WHEN DO NONLINEAR FILTERS ACHIEVE MAXIMAL ACCURACY?*

RAMON VAN HANDEL†

Abstract. The nonlinear filter for an ergodic signal observed in white noise is said to achieve maximal accuracy if the stationary filtering error vanishes as the signal to noise ratio diverges. We give a general characterization of the maximal accuracy property in terms of various systems theoretic notions. When the signal state space is a finite set explicit necessary and sufficient conditions are obtained, while the linear Gaussian case reduces to a classic result of Kwakernaak and Sivan (1972).

Key words. nonlinear filtering, maximal accuracy, systems theory, small noise limit

AMS subject classifications. 93E11, 60G10, 62M20, 93B07, 94A12

1. Introduction. Let $(X_t)_{t\geq 0}$ be a signal of interest, which we model as an ergodic Markov process. It is often the case that the detection of such a signal is imperfect: only some function of the signal may be directly observable, and the observations are additionally corrupted by additive white noise. That is, one observes in practice the integrated observation process

$$Y_t = \int_0^t h(X_s)\,ds + \kappa\, B_t,$$

where B_t is a Wiener process independent of $(X_t)_{t\geq 0}$, $\kappa > 0$ determines the strength of the corrupting noise, and h is a (possibly nonlinear and noninvertible) function of the signal. When only the imperfect observations $(Y_s)_{s\leq t}$ are available, the exact value of the signal X_t can certainly not be detected with arbitrary precision, even when t is very large (so that we have a long observation history at our disposal).

To improve the accuracy of our detector, we must decrease the strength of the corrupting noise. It is intuitively obvious that as $\kappa \to 0$, we will eventually be able to determine precisely the value of $h(X_t)$. However, when the function h is not invertible (as is the case in many engineering systems of practical interest), this does not necessarily imply that we will be able to determine precisely the value of the signal itself. The optimal estimate of the signal X_t, given the observation history $(Y_s)_{s\leq t}$, is called the nonlinear filter. We say that the filter *achieves maximal accuracy* if, as the noise strength vanishes, the stationary filtering error vanishes also—i.e., if as $t \to \infty$ and $\kappa \to 0$, we are able to determine precisely the value of the signal.

When do nonlinear filters achieve maximal accuracy? In the special linear Gaussian case, where the nonlinear filter reduces to the Kalman-Bucy filter, this question was first posed and (almost) completely resolved in a well known paper of Kwakernaak and Sivan [12]. Somewhat surprisingly, the answer is far from trivial and the proof given by Kwakernaak and Sivan is reasonably involved. In fact, Kwakernaak and Sivan chiefly study the dual deterministic control problem with 'cheap' control. Their proof is not probabilistic in nature, but is based on a delicate analysis of the associated riccati equation.

Very little appears to be known beyond the linear Gaussian case. To the best of our knowledge the only nonlinear result is due to Zeitouni and Dembo [22], who study

*The author is partially supported by the NSF RTG Grant DMS-0739195.

†Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544 (rvan@princeton.edu).

a special class of diffusion signals with nonlinear drift term and linear observations. Their result, however, also reduces to the linear Gaussian case: the key step in the proof is to estimate the filtering error by that of an auxiliary Kalman-Bucy filter.

The purpose of this paper is to investigate the maximal accuracy problem in a general setting. After setting up the problem and introducing the relevant concepts in section 2, we proceed in section 3 to relate the maximal accuracy property of the filter to several systems theoretic notions (theorem 3.1 below). The proof of our main result follows from simple probabilistic arguments. Then, in section 4, we apply our general result to provide a complete characterization of the maximal accuracy property for the case where the signal is a finite state Markov process. The resulting necessary and sufficient condition—observability of the model after time reversal, together with a condition of the graph coloring type—is easily verified, but is surprisingly quite different in nature than the result for linear Gaussian systems.

Finally, in section 5, we revisit the linear Gaussian setting and provide a complete proof of the result of Kwakernaak and Sivan using our general characterization. Though this does not lead to new results, our approach does not use the explicit form of the filtering equations and some parts of the proof are significantly simpler than that of [12]. We believe that our approach takes a little of the mystery out of the result of Kwakernaak and Sivan by placing it within a general probabilistic framework.

Acknowledgment. The problem studied in this paper was posed to me by Prof. Ofer Zeitouni during a visit to the University of Minnesota in October 2008. I am indebted to him for arranging this visit and for our many subsequent discussions on this topic, without which this paper would not have been written.

2. Preliminaries. We suppose that defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ is a stationary Markov process $(X_t)_{t\in\mathbb{R}}$ with càdlàg sample paths in the Polish state space E, and we denote its stationary measure as $\mathbf{P}(X_0 \in A) = \pi(A)$. Moreover, we presume that the probability space supports an n-dimensional two-sided Wiener process $(B_t)_{t\in\mathbb{R}}$ that is independent of $(X_t)_{t\in\mathbb{R}}$. Let us define for every $\kappa > 0$ the $\mathbb{R}^n$-valued observation process Y_t^κ according to the expression

$$Y_t^\kappa = \int_0^t h(X_s)\,ds + \kappa\, B_t,$$

where $h : E \to \mathbb{R}^n$ is a given observation function. X_t is called the signal process, and Y_t is called the observations process. In addition, we introduce the following notation. Let $\tilde{X}_t = X_{-t}$ be the time reversed signal, and note that $\tilde{X}_t$ is again a stationary Markov process under $\mathbf{P}$ with invariant distribution π. We denote

$$\begin{aligned} \mathcal{F}_I^X &= \sigma\{X_s : s \in I\}, & \mathcal{F}_I^{\tilde{X}} &= \sigma\{\tilde{X}_s : s \in I\}, \\ \mathcal{F}_I^{h(X)} &= \sigma\{h(X_s) : s \in I\}, & \mathcal{F}_I^{h(\tilde{X})} &= \sigma\{h(\tilde{X}_s) : s \in I\} \end{aligned}$$

for $I \subset \mathbb{R}$, while

$$\mathcal{F}_{[a,b]}^{Y,\kappa} = \sigma\{Y_s^\kappa - Y_a^\kappa : s \in [a,b]\}, \qquad \mathcal{F}_{[a,\infty[}^{Y,\kappa} = \bigvee_{b>a} \mathcal{F}_{[a,b]}^{Y,\kappa}, \qquad \mathcal{F}_{]-\infty,b]}^{Y,\kappa} = \bigvee_{a<b} \mathcal{F}_{[a,b]}^{Y,\kappa}$$

for $a \le b$. Finally, for any probability measure $\mu \ll \pi$, we define

$$\mathbf{P}^\mu(A) = \mathbf{E}\left(I_A\, \frac{d\mu}{d\pi}(X_0)\right).$$

Then under $\mathbf{P}^\mu$, the process X_t (and similarly $\tilde{X}_t$) is still a Markov process with the same transition probabilities as under $\mathbf{P}$, but with initial measure μ instead of π.

REMARK 2.1. If we are given a transition semigroup for the Markov process $(X_t)_{t\geq 0}$, then we can construct $\mathbf{P}^\mu|_{(X_t,Y_t)_{t\geq 0}}$ even for $\mu \not\ll \pi$. However, the transition semigroup for the time reversed process $\tilde{X}_t$ under $\mathbf{P}$ is defined implicitly only up to π-a.s. equivalence. Therefore, for the time reversed process, we can not unambiguously define $\mathbf{P}^\mu$ for $\mu \not\ll \pi$. We will therefore restrict our attention throughout to probability measures $\mu \ll \pi$, except in remark 2.11 and lemmas 2.12 and 2.13 below where only $(X_t, Y_t)_{t\geq 0}$ under $\mathbf{P}^\mu$ is considered (and not the time reversed part).

The *nonlinear filter* of the signal X_t given the noisy observations Y_s^κ, $0 \leq s \leq t$ is defined as the regular conditional probability $\mathbf{P}(X_t \in \cdot\,|\mathcal{F}^{Y,\kappa}_{[0,t]})$. By construction, the filter minimizes the mean square estimation error $e_t(f,\kappa)$ for every test function $f : E \to \mathbb{R}$ with $\int f^2 d\pi < \infty$: i.e.,

$$e_t(f,\kappa) = \mathbf{E}\left(\Big\{f(X_t) - \mathbf{E}\big(f(X_t)\big|\mathcal{F}^{Y,\kappa}_{[0,t]}\big)\Big\}^2\right)$$

is minimal ($e_t(f,\kappa) \leq \mathbf{E}(\{f(X_t) - \hat{F}\}^2)$ for every $\mathcal{F}^{Y,\kappa}_{[0,t]}$-measurable $\hat{F}$).

LEMMA 2.2. *For every test function f with $\int f^2 d\pi < \infty$ and every noise strength $\kappa \geq 0$, the mean square error $e_t(f,\kappa)$ converges to the stationary error*

$$\lim_{t\to\infty} e_t(f,\kappa) = \mathbf{E}\left(\Big\{f(X_0) - \mathbf{E}\big(f(X_0)\big|\mathcal{F}^{Y,\kappa}_{]-\infty,0]}\big)\Big\}^2\right) := e(f,\kappa).$$

Proof. The result follows directly from the stationarity of $\mathbf{P}$ and the martingale convergence theorem. □

Our interest is in the behavior of $e(f,\kappa)$ in the limit $\kappa \to 0$ where the observation noise vanishes. In particular, we aim to understand when the filter achieves *maximal accuracy*.

