in Proc. IEEE ICASSP-2001,
Salt Lake City (UT), May 2001.

SIGNAL-ADAPTIVE ROBUST TIME-VARYING WIENER FILTERS: BEST SUBSPACE SELECTION AND STATISTICAL ANALYSIS

Gerald Matz, Franz Hlawatsch, and Alfred Raidl

Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology
Gusshausstrasse 25/389, A-1040 Wien, Austria
phone: +43 1 58801 38916, fax: +43 1 58801 38999, email: g.matz@ieee.org
web: http://www.nt.tuwien.ac.at/dspgroup/time.html

ABSTRACT

We propose a signal-adaptive robust time-varying Wiener filter for nonstationary signal estimation/enhancement. This filter uses projections onto local cosine subspaces and a novel "best subspace" algorithm. It allows efficient on-line operation including stable on-line estimation of design parameters. A statistical analysis is provided, and a speech enhancement example is considered.

1. INTRODUCTION

We consider estimation of a signal $s(t)$ from a noisy observation $r(t) = s(t) + n(t)$. Signal $s(t)$ and noise $n(t)$ are uncorrelated, real-valued, *nonstationary* random processes with respective correlation operators[1] $\mathbf{R}_s$ and $\mathbf{R}_n$. The signal estimate is $\hat{s}(t) = (\mathbf{H}r)(t)$ with $\mathbf{H}$ a *time-varying* system.

The linear system $\mathbf{H}$ minimizing the mean-square error (the *time-varying Wiener filter*) [1, 2] is very sensitive to errors in modeling and/or estimating $\mathbf{R}_s$ and $\mathbf{R}_n$ [3, 4], and for long signals its design and implementation are computationally intensive. Therefore, in this paper we propose a *signal-adaptive robust time-varying Wiener filter* with efficient on-line operation and stable on-line estimation of design parameters. The filter uses orthogonal projections onto local cosine subspaces and a novel "best subspace" algorithm that extends the classical best basis algorithm [5, 6].

The paper is organized as follows. Section 2 reviews signal-adaptive minimax robust time-varying Wiener filters [3]. Section 3 proposes an algorithm for signal-adaptive subspace optimization. Section 4 provides a statistical analysis of estimated filter and error parameters. Finally, Section 5 considers the application of the novel filter to speech enhancement.

2. SIGNAL-ADAPTIVE ROBUST WIENER FILTERS

We first review the signal-adaptive version of the *minimax robust time-varying Wiener filter* introduced in [3]. This filter provides the basis for our subsequent development.

Local Cosine Subspaces. An efficient on-line version of the robust Wiener filter is based on a partition $\{\mathcal{X}_{k,l}\}_{k\in\mathbb{Z},l\in\mathbb{N}_0}$ of the real signal space $L_2(\mathbb{R})$ into orthogonal *local cosine subspaces (LCSs)*

$$\mathcal{X}_{k,l} \triangleq \text{span}\big\{u_{k,l}^{(m)}(t)\big\}_{m=1,\ldots,M}\,, \qquad k\in\mathbb{Z},\, l\in\mathbb{N}_0 \qquad (1)$$

that have dimension M and are spanned by the orthonormal *local cosine basis (LCB) functions*[2] [6, 7]

$$u_{k,l}^{(m)}(t) \triangleq w_k(t)\sqrt{\frac{2}{T_k}}\,\cos\!\left(\frac{2(lM+m)-1}{2T_k}\pi(t-t_k)\right), \quad (2)$$

with $m = 1, 2, \ldots, M$. Here, the t_k ($k \in \mathbb{Z}$) define a partition of the time axis into disjoint intervals $[t_k, t_{k+1}]$ of duration $T_k = t_{k+1} - t_k$ and $w_k(t)$ is a window associated to the kth interval $[t_k, t_{k+1}]$. For background and details on the construction of $w_k(t)$ see [6, 7].

To any interval partition $\{[t_k, t_{k+1}]\}_{k\in\mathbb{Z}}$ of the time axis, (1) and (2) associate an orthogonal subspace partition $\{\mathcal{X}_{k,l}\}_{k\in\mathbb{Z},l\in\mathbb{N}_0}$. Since the LCB function $u_{k,l}^{(m)}(t)$ is effectively supported in the time interval $[t_k, t_{k+1}] = [t_k, t_k + T_k]$ and the frequency band $[(lM + m - 1)F_k, (lM + m)F_k]$ with $F_k = 1/(2T_k)$, the LCS $\mathcal{X}_{k,l}$ is effectively supported in the time-frequency region $[t_k, t_k + T_k] \times [lMF_k, (l + 1)MF_k]$ of area M. Hence, the LCS partition $\{\mathcal{X}_{k,l}\}_{k\in\mathbb{Z},l\in\mathbb{N}_0}$ corresponds to a rectangular tiling of the time-frequency plane.

