Interpretation of the Multi-Stage Nested Wiener Filter in the Krylov Subspace Framework

M. Joham and M. D. Zoltowski
School of Electrical Engineering,
Purdue University, West Lafayette, IN 47907-1285
e-mail: mikedz@ecn.purdue.edu

Abstract

In this paper, 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 settlement of the MSNWF in the Krylov subspace framework helps to derive an alternative formulation of the already presented MSNWF composition.

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. The WF is also optimal in the *Bayesian sense* if the signals $d_0[n]$ and $\boldsymbol{x}_0[n]$ are jointly Gaussian random variables (e. g. [Sch91, MW95]).

Therefore, the WF is employed in many applications because it is easily implemented and only relies on second order statistics which can be estimated with partial knowledge of the transmitted signal.

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 [Hot33, EY36]. 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. The Wiener Filter solution with respect to the new observation can be easily obtained since the covariance matrix of the new observation is a diagonal matrix with the principal eigenvalues as its entries.

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. [GR97b] introduced the *Cross-Spectral* (CS) metric which evolved from the *Generalized Sidelobe Canceller* (GSC) [AC76, GJ82] and incorporates the similarity of the crosscorrelation vector of the observation and the desired signal with the respective eigenvector. The CS method does not choose the principal eigenvectors, but the eigenvectors which belong to the largest CS metric as was also suggested by Byerly et. al. in [BR89].

Recently, Goldstein et. al. presented the *Multi-Stage Nested Wiener Filter* (MSNWF) approach [GR97a, GRS98] 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 (see [GGR99, GGDR98] for an illustrative example). This property led to the application of the MSNWF in detection problems [GRZ99] and CDMA interference cancellation [HG00]. Honig et. al. showed that the number of necessary MSNWF stages even for a heavy loaded CDMA system is small compared to the eigenspace based methods [HX99].

Pados et. al. presented a similar result in [PB99] for the *Auxiliary Vector* (AV) method which was also developed from the GSC. In fact the earlier AV publications [KBP98, PLB99] are an implementation of the GSC method. Their formulation in [PB99] is similar to [HG00] except that the combination of the transformed observation signal to form the estimate of the desired signal is suboptimal.

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 [PB99] and especially in [HG00], 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* [Saa96] 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. Schneider et. al. [SW99] presented a similar algorithm which also included the computation of the error variance in an image processing application. However, their formulation assumes only additive noise and a few dominant eigenvalues of the covariance matrix.

In the next section, we briefly review the theory of the Wiener filter to make the reader familiar with our notation. Before we discuss the reduced rank MSNWF in Section 4, we concentrate on the original MSNWF approach in

Section 3 to motivate our reasoning in Section 5, where we show the close relationship between the MSNWF and *Krylov subspace* based methods. In Section 6 we present a new formulation of the MSNWF algorithm.

2 Wiener Filter

Figure 1 shows the principle of a *Wiener Filter* (WF). The desired signal $d_0[n] \in \mathbb{C}$ which is assumed to be a zero-mean white Gaussian process is estimated by applying the linear filter $\boldsymbol{w} \in \mathbb{C}^N$ to the observation signal $\boldsymbol{x}_0[n] \in \mathbb{C}^N$ which is a multivariate zero-mean Gaussian process. The error of the estimation can be written as

$$\varepsilon_0[n] = d_0[n] - \hat{d}_0[n] = d_0[n] - \boldsymbol{w}^{\mathrm{H}}\boldsymbol{x}_0[n], \tag{1}$$

where $(\bullet)^{\mathrm{H}}$ denotes conjugate transpose. The variance of the estimation error is the *mean squared error*

$$\mathit{MSE}_0 = \mathrm{E}\{|\varepsilon_0|^2\} = \sigma_{d_0}^2 - \boldsymbol{w}^{\mathrm{H}}\boldsymbol{r}_{\boldsymbol{x}_0,d_0} - \boldsymbol{r}_{\boldsymbol{x}_0,d_0}^{\mathrm{H}}\boldsymbol{w} + \boldsymbol{w}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{x}_0}\boldsymbol{w}, \tag{2}$$

with the covariance matrix of the observation $\boldsymbol{x}_0[n]$

$$\boldsymbol{R}_{\boldsymbol{x}_0} = \mathrm{E}\{\boldsymbol{x}_0[n]\boldsymbol{x}_0^{\mathrm{H}}[n]\} \in \mathbb{C}^{N\times N}, \tag{3}$$

the variance of the desired signal $d_0[n]$

$$\sigma_{d_0}^2 = \mathrm{E}\{|d_0[n]|^2\}, \tag{4}$$

and the crosscorrelation between the desired signal $d_0[n]$ and the observation signal $\boldsymbol{x}_0[n]$

$$\boldsymbol{r}_{\boldsymbol{x}_0,d_0} = \mathrm{E}\{\boldsymbol{x}_0[n]d_0^*[n]\} \in \mathbb{C}^N, \tag{5}$$

where $(\bullet)^*$ denotes complex conjugate.

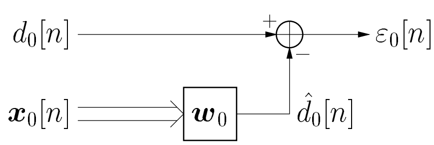

Figure 1: Wiener Filter

The Wiener filter $\boldsymbol{w}_0$ is the filter $\boldsymbol{w}$ which minimizes the *mean squared error*, thus,

$$\boldsymbol{w}_0 = \arg\min_{\boldsymbol{w}} \mathit{MSE}_0. \tag{6}$$

This criterion leads to the Wiener-Hopf equation

$$\boldsymbol{R}_{\boldsymbol{x}_0}\boldsymbol{w}_0 = \boldsymbol{r}_{\boldsymbol{x}_0,d_0} \tag{7}$$

and the Wiener filter

$$\boldsymbol{w}_0 = \boldsymbol{R}_{\boldsymbol{x}_0}^{-1}\boldsymbol{r}_{\boldsymbol{x}_0,d_0} \in \mathbb{C}^N. \tag{8}$$

Consequently, the *minimum mean squared error* (MMSE) can be written as

$$MMSE_0 = \sigma_{d_0}^2 - \boldsymbol{r}_{\boldsymbol{x}_0,d_0}^{\mathrm{H}} \boldsymbol{R}_{\boldsymbol{x}_0}^{-1} \boldsymbol{r}_{\boldsymbol{x}_0,d_0}. \tag{9}$$

Before we focus on the *Multi-Stage Nested Wiener Filter* (MSNWF) in the next section, we have to recite the following theorem (e. g. [GRS98]) which is proved in Appendix A.

Theorem 1 *If the observation $\boldsymbol{x}_0[n]$ to estimate $d_0[n]$ is pre-filtered by a full-rank matrix $\boldsymbol{T} \in \mathbb{C}^{N\times N}$, i. e., $\boldsymbol{z}_1[n] = \boldsymbol{T}\boldsymbol{x}_0[n]$, the Wiener filter $\boldsymbol{w}_{\boldsymbol{z}_1}$ to estimate $d_0[n]$ from $\boldsymbol{z}_1[n]$ leads to the same estimate $\hat{d}_0[n]$, thus, the MMSE is unchanged.*

This Theorem can be generalized to a bigger set of pre-filtering matrices by the following corollary.