DEFINITION 2.3. *The filter is said to achieve* maximal accuracy *if $e(f,\kappa) \to 0$ as $\kappa \to 0$ whenever $\int f^2 d\pi < \infty$, i.e., if the true location of the signal is revealed in the stationary limit when the observation noise is small.*

Our main results relate the maximal accuracy problem to certain structural properties of a systems theoretic flavor. We define these notions presently.

DEFINITION 2.4. *The model is said to be* reconstructible *if*

$$\mu,\nu \ll \pi \quad \textit{and} \quad \mu \neq \nu \qquad \textit{implies} \qquad \mathbf{P}^\mu|_{\mathcal{F}^{h(X)}_{]-\infty,0]}} \neq \mathbf{P}^\nu|_{\mathcal{F}^{h(X)}_{]-\infty,0]}}.$$

DEFINITION 2.5. *The model is said to be* strongly reconstructible *if*

$$\mu,\nu \ll \pi \quad \textit{and} \quad \mu \perp \nu \qquad \textit{implies} \qquad \mathbf{P}^\mu|_{\mathcal{F}^{h(X)}_{]-\infty,0]}} \perp \mathbf{P}^\nu|_{\mathcal{F}^{h(X)}_{]-\infty,0]}}.$$

DEFINITION 2.6. *The model is said to be* invertible *if for every $t \geq s$, the random variable X_t coincides $\mathbf{P}$-a.s. with a $\sigma(X_s) \vee \mathcal{F}^{h(X)}_{[s,t]}$-measurable random variable.*

DEFINITION 2.7. *The model is said to be* stably invertible *if for every t the random variable X_t coincides $\mathbf{P}$-a.s. with a $\mathcal{F}^{h(X)}_{]-\infty,t]}$-measurable random variable.*

Let us finish this section with some remarks on these definitions.

REMARK 2.8. In the setting of deterministic linear systems theory, the notion of reconstructibility dates back to Kalman [8], see also [11]. Our definition 2.4 in

the stochastic setting is close to a similar notion that plays an important role in the realization theory of stationary Gaussian processes [16, 13]. Reconstructibility is essentially the time reversed counterpart of the notion of observability [18], though as discussed above we must restrict to probability measures $\mu, \nu \ll \pi$.

REMARK 2.9. By the stationarity of $\mathbf{P}$, definitions 2.6 and 2.7 can be restricted without loss of generality to the case $t = 0$. Thus invertibility means X_0 can be written as a function of X_s at some previous time s and all the intermediate noiseless observations $h(X_r)$, i.e., $X_0 = F_s[X_s, (h(X_r))_{s \le r \le 0}]$ for any $s < 0$. This idea is well known in the deterministic setting; see, e.g., [14] in the linear case and [7] in the nonlinear case. We think of the inverse F_s as being 'stable' if it becomes independent of X_s as $s \to -\infty$; it therefore makes sense to talk of stable inversion when X_0 can be written as a function $X_0 = F_{-\infty}[(h(X_r))_{r \le 0}]$ of all past noiseless observations.

REMARK 2.10. Suppose that the model is invertible. Then certainly X_0 is $\mathbf{P}$-a.s. $\sigma\{X_u, h(X_r) : u \le s,\ r \le 0\}$-measurable for every $s < 0$. In particular,

$$X_0 \quad \text{is} \quad \mathbf{P}\text{-a.s.} \quad \bigcap_{s \le 0} \Big(\mathcal{F}^X_{]-\infty,s]} \vee \mathcal{F}^{h(X)}_{]-\infty,0]} \Big)\text{-measurable.}$$

Now suppose that the left tail σ-field $\bigcap_{s \le 0} \mathcal{F}^X_{]-\infty,s]}$ is $\mathbf{P}$-trivial, i.e., the signal is ergodic in a weak sense [17]. Then it is tempting to exchange the order of intersection and supremum, as follows:

$$X_0 \text{ is } \mathbf{P}\text{-a.s. } \bigcap_{s \le 0} \Big(\mathcal{F}^X_{]-\infty,s]} \vee \mathcal{F}^{h(X)}_{]-\infty,0]} \Big) \stackrel{?}{=} \left[\bigcap_{s \le 0} \mathcal{F}^X_{]-\infty,s]} \right] \vee \mathcal{F}^{h(X)}_{]-\infty,0]} = \mathcal{F}^{h(X)}_{]-\infty,0]}\text{-measurable.}$$

This would indicate that invertibility plus ergodicity implies stable invertibility. However, the exchange of intersection and supremum is not necessarily permitted, as an illuminating counterexample in [3] shows. This conclusion is therefore invalid. A further discussion of this problem can be found in [20]. In particular, it is evident the the present problem is closely related to the *innovations problem* which is discussed in [20]. Another closely related problem, that of the stability of the nonlinear filter, is discussed in detail in [19]; however, it should be noted that the nondegeneracy assumption made there is manifestly absent in the problems discussed here.

REMARK 2.11. Suppose the signal is not started in the stationary distribution, but in a distribution μ that is not even necessarily absolutely continuous with respect to π. In this setting, the maximal achievable accuracy problem is to determine whether

$$e^\mu_t(f, \kappa) = \mathbf{E}^\mu \left(\Big\{ f(X_t) - \mathbf{E}^\mu\big(f(X_t) \big| \mathcal{F}^{Y,\kappa}_{[0,t]}\big) \Big\}^2 \right)$$

converges to zero as $t \to \infty$, $\kappa \to 0$ (in that order). However, we will presently show that if $\|\mathbf{P}^\mu(X_t \in \cdot\,) - \pi\|_{\mathrm{TV}} \to 0$ as $t \to \infty$, this problem reduces to the stationary problem where $\mu = \pi$. In particular, *when the signal is ergodic, the maximal achievable accuracy problem always reduces to the stationary case* (by ergodic we mean $\|\mathbf{P}^\mu(X_t \in \cdot\,) - \pi\|_{\mathrm{TV}} \to 0$ for all μ). This strongly motivates our choice to study directly the stationary problem in the remainder of this paper.

Let us now make these claims precise in the form of two lemmas. For simplicity, we concentrate on bounded functions, which is not a significant restriction.

LEMMA 2.12. *Suppose that f is a bounded measurable function. Then for any κ*

$$\|\mathbf{P}^\mu(X_t \in \cdot\,) - \pi\|_{\mathrm{TV}} \xrightarrow{t \to \infty} 0 \qquad \textit{implies} \qquad \limsup_{t \to \infty} e^\mu_t(f, \kappa) \le e(f, \kappa).$$

Thus $e(f,\kappa)\to 0$ as $\kappa\to 0$ implies $e_t^\mu(f,\kappa)\to 0$ as $t\to\infty$, $\kappa\to 0$ (in that order).

LEMMA 2.13. *Suppose that $\|\mathbf{P}^\mu(X_t\in\cdot\,)-\pi\|_{\mathrm{TV}}\to 0$ for all μ (i.e., the signal is ergodic). Then $e_t^\mu(f,\kappa)\to e(f,\kappa)$ for all $\kappa>0$, μ, and bounded measurable f.*

Proof of lemmas 2.12 and 2.13. Let P_t be the Markov semigroup of the signal $\mu P_t=\mathbf{P}^\mu(X_t\in\cdot\,)$. We basically follow Kunita [10]. First, by Jensen's inequality

$$\begin{aligned}
e_{t+s}^\mu(f,\kappa)&=\mathbf{E}^\mu(f(X_{t+s})^2)-\mathbf{E}^\mu\left(\mathbf{E}^\mu\left(\mathbf{E}^\mu\big(f(X_{t+s})\big|\mathcal{F}^{Y,\kappa}_{[0,t+s]}\big)^2\middle|\mathcal{F}^{Y,\kappa}_{[s,t+s]}\right)\right)\\
&\le\mathbf{E}^\mu(f(X_{t+s})^2)-\mathbf{E}^\mu\left(\mathbf{E}^\mu\big(f(X_{t+s})\big|\mathcal{F}^{Y,\kappa}_{[s,t+s]}\big)^2\right)\\
&=\mathbf{E}^{\mu P_s}(f(X_t)^2)-\mathbf{E}^{\mu P_s}\left(\mathbf{E}^{\mu P_s}\big(f(X_t)\big|\mathcal{F}^{Y,\kappa}_{[0,t]}\big)^2\right)=e_t^{\mu P_s}(f,\kappa)
\end{aligned}$$

for $0<s<t$. We claim that if $\|\mu P_s-\pi\|_{\mathrm{TV}}\to 0$, then $e_t^{\mu P_s}(f,\kappa)\to e_t(f,\kappa)$ as $s\to\infty$. Indeed, it follows trivially (as f is bounded) that $\mathbf{E}^{\mu P_s}(f(X_t)^2)\to\mathbf{E}(f(X_t)^2)$, while convergence of the second term above is easily established by [4, theorem 3.1]. Thus

$$\begin{aligned}
\limsup_{t\to\infty}e_t^\mu(f,\kappa)&=\limsup_{t\to\infty}\limsup_{s\to\infty}e_{t+s}^\mu(f,\kappa)\\
&\le\limsup_{t\to\infty}\limsup_{s\to\infty}e_t^{\mu P_s}(f,\kappa)=\limsup_{t\to\infty}e_t(f,\kappa)=e(f,\kappa).
\end{aligned}$$

