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 $R_s$ and $R_n$. The signal estimate is $\hat{s}(t) = (Hr)(t)$ with $H$ a time-varying system.

The linear system $H$ minimizing the mean-square error (the time-varying Wiener filter) [1, 2] is very sensitive to errors in modeling and/or estimating $R_s$ and $R_n$ [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 $\{X_{k,t}\}_{k \in \mathbb{Z}, t \in \mathbb{N}_0}$ of the real signal space $L_2(\mathbb{R})$ into orthogonal local cosine subspaces (LCSs) [6, 7]

$$X_{k,t} = \text{span}\{u_{k,m}(t)\}_{m=1,\ldots,M}$$

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

$$u_{k,m}(t) = w_k(t) \cos \left( \frac{2(M+m)-1}{2T_k} \pi(t-t_k) \right).$$

Funding by FWF grant P11904-TEC.

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

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

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 $k$th 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 $\{X_{k,t}\}_{k \in \mathbb{Z}, t \in \mathbb{N}_0}$.

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

Robust Wiener Filter. Let us consider a specific LCS partition $\{X_{k,t}\}_{k \in \mathbb{Z}, t \in \mathbb{N}_0}$. The kernels of the associated orthogonal projection operators [8] $P_{k,t}$ are given by $P_{k,t}(t, t') = \sum_{m=1}^M u_{k,m}(t) u_{k,m}^*(t')$. Whereas the correlation operators $R_s$, $R_n$ are unknown, we initially assume that the expected energies of $s(t)$ and $n(t)$ within the subspaces $X_{k,t}$ are known to equal $s_{k,t} \geq 0$ and $n_{k,t} \geq 0$, respectively, i.e., $E[\|P_{k,t}s\|_2^2] \geq s_{k,t}$ and $E[\|P_{k,t}n\|_2^2] \geq n_{k,t}$. The uncertainty classes $\mathcal{S}$ and $\mathcal{N}$ are defined as the sets of all $R_s$ and $R_n$ satisfying the above properties. By definition, the minimax robust time-varying Wiener filter $W_{k,t}$ optimizes the worst-case performance within these uncertainty classes:

$$W_{k,t} = \arg \min_{H} \{ \max_{R_s \in \mathcal{S}, R_n \in \mathcal{N}} E[\|Hr-s\|_2^2] \},$$

with the mean-square error (MSE) $E[\|Hr-s\|_2^2]$. It is shown in [3, 4] that the signal estimate $\hat{s}(t) = (W_{k,t})r(t)$ equals

$$\hat{s}(t) = \sum_{k=-\infty}^{\infty} \sum_{h=0}^{M} h_{k,t} P_{k,h}(t) r(t)$$

with $h_{k,t} = \frac{s_{k,t}}{s_{k,t} + n_{k,t}}$, $P_{k,h}(t)$ is

$$P_{k,h}(t) = \sum_{m=1}^M u_{k,m}(t) u_{k,m}^*(t),$$

where

$$\sum_{k=-\infty}^{\infty} c_{k,t} = \sum_{k=-\infty}^{\infty} s_{k,t}$$

and

$$c_{k,t} = \sum_{k=-\infty}^{\infty} s_{k,t} n_{k,t}.$$
dim[Xk,t] > 1 entails a resolution loss (since all components of r(t) lying in a given subspace Xk,t are treated alike) but will turn out to be advantageous in a signal-adaptive implementation.

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

\[
    r_{k,t} \triangleq s_{k,t} + n_{k,t} = E[\|P_{k,t} r\|^2] = E[\|P_{k,t} r\|^2] - E[\|P_{k,t} n\|^2],
\]

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

\[
    \hat{h}_{k,t} = \max \left\{ 0, 1 - \frac{n_{k,t}}{r_{k,t}} \right\},
\]

where \( n_{k,t} \) and \( r_{k,t} \) are suitable estimates of \( n_{k,t} \) and \( r_{k,t} \), respectively. An unbiased estimate of \( r_{k,t} \) is provided by

\[
    \hat{r}_{k,t} = \|P_{k,t} r\|^2.
\]

The statistical properties of the estimates \( r_{k,t} \) and \( \hat{h}_{k,t} \) will be studied in Section 4.