Robust Wiener Filter. Let us consider a specific LCS partition $\{\mathcal{X}_{k,l}\}_{k\in\mathbb{Z},l\in\mathbb{N}_0}$. The kernels of the associated orthogonal projection operators [8] $\mathbf{P}_{k,l}$ are given by $p_{k,l}(t,t') = \sum_{m=1}^{M} u_{k,l}^{(m)}(t)\, u_{k,l}^{(m)}(t')$. Whereas the correlation operators $\mathbf{R}_s$, $\mathbf{R}_n$ are unknown, we initially assume that the expected energies of $s(t)$ and $n(t)$ within the subspaces $\mathcal{X}_{k,l}$ are known to equal $s_{k,l} \geq 0$ and $n_{k,l} \geq 0$, respectively, i.e., $\mathrm{E}\{\|\mathbf{P}_{k,l}s\|^2\} = \mathrm{tr}\{\mathbf{P}_{k,l}\mathbf{R}_s\} = s_{k,l}$ and $\mathrm{E}\{\|\mathbf{P}_{k,l}n\|^2\} = \mathrm{tr}\{\mathbf{P}_{k,l}\mathbf{R}_n\} = n_{k,l}$. The *uncertainty classes* $\mathcal{S}$ and $\mathcal{N}$ are defined as the sets of all $\mathbf{R}_s$ and $\mathbf{R}_n$ satisfying the above properties. By definition, the *minimax robust time-varying Wiener filter* $\mathbf{H}_\mathrm{R}$ [3, 4] optimizes the worst-case performance within these uncertainty classes:

$$\mathbf{H}_\mathrm{R} \triangleq \arg\min_{\mathbf{H}} \max_{\substack{\mathbf{R}_s\in\mathcal{S}\\ \mathbf{R}_n\in\mathcal{N}}} e(\mathbf{H};\mathbf{R}_s,\mathbf{R}_n)\,,$$

with the mean-square error (MSE) $e(\mathbf{H};\mathbf{R}_s,\mathbf{R}_n) \triangleq \mathrm{E}\{\|\mathbf{H}r - s\|^2\}$. It is shown in [3,4] that the signal estimate $\hat{s}(t) = (\mathbf{H}_\mathrm{R}\, r)(t)$ equals

$$\hat{s}(t) = \sum_{k=-\infty}^{\infty}\sum_{l=0}^{\infty} h_{k,l}\,(\mathbf{P}_{k,l}\, r)(t) \quad \text{with} \quad h_{k,l} = \frac{s_{k,l}}{s_{k,l} + n_{k,l}}\,, \qquad (3)$$

where

$$(\mathbf{P}_{k,l}\, r)(t) = \sum_{m=1}^{M} \big\langle r, u_{k,l}^{(m)}\big\rangle\, u_{k,l}^{(m)}(t)\,.$$

Based on these expressions, efficient on-line calculation of $\hat{s}(t)$ is possible using fast LCB analysis and synthesis algorithms [6]. Furthermore, for *any* $\mathbf{R}_s \in \mathcal{S}, \mathbf{R}_n \in \mathcal{N}$ the resulting MSE is

$$e(\mathbf{H}_\mathrm{R};\mathbf{R}_s,\mathbf{R}_n) = \sum_{k=-\infty}^{\infty}\sum_{l=0}^{\infty} e_{k,l} \quad \text{with} \quad e_{k,l} = \frac{s_{k,l}\, n_{k,l}}{s_{k,l} + n_{k,l}}\,. \qquad (4)$$

We note that previously proposed subband/subspace-based signal estimation and speech enhancement schemes (e.g. [9, 10]) can be cast in a similar framework with $\dim\{\mathcal{X}_{k,l}\} = 1$. Choosing

Funding by FWF grant P11904-TEC.

[1] The correlation operator $\mathbf{R}_x$ of a (generally nonstationary) random process $x(t)$ is the positive (semi-)definite linear operator whose kernel equals $r_x(t,t') = \mathrm{E}\{x(t)\, x(t')\}$.

[2] An alternative to LCBs is provided by wavelet packet bases [6, 7].

$\dim\{\mathcal{X}_{k,l}\} > 1$ entails a resolution loss (since all components of $r(t)$ lying in a given subspace $\mathcal{X}_{k,l}$ are treated alike) but will turn out to be advantageous in a signal-adaptive implementation.