This proves lemma 2.12. For lemma 2.13, note that

$$\begin{aligned}
e_{t+s}^\mu(f,\kappa)&=\mathbf{E}^\mu(f(X_{t+s})^2)-\mathbf{E}^\mu\left(\mathbf{E}^\mu\left(\mathbf{E}^\mu\big(f(X_{t+s})\big|\mathcal{F}^{Y,\kappa}_{[0,t+s]}\vee\mathcal{F}^{X}_{[0,s]}\big)\middle|\mathcal{F}^{Y,\kappa}_{[0,t+s]}\right)^2\right)\\
&\ge\mathbf{E}^\mu(f(X_{t+s})^2)-\mathbf{E}^\mu\left(\mathbf{E}^\mu\big(f(X_{t+s})\big|\mathcal{F}^{Y,\kappa}_{[0,t+s]}\vee\mathcal{F}^{X}_{[0,s]}\big)^2\right)\\
&=\mathbf{E}^\mu(f(X_{t+s})^2)-\mathbf{E}^\mu\left(\mathbf{E}^\mu\left(\mathbf{E}^\mu\big(f(X_{t+s})\big|\mathcal{F}^{Y,\kappa}_{[s,t+s]}\vee\sigma(X_s)\big)^2\middle|\sigma(X_s)\right)\right)\\
&=\mathbf{E}^{\mu P_s}\left[\mathbf{E}^{\delta_{X_0}}(f(X_t)^2)-\mathbf{E}^{\delta_{X_0}}\left(\mathbf{E}^{\delta_{X_0}}\big(f(X_t)\big|\mathcal{F}^{Y,\kappa}_{[0,t]}\big)^2\right)\right]=\mathbf{E}^{\mu P_s}\left[e_t^{\delta_{X_0}}(f,\kappa)\right].
\end{aligned}$$

Thus evidently, we can estimate

$$\liminf_{t\to\infty}e_t^\mu(f,\kappa)=\liminf_{t\to\infty}\liminf_{s\to\infty}e_{t+s}^\mu(f,\kappa)\ge\liminf_{t\to\infty}\mathbf{E}\left[e_t^{\delta_{X_0}}(f,\kappa)\right],$$

and it remains to establish that the latter limit equals $e(f,\kappa)$. But

$$\begin{aligned}
\left|\mathbf{E}\left[e_t^{\delta_{X_0}}(f,\kappa)\right]-e_t(f,\kappa)\right|&=\left|\mathbf{E}\left[\mathbf{E}\big(f(X_t)\big|\mathcal{F}^{Y,\kappa}_{[0,t]}\big)^2-\mathbf{E}^{\delta_{X_0}}\big(f(X_t)\big|\mathcal{F}^{Y,\kappa}_{[0,t]}\big)^2\right]\right|\\
&\le 2\|f\|_\infty\,\mathbf{E}\left[\left|\mathbf{E}\big(f(X_t)\big|\mathcal{F}^{Y,\kappa}_{[0,t]}\big)-\mathbf{E}^{\delta_{X_0}}\big(f(X_t)\big|\mathcal{F}^{Y,\kappa}_{[0,t]}\big)\right|\right]\xrightarrow{t\to\infty}0
\end{aligned}$$

using $\kappa>0$, ergodicity of the signal, and [19, theorem 6.6]. □

We emphasize, in particular, that in the ergodic case the maximal achievable accuracy problem is completely equivalent to the stationary maximal achievable accuracy problem for any initial measure μ. We may therefore concentrate on the stationary case without loss of generality, which we will do from now on. Note, however, that our results below do not themselves require ergodicity.

3. A General Result. The purpose of this section is to prove the following general result, which relates the maximal achievable accuracy problem to the various systems theoretic notions introduced above.

THEOREM 3.1. *The following conditions are equivalent:*

1. *The filter achieves maximal accuracy.*
2. *The filtering model is stably invertible.*
3. *The filtering model is strongly reconstructible.*

Moreover, any of these conditions implies the following:

4. *The filtering model is invertible.*
5. *The filtering model is reconstructible.*

It should be noted that often invertibility and reconstructibility are much easier to verify than stable invertibility or strong reconstructibility. However, our general result only shows that the former are necessary conditions for the filter to achieve maximal accuracy. In the next section, we will show that when the signal state space is a finite set, the filter achieves maximal accuracy if and only if the model is both invertible and reconstructible. This will allow us to give simple necessary and sufficient conditions which can be verified directly in terms of the model coefficients. In general, however, it need not be the case that invertibility and reconstructibility are sufficient for the filter to achieve maximal accuracy, see section 5.1 for a counterexample.

The remainder of this section is devoted to the proof of theorem 3.1.

3.1. Proof of $1 \Leftrightarrow 2$. The key here is that we can characterize precisely the limit of $e(f,\kappa)$ as $\kappa \to 0$.

LEMMA 3.2. *For every test function f with $\int f^2 d\pi < \infty$, we have*

$$e(f,\kappa) \xrightarrow{\kappa\to 0} e(f,0) = \mathbf{E}\left(\Big\{f(X_0) - \mathbf{E}\big(f(X_0)\big|\mathcal{F}^{h(X)}_{]-\infty,0]}\big)\Big\}^2\right) := e(f).$$

Proof. Let $\kappa_\ell \searrow 0$ as $\ell \to \infty$. It suffices to show that $e(f,\kappa_\ell) \to e(f)$ as $\ell \to \infty$ for every such sequence. We will therefore fix an arbitrary sequence $\kappa_\ell \searrow 0$ in the remainder of the proof.

Without loss of generality, we assume that $(\Omega,\mathcal{F},\mathbb{P})$ carries a countable sequence B^ℓ_t of independent n-dimensional Wiener processes (independent of $(X_t)_{t\in\mathbb{R}}$). Define

$$W^r_t = \sum_{\ell=r}^\infty \sqrt{\kappa_\ell^2 - \kappa_{\ell+1}^2}\, B^\ell_t, \qquad Z^r_t = \int_0^t h(X_s)\,ds + W^r_t.$$

Note that it is easily established that the sum in the expression for W^r_t is a.s. convergent uniformly on compact time intervals, and that the limit is a Wiener process with covariance $\kappa_r^2 I$. The process $(X_t, Y^{\kappa_r}_t)_{t\in\mathbb{R}}$ therefore has the same law as $(X_t, Z^r_t)_{t\in\mathbb{R}}$ under $\mathbf{P}$, and in particular (here $\mathcal{F}^{Z,r}_{]-\infty,0]} = \sigma\{Z^r_t : t \le 0\}$)

$$e(f,\kappa_r) = \mathbf{E}\left(\Big\{f(X_0) - \mathbf{E}\big(f(X_0)\big|\mathcal{F}^{Z,r}_{]-\infty,0]}\big)\Big\}^2\right).$$

But by the independence of B^ℓ_t and the signal, evidently

$$\mathbf{E}\big(f(X_0)\big|\mathcal{F}^{Z,r}_{]-\infty,0]}\big) = \mathbf{E}\big(f(X_0)\big|\mathcal{F}^{Z,0}_{]-\infty,0]} \vee \mathcal{F}^{B,0}_{]-\infty,0]} \vee \cdots \vee \mathcal{F}^{B,r-1}_{]-\infty,0]}\big) \quad \mathbf{P}\text{-a.s.},$$

where $\mathcal{F}^{B,\ell}_{]-\infty,0]} = \sigma\{B^\ell_t : t \le 0\}$. Therefore

$$\mathbf{E}\big(f(X_0)\big|\mathcal{F}^{Z,r}_{]-\infty,0]}\big) \xrightarrow{r\to\infty} \mathbf{E}\left(f(X_0)\,\middle|\,\mathcal{F}^{Z,0}_{]-\infty,0]} \vee \bigvee_{\ell\ge 0} \mathcal{F}^{B,\ell}_{]-\infty,0]}\right) \quad \text{in } L^2(\mathbf{P})$$

by the martingale convergence theorem. But using again independence

$$\mathbf{E}\left(f(X_0)\middle|\mathcal{F}^{Z,0}_{]-\infty,0]}\vee\bigvee_{\ell\geq 0}\mathcal{F}^{B,\ell}_{]-\infty,0]}\right)=$$
$$\mathbf{E}\left(f(X_0)\middle|\mathcal{F}^{h(X)}_{]-\infty,0]}\vee\bigvee_{\ell\geq 0}\mathcal{F}^{B,\ell}_{]-\infty,0]}\right)=\mathbf{E}\big(f(X_0)\big|\mathcal{F}^{h(X)}_{]-\infty,0]}\big)\quad\mathbf{P}\text{-a.s.},$$

and the claim follows directly. □

We can now prove the implications $1\Rightarrow 2$ and $2\Rightarrow 1$ of theorem 3.1.

