TWO–DIMENSIONAL PILOT–SYMBOL–AIDED CHANNEL ESTIMATION BY WIENER FILTERING

Peter Hoeher, Stefan Kaiser, and Patrick Robertson
Institute for Communications Technology
German Aerospace Research Establishment (DLR)
P.O. Box 1116, D-82230 Oberpfaffenhofen, Germany
E-mail: Firstname.Lastname@dlr.de

ABSTRACT

The potentials of pilot-symbol-aided channel estimation in two dimensions are explored. In order to procure this goal, the discrete shift-variant 2-D Wiener filter is derived and analyzed given an arbitrary sampling grid, an arbitrary (but possibly optimized) selection of observations, and the possibility of model mismatch. Filtering in two dimensions is revealed to outperform filtering in just one dimension with respect to overhead and mean-square error performance. However, two cascaded orthogonal 1-D filters are simpler to implement and shown to be virtually as good as true 2-D filters.

1 INTRODUCTION

ACCORDING to the sampling theorem, any band-limited 1-D stochastic process or deterministic signal is uniquely representable by samples taken at least at the Nyquist rate. If the process is observed in noise, however, a perfect reconstruction is not possible. The Wiener filter is the optimum (minimum mean-squared error) linear filter/smoother/predictor, if the noise is additive. The sampling theorem also holds for multi-dimensional processes and signals [1]. Generalizations of the Wiener filter for multiple dimensions exist as well, both for continuous and discrete signals [2, 3]. Applications considered so far include image processing, sensor-array processing, and geophysical applications.

Pilot-symbol-aided channel estimation of time- and/or frequency-selective channels is another potential application. The basic principle of 1-D pilot-symbol-aided channel estimation is to multiplex pilot (training) symbols known to the receiver into the data stream [4, 5]. Hence, the receiver is able to estimate the process (the channel) at any time given the observations at the pilot locations, assuming the pilot density is sufficient w.r.t. the channel bandwidth. 1-D pilot-symbol-aided channel estimation is well understood; different filters have been investigated [4, 5, 6, 7].

When the channel is probed simultaneously in both time and frequency domains, as proposed by the first author, the overhead of pilot symbols can be reduced significantly [8]. This scheme is particularly attractive for multi-carrier modulation such as Orthogonal Frequency-Division Multiplexing (OFDM) [8, 9]. Applications include the upcoming European Terrestrial Digital Video Broadcasting and Digital Audio Broadcasting standards, mobile radio, and line modems. However so far, to our best knowledge, only filtering with two cascaded orthogonal 1-D filters was studied for this particular application, referred to as 2× 1-D filtering. In this paper, we explore the potentials of pilot-symbol-aided channel estimation in two dimensions.

2 THE 2-D WIENER FILTER

The formulation of the 2-D discrete minimum mean-square error (MMSE) estimation problem and its solution is as follows [3]. Assume it is of interest to estimate a wide-sense stationary (WSS) 2-D stochastic process $h(k,l) \in \mathbb{C}$ disturbed by a WSS additive discrete noise process $n(k,l) \in \mathbb{C}$, where $0 \le k \le K-1$ and $0 \le l \le L-1$ are the two indices of the process (e.g., time/frequency indices for the example of channel estimation) and $K \cdot L$ is the arbitrary (finite or infinite) block size. Hence, the received bisequence is

$$r(k,l) = h(k,l) + n(k,l), \ \forall\, 0 \le k \le K-1,\ 0 \le l \le L-1. \tag{1}$$