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

\[
    \hat{n}_{k,t} = \hat{r}_{k,t} - \hat{s}_{k,t} - \hat{h}_{k,t} \cdot \hat{n}_{k,t} \quad \text{for all} \quad (k,l) \in I_i.
\]

Of practical relevance is the special case where each subset \( I_i \) corresponds to some frequency index \( f \) and some time interval \( [k_j,k_{j+1}] \) \( \subseteq I_i \), i.e., \( I_i = \{(k_j,l),(k_{j+1},l),\ldots,(k_{j+1}-1,l)\} \). That is, for each frequency index \( k \), \( n_{k,t} \) is constant on the time interval \( [k_j,k_{j+1}] \) (which may depend on \( l \)) and at least for one \((k_0,l) \in I_1 \) there is \( n_{k_0,l_0} = 0 \). Note that the first property corresponds to a “generalized stationarity” of \( n(t) \). A dual situation is the case where each subset \( I_i \) corresponds to some frequency index \( f \) and some time interval \( [l_j,l_{j+1}] \) \( \subseteq I_i \), i.e., \( I_i = \{(k,l_j),(k,l_{j+1}),\ldots,(k,l_{j+1}-1)\} \). Here, \( n_{k,t} \) is locally constant with respect to the frequency index \( f \), corresponding to a “generalized whiteness” of \( n(f) \).

3. BEST SUBSPACE SELECTION

We now propose a “best subspace” algorithm that allows to adapt the LCS partition \( \{X_{k_0,t}\}_{k_0,t} \) (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 \( T_{\text{max}} \) (\( T_{\text{max}} \) determines the finest frequency resolution of the filter). For each such interval—e.g., the interval \([0, T_{\text{max}}] \)—we construct a dyadic LCS tree (see Fig. 1) by recursively splitting the tree into admissible trees if each node has either no or two children. The statistical properties of the estimates \( \hat{n}_{k,t} \) and \( \hat{h}_{k,t} \) will be studied in Section 4.

### Cost Function

A statistical analysis of the subspace MSE estimates \( \hat{e}_{k,t} \) is given by

\[
    \hat{e}_{k,t} \triangleq \frac{\hat{s}_{k,t} - \hat{n}_{k,t} \cdot \hat{h}_{k,t}}{r_{k,t}} \quad \text{with} \quad \hat{n}_{k,t} = \max \left\{ 0, \frac{\hat{r}_{k,t} - \hat{s}_{k,t}}{\hat{h}_{k,t}} \right\}.
\]

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

### Best Subspace Algorithm

For a given observed signal \( r(t) \), the best subspace partition \( \mathcal{B}_0 \) associated to the interval \([0, T_{\text{max}}] \) is defined as the \( \mathcal{B}_0 \) minimizing the estimated MSE.

\[
    \mathcal{B}_0 \triangleq \arg \min \hat{E}(\mathcal{B}_0).
\]
We next provide a statistical analysis of the filter weight estimates $\hat{h}_{\ell,t}$ in (5) and the subspace MSE estimates $\hat{e}_{\ell,t}$ in (6) for the case where $s(t)$ and $n(t)$ (and thus also $r(t)$) are Gaussian random processes.

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

$$\hat{h}_{\ell,t} = \max \left\{ 0, 1 - \frac{n_{\ell,t}}{\|r_{\ell,t}\|} \right\}, \quad (9)$$

Since $\hat{h}_{\ell,t}$ is determined by $r_{\ell,t} = \|P_{\ell,t}r\|^2 = (P_{\ell,t}r, r)$, we first analyze $r_{\ell,t}$. Even though the probability density function (pdf) and characteristic function of $r_{\ell,t}$ can be determined [12], we here restrict ourselves to a second-order analysis. We just note that if $R_{\ell,t} = P_{\ell,t}R_tP_{\ell,t}$ is an unbiased estimate of $R_t$, then $r_{\ell,t}$ is $\chi^2$-distributed with $M$ degrees of freedom.