Proof of theorem 3.1, $1\Rightarrow 2$. Suppose the filter achieves maximal accuracy. Then

$$e(f)=0\quad\Longrightarrow\quad f(X_0)=\mathbf{E}\big(f(X_0)\big|\mathcal{F}^{h(X)}_{]-\infty,0]}\big)\quad\mathbf{P}\text{-a.s.}$$

whenever f is bounded and measurable. As the signal state space E is Polish, it is isomorphic as a measure space to a subset of the interval $[0,1]$. Denote by $\iota:E\to[0,1]$ this isomorphism. Setting $f=\iota$ above, we find that $\iota(X_0)$ coincides $\mathbf{P}$-a.s. with an $\mathcal{F}^{h(X)}_{]-\infty,0]}$-measurable random variable. Therefore so does X_0. □

Proof of theorem 3.1, $2\Rightarrow 1$. Suppose the filtering model is stably invertible. It follows immediately that $e(f)=0$ for every test function f with $\int f^2 d\pi<\infty$. □

3.2. Proof of $2\Leftrightarrow 3$.

Proof of theorem 3.1, $2\Rightarrow 3$. Let $\mu,\nu\ll\pi$ and $\mu\perp\nu$. Define the event $M=d\mu/d\pi(X_0)>0$. If the filtering model is stably invertible, then I_M coincides $\mathbf{P}$-a.s. with I_H for some $H\in\mathcal{F}^{h(X)}_{]-\infty,0]}$. But then $\mathbf{P}^\mu(H)=1$ and $\mathbf{P}^\nu(H)=0$, so the filtering model is strongly reconstructible. □

Proof of theorem 3.1, $3\Rightarrow 2$. We suppose the model is strongly reconstructible. Let $\{A_1,\ldots,A_m\}$ be a partition of E with $\pi(A_i)>0$ for all i. Define $\pi_i(B)=\pi(B\cap A_i)/\pi(A_i)$. Then $\pi_i\ll\pi$ for every i and $\pi_i\perp\pi_j$ for $i\neq j$. By strong reconstructibility, we may therefore find disjoint $H_1,\ldots,H_m\in\mathcal{F}^{h(X)}_{]-\infty,0]}$ such that $\mathbf{P}^{\pi_i}(H_j)=\delta_{ij}$ for all i,j, or, in other words, $\mathbf{P}(H_i|X_0\in A_i)=1$.

Now let $f(x)=\sum_i f_i I_{A_i}(x)$, and let $H=\sum_i f_i I_{H_i}$ (f_i are distinct). Then

$$\mathbf{P}(f(X_0)=H)=\sum_{i=1}^m\mathbf{P}(f(X_0)=H|X_0\in A_i)\,\mathbf{P}(X_0\in A_i)$$
$$=\sum_{i=1}^m\mathbf{P}(H_i|X_0\in A_i)\,\mathbf{P}(X_0\in A_i)=1.$$

Therefore $f(X_0)$ coincides $\mathbf{P}$-a.s. with the $\mathcal{F}^{h(X)}_{]-\infty,0]}$-measurable random variable H.

Evidently $f(X_0)$ coincides $\mathbf{P}$-a.s. with an $\mathcal{F}^{h(X)}_{]-\infty,0]}$-measurable random variable whenever f is a simple function. But recall that any measurable function can be approximated monotonically by a sequence simple functions, so that the claim evidently holds for any measurable function f. It suffices to note that as the signal state space E is Polish, it is isomorphic as a measure space to a subset of the interval $[0,1]$, so that we may apply our conclusion to the isomorphism ι. □

3.3. Proof of $1 \Rightarrow 4$. The proof of this implication follows from the following observation: it can be read off from the proof of lemmas 2.12 and 2.13 that

$$\mathbf{E}\left(\Big\{f(X_t)-\mathbf{E}\big(f(X_t)\big|\mathcal{F}^{h(X)}_{[0,t]}\vee\sigma(X_0)\big)\Big\}^2\right)\le e_{t+s}(f,0)$$

whenever $\int f^2 d\pi<\infty$ and $t,s\ge 0$. Therefore, if the filter achieves maximal accuracy,

$$\mathbf{E}\left(\Big\{f(X_t)-\mathbf{E}\big(f(X_t)\big|\mathcal{F}^{h(X)}_{[0,t]}\vee\sigma(X_0)\big)\Big\}^2\right)\le \lim_{s\to\infty} e_{t+s}(f,0)=e(f)=0.$$

Thus $f(X_t)$ coincides $\mathbf{P}$-a.s. with a $\mathcal{F}^{h(X)}_{[0,t]}\vee\sigma(X_0)$-measurable random variable. □

3.4. Proof of $3 \Rightarrow 5$. Suppose the model is not reconstructible. We claim that it cannot be strongly reconstructible. Indeed, if the model is not reconstructible, then

$$\text{there exist}\quad \mu,\nu\ll\pi,\quad \mu\neq\nu\quad\text{such that}\quad \mathbf{P}^\mu\big|_{\mathcal{F}^{h(X)}_{]-\infty,0]}}=\mathbf{P}^\nu\big|_{\mathcal{F}^{h(X)}_{]-\infty,0]}}.$$

Define $\mu'=(\mu-\nu)^+/(\mu-\nu)^+(E)$ and $\nu'=(\mu-\nu)^-/(\mu-\nu)^-(E)$. Clearly μ',ν' are probability measures and $\mu'-\nu'\propto\mu-\nu$ (as $(\mu-\nu)^+(E)=(\mu-\nu)^-(E)$), so

$$\mu',\nu'\ll\pi,\quad \mu'\perp\nu',\quad\text{and}\quad \mathbf{P}^{\mu'}\big|_{\mathcal{F}^{h(X)}_{]-\infty,0]}}=\mathbf{P}^{\nu'}\big|_{\mathcal{F}^{h(X)}_{]-\infty,0]}}.$$

Hence the model can certainly not be strongly reconstructible. □

4. Finite State Space. We have shown above that invertibility and reconstructibility are necessary conditions for the filter to achieve maximal accuracy. In this section, we will show that in the case where the signal state space is a finite set, these conditions together are also sufficient. This is particularly useful as invertibility and reconstructibility can be verified algebraically in terms of the model coefficients, while verifying stable invertibility or strong reconstructibility directly is difficult.

Let $(X_t)_{t\in\mathbb{R}}$ be a stationary finite state Markov process. The state space is $E=\{1,\dots,d\}$, the transition law is determined by the $d\times d$ transition intensities matrix $\Lambda=(\lambda_{ij})$, and the stationary measure is the d-dimensional vector $\pi=(\pi_i)$. The observation function is also represented as a d-dimensional vector $h=(h_i)$ (as no confusion may arise, we will make no distinction between functions and measures on E and their representing vectors). We will assume that $\pi_i>0$ for all i.

REMARK 4.1. The assumption that $\pi_i>0$ for all i is made for convenience only and does not entail any loss of generality. Indeed, if any of the entries of the stationary distribution π are zero, then we may remove the corresponding points from the state space E and apply our results below to the resulting stationary Markov process on the reduced state space. Of course, the algebraic conditions in lemmas 4.3 and 4.4 below must then be applied to the coefficients of the reduced model.

The main result in this section is the following.

THEOREM 4.2. *For the finite state filtering model, the following are equivalent:*

1. *The filter achieves maximal accuracy.*
2. *The filtering model is stably invertible.*
3. *The filtering model is strongly reconstructible.*
4. *The filtering model is invertible and reconstructible.*

Clearly, all that remains to be shown is the implication $4\Rightarrow 1$. Before we proceed to the proof, let us show how invertibility and reconstructibility can be verified in terms of the model parameters. To this end we give the following two lemmas.

LEMMA 4.3. *The finite state filtering model is invertible iff the following hold:*

1. *For any $i \neq j$ such that $\lambda_{ij} > 0$, we have $h_i \neq h_j$.*
2. *For any $i \neq j \neq k$ such that $\lambda_{ij} > 0$, $\lambda_{ik} > 0$, we have $h_j \neq h_k$.*

The proof is elementary and is therefore omitted.

LEMMA 4.4. *The finite state filtering model is reconstructible if and only if*

$$\dim\Big(\operatorname{span}\Big\{H^{n_0}\tilde\Lambda H^{n_1}\tilde\Lambda\cdots\tilde\Lambda H^{n_k}\mathbf{1} : k, n_0, \ldots, n_k \geq 0\Big\}\Big) = d.$$

Here $\mathbf{1} = (1 \;\; 1 \;\; \cdots \;\; 1)^*$ is the column vector of ones; $\tilde\Lambda = (\tilde\lambda_{ij})$ is the transition intensities matrix whose off-diagonal entries satisfy $\tilde\lambda_{ij} = \lambda_{ji}\pi_j/\pi_i$; and $H = \operatorname{diag}(h)$.

Proof. It is readily verified that the time reversed signal $\tilde X_t$ is a finite state Markov process with transition intensities matrix $\tilde\Lambda$. As we have assumed without loss of generality that every point of the state space is positively charged by π, any probability measure μ on E is absolutely continuous $\mu \ll \pi$. Therefore reconstructibility in this setting is simply observability of the reverse time system, and the condition in the lemma follows along the lines of [18, lemma 9]. □

The condition in this last lemma can always be computed in a finite number of steps; see [18, remark 11] for further comments and a simple but explicit algorithm.

