MODEL ROBUSTNESS OF FINITE STATE NONLINEAR FILTERING OVER THE INFINITE TIME HORIZON

PAVEL CHIGANSKY AND RAMON VAN HANDEL

Abstract. We investigate the robustness of nonlinear filtering for continuous time finite state Markov chains, observed in white noise, with respect to misspecification of the model parameters. It is shown that the distance between the optimal filter and that with incorrect model parameters converges to zero uniformly over the infinite time interval as the misspecified model converges to the true model, provided the signal obeys a mixing condition. The filtering error is controlled through the exponential decay of the derivative of the nonlinear filter with respect to its initial condition. We allow simultaneously for misspecification of the initial condition, of the transition rates of the signal, and of the observation function. The first two cases are treated by relatively elementary means, while the latter case requires the use of Skorokhod integrals and tools of anticipative stochastic calculus.

1. Introduction

The theory of nonlinear filtering concerns the estimation of a signal corrupted by white noise, and has diverse applications in target tracking, signal processing, automatic control, etc. The basic setting of the theory involves a Markov signal process, e.g. the solution of a (nonlinear) stochastic differential equation or a finite-state Markov process, observed in independent corrupting noise. The calculation of the resulting filters is a classical topic in stochastic analysis [12]. Of course, the filtering equations will depend explicitly on the model chosen for the signal process and observations; in almost all realistic applications, however, the model that underlies the filter is only an approximation of the true system that generates the observations. In order for the theory to be practically useful, it is important to establish that the filtered estimates are not too sensitive to the choice of underlying model.

Continuity with respect to the model parameters of nonlinear filtering estimates on a fixed *finite* time interval is well established, e.g. [2, 3, 8]; generally speaking, it is known that the error incurred in a finite time interval due to the choice of incorrect model parameters can be made arbitrarily small if the model parameters are chosen sufficiently close to those of the true model. As the corresponding error bounds grow rapidly with the length of the time interval, however, such estimates are of little use if we are interested in robustness of the filter over a long period of

2000 *Mathematics Subject Classification.* Primary 93E11; secondary 93E15, 60H07, 60J27.

Key words and phrases. nonlinear filtering, filter stability, model robustness, error bounds, markov chains, anticipative stochastic calculus.

Research of P.C. was supported by a grant of the Israeli Science Foundation.

Research of R.v.H. was supported by the ARO under Grant DAAD19-03-1-0073. This author thanks P. S. Krishnaprasad of the University of Maryland for hosting a visit to the Institute for Systems Research, during which the this work was initiated.

time. One would like to show that the approximation errors do not accumulate, so that the error remains bounded uniformly over an *infinite* time interval.

The model robustness of nonlinear filters on the infinite time horizon was investigated in discrete time in [5, 10, 11]. The key idea that allows one to control the accumulation of approximation errors is the asymptotic stability property of many nonlinear filters, which is the focus of much recent work (see [1] and the references therein) and can be summarized as follows. The optimal nonlinear filter is a recursive equation that is initialized with the true distribution of the signal process at the initial time. If the filter is initialized with a different distribution, then the resulting filtered estimates are no longer optimal (in the least-squares sense). The filter is called asymptotically stable if the solution of the wrongly initialized filter converges to the solution of the correctly initialized filter at large times; i.e., the filter "forgets" its initial condition after a period of observation.

Using an approximate filter rather than the optimal filter is equivalent to using the optimal filter where we make an approximation error after every time step. Now suppose the optimal filter forgets its initial condition at an exponential rate; then also the approximation error at each time step is forgotten at an exponential rate, and the errors cannot accumulate in time. If the approximation error at each time step is bounded (finite time robustness), then the total approximation error will be bounded uniformly in time. Model robustness on the infinite time horizon is thus a consequence of finite time robustness together with the exponential forgetting property of the filter. This is precisely the method used in [5, 10, 11], and its implementation is fairly straightforward once bounds on the exponential forgetting rate of the filter have been obtained. However, the method used there does not extend to nonlinear filtering in continuous time; even the continuous time model with point process observations studied in [5], though more involved, reduces essentially to discrete (but random) observation times. The continuous time case requires different tools, which we develop in this paper in the setting of nonlinear filtering of a finite-state Markov signal process. (We also mention [6], where a different but related problem is solved).

We consider the following filtering setup. The signal process $X = (X_t)_{t\geq 0}$ is a continuous-time, homogeneous Markov chain with values in the finite alphabet $\mathbb{S} = \{a_1, \dots, a_d\}$, transition intensities matrix $\Lambda = (\lambda_{ij})$ and initial distribution $\nu^i = \mathbf{P}(X_0 = a_i)$. The observation process $Y = (Y_t)_{t\geq 0}$ is given by

$$Y_t = \int_0^t \mathrm{h}(X_s)\, ds + B_t, \tag{1.1}$$

where $\mathrm{h} : \mathbb{S} \to \mathbb{R}$ is the observation function (we will also write $h^i = \mathrm{h}(a_i)$) and B is a Wiener process that is independent of X. The filtering problem for this model concerns the calculation of the conditional probabilities $\pi_t^i = \mathbf{P}(X_t = a_i|\mathscr{F}_t^Y)$ from the observations $\{Y_s : s \leq t\}$, where $\mathscr{F}_t^Y = \sigma\{Y_s : s \leq t\}$. It is well known that π_t satisfies the Wonham equation [18, 12]

$$d\pi_t = \Lambda^* \pi_t\, dt + (H - h^*\pi_t)\pi_t\, (dY_t - h^*\pi_t\, dt), \qquad \pi_0 = \nu, \tag{1.2}$$

where x^* denotes the transpose of x and $H = \operatorname{diag} h$. Note that the Wonham equation is initialized with the true distribution of X_0; we will denote by $\pi_t(\mu)$ the solution of the Wonham equation at time t with an arbitrary initial distribution $\pi_0 = \mu$, and by $\pi_{s,t}(\mu)$ the solution of the Wonham equation at time $t \geq s$ with the initial condition $\pi_s = \mu$. In [1] the exponential forgetting property of the

Wonham filter was established as follows: the ℓ_1-distance $|\pi_t(\mu) - \pi_t(\nu)|$ decays exponentially a.s., provided the initial distributions are equivalent $\mu \sim \nu$ and that the *mixing condition* $\lambda_{ij} > 0\ \forall i \neq j$ is satisfied. Now consider the Wonham filter with incorrect model parameters:

$$d\tilde{\pi}_t = \tilde{\Lambda}^* \tilde{\pi}_t\, dt + (\tilde{H} - \tilde{h}^* \tilde{\pi}_t)\tilde{\pi}_t\, (dY_t - \tilde{h}^* \tilde{\pi}_t\, dt), \qquad \tilde{\pi}_0 = \nu, \tag{1.3}$$

where $\tilde{\Lambda}$ and $\tilde{h}$ denote a transition intensities matrix and observation function that do not match the underlying signal-observation model (X, Y), $\tilde{H} = \operatorname{diag} \tilde{h}$, and we denote by $\tilde{\pi}_t(\mu)$ the solution of this equation with initial condition $\tilde{\pi}_0 = \mu$ and by $\tilde{\pi}_{s,t}(\mu)$ the solution with $\tilde{\pi}_s = \mu$. The following is the main result of this paper.