Corollary 1 *If the pre-filtering matrix $\boldsymbol{T} \in \mathbb{C}^{M\times N}, M \geq N$ has full column rank, i. e. $\boldsymbol{T}$ has a left inverse, the resulting estimate $\hat{d}_0[n]$ and the MMSE are unchanged.*

The proof of Corollary 1 is straightforward, the inverse of $\boldsymbol{T}$ just has to be replaced by the left inverse of $\boldsymbol{T}$ in Appendix A. However, if $\boldsymbol{T}$ is tall, i. e., $M > N$, then the Wiener filter solution is ambiguous, because a vector which is orthogonal to the column space of $\boldsymbol{T}$ can be added to $\boldsymbol{w}_{\boldsymbol{z}_1}$ in Equation (58) without changing the estimate $\hat{d}_0[n]$ and the MMSE.

3 Multi-Stage Nested Wiener Filter

The *Multi-Stage Nested Wiener Filter* (MSNWF) was developed by Goldstein et. al. [GR97a, GRS98] to find an approximate solution of the Wiener-Hopf equation (cf. Equation 7) 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

$$\boldsymbol{T}_1 = \begin{bmatrix} \boldsymbol{h}_1^{\mathrm{H}} \\ \boldsymbol{B}_1 \end{bmatrix} \in \mathbb{C}^{N\times N} \tag{10}$$

to get the new observation signal

$$\boldsymbol{z}_1 = \boldsymbol{T}_1 \boldsymbol{x}_0[n] = \begin{bmatrix} \boldsymbol{h}_1^{\mathrm{H}} \boldsymbol{x}_0[n] \\ \boldsymbol{B}_1 \boldsymbol{x}_0[n] \end{bmatrix} = \begin{bmatrix} d_1[n] \\ \boldsymbol{x}_1[n] \end{bmatrix} \in \mathbb{C}^N \tag{11}$$

which does not change the estimate $\hat{d}_0[n]$ as postulated in Theorem 1. The rows of $\boldsymbol{B}_1$ are chosen to be orthogonal to $\boldsymbol{h}_1^{\mathrm{H}}$, therefore,

$$\boldsymbol{B}_1 \boldsymbol{h}_1 = \boldsymbol{0} \qquad \text{or} \quad \boldsymbol{B}_1 = \mathrm{null}(\boldsymbol{h}_1^{\mathrm{H}})^{\mathrm{H}}. \tag{12}$$

The intuitive choice for the first row $\boldsymbol{h}_1^{\mathrm{H}}$ is 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 $d_0[n]$. Without loss of generality we assume that $\|\boldsymbol{h}_1\|_2 = 1$ and force $d_1[n]$ to be 'in-phase' with $d_0[n]$, i. e. the correlation between $d_0[n]$ and $d_1[n]$ is real, which is motivated by the trivial case when $d_1[n] = d_0[n]$. Note that this criterion is different from the one proposed in [GGDR98] since we do not optimize the absolute value of the correlation which would lead to a loss of information about the phase and, thus, Goldstein et. al. had to use Schwarz inequality to derive the same result. The formulation in [PB99, PLB99] is also based on the absolute value of the correlation, but in the proof of the result the correlation is *assumed* to be real. With our consideration we end up with following optimization problem

$$\boldsymbol{h}_1 = \arg\max_{\boldsymbol{h}} \mathrm{E}\{\mathrm{Re}(d_1[n]d_0^*[n])\} \qquad \text{or}$$

$$\boldsymbol{h}_1 = \arg\max_{\boldsymbol{h}} \frac{1}{2}(\boldsymbol{h}^{\mathrm{H}}\boldsymbol{r}_{\boldsymbol{x}_0,d_0} + \boldsymbol{r}_{\boldsymbol{x}_0,d_0}^{\mathrm{H}}\boldsymbol{h}) \qquad \text{s.t.:} \quad \boldsymbol{h}^{\mathrm{H}}\boldsymbol{h} = 1 \tag{13}$$

which results in the normalized matched filter

$$\boldsymbol{h}_1 = \frac{\boldsymbol{r}_{\boldsymbol{x}_0,d_0}}{\|\boldsymbol{r}_{\boldsymbol{x}_0,d_0}\|_2} \in \mathbb{C}^N. \tag{14}$$

Note that $d_1[n]$ contains all information about $d_0[n]$ which can be found in $\boldsymbol{x}_0[n]$ since $d_0[n]$ is a scalar, thus, the information about $d_0[n]$ lies in a one-dimensional subspace of $\mathbb{C}^N$ and with the matched filter we pick out only these portions of $\boldsymbol{x}_0[n]$ which are included in this subspace. Moreover, the second part of $\boldsymbol{z}_1[n]$, namely $\boldsymbol{x}_1[n]$, does not contain any information about $d_0[n]$ because of the orthogonality condition in Equation (12).

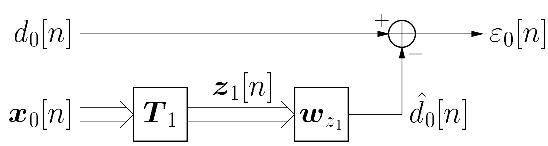

Figure 2: Wiener Filter with Pre-Filtering

Again, we have to solve the Wiener-Hopf equation of the new system depicted in Figure 2 to get

$$\boldsymbol{w}_{\boldsymbol{z}_1} = \boldsymbol{R}_{\boldsymbol{z}_1}^{-1}\boldsymbol{r}_{\boldsymbol{z}_1,d_0} \in \mathbb{C}^N, \tag{15}$$

where the covariance matrix of $\boldsymbol{z}_1[n]$ can be expressed by the statistics of $d_1[n]$ and $\boldsymbol{x}_1[n]$, i.e.,

$$\boldsymbol{R}_{\boldsymbol{z}_1} = \begin{bmatrix} \sigma_{d_1}^2 & \boldsymbol{r}_{\boldsymbol{x}_1,d_1}^{\mathrm{H}} \\ \boldsymbol{r}_{\boldsymbol{x}_1,d_1} & \boldsymbol{R}_{\boldsymbol{x}_1} \end{bmatrix} \in \mathbb{C}^{N\times N}, \tag{16}$$

with the variance of $d_1[n]$

$$\sigma_{d_1}^2 = \mathrm{E}\{|d_1[n]|^2\} = \boldsymbol{h}_1^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{x}_0}\boldsymbol{h}_1, \tag{17}$$

the crosscorrelation vector between $\boldsymbol{x}_1[n]$ and $d_1[n]$

$$\boldsymbol{r}_{\boldsymbol{x}_1,d_1} = \mathrm{E}\{\boldsymbol{x}_1[n]d_1^*[n]\} = \boldsymbol{B}_1\boldsymbol{R}_{\boldsymbol{x}_0}\boldsymbol{h}_1 \in \mathbb{C}^{N-1} \tag{18}$$

and the covariance matrix of $\boldsymbol{x}_1[n]$

$$\boldsymbol{R}_{\boldsymbol{x}_1} = \mathrm{E}\{\boldsymbol{x}_1[n]\boldsymbol{x}_1^{\mathrm{H}}[n]\} = \boldsymbol{B}_1\boldsymbol{R}_{\boldsymbol{x}_0}\boldsymbol{B}_1^{\mathrm{H}} \in \mathbb{C}^{N-1\times N-1}. \tag{19}$$