Assume further that $r(k,l) \in \mathbb{C}$ is observed only at sample locations (pilot locations) $\mathcal{P}$. The corresponding indices are denoted as k' and l', respectively, where again $0 \le k' \le K-1$ and $0 \le l' \le L-1$. The number of observations (later: pilot symbols) is $N_{grid} = ||\mathcal{P}|| \le K \cdot L$. The estimator is chosen to be discrete, linear, and generally shift-variant:

$$\hat{h}(k,l) = \sum_{\{k',l'\} \in \mathcal{P}} w(k,l;k',l')\ r(k',l'), \tag{2}$$

where $\hat{h}(k,l) \in \mathbb{C}$ is the estimated bisequence and $w(k,l;k',l') \in \mathbb{C}$ is the shift-variant impulse response of the estimator or filter. The optimal filter order, i.e. the number of filter coefficients, is $N_{tap} = N_{grid}$; a simplification is introduced in Section 5. An example is shown in Fig. 1 illustrating the three special cases of prediction, filtering and smoothing w.r.t. the actual location(s) $\{k,l\}$ given a sampling grid $\mathcal{P}$.

The mean square error is defined as

$$J(\underline{w}(k,l)) = E[|h(k,l) - \hat{h}(k,l)|^2], \tag{3}$$

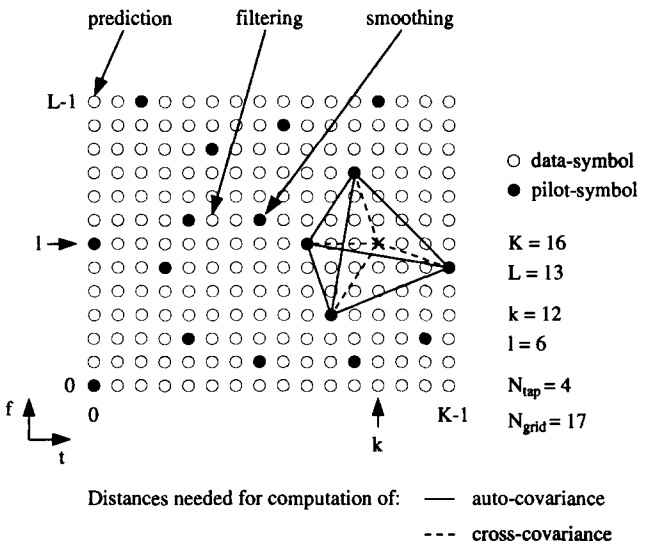

Figure 1: Random sampling grid illustrating prediction, filtering and smoothing, and illustrating the computation of cross- and auto-covariances.

where $E[.]$ denotes the expectation and $\underline{w}$ is the N_{tap} by 1 filter coefficient vector. The optimal filter in the sense of minimizing (3), i.e. the MMSE estimator or 2-D Wiener filter, is obtained by applying the orthogonal projection theorem [3]:

$$E[(h(k,l) - \hat{h}(k,l)) \cdot r^*(k'',l'')] = 0$$
$$\forall\ 0 \le k'' \le K-1,\ 0 \le l'' \le L-1, \qquad (4)$$

where $()^*$ denotes the complex conjugate. Although (4) holds for all $\{k'',l''\}$, in Section 6.2 $\{k'',l''\}$ are from the pilot grid $\mathcal{P}$. We assume that the Wiener filter is physically realizable, i.e., the coefficients $w(k,l;k',l')$ exist. The optimum filter coefficients are denoted as $w_o(k,l;k',l')$. After basic operations, the 2-D discrete Wiener-Hopf equation is obtained:

$$E[h(k,l)\ r^*(k'',l'')] = \sum_{\{k',l'\}\in\mathcal{P}} w_o(k,l;k',l')$$
$$\cdot E[r(k',l')\ r^*(k'',l'')], \qquad \forall\{k'',l''\} \in \mathcal{P}. \quad (5)$$

Let

$$\theta(k-k'',l-l'') = E[h(k,l)\ r^*(k'',l'')] \qquad (6)$$

denote the cross-covariance and

$$\Phi(k'-k'',l'-l'') = E[r(k',l')\ r^*(k'',l'')] \qquad (7)$$

denote the auto-covariance. Inserting (6) and (7) into (5) yields

$$\underline{\theta}^T(k,l) = \underline{w}_o^T(k,l)\,\underline{\Phi}, \qquad (8)$$

where $\underline{\Phi}$ is an N_{tap} by N_{tap} auto-covariance matrix and $\underline{\theta}(k,l)$ is an N_{tap} by 1 cross-covariance vector and $()^T$ denotes the transpose. Hence the optimal solution (if existent) is

$$\underline{w}_0^T(k,l) = \underline{\theta}^T(k,l)\,\underline{\Phi}^{-1}. \qquad (9)$$

