A NEW BACKWARD RECURSION FOR THE MULTI-STAGE NESTED WIENER FILTER EMPLOYING KRYLOV SUBSPACE METHODS

M. Joham[1], *Y. Sun*[2], *M. D. Zoltowski*[3], *M. Honig*[2], *J. S. Goldstein*[4]

1. Technical University of Munich, Germany, e-mail: mijo@nws.ei.tum.de
2. Northwestern University, IL, e-mail: {yksun,mh}@ece.nwu.edu
3. Purdue University, IN, e-mail: mikedz@ecn.purdue.edu
4. SAIC, VA, e-mail: sgoldstein@trg1.saic.com

Abstract

We show that the *Multi-Stage Nested Wiener Filter* (MSNWF) can be identified to be the solution of the *Wiener-Hopf equation* in the *Krylov subspace* of the covariance matrix of the observation and the crosscorrelation vector of the observation and the desired signal. This understanding leads to the conclusion that the *Arnoldi algorithm* which arises from the MSNWF development can be replaced by the *Lanczos algorithm.* Thus, the computation of the underlying basis of the Krylov subspace can be simplified. Moreover, the foundation of the MSNWF in the Krylov subspace framework helps to derive an alternative formulation of the already presented MSNWF decomposition.
The new backward recursion is used to design a linear equalizer filter in an *Enhanced Data rates for GSM Evolution* (EDGE) system. Simulation results show the ability of the MSNWF to reduce the receiver complexity while the system performance is unchanged.

1. INTRODUCTION

The *Wiener filter* (WF) is a well known approach to estimate the unknown signal $d_0[n]$ from an observation $\boldsymbol{x}_0[n]$ and is optimal in the *Minimum Mean Square Error* (MMSE) sense.

Since the resulting filter depends upon the inverse of the covariance matrix of the observation $\boldsymbol{x}_0[n]$ and the needed filter length can be very large, if the observation $\boldsymbol{x}_0[n]$ is of high dimensionality, an alternative approach which operates in a reduced space is of great interest to reduce computational complexity and the number of observations needed to estimate the statistics.

The first approach to reduce the dimension of the estimation problem is the well established *Principal Component* (PC) method [1]. The observation signal is transformed by a matrix constituted by the eigenvectors belonging to the principal eigenvalues, thus, a truncated Karhunen-Loeve transform is applied. However, the PC method only takes into account the statistics of the observation signal and does not consider the relation to the desired signal. Therefore, Goldstein et. al. [2] introduced the *Cross-Spectral* (CS) metric which evolved from the *Generalized Sidelobe Canceller* (GSC) [3] and incorporates the similarity of the crosscorrelation vector of the observation and the desired signal with the respective eigenvector.

Recently, Goldstein et. al. presented the *Multi-Stage Nested Wiener Filter* (MSNWF) approach [4] which can be seen as a chain of GSCs. This fundamental contribution showed that the reduction of the dimension of the observations signal based on the eigenvectors as proposed by the PC and CS methods is suboptimum. The MSNWF does not need the eigenvectors of the covariance matrix of the observation signal and is, thus, computationally advantageous.

Our contribution is to show that the *Multi-Stage Nested Wiener Filter* (MSNWF) can be seen as the solution of the *Wiener-Hopf* equation in the *Krylov subspace* of the covariance matrix of the observation signal and the crosscorrelation vector of the observation and the desired signal. This conclusion follows from the results in [5] for the *Auxiliary Vector* (AV) method and especially in [6], but we present the consequences of this connection. First, the *Arnoldi algorithm* which is used to find the orthonormal basis for the *Krylov subspace* can be replaced by the *Lanczos algorithm* [7] since the covariance matrix is Hermitian. Second, we develop a new formulation of the MSNWF algorithm which gives an expression for the resulting *Mean Squared Error* (MSE), although it only works in the reduced dimensional space.

In the next section, we briefly discuss the reduced rank MSNWF and concentrate on the original MSNWF approach in Section 2 to motivate our reasoning in Section 4, where we show the close relationship between the MSNWF and *Krylov subspace* based methods. In Section 5 we present a new formulation of the MSNWF algorithm.