Signal-Adaptive Implementation. With certain assumptions, the subspace projections $(\mathbf{P}_{k,l}\, r)(t)$ calculated during the analysis stage can be used for signal-adaptive[3] estimation of the filter weights $h_{k,l} = s_{k,l}/(s_{k,l} + n_{k,l})$ in (3). We first note that $h_{k,l} = 1 - n_{k,l}/r_{k,l}$ with

$$r_{k,l} \triangleq s_{k,l} + n_{k,l} = \mathrm{E}\{\|\mathbf{P}_{k,l}\, r\|^2\}.$$

Thus, a (nonnegativity-enforced) estimate of $h_{k,l}$ is given by

$$\hat{h}_{k,l} = \max\left\{0, 1 - \frac{\hat{n}_{k,l}}{\hat{r}_{k,l}}\right\}, \qquad (5)$$

where $\hat{n}_{k,l}$ and $\hat{r}_{k,l}$ are suitable estimates of $n_{k,l}$ and $r_{k,l}$, respectively. An unbiased estimate of $r_{k,l}$ is provided by

$$\hat{r}_{k,l} \triangleq \|\mathbf{P}_{k,l}\, r\|^2.$$

The statistical properties of the estimates $\hat{r}_{k,l}$ and $\hat{h}_{k,l}$ will be studied in Section 4.

An unbiased estimate of the noise energies $n_{k,l}$ can be obtained if the set of all index pairs (k,l) can be partitioned into disjoint subsets $\mathbb{I}_i$ such that: (i) the noise energies $n_{k,l}$ for all $(k,l) \in \mathbb{I}_i$ are equal and (ii) there exists at least one "noise only" index pair $(k_0, l_0) \in \mathbb{I}_i$ (equivalently, one "noise only" subspace $\mathcal{X}_{k_0,l_0}$) for which $s_{k_0,l_0} = 0$ or equivalently $r_{k_0,l_0} = n_{k_0,l_0}$. Then, $\hat{r}_{k_0,l_0}$ provides an unbiased estimate of $n_{k,l}$ within $\mathbb{I}_i$, i.e.,

$$\hat{n}_{k,l} = \hat{r}_{k_0,l_0} \qquad \text{for all } (k,l) \in \mathbb{I}_i.$$

Of practical relevance is the special case where each subset $\mathbb{I}_i$ corresponds to some frequency index l and some time interval $k_j \le k < k_{j+1}$, i.e., $\mathbb{I}_i = \{(k_j, l), (k_j+1, l), \ldots, (k_{j+1}-1, l)\}$. That is, for each frequency index l, $n_{k,l}$ is constant on the time interval $[k_j, k_{j+1}-1]$ (which may depend on l) and at least for one $(k_0, l) \in \mathbb{I}_i$ there is $s_{k_0,l} = 0$. Note that the first property corresponds to a "generalized stationarity" of $n(t)$. A dual situation is that $\mathbb{I}_i$ corresponds to some time index k and some frequency interval $l_j \le l < l_{j+1}$, i.e., $\mathbb{I}_i = \{(k, l_j), (k, l_j+1), \ldots, (k, l_{j+1}-1)\}$. Here, $n_{k,l}$ is locally constant with respect to the frequency index l, corresponding to a "generalized whiteness" of $n(t)$.

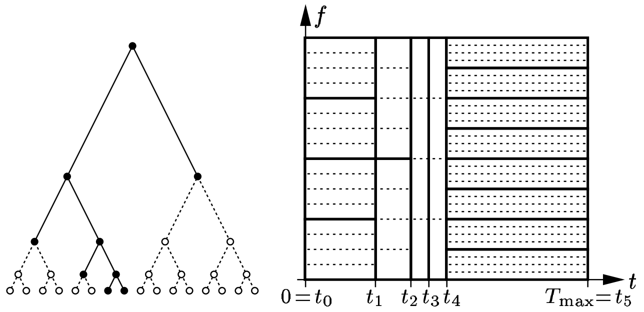
3. BEST SUBSPACE SELECTION

We now propose a "best subspace" algorithm that allows to adapt the LCS partition $\{\mathcal{X}_{k,l}\}_{k\in\mathbb{Z}, l\in\mathbb{N}_0}$ (equivalently, the associated time-frequency tiling) to the observed signal $r(t)$. This method is inspired by the well-known best basis algorithm [5, 6]. For algorithmic simplicity, we have to restrict the LCS partitions to correspond to dyadic trees.

Dyadic LCS Trees. We first split the time axis into disjoint intervals of duration $T_{\max}$. ($T_{\max}$ determines the processing delay and $1/(2T_{\max})$ is the finest frequency resolution of the filter.) For each such interval—e.g., the interval $[0, T_{\max}]$—we construct a dyadic LCS tree (see Fig. 1) by recursively splitting $[0, T_{\max}]$ into disjoint subintervals of length $T_{\max}/2^d$ with $d = 0, 1, \ldots, D-1$. (D is the depth of the tree; note that the lowest tree level $d = D-1$ corresponds to the minimal subinterval length $T_{\min} = T_{\max}/2^{D-1}$.) To the pth ($p = 0, \ldots, 2^d - 1$) node at the dth level of the tree, we associate the subinterval $\mathcal{I}_d^p \triangleq [pT_{\max}/2^d, (p+1)T_{\max}/2^d]$. A tree is admissible if each node has either no or two children.