The crosscorrelation vector of the pre-filtered observation signal $\boldsymbol{z}_1[n]$ and the desired signal $d_1[n]$ reveals the usefulness of the choice of pre-filtering:

$$\boldsymbol{r}_{\boldsymbol{z}_1,d_0} = \boldsymbol{T}_1\boldsymbol{r}_{\boldsymbol{x}_0,d_0} = \|\boldsymbol{r}_{\boldsymbol{x}_0,d_0}\|_2\boldsymbol{e}_1 \in \mathbb{R}^N, \tag{20}$$

where $\boldsymbol{e}_i$ denotes a unit norm vector with a one at the i-th position. Therefore, the Wiener filter $\boldsymbol{w}_{\boldsymbol{z}_1}$ of the pre-filtered signal $\boldsymbol{z}_1[n]$ is just a weighted version of the first column of the inverse of the covariance matrix $\boldsymbol{R}_{\boldsymbol{z}_1}$ in Equation (16). After applying the inversion lemma for partitioned matrices (e. g. [Sch91, MW95]) we end up with [GR97a, GRS98]

$$\boldsymbol{w}_{\boldsymbol{z}_1} = \alpha_1 \begin{bmatrix} 1 \\ -\boldsymbol{R}_{\boldsymbol{x}_1}^{-1}\boldsymbol{r}_{\boldsymbol{x}_1,d_1} \end{bmatrix} \in \mathbb{C}^N, \tag{21}$$

where

$$\alpha_1 = \|\boldsymbol{r}_{\boldsymbol{x}_0,d_0}\|_2(\sigma_{d_1}^2 - \boldsymbol{r}_{\boldsymbol{x}_1,d_1}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{x}_1}^{-1}\boldsymbol{r}_{\boldsymbol{x}_1,d_1})^{-1}. \tag{22}$$

Equation (21) is the key equation to understand most interpretations of the MSNWF.

The first and most important observation in Equation (21) is that the vector in brackets, when applied to $\boldsymbol{z}_1[n]$, gives the error signal $\varepsilon_1[n]$ of the Wiener filter which estimates $d_1[n]$ from $\boldsymbol{x}_1[n]$:

$$\varepsilon_1[n] = d_1[n] - \hat{d}_1[n] = d_1[n] - \boldsymbol{w}_1^{\mathrm{H}}\boldsymbol{x}_1[n] = \left[1, -\boldsymbol{w}_1^{\mathrm{H}}\right]\boldsymbol{z}_1[n] \tag{23}$$

with the Wiener filter

$$\boldsymbol{w}_1 = \boldsymbol{R}_{\boldsymbol{x}_1}^{-1}\boldsymbol{r}_{\boldsymbol{x}_1,d_1} \in \mathbb{C}^{N-1}. \tag{24}$$

This observation immediately leads to the next step in the MSNWF development. In the second step, the output of the Wiener filter $\boldsymbol{w}_1$ with dimension $N-1$ can be replaced by the weighted error signal $\varepsilon_2[n]$ of a Wiener filter which estimates the output signal $d_2[n]$ of the matched filter $\boldsymbol{h}_2$ from the blocking-matrix output $\boldsymbol{x}_2[n] = \boldsymbol{B}_2\boldsymbol{x}_1[n]$.
Moreover, this observation shows the close relationship of the MSNWF to the *Generalized Sidelobe Canceller* (GSC, [AC76, GJ82, GRZ99]). The GSC can be interpreted as a MSNWF after the first step.

Second, the factor α_1 is a scalar Wiener filter to estimate $d_0[n]$ from the scalar $\varepsilon_1[n]$. The covariance matrix or variance of $\varepsilon_1[n]$ is the MMSE of the Wiener filter $\boldsymbol{w}_1$ and the crosscorrelation between the scalar observation signal $\varepsilon_1[n]$ and the desired signal $d_0[n]$ is the norm of the matched filter $\boldsymbol{r}_{\boldsymbol{x}_0,d_0}$, thus,

$$\alpha_1 = \sigma_{\varepsilon_1}^{-1} r_{\varepsilon_1,d_0} = (\sigma_{d_1}^2 - \boldsymbol{r}_{\boldsymbol{x}_1,d_1}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{x}_1}^{-1}\boldsymbol{r}_{\boldsymbol{x}_1,d_1})^{-1}\|\boldsymbol{r}_{\boldsymbol{x}_0,d_0}\|_2. \tag{25}$$

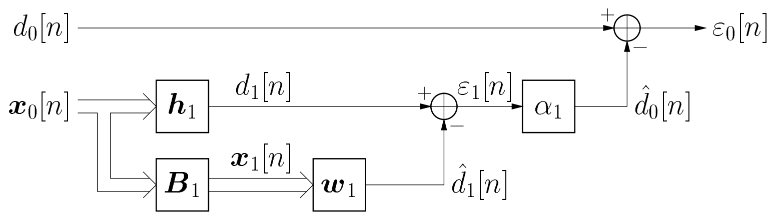

Figure 3: MSNWF after the First Step

These two interpretations lead to the well known structure in Figure 3. The Wiener filter $\boldsymbol{w}_0$ is replaced by the scalar Wiener filter α_1 which estimates $d_0[n]$ from the error signal $\varepsilon_1[n]$ of the vector Wiener filter $\boldsymbol{w}_1$.

Third, the MSNWF can be created without knowledge of the covariance matrix $\boldsymbol{R}_{\boldsymbol{x}_0}$ and, therefore, the implementation of a MSNWF is simplified, because at each step the Wiener filter is replaced by the normalized matched filter and the next stage. Since the matched filter is simply the crosscorrelation between the new observation $\boldsymbol{x}_i[n]$ and the new desired signal $d_i[n]$ at each step only an estimation of this crosscorrelation is needed. Note, however, that estimating all the crosscorrelations and estimating the covariance matrix lead to the same resulting estimate $\hat{d}_0[n]$ at the same expense.

Fourth, it is straightforward to see that each new desired signal $d_i[n], i = 1, \ldots, N$, is the output of a length N filter

$$\boldsymbol{t}_i = (\prod_{k=1}^{i-1} \boldsymbol{B}_k^{\mathrm{H}})\boldsymbol{h}_i \in \mathbb{C}^N \tag{26}$$

as depicted in Figure 4. 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 $k \neq i$. However, the filters $\boldsymbol{t}_i$ are not an orthogonal basis of $\mathbb{C}^N$ in general (cf. Section 5).

The *Auxiliary Vector* method [KBP98] also evolved from GSC considerations, but ended the iteration after the second step and is, thus, a rank two MSNWF approximation of the ideal Wiener filter. The extension of the AV method presented in [PB99] is basically the structure shown in Figure 4, but the proposed choice of the scalar filters α_i is suboptimal.

The very important property of the MSNWF is that the pre-filtered observation vector

$$\boldsymbol{d}[n] = [d_1[n], \ldots, d_N[n]]^{\mathrm{T}}, \tag{27}$$

where $(\bullet)^{\mathrm{T}}$ denotes transpose, has a tri-diagonal covariance matrix [GRS98]. This can be understood with the help of Figures 3 and 4. The matched filter $\boldsymbol{h}_i$ is designed to retrieve all information of $d_{i-1}[n]$ that can be found in $\boldsymbol{x}_{i-1}[n]$. Therefore, the output of $\boldsymbol{h}_i$, $d_i[n]$, is correlated with $d_{i-1}[n]$ and also with $d_{i+1}[n]$, because $\boldsymbol{h}_{i+1}$ is the matched filter to find $d_i[n]$. But $d_{i+1}[n]$ includes

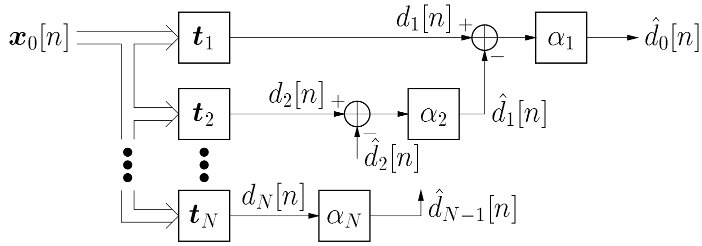

Figure 4: MSNWF as a Filter Bank

no information about $d_{i-1}[n]$, since the input of $\boldsymbol{h}_{i+1}$ was pre-filtered by the blocking matrix $\boldsymbol{B}_{i+1}$. Consequently, $d_i[n]$ is only correlated to its neighbours $d_{i-1}[n]$ and $d_{i+1}[n]$ leading to a tri-diagonal covariance matrix.

