Robust Least-Squares Filtering With a Relative Entropy Constraint

Bernard C. Levy and **Ramine Nikoukhah**

ECE Department
University of California
Davis, CA 95616
USA

INRIA
Domaine de Voluceau
Rocquencourt
France

Abstract

We formulate a robust Wiener filtering problem for wide-sense stationary (WSS) Gaussian processes in the presence of modelling errors. It requires solving a minimax problem that consists of finding the best filter for the least-favorable statistical model within a neighborhood of the nominal model. The neighborhood is formed by models whose relative entropy with respect to the nominal model is less than a fixed constant. The standard noncausal Wiener filter is optimal, but the causal Wiener filter is not optimal, and a characterization is provided for the best filter and the corresponding least-favorable model.

1 Introduction

We examine the least-squares filtering problem for a signal based on observations over an infinite interval. When exact joint second order statistics are available for the signal and observations, the signal estimate is generated by a causal or noncausal Wiener filter, depending on whether a causality constraint is imposed on the estimation filter. Unfortunately, in most situations the statistics of the signal and observations are known only imprecisely. In such situations, the estimates produced by an ordinary Wiener filter based on the nominal model may be less accurate than those generated by more conservative filters that take into account the possible existence of modelling errors. In this context, the recent prominence of H^∞ and set membership robust filtering techniques [1, 2] appears to have overshadowed another fruitful approach proposed approximately 20 years ago by Kassam, Poor and their collaborators [3, 4, 5, 6]. In this approach, which was inspired by Huber's pioneering work [7] in robust statistics, the actual joint spectral density of the signal and observations is assumed to belong to a neighborhood of the nominal model. This neighborhood or uncertainty class can be specified in a variety of ways. It can be based on an ϵ-contamination model of the type originally considered by Huber, a total variation model [4, 5], or a spectral band model [3, 6] wherein the power spectral densities (PSDs) specifying the signal and observations are required to stay within a band centered on the nominal PSD. Other classes that have been considered in the literature [6] include p-point models which allocate fixed amounts of power to certain spectral bands and moment-constrained models where some moments are fixed. The robust filtering problem then reduces to the solution of a minimax problem where one seeks to find the best filter for the worst set of statistics in the specified uncertainty class. This approach is straightforward at the conceptual level, but sometimes difficult to implement in practice, since even when the minimax problem is of convex-concave type so that a saddle point exists, it is not always possible to give a simple parametrization to the saddle point.

In this paper, the neighborhood of the nominal model we consider is formed by models whose relative entropy with respect to the nominal model is bounded by a fixed constant. Section 2 considers the robust noncausal least-squares filtering problem for WSS Gaussian processes. In this case, it is shown that the noncausal Wiener filter is robust with respect to the class of perturbations considered. On the other hand, for causal filters, it is shown in Section 3 that the standard Wiener filter is not robust and a characterization is provided for the optimum causal filter and the corresponding least-favorable statistical model. The Section 4 presents some conclusions and provides directions for further research.

2 Noncausal Wiener Filtering

Let $X(t)$ and $Y(t)$ be two jointly stationary scalar Gaussian processes over $\mathbb{Z}$. The nominal and actual Gaussian probability measures P and $\tilde{P}$ of the process

$$Z(t) = \begin{bmatrix} X(t) \\ Y(t) \end{bmatrix}, \tag{2.1}$$

are specified respectively by the mean and autocovariance functions

$$E_P[Z(t)] = m_Z = \begin{bmatrix} m_X \\ m_Y \end{bmatrix} \tag{2.2}$$

$$E_P[(Z(t+k) - m_Z)(Z(t) - m_Z)^T] = K_Z(k) = \begin{bmatrix} K_X(k) & K_{XY}(k) \\ K_{YX}(k) & K_Y(k) \end{bmatrix} \tag{2.3}$$

and

$$E_{\tilde{P}}[Z(t)] = \tilde{m}_Z = \begin{bmatrix} \tilde{m}_X \\ \tilde{m}_Y \end{bmatrix} \tag{2.4}$$

$$E_{\tilde{P}}[(Z(t+k) - m_Z)(Z(t) - m_Z)^T] = \tilde{K}_Z(k) = \begin{bmatrix} \tilde{K}_X(k) & \tilde{K}_{XY}(k) \\ \tilde{K}_{YX}(k) & \tilde{K}_Y(k) \end{bmatrix}. \tag{2.5}$$

