Time–Frequency Formulation, Design, and Implementation of Time-Varying Optimal Filters for Signal Estimation

Franz Hlawatsch, *Member, IEEE*, Gerald Matz, *Student Member, IEEE*, Heinrich Kirchauer, and Werner Kozek, *Member, IEEE*

***Abstract*—This paper presents a time–frequency framework for optimal linear filters (signal estimators) in nonstationary environments. We develop *time–frequency formulations* for the optimal linear filter (time-varying Wiener filter) and the optimal linear time-varying filter under a projection side constraint. These time–frequency formulations extend the simple and intuitive spectral representations that are valid in the stationary case to the practically important case of underspread nonstationary processes. Furthermore, we propose an approximate *time–frequency design* of both optimal filters, and we present bounds that show that for underspread processes, the time-frequency designed filters are nearly optimal. We also introduce extended filter design schemes using a weighted error criterion, and we discuss an efficient *time–frequency implementation* of optimal filters using multiwindow short-time Fourier transforms. Our theoretical results are illustrated by numerical simulations.**

***Index Terms*—Nonstationary random processes, optimal filters, signal enhancement, signal estimation, time-frequency analysis, time-varying systems, Wiener filters.**

I. Introduction

THE enhancement or estimation of signals corrupted by noise or interference is important in many signal processing applications. For stationary random processes, the mean-square error optimal linear estimator is the (time-invariant) *Wiener filter* [1]–[7], whose transfer function is given by

$$H(f) = \frac{S_s(f)}{S_s(f) + S_n(f)} \tag{1}$$

where $S_s(f)$ and $S_n(f)$ denote the power spectral densities of the signal and noise processes, respectively. The extension of the Wiener filter to nonstationary processes yields a linear, time-varying filter ("time-varying Wiener filter") [2], [4]–[7] for which the simple frequency-domain formulation in (1) is no longer valid. We may ask whether a similarly simple formulation can be obtained by introducing an explicit time dependence, i.e., whether in a joint time-frequency (TF) domain the time-varying Wiener filter can be formulated as

$$H(t,f) = \frac{S_s(t,f)}{S_s(t,f) + S_n(t,f)} \tag{2}$$

where $H(t,f)$ and $S_s(t,f), S_n(t,f)$ are suitably defined. Such a TF formulation would greatly facilitate the interpretation, analysis, design, and implementation of time-varying Wiener filters.

In this paper, we provide an answer to this and several other questions of theoretical and practical importance.

- We show that for *underspread* [8]–[12] nonstationary processes, the TF formulation in (2) is approximately valid if $H(t,f)$ and $S_s(t,f), S_n(t,f)$ are chosen as the Weyl symbol [13]–[16] and the Wigner–Ville spectrum (WVS) [12], [17]–[19], respectively. We present upper bounds on the associated approximation errors.
- We propose an efficient, intuitive, and nearly optimal TF design of signal estimators that is easily adapted to modified (weighted) error criteria.
- We discuss an efficient TF implementation of optimal filters using the multiwindow short-time Fourier transform (STFT) [8], [20].
- We also consider the TF formulation, approximate TF design, and TF implementation of an optimal projection filter.

Previous work on the TF formulation of time-varying Wiener filters [21]–[24] has mostly used Zadeh's time-varying transfer function [25], [26] for $H(t,f)$ and the evolutionary spectrum [12], [27]–[29] for $S_s(t,f)$ and $S_n(t,f)$. A Wiener filtering procedure using a time-varying spectrum based on subband AR models has been proposed in [30]. A somewhat different approach is the TF implementation of signal enhancement by means of a TF weighting applied to a linear TF signal representation such as the STFT, the wavelet transform, or a filterbank/subband representation. In [31] and [32], the STFT is multiplied by a TF weighting function similar to (2), where $S_s(t,f), S_n(t,f)$ are chosen as the *physical spectrum* [18], [19], [33]. Wiener filter-type modifications of the STFT and subband signals have long been used for speech enhancement [34]–[37]. Methods that perform a Wiener-type weighting of wavelet transform coefficients are described in [38]–[41].

Manuscript received June 17, 1998; revised September 20, 1999. This work was supported by FWF Grants P10012-ÖPH and P11904-TEC. The associate editor coordinating the review of this paper and approving it for publication was Prof. P. C. Ching.

F. Hlawatsch and G. Matz are with the Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology, Vienna, Austria.

H. Kirchauer is with Intel Corp., Santa Clara CA 95052 USA.

W. Kozek is with Siemens AG, Munich, Germany.

Publisher Item Identifier S 1053-587X(00)03289-X.

1053–587X/00$10.00 © 2000 IEEE

The TF formulations proposed in this paper differ from the above-mentioned work on several points:

- While, usually, Zadeh's time-varying transfer function and the evolutionary spectrum or physical spectrum are chosen for $H(t,f)$ and $S_s(t,f), S_n(t,f)$, respectively, our development is based on the Weyl symbol and the WVS. This has important advantages, especially regarding TF resolution. Specifically, the Weyl symbol and the WVS are much better suited to systems and processes with chirp-like characteristics. Furthermore, contrary to the evolutionary spectrum and physical spectrum, the WVS is in one-to-one correspondence with the correlation function.
- Our TF formulations are not merely heuristic but are shown to yield approximations to the optimal filters featuring nearly optimal performance in the case of jointly underspread signal and noise processes. We present bounds on the associated approximation errors, and we discuss explicit conditions for the validity of our TF formulations. In particular, we show that the usual quasi-stationarity assumption is neither a sufficient nor a necessary condition.
- Contrary to the conventional heuristic STFT enhancement schemes, the multiwindow STFT implementation discussed here is based on a theoretical analysis that shows how to choose the STFT window functions and how well the multiwindow STFT filter approximates the optimal filter.

The paper is organized as follows. Section II reviews the time-varying Wiener filter and the optimal time-varying projection filter. Section III reviews some TF representations and the concept of underspread systems and processes. In Section IV, a TF formulation of optimal filters is developed. Based on this TF formulation, Section V introduces simple and intuitive TF design methods for optimal filters. In Section VI, a multiwindow STFT implementation of optimal filters is discussed. Finally, Section VII presents numerical simulations that compare the performance of the various filters and verify the theoretical results.

II. Optimal Filters

This section reviews the theory of the time-varying Wiener filter [2], [4]–[7] and a recently proposed "projection-constrained" optimal filter [42]. For stationary processes,there exist simple frequency-domain formulations of both optimal filters. These will be generalized to the nonstationary case in Section IV.

A. Time-Varying Wiener Filter

Let $s(t)$ be a zero-mean, nonstationary, circular complex or real random process with known correlation function $r_s(t,t') = \mathrm{E}\{s(t)\, s^*(t')\}$ or, equivalently, correlation operator[1] $\mathbf{R}_s$. This signal process is contaminated by additive zero-mean, nonstationary, circular complex or real random noise $n(t)$ with known correlation operator $\mathbf{R}_n$. Signal $s(t)$ and noise $n(t)$ are assumed to be uncorrelated, i.e., $\mathrm{E}\{s(t)\, n^*(t')\} \equiv 0$. We form an estimate $\hat{s}(t)$ of the signal $s(t)$ from the noisy observation $r(t) = s(t)+n(t)$ using a linear, generally time-varying system (linear operator [43]) $\mathbf{H}$ with impulse response (kernel) $h(t,t')$,[2]

$$\hat{s}(t) = (\mathbf{H}r)(t) = \int_{t'} h(t,t') r(t')\, dt'.$$

Our performance index will be the mean-square error (MSE), i.e., the expected energy $\bar{E}_e = \mathrm{E}\{\|e\|^2\} = \mathrm{E}\{\int_t |e(t)|^2\, dt\} = \mathrm{tr}\{\mathbf{R}_e\}$ of the estimation error $e(t) = s(t) - \hat{s}(t)$, which can be shown to be given by

$$\begin{aligned}\bar{E}_e &= \bar{E}_{n_r} + \bar{E}_{s_d} \\ &= \mathrm{tr}\{\mathbf{H}\mathbf{R}_n\mathbf{H}^+\} + \mathrm{tr}\{(\mathbf{I}-\mathbf{H})\mathbf{R}_s(\mathbf{I}-\mathbf{H})^+\}. \end{aligned} \tag{3}$$

Here, $\mathrm{tr}\{\cdot\}$ denotes the trace of an operator, $\mathbf{H}^+$ denotes the adjoint of $\mathbf{H}$ [43], and $\bar{E}_{n_r}$ and $\bar{E}_{s_d}$ are the expected energies of the residual noise $n_r(t) = (\mathbf{H}n)(t)$ and the signal distortion $s_d(t) = ((\mathbf{I}-\mathbf{H})s)(t)$, respectively. The linear system minimizing the MSE $\bar{E}_e$ can be shown [2], [4]–[7] to satisfy

$$\mathbf{H}(\mathbf{R}_s + \mathbf{R}_n) = \mathbf{R}_s. \tag{4}$$

The solution of (4) with minimal rank[3] is the "time-varying Wiener filter"

$$\mathbf{H}_o = \mathbf{R}_s(\mathbf{R}_s + \mathbf{R}_n)^{-1} \tag{5}$$

where $(\mathbf{R}_s + \mathbf{R}_n)^{-1}$ denotes the (pseudo-)inverse of $\mathbf{R}_r = \mathbf{R}_s + \mathbf{R}_n$ on its range $\mathcal{S}_r = \mathrm{range}\{\mathbf{R}_r\}$. The minimum MSE achieved by the Wiener filter is given by

$$\begin{aligned}\bar{E}_{e_o} &= \mathrm{tr}\{\mathbf{R}_s(\mathbf{R}_s+\mathbf{R}_n)^{-1}\mathbf{R}_n\} \\ &= \mathrm{tr}\{\mathbf{H}_o\mathbf{R}_n\} = \mathrm{tr}\{(\mathbf{I}-\mathbf{H}_o)\mathbf{R}_s\}. \end{aligned} \tag{6}$$

In order to interpret the time-varying Wiener filter $\mathbf{H}_o$, let us define the *signal space* $\mathcal{S}_s = \mathrm{range}\{\mathbf{R}_s\}$ and the *noise space* $\mathcal{S}_n = \mathrm{range}\{\mathbf{R}_n\}$. Since $r(t) = s(t) + n(t)$, the *observation space* is given by $\mathcal{S}_r = \mathcal{S}_s + \mathcal{S}_n$. We note that[4] $s(t) \in \mathcal{S}_s, n(t) \in \mathcal{S}_n$, and $r(t) \in \mathcal{S}_r$. It can now be shown that the optimal signal estimate is an element of the signal space $\mathcal{S}_s$, i.e., $\hat{s}_o(t) \in \mathcal{S}_s$. Since $s(t) \in \mathcal{S}_s$ and $\hat{s}_o(t) \in \mathcal{S}_s$, we also have $e_o(t) \in \mathcal{S}_s$. Since it can also be shown that $e_o(t) \in \mathcal{S}_n$, we get $e_o(t) \in \mathcal{S}_s \cap \mathcal{S}_n$. These results are intuitive since any contribution to $\hat{s}_o(t)$ from outside $\mathcal{S}_s$ would unnecessarily increase the resulting error. Furthermore, if signal space $\mathcal{S}_s$ and noise space $\mathcal{S}_n$ are disjoint, $\mathcal{S}_s \cap \mathcal{S}_n = \{0\}$, then the Wiener filter performs an oblique projection [44] onto $\mathcal{S}_s$ and obtains perfect reconstruction of the signal $s(t)$, i.e., $e_o(t) = 0$.

[1]The correlation operator $\mathbf{R}_x$ of a random process $x(t)$ is the self-adjoint and positive (semi-)definite linear operator [43] whose kernel is the correlation function $r_x(t,t') = \mathrm{E}\{x(t)\, x^*(t')\}$. In a discrete-time setting, $\mathbf{R}_x$ would be a matrix.

[2]Integrals are from $-\infty$ to ∞ unless stated otherwise.

[3]The general solution of (4) is $\mathbf{H}_o + \mathbf{X}\mathbf{P}_r^\perp$, where $\mathbf{H}_o$ denotes the minimal rank solution in (5), $\mathbf{X}$ is an arbitrary operator, and $\mathbf{P}_r^\perp$ is the orthogonal projection operator on $\mathcal{S}_r^\perp$, the orthogonal complement space of $\mathcal{S}_r = \mathrm{range}\{\mathbf{R}_r\}$ with $\mathbf{R}_r = \mathbf{R}_s + \mathbf{R}_n$. (For $\mathcal{S}_r = L_2(\mathbb{R})$ [where $L_2(\mathbb{R})$ denotes the space of square-integrable functions], $\mathbf{H}_o$ becomes the unique solution of (4).

[4]Since the signals $s(t)$ etc. are random, relations like $s(t) \in \mathcal{S}_s$ are to be understood to hold with probability 1.

In some applications, e.g., if signal distortions are less acceptable than residual noise, we might consider replacing the MSE in (3) by a "weighted MSE"

$$\begin{aligned}\bar{E}_e^{(\alpha)} &\triangleq \alpha \bar{E}_{n_r} + (1-\alpha)\bar{E}_{s_d} \\ &= \alpha \, \mathrm{tr}\{\mathbf{H}\mathbf{R}_n\mathbf{H}^+\} \\ &\quad + (1-\alpha)\, \mathrm{tr}\{(\mathbf{I}-\mathbf{H})\mathbf{R}_s(\mathbf{I}-\mathbf{H})^+\} \end{aligned} \tag{7}$$

with $0 \le \alpha \le 1$. This amounts to replacing $\mathbf{R}_n$ by $\alpha\mathbf{R}_n$ and $\mathbf{R}_s$ by $(1-\alpha)\mathbf{R}_s$ so that the resulting modified Wiener filter is

$$\mathbf{H}_o^{(\alpha)} = (1-\alpha)\mathbf{R}_s[(1-\alpha)\mathbf{R}_s + \alpha\mathbf{R}_n]^{-1}.$$

B. Optimal Time-Varying Projection Filter

We next consider the estimation of $s(t)$ using a linear, time-varying filter $\mathbf{P}$ that is an *orthogonal projection operator*, i.e., it is idempotent, $\mathbf{P}^2 = \mathbf{P}$, and self-adjoint, $\mathbf{P}^+ = \mathbf{P}$ [43]. Although the projection constraint generally results in a larger MSE, it leads to an estimator that is robust in a specific sense [45]. Further advantages regarding a TF design are discussed in Subsection V-B.