Theorem 1.1. *Suppose $\nu^i, \mu^i > 0\ \forall i$ and $\lambda_{ij}, \tilde{\lambda}_{ij} > 0\ \forall i \neq j$. Then*

$$\sup_{t\geq 0} \mathbf{E}\|\tilde{\pi}_t(\mu) - \pi_t(\nu)\|^2 \leq C_1\, |\mu - \nu| + C_2\, |\tilde{h} - h| + C_3\, |\tilde{\Lambda}^* - \Lambda^*|,$$

where $|\tilde{\Lambda}^* - \Lambda^*| = \sup\{|(\tilde{\Lambda}^* - \Lambda^*)\tau| : \tau^i > 0\ \forall i,\ |\tau| = 1\}$ and the quantities C_1, C_2, C_3 are bounded on any compact subset of parameters $\{(\nu, \Lambda, h, \mu, \tilde{\Lambda}, \tilde{h}) : \nu^i, \mu^i > 0\ \forall i,\ |\nu| = |\mu| = 1,\ \lambda_{ij}, \tilde{\lambda}_{ij} > 0\ \forall i \neq j,\ \sum_j \lambda_{ij} = \sum_j \tilde{\lambda}_{ij} = 0\ \forall i\}$. Additionally we have the asymptotic estimate

$$\limsup_{t\to\infty} \mathbf{E}\|\tilde{\pi}_t(\mu) - \pi_t(\nu)\|^2 \leq C_2\, |\tilde{h} - h| + C_3\, |\tilde{\Lambda}^* - \Lambda^*|.$$

In particular, this implies that if $\nu^i > 0\ \forall i$, $\lambda_{ij} > 0\ \forall i \neq j$, then

$$\lim_{\tilde{h}\to h} \lim_{\tilde{\Lambda}\to\Lambda} \lim_{\mu\to\nu} \sup_{t\geq 0} \mathbf{E}\|\tilde{\pi}_t(\mu) - \pi_t(\nu)\| = \lim_{\tilde{h}\to h} \lim_{\tilde{\Lambda}\to\Lambda} \limsup_{t\to\infty} \mathbf{E}\|\tilde{\pi}_t(\mu) - \pi_t(\nu)\| = 0.$$

Let us sketch the basic idea of the proof. Rather than considering the Wonham filter, let us demonstrate the idea using the following simple caricature of a filtering equation. Consider a smooth "observation" y_t and a "filter" whose state x_t is propagated by the ordinary differential equation $dx_t/dt = f(x_t, y_t)$. Similarly, we consider the "approximate filter" $d\tilde{x}_t/dt = \tilde{f}(\tilde{x}_t, y_t)$ and assume that everything is sufficiently smooth, so that for fixed y both equations generate a two-parameter flow $x_t = \varphi^y_{s,t}(x_s)$, $\tilde{x}_t = \tilde{\varphi}^y_{s,t}(\tilde{x}_s)$. The following calculation is straightforward:

$$\begin{aligned}\varphi^y_{0,t}(x) - \tilde{\varphi}^y_{0,t}(x) &= \int_0^t \frac{d}{ds}\left[\tilde{\varphi}^y_{s,t}(\varphi^y_{0,s}(x))\right] ds \\ &= \int_0^t D\tilde{\varphi}^y_{s,t}(\varphi^y_{0,s}(x)) \cdot (f(\varphi^y_{0,s}(x), y_s) - \tilde{f}(\varphi^y_{0,s}(x), y_s))\, ds,\end{aligned}$$

where $D\tilde{\varphi}^y_{s,t}(x) \cdot v$ denotes the directional derivative of $\tilde{\varphi}^y_{s,t}(x)$ in the direction v. Using the triangle inequality, we obtain the estimate

$$|\varphi^y_{0,t}(x) - \tilde{\varphi}^y_{0,t}(x)| \leq \int_0^t |D\tilde{\varphi}^y_{s,t}(\varphi^y_{0,s}(x))|\, |f(\varphi^y_{0,s}(x), y_s) - \tilde{f}(\varphi^y_{0,s}(x), y_s)|\, ds.$$

Now suppose that $|f(\cdot,\cdot) - \tilde{f}(\cdot,\cdot)| \leq K$; this is an expression of finite-time robustness, as it ensures that $|\varphi^y_{0,t}(x) - \tilde{\varphi}^y_{0,t}(x)| \leq Kt \to 0$ (for fixed t) as $\tilde{f} \to f$. Suppose furthermore that $|D\tilde{\varphi}^y_{s,t}(\cdot)| \leq Ce^{-\lambda(t-s)}$, i.e. an infinitesimal perturbation to the initial condition is forgotten at an exponential rate. Then the estimate above is uniformly bounded and converges to zero uniformly in time as $\tilde{f} \to f$. Conceptually this is similar to the logic used in discrete time, but we have to replace the exponential forgetting of the initial condition by the requirement that the derivative of the filter with respect to its initial condition decays exponentially.

Returning to the Wonham filter, this procedure can be implemented in a fairly straightforward way if $\tilde{h} = h$. In this case, most of the work involves finding a suitable estimate on the exponential decay of the derivative of the filter with respect to its initial condition; despite the large number of results on filter stability, such estimates are not available in the literature to date. We obtain estimates by adapting methods from [1], together with uniform estimates of the concentration of the optimal filter near the boundary of the simplex.

The general case with $\tilde{h} \neq h$ is significantly more involved. The problem is already visible in the simple demonstration above. Note that the integrand on the right hand side of the error estimate is not adapted; it depends on the observations on the entire interval $[0, t]$. As the Wonham filter is defined in terms of an Itô-type stochastic integral, this will certainly get us into trouble. When $\tilde{h} = h$ the stochastic integral cancels in the error bound and the problems are kept to a minimum; in the general case, however, we are in no such luck. Nonetheless this problem is not prohibitive, but it requires us to use the stochastic calculus for anticipating integrands developed by Nualart and Pardoux [14, 13] using Skorokhod integrals rather than Itô integrals and using Malliavin calculus tools.

The remainder of this paper is organized as follows. In Section 2 we prove some regularity properties of the solution of the Wonham equation. We also demonstrate the error estimate discussed above in the simpler case $\tilde{h} = h$, and comment on the more general applicability of such a bound. In Section 3 we obtain exponential bounds on the derivative of the Wonham filter with respect to its initial condition. Section 4 treats the general case $\tilde{h} \neq h$ using anticipative stochastic calculus; some of the technical estimates appear in Appendix B. Finally, Appendix A contains a brief review of the results from the Malliavin calculus and anticipative stochastic calculus that are needed in the proofs.

Notation. The signal-observation pair (X, Y) is defined on the standard probability space $(\Omega, \mathscr{F}, \mathbf{P})$. The expectation with respect to $\mathbf{P}$ is denoted by $\mathbf{E}$ or sometimes $\mathbf{E_P}$. For $x \in \mathbb{R}^d$, we denote by $|x|$ the ℓ_1-norm, by $\|x\|$ the ℓ_2-norm, and by $\|x\|_p$ the ℓ_p-norm. We write $x \succ y$ (resp. $\prec, \succeq, \preceq$) if $x_i > y_i$ $(<, \geq, \leq)$ $\forall i$.

The following spaces will be used throughout. Probability distributions on $\mathbb{S}$ are elements of the simplex $\Delta^{d-1} = \{x \in \mathbb{R}^d : x \succeq 0,\ |x| = 1\}$. Usually, we will be interested in the interior of the simplex $\mathcal{S}^{d-1} = \{x \in \mathbb{R}^d : x \succ 0,\ |x| = 1\}$. The space of vectors tangent to $\mathcal{S}^{d-1}$ is denoted by $T\mathcal{S}^{d-1} = \{x \in \mathbb{R}^d : \sum_i x_i = 0\}$. Finally, we will denote the positive orthant by $\mathbb{R}^d_{++} = \{x \in \mathbb{R}^d : x \succ 0\}$.