Then if

$$R_Z(k) = m_Z m_Z^T + K_Z(k) \tag{2.6}$$

$$\tilde{R}_Z(k) = \tilde{m}_Z \tilde{m}_Z^T + \tilde{K}_Z(k) \tag{2.7}$$

denote respectively the nominal and actual autocorrelation functions of $Z(t)$, the nominal and true power spectral densities of $Z(t)$ can be written as

$$\begin{aligned} S_Z(e^{j\theta}) &= \sum_{k=-\infty}^{\infty} R_Z(k) e^{-jk\theta} \\ &= \begin{bmatrix} S_X(e^{j\theta}) & S_{XY}(e^{j\theta}) \\ S_{YX}(e^{j\theta}) & S_Y(e^{j\theta}) \end{bmatrix}. \end{aligned} \tag{2.8}$$

and

$$\begin{aligned} \tilde{S}_Z(e^{j\theta}) &= \sum_{k=-\infty}^{\infty} \tilde{R}(k) e^{-jk\theta} \\ &= \begin{bmatrix} \tilde{S}_X(e^{j\theta}) & \tilde{S}_{XY}(e^{j\theta}) \\ \tilde{S}_{YX}(e^{j\theta}) & \tilde{S}_Y(e^{j\theta}) \end{bmatrix}. \end{aligned} \tag{2.9}$$

They can be decomposed as

$$S_Z(e^{j\theta}) = 2\pi m_Z m_Z^T \delta(\theta) + \Sigma_Z(e^{j\theta}) \quad (2.10)$$

$$\tilde{S}_Z(e^{j\theta}) = 2\pi \tilde{m}_Z \tilde{m}_Z^T \delta(\theta) + \tilde{\Sigma}_Z(e^{j\theta}) \,, \quad (2.11)$$

where $\delta(.)$ denotes the impulse function, and $\Sigma_Z(e^{j\theta})$ and $\tilde{\Sigma}(e^{j\theta})$ are respectively the discrete-time Fourier transforms of $K_Z(k)$ and $\tilde{K}_Z(k)$. For stationary Gaussian processes, the relative entropy between the true and nominal models takes the form of the Itakura and Saito spectral distortion measure [8]

$$\begin{aligned} I(\tilde{m}_Z, \tilde{\Sigma}_Z; m_Z, \Sigma_Z) &\stackrel{\triangle}{=} \Delta m_Z^T \Sigma_Z^{-1}(1) \Delta m_Z \\ &+ \frac{1}{2\pi} \int_0^{2\pi} [\mathrm{tr}\{(\Sigma_Z^{-1}\tilde{\Sigma}_Z)(e^{j\theta}) - I_2\} - \ln\det(\Sigma_Z^{-1}\tilde{\Sigma}_Z)(e^{j\theta})] \,, \end{aligned} \quad (2.12)$$

where

$$\Delta m_Z \stackrel{\triangle}{=} \tilde{m}_Z - m_Z \,. \quad (2.13)$$

In order to ensure the existence of I, we assume that the nominal spectral density matrix $\Sigma_Z(e^{j\theta})$ is positive definite for all $\theta \in [0, 2\pi]$. Note that since $\ln\det M$ is a concave function of positive definite matrices M [9, pp. 501–502], I is a convex functional of $\tilde{m}_Z$ and $\tilde{\Sigma}_Z(\cdot)$.

