Connection between causal and non-causal minimum mean-square errors in optimal linear filtering

Steven R. Weller

School of Electrical Engineering and Computer Science

The University of Newcastle

Callaghan, NSW 2308, Australia.

Email: steven.weller@newcastle.edu.au

Technical Report EE05052[1]
14 December, 2005

A direct proof is given that for optimal linear filtering in additive white noise, the causal (filtering) minimum mean-square error (MMSE) is equal to the non-causal (smoothing) MMSE averaged over the signal-to-noise ratio.

Introduction: The optimal linear, minimum mean-square error (MMSE) estimation of stationary, continuous-time signals from noise-corrupted observations is a classical signal processing problem, whose solution dates back to the work of Wiener in the 1940s [1]. As shown by Wiener, a lower bound on the MMSE achievable by any realisable (namely, causal) linear filter is obtained by considering the associated unrealisable (non-causal) filter, in which access to noisy measurements at all times t satisfying $-\infty < t < \infty$ is assumed. It is then a very straight-

[1]This paper is a preprint of a paper submitted to *Electronics Letters* and is subject to IEE Copyright. If accepted, the copy of record will be available at IEE Digital Library.

forward matter to derive a closed-form expression for the resulting so-called smoothing MMSE, denoted P_{s}, directly in terms of the spectral densities of the signal and additive noise; see, e.g., equation (6) for the case of white noise with power spectral density N_0.

For the corresponding causal linear filtering problem, obtaining closed-form expressions for the filtering MMSE, denoted P_{f}, proves not to be so straightforward. For certain special classes of problems, however, formulas are known which express the filtering MMSE directly in terms of the signal and noise spectra, circumventing the need to explicitly calculate the impulse response of the optimal filter by the Wiener-Hopf spectral factorization technique. The first result in this area was established by Yovits and Jackson [2] for the case of a rational signal spectrum and additive white measurement noise; see equation (5) and [3, pp. 498–501].

A natural problem arising at this point is the quantification of the comparative advantage of (non-causal) smoothing over (causal) filtering in terms of mean-square estimation error reduction, and indeed problems of this kind have been of long-standing interest [4]. Van Trees compared the behaviour of P_{s} and P_{f} in the case of message spectral densities being inverse Butterworth polynomials, or from a Gaussian family [3, pp. 501–505]. Subsequently, Anderson and Chirarattananon derived a universal lower bound on the ratio $P_{\mathrm{s}}/P_{\mathrm{f}}$ in [5]; see also [6].

In very recent work, Guo, Shamai and Verdú [7] established a striking fundamental connection between filtering and smoothing MMSEs. This result is shown to hold under quite general conditions, and asserts that for least-squares filtering in additive white noise, the filtering MMSE is equal to the smoothing MMSE averaged over the signal-to-noise ratio. The Guo-Shamai-Verdú result has an appealing graphical interpretation as follows. In Figure 1 is shown the filtering and smoothing MMSEs as a function of signal-to-noise ratio $\gamma = 1/N_0$, obtained from equations (5) and (6), for a particular choice of message spectral density $S(\omega)$, the details of which will be presented shortly. The Guo-Shamai-Verdú result states that, for any signal-to-noise ratio $\Gamma \geq 0$, the areas of the shaded and cross-hatched regions are equal or, equivalently, that

$$P_{\mathrm{f}}(\Gamma) = \frac{1}{\Gamma} \int_0^{\Gamma} P_{\mathrm{s}}(\gamma)\, d\gamma. \tag{1}$$

As already noted, the Guo-Shamai-Verdú theorem is shown in [7] to hold under quite general conditions, including even a large class of nonlinear filtering–smoothing problems. The proof of (1) in [7] is information-theoretic, making use of input–output mutual information between continuous-time stochastic processes, an incremental Gaussian channel device introduced in [7], and Duncan's theorem [8].

In this Letter, a direct proof of (1) is given for the important special case of linear estimation and additive white measurement noise, thereby establishing a relationship between the now-classical expressions (5) and (6) by means of a non-information-theoretic proof.

