F^2.
$$

The solution is given by

$$
T = (H^T H)^{-1} H^T \tilde{H}.
$$

Hence, the MV estimate of $H$ is given by

$$
HT = H(H^T H)^{-1} H^T \tilde{H}.
$$

By means of the SVD of $H$, it is easy to show that

$$
HT = UU^T \tilde{H}.
$$

Geometrically, the MV estimate of $\tilde{H}$, given $H$, can be interpreted as the orthogonal projection of $\tilde{H}$ onto the column space of $H$ because $UU^T$ is the associated projection matrix. Observe that $\text{rank}(HT) = \text{rank}(H) = K$. In spite of the fact that $H$ is not known, it is possible to compute the MV estimate from the SVD of $H$ if Assumptions 1)–3) are satisfied. Using (3) and (6), the SVD of $H$ is then given by

$$
H = \tilde{H} + N
= U_1 \Sigma_1 V_1^T + NV_1 V_1^T + NV_2 V_2^T
= \left[ (U_1 \Sigma_1 + NV_1) (\Sigma_1^2 + \sigma_{\text{noise}}^2 I)^{-1/2} \right]
\left[ \begin{array}{c}
\Sigma_1^2 + \sigma_{\text{noise}}^2 I
0
\end{array} \right]
\left[ \begin{array}{c}
V_1^T
V_2
\end{array} \right].
$$

We observe that the middle matrix is diagonal, and the left and right matrices have orthonormal columns. Hence, (13) is a SVD of $H$, and we identify the singular values of $H$ as

$$
\Sigma_1 = (\Sigma_1^2 + \sigma_{\text{noise}}^2 I)^{1/2},
$$

$$
\Sigma_2 = \sigma_{\text{noise}} I.
$$

Thus, the singular values in $\Sigma_2$ can be interpreted as a noise threshold, which permits estimating $\sigma_{\text{noise}}$ from (15).

From (13), we also identify the submatrices

$$
U_1 = (U_1 \Sigma_1 + NV_1) \Sigma_1^{-1},
$$

$$
U_2 = \sigma_{\text{noise}}^{-1} NV_2,
$$

$$
V_1 = \tilde{V}_1,
$$

$$
V_2 = \tilde{V}_2.
$$

The last equality in (16) follows from (14). Using relations (14)–(19) together with the important relation $H^T N = 0 \Rightarrow U_1^T N = 0$, we then obtain the desired MV estimate of $\tilde{H}$:

$$
H_{\text{MV}} = UU^T \tilde{H}
= \begin{bmatrix}
U_1 U_1^T & U_2 U_1^T
\end{bmatrix}
\begin{bmatrix}
\Sigma_1^2 + \sigma_{\text{noise}}^2 I
\sigma_{\text{noise}}^2 I
\end{bmatrix}
\begin{bmatrix}
\Sigma_1 V_1^T
V_2
\end{bmatrix},
$$

$$
= U_1 \Sigma_1^2 V_1^T + U_1 \Sigma_1 \sigma_{\text{noise}}^2 V_1^T + U_2 U_1^T \Sigma_1 V_1^T
= U_1 \Sigma_1 \sigma_{\text{noise}}^{-1} U_1^T V_1^T + U_1 \Sigma_1 \sigma_{\text{noise}}^2 V_1^T
= U_1 \Sigma_1 \sigma_{\text{noise}}^{-1} U_1^T V_1^T + U_1 \Sigma_1 \sigma_{\text{noise}}^2 V_1^T
= U_1 \Sigma_1 \sigma_{\text{noise}}^{-1} U_1^T V_1^T.
$$

Equations (13) and (20) show that we do not obtain a consistent estimate of the column space of $H$ since $U_1 \neq U_1'$, i.e., the estimate is biased.