We consider the noncausal Wiener filtering problem where the goal is to estimate $X(t)$ given the observations $\{Y(s), s \in \mathbb{Z}\}$. Let

$$\hat{X}(t) = \sum_{k=-\infty}^{\infty} G_k Y(t-k) + h \quad (2.14)$$

denote the estimate of $X(t)$ produced by the pair (G, h) formed by the filter

$$G(z) = \sum_{k=-\infty}^{\infty} G_k z^{-k} \quad (2.15)$$

and the additive constant h. We assume that the filter $G(z)$ is BIBO stable, which ensures the existence of the discrete-time Fourier transform $G(e^{j\theta})$. The estimation error is given by

$$\begin{aligned} E(t) &= X(t) - \hat{X}(t) \\ &= \sum_{k=-\infty}^{\infty} \begin{bmatrix} \delta_k & -G_k \end{bmatrix} Z(t-k) - h \,, \end{aligned} \quad (2.16)$$

so that the mean and variance of the estimation error under the probability measure $\tilde{P}$ can be expressed as

$$\tilde{m}_E = E_{\tilde{P}}[E(t)] = \begin{bmatrix} 1 & -G(1) \end{bmatrix} \tilde{m}_Z - h \quad (2.17)$$

and

$$\begin{aligned} \tilde{K}_E &= E_{\tilde{P}}[(E(t) - \tilde{m}_E)^2] \\ &= \frac{1}{2\pi} \int_0^{2\pi} \begin{bmatrix} 1 & -G(e^{j\theta}) \end{bmatrix} \tilde{\Sigma}_Z(e^{j\theta}) \begin{bmatrix} 1 \\ -G(e^{-j\theta}) \end{bmatrix} d\theta \,. \end{aligned} \quad (2.18)$$

Consider now the convex ball

$$\mathcal{B} = \{(\tilde{m}_Z, \tilde{\Sigma}_Z) \ : \ I(\tilde{m}_Z, \tilde{\Sigma}_Z; m_Z, \Sigma_Z) \leq c\} \tag{2.19}$$

centered on the nominal statistics (m_Z, Σ_Z). Let also $\mathcal{S}$ be the class of estimators (G, h) such that $G(z)$ is BIBO stable. We seek to solve the minimax Wiener filtering problem

$$\min_{(G,h)\in\mathcal{S}} \max_{(\tilde{m}_Z,\tilde{\Sigma}_Z)\in\mathcal{B}} J(\tilde{m}_Z, \tilde{\Sigma}_Z; G, h) \tag{2.20}$$

where

$$J(\tilde{m}_Z, \tilde{\Sigma}_Z; G, h) \stackrel{\triangle}{=} E_{\tilde{P}}[E^2(t)] = \tilde{m}_E^2 + \tilde{K}_E \, . \tag{2.21}$$

The solution of the minimax problem is obtained by exhibiting a saddle point for the objective function J [10, Chap. 6].

Theorem 2.1 *The function $J(\tilde{m}_Z, \tilde{\Sigma}_Z; G, h)$ admits a saddle point $((\tilde{m}_0, \tilde{\Sigma}_0)), (G_0, h_0)) \in \mathcal{B} \times \mathcal{S}$. The estimator*

$$G_0(z) = \Sigma_{XY}(z)\Sigma_Y^{-1}(z) \quad , \quad h_0 = m_X - G_0(1)m_Y \tag{2.22}$$

is the optimal noncausal Wiener filter for both the nominal and least-favorable statistics of $Z(t)$, and the least-favorable statistics have the structure

$$\tilde{m}_0 = m_Z \quad , \quad \tilde{\Sigma}_0(z) = \begin{bmatrix} \tilde{\Sigma}_X(z) & \Sigma_{XY}(z) \\ \Sigma_{YX}(z) & \Sigma_Y(z) \end{bmatrix} , \tag{2.23}$$

so that only $\Sigma_X(z)$ is perturbed. If

$$\tilde{S}_E(z) = \tilde{\Sigma}_X(z) - \Sigma_{XY}(z)\Sigma_Y^{-1}(z)\Sigma_{YX}(z) \tag{2.24}$$

denotes the power spectrum of the error for the least favorable statistics and

$$S_E(z) = \Sigma_X(z) - \Sigma_{XY}(z)\Sigma_Y^{-1}(z)\Sigma_{YX}(z) \tag{2.25}$$

represents the error power spectrum for the nominal statistics, $\tilde{S}_E(z)$ can be expressed in terms of $S_E(z)$ as

$$\tilde{S}_E(z) = [S_E^{-1}(z) - \lambda^{-1}]^{-1} \tag{2.26}$$

where there exists a unique Lagrange multiplier

$$\lambda > r(S_E) \stackrel{\triangle}{=} \max_{\theta\in[0,2\pi]} S_E(e^{j\theta}) \tag{2.27}$$

such that $I(\tilde{m}_0, \tilde{\Sigma}_0, m_Z, \Sigma_Z) = c$.