Throughout the paper the covariance matrix of a vector $\boldsymbol{x}[n]$ is denoted by $\boldsymbol{R}_{\boldsymbol{x}} = \mathrm{E}\{\boldsymbol{x}[n]\boldsymbol{x}^{\mathrm{H}}[n]\}$ and the correlation between a vector $\boldsymbol{x}[n]$ and a scalar $d[n]$ is $\boldsymbol{r}_{\boldsymbol{x},d} =$

$\mathrm{E}\{x[n]d^*[n]\}$.

2. MULTI-STAGE NESTED WIENER FILTER

The *Multi-Stage Nested Wiener Filter* (MSNWF) was developed by Goldstein et. al. [4] to find an approximate solution of the Wiener-Hopf equation which does not need the inverse or the eigenvalue decomposition of the covariance matrix. The approximation for the Wiener filter is found by stopping the recursive algorithm after D steps, hence, the approximation lies in a D-dimensional subspace of $\mathbb{C}^N$.

The first step of the MSNWF algorithm is to apply a full rank pre-filtering matrix of the form

$$\boldsymbol{T}_1 = [\boldsymbol{h}_1 \;\; \boldsymbol{B}_1] \in \mathbb{C}^{N\times N} \tag{1}$$

to get the new observation signal

$$\boldsymbol{z}_1[n] = \boldsymbol{T}_1^{\mathrm{H}}\boldsymbol{x}_0[n] = \begin{bmatrix} \boldsymbol{h}_1^{\mathrm{H}}\boldsymbol{x}_0[n] \\ \boldsymbol{B}_1^{\mathrm{H}}\boldsymbol{x}_0[n] \end{bmatrix} = \begin{bmatrix} d_1[n] \\ \boldsymbol{x}_1[n] \end{bmatrix} \in \mathbb{C}^N \tag{2}$$

which does not change the estimate $\hat{d}_0[n]$ as postulated in [4, 8]. The columns of $\boldsymbol{B}_1$ are chosen to be orthogonal to $\boldsymbol{h}_1$, therefore,

$$\boldsymbol{B}_1^{\mathrm{H}}\boldsymbol{h}_1 = \boldsymbol{0} \quad \text{or} \quad \boldsymbol{B}_1 = \mathrm{null}(\boldsymbol{h}_1^{\mathrm{H}}). \tag{3}$$

The intuitive choice for the first column $\boldsymbol{h}_1$ is the matched filter [4] or the vector which, when applied to $\boldsymbol{x}_0[n]$, gives a scalar signal $d_1[n]$ that has maximum correlation with the desired signal $d_0[n]$ [5, 8]:

$$\boldsymbol{h}_1 = \frac{\boldsymbol{r}_{\boldsymbol{x}_0,d_0}}{\|\boldsymbol{r}_{\boldsymbol{x}_0,d_0}\|_2} \quad \text{and} \quad \boldsymbol{r}_{\boldsymbol{x}_0,d_0} = \mathrm{E}\{\boldsymbol{x}_0[n]d_0^*[n]\} \in \mathbb{C}^N \tag{4}$$

Solving of the Wiener-Hopf equation of the new system and employing the inversion lemma for partitioned matrices (e. g. [9]) leads to the Wiener filter to estimate $d_0[n]$ from $\boldsymbol{z}_1[n]$:

$$\boldsymbol{w}_{\boldsymbol{z}_1} = \alpha_1 \begin{bmatrix} 1 \\ -\boldsymbol{R}_{\boldsymbol{x}_1}^{-1}\boldsymbol{r}_{\boldsymbol{x}_1,d_1} \end{bmatrix} \in \mathbb{C}^N, \tag{5}$$

where

$$\alpha_1 = \|\boldsymbol{r}_{\boldsymbol{x}_0,d_0}\|_2(\sigma_{d_1}^2 - \boldsymbol{r}_{\boldsymbol{x}_1,d_1}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{x}_1}^{-1}\boldsymbol{r}_{\boldsymbol{x}_1,d_1})^{-1} \tag{6}$$

and the variance of $d_1[n]$ reads as

$$\sigma_{d_1}^2 = \mathrm{E}\{|d_1[n]|^2\} = \boldsymbol{h}_1^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{x}_0}\boldsymbol{h}_1. \tag{7}$$