Optimal linear filtering and smoothing: Let $s(t)$ be a zero-mean, wide-sense stationary scalar continuous-time random process defined on $-\infty < t < \infty$, with (two-sided) power spectral density $S(\omega) = \int_{-\infty}^{\infty} R(\tau) e^{-j\omega\tau} d\tau$, where $R(\tau) = E\left\{s(t+\tau)s(t)\right\}$ is the signal covariance. The power spectral density $S(\omega)$ is assumed Lebesgue integrable, but not necessarily rational.

Noisy measurements $z(t) = s(t) + n(t)$ are available for $-\infty < t < \infty$, where the noise process $n(t)$ is zero-mean, wide-sense stationary, independent of $s(t)$ and white, with covariance $E\left\{n(t+\tau)n(t)\right\} = N_0 \delta(\tau)$ and power spectral density N_0. Without loss of generality, it is assumed that the average power of the signal process $E\left\{s^2(t)\right\} = R(0) = 1$, and the resulting signal-to-noise power ratio $1/N_0$ is denoted by γ.

In this Letter we consider the least-squares performance of the conditional mean filtered and smoothed estimates, respectively defined as

$$\hat{s}_{\mathrm{f}}(t) = E\{s(t) \mid z(\tau), -\infty < \tau \leq t\}, \tag{2}$$

$$\hat{s}_{\mathrm{s}}(t) = E\{s(t) \mid z(\tau), -\infty < \tau < \infty\}. \tag{3}$$

The precise forms of the filter and smoother are not of immediate relevance in this Letter. Rather, we are interested in the minimum mean-square errors, defined for the filter and smoother respectively as:

$$P_{\mathrm{f}} = E\{(s(t) - \hat{s}_{\mathrm{f}}(t))^2\}, \qquad P_{\mathrm{s}} = E\{(s(t) - \hat{s}_{\mathrm{s}}(t))^2\}. \tag{4}$$

For the signal model above, closed-form expressions for P_{s} and P_{f} are well known; see [2, 9] for filtering, and [1, p. 496] for smoothing:

$$P_{\mathrm{f}}(\gamma) = \frac{1}{2\pi\gamma} \int_{-\infty}^{\infty} \log\left(1 + \gamma S(\omega)\right) d\omega, \tag{5}$$

$$P_{\mathrm{s}}(\gamma) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{S(\omega)}{1+\gamma S(\omega)} d\omega. \tag{6}$$

Connecting optimal linear filtering and smoothing: To prove (1) directly, we proceed as follows:

$$\begin{aligned}
\frac{1}{\Gamma}\int_0^{\Gamma} P_{\mathrm{s}}(\gamma) d\gamma &= \frac{1}{\Gamma}\int_0^{\Gamma} \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{S(\omega)}{1+\gamma S(\omega)}\, d\omega\, d\gamma & (7)\\
&= \frac{1}{2\pi\Gamma}\int_{-\infty}^{\infty}\int_0^{\Gamma} \frac{S(\omega)}{1+\gamma S(\omega)}\, d\gamma\, d\omega & (8)\\
&= \frac{1}{2\pi\Gamma}\int_{-\infty}^{\infty}\int_0^{\Gamma} \frac{f'(\gamma,\omega)}{f(\gamma,\omega)}\, d\gamma\, d\omega, & (9)
\end{aligned}$$

where $f(\gamma, \omega) = 1 + \gamma S(\omega)$, and the indicated derivative in (9) is taken with respect to γ. By evaluating the inner integral we have:

$$\begin{aligned}
\frac{1}{\Gamma}\int_0^{\Gamma} P_{\mathrm{s}}(\gamma) d\gamma &= \frac{1}{2\pi\Gamma}\int_{-\infty}^{\infty} \left[\log\left(1+\gamma S(\omega)\right)\right]_{\gamma=0}^{\gamma=\Gamma}\, d\omega & (10)\\
&= \frac{1}{2\pi\Gamma}\int_{-\infty}^{\infty} \log\left(1+\Gamma S(\omega)\right)\, d\omega & (11)\\
&= P_{\mathrm{f}}(\Gamma), & (12)
\end{aligned}$$

and the result is established.