Let μ_k and $v_k(t)$ denote the eigenvalues and normalized eigenfunctions, respectively, of the operator $\mathbf{D} = \mathbf{R}_s - \mathbf{R}_n$, i.e., $(\mathbf{D}v_k)(t) = \mu_k v_k(t)$. It is shown in [42] that the minimum-rank orthogonal projection operator minimizing the MSE in (3) is given by

$$\mathbf{P}_o = \sum_{k \in I_+} \mathbf{P}_{v_k}$$

where $\mathbf{P}_{v_k}$ denotes the rank-one orthogonal projection operator defined by $(\mathbf{P}_{v_k}x)(t) = \langle x, v_k\rangle v_k(t)$ with $\langle x, v_k \rangle = \int_t x(t)v_k^*(t)\,dt$, and $I_+ = \{k : \mu_k > 0\}$ is the index set corresponding to positive eigenvalues. Thus, the optimal projection filter $\mathbf{P}_o$ performs a projection onto the space $\mathcal{S}_+$ spanned by all eigenfunctions of $\mathbf{D} = \mathbf{R}_s - \mathbf{R}_n$ corresponding to positive eigenvalues. The MSE obtained with $\mathbf{P}_o$ is given by

$$\bar{E}_{e_p} = \mathrm{tr}\{\mathbf{P}_o\mathbf{R}_n\} + \mathrm{tr}\{\mathbf{P}_o^{\perp}\mathbf{R}_s\} = \bar{E}_s - \sum_{k\in I_+} \mu_k. \tag{8}$$

For an interpretation of this result, let us partition the observation space $\mathcal{S}_r$ into three orthogonal subspaces corresponding to positive, zero, and negative eigenvalues of $\mathbf{D}$:

$$\begin{aligned}\mathcal{S}_+ &= \mathrm{span}\{v_k(t)\colon \mu_k > 0\} \\ \mathcal{S}_= &= \mathrm{span}\{v_k(t) \in \mathcal{S}_r\colon \mu_k = 0\} \\ \mathcal{S}_- &= \mathrm{span}\{v_k(t)\colon \mu_k < 0\}\end{aligned}$$

so that $\mathcal{S}_+ + \mathcal{S}_= + \mathcal{S}_- = \mathcal{S}_r$ and $\mathbf{P}_+ + \mathbf{P}_= + \mathbf{P}_- = \mathbf{P}_r$, where $\mathbf{P}_+, \mathbf{P}_=, \mathbf{P}_-$, and $\mathbf{P}_r$ are the orthogonal projection operators on the spaces $\mathcal{S}_+, \mathcal{S}_=, \mathcal{S}_-$, and $\mathcal{S}_r$, respectively. Note that $\mathbf{P}_+ = \mathbf{P}_o$. We now consider the expected energies of the signal and noise processes in the space $\mathcal{S}_+$. The expected energy of any process $x(t)$ in the one-dimensional (1-D) space $\mathrm{span}\{v_k(t)\}$ is given by $\mathrm{E}\{|\langle x, v_k\rangle|^2\} = \langle \mathbf{R}_x v_k, v_k\rangle$. Hence, the difference between expected signal energy and expected noise energy in $\mathrm{span}\{v_k(t)\}$ is

$$\begin{aligned}\mathrm{E}\{|\langle s, v_k\rangle|^2\} - \mathrm{E}\{|\langle n, v_k\rangle|^2\} &= \langle \mathbf{R}_s v_k, v_k\rangle - \langle \mathbf{R}_n v_k, v_k\rangle \\ &= \langle \mathbf{D}v_k, v_k\rangle = \mu_k.\end{aligned}$$

Thus, if (and only if) $\mu_k > 0$, then in $\mathrm{span}\{v_k(t)\}$, the expected signal energy is larger than the expected noise energy. It is then easily shown that $\mathrm{E}\{|\langle s, y\rangle|^2\} > \mathrm{E}\{|\langle n, y\rangle|^2\}$ for any signal $y(t) \in \mathcal{S}_+$, i.e., in $\mathcal{S}_+$, the expected signal energy is larger than the expected noise energy. Equivalently, $\mathrm{E}\{\|\mathbf{P}_o s\|^2\} > \mathrm{E}\{\|\mathbf{P}_o n\|^2\}$, which shows that the optimal projection filter $\mathbf{P}_o = \mathbf{P}_+$ projects onto the space $\mathcal{S}_+$ where, on average, there is more signal energy than noise energy. This is intuitively reasonable.

If we use the "weighted MSE" (7) instead of the MSE (3), the resulting optimal projection filter $\mathbf{P}_o^{(\alpha)}$ is constructed as before with $\mathbf{D} = \mathbf{R}_s - \mathbf{R}_n$ replaced by $\mathbf{D}^{(\alpha)} = (1-\alpha)\mathbf{R}_s - \alpha\mathbf{R}_n$.

C. Stationary Processes

The special cases of stationary processes and nonstationary white processes allow particularly simple frequency-domain and time-domain formulations, respectively, of both optimal filters. These formulations will motivate the development of TF formulations of optimal filters in Section IV.

If $s(t)$ and $n(t)$ are wide-sense stationary processes, the corresponding correlation operators $\mathbf{R}_s$ and $\mathbf{R}_n$ are time-invariant, i.e., of convolution type, and the optimal system becomes time-invariant as well. The MSE in (3) does not exist and must be replaced by the "instantaneous MSE" $\mathrm{E}\{|e(t)|^2\}$. For any linear, time-invariant system $\mathbf{H}$ with transfer function $H_{\mathbf{H}}(f)$, the instantaneous MSE can be shown to be

$$\begin{aligned}\mathrm{E}\{|e(t)|^2\} = &\int_f |H_{\mathbf{H}}(f)|^2 S_n(f)\,df \\ &+ \int_f |1 - H_{\mathbf{H}}(f)|^2 S_s(f)\,df\end{aligned} \tag{9}$$

where $S_s(f)$ and $S_n(f)$ are the power spectral densities (PSD's) of $s(t)$ and $n(t)$, respectively. The optimal system minimizing (9) satisfies $H_{\mathbf{H}}(f)[S_s(f) + S_n(f)] = S_s(f)$ [2]–[7]. A "minimal" solution of this equation is

$$H_{\mathbf{H}_o}(f) = \begin{cases} \dfrac{S_s(f)}{S_s(f)+S_n(f)}, & f \in F_r \\ 0, & f \notin F_r \end{cases} \tag{10}$$

where $F_r = \{f : S_s(f) + S_n(f) > 0\}$ denotes the set of frequencies where the PSD of the observation $r(t)$, $S_r(f) = S_s(f) + S_n(f)$, is positive. The minimum instantaneous MSE is given by

$$\mathrm{E}\{|e_o(t)|^2\} = \int_f \frac{S_s(f)\,S_n(f)}{S_s(f)+S_n(f)}\,df. \tag{11}$$

The frequency-domain formulations (10) and (11) allow an intuitively appealing interpretation of the time-invariant Wiener filter (see Fig. 1). Let $F_s = \{f : S_s(f) > 0\}$ and $F_n = \{f : S_n(f) > 0\}$ denote the sets of frequencies where $S_s(f)$ and $S_n(f)$, respectively, are positive. Then, (10) shows

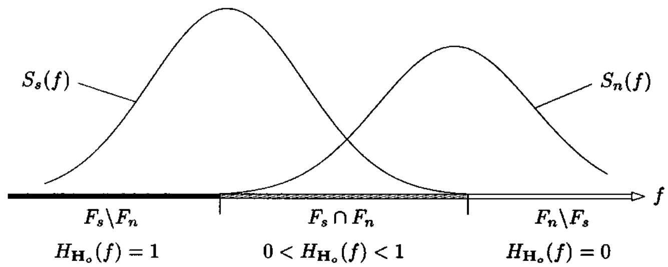


Fig. 1. Frequency-domain interpretation of the time-invariant Wiener filter.

that $H_{\mathbf{H}_o}(f) = 1$ for $f \in F_s \backslash F_n$, i.e., the Wiener filter passes all observation components in the frequency bands where only signal energy is present, which is intuitively reasonable. Furthermore, $H_{\mathbf{H}_o}(f) = 0$ for $f \notin F_s$, i.e., outside F_s, the observation is suppressed, which is again reasonable since outside F_s, there is only noise energy. In the frequency ranges $F_s \cap F_n$, where both signal and noise are present, there is $0 < H_{\mathbf{H}_o}(f) < 1$ with the values of $H_{\mathbf{H}_o}(f)$ depending on the relative values of $S_s(f)$ and $S_n(f)$ at the respective frequency.

The MSE in (9) can again be replaced by a weighted MSE. Here, the weights can even be frequency dependent, which results in the weighted MSE

$$\int_f [\alpha(f)|H_{\mathbf{H}}(f)|^2 \, S_n(f) + [1-\alpha(f)]|1 - H_{\mathbf{H}}(f)|^2 \, S_s(f)] \, df \qquad (12)$$

with $0 \leq \alpha(f) \leq 1$. The optimal filter minimizing this weighted MSE is then simply given by (10) with $S_n(f)$ replaced by $\alpha(f)S_n(f)$ and $S_s(f)$ replaced by $[1-\alpha(f)]\, S_s(f)$.

We next consider the optimal projection filter $\mathbf{P}_o$. For stationary processes, it can be shown that $\mathbf{P}_o$ is time-invariant with the zero-one valued transfer function

$$H_{\mathbf{P}_o}(f) = I_{F_+}(f) \qquad (13)$$

where $I_{F_+}(f)$ is the indicator function of $F_+ = \{f : S_s(f) > S_n(f)\}$, which is the set of frequencies where the signal PSD is larger than the noise PSD. Thus, $\mathbf{P}_o$ is an idealized bandpass filter with one or several pass bands. Note that $H_{\mathbf{P}_o}(f)$ can be obtained from $H_{\mathbf{H}_o}(f)$ by a rounding operation, i.e., $H_{\mathbf{P}_o}(f) = \mathrm{round}\{H_{\mathbf{H}_o}(f)\}$. The instantaneous MSE obtained with $\mathbf{P}_o$ can be shown to be given by

$$\mathrm{E}\{|e_p(t)|^2\} = \int_{F_+} S_n(f)\, df + \int_{F_- \cup F_=} S_s(f)\, df \qquad (14)$$

where $F_- = \{f : S_s(f) < S_n(f)\}$ and $F_= = \{f \in F_r : S_s(f) = S_n(f)\}$.

If the weighted instantaneous MSE (12) is used, then the optimal projection filter is the idealized bandpass filter with passband(s) $F_+^{(\alpha)} = \{f : [1-\alpha(f)]S_s(f) > \alpha(f)\, S_n(f)\}$.

The case of nonstationary white processes $s(t)$ and $n(t)$ is dual to the stationary case discussed above. Here, the time-varying Wiener filter and projection filter are "frequency-invariant," i.e., simple time-domain multiplications. All results and interpretations are dual to the stationary case.

III. Time-Frequency Representations and Underspread Processes

In this section, we review some fundamentals of TF analysis that will be used in Sections IV and V for the TF formulation and TF design of optimal filters.

A. Time–Frequency Representations

We first review four TF representations on which our TF formulations will be based.

- The *Weyl symbol* (WS) [13]–[16] of a linear operator (linear, time-varying system) $\mathbf{H}$ with kernel (impulse response) $h(t, t')$ is defined as

$$L_{\mathbf{H}}(t,f) \triangleq \int_\tau h\left(t+\frac{\tau}{2}, t-\frac{\tau}{2}\right) e^{-j2\pi f\tau}\, d\tau \qquad (15)$$

  where t and f denote time and frequency, respectively. For underspread systems (see Section III-B), the WS can be considered to be a "time-varying transfer function" [8], [11], [46], [47].
- The *Wigner–Ville Spectrum* (WVS) [17]–[19] of a nonstationary random process $x(t)$ is defined as the WS of the correlation operator $\mathbf{R}_x$

$$\bar{W}_x(t,f) \triangleq L_{\mathbf{R}_x}(t,f) = \int_\tau r_x\left(t+\frac{\tau}{2}, t-\frac{\tau}{2}\right) e^{-j2\pi f\tau}\, d\tau. \qquad (16)$$

  For underspread processes (see Section III-B), the WVS can be interpreted as a "time-varying PSD" or expected TF energy distribution of $x(t)$ [8], [9], [12].
- The *spreading function* [8], [13], [16], [26], [46], [48], [49] of a linear operator $\mathbf{H}$ is defined as

$$S_{\mathbf{H}}(\tau,\nu) \triangleq \int_t h\left(t+\frac{\tau}{2}, t-\frac{\tau}{2}\right) e^{-j2\pi\nu t}\, dt$$

  where τ and ν denote time lag and frequency lag, respectively. The spreading function is the 2-D Fourier transform of the WS. Any linear operator $\mathbf{H}$ admits the following expansion into TF shift operators $\mathbf{S}_{\tau,\nu}$ defined as $(\mathbf{S}_{\tau,\nu}\, x)(t) = x(t-\tau)e^{j2\pi\nu t}e^{-j\pi\tau\nu}$ [8], [13], [16], [49]:

$$(\mathbf{H}x)(t) = \int_\tau \int_\nu S_{\mathbf{H}}(\tau,\nu)(\mathbf{S}_{\tau,\nu}x)(t)\, d\tau\, d\nu. \qquad (17)$$

  Hence, the spreading function describes the TF shifts introduced by $\mathbf{H}$ (see Section III-B).
- The *expected ambiguity function* [8]–[12] of a nonstationary process $x(t)$ is defined as the spreading function of the correlation operator $\mathbf{R}_x$

$$\bar{A}_x(\tau,\nu) \triangleq S_{\mathbf{R}_x}(\tau,\nu) = \int_t r_x\left(t+\frac{\tau}{2}, t-\frac{\tau}{2}\right) e^{-j2\pi\nu t}\, dt. \qquad (18)$$

  The expected ambiguity function is the 2-D Fourier transform of the WVS. It can be interpreted as a TF correlation function in the sense that it describes the correlation of process components whose TF locations are separated by τ in time and by ν in frequency [8]–[12].

B. Underspread Systems and Processes

Since our TF formulation of optimal filters will be valid for the class of *underspread* nonstationary processes, we will now review the definition of underspread systems and processes [8]–[12], [46], [47].

According to the expansion (17), the effective support of the spreading function $S_{\mathbf{H}}(\tau,\nu)$ describes the TF shifts caused by a linear time-varying system $\mathbf{H}$. A global description of the TF shift behavior of $\mathbf{H}$ is given by the "displacement spread" $\sigma_{\mathbf{H}}$ [8], [11], [46], which is defined as the area of the smallest rectangle [centered about the origin of the (τ,ν) plane and possibly with oblique orientation] that contains the effective support of $S_{\mathbf{H}}(\tau,\nu)$. A system $\mathbf{H}$ is called *underspread* [8], [11], [46] if $\sigma_{\mathbf{H}} \ll 1$, which means that $S_{\mathbf{H}}(\tau,\nu)$ is highly concentrated about the origin of the (τ,ν) plane and, hence, that $\mathbf{H}$ causes only small TF shifts. Since $S_{\mathbf{H}}(\tau,\nu)$ is the 2-D Fourier transform of $L_{\mathbf{H}}(t,f)$, this implies that $L_{\mathbf{H}}(t,f)$ is a smooth function. Two systems $\mathbf{H}_1$ and $\mathbf{H}_2$ are called *jointly underspread* if their spreading functions are effectively supported within the *same* small rectangle of area $\sigma_{\mathbf{H}_1,\mathbf{H}_2} \ll 1$.

Let us next consider a nonstationary, zero-mean random process. The expected ambiguity function $\bar{A}_x(\tau,\nu)$ in (18) describes the correlation of process components whose TF locations are separated by τ in time and by ν in frequency [8]–[12]. Since the expected ambiguity function is the spreading function of the correlation operator $\mathbf{R}_x$, i.e., $\bar{A}_x(\tau,\nu) = S_{\mathbf{R}_x}(\tau,\nu)$, it is natural to define the *correlation spread* σ_x of a process $x(t)$ as the displacement spread of its correlation operator, i.e., $\sigma_x \triangleq \sigma_{\mathbf{R}_x}$. A nonstationary random process $x(t)$ is then called *underspread* if $\sigma_x \ll 1$ [8]–[11]. Since in this case the expected ambiguity function is highly concentrated about the origin of the (τ,ν) plane, process components at different (i.e., not too close) TF locations will be effectively uncorrelated. This also implies that the WVS of $x(t)$ is a smooth, effectively non-negative function [8], [12]. Two processes $x(t)$ and $y(t)$ are called *jointly underspread* if their expected ambiguity functions are effectively supported within the *same* small rectangle of area $\sigma_{x,y} \ll 1$.

Two special cases of underspread processes are "quasistationary" processes with limited temporal correlation width and "nearly white" processes with limited time variation of their statistics. We emphasize that quasistationarity alone is neither necessary nor sufficient for the underspread property.