The expression in brackets in Equation (5) motivates the next step in the MSNWF development. Since $\boldsymbol{R}_{\boldsymbol{x}_1}^{-1}\boldsymbol{r}_{\boldsymbol{x}_1,d_1}$ is a Wiener filter, the new observation $\boldsymbol{x}_1[n]$ can be pre-filtered with $\boldsymbol{T}_2 = [\boldsymbol{h}_2 \;\; \boldsymbol{B}_2]$ again leading to a splitting into a matched filter and a Wiener filter of lower dimension. Performing this splitting $N-1$ times leads to a set of signals $d_i[n], i = 1, \ldots, N$ which are the outputs of the length N filters

$$\boldsymbol{t}_i = (\prod_{k=1}^{i-1}\boldsymbol{B}_k)\boldsymbol{h}_i \in \mathbb{C}^N. \tag{8}$$

This interpretation of the MSNWF will be used in the following sections. Note that all following stages are orthogonal to the first stage, i.e., $\boldsymbol{t}_1^{\mathrm{H}}\boldsymbol{t}_i = \delta_{1,i}$, i=2, ... ,N, and $\delta_{k,i}$ denotes the Kronecker delta function which is 1 for $k = i$ and 0 for $k \neq i$.

3. REDUCED RANK MSNWF

The reduced rank MSNWF of rank D is easily obtained by stopping the development of the MSNWF after $D-1$ steps and replacing the last Wiener filter $\boldsymbol{w}_{D-1}$ by the respective matched filter.

For $D < N$ we get the length D observation with a tri-diagonal covariance matrix [4, 8]

$$\boldsymbol{d}^{(D)}[n] = \boldsymbol{T}^{(D),\mathrm{H}}\boldsymbol{x}_0[n] \in \mathbb{C}^D, \tag{9}$$

where the superscript $(\bullet)^{(D)}$ indicates that we use a rank D approximation and the transformation matrix

$$\boldsymbol{T}^{(D)} = [\boldsymbol{t}_1, \ldots, \boldsymbol{t}_D] \in \mathbb{C}^{N\times D} \tag{10}$$

comprises the first D filters which were already defined in Equation (8). Therefore, we have to find the Wiener filter $\boldsymbol{w}_{\boldsymbol{d}}^{(D)}$ which estimates $d_0[n]$ from $\boldsymbol{d}^{(D)}[n]$. Obviously, this Wiener filter reads as

$$\boldsymbol{w}_{\boldsymbol{d}}^{(D)} = \left(\boldsymbol{T}^{(D),\mathrm{H}}\boldsymbol{R}_{\boldsymbol{x}_0}\boldsymbol{T}^{(D)}\right)^{-1}\boldsymbol{T}^{(D),\mathrm{H}}\boldsymbol{r}_{\boldsymbol{x}_0,d_0} \tag{11}$$

and the MSNWF rank D approximation of the Wiener filter $\boldsymbol{w}_0$ can be expressed as

$$\boldsymbol{w}_0^{(D)} = \boldsymbol{T}^{(D)}\boldsymbol{w}_{\boldsymbol{d}}^{(D)}. \tag{12}$$

4. MSNWF AND KRYLOV SUBSPACE

In the following we restrict the filters $\boldsymbol{t}_i$ to be orthogonal. Moreover, without loss of generality we assume that $\|\boldsymbol{t}_i\|_2 = 1$. Recall that the first column of the pre-filtering matrix $\boldsymbol{T}_i$ at step i was chosen to be the matched filter or the filter which maximizes the correlation to the output of the previous matched filter. Together with the orthogonality conditions this leads to following optimization:

$$\begin{aligned} \boldsymbol{t}_i &= \arg\max_{\boldsymbol{t}} \mathrm{E}\{\mathrm{Re}(d_i[n]d_{i-1}^*[n])\} \qquad \text{or} \\ \boldsymbol{t}_i &= \arg\max_{\boldsymbol{t}} \frac{1}{2}(\boldsymbol{t}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{x}_0}\boldsymbol{t}_{i-1} + \boldsymbol{t}_{i-1}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{x}_0}\boldsymbol{t}) \\ &\text{s.t.:} \quad \boldsymbol{t}^{\mathrm{H}}\boldsymbol{t} = 1 \quad \text{and} \quad \boldsymbol{t}^{\mathrm{H}}\boldsymbol{t}_k = 0, k = 1, \ldots, i-1. \end{aligned} \tag{13}$$