4.1. Proof of $4 \Rightarrow 1$. As we assume invertibility, we have

$$f(X_t) = \mathbf{E}\big(f(X_t)\big|\mathcal{F}^{h(X)}_{[0,t]} \vee \sigma(X_0)\big) \quad \mathbf{P}\text{-a.s.} \qquad \text{for all functions } f.$$

Therefore, we can evidently write

$$\begin{aligned} e_t(f,0) &= \mathbf{E}\left(\Big\{\mathbf{E}\big(f(X_t)\big|\mathcal{F}^{h(X)}_{[0,t]} \vee \sigma(X_0)\big) - \mathbf{E}\big(f(X_t)\big|\mathcal{F}^{h(X)}_{[0,t]}\big)\Big\}^2\right) \\ &= \mathbf{E}\left(\Big\{\mathbf{E}\big(f(\tilde X_0)\big|\mathcal{F}^{h(\tilde X)}_{[0,t]} \vee \sigma(\tilde X_t)\big) - \mathbf{E}\big(f(\tilde X_0)\big|\mathcal{F}^{h(\tilde X)}_{[0,t]}\big)\Big\}^2\right). \end{aligned}$$

We would like to show that $e(f) = 0$, i.e., that $e_t(f,0) \to 0$ as $t \to \infty$ for any f. As

$$e_t(\alpha f + \beta g, 0) \leq 2\alpha^2\, e_t(f,0) + 2\beta^2\, e_t(g,0),$$

it clearly suffices to restrict to positive $f > 0$ such that $\int f d\pi = 1$. Fix such a function, and define the probability measure $d\nu = f d\pi$. Then

$$\begin{aligned} e_t(f,0) &\leq 2\,\|f\|_\infty\, \mathbf{E}\left(\Big|\mathbf{E}\big(f(\tilde X_0)\big|\mathcal{F}^{h(\tilde X)}_{[0,t]} \vee \sigma(\tilde X_t)\big) - \mathbf{E}\big(f(\tilde X_0)\big|\mathcal{F}^{h(\tilde X)}_{[0,t]}\big)\Big|\right) \\ &= 2\,\|f\|_\infty\, \mathbf{E}^\nu\left(\mathbf{E}\left(\left|\frac{\mathbf{E}\big(f(\tilde X_0)\big|\mathcal{F}^{h(\tilde X)}_{[0,t]} \vee \sigma(\tilde X_t)\big)}{\mathbf{E}\big(f(\tilde X_0)\big|\mathcal{F}^{h(\tilde X)}_{[0,t]}\big)} - 1\right| \,\middle|\, \mathcal{F}^{h(\tilde X)}_{[0,t]}\right)\right) \\ &= 2\,\|f\|_\infty\, \mathbf{E}^\nu\big(\big\|\mathbf{P}^\nu(\tilde X_t \in \cdot\,|\mathcal{F}^{h(\tilde X)}_{[0,t]}) - \mathbf{P}(\tilde X_t \in \cdot\,|\mathcal{F}^{h(\tilde X)}_{[0,t]})\big\|_{\mathrm{TV}}\big) \\ &= 2\,\|f\|_\infty \sum_{i=1}^d \mathbf{E}^\nu\big(\big|\mathbf{P}^\nu(\tilde X_t = i|\mathcal{F}^{h(\tilde X)}_{[0,t]}) - \mathbf{P}(\tilde X_t = i|\mathcal{F}^{h(\tilde X)}_{[0,t]})\big|\big), \end{aligned}$$

where we used the Bayes formula as in [19, lemma 5.6 and corollary 5.7] to show that

$$\begin{aligned} \big\|\mathbf{P}^\nu(\tilde X_t \in \cdot\,|\mathcal{F}^{h(\tilde X)}_{[0,t]}) - \mathbf{P}(\tilde X_t \in \cdot\,|\mathcal{F}^{h(\tilde X)}_{[0,t]})\big\|_{\mathrm{TV}} = \\ \mathbf{E}\left(\left|\frac{\mathbf{E}\big(f(\tilde X_0)\big|\mathcal{F}^{h(\tilde X)}_{[0,t]} \vee \sigma(\tilde X_t)\big)}{\mathbf{E}\big(f(\tilde X_0)\big|\mathcal{F}^{h(\tilde X)}_{[0,t]}\big)} - 1\right| \,\middle|\, \mathcal{F}^{h(\tilde X)}_{[0,t]}\right). \end{aligned}$$

But by reconstructibility and [18, corollary 1], we find that

$$e_t(f,0) \le 2\,\|f\|_\infty \sum_{i=1}^d \mathbf{E}^\nu\big(\big|\mathbf{P}^\nu(\tilde X_t = i|\mathcal{F}^{h(\tilde X)}_{[0,t]}) - \mathbf{P}(\tilde X_t = i|\mathcal{F}^{h(\tilde X)}_{[0,t]})\big|\big) \xrightarrow{t\to\infty} 0.$$

Therefore $e_t(f,0) \to 0$ as $t \to \infty$ for any f, and the proof is complete. □

REMARK 4.5. It is interesting to note that all the steps in this proof have counterparts in the general setting of section 2. In particular, it is not difficult to establish that in general, to achieve maximal accuracy it is sufficient that the model is invertible and that the time-reversed noiseless filter is stable in the sense that

$$\mathbf{E}^\nu\big(\big\|\mathbf{P}^\nu(\tilde X_t \in \cdot\,|\mathcal{F}^{h(\tilde X)}_{[0,t]}) - \mathbf{P}(\tilde X_t \in \cdot\,|\mathcal{F}^{h(\tilde X)}_{[0,t]})\big\|_{\mathrm{TV}}\big) \xrightarrow{t\to\infty} 0 \quad \forall\,\nu \ll \pi,\ \|\tfrac{d\nu}{d\pi}\|_\infty < \infty.$$

On the other hand, the results in [18] are easily adapted to show that if the model is reconstructible, then the time-reversed noiseless filter is stable in the sense that

$$\mathbf{E}^\nu\big(\big|\mathbf{E}^\nu(g(\tilde X_t)|\mathcal{F}^{h(\tilde X)}_{[0,t]}) - \mathbf{E}(g(\tilde X_t)|\mathcal{F}^{h(\tilde X)}_{[0,t]})\big|\big) \xrightarrow{t\to\infty} 0 \quad \forall\,\nu \ll \pi,\ \|\tfrac{d\nu}{d\pi}\|_\infty < \infty,\ g \in L^2(\pi).$$

Therefore, invertibility and reconstructibility imply maximal accuracy if one can close the gap between total variation stability and individual stability of the time-reversed noiseless filter. This is automatic in a finite state space (trivial) and in a countable state space (as the sequence space ℓ_1 has the Schur property [1, theorem 4.32]). When the signal state space is continuous, however, invertibility and reconstructibility is typically not sufficient to guarantee that the filter achieves maximal accuracy; a counterexample is given in the next section.

5. Linear Gaussian Models. In this section, we consider a linear Gaussian model of the following form ($t \in \mathbb{R}$):

$$X_t = X_0 + \int_0^t A X_u\,du + D\,W_t,$$
$$Y^\kappa_t = \int_0^t H X_u\,du + \kappa\,B_t.$$

Here $(B_t)_{t\in\mathbb{R}}$ and $(W_t)_{t\in\mathbb{R}}$ are independent two-sided Wiener processes of dimensions n and m, respectively, and the signal state space is $E = \mathbb{R}^p$.

We make the following assumptions:

1. $A \in \mathbb{R}^{p\times p}$ is a stable matrix and $(X_t)_{t\in\mathbb{R}}$ is stationary;
2. $D \in \mathbb{R}^{p\times m}$ and $H \in \mathbb{R}^{n\times p}$ are matrices of full rank and $m, n \le p$.

The stability of A ensures that the signal is ergodic, and in particular that the stationary solution of the signal equation exists and is unique.

REMARK 5.1. The rank assumption on D and H and the assumption on the dimensions m, n, p is made for convenience only and does not entail any loss of generality. Indeed, when the matrices are not of full rank we can trivially obtain an equivalent model of full rank by reducing the dimensions of W, B and/or Y^κ. Similarly, if $m > p$ we can obtain an equivalent model with $m = p$. Of course, the algebraic condition in theorem 5.2 must then be applied to the coefficients of the reduced model. If $n > p$ the filter is trivially seen to achieve maximal accuracy (as then H, being of full rank, has a left inverse, so the noiseless observations are fully informative).

The maximal achievable accuracy problem in the linear Gaussian setting was considered in a classic paper of Kwakernaak and Sivan [12], where an (almost) necessary

and sufficient condition was obtained. Their approach is surprisingly complicated, however, and relies on rather explicit computations of the behavior of Riccati equations in the limit of vanishing noise. In this section we give a direct proof of their theorem by verifying the stable invertibility property.

THEOREM 5.2. *In the linear Gaussian setting of this section, the filter achieves maximal accuracy if and only if the matrix* $H(\lambda I - A)^{-1}D$ *has linearly independent columns for all* $\lambda \in \mathbb{C}$ *with* $\operatorname{Re}\lambda > 0$.

The result of Kwakernaak and Sivan follows easily.

COROLLARY 5.3 (Kwakernaak-Sivan). *The following hold.*

1. *If* $m > n$, *then the filter does not achieve maximal accuracy.*
2. *If* $m = n$, *then the filter achieves maximal accuracy if and only if* $\det[H(\lambda I - A)^{-1}D]$ *is nonzero for any* $\lambda \in \mathbb{C}$ *with* $\operatorname{Re}\lambda > 0$.
3. *If* $m < n$ *and there exists* $M \in \mathbb{R}^{m\times n}$ *such that* $\det[MH(\lambda I - A)^{-1}D]$ *is nonzero for any* $\lambda \in \mathbb{C}$ *with* $\operatorname{Re}\lambda > 0$, *then the filter achieves maximal accuracy.*

Proof. We consider each case separately.