IV. TIME–FREQUENCY FORMULATION OF OPTIMAL FILTERS

In Section II, we saw that the time-varying Wiener filter and projection filter are described by equations involving linear operators and signal spaces. This description is rather abstract and leads to computationally intensive design procedures and implementations. However, for stationary or white processes, there exist simple frequency-domain or time-domain formulations that involve functions instead of operators and intervals instead of signal spaces and thus allow a substantially simplified design and implementation. We will now show that similar simplifications can be developed for *underspread* nonstationary processes [50].

A. Time–Frequency Formulation of the Time-Varying Wiener Filter

We assume that $s(t)$ and $n(t)$ are jointly underspread processes; specifically, their expected ambiguity functions are assumed to be exactly zero outside a rectangle $\mathcal{G}$ (centered about the origin of the (τ,ν) plane) of area $\sigma_{s,n} \ll 1$. The Wiener filter can then be split up as $\mathbf{H}_o = \mathbf{H}_o^{\mathcal{G}} + \mathbf{H}_o^{\bar{\mathcal{G}}}$, where the *underspread part* $\mathbf{H}_o^{\mathcal{G}}$ is defined by

$$S_{\mathbf{H}_o^{\mathcal{G}}}(\tau,\nu) = S_{\mathbf{H}_o}(\tau,\nu) I_{\mathcal{G}}(\tau,\nu) \tag{19}$$

where $I_{\mathcal{G}}(\tau,\nu)$ is the indicator function of $\mathcal{G}$, and the *overspread part* $\mathbf{H}_o^{\bar{\mathcal{G}}}$ is defined by $S_{\mathbf{H}_o^{\bar{\mathcal{G}}}}(\tau,\nu) = I_{\bar{\mathcal{G}}}(\tau,\nu) S_{\mathbf{H}_o}(\tau,\nu) = [1 - I_{\mathcal{G}}(\tau,\nu)] S_{\mathbf{H}_o}(\tau,\nu)$. That is, $S_{\mathbf{H}_o^{\mathcal{G}}}(\tau,\nu)$ and $S_{\mathbf{H}_o^{\bar{\mathcal{G}}}}(\tau,\nu)$ lie inside and outside $\mathcal{G}$, respectively. Since the spreading function is the 2-D Fourier transform of the Weyl symbol, (19) implies $L_{\mathbf{H}_o^{\mathcal{G}}}(t,f) = L_{\mathbf{H}_o}(t,f) ** \psi(t,f)$; here, $**$ denotes 2-D convolution and $\psi(t,f)$ is the 2-D Fourier transform of $I_{\mathcal{G}}(\tau,\nu)$ and, hence, a 2-D lowpass function. This means that $L_{\mathbf{H}_o^{\mathcal{G}}}(t,f)$ is a smoothed version of $L_{\mathbf{H}_o}(t,f)$.

We now recall that the Wiener filter $\mathbf{H}_o$ is characterized by the relation (4), i.e., $\mathbf{H}_o(\mathbf{R}_s + \mathbf{R}_n) = \mathbf{R}_s$. A first step toward a TF formulation of $\mathbf{H}_o$ is to notice that removing the overspread part $\mathbf{H}_o^{\bar{\mathcal{G}}}$ from $\mathbf{H}_o$ does not greatly affect the validity of this relation, i.e.,

$$\mathbf{H}_o(\mathbf{R}_s + \mathbf{R}_n) = \mathbf{R}_s \;\Rightarrow\; \mathbf{H}_o^{\mathcal{G}}(\mathbf{R}_s + \mathbf{R}_n) \approx \mathbf{R}_s.$$

Indeed, it is shown in Appendix A that the Hilbert-Schmidt (HS) norm [43] of the error $\mathbf{H}_o^{\mathcal{G}}(\mathbf{R}_s + \mathbf{R}_n) - \mathbf{R}_s$ incurred by this approximation is upper bounded as

$$\begin{aligned} &\left\|\mathbf{H}_o^{\mathcal{G}}(\mathbf{R}_s + \mathbf{R}_n) - \mathbf{R}_s\right\|_{\mathrm{HS}} \\ &\quad \le 2\left\|\mathbf{H}_o^{\bar{\mathcal{G}}}\right\|_{\mathrm{HS}} \|\mathbf{R}_s + \mathbf{R}_n\|_{\mathrm{HS}} \sqrt{\sigma_{s,n}} \end{aligned} \tag{20}$$

and, furthermore, that the "excess MSE" resulting from using $\mathbf{H}_o^{\mathcal{G}}$ instead of $\mathbf{H}_o$ is bounded as

$$0 \le \bar{E}_e^{\mathcal{G}} - \bar{E}_{e_o} \le 2\left\|\mathbf{H}_o^{\mathcal{G}}\right\|_{\mathrm{HS}} \left\|\mathbf{H}_o^{\bar{\mathcal{G}}}\right\|_{\mathrm{HS}} \|\mathbf{R}_s + \mathbf{R}_n\|_{\mathrm{HS}} \sqrt{\sigma_{s,n}}. \tag{21}$$

Here, $\bar{E}_e^{\mathcal{G}}$ denotes the MSE obtained with $\mathbf{H}_o^{\mathcal{G}}$. Hence, if $\sigma_{s,n}$ is small, i.e., if $s(t)$ and $n(t)$ are jointly underspread, $\mathbf{H}_o^{\mathcal{G}}(\mathbf{R}_s + \mathbf{R}_n) \approx \mathbf{R}_s$ and $\bar{E}_e^{\mathcal{G}} \approx \bar{E}_{e_o}$. This shows that *the underspread part* $\mathbf{H}_o^{\mathcal{G}}$ *is nearly optimal*.

This being the case, it suffices to develop a TF formulation for $\mathbf{H}_o^{\mathcal{G}}$. Intuitively, we could conjecture that a natural TF formulation of the operator relation $\mathbf{H}_o^{\mathcal{G}}(\mathbf{R}_s + \mathbf{R}_n) \approx \mathbf{R}_s$ would be $L_{\mathbf{H}_o^{\mathcal{G}}}(t,f)[\bar{W}_s(t,f) + \bar{W}_n(t,f)] \approx \bar{W}_s(t,f)$. Indeed, it is shown in Appendix B that the L_2 norm of the error

$L_{\mathbf{H}_o^{\mathcal{G}}}(t,f)[\bar{W}_s(t,f)+\bar{W}_n(t,f)]-\bar{W}_s(t,f)$ incurred by this approximation is upper bounded as

$$\begin{aligned}&\left\|L_{\mathbf{H}_o^{\mathcal{G}}}[\bar{W}_s+\bar{W}_n]-\bar{W}_s\right\| \\ &\le \|\mathbf{R}_s+\mathbf{R}_n\|_{\mathrm{HS}}\left[\pi\left\|\mathbf{H}_o^{\mathcal{G}}\right\|_{\mathrm{HS}}\sqrt{\sigma_{s,n}^3}+2\left\|\mathbf{H}_o^{\bar{\mathcal{G}}}\right\|_{\mathrm{HS}}\sqrt{\sigma_{s,n}}\right] \\ &\le \|\mathbf{H}_o\|_{\mathrm{HS}}\|\mathbf{R}_s+\mathbf{R}_n\|_{\mathrm{HS}}\left[\pi\sqrt{\sigma_{s,n}^3}+2\sqrt{\sigma_{s,n}}\right]. \end{aligned} \tag{22}$$

An upper bound involving $\sigma_{s,n}$ can furthermore be formulated for the error magnitude (L_∞ norm) as well. Hence, $L_{\mathbf{H}_o^{\mathcal{G}}}(t,f)[\bar{W}_s(t,f)+\bar{W}_n(t,f)]\approx\bar{W}_s(t,f)$ if $\sigma_{s,n}$ is small, i.e., if $s(t)$ and $n(t)$ are jointly underspread. Defining the TF region where $\bar{W}_r(t,f)=\bar{W}_s(t,f)+\bar{W}_n(t,f)$ is effectively positive by $\mathcal{R}_r=\{(t,f):\bar{W}_r(t,f)>\epsilon\}$ (with some[5] small $\epsilon>0$), we finally obtain the following TF formulation for $\mathbf{H}_o^{\mathcal{G}}$:

$$L_{\mathbf{H}_o^{\mathcal{G}}}(t,f)\approx\begin{cases}\dfrac{\bar{W}_s(t,f)}{\bar{W}_s(t,f)+\bar{W}_n(t,f)}, & (t,f)\in\mathcal{R}_r\\ 0, & (t,f)\notin\mathcal{R}_r.\end{cases} \tag{23}$$

The minimum MSE in (6) is given by $\bar{E}_{e_o}=\mathrm{tr}\{\mathbf{H}_o\mathbf{R}_n\}=\langle L_{\mathbf{H}_o},\bar{W}_n\rangle$ which, for $n(t)$ underspread, can be shown to equal $\langle L_{\mathbf{H}_o^{\mathcal{G}}},\bar{W}_n\rangle$. Inserting (23), we obtain the approximate TF formulation

$$\bar{E}_{e_o}\approx\iint_{\mathcal{R}_r}\frac{\bar{W}_s(t,f)\bar{W}_n(t,f)}{\bar{W}_s(t,f)+\bar{W}_n(t,f)}\,dt\,df. \tag{24}$$

The relations (23) and (24) constitute a TF formulation of the time-varying Wiener filter that extends the frequency-domain formulation (10), (11) that is valid for stationary processes and the analogous time-domain formulation that is valid for nonstationary white processes to the much broader class of underspread nonstationary processes. This TF formulation of the time-varying Wiener filter allows a simple and intuitively appealing interpretation. Let us define the TF support regions where the WVS of $s(t)$ and $n(t)$ are effectively positive by $\mathcal{R}_s=\{(t,f):\bar{W}_s(t,f)>\|\bar{A}_s\|_1\sigma_{s,n}\}$ and $\mathcal{R}_n=\{(t,f):\bar{W}_n(t,f)>\|\bar{A}_n\|_1\sigma_{s,n}\}$. The TF regions $\mathcal{R}_s$, $\mathcal{R}_n$ and $\mathcal{R}_r$ correspond to the signal spaces $\mathcal{S}_s$, $\mathcal{S}_n$ and $\mathcal{S}_r$ underlying the respective processes [42], [51]. With these definitions at hand, the Wiener filter can be interpreted in the TF domain as follows (see Fig. 2).

- In the "signal only" TF region $\mathcal{R}_s\backslash\mathcal{R}_n$, i.e., in the TF region where only signal energy is present, it follows from (23) that the WS of the Wiener filter is approximately one:

$$L_{\mathbf{H}_o^{\mathcal{G}}}(t,f)\approx 1,\quad (t,f)\in\mathcal{R}_s\backslash\mathcal{R}_n.$$

Thus, $\mathbf{H}_o^{\mathcal{G}}$ passes all "noise-free" observation components without attenuation or distortion.

[5] An analysis of the WS of positive operators ([8] and [47]) suggests that we use $\epsilon\approx\|\bar{A}_r\|_1\sigma_r=\|\bar{A}_r\|_1\sigma_{s,n}$.

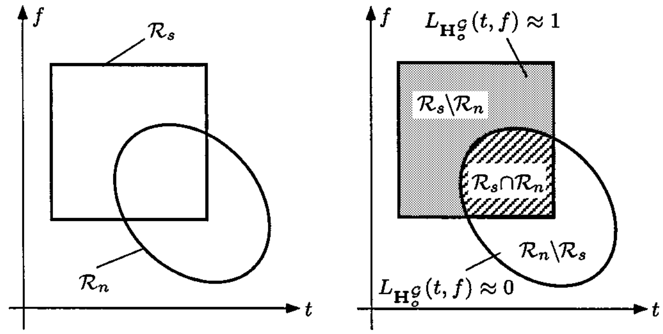


Fig. 2. TF interpretation of the time-varying Wiener filter.

- In the "no signal" TF region, where no signal energy is present, i.e., for $(t,f)\notin\mathcal{R}_s$, it follows from (23) that the WS of the Wiener filter is approximately zero:

$$L_{\mathbf{H}_o^{\mathcal{G}}}(t,f)\approx 0,\quad (t,f)\notin\mathcal{R}_s.$$

This "stop region" of the Wiener filter consists of two parts: i) the "noise only" TF region $\mathcal{R}_n\backslash\mathcal{R}_s$, where only noise energy is present and ii) the outside (complement) of the observation TF region $\mathcal{R}_r$, i.e., the TF region where neither signal nor noise energy is present; here, $L_{\mathbf{H}_o^{\mathcal{G}}}(t,f)\approx 0$ since (23) corresponds to the *minimal* Wiener filter.

- In the "signal plus noise" TF region $\mathcal{R}_s\cap\mathcal{R}_n$, where both signal energy and noise energy are present, the WS of the Wiener filter assumes values approximately between 0 and 1

$$0\lesssim L_{\mathbf{H}_o^{\mathcal{G}}}(t,f)\lesssim 1,\quad (t,f)\in\mathcal{R}_s\cap\mathcal{R}_n.$$

Here, $\mathbf{H}_o^{\mathcal{G}}$ performs a TF weighting that depends on the relative signal and noise energy at the respective TF point. Indeed, it follows from (23) that the ratio of $L_{\mathbf{H}_o^{\mathcal{G}}}(t,f)$ and $1-L_{\mathbf{H}_o^{\mathcal{G}}}(t,f)$ is given by the "local signal-to-noise ratio" $\mathrm{SNR}(t,f)\triangleq\bar{W}_s(t,f)/\bar{W}_n(t,f)$:

$$\frac{L_{\mathbf{H}_o^{\mathcal{G}}}(t,f)}{1-L_{\mathbf{H}_o^{\mathcal{G}}}(t,f)}\approx\frac{\bar{W}_s(t,f)}{\bar{W}_n(t,f)}=\mathrm{SNR}(t,f).$$

In particular, for equal signal and noise energy, i.e., $\mathrm{SNR}(t,f)=1$, there is $L_{\mathbf{H}_o^{\mathcal{G}}}(t,f)\approx 1/2$.

This TF interpretation of the time-varying Wiener filter extends the frequency-domain interpretation valid in the stationary case (see Fig. 1) to the broader class of underspread processes.

B. Time–Frequency Formulation of the Time-Varying Projection Filter

We recall from Section II-B that the optimal projection filter $\mathbf{P}_o$ performs a projection onto the space $\mathcal{S}_+=\mathrm{span}\{v_k(t):\mu_k>0\}$ spanned by all eigenfunctions of the operator $\mathbf{D}=\mathbf{R}_s-\mathbf{R}_n$ associated with positive eigenvalues. In order to obtain a TF formulation of this result, we consider the following three

TF regions on which the signal energy is, respectively, larger than, equal to, or smaller than the noise energy:

$$\begin{aligned}\mathcal{R}_+ &= \{(t,f) \in \mathcal{R}_r : \bar{W}_s(t,f) > \bar{W}_n(t,f)\}\\ \mathcal{R}_= &= \{(t,f) \in \mathcal{R}_r : \bar{W}_s(t,f) = \bar{W}_n(t,f)\}\\ \mathcal{R}_- &= \{(t,f) \in \mathcal{R}_r : \bar{W}_s(t,f) < \bar{W}_n(t,f)\}.\end{aligned}$$

With $L_{\mathbf{D}}(t,f) = L_{\mathbf{R}_s}(t,f) - L_{\mathbf{R}_n}(t,f) = \bar{W}_s(t,f) - \bar{W}_n(t,f)$, these TF regions can alternatively be written as

$$\begin{aligned}\mathcal{R}_+ &= \{(t,f) \in \mathcal{R}_r : L_{\mathbf{D}}(t,f) > 0\}\\ \mathcal{R}_= &= \{(t,f) \in \mathcal{R}_r : L_{\mathbf{D}}(t,f) = 0\}\\ \mathcal{R}_- &= \{(t,f) \in \mathcal{R}_r : L_{\mathbf{D}}(t,f) < 0\}.\end{aligned}$$

Note that $\mathcal{R}_+ \cup \mathcal{R}_= \cup \mathcal{R}_- = \mathcal{R}_r$.