3) A Unified Notation The left and right singular vectors of the LS and MV estimate, respectively, are the same, but the singular values are different. For ease of notation, we therefore introduce the following $K \times K$ filter matrices:

$$
F_{LS} = I_K,
$$

$$
F_{MV} = \text{diag}\left( \left( 1 - \frac{\sigma_{\text{noise}}^2}{\sigma_1^2} \right), \ldots, \left( 1 - \frac{\sigma_{\text{noise}}^2}{\sigma_K^2} \right) \right).
$$

which allow us to unify our notation and write the LS and MV estimate of $\tilde{H}$ as

$$
\tilde{H}_{LS} = U_1 (F_{LS} \Sigma_1) V_1^T,
$$

$$
\tilde{H}_{MV} = U_1 (F_{MV} \Sigma_1) V_1^T.
$$

In Section IV, we return to the actual choice of the rank $K$ in a practical algorithm.

We remark that both approaches are equivalent to regularization of the problem in the sense that the influence of noise is suppressed by discarding the SVD components associated with the smallest singular values. In particular, the LS technique is identical to “truncated SVD.” Regularization techniques are commonly used in connection with deconvolution problems in image reconstruction, seismology, and other areas that involve inverse problems; see [10] for more details.

C. The Colored Noise Case

If the noise is not white, $N^T N \neq \sigma_{\text{noise}}^2 I$, then a prewhitening matrix $R^{-1}$ can always be applied to $H$ in (3). If the noise covariance matrix of the type $N^T N$ is known or can be estimated, then we can factor $N^T N = R^T R$, where $R$ is the Cholesky factor of $N^T N$. Moreover, if the noise matrix $N$ is directly available, then we can factor $N = QR$, where $Q^T Q = I$ and $R$ is the same Cholesky factor as before. In the colored noise case, we then consider the matrix

$$
X = HR^{-1}.
$$

Substituting (3) into (25) yields

$$
X = HR^{-1} + NR^{-1}.
$$

From $N = QR$, we obtain $NR^{-1} = Q$, and thus, $(NR^{-1})^T (NR^{-1}) = Q^T Q = I$, which is the identity matrix. Hence, (26) is now in the same form as (3), and Assumptions 1)–3) apply to the transformed signal matrix $HR^{-1}$ and the transformed noise matrix $NR^{-1}$. We can therefore apply the technique from (5)–(20) to the transformed matrices.

The requirement that the noise covariance matrix $N^T N$ or an estimate should be known is not always trivial. In speech processing, we are fortunate that $N$ is known for “silent” periods in the speech and that this can be used to estimate the noise covariance matrix for data where speech is present.
D. Signal Reconstruction by Hankel Matrix Approximation

The LS and MV estimates $\mathbf{H}_{LS}$ (23) and $\mathbf{H}_{MV}$ (24) of $\mathbf{H}$ spoil the Hankel structure. Therefore, it is not easy to directly obtain a signal vector corresponding to the LS and MV estimates. We need to make a Hankel matrix approximation to $\mathbf{H}_{LS}$ or $\mathbf{H}_{MV}$.

A simple way to compute a Hankel matrix approximation is to arithmetically average every antidiagonal of these matrices and put each average-value as a common element in the corresponding diagonal of a new Hankel-structured matrix of the same dimension:

$$
\tilde{\mathbf{H}} = \begin{bmatrix}
\hat{x}_0 & \hat{x}_1 & \ldots & \hat{x}_{M-1} \\
\hat{x}_1 & \hat{x}_2 & \ldots & \hat{x}_M \\
\vdots & \vdots & \ddots & \vdots \\
\hat{x}_{L-1} & \hat{x}_L & \ldots & \hat{x}_{N-1}
\end{bmatrix}
$$

(27)

where

$$
\hat{x}_i = \frac{1}{\beta - \alpha + 1} \sum_{k=\alpha}^{\beta} \tilde{\mathbf{H}}(i-k+2,k),
$$

(28)

$$
\alpha = \max(1,i-L+2),
$$

(29)

$$
\beta = \min(M,i+1).
$$

(30)

A similar transformation is used by Tufts et al. in [17], but here, a definition that leads to matrices with block Toeplitz-Hankel structure is used.