2. Preliminaries

Eq. (1.2) is a nonlinear equation for the conditional distribution π_t. It is well known however (e.g. [7]) that π_t can also be calculated in a linear fashion: $\pi_t = \rho_t/|\rho_t|$, where the unnormalized density ρ_t is propagated by the Zakai equation

$$d\rho_t = \Lambda^* \rho_t \, dt + H \rho_t \, dY_t, \qquad \rho_0 = \nu. \tag{2.1}$$

We will repeatedly exploit this representation in what follows. As before $\rho_t(\mu)$ and $\rho_{s,t}(\mu)$ $(t \geq s)$ denote the solution of the Zakai equation at time t with the initial condition $\rho_0 = \mu$ and $\rho_s = \mu$, respectively, and $\pi_{s,t}(\mu) = \rho_{s,t}(\mu)/|\rho_{s,t}(\mu)|$.

We also recall the following interpretation of the norm $|\rho_t|$ of the unnormalized conditional distribution. If we define a new measure $\mathbf{Q} \sim \mathbf{P}$ through

$$\frac{d\mathbf{P}}{d\mathbf{Q}} = |\rho_t(\nu)| = |\rho_t|, \tag{2.2}$$

then under $\mathbf{Q}$ the observation process Y_t is an $\mathscr{F}_t^Y$-Wiener process. This observation will be used in Section 4 to apply the Malliavin calculus.

The main goal of this section is to establish some regularity properties of the solutions of the Wonham and Zakai equations. In particular, as we will want to calculate the derivative of the filter with respect to its initial condition, we have to establish that $\pi_{s,t}(\mu)$ is in fact differentiable. We will avoid problems at the boundary of the simplex by disposing of it alltogether: we begin by proving that if $\mu \in \mathcal{S}^{d-1}$, then a.s. $\pi_{s,t}(\mu) \in \mathcal{S}^{d-1}$ for all times $t > s$.

Lemma 2.1. $\mathbf{P}(\rho_{s,t}(\mu) \in \mathbb{R}^d_{++}$ for all $\mu \in \mathbb{R}^d_{++},\ 0 \le s \le t < \infty) = 1$.

Proof. The following variant on the pathwise filtering method reduces the Zakai equation to a random differential equation. First, we write $\Lambda^* = S + T$ where S is the diagonal matrix with $S_{ii} = \lambda_{ii}$. Note that the matrix T has only nonnegative entries. We now perform the transformation $f_{s,t}(\mu) = L_{s,t}\rho_{s,t}(\mu)$ where

$$L_{s,t} = \exp((\tfrac{1}{2}H^2 - S)(t-s) - H(Y_t - Y_s)).$$

Then $f_{s,t}(\mu)$ satisfies

$$\frac{df_{s,t}}{dt} = L_{s,t} T L_{s,t}^{-1} f_{s,t}, \qquad f_{s,s} = \mu. \tag{2.3}$$

Let $\Omega_c \subset \Omega$, $\mathbf{P}(\Omega_c) = 1$ be a set such that $t \mapsto B_t(\omega)$ is continuous for every $\omega \in \Omega_c$. Then $t \mapsto L_{s,t}$, $t \mapsto L_{s,t}^{-1}$ are continuous in t and have strictly positive diagonal elements for every $\omega \in \Omega_c$. By standard arguments, there exists for every $\omega \in \Omega_c$, $\mu \in \mathbb{R}^d$ and $s \ge 0$ a unique solution $f_{s,t}(\mu)$ to Eq. (2.3) where $t \mapsto f_{s,t}(\mu)$ is a C^1-curve. Moreover, note that $L_{s,t}TL_{s,t}^{-1}$ has nonnegative matrix elements for every $\omega \in \Omega_c$, $s \le t < \infty$. Hence if $\mu \in \mathbb{R}^d_{++}$ then clearly $f_{s,t}(\mu)$ must be nondecreasing, i.e. $f_{s,t} \succeq f_{s,r}$ for every $t \ge r \ge s$ and $\omega \in \Omega_c$. But then $\mathbb{R}^d_{++}$ must be forward invariant under Eq. (2.3) for every $\omega \in \Omega_c$, and as $L_{s,t}$ has strictly positive diagonal elements the result follows. □

Corollary 2.2. $\mathbf{P}(\pi_{s,t}(\mu) \in \mathcal{S}^{d-1}$ for all $\mu \in \mathcal{S}^{d-1},\ 0 \le s \le t < \infty) = 1$.

Let us now investigate the map $\rho_{s,t}(\mu)$. As this map is linear in μ, we can write $\rho_{s,t}(\mu) = U_{s,t}\mu$ a.s. where the $d \times d$ matrix $U_{s,t}$ is the solution of

$$dU_{s,t} = \Lambda^* U_{s,t}\,dt + HU_{s,t}\,dY_t, \qquad U_{s,s} = I. \tag{2.4}$$

The following lemma establishes that $U_{s,t}$ defines a linear stochastic flow in $\mathbb{R}^d$.

Lemma 2.3. *For a.e. $\omega \in \Omega$ (i) $\rho_{s,t}(\mu) = U_{s,t}\mu$ for all $s \le t$; (ii) $U_{s,t}$ is continuous in (s,t); (iii) $U_{s,t}$ is invertible for all $s \le t$, where $U_{s,t}^{-1}$ is given by*

$$dU_{s,t}^{-1} = -U_{s,t}^{-1}\Lambda^*\,dt + U_{s,t}^{-1}H^2\,dt - U_{s,t}^{-1}H\,dY_t, \qquad U_{s,s}^{-1} = I; \tag{2.5}$$

(iv) $U_{r,t}U_{s,r} = U_{s,t}$ (and hence $U_{s,t}U_{s,r}^{-1} = U_{r,t}$) for all $s \le r \le t$.

Proof. Continuity of $U_{s,t}$ (and $U_{s,t}^{-1}$) is a standard property of solution of Lipschitz stochastic differential equations. Invertibility of $U_{0,t}$ for all $0 \le t < \infty$ is established in [17, p. 326], and it is evident that $U_{s,t} = U_{0,t}U_{0,s}^{-1}$ satisfies Eq. (2.4). The remaining statements follow, where we can use continuity to remove the time dependence of the exceptional set as in the proof of [17, p. 326]. □

We now turn to the properties of the map $\pi_{s,t}(\mu)$.

Lemma 2.4. *The Wonham filter generates a smooth stochastic semiflow in $\mathcal{S}^{d-1}$, i.e. the solutions $\pi_{s,t}(\mu)$ satisfy the following conditions:*

(1) *For a.e. $\omega \in \Omega$, $\pi_{s,t}(\mu) = \pi_{r,t}(\pi_{s,r}(\mu))$ for all $s \le r \le t$ and μ.*
(2) *For a.e. $\omega \in \Omega$, $\pi_{s,t}(\mu)$ is continuous in (s,t,μ).*
(3) *For a.e. $\omega \in \Omega$, the injective map $\pi_{s,t}(\cdot) : \mathcal{S}^{d-1} \to \mathcal{S}^{d-1}$ is C^∞ for all $s \le t$.*

Proof. For $x \in \mathbb{R}^d_{++}$ define $\Sigma(x) = x/|x|$, so that $\pi_{s,t}(\mu) = \Sigma(\rho_{s,t}(\mu))$ $(\mu \in \mathcal{S}^{d-1})$. Note that Σ is smooth on $\mathbb{R}^d_{++}$. Hence continuity in (s,t,μ) and smoothness with respect to μ follow directly from the corresponding properties of $\rho_{s,t}(\mu)$. The semiflow property $\pi_{s,t}(\mu) = \pi_{r,t}(\pi_{s,r}(\mu))$ follows directly from Lemma 2.3. It remains to prove injectivity.