There exists a correspondence between the "pass space" $\mathcal{S}_+ = \operatorname{span}\{v_k(t) : \mu_k > 0\}$ of $\mathbf{P}_o$ and the TF region $\mathcal{R}_+ = \{(t,f) \in \mathcal{R}_r : L_{\mathbf{D}}(t,f) > 0\}$. Indeed, if $s(t)$ and $n(t)$ are jointly underspread, then $\mathbf{D} = \mathbf{R}_s - \mathbf{R}_n$ is an underspread system. It is here known [8], [11], [47] that TF shifted versions $g_{t',f'}(t) = g(t-t')e^{j2\pi f' t}$ of a well-TF localized function $g(t)$ are approximate eigenfunctions of $\mathbf{D}$, and the WS values $L_{\mathbf{D}}(t,f)$ are the corresponding approximate eigenvalues, i.e.,

$$(\mathbf{D}g_{t',f'})(t) \approx L_{\mathbf{D}}(t',f')g_{t',f'}(t).$$

Hence, the optimal pass space $\mathcal{S}_+$ (spanned by all eigenfunctions $v_k(t)$ with $\mu_k > 0$) corresponds to the TF region $\mathcal{R}_+ = \{(t,f) \in \mathcal{R}_r : L_{\mathbf{D}}(t,f) > 0\}$ (comprising the TF locations (t',f') of all $g_{t',f'}(t)$ such that $L_{\mathbf{D}}(t',f') > 0$). The action of $\mathbf{P}_o$—passing signals in $\mathcal{S}_+$ and suppressing signals in $\mathcal{S}_+^\perp$—can therefore be reformulated in the TF plane as passing signals in the TF region $\mathcal{R}_+$ and suppressing signals outside $\mathcal{R}_+$. Thus, we conclude that $\mathbf{P}_o$ passes signal components in the TF region where the local TF SNR is larger than one and suppresses signal components in the complementary TF region, which is intuitively reasonable.

The optimal projection filter $\mathbf{P}_o$ performs a projection onto the space $\mathcal{S}_+$; for jointly underspread signal and noise, this space is a "simple" space as defined in [42]. The WS of the projection operator on a simple space essentially assumes the values 1 (on the TF pass region, in our case $\mathcal{R}_+$) and 0 (on the TF stop region) [42]. Hence, the WS of $\mathbf{P}_o$ can be approximated as[6]

$$L_{\mathbf{P}_o}(t,f) \approx I_{\mathcal{R}_+}(t,f) \tag{25}$$

where $I_{\mathcal{R}_+}(t,f)$ is the indicator function of $\mathcal{R}_+$. This extends the relation (13) to general underspread processes. Note that the projection property of $\mathbf{P}_o$ only allows $\mathbf{P}_o$ to *pass* or *suppress* components of $r(t)$; no intermediate type of attenuation (which would correspond to values of $L_{\mathbf{P}_o}(t,f)$ between 0 and 1) is possible.

[6]Experiments show that this approximation is valid apart from oscillations of $L_{\mathbf{P}_o}(t,f)$ about the values 0 or 1. Hence, the approximation can be improved by a slight smoothing of $L_{\mathbf{P}_o}(t,f)$, which corresponds to the removal of overspread components from $\mathbf{P}_o$ (cf. Section IV-A).

It can finally be shown that the MSE obtained with $\mathbf{P}_o$ [cf. (8)] is approximately given by

$$\bar{E}_{e_p} \approx \iint_{\mathcal{R}_+} \bar{W}_n(t,f)\,dt\,df + \iint_{\mathcal{R}_- \cup \mathcal{R}_=} \bar{W}_s(t,f)\,dt\,df$$

which extends the relation (14) to general underspread processes.

V. Approximate Time–Frequency Design of Optimal Filters

In the previous section, we have shown that the time-varying Wiener and projection filters can be reformulated in the TF domain if the nonstationary random processes $s(t)$ and $n(t)$ are jointly underspread. We will now show that this TF formulation of optimal filters leads to simple *design procedures* that operate in the TF domain and yield nearly optimal filters.

A. Time–Frequency Pseudo-Wiener Filter

We recall that for jointly underspread processes, an approximate expression for the Wiener filter's WS is given by (23). Let us now define another linear, time-varying filter $\mathbf{H}_1$ by setting its WS equal to the right-hand side of (23) [50]:

$$L_{\mathbf{H}_1}(t,f) \triangleq \begin{cases} \dfrac{\bar{W}_s(t,f)}{\bar{W}_s(t,f) + \bar{W}_n(t,f)}, & (t,f) \in \mathcal{R}_r \\ 0, & (t,f) \notin \mathcal{R}_r. \end{cases} \tag{26}$$

We refer to $\mathbf{H}_1$ as the *TF pseudo-Wiener filter*. For jointly underspread processes $s(t)$, $n(t)$, where (23) is a good approximation, a combination of (26) and (23) yields $L_{\mathbf{H}_1}(t,f) \approx L_{\mathbf{H}_o^{\mathcal{G}}}(t,f)$, and thus, $\mathbf{H}_1 \approx \mathbf{H}_o^{\mathcal{G}}$. Hence, the TF pseudo-Wiener filter $\mathbf{H}_1$ is a close approximation to the underspread part $\mathbf{H}_o^{\mathcal{G}}$ of the Wiener filter $\mathbf{H}_o$ and, therefore, is nearly optimal; furthermore, the TF interpretation of $\mathbf{H}_o^{\mathcal{G}}$ (see Section IV-A) applies equally well to $\mathbf{H}_1$. The deviation from optimality is characterized by the error bounds in (20)–(22). For processes that are not jointly underspread, however, $\mathbf{H}_1$ must be expected to perform poorly. Note that $\mathbf{H}_1$ is a self-adjoint operator since according to (26), the WS of $\mathbf{H}_1$ is real-valued.

For jointly *strictly* underspread processes, i.e., underspread processes whose support rectangle $\mathcal{G}$ is oriented parallel to the τ and ν axes [8], [29], the WVS occurring on the right-hand side of (26) can be replaced by the generalized WVS [17]–[19], the (generalized) evolutionary spectrum [23], [24], [28], [29], or the physical spectrum (expected spectrogram) using an appropriate analysis window [18], [19], [33]. This is possible since in the *strictly* underspread case, these spectra become nearly equivalent [8], [9], [12], [19], [29].

While the TF pseudo-Wiener filter $\mathbf{H}_1$ is *defined* in the TF domain, the actual calculation of the signal estimate can be performed directly in the time domain according to

$$\hat{s}_1(t) = (\mathbf{H}_1 r)(t) = \int_{t'} h_1(t,t')r(t')\,dt'$$

where $h_1(t,t')$, which is the impulse response of $\mathbf{H}_1$, can be obtained from the WS in (26) as [cf. (15)]

$$h_1(t,t') = \int_f L_{\mathbf{H}_1}\left(\frac{t+t'}{2}, f\right) e^{j2\pi f(t-t')}\, df. \tag{27}$$

An efficient approximate multiwindow STFT implementation will be discussed in Section VI.

Compared with the Wiener filter $\mathbf{H}_o$, the TF pseudo-Wiener filter $\mathbf{H}_1$ possesses two practical advantages:

- *Modified a priori knowledge:* Ideally, the calculation (design) of $\mathbf{H}_o$ requires knowledge of the correlation operators $\mathbf{R}_s$ and $\mathbf{R}_n$ [see (5)], whereas the design of $\mathbf{H}_1$ requires knowledge of the WVS $\bar{W}_s(t,f)$ and $\bar{W}_n(t,f)$ [see (26)]. Although correlation operators and WVS are mathematically equivalent due to the one-to-one mapping (16), the WVS are much easier and more intuitive to handle than the correlation operators or correlation functions since WVS have a more immediate physical significance and interpretation. In practice, the quantities constituting the *a priori* knowledge (correlation operators or WVS) are usually replaced by estimated or schematic/idealized versions, which are much more easily controlled or designed in the TF domain than in the correlation domain. For example, an approximate or partial knowledge of the WVS will often suffice for a reasonable filter design.
- *Reduced computation:* The calculation (design) of $\mathbf{H}_o$ requires a computationally intensive and potentially unstable operator inversion (or, in a discrete-time setting, a matrix inversion). In the TF design (26), this operator inversion is replaced by simple and easily controllable pointwise divisions of functions. Assuming discrete-time signals of length N, the computational cost of the design of $\mathbf{H}_o$ grows with N^3, whereas that of $\mathbf{H}_1$ (using divisions and FFT's) grows only with $N^2 \log_2 N$.

The TF pseudo-Wiener filter satisfies an approximate "TF optimality." For $s(t), n(t)$ jointly underspread, it can be shown that the MSE obtained with *any* underspread system $\mathbf{H}$ is approximately given by

$$\begin{aligned}\bar{E}_e \approx \varepsilon_{\mathbf{H}} \triangleq \iint_{\mathcal{R}_r} [&|L_{\mathbf{H}}(t,f)|^2 \bar{W}_n(t,f) \\ &+ |1 - L_{\mathbf{H}}(t,f)|^2 \bar{W}_s(t,f)]\, dt\, df.\end{aligned} \tag{28}$$

Minimizing $\varepsilon_{\mathbf{H}}$ (assuming a minimal solution, i.e., $L_{\mathbf{H}}(t,f) = 0$ for $(t,f) \notin \mathcal{R}_r$) then yields the $L_{\mathbf{H}}(t,f)$ in (26) and, thus, the TF pseudo-Wiener filter $\mathbf{H}_1$. This interpretation suggests an *extended TF design* that is based on the following weighted MSE with smooth TF weighting function $\alpha(t,f)$ [extending (12)]

$$\begin{aligned}\varepsilon_{\mathbf{H}}^{(\alpha)} = \iint_{\mathcal{R}_r} &[\alpha(t,f) |L_{\mathbf{H}}(t,f)|^2 \bar{W}_n(t,f) \\ &+ [1-\alpha(t,f)] |1 - L_{\mathbf{H}}(t,f)|^2 \bar{W}_s(t,f)]\, dt\, df\end{aligned} \tag{29}$$

with $0 \leq \alpha(t,f) \leq 1$. The filter minimizing this weighted TF error is given by (26) with $\bar{W}_n(t,f)$ replaced by $\alpha(t,f)\bar{W}_n(t,f)$ and $\bar{W}_s(t,f)$ replaced by $[1-\alpha(t,f)]\bar{W}_s(t,f)$.

B. Time–Frequency Pseudo-Projection Filter

Next, we consider a TF filter design that is motivated by the approximate expression (25) for the WS of $\mathbf{P}_o$. Let us define a new linear system $\tilde{\mathbf{P}}$ by setting its WS equal to the right-hand side of (25):

$$L_{\tilde{\mathbf{P}}}(t,f) \triangleq I_{\mathcal{R}_+}(t,f).$$

Since $I_{\mathcal{R}_+}(t,f)$ is the indicator function of the TF region $\mathcal{R}_+ = \{(t,f) \in \mathcal{R}_r : \bar{W}_s(t,f) > \bar{W}_n(t,f)\}$, the WS of $\tilde{\mathbf{P}}$ is 1 for those TF points where there is more signal energy than noise energy and 0 elsewhere. It is interesting that the WS of $\tilde{\mathbf{P}}$ is a rounded version of the WS of the TF pseudo-Wiener filter $\mathbf{H}_1$ in (26), i.e., $L_{\tilde{\mathbf{P}}}(t,f) = \text{round}\{L_{\mathbf{H}_1}(t,f)\}$, which is consistent with the stationary case. Furthermore, $L_{\tilde{\mathbf{P}}}(t,f)$ differs from $L_{\mathbf{H}_1}(t,f)$ only on the signal-plus-noise TF region $\mathcal{R}_s \cap \mathcal{R}_n$; in the TF pass region $\mathcal{R}_s \backslash \mathcal{R}_n$, we have $L_{\tilde{\mathbf{P}}}(t,f) \approx L_{\mathbf{H}_1}(t,f) \approx 1$, and in the TF stop region outside $\mathcal{R}_s$, there is $L_{\tilde{\mathbf{P}}}(t,f) \approx L_{\mathbf{H}_1}(t,f) \approx 0$.

In general, $\tilde{\mathbf{P}}$ is *not* exactly an orthogonal projection operator. However, for jointly underspread processes $s(t)$ and $n(t)$, where (25) is a good approximation, there is $L_{\tilde{\mathbf{P}}}(t,f) \approx L_{\mathbf{P}_o}(t,f)$, and thus, $\tilde{\mathbf{P}} \approx \mathbf{P}_o$. Hence, the TF-designed filter $\tilde{\mathbf{P}}$ is a reasonable approximation to the optimal projection filter $\mathbf{P}_o$, and we will thus call it *TF pseudo-projection filter*. However, for processes that are not jointly underspread, $\tilde{\mathbf{P}}$ must be expected to perform poorly and to be very different from an orthogonal projection operator.

Although $\tilde{\mathbf{P}}$ is defined in the TF domain, the actual calculation of the signal estimate can be done in the time domain using the impulse response of $\tilde{\mathbf{P}}$ that is derived from $I_+(t,f)$ as (cf. (27))

$$\tilde{p}(t,t') = \int_f I_{\mathcal{R}_+}\left(\frac{t+t'}{2}, f\right) e^{j2\pi f(t-t')}\, df. \tag{30}$$

An efficient approximate multiwindow STFT implementation will be discussed in Section VI.

Compared with the optimal projection filter $\mathbf{P}_o$, the TF pseudo-projection filter $\tilde{\mathbf{P}}$ has two advantages:

- *Modified/reduced* a priori *knowledge:* The design of $\mathbf{P}_o$ requires knowledge of the space that is spanned by the eigenfunctions of the operator $\mathbf{D} = \mathbf{R}_s - \mathbf{R}_n$ corresponding to positive eigenvalues (see Section II-B), whereas the design of $\tilde{\mathbf{P}}$ merely presupposes that we know the TF region $\mathcal{R}_+$ in which the signal dominates the noise. This knowledge is of a much simpler and more intuitive form, thus facilitating the use of approximate or partial information about the WVS.
- *Reduced computation:* The design of $\mathbf{P}_o$ requires the solution of an eigenvalue problem, which is computationally intensive. In contrast, the proposed TF design only requires an inverse WS transformation [see (30)]. Assuming discrete-time signals of length N, the computational cost of the design of $\mathbf{P}_o$ grows with N^3, whereas that of $\tilde{\mathbf{P}}$ (using FFT's) grows with $N^2 \log_2 N$.

The TF pseudo-projection filter $\tilde{\mathbf{P}}$ is furthermore advantageous as compared with the Wiener-type filters $\mathbf{H}_o$ or $\mathbf{H}_1$ since its design is less expensive and more robust (especially with respect to errors in estimating the *a priori* knowledge and with respect to potential correlations of signal and noise).

Similar to the TF pseudo-Wiener filter, the TF pseudo-projection filter satisfies a "TF optimality" in that it minimizes the TF MSE $\varepsilon_{\mathbf{H}}$ in (28) under the constraint of a 0/1-valued WS, i.e., $L_{\mathbf{H}}(t,f) \in \{0,1\}$. Minimizing the TF weighted MSE $\varepsilon_{\mathbf{H}}^{(\alpha)}$ in (29) rather than $\varepsilon_{\mathbf{H}}$, again under the constraint $L_{\mathbf{H}}(t,f) \in \{0,1\}$, results in a generalized TF pseudo-projection filter $\tilde{\mathbf{P}}^{(\alpha)}$ that is given by $L_{\tilde{\mathbf{P}}^{(\alpha)}}(t,f) = I_{\mathcal{R}_+^{(\alpha)}}(t,f)$, where $\mathcal{R}_+^{(\alpha)} = \{(t,f) \in \mathcal{R}_r : [1-\alpha(t,f)]\bar{W}_s(t,f) > \alpha(t,f)\bar{W}_n(t,f)\}$.