Proof: We need to show that

$$J(\tilde{m}_Z, \tilde{\Sigma}_Z; G_0, h_0) \leq J(\tilde{m}_0, \tilde{\Sigma}_0; G_0, h_0) \leq J(\tilde{m}_0, \tilde{\Sigma}_0; G, h) \tag{2.28}$$

for all $(\tilde{m}_Z, \tilde{\Sigma}_Z) \in \mathcal{B}$ and $(G, h) \in \mathcal{S}$. The second inequality is just a consequence of the fact that (G_0, h_0) is the optimum noncausal Wiener filter for the pair $(\tilde{m}_0, \tilde{\Sigma}_0)$.

To verify that $(\tilde{m}_0, \tilde{\Sigma}_0)$ maximizes $J(\tilde{m}_Z, \tilde{\Sigma}_Z; G_0, h_0)$ in the ball $\mathcal{B}$, we form the Lagrangian

$$L(\tilde{m}_Z, \tilde{\Sigma}_Z, \lambda) \stackrel{\triangle}{=} J(\tilde{m}_Z, \tilde{\Sigma}_Z; G_0, h_0) + \lambda(c - I(\tilde{m}_Z, \tilde{\Sigma}_Z; m_Z, \Sigma_Z)) \,. \tag{2.29}$$

The first order Gateaux derivatives of L with respect to $\tilde{m}_Z$ in the direction of $u \in \mathbb{R}^2$ and with respect to $\tilde{\Sigma}(z)$ in the direction of a para-symmetric 2×2 matrix function $V(z)$ are given respectively by

$$\nabla_{\tilde{m}_Z,u} L = 2\Delta m_Z^T \Big(\begin{bmatrix} I_n \\ -G_0(1) \end{bmatrix} \begin{bmatrix} I_n & -G_0(1) \end{bmatrix} - \lambda \Sigma_Z^{-1}(1) \Big) u \tag{2.30}$$

$$\begin{aligned} \nabla_{\tilde{\Sigma}_Z,V} L \;&=\; \frac{1}{2\pi} \int_0^{2\pi} \mathrm{tr} \Big\{ \Big(\begin{bmatrix} I_n \\ -G_0(e^{-j\theta}) \end{bmatrix} \begin{bmatrix} I_n & -G_0(e^{j\theta}) \end{bmatrix} \\ &\quad -\lambda(\Sigma_Z^{-1}(e^{j\theta}) - \tilde{\Sigma}_Z^{-1}(e^{j\theta})) \Big) V(e^{j\theta}) \Big\} d\theta \,. \end{aligned} \tag{2.31}$$

Recall that a matrix function $V(z)$ is said to be para-symmetric if it admits the symmetry $V(z) = V^{\#}(z)$ with

$$V^{\#}(z) \stackrel{\triangle}{=} V^T(z^{-1}) \,. \tag{2.32}$$

We have also

$$(\nabla_{\tilde{m}_Z,u})^2 L = 2u^T \Big(\begin{bmatrix} I_n \\ -G_0(1) \end{bmatrix} \begin{bmatrix} I_n & -G_0(1) \end{bmatrix} - \lambda \Sigma_Z^{-1}(1) \Big) u \tag{2.33}$$