The process for making a Hankel matrix approximation to $\mathbf{H}_{LS}$ and $\mathbf{H}_{MV}$ implies that the exact rank is lost in $\tilde{\mathbf{H}}$. However, $\tilde{\mathbf{H}}$ is closer to $\mathbf{H}_{LS}$ and $\mathbf{H}_{MV}$ than was the original data matrix $\mathbf{H}$. Hence, the elements of $\tilde{\mathbf{H}}$ comprise signal elements that are more compatible with the $K$th-order signal model than the elements of $\mathbf{H}$. In speech applications, this simple transformation is sufficient to reduce noise that contaminates the signal, cf. [4]. This was also confirmed by our experiments.

To obtain a signal vector that is exactly represented by a $K$th-order signal model, we can repeat the process a number of times starting from $\tilde{\mathbf{H}}$. This technique was proposed by Cadzow in [2] and used, e.g., by Van Huffel in [19]. For more details, we refer to [2].

III. NOISE REDUCTION ALGORITHM

A. Formulation of the Algorithm

We can now formulate the complete algorithm for reduction of broadband noise in speech, based on subspace estimation, prewhitening, and signal reconstruction.

Algorithm

0) Form the $L \times M$ Hankel matrices $\mathbf{H}$ and $\mathbf{N}$, $L \geq M > K$ and $L + M = N + 1$ from the available signal and noise vectors.

1) Compute the QR decomposition of $\mathbf{N}$:

$$
\mathbf{N} = \mathbf{QR}.
$$

(31)

2) Perform a prewhitening of $\mathbf{H}$:

$$
\mathbf{X} = \mathbf{HR}^{-1}.
$$

(32)

3) Modify the matrix $\mathbf{X}$ as follows:

a. Compute the SVD of $\mathbf{X}$

$$
\mathbf{X} = [\mathbf{U}_1 \quad \mathbf{U}_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{bmatrix}.
$$

(33)

b. Truncate $\mathbf{X}$ to rank $K$ and correct the retained singular values by applying a $K \times K$ filter matrix $\mathbf{F}$, cf. Section II.B.3:

$$
\mathbf{X}_K = \mathbf{U}_1 (\mathbf{F} \Sigma_1) \mathbf{V}_1^T.
$$

(34)

4) Perform a dewhitening of $\mathbf{X}_K$:

$$
\mathbf{Z}_K = \mathbf{X}_K \mathbf{R}.
$$

(35)

5) Compute the signal vector $\mathbf{x}$ from $\mathbf{Z}_K$ by arithmetic averaging along its antidiagonals:

$$
\hat{x}_i^{(t)} = \frac{1}{\beta - \alpha + 1} \sum_{k=\alpha}^{\beta} \mathbf{Z}_K(i-k+2,k),
$$

(36)

$$
\alpha = \max(1,i-L+2),
$$

(37)

$$
\beta = \min(M,i+1).
$$

(38)

If desired, Steps 2–5 can be repeated iteratively by starting from the reconstructed signal vector $\mathbf{x}$ obtained in the previous iteration step, cf. Section II.D.

A challenging problem in the noise reduction algorithm is the determination of $K$, i.e., the effective rank of the data matrix. In a recursive algorithm, it is convenient if a fixed effective rank can be used. This problem is addressed in Section IV.

B. Implementation by Means of QSVD

The explicit use of the matrix $\mathbf{R}$ may result in a loss of accuracy in the data. This can be avoided by working directly with $\mathbf{H}$ and $\mathbf{N}$, i.e., by using the QSVD of the pair of matrices $(\mathbf{H}, \mathbf{N})$ that, together with the matrix $\mathbf{Q}$, delivers the required factorization without forming quotients and products. Another issue, which may be even more important in practice, is that it is very complicated to update the matrix $\mathbf{X} = \mathbf{HR}^{-1}$ when $\mathbf{H}$ and $\mathbf{N}$ are updated, e.g., in a recursive application. Therefore, it is advantageous to use a decomposition of the matrix pair $(\mathbf{H}, \mathbf{N})$ instead, which allows each matrix to be updated individually. We return to this subject in the next section.