C. Time–Frequency Projection Filter

For jointly underspread processes, the TF pseudo-projection filter $\tilde{\mathbf{P}}$ approximates the optimal projection filter $\mathbf{P}_o$, but it is not exactly an orthogonal projection operator. If an orthogonal projection operator is desired, we may calculate the orthogonal projection operator $\mathbf{P}$ that optimally approximates $\tilde{\mathbf{P}}$ in the sense of minimizing $\|\mathbf{P}-\tilde{\mathbf{P}}\|_{\mathrm{HS}}$. This can be shown to be equivalent to minimizing $\|L_{\mathbf{P}} - I_+\|^2 = \int_t \int_f [L_{\mathbf{P}}(t,f) - I_+(t,f)]^2 \, dt\, df$ [42], [51]. That is, the TF designed projection filter (hereafter denoted by $\mathbf{P}_1$) is the orthogonal projection operator whose WS is closest to the indicator function $I_+(t,f)$. It is shown in [42] and [51] that $\mathbf{P}_1$ is given by the orthogonal projection operator on the space $\mathcal{S}_1 = \mathrm{span}\{u_k(t) : \lambda_k > 1/2\}$ spanned by all eigenfunctions $u_k(t)$ of $\tilde{\mathbf{P}}$ with associated eigenvalues $\lambda_k > 1/2$, i.e.,

$$\mathbf{P}_1 = \sum_{k \in I_{1/2}} \mathbf{P}_{u_k} \quad \text{with } I_{1/2} = \{k : \lambda_k > 1/2\}.$$

For jointly underspread processes where $\tilde{\mathbf{P}}$ is close to an orthogonal projection operator, $\mathbf{P}_1$ is close to $\tilde{\mathbf{P}}$ and, hence, also close to the optimal projection filter $\mathbf{P}_o$. However, due to the inherent projection constraint, $\mathbf{P}_1$ may lead to a larger MSE than $\tilde{\mathbf{P}}$. In addition, the derivation of $\mathbf{P}_1$ from $\tilde{\mathbf{P}}$ requires the solution of an eigenproblem. Therefore, the TF projection filter $\mathbf{P}_1$ appears to be less attractive than the TF pseudo-projection filter $\tilde{\mathbf{P}}$.

VI. MULTIWINDOW STFT IMPLEMENTATION

Following [8] and [20], we now discuss a TF implementation of the Wiener and projection filters that is based on the multiwindow short-time Fourier transform (STFT) and that is valid in the underspread case.

A simple TF filter method is an *STFT filter* that consists of the following steps [8], [13], [20], [52]–[57]:

- *STFT analysis*: Calculation of the STFT [19], [54], [55], [58] of the input signal $x(t)$

$$\mathrm{STFT}_x^{(g)}(t,f) = \int_{t'} x(t')\, g_{t,f}^*(t')\, dt'.$$

Here, $g_{t,f}(t') = g(t'-t)e^{j2\pi f t'}$ where $g(t)$ is a normalized window.

- *STFT weighting*: Multiplication of the STFT by a weighting function $M(t,f)$

$$\mathrm{STFT}_x^{(g)}(t,f) \to M(t,f)\mathrm{STFT}_x^{(g)}(t,f).$$

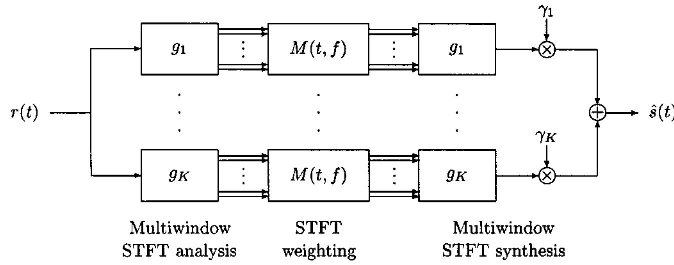


Fig. 3. Multiwindow STFT implementation of the Wiener filter.

- *STFT synthesis*: The output signal $y(t)$ is the inverse STFT [54], [55], [58] of the weighted STFT:

$$y(t) = \int_{t'} \int_{f'} \left[M(t',f')\mathrm{STFT}_x^{(g)}(t',f') \right] g_{t',f'}(t)\, dt'\, df'.$$

We note that $y(t)$ is the signal whose STFT is closest to $M(t,f)\mathrm{STFT}_x^{(g)}(t,f)$ in a least-squares sense [57].

These steps implement a linear, time-varying filter (hereafter denoted $\mathbf{Q}$) that depends on $M(t,f)$ and $g(t)$.

We recall from Section IV-A that for jointly underspread processes, the underspread part $\mathbf{H}_o^{\mathcal{G}}$ of the Wiener filter $\mathbf{H}_o$ is nearly optimal. Since $S_{\mathbf{H}_o^{\mathcal{G}}}(\tau,\nu) = 0$ for $(\tau,\nu) \notin \mathcal{G}$, it follows that

$$S_{\mathbf{H}_o^{\mathcal{G}}}(\tau,\nu) = S_{\mathbf{H}_o^{\mathcal{G}}}(\tau,\nu)\Psi(\tau,\nu)$$

where $\Psi(\tau,\nu)$ is any real-valued function that is 1 on $\mathcal{G}$ [a "minimal" choice for $\Psi(\tau,\nu)$ is the indicator function $I_{\mathcal{G}}(\tau,\nu)$, cf. (19)]. Let $\mathbf{T}$ denote the linear system defined by $S_{\mathbf{T}}(\tau,\nu) \triangleq \Psi(\tau,\nu)$. If $\Psi(\tau,\nu)$ is chosen such that $\Psi(-\tau,-\nu) = \Psi(\tau,\nu)$, $\mathbf{T}$ is a self-adjoint operator with real-valued eigenvalues γ_k and orthonormal eigenfunctions $g_k(t)$. It is shown in [8] and [20] that $\mathbf{H}_o^{\mathcal{G}}$ can be expanded as

$$\mathbf{H}_o^{\mathcal{G}} = \sum_{k=1}^{\infty} \gamma_k \mathbf{Q}_k \tag{31}$$

where the $\mathbf{Q}_k$ are STFT filters with TF weighting function $M(t,f) = L_{\mathbf{H}_o^{\mathcal{G}}}(t,f)$ and STFT windows $g_k(t)$. Thus, $\mathbf{H}_o^{\mathcal{G}}$ is represented as a weighted sum of STFT filters with identical TF weighting function $M(t,f)$ and orthonormal windows $g_k(t)$. In practice, the eigenvalues γ_k often decay quickly so that it suffices to use only a few terms of (31) corresponding to the largest eigenvalues (an example will be presented in Section VII). Assuming that the eigenvalues γ_k are arranged in nonincreasing order, we then obtain

$$\mathbf{H}_o^{\mathcal{G}} \approx \mathbf{Q}^{(K)} \triangleq \sum_{k=1}^{K} \gamma_k \mathbf{Q}_k.$$

This yields the approximate *multiwindow STFT implementation* of $\mathbf{H}_o^{\mathcal{G}}$ that is depicted in Fig. 3. It can be shown [8] that the approximation error $\mathbf{H}_o^{\mathcal{G}} - \mathbf{Q}^{(K)}$ is bounded as

$$\left\| \mathbf{H}_o^{\mathcal{G}} - \mathbf{Q}^{(K)} \right\|_{\mathrm{HS}} \le \left\| \mathbf{H}_o^{\mathcal{G}} \right\|_{\mathrm{HS}} \sum_{k=K+1}^{\infty} |\gamma_k|$$

which can be used to estimate the appropriate order K. Since $s(t)$ and $n(t)$ are assumed jointly underspread, (23) holds, and we obtain the following simple approximation to the STFT weighting function $M(t,f) = L_{\mathbf{H}_o^{\mathcal{G}}}(t,f)$:

$$M(t,f) \approx \begin{cases} \dfrac{\bar{W}_s(t,f)}{\bar{W}_s(t,f)+\bar{W}_n(t,f)}, & (t,f) \in \mathcal{R}_r \\ 0, & (t,f) \notin \mathcal{R}_r. \end{cases} \tag{32}$$

A particularly efficient discrete-time/discrete-frequency version of the multiwindow STFT implementation that uses filter banks is discussed in [8]. Furthermore, it is shown in [8] and [20] that in the case of white noise $n(t)$, the STFT's used in the multiwindow STFT implementation can additionally be used to estimate the WVS of $s(t)$, which is required to calculate $M(t,f)$ according to (32).

A heuristic, approximate TF implementation of time-varying Wiener filters proposed in [31] and [32] uses a single STFT filter (i.e., $K = 1$) with TF weighting function

$$M(t,f) = \frac{P_s(t,f)}{P_s(t,f)+P_n(t,f)}$$

where $P_s(t,f)$ and $P_n(t,f)$ are the *physical spectra* [18], [19], [33] of $s(t)$ and $n(t)$, respectively. Compared with the multiwindow STFT implementation discussed above, this suffers from a threefold performance loss since

i) Only one STFT filter is used.
ii) The physical spectrum is a smoothed version of the WVS [18], [33].
iii) The STFT window is not chosen as the eigenfunction of $\mathbf{T}$ corresponding to the largest eigenvalue.

An approximate multiwindow STFT implementation can also be developed for the TF pseudo-projection filter $\tilde{\mathbf{P}}$ (which approximates the optimal projection filter $\mathbf{P}_o$). Here, the TF weighting function is

$$M(t,f) = L_{\tilde{\mathbf{P}}}(t,f) = I_+(t,f).$$

Since $\tilde{\mathbf{P}}$ is not exactly underspread, the function $\Psi(\tau,\nu)$ defining the windows $g_k(t)$ and the coefficients γ_k must be chosen such that it covers the *effective* support of $S_{\tilde{\mathbf{P}}}(\tau,\nu)$, and hence

$$S_{\tilde{\mathbf{P}}}(\tau,\nu) \approx S_{\tilde{\mathbf{P}}}(\tau,\nu)\Psi(\tau,\nu). \tag{33}$$

Subsequently, the γ_k and $g_k(t)$ can be constructed from $\Psi(\tau,\nu)$ as explained above. The resulting multiwindow STFT filter $\mathbf{Q}^{(K)}$ leads to an approximation error $\tilde{\mathbf{P}} - \mathbf{Q}^{(K)}$ that can be bounded as

$$\left\|\tilde{\mathbf{P}} - \mathbf{Q}^{(K)}\right\|_{\mathrm{HS}} \le \left[\int_\tau\int_\nu |1 - S_{\mathbf{T}_K}(\tau,\nu)|^2 |S_{\tilde{\mathbf{P}}}(\tau,\nu)|^2 \, d\tau \, d\nu\right]^{1/2}$$

with $\mathbf{T}_K = \sum_{k=1}^K \gamma_k \mathbf{P}_{g_k}$. This bound is a measure of the extension of $S_{\tilde{\mathbf{P}}}(\tau,\nu)$ and will thus be small in the underspread case. However, $\|\tilde{\mathbf{P}} - \mathbf{Q}^{(K)}\|_{\mathrm{HS}} \neq 0$ even for $K \to \infty$ since (33) is only an approximation.

VII. Numerical Simulations

This section presents numerical simulations that illustrate and corroborate our theoretical results. Using the TF synthesis technique introduced in [59], we synthesized random processes $s(t)$ and $n(t)$ with expected energies $\bar{E}_s = 9.09$ and $\bar{E}_n = 11.89$, respectively. Thus, the mean input SNR is $\mathrm{SNR_{in}} = \bar{E}_s/\bar{E}_n = -1.17$ dB. The WVS and expected ambiguity functions of $s(t)$ and $n(t)$ are shown in Fig. 4. From the expected ambiguity functions, it is seen that the processes are jointly underspread. The WS's of the Wiener filter $\mathbf{H}_o$, its underspread part $\mathbf{H}_o^{\mathcal{G}}$, and the TF pseudo-Wiener filter $\mathbf{H}_1$ are shown in Fig. 5(a)–(c). The TF pass, stop, and transition regions of the filters are easily recognized. It is verified that the WS of $\mathbf{H}_o^{\mathcal{G}}$ is a smoothed version of the WS of $\mathbf{H}_o$ and that the WS of $\mathbf{H}_1$ closely approximates that of $\mathbf{H}_o^{\mathcal{G}}$. Fig. 5(d)–(f) compare the WS's of the optimal projection filter $\mathbf{P}_o$, TF pseudo-projection filter $\tilde{\mathbf{P}}$, and TF projection filter $\mathbf{P}_1$. It is seen that the WS's of these filters are similar, except for small-scale oscillations, and that the WS of $\tilde{\mathbf{P}}$ is a rounded version of the WS of $\mathbf{H}_1$.

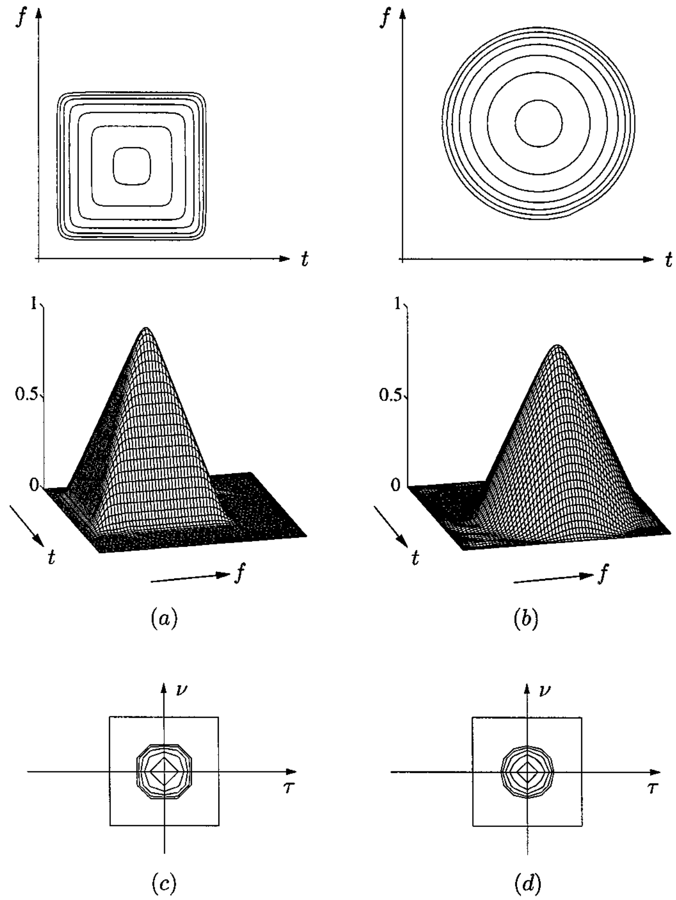


Fig. 4. Second-order statistics of signal and noise. (a) WVS of $s(t)$. (b) WVS of $n(t)$. (c) Expected ambiguity function of $s(t)$. (d) Expected ambiguity function of $n(t)$. The rectangles in parts (c) and (d) have area 1 and thus allow assessment of the underspread property of $s(t)$ and $n(t)$. The signal length is 128 samples.