The remaining part of the MSNWF (summers and scalar Wiener filters α_i) is a Wiener filter which estimates $d_0[n]$ of $\boldsymbol{d}[n]$. Because only $d_1[n]$ is correlated with $d_0[n]$, the crosscorrelation vector is simply a weighted version of $\boldsymbol{e}_1$, thus, again only the first column of the inverse of the covariance matrix of $\boldsymbol{d}[n]$ is of interest. The structure with summers and scalar Wiener filters follows from the tri-diagonal property of the covariance matrix of $\boldsymbol{d}[n]$.

4 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. Thus, the MSNWF of rank 1 is simply the matched filter $\boldsymbol{r}_{\boldsymbol{x}_0,d_0}$.

We restrict ourselves to the MSNWF interpretation shown in Figure 4, hence, for $D < N$ we get the length D observation with a tri-diagonal covariance matrix

$$\boldsymbol{d}^{(D)}[n] = \boldsymbol{T}^{(D),\mathrm{H}} \boldsymbol{x}_0[n] \in \mathbb{C}^D, \tag{28}$$

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

$$\boldsymbol{T}^{(D)} = [\boldsymbol{t}_1, \ldots, \boldsymbol{t}_D] \in \mathbb{C}^{N \times D} \tag{29}$$

comprises the first D filters which were already defined in Equation (26). 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

$$\boldsymbol{w}_{\boldsymbol{d}}^{(D)} = \left(\boldsymbol{R}_{\boldsymbol{d}}^{(D)}\right)^{-1} \boldsymbol{r}_{\boldsymbol{d},d_0}^{(D)} = \left(\boldsymbol{T}^{(D),\mathrm{H}} \boldsymbol{R}_{\boldsymbol{x}_0} \boldsymbol{T}^{(D)}\right)^{-1} \boldsymbol{T}^{(D),\mathrm{H}} \boldsymbol{r}_{\boldsymbol{x}_0,d_0} \tag{30}$$

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

$$\boldsymbol{w}_0^{(D)} = \boldsymbol{T}^{(D)} \boldsymbol{w}_{\boldsymbol{d}}^{(D)} = \boldsymbol{T}^{(D)} \left(\boldsymbol{T}^{(D),\mathrm{H}} \boldsymbol{R}_{\boldsymbol{x}_0} \boldsymbol{T}^{(D)}\right)^{-1} \boldsymbol{T}^{(D),\mathrm{H}} \boldsymbol{r}_{\boldsymbol{x}_0,d_0} \tag{31}$$

and has following *mean squared error*:

$$MSE^{(D)} = \sigma_{d_0}^2 - \boldsymbol{r}_{\boldsymbol{x}_0,d_0}^{\mathrm{H}} \boldsymbol{T}^{(D)} \left(\boldsymbol{T}^{(D),\mathrm{H}} \boldsymbol{R}_{\boldsymbol{x}_0} \boldsymbol{T}^{(D)} \right)^{-1} \boldsymbol{T}^{(D),\mathrm{H}} \boldsymbol{r}_{\boldsymbol{x}_0,d_0}. \tag{32}$$

5 MSNWF and Krylov Subspace

In Section 3 we stated that the filters $\boldsymbol{t}_i$ in Figure 4 are not orthogonal in general. It can be easily shown that they are orthogonal for the special choice of the blocking matrices $\boldsymbol{B}_i$ to have the property that all singular values unequal to 0 are the same, i. e., $\sigma_1 = \ldots = \sigma_{N-i} = \sigma$ and $\sigma_k = 0, k > N - i$.

In the following we restrict ourselves to this class of blocking matrices, therefore, the filters $\boldsymbol{t}_i$ are orthogonal. Moreover, without loss of generality we assume that $\|\boldsymbol{t}_i\|_2 = 1$.

Now, recall the MSNWF development (cf. Figure 3). At step i the signal $\boldsymbol{x}_{i-1}[n]$ at the blocking matrix output of the previous stage was the input for the matched filter $\boldsymbol{h}_i$ and the blocking matrix $\boldsymbol{B}_i$. Again, the matched filter was chosen, because its output $d_i[n]$ has the maximum correlation with the output of the previous stage $d_{i-1}[n]$. Because we can substitute the chain of blocking matrices and the matched filter $\boldsymbol{h}_i \in \mathbb{C}^{N-i+1}$ by a filter $\boldsymbol{t}_i \in \mathbb{C}^N$ (cf. Equation 26) and since we restrict ourselves to use orthonormal filters $\boldsymbol{t}_i$, we can compute the filters $\boldsymbol{t}_i$ directly. At the i-th step we get the additional output signal $d_i[n] = \boldsymbol{t}_i^{\mathrm{H}} \boldsymbol{x}_0[n]$ which has to be maximally correlated with the output signal of the previous stage $d_{i-1}[n] = \boldsymbol{t}_{i-1}^{\mathrm{H}} \boldsymbol{x}_0[n]$. Together with the orthogonality conditions this leads to following optimization:

$$\begin{aligned} &\boldsymbol{t}_i = \arg\max_{\boldsymbol{t}} \mathrm{E}\{\mathrm{Re}(d_i[n] d_{i-1}^*[n])\} \qquad \text{or} \\ &\boldsymbol{t}_i = \arg\max_{\boldsymbol{t}} \frac{1}{2} (\boldsymbol{t}^{\mathrm{H}} \boldsymbol{R}_{\boldsymbol{x}_0} \boldsymbol{t}_{i-1} + \boldsymbol{t}_{i-1}^{\mathrm{H}} \boldsymbol{R}_{\boldsymbol{x}_0} \boldsymbol{t}) \\ &\text{s.t.:} \quad \boldsymbol{t}^{\mathrm{H}} \boldsymbol{t} = 1 \quad \text{and} \quad \boldsymbol{t}^{\mathrm{H}} \boldsymbol{t}_k = 0, k = 1, \ldots, i-1. \end{aligned} \tag{33}$$

The result which is easily obtained (e.g. with Langrange multipliers, cf. Appendix B) reads as

$$\boldsymbol{t}_i = \frac{\left(\prod_{k=i-1}^{1} \boldsymbol{P}_k \right) \boldsymbol{R}_{\boldsymbol{x}_0} \boldsymbol{t}_{i-1}}{\| \left(\prod_{k=i-1}^{1} \boldsymbol{P}_k \right) \boldsymbol{R}_{\boldsymbol{x}_0} \boldsymbol{t}_{i-1} \|_2}, \tag{34}$$

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

$$\boldsymbol{P}_k = \mathbf{1}_N - \boldsymbol{t}_k \boldsymbol{t}_k^{\mathrm{H}} \tag{35}$$

and $\mathbf{1}_N$ denotes the $N \times N$ identity matrix.