By this construction, the ith admissible tree corresponds to an interval partition $\big\{[t_k^{(i)}, t_{k+1}^{(i)}]\big\}_{k=0,\ldots,K_i-1}$ of $[0, T_{\max}]$. Here, K_i denotes the number of subintervals $[t_k^{(i)}, t_{k+1}^{(i)}]$ into which $[0, T_{\max}]$ is split. The disjoint subintervals $[t_k^{(i)}, t_{k+1}^{(i)}] \subseteq [0, T_{\max}]$ are of the form $[pT_{\max}/2^d, (p+1)T_{\max}/2^d]$ with variable lengths $T_{\max}/2^d$. Via (1) and (2), the interval partition $\big\{[t_k^{(i)}, t_{k+1}^{(i)}]\big\}_{k=0,\ldots,K_i-1}$ in turn corresponds to an LCS partition $\mathcal{B}_i \triangleq \{\mathcal{X}_{k,l}^{(i)}\}_{k=0,\ldots,K_i-1, l\in\mathbb{N}_0}$ of the subspace corresponding to $[0, T_{\max}]$ into M-dimensional LCSs $\mathcal{X}_{k,l}^{(i)}$. Note that the kth subinterval $[t_k^{(i)}, t_{k+1}^{(i)}] \subseteq [0, T_{\max}]$ corresponds to the subspace $\bigoplus_{l=0}^{\infty} \mathcal{X}_{k,l}^{(i)}$. Also note that higher levels d correspond to longer subintervals and, hence, to subspaces $\mathcal{X}_{k,l}^{(i)}$ with longer duration and smaller bandwidth (resulting in poorer time resolution and better frequency resolution).

[3] A related discussion of SVD-based data-adaptive minimum MSE estimation is given in [11].



(a) (b)

Figure 1: *Example of an LCS partition. (a) Admissible LCS tree (solid) based on a dyadic tree with $D = 5$ levels (dotted). (b) Time-frequency tiling corresponding to the LCS tree in (a) (LCS dimension $M = 4$; the dotted lines indicate the time-frequency support of the individual LCB functions spanning the respective LCS).*

Cost Function. We next assign a cost to each LCS partition $\mathcal{B}_i$. Since our goal is signal estimation with small MSE, we will use an estimate of the MSE as our cost function (a related approach for denoising is described in [6, 10]). An estimate of the MSE $e_{k,l}$ in (4) contributed by the subspace $\mathcal{X}_{k,l}^{(i)}$ is obtained by replacing $r_{k,l} = s_{k,l} + n_{k,l}$, $n_{k,l}$, and $s_{k,l}$ with estimates $\hat{r}_{k,l}$, $\hat{n}_{k,l}$ (cf. Section 2), and $\hat{s}_{k,l}$, respectively:

$$\hat{e}_{k,l}^{(i)} \triangleq \frac{\hat{s}_{k,l}\hat{n}_{k,l}}{\hat{r}_{k,l}} \qquad \text{with} \quad \hat{s}_{k,l} \triangleq \max\{0, \hat{r}_{k,l} - \hat{n}_{k,l}\}. \qquad (6)$$

(Note that the estimates $\hat{r}_{k,l}$, $\hat{n}_{k,l}$, $\hat{s}_{k,l}$ depend on i since they are based on $\mathcal{X}_{k,l}^{(i)}$.) In view of (4), an estimate of the total MSE corresponding to the ith partition $\mathcal{B}_i = \{\mathcal{X}_{k,l}^{(i)}\}_{k=0,\ldots,K_i-1, l\in\mathbb{N}_0}$ is then obtained by adding all subspace MSE estimates $\hat{e}_{k,l}^{(i)}$,

$$\hat{e}(\mathcal{B}_i) \triangleq \sum_{k=0}^{K_i-1} \sum_{l=0}^{\infty} \hat{e}_{k,l}^{(i)} = \sum_{k=0}^{K_i-1} \sum_{l=0}^{\infty} \frac{\hat{s}_{k,l}\hat{n}_{k,l}}{\hat{r}_{k,l}}. \qquad (7)$$

A statistical analysis of the subspace MSE estimates $\hat{e}_{k,l}^{(i)}$ will be provided in Section 4.