The result reads as (cf. [8])

$$\boldsymbol{t}_i = \frac{\left(\prod_{k=1}^{i-1} \boldsymbol{P}_k\right) \boldsymbol{R}_{\boldsymbol{x}_0} \boldsymbol{t}_{i-1}}{\| \left(\prod_{k=1}^{i-1} \boldsymbol{P}_k\right) \boldsymbol{R}_{\boldsymbol{x}_0} \boldsymbol{t}_{i-1} \|_2}, \tag{14}$$

where $\boldsymbol{P}_k$ denotes the projector onto the space orthogonal to $\boldsymbol{t}_k$, i.e.,

$$\boldsymbol{P}_k = \mathbf{1}_N - \boldsymbol{t}_k \boldsymbol{t}_k^{\mathrm{H}} \tag{15}$$

and $\mathbf{1}_N$ denotes the $N \times N$ identity matrix. Note that the recursive algorithm described in Equation (14) is the well known Gram-Schmidt *Arnoldi algorithm* [7]. The *Arnoldi recursion* is the basic algorithm to compute the orthonormal basis of the Krylov subspace $\mathcal{K}^{(D)}$ of the square matrix $\boldsymbol{A} \in \mathbb{C}^{M \times M}$ and the column vector $\boldsymbol{b} \in \mathbb{C}^M$. The Krylov subspace of dimension D is defined as follows [7]:

$$\mathcal{K}^{(D)} = \operatorname{span}\left(\left[\boldsymbol{b}, \boldsymbol{A}\boldsymbol{b}, \ldots, \boldsymbol{A}^{D-1}\boldsymbol{b}\right]\right). \tag{16}$$

Honig et. al. [10] made this observation and proved that for the choice $\boldsymbol{B}_i = \boldsymbol{P}_i$ the filters $\boldsymbol{t}_i$ are an orthonormal basis of the Krylov subspace. However, they did not see the fundamental implications of this result which can be found in following theorem [7].

Theorem 1 *If the columns of the matrix* $\boldsymbol{T}^D = [\boldsymbol{t}_1, \ldots, \boldsymbol{t}_D]$ *were computed using the recursion*

$$\boldsymbol{t}_i = \frac{\left(\prod_{k=i-1}^{1} \boldsymbol{P}_k\right) \boldsymbol{A} \boldsymbol{t}_{i-1}}{\| \left(\prod_{k=i-1}^{1} \boldsymbol{P}_k\right) \boldsymbol{A} \boldsymbol{t}_{i-1} \|_2}, \qquad \boldsymbol{t}_1 = \frac{\boldsymbol{b}}{\|\boldsymbol{b}\|_2}, \tag{17}$$

where $\boldsymbol{A} \in \mathbb{C}^{N \times N}$ *is an arbitrary square matrix and* $\boldsymbol{b} \in \mathbb{C}^N$ *is an arbitrary column vector, then the following equality holds:*

$$\boldsymbol{H}^{(D)} = \boldsymbol{T}^{(D),H} \boldsymbol{A} \boldsymbol{T}^{(D)}, \tag{18}$$

where $\boldsymbol{H}^{(D)}$ *is a* $D \times D$ *Hessenberg matrix.*

To see the value of Theorem 1 we need to specialize $\boldsymbol{A}$ to be Hermitian [7].

Corollary 1 *Given an Hermitian square matrix* $\boldsymbol{A} \in \mathbb{C}^{N \times N}$*, i.e.,* $\boldsymbol{A} = \boldsymbol{A}^H$*, the recursion in Equation (17) leads to a transformation matrix* $\boldsymbol{T}^{(D)}$ *which tri-diagonalizes* $\boldsymbol{A}$*, i.e.,*

$$\boldsymbol{H}^{(D)} = \boldsymbol{T}^{(D),H} \boldsymbol{A} \boldsymbol{T}^{(D)}.$$

is a Hermitian tri-diagonal matrix.