1. $H(\lambda I - A)^{-1}D$ has m columns each of which is an n-dimensional vector. Therefore if $m > n$, the columns cannot be linearly independent for any λ.

2. When $m = n$ the matrix $H(\lambda I - A)^{-1}D$ is square, so that it has linearly independent columns if and only if $\det[H(\lambda I - A)^{-1}D] \neq 0$.

3. If $\det[MH(\lambda I - A)^{-1}D]$ is nonzero, the square matrix $MH(\lambda I - A)^{-1}D$ has linearly independent columns. Then certainly the columns of $H(\lambda I - A)^{-1}D$ are linearly independent.

In view of these facts, the corollary follows by applying theorem 5.2. ∎

REMARK 5.4. As is pointed out by Godbole [6], the condition of theorem 3.1 corresponds to the requirement that the model is invertible and that the inverse has no unstable modes (in the sense of linear systems). Indeed, note that $H(\lambda I - A)^{-1}D$ is the transfer function associated to our filtering model, so that invertibility holds in this setting if and only if the matrix $H(\lambda I - A)^{-1}D$ has a left inverse for all but a finite number of $\lambda \in \mathbb{C}$ (see, e.g., [14, theorem 5]). The inverse is again a linear system whose transfer function is the left inverse of $H(\lambda I - A)^{-1}D$, so that the lack of right halfplane zeros of $H(\lambda I - A)^{-1}D$ ensures that the inverse system does not have any unstable poles. If there are additionally no zeros on the imaginary axis, then the inverse system is even asymptotically stable and the heuristic outlined in remark 2.9 can be rigorously implemented.

However, it is not immediately obvious from such arguments that stable invertibility follows even when there are zeros on the imaginary axis, or that the model cannot be stably invertible when there are right halfplane zeros. In the proof of theorem 5.2, the former problem is circumvented by using the idea in [12] of using an approximate, rather than exact, inverse system. The latter problem is easily resolved directly in our setting, and the proof of this part of the theorem is substantially simpler than the corresponding arguments of Kwakernaak and Sivan.

5.1. An example. Before we proceed to the proof of theorem 5.2, let us demonstrate by means of an example that, unlike in the finite state setting, invertibility and reconstructibility are not always sufficient to ensure that the filter achieves maximal accuracy. This implies the existence of a gap between total variation and individual stability of the time reversed filter, discussed in remark 4.5.

For our example, let $m = n = 1$ and $p = 2$, and we set

$$A = \begin{bmatrix} -1 & 0 \\ 0 & -4 \end{bmatrix}, \qquad D = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad H = [1 \quad -2].$$

This model satisfies all the assumptions of this section. We now compute

$$H(\lambda I - A)^{-1}D = [1 \quad -2] \begin{bmatrix} (\lambda+1)^{-1} & 0 \\ 0 & (\lambda+4)^{-1} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = (\lambda+1)^{-1} - 2(\lambda+4)^{-1}.$$

Therefore $H(\lambda I - A)^{-1}D = 0$ for $\lambda = 2$, so by theorem 5.2 the filter does not achieve maximal accuracy. However, we claim that the model is both invertible and reconstructible.

To prove invertibility, it suffices to note that $H(\lambda I - A)^{-1}D$ is nonzero (hence left invertible) for all but a finite number of $\lambda \in \mathbb{C}$. To prove reconstructibility, we use the fact that the reverse time signal $\tilde{X}_t$ satisfies an equation of the form [2]

$$\tilde{X}_t = \tilde{X}_0 + \int_0^t \Sigma A^* \Sigma^{-1} \tilde{X}_s \, ds + D\,\tilde{W}_t,$$

where $\tilde{W}_t$ is a suitably defined Wiener process and Σ denotes the stationary covariance matrix of the signal. The matrix Σ can be computed as the unique solution of the Lyapunov equation:

$$A\Sigma + \Sigma A^* + DD^* = 0 \qquad \Longrightarrow \qquad \Sigma = \begin{bmatrix} 1/2 & 1/5 \\ 1/5 & 1/8 \end{bmatrix}.$$

Note that Σ is a strictly positive matrix. Therefore, the model is evidently reconstructible if $H\Sigma$ and $H\Sigma A^*$ are linearly independent (so that the time reversed model is observable). But this is easily established to be the case by explicit computation.

5.2. Proof of Theorem 5.2. We will show that the condition of the theorem is necessary and sufficient for stable invertibility. The necessity part of the proof is closely related to a problem of Karhunen [9], while sufficiency is proved along the lines of [12].

In the following, let us write $Z_t = HX_t$ ($t \in \mathbb{R}$) for notational simplicity. We introduce the Hilbert spaces of random variables $\mathcal{L}_X, \mathcal{L}_Z, \mathcal{L}_W \subset L^2(\mathbf{P})$ as follows:

$$\begin{aligned}
\mathcal{L}_X &= L^2(\mathbf{P})\text{-cl}\,\{v_1^* X_{t_1} + \cdots + v_k^* X_{t_k} : k \in \mathbb{N},\ t_1, \ldots, t_k \le 0,\ v_1, \ldots, v_k \in \mathbb{R}^p\}, \\
\mathcal{L}_Z &= L^2(\mathbf{P})\text{-cl}\,\{v_1^* Z_{t_1} + \cdots + v_k^* Z_{t_k} : k \in \mathbb{N},\ t_1, \ldots, t_k \le 0,\ v_1, \ldots, v_k \in \mathbb{R}^n\}, \\
\mathcal{L}_W &= L^2(\mathbf{P})\text{-cl}\,\{v_1^* W_{t_1} + \cdots + v_k^* W_{t_k} : k \in \mathbb{N},\ t_1, \ldots, t_k \le 0,\ v_1, \ldots, v_k \in \mathbb{R}^m\}.
\end{aligned}$$

For an $\mathbb{R}^k$-valued random variable K we will write, e.g., $K \in \mathcal{L}_X$ when $v^*K \in \mathcal{L}_X$ for every $v \in \mathbb{R}^k$.

As the joint process $(X_t, Z_t, W_t)_{t\in\mathbb{R}}$ is Gaussian, the stable invertibility problem is essentially linear and can be reduced to the investigation of the spaces $\mathcal{L}_X, \mathcal{L}_Z, \mathcal{L}_W$.

LEMMA 5.5. *The model is stably invertible if and only if $\mathcal{L}_Z = \mathcal{L}_W$.*

Proof. By definition, the model is stably invertible iff X_0 coincides $\mathbf{P}$-a.s. with a $\sigma\{Z_s : s \le 0\}$-measurable random variable, i.e., iff $\mathbf{E}(X_0|Z_{]-\infty,0]}) = X_0$. However, as $(X_0, Z_s : s \le 0)$ is Gaussian, it is well known that $\mathbf{E}(X_0|Z_{]-\infty,0]}) \in \mathcal{L}_Z$. The model is therefore stably invertible iff $X_0 \in \mathcal{L}_Z$.

To proceed, note that

$$X_t = \int_{-\infty}^t e^{A(t-s)} D\, dW_s, \qquad Z_t = \int_{-\infty}^t H e^{A(t-s)} D\, dW_s \qquad (t \in \mathbb{R})$$

as A is stable. Therefore clearly $X_0 \in \mathcal{L}_W$ and $\mathcal{L}_Z \subset \mathcal{L}_W$. If $\mathcal{L}_W = \mathcal{L}_Z$, it then follows immediately that $X_0 \in \mathcal{L}_Z$. It therefore remains to show that $X_0 \in \mathcal{L}_Z$

implies $\mathcal{L}_W \subset \mathcal{L}_Z$. To this end, assume $X_0 \in \mathcal{L}_Z$. By stationarity, $X_t \in \mathcal{L}_Z$ also for $t \le 0$. Therefore evidently $\mathcal{L}_X \subset \mathcal{L}_Z$. But

$$DW_s = X_s - X_0 + \int_s^0 AX_u\,du, \qquad s \le 0.$$

As D has full rank, we find that $\mathcal{L}_W \subset \mathcal{L}_X$. Therefore $\mathcal{L}_W \subset \mathcal{L}_Z$ as required. □

We can now complete the proof of theorem 5.2.

Proof of theorem 5.2. Suppose $H(\lambda I - A)^{-1}D$ does not have linearly independent columns for some $\lambda \in \mathbb{C}$ with $\operatorname{Re}\lambda > 0$. Then there exists $0 \neq w \in \mathbb{C}^m$ such that $H(\lambda I - A)^{-1}Dw = 0$, and we define

$$U = U_1 + iU_2 := \int_{-\infty}^0 (e^{\lambda s}w)^* dW_s, \qquad U_1, U_2 \in \mathcal{L}_W.$$

We can now compute

$$\mathbf{E}(U^* v^* Z_u) = \int_{-\infty}^u e^{\lambda s}\, v^* H e^{A(u-s)} Dw\,ds = e^{\lambda u}\, v^* H(\lambda I - A)^{-1} Dw = 0$$

for all $u \le 0$, $v \in \mathbb{R}^n$. In particular, as $v^* Z_u$ is real-valued, $\langle U_1, v^* Z_u\rangle_{L^2(\mathbf{P})} = \langle U_2, v^* Z_u\rangle_{L^2(\mathbf{P})} = 0$ for $u \le 0$, $v \in \mathbb{R}^n$, so that $U_1, U_2 \perp \mathcal{L}_Z$. But as U is nonzero, evidently $\mathcal{L}_Z \neq \mathcal{L}_W$ and the model is not stably invertible.