3 PERFORMANCE ANALYSIS

For a given set of filter coefficients $\underline{w}$, (2) can be written in matrix form as

$$\hat{h}(k,l) = \underline{w}^T(k,l)\,\underline{r}. \qquad (10)$$

Substitution into (3) yields the MSE [3]

$$J(\underline{w}(k,l)) = \sigma_h^2 - \underline{\theta}^T(k,l)\,\underline{w}^*(k,l) - \underline{w}^T(k,l)\,\underline{\theta}^*(k,l)$$
$$+\underline{w}^T(k,l)\,\underline{\Phi}\,\underline{w}^*(k,l); \quad J(.) \in I\!R, \qquad (11)$$

where $()^H$ is the hermitian operator (conjugate transpose). This equation is valid for *any* FIR filter $\underline{w}$, i.e., also for model mismatch.

If the cross- and auto-covariance functions are known, the MMSE is obtained by inserting (9) into (11):

$$J_o(k,l) = \sigma_h^2 - \underline{\theta}^T(k,l)\underline{\Phi}^{-1}\underline{\theta}^*(k,l). \qquad (12)$$

4 GENERALIZED TWO–DIM. SAMPLING THEOREM

It is interesting to recall that the multi-dimensional Wiener theory is a generalization of the sampling theorem, as pointed out in [7] for the 1-D case. The multi-dimensional Wiener filter is the optimal predictor/filter/smoother in the presence of additive noise, for finite or infinite (bi)sequences when given an arbitrary sampling grid $\mathcal{P}$. It provides a compact tool to analyze the MSE including model mismatch and aliasing effects.

Equations (9) and (12) (respectively and (11)) manifest the *generalized 2-D sampling theorem*. The conventional 2-D sampling theorem is given by the special case where $J(\underline{w}(k,l)) = 0, \forall\ \{k,l\}$. This implies that the bisequence $h(k,l)$ is bandlimited in both dimensions, sampling is sufficient in both dimensions (according to the known bandwidths of h), and the additive noise process $n(k,l)$ is zero.

5 SIMPLIFICATIONS

The 2-D Wiener filter exists iff $\underline{\Phi}$ is invertible. Coefficient sets can be pre-computed and stored. So far, the number of coefficients per set, N_{tap}, was assumed to be equal to the number of observations, N_{grid}. The complexity can be significantly reduced, however, if only a subset $\mathcal{T}(k,l) \in \mathcal{P}$ of $N_{tap} < N_{grid}$ observations is chosen, where $N_{tap} = ||\mathcal{T}(k,l)||$. The performance/complexity trade-off is adjustable. Suitable selection criteria are: **1.)** Optimize $\mathcal{T}(k,l)$ for each location $\{k,l\}$ so that $J(\underline{w}(k,l))$ is minimized. This optimum rule is rather complex for practical block sizes: $\binom{N_{grid}}{N_{tap}}$ possibilities must be checked. **2.)** Search the N_{tap} sample locations $\{k',l'\}$ "nearest" to the actual location $\{k,l\}$. "Nearest"

could be w.r.t. **A.**) the sum of horizontal and vertical distances, $|k-k'|+|l-l'|$ (suitable if the normalized channel bandwidths are similar in both dimensions; using the Euclidean distance $\sqrt{(k-k')^2+(l-l')^2}$ appeared to be worse), or **B.**) some form of "weighted" distances, such as $|k-k'|f_{D_{max}}T_s + |l-l'|\tau_{max}\Delta F$ for the example in Section 6.2.