**Theorem 2**—(Quotient Singular Value Decomposition (QSVD)) If $\mathbf{H}, \mathbf{N} \in \mathbb{R}^{L \times M}$ with $L \geq M$, then there exist matrices $\mathbf{U}, \mathbf{V} \in \mathbb{R}^{L \times M}$, with $\mathbf{U}^T \mathbf{U} = \mathbf{I}$, $\mathbf{V}^T \mathbf{V} = \mathbf{I}$, and a nonsingular $\mathbf{\Theta} \in \mathbb{R}^{M \times M}$ such that

$$
\mathbf{H} = \mathbf{U} \Delta \mathbf{\Theta}^{-1},
$$

(39)

$$
\mathbf{N} = \mathbf{V} \mathbf{M} \mathbf{\Theta}^{-1},
$$

(40)

where

$$
\Delta = \text{diag}(\delta_1, \delta_2, \ldots, \delta_M),
$$

(41)

$$
\mathbf{M} = \text{diag}(\mu_1, \mu_2, \ldots, \mu_M)
$$

(42)

and $\delta_1 \geq \delta_2 \geq \ldots \geq \delta_M$ and $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_M$. 

Proof: For the proof, see [8, p. 471], where the phrase “GSVD” is used instead of “SVD.”

The matrix $R$ may be written as $R = Q^T N = Q^T V M \Theta^{-1}$. By substituting this and (39) into $X = HR^{-1}$, we obtain

$$X = HR^{-1} = U \Delta \Theta^{-1} (Q^T V M \Theta^{-1})^{-1} = U \Delta \Theta^{-1} \Theta M^{-1} V^T Q = U (\Delta M^{-1}) (Q^T V)^T. \tag{43}$$

This shows that the three matrices $U$, $\Delta M^{-1} = \text{diag}(\delta_1/\mu_1, \delta_2/\mu_2, \ldots, \delta_M/\mu_M)$, and $Q^T V$ are identical to the SVD of $X = HR^{-1}$, with $\Sigma = \Delta M^{-1}$. Hence, working with the QSVD of $(H, N)$ and the matrix $Q$ is mathematically equivalent to working with the SVD of $HR^{-1}$.

Using the QSVD formulation, Steps 1–4) of the noise reduction algorithm reduce to the following two steps:

1) Compute the QSVD of $(H, N)$:

$$H = U \Delta \Theta^{-1}, \tag{44}$$

$$N = VM \Theta^{-1}. \tag{45}$$

2) Obtain $Z_K$ as follows:

$$\Delta_1 = \text{diag}(\delta_1, \delta_2, \ldots, \delta_K), \tag{46}$$

$$Z_K = U \begin{bmatrix} \Delta_1 & 0 \\ 0 & 0 \end{bmatrix} \Theta^{-1}. \tag{47}$$

We see that the prewhitening is now an integral part of the algorithm, and we refer to the complete algorithm based on the QSVD implementation as the truncated QSVD algorithm. A similar use of the truncated QSVD is suggested in connection with regularization problems by Hansen in [9].

C. Updating Issues

In some applications, one would like to update the QSVD instead of recomputing it. This is desirable in real-time signal processing applications. Updating of the QSVD is a topic of current research. In connection with the truncated QSVD algorithm, we see from (47) that it is only the matrices $U$, $\Delta$, $M$, and $\Theta^{-1}$ that we need to update.

We mention in passing that it may be more convenient to update the matrix $V$ instead of the matrix $\Theta^{-1}$ because $V$ is orthogonal, whereas $\Theta^{-1}$ is merely nonsingular. In this case, one can obtain $Z_K$ as follows:

$$\Delta' = \text{diag}(\delta_1, \delta_2, \ldots, \delta_K), \tag{48}$$

$$M' = \text{diag}(\mu_1, \mu_2, \ldots, \mu_K), \tag{49}$$

$$Z_K = U \begin{bmatrix} \Delta' & M' \\ 0 & 0 \end{bmatrix} (N^T V)^T. \tag{50}$$

That is, the cost of this approach is an extra matrix multiplication and $K$ extra divisions.