Conversely, suppose that the matrix $H(\lambda I - A)^{-1}D$ has linearly independent columns for all $\lambda \in \mathbb{C}$ with $\operatorname{Re}\lambda > 0$. We will prove that for any $\varepsilon > 0$, there is a random variable X_0^ε of the form

$$X_0^\varepsilon = \int_{-\infty}^0 m_\varepsilon(s)\, Z_s\,ds \in \mathcal{L}_Z$$

such that $\|X_0^\varepsilon - X_0\|_{L^2(\mathbf{P})} < \varepsilon$. Then $X_0 \in \mathcal{L}_Z$ (as $\mathcal{L}_Z$ is closed), so the model is stably invertible.

To prove the claim, fix $\varepsilon > 0$. Then

$$X_0^\varepsilon = \int_{-\infty}^0 m_\varepsilon(s) \int_{-\infty}^0 I_{s\ge u} H e^{A(s-u)} D\,dW_u\,ds = \int_{-\infty}^0 \int_u^0 m_\varepsilon(s)\, H e^{A(s-u)} D\,ds\,dW_u,$$

provided that m_ε is bounded and the function

$$T_\varepsilon(u) := \int_u^0 m_\varepsilon(s)\, H e^{A(s-u)} D\,ds$$

is square integrable (this is justified by truncating the lower bounds on the integrals and applying Fubini's theorem for stochastic integrals [15, theorem IV.64]). Note that

$$\|X_0^\varepsilon - X_0\|^2_{L^2(\mathbf{P})} = \int_{-\infty}^0 \|T_\varepsilon(u) - e^{-Au}D\|_F^2\,du,$$

where $\|C\|_F^2 = \operatorname{Tr}[CC^*]$ is the Frobenius norm. Define for $\operatorname{Re}\lambda > 0$ the Laplace transforms

$$\hat m_\varepsilon(\lambda) = \int_{-\infty}^0 e^{\lambda s}\, m_\varepsilon(s)\,ds, \qquad \hat T_\varepsilon(\lambda) = \int_{-\infty}^0 e^{\lambda s}\, T_\varepsilon(s)\,ds = \hat m_\varepsilon(\lambda)\, H(\lambda I - A)^{-1} D.$$

By Plancherel's theorem, we can write (see, e.g., [21, pp. 162–163])

$$\|X_0^\varepsilon - X_0\|_{L^2(\mathbf{P})}^2 = \frac{1}{2\pi} \lim_{x \searrow 0} \int_{-\infty}^{\infty} \|\hat{m}_\varepsilon(x+iy)\, H(\{x+iy\}I - A)^{-1}D - (\{x+iy\}I - A)^{-1}D\|_F^2 \, dy.$$

It remains to choose m_ε with the required properties such that this expression is smaller than ε^2.

By our assumption, the left inverse $V(\lambda)$ of the matrix $H(\lambda I - A)^{-1}D$ is defined on the right halfplane, i.e., $V(\lambda)H(\lambda I - A)^{-1}D = I$ for $\operatorname{Re}\lambda > 0$. The above expression for $\|X_0^\varepsilon - X_0\|_{L^2(\mathbf{P})}$ is therefore identically zero if we choose $\hat{m}_\varepsilon(\lambda)$ as $(\lambda I - A)^{-1}DV(\lambda)$. The problem is that the latter may not be the Laplace transform of a function m_ε with the required properties. We therefore regularize as follows:

$$\hat{m}_\varepsilon(\lambda) = \frac{\gamma^\ell}{(\lambda+\gamma)^\ell}\,(\lambda I - A)^{-1}D\,V(\lambda)$$

for some $\gamma > 0$, $\ell \in \mathbb{N}$ to be chosen presently. As $\lambda \mapsto H(\lambda I - A)^{-1}D$ is a rational function, $\hat{m}_\varepsilon$ is rational also. We choose ℓ sufficiently large that the degree of the denominator is larger than the degree of the numerator. Then $\hat{m}_\varepsilon$ is strictly proper with poles in the closed left halfplane $\operatorname{Re}\lambda \le 0$ only, and is therefore the Laplace transform of some bounded function m_ε. Moreover, as

$$\sup_{x>0} \int_{-\infty}^{\infty} \|\hat{T}_\varepsilon(x+iy)\|_F^2 \, dy = \sup_{x>0} \int_{-\infty}^{\infty} \left|\frac{\gamma^\ell}{(x+iy+\gamma)^\ell}\right|^2 \|(\{x+iy\}I - A)^{-1}D\|_F^2 \, dy \le 2\pi \, \|X_0\|_{L^2(\mathbf{P})}^2,$$

the function T_ε is square integrable by the Paley-Wiener theorem. Finally, as

$$\|X_0^\varepsilon - X_0\|_{L^2(\mathbf{P})}^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left|\frac{\gamma^\ell}{(iy+\gamma)^\ell} - 1\right|^2 \|(iyI - A)^{-1}D\|_F^2 \, dy \xrightarrow{\gamma\to\infty} 0$$

by dominated convergence, we may choose γ such that $\|X_0^\varepsilon - X_0\|_{L^2(\mathbf{P})} < \varepsilon$. □

5.3. The unstable case. To be fair, it should be noted that we have not entirely reproduced the result of Kwakernaak and Sivan as we have assumed that the matrix A is stable. The result of Kwakernaak and Sivan requires only that the filtering model is detectable and stabilizable. Unfortunately, our approach relies crucially on the stationarity (or ergodicity) of the signal process, so that one could never expect to obtain general results in the setting where the signal may be transient. On the other hand, in the linear Gaussian case, the special structure of the model allows us to reduce the detectable/stabilizable case to the stationary case considered above. We therefore recover the result of Kwakernaak and Sivan in its entirety.

We develop the relevant argument presently. Let us emphasize, however, that the following argument is very specific to the linear Gaussian setting.

We consider again a linear Gaussian model of the form

$$X_t = X_0 + \int_0^t AX_u \, du + D\,W_t,$$
$$Y_t^\kappa = \int_0^t HX_u \, du + \kappa\, B_t,$$

where $(B_t)_{t\in\mathbb{R}}$ and $(W_t)_{t\in\mathbb{R}}$ are independent two-sided Wiener processes of dimensions n and m, respectively, and the signal state space is $E = \mathbb{R}^p$. As above, we will assume $D \in \mathbb{R}^{p\times m}$ and $H \in \mathbb{R}^{n\times p}$ are matrices of full rank and $m, n \le p$. We do not assume, however, that $A \in \mathbb{R}^{p\times p}$ is stable; instead, we assume only that (A, D) is stabilizable and (A, H) is detectable. The law of X_0 may be chosen arbitrarily.

As (A, H) is detectable, it is well known that there exists a matrix $K \in \mathbb{R}^{p\times n}$ such that $\bar{A} := A - KH$ is a stable matrix. Fix such a matrix K (it may not be unique, but this will not affect our final result). Using Itô's rule, we compute

$$e^{-\bar{A}t}X_t = X_0 + \int_0^t e^{-\bar{A}s}KHX_s\,ds + \int_0^t e^{-\bar{A}s}D\,dW_s = X_0 + \int_0^t e^{-\bar{A}s}K\,dY_s^\kappa - \kappa\int_0^t e^{-\bar{A}s}K\,dB_s + \int_0^t e^{-\bar{A}s}D\,dW_s.$$

Now define

$$\bar{X}_t^\kappa := X_t - \int_0^t e^{\bar{A}(t-s)}K\,dY_s^\kappa, \qquad \bar{Y}_t^\kappa := \int_0^t H\bar{X}_u^\kappa\,du + \kappa\,B_t.$$

Then evidently $\bar{X}_t^\kappa$ satisfies the stochastic differential equation

$$\bar{X}_t^\kappa = X_0 + \int_0^t \bar{A}\bar{X}_s^\kappa\,ds + D\,W_t - \kappa\,K\,B_t.$$

Moreover, we can compute

$$e^{-At}\bar{X}_t^\kappa = X_0 + \int_0^t e^{-As}D\,dW_s - \int_0^t e^{-As}KH\bar{X}_s^\kappa\,ds - \kappa\int_0^t e^{-As}K\,dB_s = X_0 + \int_0^t e^{-As}D\,dW_s - \int_0^t e^{-As}K\,d\bar{Y}_s^\kappa.$$

Thus evidently

$$X_t = \bar{X}_t^\kappa + \int_0^t e^{A(t-s)}K\,d\bar{Y}_s^\kappa.$$

The following lemma is therefore immediate.