Suppose that $\pi_{s,t}(\mu) = \pi_{s,t}(\nu)$ for some $\mu, \nu \in \mathcal{S}^{d-1}$. Then $U_{s,t}\mu/|U_{s,t}\mu| = U_{s,t}\nu/|U_{s,t}\nu|$, and as $U_{s,t}$ is invertible we have $\mu = (|U_{s,t}\mu|/|U_{s,t}\nu|)\nu$. But as μ and ν must lie in $\mathcal{S}^{d-1}$, it follows that $\mu = \nu$. Hence $\pi_t(\cdot)$ is injective. □

Remark 2.5. The results in this section hold identically if we replace Λ by $\tilde{\Lambda}$, h by $\tilde{h}$. We will use the obvious notation $\tilde{\pi}_{s,t}(\mu)$, $\tilde{\rho}_{s,t}(\mu)$, $\tilde{U}_{s,t}$, etc.

We finish this section by obtaining an expression for the approximation error in the case $\tilde{h} = h$; in fact, we will demonstrate the bound for this simple case in a more general setting than is considered in the following. Rather than considering the approximate Wonham filter with modified Λ, we will allow the approximate filter to have an arbitrary finite variation term, provided $\mathcal{S}^{d-1}$ is left invariant.

Proposition 2.6. *Let $\breve{\pi}_t$ be a process with continuous paths in $\mathcal{S}^{d-1}$ where*

$$d\breve{\pi}_t = f(\breve{\pi}_t)\,dt + (H - h^*\breve{\pi}_t)\breve{\pi}_t\,(dY_t - h^*\breve{\pi}_t\,dt), \qquad \breve{\pi}_0 = \mu \in \mathcal{S}^{d-1}. \tag{2.6}$$

Then the difference between $\breve{\pi}_t$ and the Wonham filter started at μ is a.s. given by

$$\breve{\pi}_t - \pi_t(\mu) = \int_0^t D\pi_{s,t}(\breve{\pi}_s) \cdot (f(\breve{\pi}_s) - \Lambda^*\breve{\pi}_s)\,ds,$$

where $D\pi_{s,t}(\mu) \cdot v$ is the derivative of $\pi_{s,t}(\mu)$ in the direction $v \in T\mathcal{S}^{d-1}$.

Proof. Define the (scalar) process Γ_t by

$$\Gamma_t = \exp\left(\int_0^t h^*\breve{\pi}_s\,dY_s - \frac{1}{2}\int_0^t (h^*\breve{\pi}_s)^2\,ds\right).$$

Using Itô's rule, we evaluate

$$\frac{d}{ds}(\Gamma_s U_{0,s}^{-1}\breve{\pi}_s) = \Gamma_s U_{0,s}^{-1}(f(\breve{\pi}_s) - \Lambda^*\breve{\pi}_s). \tag{2.7}$$

Multiplying both sides by $U_{0,t}$, we obtain

$$\frac{d}{ds}(\Gamma_s U_{s,t}\breve{\pi}_s) = \Gamma_s U_{s,t}(f(\breve{\pi}_s) - \Lambda^*\breve{\pi}_s).$$

Now introduce as before the map $\Sigma:\mathbb{R}^d_{++}\to\mathcal{S}^{d-1}$, $\Sigma(x)=x/|x|$, which is smooth on $\mathbb{R}^d_{++}$. Define the matrix $D\Sigma(x)$ with elements

$$[D\Sigma(x)]^{ij}=\frac{\partial\Sigma^i(x)}{\partial x^j}=\frac{1}{|x|}\left[\delta_{ij}-\Sigma^i(x)\right].$$

Note that $\Sigma(\alpha x)=\Sigma(x)$ for any $\alpha>0$. Hence

$$\frac{d}{ds}\Sigma(U_{s,t}\breve{\pi}_s)=\frac{d}{ds}\Sigma(\Gamma_sU_{s,t}\breve{\pi}_s)=D\Sigma(\Gamma_sU_{s,t}\breve{\pi}_s)\frac{d}{ds}(\Gamma_sU_{s,t}\breve{\pi}_s).$$

But then we have, using $D\Sigma(\alpha x)=\alpha^{-1}D\Sigma(x)$ $(\alpha>0)$,

$$\frac{d}{ds}\Sigma(U_{s,t}\breve{\pi}_s)=D\Sigma(\Gamma_sU_{s,t}\breve{\pi}_s)\Gamma_sU_{s,t}(f(\breve{\pi}_s)-\Lambda^*\breve{\pi}_s)$$
$$=D\Sigma(U_{s,t}\breve{\pi}_s)U_{s,t}(f(\breve{\pi}_s)-\Lambda^*\breve{\pi}_s).$$

On the other hand, we obtain from the representation $\pi_{s,t}(\mu)=\Sigma(U_{s,t}\mu)$

$$D\pi_{s,t}(\mu)\cdot v=D\Sigma(U_{s,t}\mu)U_{s,t}v,\qquad \mu\in\mathcal{S}^{d-1},\ v\in T\mathcal{S}^{d-1}.$$

Note that $f(\breve{\pi}_s)-\Lambda^*\breve{\pi}_s$ is neccessarily in $T\mathcal{S}^{d-1}$ as $\breve{\pi}_t$ evolves in $\mathcal{S}^{d-1}$, so that $D\Sigma(U_{s,t}\breve{\pi}_s)U_{s,t}(f(\breve{\pi}_s)-\Lambda^*\breve{\pi}_s)=D\pi_{s,t}(\breve{\pi}_s)\cdot(f(\breve{\pi}_s)-\Lambda^*\breve{\pi}_s)$. Finally, note that

$$\int_0^t\frac{d}{ds}\Sigma(U_{s,t}\breve{\pi}_s)\,ds=\Sigma(\breve{\pi}_t)-\Sigma(U_{0,t}\breve{\pi}_0)=\breve{\pi}_t-\pi_t(\mu),$$

and the proof is complete. □

Corollary 2.7. *Using the triangle inequality we obtain*

$$|\breve{\pi}_t-\pi_t(\mu)|\le\int_0^t|D\pi_{s,t}(\breve{\pi}_s)|\,|f(\breve{\pi}_s)-\Lambda^*\breve{\pi}_s|\,ds,$$

where $|D\pi_{s,t}(\mu)|=\sup\{|D\pi_{s,t}(\mu)\cdot v|:v\in T\mathcal{S}^{d-1},\ |v|=1\}$. *Moreover*

$$|\breve{\pi}_t-\pi_t(\nu)|\le|\pi_t(\mu)-\pi_t(\nu)|+\int_0^t|D\pi_{s,t}(\breve{\pi}_s)|\,|f(\breve{\pi}_s)-\Lambda^*\breve{\pi}_s|\,ds.$$

Remark 2.8. Corollary 2.7 suggests that the method used here could be applicable to a wider class of filter approximations than those obtained by misspecification of the underlying model. In particular, in the infinite-dimensional setting it is known [4] that by projecting the filter onto a properly chosen finite-dimensional manifold, one can obtain finite-dimensional approximate filters that take a form very similar to Eq. (2.6). In order to obtain useful error bounds for such approximations one would need to have a fairly tight estimate on the derivative of the filter with respect to its initial condition. Unfortunately, worst-case estimates of the type developed in Section 3 are not sufficiently tight to give quantitative results on the approximation error, even in the finite-state case. In the remainder of the article we will restrict ourselves to studying the robustness problem.