Illustrative example: To illustrate equation (1), consider a class of message processes $s(t)$ whose average power $E\{s^2(t)\} = 1$, and whose spectral densities are inverse Butterworth polynomials of order $2n$ and break-point frequency k rad/s [3, p. 502]:

$$S(\omega) = \frac{2n}{k}\frac{\sin\left(\pi/2n\right)}{1+(\omega/k)^{2n}}. \tag{13}$$

For $n = 2$ and $k = 1$, Figure 1 shows the filtering and smoothing MMSEs $P_{\mathrm{f}}(\gamma)$ and $P_{\mathrm{s}}(\gamma)$, computed by numerical integration of equations (5) and (6), respectively. Also shown are regions whose areas are $\Gamma P_{\mathrm{f}}\left(\Gamma\right)$ (shaded) and $\int_0^{\Gamma} P_{\mathrm{s}}(\gamma)d\gamma$ (cross-hatched) for a particular choice of Γ; by equation (1) these areas are equal for any $\Gamma \geq 0$.

Smoothing MMSE from filtering MMSE: By application of Leibniz' rule to equation (1), it readily follows that

$$P_{\mathrm{s}}(\gamma) = \gamma \frac{dP_{\mathrm{f}}(\gamma)}{d\gamma} + P_{\mathrm{f}}(\gamma), \tag{14}$$

thereby enabling the calculation of the smoothing MMSE from knowledge of the filtering MMSE. Since (14) holds for the class of nonlinear filtering–smoothing problems considered in [7], it may find application in situations where performance assessment of a smoother is desirable prior to obtaining the explicit form of the smoother. Continuing the previous example, Figure 1 shows the smoothing MMSE P_{s} computed by (14), using numerical differentiation of the filtering MMSE P_{f} from (5). Excellent agreement is obtained between P_{s} calculated in this way, and the evaluation of P_{s} found by numerical integration as per equation (6).

Acknowledgements: This work was supported by the Australian Research Council (ARC) under Discovery Project Grant DP0449627 and Linkage Project Grant LP0561092.

References

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

[2] M. C. Yovits and J. L. Jackson, "Linear filter optimization with game theory considerations," in *IRE Natl. Conv. Rec., part 4*, 1955, pp. 193–199.

[3] H. L. van Trees, *Detection, Estimation and Modulation Theory,* Part I. New York: John Wiley and Sons, 1968.

[4] B. D. O. Anderson, "From Wiener to hidden Markov models," *IEEE Control Syst. Mag.*, vol. 19, no. 3, pp. 41–51, June 1999.

[5] B. D. O. Anderson and S. Chirarattananon, "Smoothing as an improvement on filtering: A universal bound," *Electron. Lett.*, vol. 7, no. 18, pp. 524–525, September 1971.

[6] J. B. Moore and K. L. Teo, “Smoothing as an improvement on filtering in high noise,” *Systems & Control Letters*, vol. 8, no. 1, pp. 51–54, October 1986.

[7] D. Guo, S. Shamai (Shitz), and S. Verdú, “Mutual information and minimum mean-square error in Gaussian channels,” *IEEE Trans. Inform. Theory*, vol. 51, no. 4, pp. 1261–1282, April 2005.

[8] T. E. Duncan, “On the calculation of mutual information,” *SIAM J. Appl. Math.*, vol. 19, no. 1, pp. 215–220, July 1970.

[9] A. J. Viterbi, “On the minimum mean square error resulting from linear filtering of stationary signals in white noise,” *IEEE Trans. Inform. Theory*, vol. IT-11, no. 4, pp. 594–595, October 1965.