At any rate, the QSVD is rather difficult to update and may not be the best choice in a practical application. A promising alternative is to use a related decomposition: the rank-revealing ULLV decomposition proposed by Luk and Qiao in [14]. Here, the matrix pair $(H, N)$ is written as follows:

$$H = \bar{U} L_H LW^T, \tag{51}$$

$$N = \bar{V} L W^T \tag{52}$$

where $U$, $V$, and $W$ are matrices with orthonormal columns, whereas $L_H$ and $L$ are lower triangular matrices. Since $HR^{-1} = \bar{U} L_H \left((Q^T V)^T\right)$, we see that the rank $K$ of $H$ is displayed in the matrix $L_H$ in that the leading $K \times K$ submatrix of $L_H$ is well conditioned, and the bottom $(M-K) \times (M-K)$ submatrix of $L_H$ has small elements of size $\sigma_{K+1}$.

Moreover, if we partition $\bar{U}$ accordingly with $U$ in (5), then $U_1$ represents approximately the same space as $U_1$ and similarly with $U_2$ and $U_3$. The same is true for $\bar{V}$ and $V$. Hence, the ULLV decomposition in (51) and (52) yields essentially the same rank and subspace information as the QSVD does, and the approximate subspaces are typically very accurate.

The chief advantage of the ULLV decomposition is that it is much easier to update than the QSVD, as demonstrated in [14] (because it is much easier to maintain the triangular structure in $L_H$ and $L$ than the diagonal structure in $\Delta$ and $M$). Thus, we can sacrifice exact subspaces for approximate subspaces and gain easy updatable ability.

It is a subject of current research to implement the truncated QSVD algorithm by means of the ULLV decomposition and to compare it with the existing implementation.

IV. APPLICATION OF TRUNCATED QSVD IN SPEECH PROCESSING

Our truncated QSVD algorithm was programmed in Matlab, and the QSVD step was implemented along the lines described in [20]. All numerical experiments were carried out in floating point with a machine precision of approximately $10^{-16}$. The results presented in this section are from [11], where more details and results can be found.

The speech materials used in our experiments are four phonetically balanced sentences (see Table I) uttered by two male speakers and two female speakers. The recording of the speech was accomplished in a noise-isolated room at a loudness level typical of normal speech communication. The speech was filtered by a $3.6$ kHz lowpass filter, then sampled at $8$ kHz, and represented by $13$ bit words.

A. Order Determination of Noise-Free Speech

In a practical application of the truncated QSVD algorithm in speech processing, a precise rank determination is irrelevant.
TABLE II
PARSIMONIOUS ORDER OF SPEECH SEGMENTS UNDER THE
CONSTRAINT OF FIVE SNR THRESHOLDS. PARAMETER
CHOICE: MALE SPEAKER, \( N = 160 \) AND \( M = 20 \).

<table>
<thead>
<tr>
<th>Segment type</th>
<th>SNR threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10 dB</td>
</tr>
<tr>
<td>Voiced speech sound</td>
<td>5</td>
</tr>
<tr>
<td>Unvoiced speech sound</td>
<td>9</td>
</tr>
</tbody>
</table>

Since there is no physical reason for a specific rank to exist. Instead, we wish to truncate the QSVD for some properly chosen integer \( K \), and it is convenient to use the concept of “parsimonious order” defined in [4].

As a case study, we consider two typical speech segments containing voiced and unvoiced speech sounds, respectively. Both segments are taken from Sentence 1 uttered by a male speaker. The segment length \( N \) is set to 160 samples (20 ms), which is a standard length in many communication systems, and the number of columns \( M \) in the Hankel data matrix is chosen to be 20. This choice of \( M \) is based on experiments, and it results in a quite efficient implementation (the “skinnier” the matrices, the faster the algorithm) while not degrading the reconstructed speech significantly.

To study the effect of truncation on the reconstructed signal when applied to isolated segments of noise-free speech, we use two objective measures: SNR of reconstructed speech segments in Section IV-A-1, RMS log spectral distortion in Section IV-A-2, and informal listening tests in Section IV-A-3.