In the following, it will be convenient to turn around the role of the exact and approximate filters in Corollary 2.7, i.e. we will use the estimate

$$|\pi_t-\tilde{\pi}_t(\mu)|\le|\tilde{\pi}_t(\nu)-\tilde{\pi}_t(\mu)|+\int_0^t|D\tilde{\pi}_{s,t}(\pi_s)|\,|(\Lambda^*-\tilde{\Lambda}^*)\pi_s|\,ds,\qquad(2.8)$$

which holds provided $\tilde{h}=h$. The proof is identical to the one given above.

3. Exponential estimates for the derivative of the filter

In order for the bound Eq. (2.8) to be useful, we must have an exponential estimate for $|D\tilde{\pi}_{s,t}(\cdot)|$. The goal of this section is to obtain such an estimate. We proceed in two steps. First, we use native filtering arguments as in [1] to obtain an a.s. exponential estimate for $|D\pi_{0,t}(\nu)|$. As the laws of the observation processes generated by signals with different initial distributions and jump rates are equivalent, we can extend this a.s. bound to $|D\tilde{\pi}_{s,t}(\mu)|$. We find, however, that the proportionality constant in the exponential estimate depends on μ and diverges as μ approaches the boundary of the simplex. This makes a pathwise bound on $|D\tilde{\pi}_{s,t}(\pi_s)|$ difficult to obtain, as π_s can get arbitrarily close to the boundary of the simplex on the infinite time interval. Instead, we proceed to find a uniform bound on $\mathbf{E}|D\tilde{\pi}_{s,t}(\pi_s)|$.

We begin by recalling a few useful results from [1].

Lemma 3.1. *Assume μ,ν are in the interior of the simplex. Then*

$$\pi_t^i(\mu) = \frac{\sum_j (\mu^j/\nu^j)\,\mathbf{P}(X_0 = a_j,\, X_t = a_i|\mathscr{F}_t^Y)}{\sum_j (\mu^j/\nu^j)\,\mathbf{P}(X_0 = a_j|\mathscr{F}_t^Y)}. \tag{3.1}$$

Proof. Define a new measure $\mathbf{P}^\mu \sim \mathbf{P}$ through

$$\frac{d\mathbf{P}^\mu}{d\mathbf{P}} = \frac{\mu^{X_0}}{\nu^{X_0}}.$$

It is not difficult to verify that under $\mathbf{P}^\mu$, X_t is still a finite-state Markov process with intensities matrix Λ but with initial distribution $\mathbf{P}^\mu(X_0 = a_i) = \mu^i$. Hence evidently $\pi_t^i(\mu) = \mathbf{P}^\mu(X_t = a_i|\mathscr{F}_t^Y)$. Using the usual change of measure formula for conditional expectations, we can write

$$\pi_t^i(\mu) = \mathbf{E}_{\mathbf{P}^\mu}(I_{X_t=a_i}|\mathscr{F}_t^Y) = \frac{\mathbf{E}(I_{X_t=a_i}\,\mu^{X_0}/\nu^{X_0}|\mathscr{F}_t^Y)}{\mathbf{E}(\mu^{X_0}/\nu^{X_0}|\mathscr{F}_t^Y)}.$$

The result now follows immediately. □

For the proof of the following Lemma we refer to [1, Lemma 5.7, page 662].

Lemma 3.2. *Define $\rho_t^{ji} = \mathbf{P}(X_0 = a_j|\mathscr{F}_t^Y,\, X_t = a_i)$. Assume that $\lambda_{ij} > 0\ \forall i \neq j$. Then for any $t \geq 0$ we have the a.s. bound*

$$\max_{j,k,\ell} |\rho_t^{jk} - \rho_t^{j\ell}| \leq \exp\left(-2t \min_{p,q\neq p} \sqrt{\lambda_{pq}\lambda_{qp}}\right).$$

We are now ready to obtain some useful estimates.

Proposition 3.3. *Let $\lambda_{ij} > 0\ \forall i \neq j$ and $\nu \in \mathcal{S}^{d-1}$, $v \in T\mathcal{S}^{d-1}$. Then a.s.*

$$|D\pi_t(\nu)\cdot v| \leq \sum_k \frac{|v^k|}{\nu^k}\, \exp\left(-2t \min_{p,q\neq p} \sqrt{\lambda_{pq}\lambda_{qp}}\right).$$

Proof. We can calculate directly the directional derivative of (3.1):

$$(D\pi_t(\mu)\cdot v)^i = \frac{\sum_j \frac{v^j}{\nu^j}\,(\mathbf{P}(X_0 = a_j,\, X_t = a_i|\mathscr{F}_t^Y) - \pi_t^i(\mu)\,\mathbf{P}(X_0 = a_j|\mathscr{F}_t^Y))}{\sum_j \frac{\mu^j}{\nu^j}\,\mathbf{P}(X_0 = a_j|\mathscr{F}_t^Y)}.$$

Setting $\mu = \nu$, we obtain after some simple manipulations

$$(D\pi_t(\nu)\cdot v)^i = \pi_t^i(\nu) \sum_{j,k} (v^j/\nu^j)\, \pi_t^k(\nu)\,(\rho_t^{ji} - \rho_t^{jk}).$$

The result follows from the triangle inequality and Lemma 3.2. □

To obtain this bound we had to use the true initial distribution ν, jump rates λ_{ij} and observation function h. However, the almost sure nature of the result allows us to drop these requirements.

Corollary 3.4. *Let $\tilde{\lambda}_{ij} > 0$ $\forall i \neq j$ and $\mu \in \mathcal{S}^{d-1}$, $v \in T\mathcal{S}^{d-1}$. Then a.s.*

$$|D\tilde{\pi}_{s,t}(\mu) \cdot v| \leq \sum_k \frac{|v^k|}{\mu^k} \exp\left(-2(t-s) \min_{p,q\neq p} \sqrt{\tilde{\lambda}_{pq}\tilde{\lambda}_{qp}}\right). \tag{3.2}$$

Moreover, the result still holds if μ, v are $\mathscr{F}_s^Y$-measurable random variables with values a.s. in $\mathcal{S}^{d-1}$ and $T\mathcal{S}^{d-1}$, respectively.

Proof. Let $\tilde{X}_t$ be a finite-state Markov jump process with transition intensities matrix $\tilde{\Lambda}$ and initial distribution μ, defined on the same probability space Ω (we could extend Ω if neccessary to admit such a process). Using Girsanov's theorem we can find an equivalent measure under which the observations process Y_t has same law as does the process $d\tilde{Y}_t = \tilde{\mathrm{h}}(\tilde{X}_t)\,dt + dB_t$ under $\mathbf{P}$ (note that $\tilde{\mathrm{h}}(\tilde{X}_t) - \mathrm{h}(X_t)$ is a bounded process, so Novikov's condition is satisfied). But then $\tilde{\pi}_t^i(\mu)$, which depends only on Y, is equivalent in law under the new measure to the process $\mathbf{P}(\tilde{X}_t = a_i|\mathscr{F}_t^{\tilde{Y}})$ under $\mathbf{P}$. Hence we can invoke Proposition 3.3 under the new measure, and use equivalence of the measures to establish the result for $s = 0$. The result for $s > 0$ follows directly as the Wonham equation is time homogeneous.