The expression in Equation (34) implies that the best choice for the blocking matrix $\boldsymbol{B}_i$ is the projector $\boldsymbol{P}_i$ onto the orthogonal space of the matched filter $\boldsymbol{t}_i$, if the filters $\boldsymbol{t}_i$ are orthogonal to each other. This result suggests that the reduction of the dimension of the solution space at each step of the MSNWF iteration does not necessarily lead to a reduction of the length of the filter as proposed by Goldstein et. al. in [GRS98]. This result is consistent since Corollary 1 allows to use a tall pre-filtering matrix $\boldsymbol{T}_i$ at each step. Honig et. al. [HX99, HG00] already developed the same result, but they gave no motivation for this special choice of $\boldsymbol{B}_i$. Also Pados et. al. [PB99, LUN99] ended up with a similar expression for the *Auxiliary Vector* (AV) method, but did not consider the orthogonality of the filters $\boldsymbol{t}_i$ which they incorporated into the optimization, leading to a more complicated expression.

However, both contributions did not make the observation that the recursive algorithm described in Equation (34) is the well known Gram-Schmidt *Arnoldi algorithm* [Arn51, Saa96]. 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 [Saa96, vdV00]:

$$\mathcal{K}^{(D)} = \operatorname{span}\left([\boldsymbol{b}, \boldsymbol{A}\boldsymbol{b}, \ldots, \boldsymbol{A}^{D-1}\boldsymbol{b}]\right). \tag{36}$$

Again, also Honig et. al. [HX99] 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 excavate the fundamental implications of this result which can be found in following theorem [Arn51, Saa96].

Theorem 2 *If the columns of the matrix $\boldsymbol{T}^D = [\boldsymbol{t}_1, \ldots, \boldsymbol{t}_D]$ were computed using the recursion*

$$\boldsymbol{t}_i = \frac{\left(\prod_{k=i-1}^{1} \boldsymbol{P}_k\right) \boldsymbol{A}\boldsymbol{t}_{i-1}}{\|\left(\prod_{k=i-1}^{1} \boldsymbol{P}_k\right) \boldsymbol{A}\boldsymbol{t}_{i-1}\|_2}, \qquad \boldsymbol{t}_1 = \frac{\boldsymbol{b}}{\|\boldsymbol{b}\|_2}, \tag{37}$$

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:

$$\boldsymbol{A}\boldsymbol{T}^{(D)} = \boldsymbol{T}^{(D)}\boldsymbol{H}^{(D)} + h_{D+1,D}\boldsymbol{t}_{D+1}\boldsymbol{e}_D^T, \tag{38}$$

where $\boldsymbol{H}^{(D)}$ is a $D \times D$ Hessenberg matrix and $\boldsymbol{t}_{D+1}$ denotes the next vector of the recursion. Obviously, since the vectors $\boldsymbol{t}_i$ are orthonormal, Equation (38) can be rewritten to get

$$\boldsymbol{H}^{(D)} = \boldsymbol{T}^{(D),H}\boldsymbol{A}\boldsymbol{T}^{(D)}. \tag{39}$$

The proof of Theorem 2 which shows the actual values of the entries of $\boldsymbol{H}^{(D)}$ can be easily obtained from the recursion in Equation (37) and is outlined in Appendix C. To see the value of Theorem 2 we need to specialize $\boldsymbol{A}$ to be Hermitian.

Corollary 2 *Given an Hermitian square matrix $\boldsymbol{A} \in \mathbb{C}^{N\times N}$, i.e., $\boldsymbol{A} = \boldsymbol{A}^H$, the recursion in Equation (37) 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.

Proof. Theorem 2 proposes that $\boldsymbol{H}^{(D)}$ is a Hessenberg matrix. Since $\boldsymbol{A}$ is Hermitian,

$$\boldsymbol{H}^{(D),\mathrm{H}} = \boldsymbol{T}^{(D),\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{T}^{(D)} = \boldsymbol{T}^{(D),\mathrm{H}}\boldsymbol{A}\boldsymbol{T}^{(D)} = \boldsymbol{H}^{(D)}, \tag{40}$$

therefore, $\boldsymbol{H}^{(D)}$ is a Hermitian Hessenberg matrix or Hermitian tri-diagonal.

Equation (34), Theorem 2, and Corollary 2 disclose the connection between the MSNWF approach and Krylov subspace methods to solve linear equation systems. 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 $D-1$.

We showed that the MSNWF procedure delivers filters $\boldsymbol{t}_i$ which form an orthonormal basis of the Krylov subspace. Henceforth, we can base our reasoning on the Krylov subspace. With Corollary 2 it is straightforward to understand that the covariance matrix $\boldsymbol{R}_{\boldsymbol{d}}$ of the pre-filtered observation $\boldsymbol{d}[n]$ (cf. Equation 27) is tri-diagonal, because the covariance matrix of the original observation $\boldsymbol{x}_0[n]$ is Hermitian. Moreover, the Hermitian property of $\boldsymbol{R}_{\boldsymbol{x}_0}$ can be exploited to compute the orthogonal basis $\boldsymbol{t}_i$ of the Krylov subspace $\mathcal{K}^{(D)}$ of $(\boldsymbol{R}_{\boldsymbol{x}_0}, \boldsymbol{r}_{\boldsymbol{x}_0,d_0})$. In the case of Hermitian matrices the *Lanczos algorithm* [Lan50, Lan52, Saa96] can be applied to find the orthogonal basis:

$$\begin{aligned}
\boldsymbol{t}_i &= \frac{\boldsymbol{P}_{i-1}\boldsymbol{P}_{i-2}\boldsymbol{R}_{\boldsymbol{x}_0}\boldsymbol{t}_{i-1}}{\|\boldsymbol{P}_{i-1}\boldsymbol{P}_{i-2}\boldsymbol{R}_{\boldsymbol{x}_0}\boldsymbol{t}_{i-1}\|_2} \\
&= \frac{\boldsymbol{R}_{\boldsymbol{x}_0}\boldsymbol{t}_{i-1} - \boldsymbol{t}_{i-2}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{x}_0}\boldsymbol{t}_{i-1}\boldsymbol{t}_{i-2} - \boldsymbol{t}_{i-1}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{x}_0}\boldsymbol{t}_{i-1}\boldsymbol{t}_{i-1}}{\|\boldsymbol{R}_{\boldsymbol{x}_0}\boldsymbol{t}_{i-1} - \boldsymbol{t}_{i-2}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{x}_0}\boldsymbol{t}_{i-1}\boldsymbol{t}_{i-2} - \boldsymbol{t}_{i-1}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{x}_0}\boldsymbol{t}_{i-1}\boldsymbol{t}_{i-1}\|_2}.
\end{aligned} \tag{41}$$

Thus, we reduced the complexity of the MSNWF iteration by exploiting the fact that the MSNWF is equivalent to finding the solution of the Wiener-Hopf equation in the Krylov subspace of $(\boldsymbol{R}_{\boldsymbol{x}_0}, \boldsymbol{r}_{\boldsymbol{x}_0,d_0})$. 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*.