Our conclusion is that the MSNWF approach, although it is only motivated by statistical reasoning, is simply the solution of the Wiener-Hopf equation by employing the Krylov subspace of the matrix vector pair $(\boldsymbol{R}_{\boldsymbol{x}_0}, \boldsymbol{r}_{\boldsymbol{x}_0,d_0})$, if the filters $\boldsymbol{t}_i$ are orthogonal. Furthermore, if a reduced rank MSNWF with rank D is computed, this is fully equivalent to solving the Wiener-Hopf equation in the D-dimensional Krylov subspace $\mathcal{K}^{(D)}$. And more descriptively, the inverse of the covariance matrix $\boldsymbol{R}_{\boldsymbol{x}_0}^{-1}$ is approximated by a matrix polynomial $q(\boldsymbol{R}_{\boldsymbol{x}_0})$ of order $D-1$.

With Corollary 1 it is straightforward to understand that the covariance matrix $\boldsymbol{R}_{\boldsymbol{d}}^{(D)}$ of the pre-filtered observation $\boldsymbol{d}[n]$ (cf. Equation 9) is tri-diagonal, because the covariance matrix of the original observation $\boldsymbol{x}_0[n]$ is Hermitian. Moreover, the Hermitian property of $\boldsymbol{R}_{\boldsymbol{x}_0}$ allows to use the *Lanczos algorithm* [7] to compute the orthogonal basis $\boldsymbol{t}_i$ of the Krylov subspace $\mathcal{K}^{(D)}$ of $(\boldsymbol{R}_{\boldsymbol{x}_0}, \boldsymbol{r}_{\boldsymbol{x}_0,d_0})$:

$$\boldsymbol{t}_i = \frac{\boldsymbol{P}_{i-1}\boldsymbol{P}_{i-2}\boldsymbol{R}_{\boldsymbol{x}_0}\boldsymbol{t}_{i-1}}{\|\boldsymbol{P}_{i-1}\boldsymbol{P}_{i-2}\boldsymbol{R}_{\boldsymbol{x}_0}\boldsymbol{t}_{i-1}\|_2} \tag{19}$$

Note that the *Lanczos algorithm* reduces the complexity to compute the orthogonal basis. However, the *Lanczos algorithm* is sensitive to rounding errors, hence, the filters $\boldsymbol{t}_i$ are not orthogonal anymore for large i. In the sequel, we assume that the necessary rank D to find an good approximation of the Wiener filter is small enough to be able to apply the *Lanczos algorithm.*

5. A NEW MSNWF ITERATION

In this section, we develop a new algorithm which computes the rank D MSNWF, but works only within the Krylov subspace of dimension D. Therefore, we assume that the orthonormal basis $\boldsymbol{T}^{(D)} \in \mathbb{C}^{N \times D}$ of the Krylov subspace $\mathcal{K}^{(D)}$ was found by the *Arnoldi algorithm* (cf. Equation 14) or by the *Lanczos algorithm* (cf. Equation 19). The resulting rank D MSNWF is given in Equation (12) by the means of the Wiener filter $\boldsymbol{w}_{\boldsymbol{d}}^{(D)}$ which is applied to the pre-filtered observation

$$\boldsymbol{d}^{(D)}[n] = \boldsymbol{T}^{(D),\mathrm{H}} \boldsymbol{x}_0[n] \in \mathbb{C}^D, \tag{20}$$

with the tri-diagonal covariance matrix

$$\boldsymbol{R}_{\boldsymbol{d}}^{(D)} = \boldsymbol{T}^{(D),\mathrm{H}} \boldsymbol{R}_{\boldsymbol{x}_0} \boldsymbol{T}^{(D)} = \left[\begin{array}{cc|c} \multicolumn{2}{c|}{\boldsymbol{T}^{(D-1),\mathrm{H}} \boldsymbol{R}_{\boldsymbol{x}_0} \boldsymbol{T}^{(D-1)}} & \begin{matrix} \mathbf{0} \\ r_{D-1,D} \end{matrix} \\ \hline \mathbf{0}^{\mathrm{T}} & r_{D-1,D}^{*} & r_{D,D} \end{array}\right] \in \mathbb{C}^{D \times D} \tag{21}$$

and the crosscorrelation vector with respect to the desired signal $d_0[n]$

$$\boldsymbol{r}_{\boldsymbol{d},d_0}^{(D)} = \boldsymbol{T}^{(D),\mathrm{H}} \boldsymbol{r}_{\boldsymbol{x}_0,d_0} = \begin{bmatrix} \|\boldsymbol{r}_{\boldsymbol{x}_0,d_0}\|_2 \\ \mathbf{0} \end{bmatrix} \in \mathbb{R}^D. \tag{22}$$