1) SNR of Reconstructed Speech Segments for Noise-Free Case: Consider the original speech segment \( x \) and the reconstructed speech segment \( \hat{x} \). The SNR of the reconstructed signal in decibels is then given by

\[
\text{SNR}(x, \hat{x}) = 10 \log \left( \frac{\|x\|^2}{\|x - \hat{x}\|^2} \right) \quad [\text{dB}].
\]  

(53)

To study the effect of truncation on the SNR of the reconstructed signal when applied to isolated segments of noise-free speech, successive reconstructions of isolated segments of voiced and unvoiced speech sounds were produced using the truncated QSVD algorithm (which for the no-noise case reduces to ordinary SVD) with increasing values of \( K \), and the SNR of the reconstructed speech segments have been calculated for each \( K \). For example, for \( K = 14 \), the SNR of the reconstructed segment of voiced and unvoiced speech sounds have been calculated to 28.2 and 19.3 dB, respectively.

Moreover, the parsimonious order \( K \) was determined to be the smallest \( K \) used in the truncated QSVD algorithm for which the SNR of the reconstructed speech segment exceeds a specified SNR threshold. Five SNR thresholds (10, 15, 20, 25, and 30 dB) were set resulting in five different parsimonious orders \( K \) for the segments of voiced and unvoiced speech sounds, respectively (see Table II). These results show that for a specified SNR threshold, the parsimonious order for voiced speech sounds is significantly lower than for unvoiced speech sounds.

2) Spectral Distances for Noise-Free Case: Consider the \( p \)-th order LPC model spectrum \( \| \hat{H}_p(e^{i\omega}) \| \) of the original speech segment \( \tilde{x} \) and the LPC model spectrum \( \| \hat{H}_p(e^{i\omega}) \| \) of the reconstructed speech segment \( \tilde{x} \). Define \( V(e^{i\omega}) \) as the log magnitude of the ratio between the two spectra versus frequency:

\[
V(e^{i\omega}) = \log |\hat{H}_p(e^{i\omega})| - \log |\hat{H}_p(e^{i\omega})|.
\]  

(54)

In our application, a natural choice for a distance between the two spectra is then given by

\[
d_2(|\hat{H}_p(e^{i\omega})|, |\hat{H}_p(e^{i\omega})|) = 20 \left[ \int_{-\omega_{0}}^{\omega_{0}} V(e^{i\omega})^2 \, d\omega \right]^{1/2} \quad [\text{dB}],
\]  

(55)

where \( \omega_0 = \frac{2 \pi}{40} \), i.e., the cut-off frequency of the lowpass filter used in connection with the A/D conversion. This measure is called the RMS log spectral distortion and is widely used in many speech processing systems.

The LPC model spectrum of the original and the reconstructed signal, respectively, have been computed for the isolated segments of voiced and unvoiced speech sounds, and the dissimilarity between them was measured by the RMS log spectral distortion.

Fig. 1 shows the 10th-order LPC model spectrum of the segment containing voiced speech sounds and the reconstructed
speech segment using 14 singular values. The reconstructed signal gives a spectrum close to the original, cf. Fig. 1(c). The RMS log spectral distortion was calculated to 0.79 dB.

The 10th-order LPC model spectrum of the segment containing unvoiced speech sound and the reconstructed speech segment using 14 singular values are shown in Fig. 2. Also here, the reconstructed signal gives a spectrum close to the original, cf. Fig. 2(c). The RMS log spectral distortion was calculated to 0.80 dB.

3) Informal Listening Test for Noise-Free Case: The final evaluation of the truncated QSVD algorithm must be done using continuous speech and later on with broadband noise added. Informal listening tests have been carried out for many sentences free from noise, demonstrating the capability of this reconstruction approach for removing redundancy. The tests showed that for $N = 160$, $M = 20$, and fixed values of $K \geq 14$, there are practically no difference between the quality of the original and the reconstructed speech signal. The quality is satisfactory for $K$ values down to 12.

B. Reconstruction of Speech Contaminated by Noise

We now focus on the recovery of speech corrupted by additive broadband noise. We consider the same two speech segments (voiced and unvoiced speech sounds) as in Section IV-A and noise with the following spectral density:

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

(56)

To study the effect of truncation on the reconstructed signal when applied to isolated segments of noisy speech, we will again use the SNR of reconstructed speech segments, spectral distances, and informal listening tests.