6 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 34) or by the *Lanczos algorithm* (cf. Equation 41). The resulting rank D MSNWF is given in Equation (31) by the means of the Wiener filter $\boldsymbol{w}_{\boldsymbol{d}}^{(D)}$ which is applied to the pre-filtered observation

$$\boldsymbol{d}^{(D)}[n] = \boldsymbol{T}^{(D),\mathrm{H}}\boldsymbol{x}_0[n] \in \mathbb{C}^D, \tag{42}$$

with the tri-diagonal covariance matrix

$$\boldsymbol{R}_{\boldsymbol{d}}^{(D)} = \boldsymbol{T}^{(D),\mathrm{H}}\boldsymbol{R}_{\boldsymbol{x}_0}\boldsymbol{T}^{(D)} = \left[\begin{array}{cc|c} \multicolumn{2}{c|}{\boldsymbol{T}^{(D-1),\mathrm{H}}\boldsymbol{R}_{\boldsymbol{x}_0}\boldsymbol{T}^{(D-1)}} & \begin{matrix}\boldsymbol{0}\\ r_{D-1,D}\end{matrix} \\ \hline \boldsymbol{0}^{\mathrm{T}} & r_{D-1,D}^{*} & r_{D,D} \end{array}\right] \in \mathbb{C}^{D\times D} \tag{43}$$

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

$$\boldsymbol{r}_{\boldsymbol{d},d_0}^{(D)} = \boldsymbol{T}^{(D)}\boldsymbol{r}_{\boldsymbol{x}_0,d_0} = \begin{bmatrix} \|\boldsymbol{r}_{\boldsymbol{x}_0,d_0}\|_2 \\ \boldsymbol{0} \end{bmatrix} \in \mathbb{R}^D. \tag{44}$$

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

$$r_{D-1,D} = \boldsymbol{t}_{D-1}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{x}_0}\boldsymbol{t}_D \quad \text{and} \quad r_{D,D} = \boldsymbol{t}_{D}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{x}_0}\boldsymbol{t}_D. \tag{45}$$

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

$$\boldsymbol{w}_{\boldsymbol{d}}^{(D)} = \boldsymbol{R}_{\boldsymbol{d}}^{(D),-1}\boldsymbol{r}_{\boldsymbol{d},d_0}^{(D)} \in \mathbb{C}^D. \tag{46}$$

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

$$\boldsymbol{C}^{(D)} = \boldsymbol{R}_{\boldsymbol{d}}^{(D),-1} = [\boldsymbol{c}_1^{(D)}, \ldots, \boldsymbol{c}_D^{(D)}] \in \mathbb{C}^{D\times D} \tag{47}$$

and the inversion lemma for partitioned matrices (e. g. [Sch91, MW95]) leads to

$$\boldsymbol{C}^{(D)} = \begin{bmatrix} \boldsymbol{C}^{(D-1)} & \boldsymbol{0} \\ \boldsymbol{0}^{\mathrm{T}} & 0 \end{bmatrix} + \beta_D^{-1}\boldsymbol{b}^{(D)}\boldsymbol{b}^{(D),\mathrm{H}}, \tag{48}$$

where the additional terms read as

$$\boldsymbol{b}^{(D)} = \begin{bmatrix} -\boldsymbol{C}^{(D-1)}\begin{bmatrix}\boldsymbol{0}\\ r_{D-1,D}\end{bmatrix} \\ 1 \end{bmatrix} = \begin{bmatrix} -r_{D-1,D}\boldsymbol{c}_{D-1}^{(D-1)} \\ 1 \end{bmatrix} \in \mathbb{C}^D \tag{49}$$

and

$$\beta_D = r_{D,D} - [\boldsymbol{0}^{\mathrm{T}}, r_{D-1,D}^{*}]\boldsymbol{C}^{(D-1)}\begin{bmatrix}\boldsymbol{0}\\ r_{D-1,D}\end{bmatrix} = r_{D,D} - |r_{D-1,D}|^2 c_{D-1,D-1}^{(D-1)} \tag{50}$$

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 $\boldsymbol{c}^{(D)}_1$ can be written as

$$\boldsymbol{c}^{(D)}_1 = \begin{bmatrix} \boldsymbol{c}^{(D-1)}_1 \\ 0 \end{bmatrix} + \beta_D^{-1} c^{(D-1),*}_{1,D-1} \begin{bmatrix} |r_{D-1,D}|^2 \boldsymbol{c}^{(D-1)}_{D-1} \\ -r^*_{D-1,D} \end{bmatrix} \in \mathbb{C}^D, \tag{51}$$

where $c^{(D-1)}_{1,D-1}$ denotes the first element of $\boldsymbol{c}^{(D-1)}_{D-1}$. Obviously, the first column of $\boldsymbol{C}^{(D)}$ and, thus, the Wiener filter $\boldsymbol{w}^{(D)}_{\boldsymbol{d}}$ at step D depends upon the first column $\boldsymbol{c}^{(D-1)}_1$ 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 (48) we get

$$\boldsymbol{c}^{(D)}_D = \beta_D^{-1} \begin{bmatrix} -r_{D-1,D} \boldsymbol{c}^{(D-1)}_{D-1} \\ 1 \end{bmatrix} \tag{52}$$

which only depends on the previous last column and the new entries of $\boldsymbol{R}^{(D)}_{\boldsymbol{d}}$. So, we found an iteration that only updates two vectors $\boldsymbol{c}^{(D)}_1$ and $\boldsymbol{c}^{(D)}_D$ at each step and, moreover, the *mean squared error* at step D can be expressed with the first entry $c^{(D)}_{1,1}$ of $\boldsymbol{c}^{(D)}_1$ (cf. Equation 32):

$$MSE^{(D)} = \sigma^2_{d_0} - \|\boldsymbol{r}_{\boldsymbol{x}_0,d_0}\|_2^2 c^{(D)}_{1,1}. \tag{53}$$

The iteration in Equation (51) and (52) only operates with scalars and vectors. However, the scalars $r_{D-1,D}$ and $r_{D,D}$ (cf. Equation 45) 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 (41) 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 $\boldsymbol{c}^{(i)}_1$ and $\boldsymbol{c}^{(i)}_i$ by $\boldsymbol{c}^{(i)}_{\text{first}}$ and $\boldsymbol{c}^{(i)}_{\text{last}}$, respectively. The resulting computational complexity for a rank D MSNWF is $O(N^2 D)$, since a matrix vector multiplication with $O(N^2)$ has to be performed at each step.

Note that the algorithm in Table 1 is just a version of the *Conjugate Gradient algorithm* [HS52, Saa96]. In fact it is a direct version of the *Lanczos algorithm* [Saa96] for linear systems.

A Wiener Filter with Pre-Filtering

The new observation signal after pre-filtering with a full-rank matrix $\boldsymbol{T} \in \mathbb{C}^{N\times N}$ reads as

$$\boldsymbol{z}_1[n] = \boldsymbol{T}\boldsymbol{x}_0[n] \tag{54}$$

leading to a new *mean squared error*

$$MSE_{\boldsymbol{z}_1} = \sigma^2_{d_0} - \boldsymbol{w}^{\mathrm{H}}_{\boldsymbol{z}} \boldsymbol{r}_{\boldsymbol{z}_1,d_0} - \boldsymbol{r}^{\mathrm{H}}_{\boldsymbol{z}_1,d_0} \boldsymbol{w}_{\boldsymbol{z}} + \boldsymbol{w}^{\mathrm{H}}_{\boldsymbol{z}} \boldsymbol{R}_{\boldsymbol{z}_1} \boldsymbol{w}_{\boldsymbol{z}}, \tag{55}$$