The mean of $r_{\ell,t}$ is

$$E[r_{\ell,t}] = r_{\ell,t} = \text{tr} \left( R_{\ell,t}^{(h,t)} \right) = \sum_{m=1}^{M} \lambda_{\ell,t}^{(m)},$$

where the $\lambda_{\ell,t}^{(m)}$ are the eigenvalues of the projected correlation operator $R_{\ell,t}^{(h,t)} = P_{\ell,t}R_tP_{\ell,t}$. The variance of $r_{\ell,t}$ can be shown to be

$$\text{var}(r_{\ell,t}) = 2 \text{tr} \left( \left( R_{\ell,t}^{(h,t)} \right)^2 \right) = 2 \sum_{m=1}^{M} \left( \lambda_{\ell,t}^{(m)} \right)^2.$$  

A measure of the reliability of the estimate $\hat{h}_{\ell,t}$ is provided by the relative variance

$$\nu_{\ell,t}^2 \triangleq \frac{\text{var}(\hat{r}_{\ell,t})}{(E[r_{\ell,t}])^2} = \frac{2 \sum_{m=1}^{M} \left( \lambda_{\ell,t}^{(m)} \right)^2}{(\sum_{m=1}^{M} \lambda_{\ell,t}^{(m)})^2},$$

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

$$\frac{2}{M} \leq \nu_{\ell,t}^2 \leq 2.$$  

The lower bound is attained iff $\lambda_{\ell,t}^{(m)} = \eta_{\ell,t}$, or equivalently $R_{\ell,t}^{(h,t)} = \eta_{\ell,t}P_{\ell,t}$. Thus, “whiteness of $r(t)$” within $X_{\ell,t}$ is desirable. The upper bound is attained iff $R_{\ell,t}^{(h,t)}$ has rank one, i.e., $\lambda_{\ell,t}^{(m)} = \eta_{\ell,t} + n_{\ell,t}$, and $\lambda_{\ell,t}^{(2)} = \cdots = \lambda_{\ell,t}^{(M)} = 0$. This means that, within $X_{\ell,t}$, $r(t)$ is almost deterministic. Clearly, this case is undesirable.

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

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

$$\hat{h}_{\ell,t} = f(\hat{r}_{\ell,t}) = f(x) = \max \left\{ 0, 1 - \frac{n_{\ell,t}}{x} \right\},$$

Using truncated Taylor series expansions of $f(x)$ about $E[\hat{r}_{\ell,t}] = r_{\ell,t}$, we obtain the following approximations for the bias and variance of $\hat{h}_{\ell,t}$:

$$E[\hat{h}_{\ell,t}] - h_{\ell,t} \approx -\frac{\nu_{\ell,t}^2}{\text{SNR}_{\ell,t} + 1}, \quad \text{var}(\hat{h}_{\ell,t}) \approx \frac{\nu_{\ell,t}^2}{\text{SNR}_{\ell,t} + 1}.$$  

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

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

$$E[\hat{e}_{\ell,t}] - e_{\ell,t} \approx -\frac{\nu_{\ell,t}^2}{\text{SNR}_{\ell,t} + 1}, \quad \text{var}(\hat{e}_{\ell,t}) \approx \frac{3 \nu_{\ell,t}^2}{\text{SNR}_{\ell,t} + 1}.$$  

Here, we again assumed that $n_{\ell,t}$ 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 $\hat{h}_{\ell,t}$ and the subspace MSEs $\hat{e}_{\ell,t}$, 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.
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 1025 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 $T_k = 92.9$ ms. Fig. 2(a) shows the SNR improvement $\Delta \text{SNR} = \text{SNR}_{\text{out}} - \text{SNR}_{\text{in}}$ achieved by the filter versus $M = 2^n$ ($n = 0, \ldots, 10$) for four different input SNRs. In all four cases, $M \geq 1$ 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 ($T_{\text{max}} = 92.9$ ms, $D = 5$) as described in Section 3. For comparison, we additionally show the results obtained with clairvoyant 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 $M = 1, 16, 256$. 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 speech block of length $3T_{\text{max}} = 279$ 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.

---

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

---

**REFERENCES**