To show that the result still holds when μ, v are random, note that $\tilde{\pi}_{s,t}$ only depends on the observation increments in the interval $[s,t]$, i.e. $D\tilde{\pi}_{s,t}(\mu) \cdot v$ is $\mathscr{F}_{[s,t]}^Y$-measurable where $\mathscr{F}_{[s,t]}^Y = \sigma\{Y_r - Y_s : s \leq r \leq t\}$. Under the equivalent measure $\mathbf{Q}$ introduced in Section 2, Y is a Wiener process and hence $\mathscr{F}_{[s,t]}^Y$ and $\mathscr{F}_s^Y$ are independent. It follows from the bound with constant μ, v that

$$\mathbf{E}_{\mathbf{Q}}(I_{|D\tilde{\pi}_{s,t}(\mu)\cdot v| \leq (*)}|\sigma\{\mu, v\}) = 1 \qquad \mathbf{Q}\text{-a.s.},$$

where $(*)$ is the right hand side of (3.2). Hence $\mathbf{E}_{\mathbf{Q}}(I_{|D\tilde{\pi}_{s,t}(\mu)\cdot v| \leq (*)}) = 1$, and the statement follows from $\mathbf{P} \sim \mathbf{Q}$. □

Proposition 3.5. *Let $\tilde{\lambda}_{ij} > 0$ $\forall i \neq j$ and $\mu_1, \mu_2 \in \mathcal{S}^{d-1}$. Then a.s.*

$$|\tilde{\pi}_{s,t}(\mu_2) - \tilde{\pi}_{s,t}(\mu_1)| \leq C\,|\mu_2 - \mu_1| \exp\left(-2(t-s) \min_{p,q\neq p} \sqrt{\tilde{\lambda}_{pq}\tilde{\lambda}_{qp}}\right),$$

where $C = \max\{1/\mu_1^k, 1/\mu_2^k : k = 1, \ldots, d\}$.

Proof. Define $\gamma(u) = \tilde{\pi}_{s,t}(\mu_1 + u(\mu_2 - \mu_1))$, $u \in [0,1]$. Then

$$\tilde{\pi}_{s,t}(\mu_2) - \tilde{\pi}_{s,t}(\mu_1) = \int_0^1 \frac{d\gamma}{du}\,du = \int_0^1 D\tilde{\pi}_{s,t}(\mu_1 + u(\mu_2 - \mu_1)) \cdot (\mu_2 - \mu_1)\,du.$$

Using the triangle inequality, we obtain

$$|\tilde{\pi}_{s,t}(\mu_2) - \tilde{\pi}_{s,t}(\mu_1)| \leq \sup_{u\in[0,1]} |D\tilde{\pi}_{s,t}(\mu_1 + u(\mu_2 - \mu_1)) \cdot (\mu_2 - \mu_1)|.$$

The result now follows from Corollary 3.4. □

Corollary 3.4 and Proposition 3.5 are exactly what we need to establish boundedness of Eq. (2.8). Note, however, that the right hand side of (3.2) is proportional to $1/\mu^i$, and we must estimate $|D\tilde{\pi}_{s,t}(\pi_s)|$. Though we established in Section 2 that π_s cannot hit the boundary of the simplex in finite time, it can get arbitrarily close to the boundary during the infinite time interval, thus rendering the right hand side of Eq. (3.2) arbitrarily large. If we can establish that $\sup_{s\geq 0}\mathbf{E}(1/\min_k \pi_s^k)<\infty$, however, then we can control $\mathbf{E}|D\tilde{\pi}_{s,t}(\pi_s)|$ to obtain a useful bound.

We begin with an auxiliary integrability property of π_t:

Lemma 3.6. *Let $\nu\in\mathcal{S}^{d-1}$ and $T<\infty$. Then*

$$\mathbf{E}\int_0^T (\pi_s^i)^{-k}\,ds<\infty,\qquad \forall\, i=1,...,d,\ \ k\geq 1.$$

Proof. Applying Itô's rule to the Wonham equation gives

$$d\log\pi_t^i=\left(\lambda_{ii}-\frac{1}{2}(h^i-h^*\pi_t)^2\right)\,dt+\sum_{j\neq i}\lambda_{ji}\frac{\pi_t^j}{\pi_t^i}\,dt+(h^i-h^*\pi_t)\,dW_t,$$

where the innovation $dW_t=dY_t-h^*\pi_t\,dt$ is an $\mathscr{F}_t^Y$-Wiener process. The application of Itô's rule is justified by a standard localization argument, as π_t is in $\mathcal{S}^{d-1}$ for all $t\geq 0$ a.s. and $\log x$ is smooth in $(0,1)$. As $\lambda_{ij}\geq 0$ for $j\neq i$, we estimate

$$-k\log\pi_t^i\leq -k\log\nu^i-k\lambda_{ii}t+\frac{k}{2}\max_j(h^i-h^j)^2\,t-k\int_0^t(h^i-h^*\pi_s)\,dW_s.$$

But as $h^i-h^*\pi_t$ is bounded, Novikov's condition is satisfied and hence

$$\mathbf{E}\,\exp\left(-k\int_0^t(h^i-h^*\pi_s)\,dW_s-\frac{k^2}{2}\int_0^t(h^i-h^*\pi_s)^2\,ds\right)=1.$$

Estimating the time integral, we obtain

$$\mathbf{E}(\pi_t^i)^{-k}\leq(\nu^i)^{-k}\exp\left(-k\lambda_{ii}t+\frac{1}{2}k(k+1)\max_j(h^i-h^j)^2\,t\right).$$

The Lemma now follows by the Fubini-Tonelli theorem, as $(\pi_s^i)^{-k}\geq 0$ a.s. □

We are now in a position to bound $\sup_{t\geq 0}\mathbf{E}(1/\min_i\pi_t^i)$.

Proposition 3.7. *Let $\nu\in\mathcal{S}^{d-1}$ and suppose that $\lambda_{ij}>0$ $\forall i\neq j$. Then*

$$\sup_{t\geq 0}\mathbf{E}\left(\frac{1}{\min_i\pi_t^i}\right)<\infty.$$

Proof. By Itô's rule and using the standard localization argument, we obtain

$$(\pi_t^i)^{-1}=(\nu^i)^{-1}-\int_0^t\lambda_{ii}(\pi_s^i)^{-1}\,ds-\int_0^t(\pi_s^i)^{-2}\sum_{j\neq i}\lambda_{ji}\pi_s^j\,ds$$
$$-\int_0^t(\pi_s^i)^{-1}(h^i-h^*\pi_s)\,dW_s+\int_0^t(\pi_s^i)^{-1}(h^i-h^*\pi_s)^2\,ds$$

where W_t is the innovations Wiener process. Using Lemma 3.6 we find

$$\mathbf{E}\int_0^t(\pi_s^i)^{-2}(h^i-h^*\pi_s)^2\,ds\leq\max_j(h^i-h^j)^2\,\mathbf{E}\int_0^t(\pi_s^i)^{-2}\,ds<\infty,$$

so the expectation of the stochastic integral term vanishes. Using the Fubini-Tonelli theorem, we can thus write

$$\mathbf{E}((\pi_t^i)^{-1}) = (\nu^i)^{-1} - \int_0^t \lambda_{ii}\, \mathbf{E}((\pi_s^i)^{-1})\, ds \\ - \int_0^t \mathbf{E}\left((\pi_s^i)^{-2} \sum_{j\neq i} \lambda_{ji}\pi_s^j \right) ds + \int_0^t \mathbf{E}((\pi_s^i)^{-1}(h^i - h^*\pi_s)^2)\, ds.$$