Using knowledge of the covariance matrix $\boldsymbol{R}_{\boldsymbol{d}}^{(D-1)}$, the new entries of $\boldsymbol{R}_{\boldsymbol{d}}^{(D)}$ are simply

$$r_{D-1,D} = \boldsymbol{t}_{D-1}^{\mathrm{H}} \boldsymbol{R}_{\boldsymbol{x}_0} \boldsymbol{t}_D \quad \text{and} \quad r_{D,D} = \boldsymbol{t}_D^{\mathrm{H}} \boldsymbol{R}_{\boldsymbol{x}_0} \boldsymbol{t}_D. \tag{23}$$

Because $\boldsymbol{r}_{\boldsymbol{d},d_0}$ has the property that only the first element is not equal to 0, only the first column of the inverse of $\boldsymbol{R}_{\boldsymbol{d}}$ is needed to compute

$$\boldsymbol{w}_{\boldsymbol{d}}^{(D)} = \boldsymbol{R}_{\boldsymbol{d}}^{(D),-1} \boldsymbol{r}_{\boldsymbol{d},d_0}^{(D)} \in \mathbb{C}^D. \tag{24}$$

Consequently, we are only interested in the first column $\boldsymbol{c}_1^{(D)} \in \mathbb{C}^D$ of

$$\boldsymbol{C}^{(D)} = \boldsymbol{R}_{\boldsymbol{d}}^{(D),-1} = [\boldsymbol{c}_1^{(D)}, \ldots, \boldsymbol{c}_D^{(D)}] \in \mathbb{C}^{D\times D} \tag{25}$$

and the inversion lemma for partitioned matrices (e. g. [9, 11]) leads to

$$\boldsymbol{C}^{(D)} = \begin{bmatrix} \boldsymbol{C}^{(D-1)} & \boldsymbol{0} \\ \boldsymbol{0}^{\mathrm{T}} & 0 \end{bmatrix} + \beta_D^{-1} \boldsymbol{b}^{(D)} \boldsymbol{b}^{(D),\mathrm{H}}, \tag{26}$$

where the additional terms read as

$$\boldsymbol{b}^{(D)} = \begin{bmatrix} -\boldsymbol{C}^{(D-1)} \begin{bmatrix} \boldsymbol{0} \\ r_{D-1,D} \end{bmatrix} \\ 1 \end{bmatrix} = \begin{bmatrix} -r_{D-1,D}\boldsymbol{c}_{D-1}^{(D-1)} \\ 1 \end{bmatrix} \in \mathbb{C}^D \tag{27}$$

and

$$\beta_D = r_{D,D} - [\boldsymbol{0}^{\mathrm{T}}, r_{D-1,D}^*]\boldsymbol{C}^{(D-1)} \begin{bmatrix} \boldsymbol{0} \\ r_{D-1,D} \end{bmatrix} = r_{D,D} - |r_{D-1,D}|^2 c_{D-1,D-1}^{(D-1)} \tag{28}$$

with $c_{D-1,D-1}^{(D-1)}$ being the last element of the last column $\boldsymbol{c}_{D-1}^{(D-1)}$ of $\boldsymbol{C}^{(D)}$ at the previous step. Therefore, the new first column $\boldsymbol{c}_1^{(D)}$ can be written as

$$\boldsymbol{c}_1^{(D)} = \begin{bmatrix} \boldsymbol{c}_1^{(D-1)} \\ 0 \end{bmatrix} + \beta_D^{-1} c_{1,D-1}^{(D-1),*} \begin{bmatrix} |r_{D-1,D}|^2 \boldsymbol{c}_{D-1}^{(D-1)} \\ -r_{D-1,D}^* \end{bmatrix} \in \mathbb{C}^D, \tag{29}$$