The mean SNR improvements $\Delta\mathrm{SNR} \triangleq \mathrm{SNR_{out}} - \mathrm{SNR_{in}}$ (where $\mathrm{SNR_{out}} = \bar{E}_s/\bar{E}_e$ with $e(t) = s(t) - \hat{s}(t)$) achieved with the Wiener-type and projection-type filters are listed in Table I. The performance of the TF pseudo-Wiener filter $\mathbf{H}_1$ is seen to be very close to that of the Wiener filter $\mathbf{H}_o$. Similarly, the performance of the TF pseudo-projection filter $\tilde{\mathbf{P}}$ is close to that of the optimal projection filter $\mathbf{P}_o$. In fact, $\tilde{\mathbf{P}}$ performs even slightly better than both $\mathbf{P}_o$ and the TF projection filter $\mathbf{P}_1$,

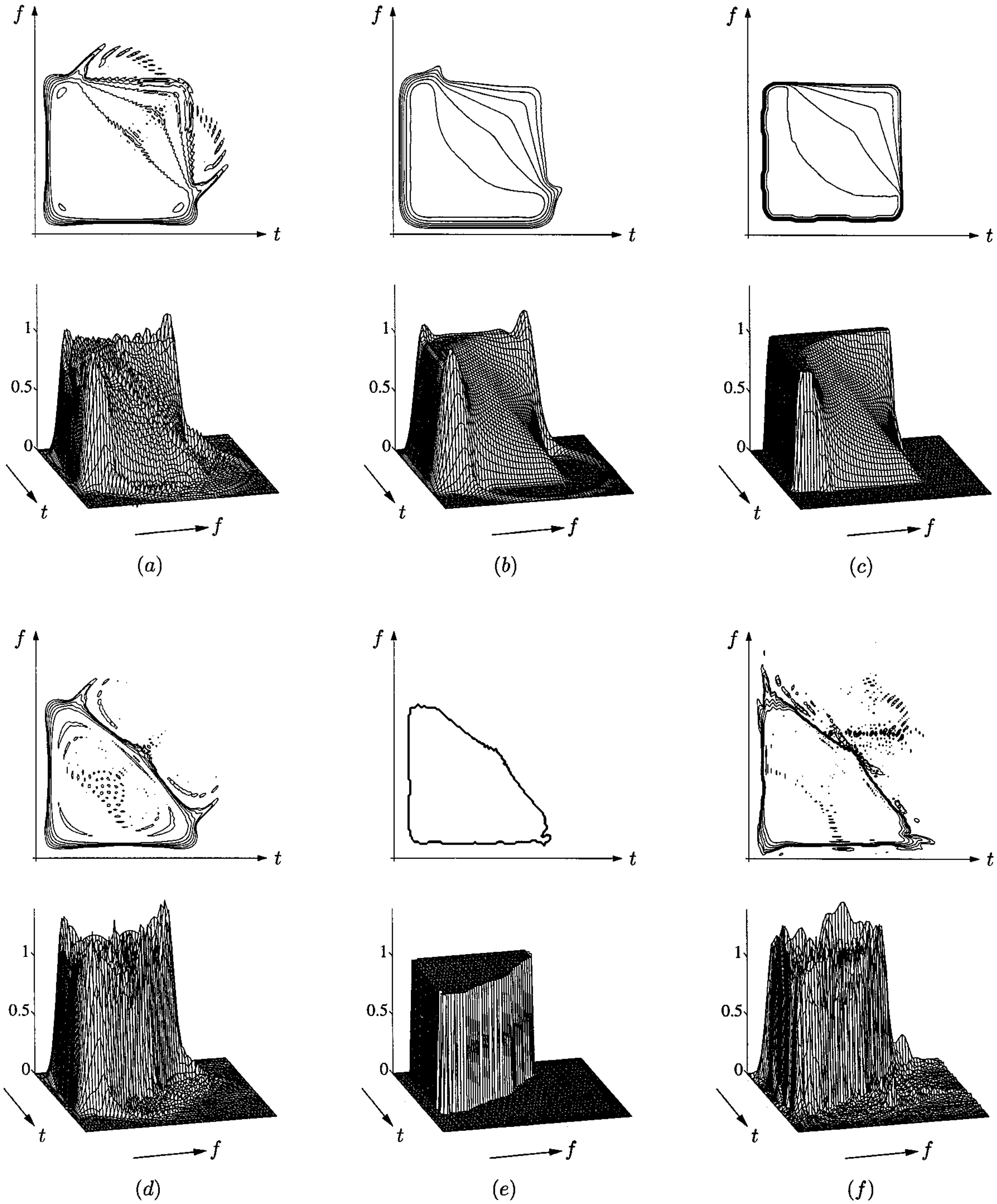


Fig. 5. WS's of (a) Wiener filter $\mathbf{H}_o$. (b) Underspread part $\mathbf{H}_o^{\mathcal{G}}$ of Wiener filter. (c) TF pseudo-Wiener filter $\mathbf{H}_1$. (d) Optimal projection filter $\mathbf{P}_o$. (e) TF pseudo-projection filter $\tilde{\mathbf{P}}$. (f) TF projection filter $\mathbf{P}_1$.

TABLE I
MEAN SNR IMPROVEMENT ACHIEVED BY WIENER-TYPE AND PROJECTION-TYPE FILTERS

Filter type	ΔSNR
Wiener filter $\mathbf{H}_o$	6.14 dB
Underspread part $\mathbf{H}_o^{\mathcal{G}}$	6.10 dB
TF pseudo-Wiener filter $\mathbf{H}_1$	6.11 dB
Optimal projection filter $\mathbf{P}_o$	4.37 dB
TF pseudo-projection filter $\tilde{\mathbf{P}}$	4.41 dB
TF projection filter $\mathbf{P}_1$	4.34 dB

which can be attributed to the orthogonal projection constraint underlying $\mathbf{P}_o$ and $\mathbf{P}_1$ (see Section V-C).

The multiwindow STFT implementation discussed in Section VI, with $M(t, f)$ approximated according to (32), is considered in Fig. 6. The signal and noise processes are as before. Fig. 6(a)–(c) show the WS's of the multiwindow STFT filters $\mathbf{Q}^{(K)}$ using filter order $K = 1, 4,$ and 10. For growing K, $\mathbf{Q}^{(K)}$ becomes increasingly similar to the Wiener filter or the TF pseudo-Wiener filter [cf. Fig. 5(a)–(c)]. Fig. 6(d) shows how the mean SNR improvement achieved with $\mathbf{Q}^{(K)}$ depends on K. Although the single-window STFT filter $(K = 1)$ is seen

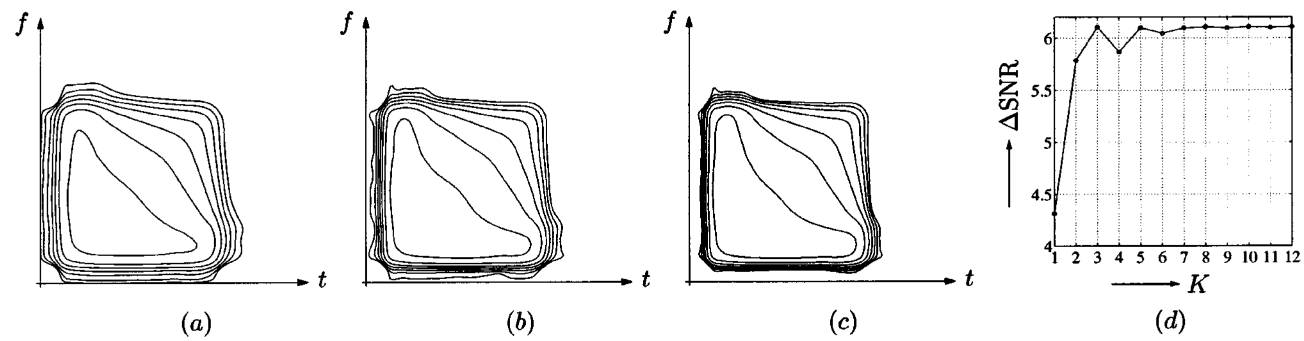


Fig. 6. Multiwindow STFT implementation of the Wiener filter. (a)–(c) WS's of multiwindow STFT filters with filter order $K = 1, 4,$ and 10. (d) Mean SNR improvement versus STFT filter order K.

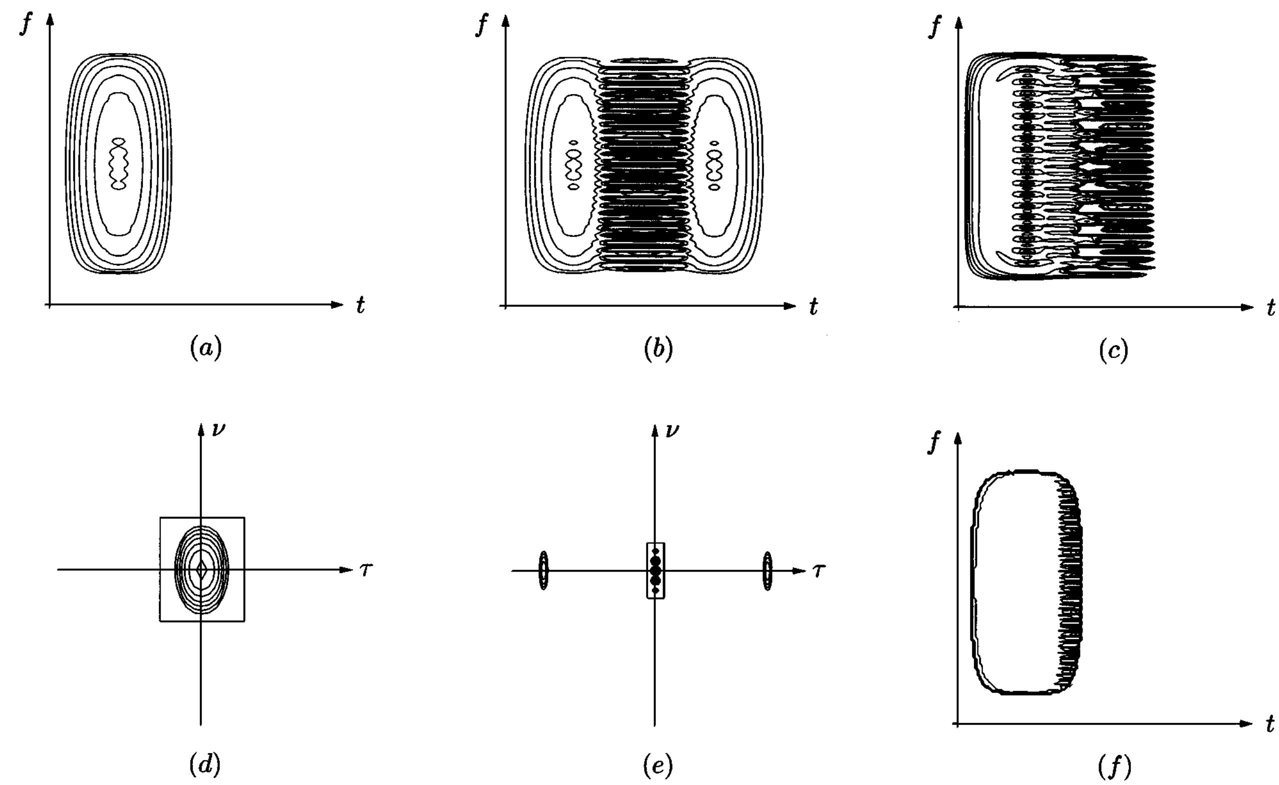


Fig. 7. Filtering experiment involving an overspread noise process. (a) WVS of $s(t)$. (b) WVS of $n(t)$. (c) WS of Wiener filter $\mathbf{H}_o$. (d) Expected ambiguity function of $s(t)$. (e) Expected ambiguity function of $n(t)$. (f) WS of TF pseudo-Wiener filter $\mathbf{H}_1$. The rectangles in parts (d) and (e) have area 1 and thus allow assessment of the under-/overspread property of $s(t)$ and $n(t)$. (In particular, part (e) shows that $n(t)$ is overspread.) The signal length is 128 samples.

to result in a significant performance loss, a modest filter order K already yields practically optimal performance.

An experiment in which the underspread assumption is violated is shown in Fig. 7. The signal $s(t)$ is again underspread (and, in addition, reasonably quasistationary), but the noise $n(t)$ is *not* underspread—it is reasonably quasistationary, but its temporal correlation width is too large, as can be seen from Fig. 7(e). A comparison of Fig. 7(c) and (f) shows that the TF pseudo-Wiener filter $\mathbf{H}_1$ is significantly different from the Wiener filter $\mathbf{H}_o$. The heavy oscillations of the WS of $\mathbf{H}_o$ in Fig. 7(c) indicate that $\mathbf{H}_o$ is far from being an underspread system. The processes $s(t)$ and $n(t)$ were constructed such that they lie in linearly independent (disjoint) signal spaces. Here, $\mathbf{H}_o$ is an oblique projection operator [44] that *perfectly* reconstructs $s(t)$, thereby achieving zero MSE and infinite SNR improvement. On the other hand, due to the significant overlap of the WVS of $s(t)$ and $n(t)$, the TF pseudo-Wiener filter $\mathbf{H}_1$ merely achieves an SNR improvement of 4.79 dB. This example is important since it shows that mere quasistationarity of $s(t)$ and $n(t)$ is not sufficient as an assumption permitting a TF design of optimal filters. It also shows that in certain cases—specifically, when signal and noise have significantly overlapping WVS but belong to linearly independent signal spaces—the TF designed filter performs poorly, even though the optimal Wiener filter achieves perfect signal reconstruction. We stress that this situation implies that the processes are *overspread* (i.e., not underspread [12]), thereby prohibiting beforehand the successful use of TF filter methods.

Finally, we present the results of a real-data experiment illustrating the application of the TF pseudo-Wiener filter to speech enhancement. The speech signal used consists of 246 pitch periods of the German vowel "a" spoken by a male speaker. This

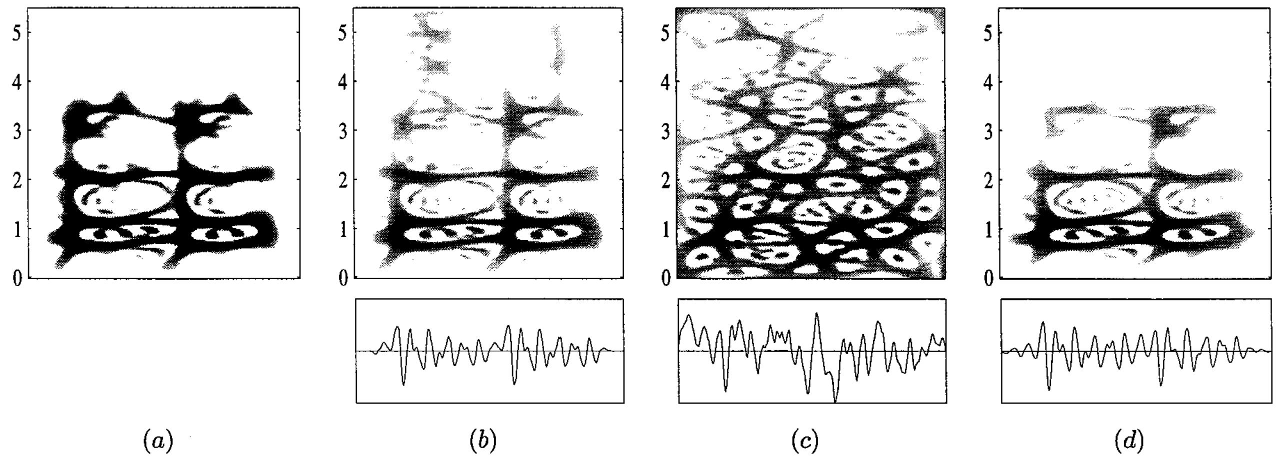


Fig. 8. Speech enhancement experiment using TF pseudo-Wiener filter. (a) WS of TF pseudo-Wiener filter $\mathbf{H}_1$. (b) Two pitch periods of noise-free speech signal. (c) Noise-corrupted speech signal. (d) Enhanced speech signal. All signals are represented by their time-domain waveforms and their (slightly smoothed) Wigner distributions [19], [58], [60], [61]. Horizontal axis: time (signal duration is 256 signal samples, corresponding to 2.32 ms). Vertical axis: frequency in kilohertz.

signal was sampled at 11.03 kHz and prefiltered in order to emphasize higher frequency content. The resulting discrete-time signal was split into 123 segments of length 256 samples (corresponding to 2.32 ms), each of which contains two pitch periods. Sixty four of these segments were considered as individual realizations of a nonstationary process of length 256 samples and were used to estimate the WVS of this process. Furthermore, we used a stationary $\mathrm{AR}(2)$ noise process that provides a reasonable model for car noise. Due to stationarity, the noise WVS equals the PSD for which a closed-form expression exists; hence, there was no need to estimate the noise WVS. Using the estimated signal WVS and the exact noise WVS, the TF pseudo-Wiener filter $\mathbf{H}_1$ was then designed according to (26). The WS of $\mathbf{H}_1$ is shown in Fig. 8(a).