$$\nabla_{\tilde{m}_Z,u} \nabla_{\tilde{\Sigma}_Z,V} L = 0 \tag{2.34}$$

$$\begin{aligned} (\nabla_{\tilde{\Sigma}_Z,V})^2 L \;&=\; -\frac{\lambda}{2\pi} \int_0^{2\pi} \mathrm{tr}\{(\tilde{\Sigma}_Z^{-1} V \tilde{\Sigma}_Z^{-1} V)(e^{j\theta})\} d\theta \\ &=\; -\frac{\lambda}{2\pi} \int_0^{2\pi} \mathrm{tr}\{(\tilde{F} V \tilde{\Sigma}_Z^{-1} V \tilde{F}^{\#})(e^{j\theta})\} d\theta \end{aligned} \tag{2.35}$$

where $\tilde{F}(z)$ is an arbitrary matrix spectral factor of $\tilde{\Sigma}_Z^{-1}(z)$, i.e.,

$$\tilde{\Sigma}_Z^{-1}(z) = \tilde{F}^{\#}(z) \tilde{F}(z) \,. \tag{2.36}$$

Thus, provided that the Lagrange multiplier λ is such that the matrix

$$W = \begin{bmatrix} I_n \\ -G_0(1) \end{bmatrix} \begin{bmatrix} I_n & -G_0(1) \end{bmatrix} - \lambda \Sigma_Z^{-1}(1) \tag{2.37}$$

is negative definite, the Hessian

$$H_{\tilde{m}_Z, \tilde{\Sigma}_Z}(u, V) = [\nabla^2_{\tilde{m}_Z,u} + 2\nabla_{\tilde{m}_Z,u} \nabla_{\tilde{\Sigma}_Z,V} + \nabla^2_{\tilde{\Sigma}_Z,V}] L \tag{2.38}$$

will be negative definite for all $\tilde{m}_Z$ and $\tilde{\Sigma}_Z$, so that L is a concave function of $\tilde{m}_Z$ and $\tilde{\Sigma}_Z$.

Let Q be an arbitrary matrix square-root of $\Sigma_Z(1)$, i.e., $\Sigma_Z(1) = QQ^T$. Premultiplying W on the left by Q^T and on the right by Q, and denoting

$$M = \begin{bmatrix} 1 & -G_0(1) \end{bmatrix} Q \,, \tag{2.39}$$

we find that W is congruent to

$$\bar{W} = M^T M - \lambda I_2 \,, \tag{2.40}$$

which will be negative definite provided

$$\lambda > M M^T = S_E(1) \,. \tag{2.41}$$

The Lagrangian $L(\tilde{m}_Z, \tilde{\Sigma}_Z, \lambda)$ is then maximized by setting $\nabla_{\tilde{m}_Z,u} L = 0$ and $\nabla_{\tilde{\Sigma}_Z,V} L = 0$ for all u and $V(z)$, which yields $\Delta m_Z = 0$ and

$$\tilde{\Sigma}_0^{-1}(z) = \Sigma_Z^{-1}(z) - \frac{1}{\lambda} \begin{bmatrix} 1 \\ -G_0(z^{-1}) \end{bmatrix} \begin{bmatrix} 1 & -G_0(z) \end{bmatrix} . \tag{2.42}$$

Noting that $\Sigma_Z(z)$ admits the block spectral factorization

$$\Sigma_Z(z) = \begin{bmatrix} 1 & G_0(z) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} S_E(z) & 0 \\ 0 & \Sigma_Y(z) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ G_0(z^{-1}) & 1 \end{bmatrix} , \tag{2.43}$$

the identity (2.42) implies that $\tilde{\Sigma}_0(z)$ admits the factorization

$$\tilde{\Sigma}_0(z) = \begin{bmatrix} 1 & G_0(z) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \tilde{S}_E(z) & 0 \\ 0 & \Sigma_Y(z) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ G_0(z^{-1}) & 1 \end{bmatrix} \tag{2.44}$$

where the error spectrum $\tilde{S}_E(z)$ obeys (2.26).