Taking the derivative and estimating each of the terms, we obtain

$$\frac{dM_t^i}{dt} \leq -\min_{j\neq i} \lambda_{ji}\, (M_t^i)^2 + \left(|\lambda_{ii}| + \min_{j\neq i} \lambda_{ji} + \max_j (h^i - h^j)^2 \right) M_t^i,$$

where we have written $M_t^i = \mathbf{E}((\pi_t^i)^{-1})$ and we have used $(M_t^i)^2 \leq \mathbf{E}(\pi_t^i)^{-2}$ by Jensen's inequality. Using the estimate

$$-K_1^i(M_t^i)^2 + K_2^i M_t^i \leq -K_2^i M_t^i + \frac{(K_2^i)^2}{K_1^i} \quad \text{for} \quad K_1^i > 0,$$

we now obtain

$$\frac{dM_t^i}{dt} \leq K_2^i \left(\frac{K_2^i}{K_1^i} - M_t^i \right), \qquad K_2^i = |\lambda_{ii}| + \min_{j\neq i} \lambda_{ji} + \max_j (h^i - h^j)^2,$$

where $K_1^i = \min_{j\neq i} \lambda_{ji} > 0$. Consequently we obtain

$$M_t^i \leq e^{-K_2^i t}(\nu^i)^{-1} + \frac{(K_2^i)^2}{K_1^i} e^{-K_2^i t} \int_0^t e^{K_2^i s} ds = e^{-K_2^i t}(\nu^i)^{-1} + \frac{K_2^i}{K_1^i}(1 - e^{-K_2^i t}).$$

We can now estimate

$$\sup_{t\geq 0} \mathbf{E}\left(\frac{1}{\min_i \pi_t^i} \right) \leq \sum_{i=1}^d \sup_{t\geq 0} \mathbf{E}\left(\frac{1}{\pi_t^i} \right) \leq \sum_{i=1}^d \left(\frac{1}{\nu^i} \vee \frac{K_2^i}{K_1^i} \right) < \infty,$$

which is what we set out to prove. □

We can now prove Theorem 1.1 for the special case $\tilde{h} = h$. Using Eq. (2.8), Corollary 3.4, Proposition 3.5, and Proposition 3.7, we obtain

$$\mathbf{E}|\pi_t - \tilde{\pi}_t(\mu)| \leq |\mu - \nu| \max_k \left\{ \frac{1}{\mu^k} \vee \frac{1}{\nu^k} \right\} \exp\left(-2t \min_{p,q\neq p} \sqrt{\tilde{\lambda}_{pq}\tilde{\lambda}_{qp}} \right) \\ + |\Lambda^* - \tilde{\Lambda}^*| \sup_{s\geq 0} \mathbf{E}(1/\min_k \pi_s^k) \int_0^t \exp\left(-2(t-s) \min_{p,q\neq p} \sqrt{\tilde{\lambda}_{pq}\tilde{\lambda}_{qp}} \right) ds,$$

where $|\Lambda^* - \tilde{\Lambda}^*| = \sup\{|(\Lambda^* - \tilde{\Lambda}^*)\mu| : \mu \in \mathcal{S}^{d-1}\}$. Thus

$$\mathbf{E}|\pi_t - \tilde{\pi}_t(\mu)| \leq |\mu - \nu| \max_k \left\{ \frac{1}{\mu^k} \vee \frac{1}{\nu^k} \right\} e^{-\beta t} + |\Lambda^* - \tilde{\Lambda}^*| \frac{\sup_{s\geq 0} \mathbf{E}(1/\min_k \pi_s^k)}{\beta},$$

where we have written $\beta = 2\min_{p,q\neq p}(\tilde{\lambda}_{pq}\tilde{\lambda}_{qp})^{1/2}$. The result follows directly using $\|\pi_t - \tilde{\pi}_t(\mu)\|^2 \leq |\pi_t - \tilde{\pi}_t(\mu)|$ (as $|\pi_t^i - \tilde{\pi}_t(\mu)^i| \leq 1$).

4. Model robustness of the Wonham filter

We are now ready to proceed to the general case where the initial density, the transition intensities matrix and the observation function can all be misspecified. The simplicity of the special case $\tilde{h} = h$ that we have treated up to this point is due to the fact that in the calculation of Eq. (2.7), the stochastic integral term drops out and we can proceed with the calculation using only ordinary calculus. In the general case we can not get rid of the stochastic integral, and hence we run into anticipativity problems in the next step of the calculation.

We solve this problem by using anticipative stochastic integrals in the sense of Skorokhod, rather than the usual Itô integral (which is a special case of the Skorokhod integral defined for adapted processes only). Though the Skorokhod integral is more general than the Itô integral in the sense that it allows some anticipating integrands, it is less general in that we have to integrate against a Wiener process (rather than against an arbitrary semimartingale), and that the integrands should be functionals of the driving Wiener process. In our setup, the most convenient way to deal with this is to operate exclusively under the measure $\mathbf{Q}$ of Section 2, under which the observation process Y is a Wiener process. At the end of the day we can calculate the relevant expectation with respect to the measure $\mathbf{P}$ by using the explicit expression for the Radon-Nikodym derivative $d\mathbf{P}/d\mathbf{Q}$. The fact that the integrands must be functionals of the underlying Wiener process is not an issue, as both the approximate and exact filters are functionals of the observations only.

Our setup is further detailed in Appendix A, together with a review of the relevant results from the Malliavin calculus and anticipative stochastic calculus. Below we will use the notations and results from this Appendix without further comment. We will also refer to Appendix B for some results on smoothness of the various integrands we encounter; these results are not central to the calculations, but are required for the application of the theory in Appendix A.

We begin by obtaining an anticipative version of Proposition 2.6. Note that this result is precisely of the form one would expect. The first two lines follow the formula for the distance between two flows as one would guess e.g. from the discussion in the Introduction; the last line is an Itô correction term which contains second derivatives of the filter with respect to its initial condition.

Proposition 4.1. *The difference between π_t and $\tilde{\pi}_t$ satisfies*

$$
\begin{aligned}
\pi_t - \tilde{\pi}_t = &\int_0^t D\tilde{\pi}_{r,t}(\pi_r) \cdot \Delta_\Lambda \pi_r \, dr + \int_0^t D\tilde{\pi}_{r,t}(\pi_r) \cdot \Delta_H(\pi_r) \, dY_r \\
&- \int_0^t D\tilde{\pi}_{r,t}(\pi_r) \cdot \Big[h^*\pi_r \, (H - h^*\pi_r)\pi_r - \tilde{h}^*\pi_r \, (\tilde{H} - \tilde{h}^*\pi_r)\pi_r \Big] \, dr \\
&+ \frac{1}{2} \int_0^t \Big[D^2\tilde{\pi}_{r,t}(\pi_r) \cdot (H - h^*\pi_r)\pi_r - D^2\tilde{\pi}_{r,t}(\pi_r) \cdot (\tilde{H} - \tilde{h}^*\pi_r)\pi_r \Big] \, dr,
\end{aligned}
$$

where the stochastic integral is a Skorokhod integral and we have written $\Delta_\Lambda = \Lambda^* - \tilde{\Lambda}^*$, $\Delta_H(\pi) = (H - h^*\pi)\pi - (\tilde{H} - \tilde{h}^*\pi)\pi$, and $D^2\tilde{\pi}_{r,t}(\mu) \cdot v$ is the directional derivative of $D\tilde{\pi}_{r,t}(\mu) \cdot v$ with respect to $\mu \in \mathcal{S}^{d-1}$ in the direction $v \in T\mathcal{S}^{d-1}$.