1) SNR of Reconstructed Speech Segments for Noise-Added Case: In connection with the LS estimate (8), the proper choice of $K$ is very important because a value of $K$ that is too low results in information loss, whereas a value that is too large leads to a matrix with unnecessary noise. For the MV estimate (20), it is also important not to choose a $K$ that is too small. On the other hand, once $K$ exceeds a certain value, the SNR of the reconstructed signal only decays slowly with $K$. This will be demonstrated shortly and is so because the retained singular values are corrected by applying the filter matrix $F_{MV}$, cf. (24).

First, we consider the segment with voiced speech sounds. Using the truncated QSVD algorithm with, respectively, the LS and MV estimate, for each $K$ between 1 and 19, we calculate the average SNR of the reconstructed speech segment using 100 noise realizations.

Fig. 3 shows plots of the average SNR of the reconstructed speech segment as a function of the retained number of singular values $K$ for voiced speech sounds contaminated with broadband noise; the SNR = 5 dB. We see from Fig. 3 that the optimal $K$ is 5 for LS and 7 for MV. The average SNR of the reconstructed speech segments becomes 7.3 and 7.8 dB, respectively. For $K = 14$, the average SNR is 5.7 dB for LS and 7.5 dB for MV. The LS curve has a clear maximum, which means that a proper choice of $K$ is important for this method. For the MV estimate, we see that the choice of $K$ is not so critical once $K$ is large enough. This means that the
choice of $K$ in the truncated QSD algorithm using the MV estimate does not affect the average SNR of the reconstructed signal seriously as long as $K$ is chosen sufficiently large.

The same experiments were repeated for the speech segment containing unvoiced speech sounds. The results are summarized below. Fig. 4 shows the plots of the average SNR of the reconstructed speech signal as a function of the retained number of singular values $K$ for unvoiced speech sounds corrupted by broadband noise; the SNR = 5. We see from Fig. 4 that the optimal $K$ is 11 for LS and 16 for MV. The average SNR of the reconstructed speech segments becomes 7.5 and 8.4 dB, respectively. For $K = 14$, the average SNR is 6.9 dB for LS and 8.3 dB for MV. Again, we see that the SNR of the reconstructed signal does not deteriorate seriously when the MV estimate is used.

Moreover, the local SNR's of reconstructed speech signals have been calculated and compared with local SNR's of original noisy speech signals. Fig. 5(a) shows the segment energy in decibels $20 \log ||x||_2$ for all 50 segments of Sentence 4 uttered by a male speaker, and Fig. 5(b) shows the local SNR’s of the noisy (global SNR of 5 dB) and the reconstructed speech ($K = 14$) using the MV estimate for the segments of the same sentence. We see from Fig. 5(b) that the local SNR's have been improved in most cases and that the great variations among the local SNR's of the various segments are reduced. The same observation was done in [4] for white noise using their algorithm.

2) Spectral Distances for Noise-Added Case: The behavior of the reconstructed segments in the frequency domain was examined as well. In most cases, we found that the model spectra of reconstructed speech from noisy segments of voiced speech sounds were closer (measured by the RMS log spectral distortion) to the model spectra of the noise-free speech than the noisy ones. Fig. 6 shows the LPC model spectrum of the speech segment containing voiced speech sounds together with a noisy speech segment; the SNR = 5 dB, and the reconstructed speech segment using 14 singular values. It is clearly seen that the performance of the LPC analysis used on the noisy speech segment is poor compared with the LPC analysis of the reconstructed speech segment—the LPC spectrum of the reconstructed speech segment matches the LPC spectrum of the noise-free speech segment much more closely in the regions near the formants. Notice that the highest formant is absent from the reconstructed speech segment. The same observation was done for many segments of voiced speech sounds where the local SNR was low.

For segments of unvoiced speech sounds, we found that the model spectra of noisy segments of unvoiced speech sounds were often closer to the model spectra of the noise-free speech than the reconstructed ones. In Fig. 7, the LPC model spectra of the original, the noisy, and the reconstructed speech segment containing unvoiced speech sounds are shown. The LPC analysis of the noisy speech appears to be a better match than the LPC analysis of the reconstructed speech.