choose desired MSE: σ_ε^2
choose maximum dimension: D
$\boldsymbol{t}_0 = \boldsymbol{0}, \quad \boldsymbol{t}_1 = \boldsymbol{r}_{\boldsymbol{x}_0,d_0}/\lVert \boldsymbol{r}_{\boldsymbol{x}_0,d_0} \rVert_2$
$\boldsymbol{u} = \boldsymbol{R}_{\boldsymbol{x}_0}\boldsymbol{t}_1$
$r_{0,1} = 0, \quad r_{1,1} = \boldsymbol{t}_1^{\mathrm{H}}\boldsymbol{u}$
$c_{\mathrm{first}}^{(1)} = r_{1,1}^{-1}, \quad c_{\mathrm{last}}^{(1)} = r_{1,1}^{-1}$
$MSE^{(1)} = \sigma_{d_0}^2 - \lVert \boldsymbol{r}_{\boldsymbol{x}_0,d_0} \rVert_2^2 c_{\mathrm{first}}^{(1)}$
$\Delta = 1$
for $i = 2, \ldots, D$
if $MSE^{(i)} < \sigma_\varepsilon^2$ then $\Delta = i-1$; break
$\boldsymbol{v} = \boldsymbol{u} - r_{i-1,i-1}\boldsymbol{t}_{i-1} - r_{i-2,i-1}\boldsymbol{t}_{i-2}$
$r_{i-1,i} = \lVert \boldsymbol{v} \rVert_2$
if $r_{i-1,i} = 0$ then $\Delta = i-1$; break
$\boldsymbol{t}_i = \boldsymbol{v}/r_{i-1,i}$
$\boldsymbol{u} = \boldsymbol{R}_{\boldsymbol{x}_0}\boldsymbol{t}_i$
$r_{i,i} = \boldsymbol{t}_i^{\mathrm{H}}\boldsymbol{u}$
$\beta_i = r_{i,i} - \lvert r_{i-1,i}\rvert^2 c_{\mathrm{last},i-1}^{(i-1)}$
$\boldsymbol{c}_{\mathrm{first}}^{(i)} = \begin{bmatrix} \boldsymbol{c}_{\mathrm{first}}^{(i-1)} \\ 0 \end{bmatrix} + \beta_i^{-1} c_{\mathrm{last},1}^{(i-1),*} \begin{bmatrix} \lvert r_{i-1,i}\rvert^2 \boldsymbol{c}_{\mathrm{last}}^{(i-1)} \\ -r_{i-1,i}^* \end{bmatrix}$
$\boldsymbol{c}_{\mathrm{last}}^{(i)} = \beta_i^{-1} \begin{bmatrix} -r_{i-1,i}\boldsymbol{c}_{\mathrm{last}}^{(i-1)} \\ 1 \end{bmatrix}$
$MSE^{(i)} = \sigma_{d_0}^2 - \lVert \boldsymbol{r}_{\boldsymbol{x}_0,d_0} \rVert_2^2 c_{\mathrm{first},1}^{(i)}$
$\boldsymbol{T}^{(D)} = [\boldsymbol{t}_1, \ldots, \boldsymbol{t}_\Delta]$
$\boldsymbol{w}_0^{(D)} = \lVert \boldsymbol{r}_{\boldsymbol{x}_0,d_0} \rVert_2 \boldsymbol{T}^{(D)} \boldsymbol{c}_{\mathrm{first}}^{(\Delta)}$

Table 1: Lanczos MSNWF

where the covariance matrix of the new observation signal $\boldsymbol{z}_1[n]$ can be expressed as

$$\boldsymbol{R}_{\boldsymbol{z}_1} = \mathrm{E}\{\boldsymbol{z}_1[n]\boldsymbol{z}_1^{\mathrm{H}}[n]\} = \boldsymbol{T}\boldsymbol{R}_{\boldsymbol{x}_0}\boldsymbol{T}^{\mathrm{H}} \tag{56}$$

and the new crosscorrelation vector is

$$\boldsymbol{r}_{\boldsymbol{z}_1,d_0} = \mathrm{E}\{\boldsymbol{z}_1[n]d_0^*[n]\} = \boldsymbol{T}\boldsymbol{r}_{\boldsymbol{x}_0,d_0}. \tag{57}$$

If we apply the resulting Wiener filter (cf. Equation 8)

$$\boldsymbol{w}_{\boldsymbol{z}_1} = \boldsymbol{R}_{\boldsymbol{z}_1}^{-1}\boldsymbol{r}_{\boldsymbol{z}_1,d_0} = \boldsymbol{T}^{-1,\mathrm{H}}\boldsymbol{R}_{\boldsymbol{x}_0}^{-1}\boldsymbol{T}^{-1}\boldsymbol{T}\boldsymbol{r}_{\boldsymbol{x}_0,d_0} \tag{58}$$

to the new observation $\boldsymbol{z}_1[n]$ to get the new estimate

$$\hat{d}_{0,\boldsymbol{z}_1}[n] = \boldsymbol{w}_{\boldsymbol{z}_1}^{\mathrm{H}}\boldsymbol{z}_1[n] = \boldsymbol{r}_{\boldsymbol{x}_0,d_0}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{x}_0}^{-1}\boldsymbol{T}^{-1}\boldsymbol{T}\boldsymbol{x}_0[n] = \boldsymbol{w}_0^{\mathrm{H}}\boldsymbol{x}_0[n], \tag{59}$$

we observe that $\hat{d}_{0,\boldsymbol{z}_1}[n] = \hat{d}_0[n]$ and, thus, the estimate is unchanged. Consequently, the new *minimum mean squared error*

$$MMSE_{\boldsymbol{z}_1} = \sigma_{d_0}^2 - \boldsymbol{w}_{\boldsymbol{z}_1}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{z}_1}^{-1}\boldsymbol{w}_{\boldsymbol{z}_1} = \sigma_{d_0}^2 - \boldsymbol{r}_{\boldsymbol{x}_0,d_0}^{\mathrm{H}}\boldsymbol{T}^{\mathrm{H}}\boldsymbol{T}^{-1,\mathrm{H}}\boldsymbol{R}_{\boldsymbol{x}_0}\boldsymbol{T}^{-1}\boldsymbol{T}\boldsymbol{r}_{\boldsymbol{x}_0,d_0} \tag{60}$$

is the same as before (cf. Equation 9) which completes the proof of Theorem 1.

B Basis of MSNWF is Krylov Subspace

The Langrange function for the optimization in Equation (33) reads as

$$L(\boldsymbol{t}, \lambda_1, \ldots, \lambda_i) = \frac{1}{2}(\boldsymbol{t}^{\mathrm{H}} \boldsymbol{R}_{\boldsymbol{x}_0} \boldsymbol{t}_{i-1} + \boldsymbol{t}_{i-1}^{\mathrm{H}} \boldsymbol{R}_{\boldsymbol{x}_0} \boldsymbol{t}) - \sum_{k=1}^{i-1} \lambda_k \boldsymbol{t}^{\mathrm{H}} \boldsymbol{t}_k - \lambda_i (\boldsymbol{t}^{\mathrm{H}} \boldsymbol{t} - 1). \tag{61}$$

The derivation with respect to the complex conjugate of $\boldsymbol{t}$ must be zero, thus,

$$\frac{\partial L(\boldsymbol{t}, \lambda_1, \ldots, \lambda_i)}{\partial \boldsymbol{t}^*} = \frac{1}{2} \boldsymbol{R}_{\boldsymbol{x}_0} \boldsymbol{t}_{i-1} - \sum_{k=1}^{i-1} \lambda_k \boldsymbol{t}_k - \lambda_i \boldsymbol{t} = \boldsymbol{0}. \tag{62}$$