6 EXAMPLE

6.1 Channel Model

A wide-sense stationary uncorrelated scattering (WSSUS) mobile radio channel model can be expressed as

$$h(k,l) = \lim_{N\to\infty} \frac{1}{\sqrt{N}} \sum_{n=1}^{N} e^{j(\phi_n + 2\pi f_{D_n} T_s \cdot k + 2\pi\tau_n \Delta F \cdot l)}, \tag{13}$$

with $E[|h(k,l)|^2] = \sigma_h^2 = 1$ [10]. N echoes superpose incoherently. Each path is characterized by a random phase ϕ_n, a random Doppler shift f_{D_n} and a random delay τ_n, where $1 \le n \le N$. The corresponding joint probability density function $p_{\phi,f_D,\tau}(\phi, f_D, \tau)$ is needed to randomly choose ϕ_n, f_{D_n} and τ_n. The symbol duration, T_s, is the spacing in the time domain and the carrier spacing, ΔF, is the spacing in the frequency domain. $f_{D_{max}}T_s$ and $\tau_{max}\Delta F$ are the normalized one-sided channel bandwidths in time and frequency domains, respectively.

Due to the assumption of uncorrelated scattering and for independent phases

$$p_{\phi,f_D,\tau}(\phi, f_D, \tau) = p_\phi(\phi) p_{f_D}(f_D) p_\tau(\tau), \tag{14}$$

where $p_{f_D}(f_D)$ and $p_\tau(\tau)$ are proportional to the so called Doppler power spectrum and delay power spectrum, respectively [10].

Example: Consider a uniform Doppler power spectrum (with $f_{D_{max}}$ being the one-sided maximum Doppler frequency) and a uniform delay power spectrum (with τ_{max} being the one-sided maximum echo delay). Assume the received signal $r(k,l)$ is observed in white Gaussian noise with average SNR $= E_s/N_0$, where E_s is the average energy per symbol (data and pilot symbols are here assumed to be transmitted with the same energy) and N_0 is the one-sided noise spectral density. Then, the cross-covariance is

$$\begin{aligned} \theta_{\Delta t,\Delta f}(k-k'', l-l'') &= \theta_{\Delta t}(k-k'') \cdot \theta_{\Delta f}(l-l'') \\ \theta_{\Delta t}(k-k'') &= \mathrm{si}(2\pi f_{D_{max}} T_s (k-k'')) \\ \theta_{\Delta f}(l-l'') &= \mathrm{si}(2\pi\tau_{max}\Delta F(l-l'')) \end{aligned} \tag{15}$$

and the auto-covariance is

$$\begin{aligned} \Phi(k'-k'', l'-l'') &= \frac{N_0}{E_s}\delta(k'-k'', l'-l'') \\ &+ \theta_{\Delta t}(k'-k'') \cdot \theta_{\Delta f}(l'-l''). \end{aligned} \tag{16}$$

6.2 Pilot-Aided Channel Estimation

Let us denote the spacing between pilot symbols in the time domain by N_K and in the frequency domain by N_L. Given the normalized channel bandwidths $f_{D_{max}}T_s$ and $\tau_{max}\Delta F$, the sampling theorem states that

$$f_{D_{max}}T_s \cdot N_K \le 1/2 \quad \text{and} \quad \tau_{max}\Delta F \cdot N_L \le 1/2. \tag{17}$$

We define a "balanced design" as $f_{D_{max}}T_s \cdot N_K \approx \tau_{max}\Delta F \cdot N_L$. A "rule of thumb" is 2× oversampling to achieve reasonable complexity (filter length) and performance, i.e.,

$$f_{D_{max}}T_s \cdot N_K \approx 1/4 \quad \text{and} \quad \tau_{max}\Delta F \cdot N_L \approx 1/4. \tag{18}$$

Example: Consider $f_{D_{max}} = 320$ Hz (e.g., $v = 192$ km/h @ $f_0 = 1.8$ GHz) and $\tau_{max} = 10\mu$s (one-sided, i.e., the maximum delay is 20μs) as worst-case parameters of a typical land mobile radio system.