Best Subspace Algorithm. For a given observed signal $r(t)$, the best subspace partition $\mathcal{B}_{\text{opt}}$ associated to the interval $[0, T_{\max}]$ is defined as the $\mathcal{B}_i$ minimizing the estimated MSE,

$$\mathcal{B}_{\text{opt}} \triangleq \arg\min_i \hat{e}(\mathcal{B}_i). \qquad (8)$$

An efficient minimization method is provided by the following *best subspace algorithm* (a variation of the best basis algorithm [5, 6]) that exploits the additivity of the cost function $\hat{e}(\mathcal{B}_i)$ in (7) to solve (8) by means of a recursive bottom-up strategy. Hereafter, let $\{\mathcal{X}\}_d^p$ denote the set of LCSs associated to the subinterval

$\mathcal{I}_d^p = [pT_{\max}/2^d, (p+1)T_{\max}/2^d]$ via (1) and (2), and let $\hat{e}(\{\mathcal{X}\}_d^p)$ denote the sum of the estimated subspace MSEs for all LCSs in $\{\mathcal{X}\}_d^p$.

Start at the lowest level $d = D-1$ of the LCS tree (corresponding to the finest interval partition, i.e., minimal subinterval length $T_{\min} = T_{\max}/2^{D-1}$) and compute the cost $\hat{e}(\{\mathcal{X}\}_{D-1}^p)$ for all $p = 0, \ldots, 2^{D-1}-1$. At this level $d = D-1$, the *best partial subspace partition* associated to the subinterval $\mathcal{I}_{D-1}^p$ is simply defined as $\mathcal{P}_{D-1}^p = \{\mathcal{X}\}_{D-1}^p$.

Next, consider the second lowest level $d = D-2$ and compute the cost $\hat{e}(\{\mathcal{X}\}_{D-2}^p)$ for all $p = 0, \ldots, 2^{D-2}-1$. A given node $(D-2, p)$ at level $d = D-2$ corresponds to the subinterval $\mathcal{I}_{D-2}^p$; it has two children $(D-1, 2p)$ and $(D-1, 2p+1)$ at level $d = D-1$ that correspond to the subintervals $\mathcal{I}_{D-1}^{2p}$ and $\mathcal{I}_{D-1}^{2p+1}$, respectively (whose union is $\mathcal{I}_{D-2}^p$). The best partial subspace partition associated to the subinterval $\mathcal{I}_{D-2}^p$ is now defined as $\mathcal{P}_{D-2}^p = \{\mathcal{X}\}_{D-2}^p$ (corresponding to the "long" subinterval $\mathcal{I}_{D-2}^p$) if $\hat{e}(\{\mathcal{X}\}_{D-2}^p) < \hat{e}(\{\mathcal{X}\}_{D-1}^{2p}) + \hat{e}(\{\mathcal{X}\}_{D-1}^{2p+1})$ or as $\mathcal{P}_{D-2}^p = \{\mathcal{X}\}_{D-1}^{2p} \cup \{\mathcal{X}\}_{D-1}^{2p+1}$ (corresponding to the two "short" subintervals $\mathcal{I}_{D-1}^{2p}$ and $\mathcal{I}_{D-1}^{2p+1}$) if $\hat{e}(\{\mathcal{X}\}_{D-2}^p) \geq \hat{e}(\{\mathcal{X}\}_{D-1}^{2p}) + \hat{e}(\{\mathcal{X}\}_{D-1}^{2p+1})$.

This process of comparing the cost of the partial subspace partition associated to a parent node with the total cost of the best partial subspace partitions associated to its two children nodes and adopting the partial subspace partition with the smaller cost is continued for $d = D-3, D-4, \ldots, 0$. For the pth node ($p = 0, 1, \ldots, 2^d-1$) at a given level $d \in \{D-2, \ldots, 0\}$, the best partial subspace partition $\mathcal{P}_d^p$ is thus determined as

$$\mathcal{P}_d^p = \begin{cases} \{\mathcal{X}\}_d^p & \text{if } \hat{e}(\{\mathcal{X}\}_d^p) \leq \hat{e}(\mathcal{P}_{d+1}^{2p}) + \hat{e}(\mathcal{P}_{d+1}^{2p+1}), \\ \mathcal{P}_{d+1}^{2p} \cup \mathcal{P}_{d+1}^{2p+1} & \text{if } \hat{e}(\{\mathcal{X}\}_d^p) > \hat{e}(\mathcal{P}_{d+1}^{2p}) + \hat{e}(\mathcal{P}_{d+1}^{2p+1}). \end{cases}$$

Finally, the best total subspace partition defined in (8) is given by $\mathcal{B}_{\text{opt}} = \mathcal{P}_0^0$; it can be determined with complexity $\mathcal{O}(2^{D-1})$. However, it also requires computation of the MSE estimates $\hat{e}(\{\mathcal{X}\}_d^p)$ which, in a discrete-time implementation, can be achieved with complexity $\mathcal{O}(2^D N \log N)$ where N is the block length corresponding to $T_{\max}$.

4. STATISTICAL ANALYSIS

We next provide a statistical analysis of the filter weight estimates $\hat{h}_{k,l}$ in (5) and the subspace MSE estimates $\hat{e}_{k,l}$ in (6) for the case where $s(t)$ and $n(t)$ (and thus also $r(t)$) are Gaussian random processes.