Figure caption

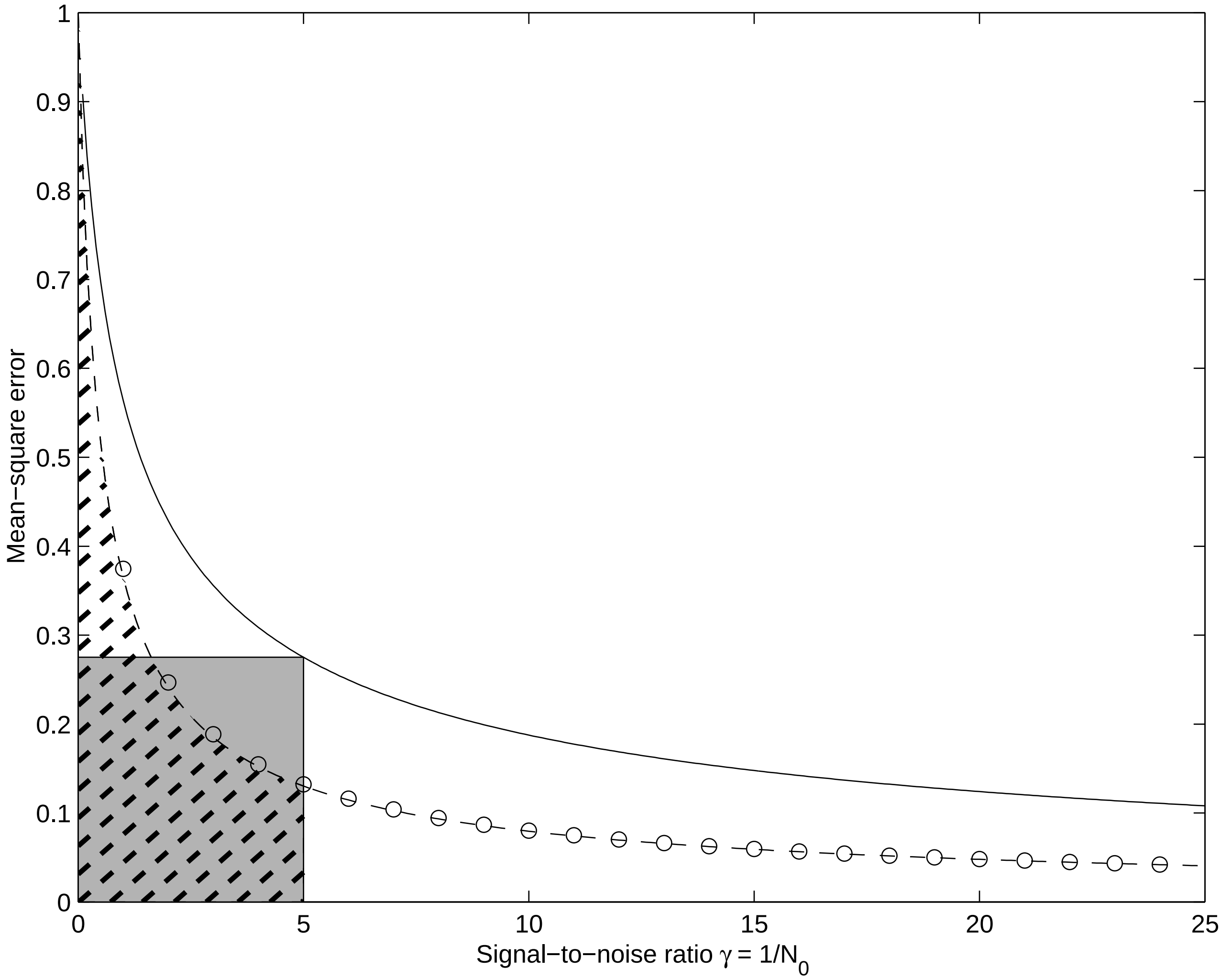
Fig. 1 Filtering and smoothing MMSEs for 4-th order inverse Butterworth polynomial message power spectral density, with break-point frequency 1 rad/s

——— $P_{\rm f}$ using eqn. (5)

- - - - - - - - $P_{\rm s}$ using eqn. (6)

$\circ$ $P_{\rm s}$ using eqn. (14)

Figure 1



1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Mean−square error
0
5
10
15
20
25
Signal−to−noise ratio γ = 1/N0

Figure 1 (unlettered)