The remaining 59 speech signal segments were corrupted by 59 different noise realizations and input to the TF pseudo-Wiener filter $\mathbf{H}_1$. The input SNR (averaged over the 59 realizations) was $\mathrm{SNR}_{\mathrm{in}} \approx 1.7$ dB, and the average SNR of the enhanced signal segments obtained at the filter output was $\mathrm{SNR}_{\mathrm{out}} \approx 8.4$ dB, corresponding to an average SNR improvement of about 6.7 dB. Results for one particular realization are shown in Fig. 8(b)–(d).

VIII. Conclusion

We have developed a time-frequency (TF) formulation and an approximate TF design of optimal filters that is valid for *underspread*, nonstationary random processes, i.e., nonstationary processes with limited TF correlation. We considered two types of optimal filters: the classical time-varying Wiener filter and a projection-constrained optimal filter. The latter will produce a larger estimation error but has practically important advantages concerning robustness and *a priori* knowledge.

The TF formulation of optimal filters was seen to allow an intuitively appealing interpretation of optimal filters in terms of passing, attenuating, or suppressing signal components located in different TF regions. The approximate TF design of optimal filters is practically attractive since it is computationally efficient and uses a more intuitive form of *a priori* knowledge. We furthermore discussed an efficient TF implementation of time-varying optimal filters using multiwindow STFT filters.

Our TF formulation and design was based on the Weyl symbol (WS) of linear, time-varying systems and the WVS of nonstationary random processes. However, if the processes satisfy a more restrictive type of underspread property (if they are *strictly underspread* [8], [29]), then our results can be extended in that the WS can be replaced by other members of the class of *generalized WSs* [48] (such as Zadeh's time-varying transfer function [25]), and the WVS can be replaced by other members of the class of *generalized WVS* [18], [19] or *generalized evolutionary spectra* [29]. This extension is possible since for strictly underspread systems, all generalized WS's are essentially equivalent and for strictly underspread processes, all generalized WVS and generalized evolutionary spectra are essentially equivalent [8], [9], [12], [19], [29], [47].

Appendix A
Underspread Approximation of the Wiener Filter: Proof of the Bounds (20) and (21)

We assume that the expected ambiguity functions of $s(t)$ and $n(t)$ are exactly zero outside the same (possibly obliquely oriented) centered rectangle $\mathcal{G}$ with area (joint correlation spread) $\sigma_{s,n}$.

We first prove the bound (20). Using $\mathbf{H}_o(\mathbf{R}_s + \mathbf{R}_n) = \mathbf{H}_o\mathbf{R}_r = \mathbf{R}_s$ and $\mathbf{H}_o = \mathbf{H}_o^{\mathcal{G}} + \mathbf{H}_o^{\bar{\mathcal{G}}}$, the approximation error $\|\mathbf{H}_o^{\mathcal{G}}\mathbf{R}_r - \mathbf{R}_s\|_{\mathrm{HS}}$ can be developed as

$$\begin{aligned}\left\|\mathbf{H}_o^{\mathcal{G}}\mathbf{R}_r - \mathbf{R}_s\right\|_{\mathrm{HS}}^2 &= \left\|\mathbf{H}_o^{\mathcal{G}}\mathbf{R}_r - \mathbf{H}_o\mathbf{R}_r\right\|_{\mathrm{HS}}^2 = \left\|\mathbf{H}_o^{\bar{\mathcal{G}}}\mathbf{R}_r\right\|_{\mathrm{HS}}^2 \\ &= \int_\tau \int_\nu \left|S_{\mathbf{H}_o^{\bar{\mathcal{G}}}\mathbf{R}_r}(\tau,\nu)\right|^2 d\tau\, d\nu. \qquad (34)\end{aligned}$$

We will now show that $S_{\mathbf{H}_o^{\bar{\mathcal{G}}}\mathbf{R}_r}(\tau,\nu)$ is zero outside the centered rectangle $\mathcal{G}_2$ that is oriented as $\mathcal{G}$ but has sides that are twice as long. Using $S_{\mathbf{R}_r}(\tau,\nu) = \bar{A}_r(\tau,\nu)$, it can be shown [13] that $S_{\mathbf{H}_o^{\bar{\mathcal{G}}}\mathbf{R}_r}(\tau,\nu) = (S_{\mathbf{H}_o^{\bar{\mathcal{G}}}} \natural \bar{A}_r)(\tau,\nu)$, where

$$\left(S_{\mathbf{H}_o^{\bar{\mathcal{G}}}} \natural \bar{A}_r\right)(\tau,\nu) \triangleq \int_{\tau_1}\int_{\nu_1} S_{\mathbf{H}_o^{\bar{\mathcal{G}}}}(\tau_1,\nu_1)\bar{A}_r(\tau-\tau_1,\nu-\nu_1) \cdot e^{j\pi(\tau\nu_1-\tau_1\nu)}\, d\tau_1\, d\nu_1 \tag{35}$$

is known as the *twisted convolution* of $S_{\mathbf{H}_o^{\bar{\mathcal{G}}}}(\tau,\nu)$ and $\bar{A}_r(\tau,\nu)$ [13]. In the (τ,ν) domain, $\mathbf{H}_o\mathbf{R}_r = \mathbf{R}_s$ corresponds to $S_{\mathbf{H}_o\mathbf{R}_r}(\tau,\nu) = S_{\mathbf{R}_s}(\tau,\nu)$, which can be rewritten as $(S_{\mathbf{H}_o}\natural\bar{A}_r)(\tau,\nu) = \bar{A}_s(\tau,\nu)$. With $\mathbf{H}_o = \mathbf{H}_o^{\mathcal{G}} + \mathbf{H}_o^{\bar{\mathcal{G}}}$, we obtain

$$\left(S_{\mathbf{H}_o^{\mathcal{G}}} \natural \bar{A}_r\right)(\tau,\nu) + \left(S_{\mathbf{H}_o^{\bar{\mathcal{G}}}} \natural \bar{A}_r\right)(\tau,\nu) = \bar{A}_s(\tau,\nu). \tag{36}$$

The twisted convolution [cf. (35)] implies that

$$\begin{aligned}\left|\left(S_{\mathbf{H}_o^{\mathcal{G}}} \natural \bar{A}_r\right)(\tau,\nu)\right| &\\ \le \int_{\tau_1}\int_{\nu_1} &\left|S_{\mathbf{H}_o^{\mathcal{G}}}(\tau_1,\nu_1)\right| \left|\bar{A}_r(\tau-\tau_1,\nu-\nu_1)\right| d\tau_1\, d\nu_1\end{aligned}$$

and, hence, that the support of $(S_{\mathbf{H}_o^{\mathcal{G}}}\natural\bar{A}_r)(\tau,\nu)$ is confined to $\mathcal{G}_2$ [recall that both $S_{\mathbf{H}_o^{\mathcal{G}}}(\tau,\nu)$ and $\bar{A}_r(\tau,\nu)$ are confined to $\mathcal{G}$]. The crucial observation now is that since $\bar{A}_s(\tau,\nu)$ is confined to $\mathcal{G} \subset \mathcal{G}_2$ and $(S_{\mathbf{H}_o^{\mathcal{G}}}\natural\bar{A}_r)(\tau,\nu)$ is confined to $\mathcal{G}_2$, (36) implies that $(S_{\mathbf{H}_o^{\bar{\mathcal{G}}}}\natural\bar{A}_r)(\tau,\nu)$ must also be confined to $\mathcal{G}_2$ since any contributions of $(S_{\mathbf{H}_o^{\bar{\mathcal{G}}}}\natural\bar{A}_r)(\tau,\nu)$ outside $\mathcal{G}_2$ cannot be canceled by $(S_{\mathbf{H}_o^{\mathcal{G}}}\natural\bar{A}_r)(\tau,\nu)$ and, thus, would contradict (36). Hence, we can write (34) as

$$\left\|\mathbf{H}_o^{\mathcal{G}}\mathbf{R}_r - \mathbf{R}_s\right\|_{\mathrm{HS}}^2 = \iint_{\mathcal{G}_2} \left|\left(S_{\mathbf{H}_o^{\bar{\mathcal{G}}}} \natural \bar{A}_r\right)(\tau,\nu)\right|^2 d\tau\, d\nu. \tag{37}$$

Applying the Schwarz inequality to (35) yields the bound $|(S_{\mathbf{H}_o^{\bar{\mathcal{G}}}}\natural\bar{A}_r)(\tau,\nu)| \le \|S_{\mathbf{H}_o^{\bar{\mathcal{G}}}}\|\|S_{\mathbf{R}_r}\| = \|\mathbf{H}_o^{\bar{\mathcal{G}}}\|_{\mathrm{HS}}\|\mathbf{R}_r\|_{\mathrm{HS}}$. Inserting this into (37) and using the fact that the area of $\mathcal{G}_2$ is $4\sigma_{s,n}$ (i.e., four times the area of $\mathcal{G}$), we obtain (20):

$$\begin{aligned}\left\|\mathbf{H}_o^{\mathcal{G}}\mathbf{R}_r - \mathbf{R}_s\right\|_{\mathrm{HS}}^2 &\le \left\|\mathbf{H}_o^{\bar{\mathcal{G}}}\right\|_{\mathrm{HS}}^2 \|\mathbf{R}_r\|_{\mathrm{HS}}^2 \iint_{\mathcal{G}_2} d\tau\, d\nu \\ &= \left\|\mathbf{H}_o^{\bar{\mathcal{G}}}\right\|_{\mathrm{HS}}^2 \|\mathbf{R}_r\|_{\mathrm{HS}}^2 4\sigma_{s,n}.\end{aligned} \tag{38}$$

We next prove the bound (21). The lower bound $\bar{E}_e^{\mathcal{G}} - \bar{E}_{e_o} \ge 0$ is trivial since $\bar{E}_{e_o}$ is the minimum possible MSE. Let us consider the upper bound. With (3), the MSE achieved by $\mathbf{H}_o^{\mathcal{G}}$ can be written as

$$\begin{aligned}\bar{E}_e^{\mathcal{G}} &= \mathrm{tr}\left\{\mathbf{H}_o^{\mathcal{G}}\mathbf{R}_n\mathbf{H}_o^{\mathcal{G}+} + \left(\mathbf{I} - \mathbf{H}_o^{\mathcal{G}}\right)\mathbf{R}_s\left(\mathbf{I} - \mathbf{H}_o^{\mathcal{G}}\right)^+\right\} \\ &= \mathrm{tr}\left\{\mathbf{H}_o^{\mathcal{G}}(\mathbf{R}_n + \mathbf{R}_s)\mathbf{H}_o^{\mathcal{G}+} + \mathbf{R}_s - \mathbf{H}_o^{\mathcal{G}}\mathbf{R}_s - \mathbf{R}_s\mathbf{H}_o^{\mathcal{G}+}\right\} \\ &= \mathrm{tr}\left\{\mathbf{R}_s + \mathbf{H}_o^{\mathcal{G}}\mathbf{R}_r\mathbf{H}_o^{\mathcal{G}+} - \mathbf{H}_o^{\mathcal{G}}\mathbf{R}_s - \mathbf{R}_s\mathbf{H}_o^{\mathcal{G}+}\right\} \\ &= \mathrm{tr}\left\{\left(\mathbf{I} - \mathbf{H}_o^{\mathcal{G}}\right)\mathbf{R}_s\right\} + \mathrm{tr}\left\{\left(\mathbf{H}_o^{\mathcal{G}}\mathbf{R}_r - \mathbf{R}_s\right)\mathbf{H}_o^{\mathcal{G}+}\right\}.\end{aligned}$$

Furthermore, $\bar{E}_{e_o} = \mathrm{tr}\{(\mathbf{I}-\mathbf{H}_o)\mathbf{R}_s\}$ according to (6). We therefore obtain

$$\begin{aligned}&\bar{E}_e^{\mathcal{G}} - \bar{E}_{e_o} \\ &\quad = \mathrm{tr}\left\{\left(\mathbf{I} - \mathbf{H}_o^{\mathcal{G}}\right)\mathbf{R}_s\right\} + \mathrm{tr}\left\{\left(\mathbf{H}_o^{\mathcal{G}}\mathbf{R}_r - \mathbf{R}_s\right)\mathbf{H}_o^{\mathcal{G}+}\right\} \\ &\qquad - \mathrm{tr}\left\{(\mathbf{I} - \mathbf{H}_o)\mathbf{R}_s\right\} \\ &\quad = \mathrm{tr}\left\{\left(\mathbf{H}_o - \mathbf{H}_o^{\mathcal{G}}\right)\mathbf{R}_s\right\} + \mathrm{tr}\left\{\left(\mathbf{H}_o^{\mathcal{G}}\mathbf{R}_r - \mathbf{R}_s\right)\mathbf{H}_o^{\mathcal{G}+}\right\} \\ &\quad = \mathrm{tr}\left\{\mathbf{H}_o^{\bar{\mathcal{G}}}\mathbf{R}_s\right\} + \mathrm{tr}\left\{\left(\mathbf{H}_o^{\mathcal{G}}\mathbf{R}_r - \mathbf{R}_s\right)\mathbf{H}_o^{\mathcal{G}+}\right\}.\end{aligned}$$

The first term is zero, i.e., $\mathrm{tr}\{\mathbf{H}_o^{\bar{\mathcal{G}}}\mathbf{R}_s\} = \langle S_{\mathbf{H}_o^{\bar{\mathcal{G}}}}, \bar{A}_s\rangle = 0$, since $\bar{A}_s(\tau,\nu) = 0$ outside $\mathcal{G}$. The second term can be bounded by applying Schwarz' inequality:

$$\mathrm{tr}\left\{\left(\mathbf{H}_o^{\mathcal{G}}\mathbf{R}_r - \mathbf{R}_s\right)\mathbf{H}_o^{\mathcal{G}+}\right\} \le \left\|\mathbf{H}_o^{\mathcal{G}}\mathbf{R}_r - \mathbf{R}_s\right\|_{\mathrm{HS}} \left\|\mathbf{H}_o^{\mathcal{G}}\right\|_{\mathrm{HS}}.$$

Using (38), the upper bound in (21) follows.