where $c_{1,D-1}^{(D-1)}$ denotes the first element of $\boldsymbol{c}_{D-1}^{(D-1)}$. Obviously, the first column of $\boldsymbol{C}^{(D)}$ and, thus, the Wiener filter $\boldsymbol{w}_{\boldsymbol{d}}^{(D)}$ at step D depends upon the first column $\boldsymbol{c}_1^{(D-1)}$ at step $D-1$ and the new entries of the covariance matrix $r_{D-1,D}$ and $r_{D,D}$. However, we also observe a dependency on the previous last column $\boldsymbol{c}_{D-1}^{(D-1)}$. Hence, we have to find an expression for the last column of $\boldsymbol{C}^{(D)}$ and with Equation (26) we get

$$\boldsymbol{c}_D^{(D)} = \beta_D^{-1} \begin{bmatrix} -r_{D-1,D}\boldsymbol{c}_{D-1}^{(D-1)} \\ 1 \end{bmatrix} \tag{30}$$

which only depends on the previous last column and the new entries of $\boldsymbol{R}_{\boldsymbol{d}}^{(D)}$. So, we found an iteration that only updates two vectors $\boldsymbol{c}_1^{(D)}$ and $\boldsymbol{c}_D^{(D)}$ at each step and, moreover, the *mean squared error* at step D can be expressed with the first entry $c_{1,1}^{(D)}$ of $\boldsymbol{c}_1^{(D)}$ (cf. Equation **??**):

$$MSE^{(D)} = \sigma_{d_0}^2 - \|\boldsymbol{r}_{\boldsymbol{x}_0,d_0}\|_2^2 c_{1,1}^{(D)}. \tag{31}$$

The iteration in Equation (29) and (30) only operates with scalars and vectors. However, the scalars $r_{D-1,D}$ and $r_{D,D}$ (cf. Equation 23) are needed which are quadratic forms with the $N\times N$ covariance matrix $\boldsymbol{R}_{\boldsymbol{x}_0}$. But the matrix vector multiplication $\boldsymbol{R}_{\boldsymbol{x}_0}\boldsymbol{t}_i$ with $O(N^2)$ which can be found in the expression for $r_{i-1,i}$ and $r_{i,i}$ has already been used for the *Lanczos algorithm* in Equation (19) to find the orthonormal basis $\boldsymbol{T}^{(i)}$. Thus, it is worth to include the *Lanczos recursion* and the resulting algorithm is shown in Table 1, where we substituted $\boldsymbol{c}_1^{(i)}$ and $\boldsymbol{c}_i^{(i)}$ by $\boldsymbol{c}_{\text{first}}^{(i)}$ and $\boldsymbol{c}_{\text{last}}^{(i)}$, respectively. The resulting computational complexity for a rank D MSNWF is $O(N^2D)$, since a matrix vector multiplication with $O(N^2)$ has to be performed at each step.