We first consider $\hat{h}_{k,l}$ in (5). For simplicity, we assume that $n_{k,l}$ is known or has been estimated very reliably (this assumption is justified for $M \gg 1$, see below). Thus, we can replace $\hat{n}_{k,l}$ by $n_{k,l}$, whence (5) becomes

$$\hat{h}_{k,l} = \max\left\{0, 1 - \frac{n_{k,l}}{\hat{r}_{k,l}}\right\}. \tag{9}$$

Since $\hat{h}_{k,l}$ is determined by $\hat{r}_{k,l} = \|\mathbf{P}_{k,l} r\|^2 = \langle \mathbf{P}_{k,l} r, r\rangle$, we first analyze $\hat{r}_{k,l}$. Even though the probability density function (pdf) and characteristic function of $\hat{r}_{k,l}$ can be determined [12], we here restrict ourselves to a second-order analysis. We just note that if $\mathbf{R}_r^{(k,l)} \triangleq \mathbf{P}_{k,l}\mathbf{R}_r\mathbf{P}_{k,l} = \eta_{k,l}\,\mathbf{P}_{k,l}$ (with $\mathbf{R}_r = \mathbf{R}_s + \mathbf{R}_n$ and $\eta_{k,l} = (s_{k,l} + n_{k,l})/M$), i.e., if $r(t)$ is "white within $\mathcal{X}_{k,l}$," then $\hat{r}_{k,l}/\eta_{k,l}$ is χ^2-distributed with M degrees of freedom. The mean of $\hat{r}_{k,l}$ is

$$\mathrm{E}\{\hat{r}_{k,l}\} = r_{k,l} = \mathrm{tr}\{\mathbf{R}_r^{(k,l)}\} = \sum_{m=1}^{M} \lambda_{k,l}^{(m)},$$

where the $\lambda_{k,l}^{(m)}$ are the eigenvalues of the projected correlation operator $\mathbf{R}_r^{(k,l)} = \mathbf{P}_{k,l}\mathbf{R}_r\mathbf{P}_{k,l}$. The variance of $\hat{r}_{k,l}$ can be shown to be given by

$$\mathrm{var}\{\hat{r}_{k,l}\} = 2\,\mathrm{tr}\{(\mathbf{R}_r^{(k,l)})^2\} = 2\sum_{m=1}^{M}\left(\lambda_{k,l}^{(m)}\right)^2.$$

A measure of the reliability of the estimate $\hat{r}_{k,l}$ is provided by the *relative variance*

$$v_{k,l}^2 \triangleq \frac{\mathrm{var}\{\hat{r}_{k,l}\}}{\left(\mathrm{E}\{\hat{r}_{k,l}\}\right)^2} = \frac{2\sum_{m=1}^{M}\left(\lambda_{k,l}^{(m)}\right)^2}{\left(\sum_{m=1}^{M}\lambda_{k,l}^{(m)}\right)^2},$$

which should be as small as possible. It can be shown that

$$\frac{2}{M} \leq v_{k,l}^2 \leq 2.$$

The lower bound is attained iff $\lambda_{k,l}^{(m)} \equiv \eta_{k,l}$ or equivalently $\mathbf{R}_r^{(k,l)} = \eta_{k,l}\,\mathbf{P}_{k,l}$. Thus, "whiteness of $r(t)$ within $\mathcal{X}_{k,l}$" is desirable. The upper bound is attained iff $\mathbf{R}_r^{(k,l)}$ has rank one, i.e., $\lambda_{k,l}^{(1)} = s_{k,l} + n_{k,l}$ and $\lambda_{k,l}^{(2)} = \cdots = \lambda_{k,l}^{(M)} = 0$. This means that, within $\mathcal{X}_{k,l}$, $r(t)$ is almost deterministic. Clearly, this case is undesirable.

In view of the above discussion, we can expect that for typical situations, the pdf of $\hat{r}_{k,l}$ will become increasingly concentrated about the true value $r_{k,l} = \mathrm{E}\{\hat{r}_{k,l}\}$ for growing subspace dimension M. Thus, the estimate $\hat{r}_{k,l}$ will be more reliable for larger subspace dimensions.