Given these constraints, a balanced design with 2x oversampling applied to an OFDM system is as follows:

$$\begin{aligned} f_{D_{max}}T_s N_K &= 320\ \text{Hz} \cdot 260\mu\text{s} \cdot 3 = 1/4 \\ \tau_{max}\Delta F N_L &= 10\mu\text{s} \cdot 4.1667\ \text{kHz} \cdot 6 = 1/4 \end{aligned} \tag{19}$$

where $N_K = 3$, $N_L = 6$, $T = T_s - \Delta = 260\ \mu$s$-20\ \mu$s $= 240\ \mu$s is the useful symbol duration, $\Delta = 2\tau_{max} = 20\ \mu$s is the guard interval, and $1/T = \Delta F = 4.1667$ kHz is the carrier spacing.

This scheme was investigated for short blocks with $K = 58$ and $L = 115$. A rectangular sampling grid with $N_K = 3$ and $N_L = 6$ consists of $N_{grid} = 400$ pilot symbols; the overhead (pilot density) is $1/(N_K N_L)$ excluding edge effects. Other grids were examined as well. Performance criterion is the MSE averaged over all symbols; edge effects are hence included. Perfect knowledge of the covariances is assumed to avoid any spurious effects. Analytical results were verified by simulations.

1-D channel estimation serves as a benchmark. We looked at filtering in the frequency domain; the overhead is $1/N_L$ excluding edge effects. The results are reported in Fig. 2 given various numbers of coefficients, N_{tap}. About five filter coefficients are sufficient.

The second strategy under investigation is 2× 1-D channel estimation [8]: Filtering in the frequency domain is followed by filtering in the time domain, both using 1-D Wiener filters. The actual order of filtering (time or frequency first) is arbitrary due to linearity. The corresponding results given a rectangular grid are plotted in Fig. 3. The MSE is better than for 1-D filtering, although the total pilot overhead is reduced by a factor of three.

Finally, 2-D channel estimation given the rectangular grid is investigated. 10, 25 and 100 coefficients were chosen: seen from the standpoint of complexity, 10 taps

corresponds to 2 × 5 taps for 2× 1-D, 25 taps corresponds to 5 × 5 taps, and 100 taps serves as a benchmark. The results are reported in Fig. 4. Since $N_{tap} < N_{grid}$ and $\binom{N_{grid}}{N_{tap}}$ is very large, the weighted distance $|k-k'|f_{D_{max}}T_s + |l-l'|\tau_{max}\Delta F$ was chosen as the favored selection criteria, see Section 5. If $N_{tap} = 10 = 2 \cdot 5$, the MSE is similar to the MSE of 2× 1-D filtering for practical SNR. If $N_{tap} = 25 = 5 \cdot 5$, the MSE of 2-D filtering is always better.

7 CONCLUSIONS

The discrete shift-variant 2-D Wiener filter was derived and analyzed given an arbitrary sampling grid $\mathcal{P}$ and including model mismatch. To reduce complexity, N_{tap} can be made smaller than N_{grid}. If so, the set of observations $\mathcal{T}(k,l)$ should be optimally chosen from $\mathcal{P}$. The MSE was expressed in closed form; the analysis is valid for any FIR filter. A generalized 2-D sampling theorem was proposed.

Application was presented for pilot-symbol-aided channel estimation. 2-D Wiener filtering was compared with 2× 1-D Wiener filtering and 1-D Wiener filtering. Filtering in two dimensions was shown to outperform filtering in just one dimension with respect to overhead and mean-square error performance. However, for a similar computational effort, the performance of 2× 1-D Wiener filtering is similar to 2-D Wiener filtering.

References

[1] R.M. Mersereau and T.C. Speake, "The processing of periodically sampled multidimensional signals," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. ASSP-31, no. 2, pp. 188-194, Feb. 1983.

[2] C.W. Helstrom, "Image restoration by the method of least squares," *J. Opt. Soc. Amer.*, vol. 57, pp. 297-303, Mar. 1967.

[3] M.P. Ekstrom, "Realizable Wiener filtering in two dimensions," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. ASSP-30, no. 2, pp. 31-40, Feb. 1982.