The structure (2.43) and (2.44) of Σ_Z and $\tilde{\Sigma}_0$ implies

$$(\Sigma_Z^{-1}\tilde{\Sigma}_0)(z) = \begin{bmatrix} 1 & 0 \\ -G_0(z^{-1}) & 1 \end{bmatrix} \begin{bmatrix} S_E^{-1}\tilde{S}_E & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ G_0(z^{-1}) & 1 \end{bmatrix} \tag{2.45}$$

so that

$$\mathrm{tr}(\Sigma_Z \tilde{\Sigma}_0 - I_2) = S_E^{-1}\tilde{S}_E - 1 \tag{2.46}$$

$$\det(\Sigma_Z \tilde{\Sigma}_0) = S_E^{-1}\tilde{S}_E \tag{2.47}$$

and thus

$$I(\tilde{m}_0, \tilde{\Sigma}_0; m_Z, \Sigma_Z) = \frac{1}{2\pi} \int_0^{2\pi} [(S_E^{-1}\tilde{S}_E)(e^{j\theta}) - 1 - \ln(S_E^{-1}\tilde{S}_E(e^{j\theta}))] d\theta \,. \tag{2.48}$$

Substituting (2.26), this gives

$$I(\tilde{m}_0, \tilde{\Sigma}_0; m_Z, \Sigma_Z) = \gamma(\lambda) \tag{2.49}$$

with

$$\gamma(\lambda) \triangleq \frac{1}{2\pi} \int_0^{2\pi} \gamma(\lambda, S_E(e^{j\theta})) d\theta \tag{2.50}$$

where

$$\gamma(\lambda, d) \triangleq \frac{\lambda}{\lambda - d} - 1 - \ln(\frac{\lambda}{\lambda - d}) \,. \tag{2.51}$$

Note that the integral on the right hand side of (2.50) converges only if $\lambda > r(S_E)$.

For $\lambda > d > 0$, we have

$$\frac{d}{d\lambda}\gamma(\lambda, d) = -\frac{d^2}{\lambda(\lambda - d)^2} < 0 \tag{2.52}$$

and

$$\lim_{\lambda \to \infty} \gamma(\lambda, d) = 0 \quad , \quad \lim_{\lambda \to d} \gamma(\lambda, d) = +\infty \, , \tag{2.53}$$

so that $\gamma(\lambda, d)$ is a monotone decreasing function of λ. Since $\gamma(\lambda)$ is expressed as an integral of this function, it is itself monotone decreasing with

$$\lim_{\lambda \to \infty} \gamma(\lambda) = 0 \quad , \quad \lim_{\lambda \to r(S_E)} \gamma(\lambda) = \infty \, . \tag{2.54}$$

For any positive c, there exists therefore a unique λ_0 such that $\gamma(\lambda_0) = c$. For this value of the Lagrange multiplier the condition (2.41) is satisfied, so that the saddle point identity (2.28) holds with

$$I(\tilde{m}_0, \tilde{\Sigma}_0, m_Z, \Sigma_Z) = c \, . \tag{2.55}$$

□

Thus, the noncausal Wiener filter is robust with respect to a relative entropy mis-modelling criterion. This robustness property should probably not come as a complete surprise since it is already known that H^2 smoothers are also optimal with respect to the H^∞ criterion [11].

However, the performance of the Wiener filter is affected by model errors. Specifically, the MSE corresponding to the nominal model is given by

$$\mathrm{MSE} = \frac{1}{2\pi} \int_0^{2\pi} S_E(e^{j\theta}) d\theta \tag{2.56}$$

whereas the MSE for the least favorable statistics takes the form

$$\widetilde{\mathrm{MSE}} = \frac{1}{2\pi} \int_0^{2\pi} \tilde{S}_E(e^{j\theta}) d\theta \, . \tag{2.57}$$