Appendix B
TF Formulation of the Wiener Filter: Proof of the Bound (22)

We consider the approximation error

$$\Delta(t,f) \triangleq L_{\mathbf{H}_o^{\mathcal{G}}}(t,f)\bar{W}_r(t,f) - \bar{W}_s(t,f).$$

Subtracting and adding $L_{\mathbf{H}_o^{\mathcal{G}}\mathbf{R}_r}(t,f)$, this can be rewritten as $\Delta(t,f) = \Delta_1(t,f) + \Delta_2(t,f)$ with

$$\begin{aligned}\Delta_1(t,f) &= L_{\mathbf{H}_o^{\mathcal{G}}}(t,f)\bar{W}_r(t,f) - L_{\mathbf{H}_o^{\mathcal{G}}\mathbf{R}_r}(t,f), \\ \Delta_2(t,f) &= L_{\mathbf{H}_o^{\mathcal{G}}\mathbf{R}_r}(t,f) - \bar{W}_s(t,f).\end{aligned}$$

According to the triangle inequality, the L_2 norm of $\Delta(t,f)$ is bounded as $\|\Delta\| \le \|\Delta_1\| + \|\Delta_2\|$. It is shown in [8], [11], and [46] that $\|\Delta_1\|$ is bounded as

$$\begin{aligned}\|\Delta_1\| &= \left\|L_{\mathbf{H}_o^{\mathcal{G}}}\bar{W}_r - L_{\mathbf{H}_o^{\mathcal{G}}\mathbf{R}_r}\right\| \\ &\le \pi\left\|\mathbf{H}_o^{\mathcal{G}}\right\|_{\mathrm{HS}}\|\mathbf{R}_r\|_{\mathrm{HS}}\sqrt{\sigma^3_{\mathbf{H}_o^{\mathcal{G}},\mathbf{R}_r}} \\ &= \pi\left\|\mathbf{H}_o^{\mathcal{G}}\right\|_{\mathrm{HS}}\|\mathbf{R}_r\|_{\mathrm{HS}}\sqrt{\sigma^3_{s,n}}\end{aligned}$$

where we have used $\sigma_{\mathbf{H}_o^{\mathcal{G}},\mathbf{R}_r} = \sigma_{s,n}$. Furthermore, $\|\Delta_2\|$ is bounded as

$$\begin{aligned}\|\Delta_2\| &= \left\|L_{\mathbf{H}_o^{\mathcal{G}}\mathbf{R}_r} - \bar{W}_s\right\| = \left\|\mathbf{H}_o^{\mathcal{G}}\mathbf{R}_r - \mathbf{R}_s\right\|_{\mathrm{HS}} \\ &\le \left\|\mathbf{H}_o^{\bar{\mathcal{G}}}\right\|_{\mathrm{HS}}\|\mathbf{R}_r\|_{\mathrm{HS}} 2\sqrt{\sigma_{s,n}}\end{aligned}$$

where we have used (38). Inserting these two bounds into $\|\Delta\| \le \|\Delta_1\| + \|\Delta_2\|$ gives (22).

References

[1] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. Cambridge, MA: MIT Press, 1949.

[2] T. Kailath, *Lectures on Wiener and Kalman Filtering*. Wien, Austria: Springer, 1981.

[3] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 1984.

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

[5] C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1992.

[6] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer, 1988.

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

[8] W. Kozek, "Matched Weyl-Heisenberg Expansions of Nonstationary Environments," Ph.D. dissertation, Vienna Univ. Technol., Vienna, Austria, Mar. 1997.

[9] W. Kozek, F. Hlawatsch, H. Kirchauer, and U. Trautwein, "Correlative time-frequency analysis and classification of nonstationary random processes," in *Proc. IEEE-SP Int. Symp. Time-Frequency Time-Scale Anal.*, Philadelphia, PA, Oct. 1994, pp. 417–420.

[10] W. Kozek, "On the underspread/overspread classification of nonstationary random processes," in *Proc. Int. Conf. Ind. Appl. Math.*, K. Kirchgässner, O. Mahrenholtz, and R. Mennicken, Eds. Berlin, Germany, 1996, pp. 63–66.

[11] ——, "Adaptation of Weyl-Heisenberg frames to underspread environments," in *Gabor Analysis and Algorithms: Theory and Applications*, H. G. Feichtinger and T. Strohmer, Eds. Boston, MA: Birkhäuser, 1998, ch. 10, pp. 323–352.

[12] G. Matz and F. Hlawatsch, "Time-varying spectra for underspread and overspread nonstationary processes," in *Proc. 32nd Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 1998, pp. 282–286.

[13] G. B. Folland, *Harmonic Analysis in Phase Space*. Princeton, NJ: Princeton Univ. Press, 1989.

[14] A. J. E. M. Janssen, "Wigner weight functions and Weyl symbols of non-negative definite linear operators," *Philips J. Res.*, vol. 44, pp. 7–42, 1989.

[15] W. Kozek, "Time-frequency signal processing based on the Wigner-Weyl framework," *Signal Process.*, vol. 29, pp. 77–92, Oct. 1992.

[16] R. G. Shenoy and T. W. Parks, "The Weyl correspondence and time-frequency analysis," *IEEE Trans. Signal Processing*, vol. 42, pp. 318–331, Feb. 1994.

[17] W. Martin and P. Flandrin, "Wigner–Ville spectral analysis of nonstationary processes," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 1461–1470, Dec. 1985.

[18] P. Flandrin and W. Martin, "The Wigner–Ville spectrum of nonstationary random signals," in *The Wigner Distribution—Theory and Applications in Signal Processing*, W. Mecklenbräuker and F. Hlawatsch, Eds. Amsterdam, The Netherlands: Elsevier, 1997, pp. 211–267.

[19] P. Flandrin, *Time-Frequency/Time-Scale Analysis*. San Diego, CA: Academic, 1999.

[20] W. Kozek, H. G. Feichtinger, and J. Scharinger, "Matched multiwindow methods for the estimation and filtering of nonstationary processes," in *Proc. IEEE ISCAS*, Atlanta, GA, May 1996, pp. 509–512.

[21] N. A. Abdrabbo and M. B. Priestley, "Filtering non-stationary signals," *J. R. Stat. Soc. Ser. B.*, vol. 31, pp. 150–159, 1969.

[22] J. A. Sills, "Nonstationary signal modeling, filtering, and parameterization," Ph.D. dissertation, Georgia Inst. Technol., Atlanta, Mar. 1995.

[23] J. A. Sills and E. W. Kamen, "Wiener filtering of nonstationary signals based on spectral density functions," in *Proc. 34th IEEE Conf. Decision Contr.*, Kobe, Japan, Dec. 1995, pp. 2521–2526.

[24] H. A. Khan and L. F. Chaparro, "Nonstationary Wiener filtering based on evolutionary spectral theory," in *Proc. IEEE ICASSP*, Munich, Germany, May 1997, pp. 3677–3680.

[25] L. A. Zadeh, "Frequency analysis of variable networks," *Proc. IRE*, vol. 76, pp. 291–299, Mar. 1950.

[26] P. A. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun. Syst.*, vol. COMM-11, pp. 360–393, 1963.

[27] M. B. Priestley, "Evolutionary spectra and nonstationary processes," *J. R. Stat. Soc. Ser. B.*, vol. 27, no. 2, pp. 204–237, 1965.

[28] ——, *Spectral Analysis and Time Series—Part II*. London, U.K.: Academic, 1981.

[29] G. Matz, F. Hlawatsch, and W. Kozek, "Generalized evolutionary spectral analysis and the Weyl spectrum of nonstationary random processes," *IEEE Trans. Signal Processing*, vol. 45, pp. 1520–1534, June 1997.

[30] A. A. Beex and M. Xie, "Time-varying filtering via multiresolution parametric spectral estimation," in *Proc. IEEE ICASSP*, Detroit, MI, May 1995, pp. 1565–1568.

[31] P. Lander and E. J. Berbari, "Enhanced ensemble averaging using the time-frequency plane," in *Proc. IEEE-SP Int. Symp. Time-Frequency Time-Scale Anal.*, Philadelphia, PA, Oct. 1994, pp. 241–243.

[32] A. M. Sayeed, P. Lander, and D. L. Jones, "Improved time-frequency filtering of signal-averaged electrocardiograms," *J. Electrocardiol.*, vol. 28, pp. 53–58, 1995.

[33] W. D. Mark, "Spectral analysis of the convolution and filtering of nonstationary stochastic processes," *J. Sound Vibr.*, vol. 11, no. 1, pp. 19–63, 1970.

[34] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proc. IEEE*, vol. 67, pp. 1586–1604, Dec. 1979.

[35] 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. ASSP-32, pp. 1109–1121, Dec. 1984.

[36] ——, "Speech enhancement using a minimum mean-square error log spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 443–445, Apr. 1985.

[37] G. Doblinger, "Computationally efficient speech enhancement by spectral minima tracking in subbands," in *Proc. Eurospeech*, Madrid, Spain, Sept. 1995, pp. 1513–1516.

[38] G. W. Wornell and A. V. Oppenheim, "Estimation of fractal signals from noisy measurements using wavelets," *IEEE Trans. Signal Processing*, vol. 40, pp. 611–623, Mar. 1992.

[39] G. W. Wornell, *Signal Processing with Fractals: A Wavelet-Based Approach*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[40] M. Unser, "Wavelets, statistics, and biomedical applications," in *Proc. IEEE SP Workshop Stat. Signal Array Process.*, Corfu, Greece, June 1996, pp. 244–249.

[41] S. P. Ghael, A. M. Sayeed, and R. G. Baraniuk, "Improved wavelet denoising via empririical Wiener filtering," in *Proc. SPIE Wavelet Appl. Signal Image Process. V*, San Diego, CA, July 1997, pp. 389–399.

[42] F. Hlawatsch, *Time-Frequency Analysis and Synthesis of Linear Signal Spaces: Time-Frequency Filters, Signal Detection and Estimation, and Range-Doppler Estimation*. Boston, MA: Kluwer, 1998.

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

[44] R. T. Behrens and L. L. Scharf, "Signal processing applications of oblique projection operators," *IEEE Trans. Signal Processing*, vol. 42, pp. 1413–1424, June 1994.

[45] G. Matz and F. Hlawatsch, "Robust time-varying Wiener filters: Theory and time-frequency formulation," in *Proc. IEEE-SP Int. Symp. Time-Frequency Time-Scale Anal.*, Pittsburgh, PA, Oct. 1998, pp. 401–404.

[46] W. Kozek, "On the transfer function calculus for underspread LTV channels," *IEEE Trans. Signal Processing*, vol. 45, pp. 219–223, Jan. 1997.

[47] G. Matz and F. Hlawatsch, "Time-frequency transfer function calculus (symbolic calculus) of linear time-varying systems (linear operators) based on a generalized underspread theory," *J. Math. Phys., Special Issue on Wavelet and Time-Frequency Analysis*, vol. 39, pp. 4041–4071, Aug. 1998.

[48] W. Kozek, "On the generalized Weyl correspondence and its application to time-frequency analysis of linear time-varying systems," in *Proc. IEEE-SP Int. Symp. Time-Frequency Time-Scale Anal.*, Victoria, Ont., Canada, Oct. 1992, pp. 167–170.

[49] K. A. Sostrand, "Mathematics of the time-varying channel," *Proc. NATO Adv. Study Inst. Signal Process. Emphasis Underwater Acoust.*, vol. 2, pp. 25.1–25.20, 1968.

[50] H. Kirchauer, F. Hlawatsch, and W. Kozek, "Time-frequency formulation and design of nonstationary Wiener filters," in *Proc. IEEE ICASSP*, Detroit, MI, May 1995, pp. 1549–1552.

[51] F. Hlawatsch and W. Kozek, "Time-frequency projection filters and time-frequency signal expansions," *IEEE Trans. Signal Processing*, vol. 42, pp. 3321–3334, Dec. 1994.

[52] I. Daubechies, "Time-frequency localization operators: A geometric phase space approach," *IEEE Trans. Inf. Theory*, vol. 34, pp. 605–612, July 1988.

[53] W. Kozek and F. Hlawatsch, "A comparative study of linear and nonlinear time-frequency filters," in *Proc. IEEE-SP Int. Sympos. Time-Frequency Time-Scale Analysis*, Victoria, B.C., Canada, Oct. 1992, pp. 163–166.

[54] M. R. Portnoff, "Time-frequency representation of digital signals and systems based on short-time Fourier analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 55–69, Feb. 1980.

[55] S. H. Nawab and T. F. Quatieri, "Short-time Fourier transform," in *Advanced Topics in Signal Processing*, J. S. Lim and A. V. Oppenheim, Eds. Englewood Cliffs, NJ: Prentice-Hall, 1988, ch. 6, pp. 289–337.

[56] R. Bourdier, J. F. Allard, and K. Trumpf, "Effective frequency response and signal replica generation for filtering algorithms using multiplicative modifications of the STFT," *Signal Process.*, vol. 15, pp. 193–201, Sept. 1988.

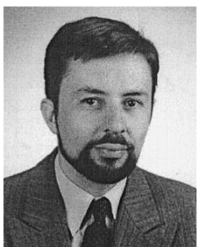
[57] M. L. Kramer and D. L. Jones, "Improved time-frequency filtering using an STFT analysis-modification-synthesis method," in *Proc. IEEE-SP Int. Symp. Time-Frequency Time-Scale Anal.*, Philadelphia, PA, Oct. 1994, pp. 264–267.

[58] F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations," *IEEE Signal Processing Mag.*, vol. 9, pp. 21–67, Apr. 1992.

[59] F. Hlawatsch and W. Kozek, "Second-order time-frequency synthesis of nonstationary random processes," *IEEE Trans. Inform. Theory*, vol. 41, pp. 255–267, Jan. 1995.

[60] T. A. C. M. Claasen and W. F. G. Mecklenbräuker, "The Wigner distribution—A tool for time-frequency signal analysis; Parts I–III," *Philips J. Res.*, vol. 35, pp. 217–250; 276–300; 372–389, 1980.

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

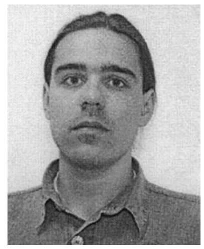


Franz Hlawatsch (S'85–M'88) received the Dipl.-Ing., Dr.techn., and Univ.-Dozent degrees in electrical engineering communications engineering/signal processing from the Vienna University of Technology, Vienna, Austria, in 1983, 1988, and 1996, respectively.

Since 1983, he has been with the Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology. From 1991 to 1992, he spent a sabbatical year with the Department of Electrical Engineering, University of Rhode Island, Kingston. He authored the book *Time-Frequency Analysis and Synthesis of Linear Signal Spaces—Time-Frequency Filters, Signal Detection and Estimation, and Range-Doppler Estimation* (Boston, MA: Kluwer, 1998) and co-edited the book *The Wigner Distribution—Theory and Applications in Signal Processing* (Amsterdam, The Netherlands: Elsevier, 1997). His research interests are in signal processing with emphasis on time-frequency methods and communications applications.

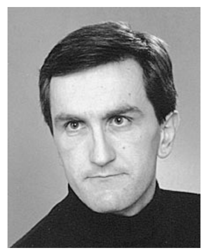
Gerald Matz (S'95) received the Dipl.-Ing. degree in electrical engineering from the Vienna University of Technology, Vienna, Austria, in 1994.

Since 1995, he has been with the Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology. His research interests include the application of time-frequency methods to statistical signal processing and wireless communications.



Heinrich Kirchauer was born in Vienna, Austria, in 1969. He received the Dip.-Ing. degree in communications engineering and the doctoral degree in microelectronics engineering from the Vienna University of Technology, Vienna, Austria, in 1994 and 1998, respectively.

In the summer of 1997 , he held a visiting research position at LSI Logic, Milpitas, CA. In 1998, he joined Intel Corporation, Santa Clara, CA. His current interests are modeling and simulation of problems for microelectronics engineering with special emphasis on lithography simulation.



Werner Kozek (M'94) received the Dipl.-Ing. and Ph.D. degrees in electrical engineering from the Vienna University of Technology, Vienna, Austria, in 1990 and 1997, respectively.

From 1990 to 1994 he was with the Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology, and from 1994 to 1998, he was with the Department of Mathematics, University of Vienna, both as a Research Assistant. Since 1998, he has been with Siemens AG, Munich, Germany, working on advanced xDSL technology. His research interests include the signal processing and information theoretic aspects of digital communication systems.