[4] M.L. Moher and J.H. Lodge, "TCMP – A modulation and coding strategy for Rician fading channels," *IEEE J. Sel. Areas Commun.*, vol. 7, pp. 1347-1355, Dec. 1989.

[5] A. Aghamohammadi, H. Meyr, and G. Ascheid, "A new method for phase synchronization and automatic gain control of linearly modulated signals on frequency-flat fading channels," *IEEE Trans. Commun.*, vol. COM-39, pp. 25-29, Jan. 1991.

[6] J.K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Vehicular Techn.*, vol. VT-40, pp. 686-693, Nov. 1991.

[7] S. Fechtel, "Verfahren und Algorithmen der robusten Synchronisation für die Datenübertragung über dispersive Schwundkanäle," Ph.D. thesis, RWTH Aachen, Germany, 1993.

[8] P. Hoeher, "TCM on frequency-selective land-mobile fading channels," in *Proc. 5th Tirrenia Int. Workshop on Dig. Comm.*, Tirrenia, Italy, pp. 317-328, Sept. 1991.

[9] T. Mueller, K. Brueninghaus, and H. Rohling, "Performance of coherent OFDM-CDMA for broadband mobile communications," *Wireless Personal Communications*, Kluver Academic Publishers, vol. 2, pp. 295-305, 1996.

[10] P. Hoeher, "A statistical discrete-time model for the WSSUS multipath channel," *IEEE Trans. Vehicular Techn.*, vol. 41, no. 4, pp. 461-468, Nov. 1992.

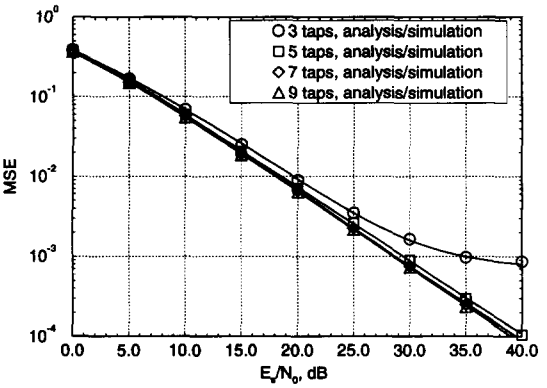

Figure 2: Average MSE versus SNR for 1-D Wiener filtering.

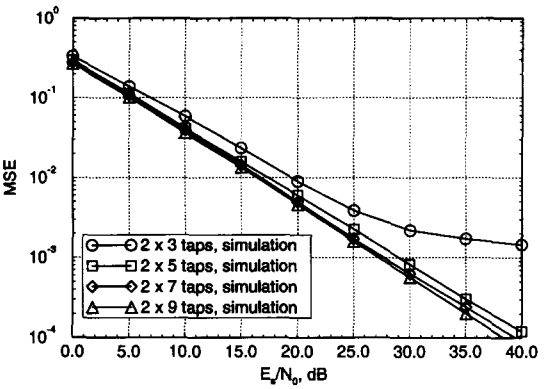

Figure 3: Average MSE versus SNR for 2× 1-D Wiener filtering given a rectangular grid.

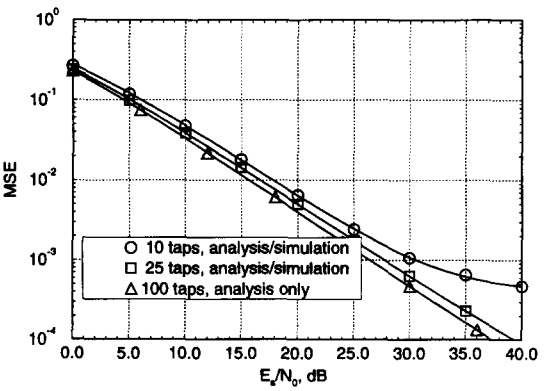

Figure 4: Average MSE versus SNR for 2-D Wiener filtering given a rectangular grid.