Taking into account (2.26), the excess MSE, i.e., the additional mean-square error occasioned by the least-favorable model perturbation, can be expressed as

$$\begin{aligned} \mathrm{MSE}_{\mathrm{exc}} &= \widetilde{\mathrm{MSE}} - \mathrm{MSE} \\ &= \frac{1}{2\pi} \int_0^{2\pi} \frac{S_E^2(e^{j\theta})}{\lambda_0 - S_E(e^{j\theta})} d\theta \end{aligned} \tag{2.58}$$

where λ_0 satisfies $\gamma(\lambda_0) = c$. Then, using the inequality

$$\frac{x^2}{1-x} > \frac{x}{1-x} + \ln(1-x) \tag{2.59}$$

for $0 < x < 1$ with

$$x = S_E(e^{j\theta}) / \lambda_0 \, , \tag{2.60}$$

and taking into account the expression (2.48) for I, one finds (see [12] for details)

$$\mathrm{MSE}_{\mathrm{exc}} > \lambda_0 c \, . \tag{2.61}$$

3 Robust Causal Estimation

Up to this point, we have not imposed any causality constraint on the estimators under consideration. The class $\mathcal{C}$ of causal estimators is formed by the pairs $(G(z),\, h)$, such that $G(z)$ is stable and

$$G_k = 0 \tag{3.1}$$

for $k < 0$, so that $G(z)$ is analytical on and outside the unit circle. Then, consider the minimax problem (2.20)–(2.21) with $\mathcal{S}$ replaced by $\mathcal{C}$. Applying an analysis similar to the one of Section 2, it is not difficult to find that a saddle point $(\tilde{m}_0\ ,\, \tilde{\Sigma}_0)$, $(G_0(z),\, h_0)$ will have the following structure. The least-favorable statistics take the form by

$$\tilde{m}_0 = m_Z \quad , \quad \tilde{\Sigma}_0(z) = \begin{bmatrix} \tilde{\Sigma}_X(z) & \tilde{\Sigma}_{XY}(z) \\ \tilde{\Sigma}_{YX}(z) & \tilde{\Sigma}_Y(z) \end{bmatrix} \tag{3.2}$$

Let

$$\tilde{\Sigma}_Y(z) = \tilde{F}(z^{-1})\tilde{F}(z) \tag{3.3}$$

be a spectral factorization of $\tilde{\Sigma}_Y(z)$ where the filter $\tilde{F}(z)$ has minimum phase. Let also $\{K(z)\}_+$ represent the causal part of an arbitrary filter $K(z)$. Then, the estimator

$$G_0(z) = \{\tilde{\Sigma}_{XY}\tilde{F}^{-1}(z^{-1})\}_+\tilde{F}^{-1}(z) \quad , \quad h_0 = m_X - G_0(1)m_Y \tag{3.4}$$

is the optimal causal Wiener filter for the statistics $(\tilde{m}_0,\, \tilde{\Sigma}_0(z))$, and $\tilde{\Sigma}_0(z)$ satisfies the identity

$$\tilde{\Sigma}_0^{-1}(z) = \Sigma_Z^{-1}(z) - \frac{1}{\lambda}\begin{bmatrix} 1 \\ -G_0(z^{-1}) \end{bmatrix}\begin{bmatrix} 1 & -G_0(z) \end{bmatrix}, \tag{3.5}$$

where the Lagrange multiplier is selected such that (2.55) holds. To see how the above conditions were derived, note that given the least-favorable model $(\tilde{m}_0,\, \tilde{\Sigma}_0)$, the second inequality in (2.28) implies that $(G_0,\, h_0)$ must be the optimal causal Wiener filter for this model. Given the estimator $(G_0,\, h_0)$, the maximization of $J(\tilde{m}_Z, \tilde{\Sigma}_Z; G_0, h_0)$ can be performed by using the approach of Theorem 2.1. This yields $\tilde{m}_0 = m_Z$ and the identity (3.5) for $\tilde{\Sigma}_0(z)$.

To ensure that the filter $(G_0,\, h_0)$ and statistics $(\tilde{m}_0,\, \tilde{\Sigma}_0)$ represent an actual saddle point of $J(\tilde{m}_Z, \tilde{\Sigma}_Z; G, h)$, we need only to prove that after substitution of (3.4) inside (3.5), the resulting equation admits a solution $\tilde{\Sigma}_0(z)$ which is positive definite on the unit circle for a value of λ such that $I(\tilde{m}_0, \tilde{\Sigma}_0, m_Z, \Sigma_Z) = c$. Unfortunately, unlike the causal case, the equation (3.5) does not appear to admit a closed form solution. If $\tilde{S}_E(z)$ denotes the Schur complement of $\tilde{\Sigma}_Y(z)$ inside $\tilde{\Sigma}_0(z)$ and if $S_E(z)$ is the Schur complement of $\Sigma_Y(z)$ inside $\Sigma_Z(z)$, by matching the $(1,1)$ blocks on both sides of (3.5), it is easy to verify that (2.26) still holds. However, the other blocks of $\tilde{\Sigma}_0(z)$ are harder to evaluate.