choose desired MSE: σ_ε^2
choose maximum dimension: D
$\boldsymbol{t}_0 = \boldsymbol{0}, \quad \boldsymbol{t}_1 = \boldsymbol{r}_{\boldsymbol{x}_0,d_0}/\|\boldsymbol{r}_{\boldsymbol{x}_0,d_0}\|_2$
$\boldsymbol{u} = \boldsymbol{R}_{\boldsymbol{x}_0}\boldsymbol{t}_1$
$r_{0,1} = 0, \; r_{1,1} = \boldsymbol{t}_1^{\mathrm{H}}\boldsymbol{u}$
$c_{\text{first}}^{(1)} = r_{1,1}^{-1}, \quad c_{\text{last}}^{(1)} = r_{1,1}^{-1}$
$MSE^{(1)} = \sigma_{d_0}^2 - \|\boldsymbol{r}_{\boldsymbol{x}_0,d_0}\|_2^2 c_{\text{first}}^{(1)}$
$\Delta = 1$
for $i = 2, \ldots, D$
if $MSE^{(i)} < \sigma_\varepsilon^2$ then $\Delta = i-1$; break
$\boldsymbol{v} = \boldsymbol{u} - r_{i-1,i-1}\boldsymbol{t}_{i-1} - r_{i-2,i-1}\boldsymbol{t}_{i-2}$
$r_{i-1,i} = \|\boldsymbol{v}\|_2$
if $r_{i-1,i} = 0$ then $\Delta = i-1$; break
$\boldsymbol{t}_i = \boldsymbol{v}/r_{i-1,i}$
$\boldsymbol{u} = \boldsymbol{R}_{\boldsymbol{x}_0}\boldsymbol{t}_i$
$r_{i,i} = \boldsymbol{t}_i^{\mathrm{H}}\boldsymbol{u}$
$\beta_i = r_{i,i} - \|r_{i-1,i}\|^2 c_{\text{last},i-1}^{(i-1)}$
$\boldsymbol{c}_{\text{first}}^{(i)} = \begin{bmatrix} \boldsymbol{c}_{\text{first}}^{(i-1)} \\ 0 \end{bmatrix} + \beta_i^{-1} c_{\text{last},1}^{(i-1),*} \begin{bmatrix} \|r_{i-1,i}\|^2 \boldsymbol{c}_{\text{last}}^{(i-1)} \\ -r_{i-1,i}^* \end{bmatrix}$
$\boldsymbol{c}_{\text{last}}^{(i)} = \beta_i^{-1} \begin{bmatrix} -r_{i-1,i}\boldsymbol{c}_{\text{last}}^{(i-1)} \\ 1 \end{bmatrix}$
$MSE^{(i)} = \sigma_{d_0}^2 - \|\boldsymbol{r}_{\boldsymbol{x}_0,d_0}\|_2^2 c_{\text{first},1}^{(i)}$
$\boldsymbol{T}^{(D)} = [\boldsymbol{t}_1, \ldots, \boldsymbol{t}_\Delta]$
$\boldsymbol{w}_0^{(D)} = \|\boldsymbol{r}_{\boldsymbol{x}_0,d_0}\|_2 \boldsymbol{T}^{(D)} \boldsymbol{c}_{\text{first}}^{(\Delta)}$

Table 1. Lanczos MSNWF

Note that the algorithm in Table 1 is just a version of the *Conjugate Gradient algorithm* [12, 7]. In fact it is a direct version of the *Lanczos algorithm* [7] for linear systems.

6. REFERENCES

[1] H. Hotelling, "Analysis of a Complex of Statistical Variables into Principal Components," *Journal of Educational Psychology*, vol. 24, no. 6/7, pp. 417–441, 498–520, September/October 1933.

[2] J. S. Goldstein and I. S. Reed, "Subspace Selection for Partially Adaptive Sensor Array Processing," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 33, no. 2, pp. 539–543, April 1997.

[3] L. J. Griffiths and C. W. Jim, "An Alternative Approach to Linearly Constrained Adaptive Beamforming," *IEEE Transactions on Antennas and Propagation*, vol. 30, no. 1, January 1982.

[4] J. S. Goldstein, I. S. Reed, and L. L. Scharf, "A Multistage Representation of the Wiener Filter Based on Orthogonal Projections," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2943–2959, November 1998.

[5] D. A. Pados and S. N. Batlama, "Joint Space-Time Auxiliary-Vector Filtering for DS/CDMA Systems with Antenna Arrays," *IEEE Transactions on Communications*, vol. 47, no. 9, September 1999.

[6] M. L. Honig and J. S. Goldstein, "Adaptive Reduced-Rank Interference Suppression Based on the Multistage Wiener Filter," submitted to *IEEE Transactions on Communications*, March 2000.

[7] Y. Saad, *Iterative Methods for Sparse Linear Systems*, PWS – *out of print*, 1996, http://www-users.cs.umn.edu/ saad/books.html.

[8] M. Joham and M. D. Zoltowski, "Interpretation of the Multi-Stage Nested Wiener Filter in the Krylov Subspace Framework," Tech. Rep. TUM-LNS-TR-00-6, Munich University of Technology, November 2000.

[9] L. L. Scharf, *Statistical Signal Processing*, Addison-Wesley, 1991.

[10] M. L. Honig and W. Xiao, "Performance of Reduced-Rank Linear Interference Suppression for DS-CDMA," submitted to *IEEE Transactions on Information Theory*, December 1999.

[11] R. N. McDonough and A. D. Whalen, *Detection of Signals in Noise*, Academic Press, 1995.

[12] M. R. Hestenes and E. Stiefel, "Methods of Conjugate Gradients for Solving Linear Systems," *Journal of Research of the National Bureau of Standards*, vol. 49, no. 6, pp. 409–436, December 1952.