At this point, something must be said about the signal level of speech segments containing voiced and unvoiced speech sounds, respectively. In general, the signal level is much lower in segments of unvoiced sounds than in segments of voiced sounds. Hence, in environments with even moderate SNR’s, the local SNR in the segments of unvoiced speech is excessively low. Independent of the algorithm, this makes their reconstruction practically impossible and introduces distortions.

3) Informal Listening Test for Noise-Added Case: Informal listening tests have been carried out for many sentences corrupted by additive broadband noise, demonstrating the inherent capacity of this signal subspace approach for removing general broadband noise from speech. Using the MV estimate and parameter settings of $(N, M, K) = (160, 20, 12)$ and $(N, M,
$K) = (160, 20, 14)$, the majority of the listeners declared that the processed speech was more intelligible. At higher noise levels (global SNR < 10 dB), some “musical noise” appears in the reconstructed speech without seriously affecting the overall intelligibility. This is a well-known phenomenon in most speech enhancement algorithms.

V. CONCLUSION

An approach for noise reduction of speech signals is presented and applied successfully to real speech embedded in noise. Our approach is based on signal subspaces, in contrast with previous algorithms based on spectral decomposition and adaptive filtering. Using a minimum-variance estimate of the signal-only data matrix (instead of a least-square estimate), we have found that computationally expensive methods for finding the number of retained singular values $K$ can be reduced to the problem of finding a proper choice of a fixed value of $K$.

The algorithm is formulated by means of the QSVD. With this formulation, the prewhitening becomes an integral part of the algorithm and not a separate step. This is essential in connection with updating issues in real-time recursive applications.

The truncated QSVD algorithm has a significant computational load, and for the time being, it is only useful as an off-line algorithm. We believe that a recursive implementation is possible using the rank-revealing ULLV decomposition, which is much easier to update than the QSVD. This is an object of current research.

ACKNOWLEDGMENT

The authors thank the reviewers for many suggestions that have greatly improved this paper’s overall presentation.

REFERENCES

Søren Holdt Jensen was born in Denmark in 1964, and he received the M.Sc. degree in electrical engineering from Aalborg University, Denmark, and the Ph.D. degree in electrical engineering from the Technical University of Denmark. He is currently a Research Fellow of the EC at Katholieke Universiteit Leuven, Belgium. Before joining the Electrical Engineering Department of Katholieke Universiteit Leuven, he was with the Telecommunications Laboratory of Telecom Denmark and the Electronics Institute of the Technical University of Denmark. He was also a Guest Scientist for short periods at Katholieke Universiteit Leuven and the Danish Computing Center for Research and Education (UNIC). His main research interests are in digital signal processing, numerical algorithms, and speech processing for mobile communication applications.

Per Christian Hansen was born in Denmark in 1957. He received the M.Sc. degree in electrical engineering in 1982 and the Ph.D. degree in numerical analysis in 1985, both from the Technical University of Denmark.

In 1985, he joined Copenhagen University’s Astronomical Observatory, and since 1988, he has been Associate Professor at UNIC, the Danish Computing Center for Research and Education. His primary research interest is numerical linear algebra and, in particular, numerical methods for rank deficient and ill-posed problems arising in such areas as signal processing and seismology. Dr. Hansen received the BIT award in 1990 and the Statel award in 1994.

Steffen Duus Hansen was born in Copenhagen, Denmark, on May 1, 1936. He received the M.Sc. and Ph.D. degrees in electrical engineering from the Technical University of Denmark in 1962 and 1972, respectively.

Early in his professional career, he was engaged in industrial research concerning television systems and measurement equipment. Later, he turned his attention to integrated circuits and to aids for the handicapped. He is now employed at the Technical University of Denmark as an associate professor. His present interests of research are digital signal processing, especially low rate speech coding and speech enhancement based on perceptual relationships. He is also involved in medical electronics.

John Aasted Sørensen received the M.Sc. and Ph.D. degrees from the Technical University of Denmark in 1975 and 1982, respectively.

He is currently employed at the Technical University of Denmark as an associate professor. His research interests are digital signal processing, parallel DSP architectures, DSP applications (speech, monitoring of systems), and pattern recognition.