4 Conclusions

In this paper, we have proposed a methodology for robust filtering that employs the relative entropy as a measure of proximity between two statistical models. By examining the resulting minimax problem, it was shown that the standard noncausal Wiener filter is optimal. However, the causal

Wiener filter is not optimal and a characterization was given for the structure of an optimal filter and the matching least-favorable statistical model. The higher level of difficulty of robust causal filtering problems should not come as a surprise, since previous work on robust Wiener filtering [4, 5, 6] had also this feature.

The report [12] extends the results described here to least-squares filtering problems defined over a finite interval. In this case, the relative entropy is expressed as the Kullback-Leibler (KL) divergence [13]. The formulation outlined here and in [12] provides only a starting point for the investigation of robust filtering from a KL divergence viewpoint. Several topics appear to deserve further investigation. First, it would be of interest to develop numerical techniques, possibly iterative, to solve the coupled spectral equations (3.4)–(3.5) for the optimal causal filter and the associated least-favorable model. Also, up to this point, we have considered unstructured perturbations for the joint statistics of the process to be estimated and the observations. However, there exists situations where it makes sense to consider structured perturbations. For example, given observations of the form

$$Y(t) = X(t) + N(t) \,, \tag{4.1}$$

if the noise process $N(t)$ is independent of $X(t)$, it would be of interest to restrict the class of model perturbations to those that maintain this property. Finally, we have limited our attention to external statistical descriptions of the observations and process to be estimated. It would be of interest to extend our results to situations where an internal state-space model of the system is available.

References

[1] B. Hassibi, A. H. Sayed, and T. Kailath, *Indefinite-Quadratic Estimation and Control.* Philadelphia, PA: Soc. for Indust. Applied Math., 1999.

[2] I. R. Petersen and A. V. Savkin, *Robust Kalman Filtering for Signals and Systems with Large Uncertainties.* Boston, MA: Birkhäuser, 1999.

[3] S. A. Kassam and T. L. Lim, "Robust Wiener filters," *J. Franklin Institute*, vol. 304, pp. 171–185, 1977.

[4] H. V. Poor, "On robust Wiener filtering," *IEEE Trans. Automat. Control*, vol. 25, pp. 531–536, June 1980.

[5] K. S. Vastola and H. V. Poor, "Robust Wiener-Kolmogorov theory," *IEEE Trans. Informat. Theory*, vol. 30, pp. 316–327, Mar. 1984.

[6] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing: a survey," *Proc. IEEE*, vol. 73, pp. 433–481, Mar. 1985.

[7] P. J. Huber, *Robust Statistics.* New York, NY: J. Wiley, 1981.

[8] D. Kazakos and P. Papantoni-Kazakos, "Spectral distance measures between Gaussian processes," *IEEE Trans. Automat. Control*, pp. 950–959, Oct. 1980.

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* J. Wiley, 1991.

[10] I. Ekeland and R. Temam, *Convex Analysis and Variational Problems.* Philadelphia, PA: Soc. Indstrial Applied Math., 1999.

[11] K. N. Nagpal and P. P. Khargonekar, "Filtering and smoothing in an H^∞ setting," *IEEE Trans. Automat. Control*, vol. 36, pp. 152–166, Feb. 1991.

[12] B. C. Levy and R. Nikoukhah, "Robust least-squares estimation with a relative entropy constraint," Tech. Rep., Institut de Recherche en Informatique et Automatique, Rocquencourt, France, May 2002.

[13] S. Kullback, *Information Theory and Statistics.* New York: John Wiley, 1959.