We next analyze $\hat{h}_{k,l}$ in (9). We have

$$\hat{h}_{k,l} = f(\hat{r}_{k,l}) \qquad \text{with } f(x) = \max\left\{0, 1 - \frac{n_{k,l}}{x}\right\}.$$

Using truncated Taylor series expansions of $f(\hat{r}_{k,l})$ about $\mathrm{E}\{\hat{r}_{k,l}\} = r_{k,l}$, we obtain the following approximations for the bias and variance of $\hat{h}_{k,l}$,

$$\mathrm{E}\{\hat{h}_{k,l}\} - h_{k,l} \approx -\frac{v_{k,l}^2}{\mathrm{SNR}_{k,l} + 1}, \qquad \mathrm{var}\{\hat{h}_{k,l}\} \approx n_{k,l} v_{k,l}^2,$$

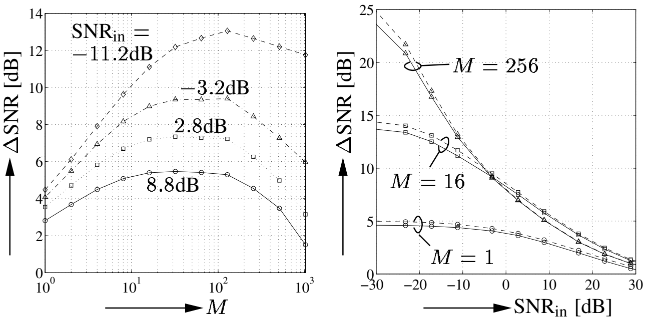
with the "subspace signal-to-noise ratio" $\mathrm{SNR}_{k,l} \triangleq s_{k,l}/n_{k,l}$. (These approximations are the better the more the pdf of $\hat{r}_{k,l}$ is concentrated about the true value $r_{k,l}$.) Thus, even though $\hat{r}_{k,l}$ is an unbiased estimate, $\hat{h}_{k,l}$ is biased. The bias will be small if $\mathrm{SNR}_{k,l}$ is large and/or $v_{k,l}^2$ is small (the latter will be particularly true for $M \gg 1$). Furthermore, the variance of $\hat{h}_{k,l}$ will be small if $n_{k,l}$ and/or $v_{k,l}^2$ is small. Thus, small noise and/or small relative variance of $\hat{r}_{k,l}$ ensure that the estimate $\hat{h}_{k,l}$ is reliable.

In a similar manner, it can be shown that the bias and variance of the subspace MSE estimates $\hat{e}_{k,l}$ in (6) can be approximated as

$$\mathrm{E}\{\hat{e}_{k,l}\} - e_{k,l} \approx -\frac{v_{k,l}^2\, n_{k,l}}{\mathrm{SNR}_{k,l} + 1}, \qquad \mathrm{var}\{\hat{e}_{k,l}\} \approx n_{k,l}^3 v_{k,l}^2.$$

Here, we again assumed that $n_{k,l}$ is known or has been estimated very reliably.

This statistical analysis shows that in a signal-adaptive implementation of robust time-varying Wiener filters, a large subspace dimension M is advantageous as it tends to yield more reliable estimates of the Wiener filter weights $h_{k,l}$ and the subspace MSEs $e_{k,l}$, which can be expected to result in a smaller signal estimation MSE. On the other hand, a larger M also results in a loss of filtering resolution and, thus, potentially in a larger signal estimation MSE. In view of this tradeoff, one may expect that there is an optimal dimension M that yields minimal signal estimation MSE. This will be verified experimentally in Section 5.



(a) (b)

Figure 2: *SNR improvement achieved by the signal-adaptive robust Wiener filter with (a) fixed interval length $T_k \equiv 92.9$ ms and various subspace dimensions M and (b) best subspace selection using $T_{\max} = 92.9$ ms and $D = 5$ (- - clairvoyant costs, — estimated costs).*

5. NUMERICAL EXPERIMENTS

We finally study the performance of our filter in a speech enhancement context. Our speech signal consisted of three German sentences spoken by a male and a female speaker (total duration 7.68 s, sampling rate 11025 Hz). The speech signal was corrupted by stationary white Gaussian noise.

Simulation 1. In order to show the advantage of subspace dimensions $M > 1$, we first used a filter based on a fixed LCS partition corresponding to fixed interval length $T_k \equiv 92.9$ ms. Fig. 2(a) shows the SNR improvement[4] $\Delta\text{SNR} = \text{SNR}_{\text{out}} - \text{SNR}_{\text{in}}$ achieved by the filter versus $M = 2^m$ ($m = 0, \ldots, 10$) for four different input SNRs. In all four cases, $M \gg 1$ offers significant performance gains with $M = 32 \ldots 64$ being a good choice for most SNR levels. However, for M growing further, the performance starts to deteriorate since the loss in filtering resolution becomes dominant.

Simulation 2. We next used the filter with best subspace selection ($T_{\max} = 92.9$ ms, $D = 5$) as described in Section 3. For comparison, we additionally show the results obtained with clairvoyant best subspace selection using the "true" subspace MSE costs (estimated using the clean signals) instead of the subspace MSE costs estimated according to Section 3 (the filtering is still signal-adaptive, however). Fig. 2(b) shows the SNR improvement versus input SNR for subspace dimension $M = 1$, 16, and 256. At low input SNR, a larger M is seen to be advantageous. For $\text{SNR}_{\text{in}} > -5$ dB, however, $M = 256$ performs worse than $M = 16$ due to the poorer filtering resolution. Furthermore, using the estimated MSE costs instead of the clairvoyant MSE costs is seen to result only in a small performance degradation.