Because $\boldsymbol{t}$ is orthogonal to $\boldsymbol{t}_k, k = 1, \ldots, i-1$ and the filters $\boldsymbol{t}_k$ are orthonormal, i.e., $\boldsymbol{t}_k^{\mathrm{H}} \boldsymbol{t}_l = \delta_{k,l}$, the Langrange multipliers can be expressed as

$$\lambda_k = \frac{1}{2} \boldsymbol{t}_k^{\mathrm{H}} \boldsymbol{R}_{\boldsymbol{x}_0} \boldsymbol{t}_k, \quad k = 1, \ldots, i-1, \tag{63}$$

and

$$\lambda_i \boldsymbol{t} = \frac{1}{2} \boldsymbol{R}_{\boldsymbol{x}_0} \boldsymbol{t}_{i-1} - \frac{1}{2} \sum_{k=1}^{i-1} \boldsymbol{t}_k^{\mathrm{H}} \boldsymbol{R}_{\boldsymbol{x}_0} \boldsymbol{t}_{i-1} \boldsymbol{t}_k = \frac{1}{2}(\boldsymbol{1}_N - \sum_{k=1}^{i-1} \boldsymbol{t}_k \boldsymbol{t}_k^{\mathrm{H}}) \boldsymbol{R}_{\boldsymbol{x}_0} \boldsymbol{t}_{i-1}. \tag{64}$$

If we rewrite this expression by substituting the sum with a product of projection matrices (cf. Equation 35), we end up with

$$\boldsymbol{t} = \frac{1}{2\lambda_i} \left(\prod_{k=i-1}^{1} \boldsymbol{P}_k \right) \boldsymbol{R}_{\boldsymbol{x}_0} \boldsymbol{t}_{i-1} \tag{65}$$

and since λ_i has to be chosen to give a unit norm vector, we proved the result in Equation (34).

C Tri-Diagonalization with Arnoldi Algorithm

The recursion formula of Theorem 2 in Equation 37 can be rewritten to get

$$h_{i+1,i} \boldsymbol{t}_{i+1} = \left(\prod_{k=i}^{1} \boldsymbol{P}_k \right) \boldsymbol{A} \boldsymbol{t}_i, \tag{66}$$

where we introduced the abbreviation

$$h_{i+1,i} = \| \left(\prod_{k=i}^{1} \boldsymbol{P}_k \right) \boldsymbol{A} \boldsymbol{t}_i \|_2. \tag{67}$$

If the product of the projectors $\boldsymbol{P}_k$ is converted to a sum and the orthogonality of the $\boldsymbol{t}_i$ is considered, Equation (66) reads as

$$h_{i+1,i}\boldsymbol{t}_{i+1} = \boldsymbol{A}\boldsymbol{t}_i - \sum_{k=1}^{i} \boldsymbol{t}_k^{\mathrm{H}} \boldsymbol{A}\boldsymbol{t}_i \boldsymbol{t}_k \tag{68}$$

and by substituting

$$h_{k,i} = \boldsymbol{t}_k^{\mathrm{H}} \boldsymbol{A}\boldsymbol{t}_i \tag{69}$$

we end up with

$$\boldsymbol{A}\boldsymbol{t}_i = h_{i+1,i}\boldsymbol{t}_{i+1} + \sum_{k=1}^{i} h_{k,i}\boldsymbol{t}_k \tag{70}$$

which is the i-th column of the matrix equality in Equation (38).

References

[AC76] S. P. Applebaum and D. J. Chapman. Adaptive Arrays with Main Beam Constraints. *IEEE Transactions on Antennas and Propagation*, 24(5), September 1976.

[Arn51] W. E. Arnoldi. The Principle of Minimized Iterations in the Solution of the Matrix Eigenvalue Problem. *Quarterly of Applied Mathematics*, 9(1):17–29, January 1951.

[BR89] K. A. Byerly and R. A. Roberts. Output Power Based Partial Adaptive Array Design. In *Proc. 23rd Asilomar Conference on Signals, Systems & Computers*, volume 2, pages 576–580, October 1989.

[EY36] C. Eckart and G. Young. The Approximation of One Matrix by Another of Lower Rank. *Psychometrika*, 1(3):211–218, September 1936.

[GGDR98] J. S. Goldstein, J. R. Guerci, D. E. Dudgeon, and I. S. Reed. Theory of Signal Representation. submitted to *IEEE Transactions on Signal Processing*, 1998.

[GGR99] J. S. Goldstein, J. R. Guerci, and I. S. Reed. An Optimal Generalized Theory of Signal Representation. In *Proc. ICASSP'99*, volume 3, pages 1357–1360, March 1999.

[GJ82] L. J. Griffiths and C. W. Jim. An Alternative Approach to Linearly Constrained Adaptive Beamforming. *IEEE Transactions on Antennas and Propagation*, 30(1), January 1982.

[GR97a] J. S. Goldstein and I. S. Reed. A New Method of Wiener Filtering and its Application to Interference Mitigation for Communications. In *Proc. MILCOM 1997*, volume 3, pages 1087–1091, November 1997.

[GR97b] J. S. Goldstein and I. S. Reed. Subspace Selection for Partially Adaptive Sensor Array Processing. *IEEE Transactions on Aerospace and Electronic Systems*, 33(2):539–543, April 1997.

[GRS98] 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*, 44(7):2943–2959, November 1998.

[GRZ99] J. S. Goldstein, I. S. Reed, and P. A. Zulch. Multistage Partially Adaptive STAP CFAR Detection Algorithm. *IEEE Transactions on Aerospace and Electronic Systems*, 35(2):645–661, April 1999.

[HG00] 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.

[Hot33] H. Hotelling. Analysis of a Complex of Statistical Variables into Principal Components. *Journal of Educational Psychology*, 24(6/7):417–441, 498–520, September/October 1933.

[HS52] M. R. Hestenes and E. Stiefel. Methods of Conjugate Gradients for Solving Linear Systems. *Journal of Research of the National Bureau of Standards*, 49(6):409–436, December 1952.

[HX99] 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.

[KBP98] A. Kansal, S. N. Batlama, and D. A. Pados. Adaptive Maximum SINR RAKE Filtering for DS-CDMA Multipath Fading Channels. *IEEE Journal on Selected Areas in Communications*, 16(9), December 1998.

[Lan50] C. Lanczos. An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and Integral Operators. *Journal of Research of the National Bureau of Standards*, 45(4):255–282, October 1950.

[Lan52] C. Lanczos. Solution of Systems of Linear Equations by Minimized Iterations. *Journal of Research of the National Bureau of Standards*, 49(1):33–53, July 1952.

[LUN99] P. Li, W. Utschick, and J. A. Nossek. A New DS-CDMA Interference Cancellation Scheme for Space-Time rake Processing. In *Proc. European Wireless 99 & ITG Mobile Communications*, pages 385–389, October 1999.

[MW95] R. N. McDonough and A. D. Whalen. *Detection of Signals in Noise*. Academic Press, 1995.

[PB99] D. A. Pados and S. N. Batlama. Joint Space-Time Auxiliary-Vector Filtering for DS/CDMA Systems with Antenna Arrays. *IEEE Transactions on Communications*, 47(9), September 1999.

[PLB99] D. A. Pados, F. J. Lombardo, and S. N. Batalama. Auxiliary-Vector Filters and Adaptive Steering for DS/CDMA Single-User Detection. *IEEE Transactions on Vehicular Technology*, 48(6), November 1999.

[Saa96] Y. Saad. *Iterative Methods for Sparse Linear Systems*. PWS – *out of print*, 1996. http://www-users.cs.umn.edu/ saad/books.html.

[Sch91] L. L. Scharf. *Statistical Signal Processing*. Addison-Wesley, 1991.

[SW99] M. K. Schneider and A. S. Willsky. Krylov Subspace Estimation. submitted to *SIAM Journal on Scientific Computing*, June 1999.

[vdV00] H. A. van der Vorst. Krylov Subspace Iteration. *Computing in Science & Engineering*, 2(1):32–37, January/February 2000.