LEMMA 5.6. *$\sigma\{Y_t^\kappa : t \in [0,T]\} = \sigma\{\bar{Y}_t^\kappa : t \in [0,T]\}$ $\mathbf{P}$-a.s. $\forall\ T \le \infty$, $\kappa \ge 0$.*

Proof. This follows directly from

$$\bar{Y}_t^\kappa = Y_t^\kappa - \int_0^t\int_0^s He^{\bar{A}(s-u)}K\,dY_u^\kappa\,ds, \qquad Y_t^\kappa = \bar{Y}_t^\kappa + \int_0^t\int_0^s He^{A(s-u)}K\,d\bar{Y}_u^\kappa\,ds.$$

The proof is complete. □

We now see immediately that for every $t, \kappa \ge 0$

$$\mathbf{E}\left(\left\|X_t - \mathbf{E}(X_t|\mathcal{F}_{[0,t]}^{Y,\kappa})\right\|^2\right) = \mathbf{E}\left(\left\|\bar{X}_t^\kappa - \mathbf{E}(\bar{X}_t^\kappa|\mathcal{F}_{[0,t]}^{\bar{Y},\kappa})\right\|^2\right),$$

where $\mathcal{F}_{[0,t]}^{\bar{Y},\kappa}$ is defined in the obvious fashion. But $\bar{X}_t^\kappa$ is an ergodic Markov process (as $\bar{A}$ is stable), which brings us back—in principle—to the setting employed throughout this paper. However, note that the driving noise of $\bar{X}_t^\kappa$ is correlated with the observation noise, so that we can not immediately apply our previous results.

LEMMA 5.7. *Define for $t \in \mathbb{R}$*

$$X_t^0 := \int_{-\infty}^t e^{\bar{A}(t-s)}D\,dW_s.$$

Then we have

$$\lim_{\kappa\to 0}\lim_{t\to\infty}\mathbf{E}\left(\left\|X_t-\mathbf{E}(X_t|\mathcal{F}^{Y,\kappa}_{[0,t]})\right\|^2\right)=\mathbf{E}\left(\left\|X_0^0-\mathbf{E}(X_0^0|\sigma\{HX_s^0:s\le 0\})\right\|^2\right).$$

Proof. Let us define, for $\kappa\ge 0$ and $t\ge 0$, the processes

$$x_t^\kappa=x_0^\kappa+\int_0^t Ax_u^\kappa\,du+D\,W_t,\qquad x_0^\kappa=\int_{-\infty}^0 e^{-\bar{A}s}D\,dW_s-\kappa\int_{-\infty}^0 e^{-\bar{A}s}K\,dB_s,$$

$$y_t^\kappa=\int_0^t Hx_u^\kappa\,du+\kappa\,B_t,\qquad \bar{x}_t^\kappa=x_t^\kappa-\int_0^t e^{\bar{A}(t-s)}K\,dy_s^\kappa,\qquad \bar{y}_t^\kappa=\int_0^t H\bar{x}_u^\kappa\,du+\kappa B_t.$$

Then (x_t^κ,y_t^κ) is a Markov process with the same transition law as (X_t,Y_t^κ), except that we have chosen a specific initial law for x_0^κ in a manner that depends on κ. However, as the model is stabiliziable and detectable, it is well known that the stationary filtering error exists and is independent of the initial law. Therefore

$$\lim_{t\to\infty}\mathbf{E}\left(\left\|X_t-\mathbf{E}(X_t|\mathcal{F}^{Y,\kappa}_{[0,t]})\right\|^2\right)=\lim_{t\to\infty}\mathbf{E}\left(\left\|x_t^\kappa-\mathbf{E}(x_t^\kappa|\mathcal{F}^{y,\kappa}_{[0,t]})\right\|^2\right).$$

From the above discussion, it is now easily seen that in fact also

$$\lim_{t\to\infty}\mathbf{E}\left(\left\|X_t-\mathbf{E}(X_t|\mathcal{F}^{Y,\kappa}_{[0,t]})\right\|^2\right)=\lim_{t\to\infty}\mathbf{E}\left(\left\|\bar{x}_t^\kappa-\mathbf{E}(\bar{x}_t^\kappa|\mathcal{F}^{\bar{y},\kappa}_{[0,t]})\right\|^2\right).$$

Here we have defined $\mathcal{F}^{y,\kappa}_{[0,t]}$ and $\mathcal{F}^{\bar{y},\kappa}_{[0,t]}$ in the obvious fashion.

Now note that $\bar{x}_t^\kappa$ is a stationary Markov process with the explicit representation

$$\bar{x}_t^\kappa=\int_{-\infty}^t e^{\bar{A}(t-s)}D\,dW_s-\kappa\int_{-\infty}^t e^{\bar{A}(t-s)}K\,dB_s.$$

Therefore (x_t^κ,y_t^κ) immediately extend to all $t\in\mathbb{R}$, and by stationarity

$$\lim_{t\to\infty}\mathbf{E}\left(\left\|X_t-\mathbf{E}(X_t|\mathcal{F}^{Y,\kappa}_{[0,t]})\right\|^2\right)=\mathbf{E}\left(\left\|\bar{x}_0^\kappa-\mathbf{E}(\bar{x}_0^\kappa|\mathcal{F}^{\bar{y},\kappa}_{]-\infty,0]})\right\|^2\right).$$

The proof is completed by following the same steps as in the proof of lemma 3.2, and noting that the Wiener process B enters linearly in the expression for $\bar{x}_0^\kappa$. □

COROLLARY 5.8. *The filter achieves maximal accuracy if and only if the matrix* $H(\lambda I-\bar{A})^{-1}D$ *has linearly independent columns for all* $\lambda\in\mathbb{C}$ *with* $\operatorname{Re}\lambda>0$.

Proof. Immediate from theorem 5.2. □

COROLLARY 5.9. *The filter achieves maximal accuracy if and only if the matrix* $H(\lambda I-A)^{-1}D$ *has linearly independent columns for all* $\lambda\in\mathbb{C}$ *with* $\operatorname{Re}\lambda>0$.

Proof. By [5, proposition 2], $H(\lambda I-A)^{-1}D$ has linearly independent columns iff $H(\lambda I-A+KH)^{-1}D$ has linearly independent columns, for any matrix K. □

REFERENCES

[1] C. D. ALIPRANTIS AND O. BURKINSHAW, *Positive operators*, Springer, Dordrecht, 2006.
[2] B. D. O. ANDERSON, *Reverse-time diffusion equation models*, Stochastic Process. Appl., 12 (1982), pp. 313–326.

[3] P. Baxendale, P. Chigansky, and R. Liptser, *Asymptotic stability of the Wonham filter: Ergodic and nonergodic signals*, SIAM J. Control Optim., 43 (2004), pp. 643–669.
[4] I. Crimaldi and L. Pratelli, *Two inequalities for conditional expectations and convergence results for filters*, Statist. Probab. Lett., 74 (2005), pp. 151–162.
[5] E. G. Gilbert, *The decoupling of multivariable systems by state feedback*, SIAM J. Control, 7 (1969), pp. 50–63.
[6] S. S. Godbole, *Comments on: "The maximally achievable accuracy of linear optimal regulators and linear optimal filters" (IEEE Trans. Automatic Control* **ac-17** *(1972), 79–86) by Huibert Kwakernaak and Raphael Sivan*, IEEE Trans. Automatic Control, AC-17 (1972), pp. 577–578. With a reply by H. Kwakernaak and R. S. Sivan.
[7] R. M. Hirschorn, *Invertibility of nonlinear control systems*, SIAM J. Control Optim., 17 (1979), pp. 289–297.
[8] R. E. Kalman, P. L. Falb, and M. A. Arbib, *Topics in mathematical system theory*, McGraw-Hill Book Co., New York, 1969.
[9] K. Karhunen, *Über die Struktur stationärer zufälliger Funktionen*, Ark. Mat., 1 (1950), pp. 141–160.
[10] H. Kunita, *Asymptotic behavior of the nonlinear filtering errors of Markov processes*, J. Multivar. Anal., 1 (1971), pp. 365–393.
[11] H. Kwakernaak and R. Sivan, *Linear optimal control systems*, Wiley-Interscience [John Wiley & Sons], New York, 1972.
[12] ———, *The maximally achievable accuracy of linear optimal regulators and linear optimal filters*, IEEE Trans. Automatic Control, AC-17 (1972), pp. 79–86.
[13] A. Lindquist and G. Picci, *Realization theory for multivariate stationary Gaussian processes*, SIAM J. Control Optim., 23 (1985), pp. 809–857.
[14] P. J. Moylan, *Stable inversion of linear systems*, IEEE Trans. Automatic Control, AC-22 (1977), pp. 74–78.
[15] P. E. Protter, *Stochastic integration and differential equations*, vol. 21 of Applications of Mathematics (New York), Springer-Verlag, Berlin, 2004.
[16] G. Ruckebusch, *On the theory of Markovian representation*, in Measure theory applications to stochastic analysis (Proc. Conf., Res. Inst. Math., Oberwolfach, 1977), vol. 695 of Lecture Notes in Math., Springer, Berlin, 1978, pp. 77–87.
[17] H. Totoki, *On a class of special flows*, Z. Wahrsch. verw. Gebiete, 15 (1970), pp. 157–167.
[18] R. van Handel, *Observability and nonlinear filtering*, Probab. Th. Rel. Fields, (2009). To appear.
[19] ———, *The stability of conditional Markov processes and Markov chains in random environments*, Ann. Probab., (2009). To appear.
[20] H. von Weizsäcker, *Exchanging the order of taking suprema and countable intersections of σ-algebras*, Ann. Inst. H. Poincaré Sect. B (N.S.), 19 (1983), pp. 91–100.
[21] K. Yosida, *Functional analysis*, vol. 123 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, sixth ed., 1980.
[22] O. Zeitouni and A. Dembo, *On the maximal achievable accuracy in nonlinear filtering problems*, IEEE Trans. Automat. Control, 33 (1988), pp. 965–967.