Fig. 3 shows the time-frequency tiling obtained by the best subspace algorithm (subspace dimension $M = 32$) for a signal block of length $3T_{\max} = 279$ ms. Clearly, the algorithm succeeds in adapting to both short broadband components and long narrowband portions of the speech signal.

6. CONCLUSION

We proposed a robust time-varying Wiener filter that is based on a local cosine subspace partition and allows efficient on-line operation. The filter selects the best subspace partition and estimates the Wiener filter weights in a signal-adaptive manner, with only a minimal amount of prior knowledge required. A statistical analysis and simulation results showed the advantage of choosing subspace dimensions larger than one. This suggests potential improvements of current speech enhancement techniques.

[4]The input and output SNRs are defined as $\text{SNR}_{\text{in}} = ||s||^2/||n||^2$ and $\text{SNR}_{\text{out}} = ||s||^2/||s - \hat{s}||^2$, respectively.

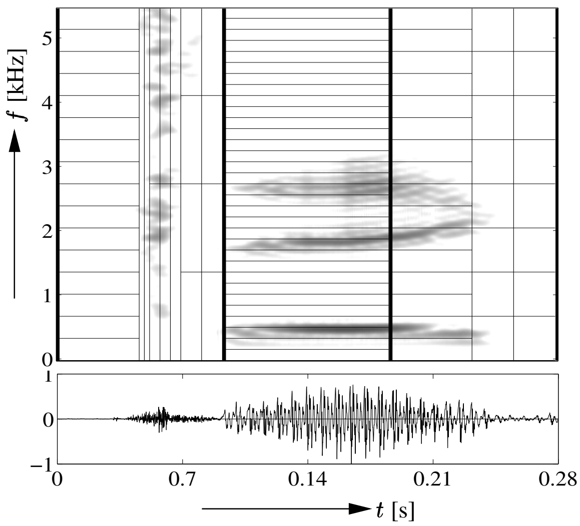


Figure 3: *Best subspace partition for a speech signal block of length $3T_{\max} = 279$ ms. The lower part shows the clean speech signal. The upper part shows a smoothed pseudo-Wigner distribution [13] of the clean speech signal and the time-frequency tiling corresponding to the best subspace partition ($M = 32$, $D = 5$) obtained from the noisy speech signal ($\text{SNR}_{\text{in}} = 2.8$ dB). The thick vertical lines indicate the basic intervals of length $T_{\max} = 92.9$ ms.*

REFERENCES

[1] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory*. New York: Wiley, 1968.

[2] F. Hlawatsch, G. Matz, H. Kirchauer, and W. Kozek, "Time-frequency formulation, design, and implementation of time-varying optimal filters for signal estimation," *IEEE Trans. Signal Processing*, vol. 48, pp. 1417–1432, May 2000.

[3] G. Matz and F. Hlawatsch, "Minimax robust nonstationary signal estimation based on a p-point uncertainty model," *J. Franklin Inst.*, vol. 337, pp. 403–419, July 2000.

[4] G. Matz and F. Hlawatsch, "Minimax robust time-frequency filters for nonstationary signal estimation," in *Proc. IEEE ICASSP-99*, (Phoenix, AZ), pp. 1333–1336, March 1999.

[5] R. R. Coifman and M. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. Inf. Theory*, vol. 38, pp. 713–718, March 1992.

[6] S. G. Mallat, *A Wavelet Tour of Signal Processing*. San Diego: Academic Press, 1998.

[7] R. R. Coifman and Y. Meyer, "Remarques sur l'analyse de Fourier à fenêtre," *C. R. Acad. Sci.*, vol. 312, no. 1, pp. 259–261, 1991.

[8] A. W. Naylor and G. R. Sell, *Linear Operator Theory in Engineering and Science*. New York: Springer, 2nd ed., 1982.

[9] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 32, pp. 1109–1121, Dec. 1984.

[10] H. Krim, D. Tucker, S. Mallat, and D. Donoho, "On denoising and best signal representation," *IEEE Trans. Inf. Theory*, vol. 45, pp. 2225–2238, Nov. 1999.

[11] A. J. Thorpe and L. L. Scharf, "Data adaptive rank shaping methods for solving least squares problems," *IEEE Trans. Signal Processing*, vol. 43, pp. 1591–1601, Nov. 1995.

[12] L. L. Scharf, *Statistical Signal Processing*. Reading (MA): Addison Wesley, 1991.

[13] W. Mecklenbräuker and F. Hlawatsch, eds., *The Wigner Distribution — Theory and Applications in Signal Processing*. Amsterdam (The Netherlands): Elsevier, 1997.