Proof. Fix some $T > t$. We begin by evaluating, using Itô's rule and Eq. (2.5),

$$\tilde{U}_{0,s}^{-1}U_{0,s}\nu = \nu + \int_0^s \tilde{U}_{0,r}^{-1}(\Lambda^* - \tilde{\Lambda}^*)U_{0,r}\nu\,dr \\ - \int_0^s \tilde{U}_{0,r}^{-1}\tilde{H}(H-\tilde{H})U_{0,r}\nu\,dr + \int_0^s \tilde{U}_{0,r}^{-1}(H-\tilde{H})U_{0,r}\nu\,dY_r.$$

Now multiply from the left by $\tilde{U}_{0,t}$; we wish to use Lemma A.6 to bring $\tilde{U}_{0,t}$ into the Skorokhod integral term, i.e. we claim that

$$\tilde{U}_{s,t}U_{0,s}\nu = \tilde{U}_{0,t}\nu + \int_0^s \tilde{U}_{r,t}(\Lambda^* - \tilde{\Lambda}^*)U_{0,r}\nu\,dr - \int_0^s \tilde{U}_{r,t}\tilde{H}(H-\tilde{H})U_{0,r}\nu\,dr \\ + \int_0^s \tilde{U}_{r,t}(H-\tilde{H})U_{0,r}\nu\,dY_r + \int_0^s (\mathbf{D}_r\tilde{U}_{0,t})\tilde{U}_{0,r}^{-1}(H-\tilde{H})U_{0,r}\nu\,dr.$$

To justify this expression we need to verify the integrability conditions of Lemma A.6. Note that all matrix elements of $\tilde{U}_{s,t}$ are in $\mathbb{D}^\infty$ $\forall\, 0 \le s \le t < T$, and that

$$\mathbf{D}_r\tilde{U}_{s,t} = \begin{cases} 0 & \text{a.e. } r \notin [s,t], \\ \tilde{U}_{r,t}\tilde{H}\tilde{U}_{s,r} & \text{a.e. } r \in [s,t]. \end{cases}$$

This follows directly from Proposition A.4 and Lemma 2.3 (note that the same result holds for $U_{s,t}$ if we replace $\tilde{H}$ by H and $\tilde{U}$ by U). Once we plug this result into the expression above, the corresponding integrability conditions can be verified explicitly, see Lemma B.1, and hence we have verified that

$$\tilde{U}_{s,t}U_{0,s}\nu = \tilde{U}_{0,t}\nu + \int_0^s \tilde{U}_{r,t}(\Lambda^* - \tilde{\Lambda}^*)U_{0,r}\nu\,dr + \int_0^s \tilde{U}_{r,t}(H-\tilde{H})U_{0,r}\nu\,dY_r.$$

Next we would like to apply the anticipating Itô rule, Proposition A.7, with the function $\Sigma : \mathbb{R}^d_{++} \to \mathcal{S}^{d-1}$, $\Sigma(x) = x/|x|$. To this end we have to verify a set of technical conditions, see Lemma B.2. We obtain

$$\Sigma(\tilde{U}_{s,t}U_{0,s}\nu) = \Sigma(\tilde{U}_{0,t}\nu) + \int_0^s D\Sigma(\tilde{U}_{r,t}U_{0,r}\nu)\tilde{U}_{r,t}(\Lambda^* - \tilde{\Lambda}^*)U_{0,r}\nu\,dr \\ + \frac{1}{2}\sum_{k,\ell}\int_0^s \frac{\partial^2\Sigma}{\partial x^k\partial x^\ell}(\tilde{U}_{r,t}U_{0,r}\nu)(\boldsymbol{\nabla}_r\tilde{U}_{r,t}U_{0,r}\nu)^k(\tilde{U}_{r,t}(H-\tilde{H})U_{0,r}\nu)^\ell\,dr \\ + \int_0^s D\Sigma(\tilde{U}_{r,t}U_{0,r}\nu)\tilde{U}_{r,t}(H-\tilde{H})U_{0,r}\nu\,dY_r.$$

We need to evaluate $\boldsymbol{\nabla}_r\tilde{U}_{r,t}U_{0,r}\nu$. Using Prop. A.2 and Lemma A.3, we calculate

$$\lim_{\varepsilon\searrow 0}\mathbf{D}_r\tilde{U}_{r+\varepsilon,t}U_{0,r+\varepsilon}\nu = \lim_{\varepsilon\searrow 0}\tilde{U}_{r+\varepsilon,t}U_{r,r+\varepsilon}HU_{0,r}\nu = \tilde{U}_{r,t}HU_{0,r}\nu,$$

and similarly

$$\lim_{\varepsilon\searrow 0}\mathbf{D}_r\tilde{U}_{r-\varepsilon,t}U_{0,r-\varepsilon}\nu = \lim_{\varepsilon\searrow 0}\tilde{U}_{r,t}\tilde{H}\tilde{U}_{r-\varepsilon,r}U_{0,r-\varepsilon}\nu = \tilde{U}_{r,t}\tilde{H}U_{0,r}\nu.$$

After some rearranging, we obtain

$$\begin{aligned}\Sigma(\tilde U_{s,t}U_{0,s}\nu) &= \Sigma(\tilde U_{0,t}\nu) + \int_0^s D\Sigma(\tilde U_{r,t}U_{0,r}\nu)\tilde U_{r,t}(\Lambda^* - \tilde\Lambda^*)U_{0,r}\nu\, dr \\ &+ \frac{1}{2}\sum_{k,\ell}\int_0^s \frac{\partial^2\Sigma}{\partial x^k\partial x^\ell}(\tilde U_{r,t}U_{0,r}\nu)(\tilde U_{r,t}HU_{0,r}\nu)^k(\tilde U_{r,t}HU_{0,r}\nu)^\ell\, dr \\ &- \frac{1}{2}\sum_{k,\ell}\int_0^s \frac{\partial^2\Sigma}{\partial x^k\partial x^\ell}(\tilde U_{r,t}U_{0,r}\nu)(\tilde U_{r,t}\tilde HU_{0,r}\nu)^k(\tilde U_{r,t}\tilde HU_{0,r}\nu)^\ell\, dr \\ &+ \int_0^s D\Sigma(\tilde U_{r,t}U_{0,r}\nu)\tilde U_{r,t}(H-\tilde H)U_{0,r}\nu\, dY_r.\end{aligned}$$

From this point onwards we will set $s = t$. We will need (on $\mathbb{R}^d_{++}$)

$$D^2\Sigma^{ik\ell}(x) = \frac{\partial^2\Sigma^i(x)}{\partial x^k\partial x^\ell} = -\frac{1}{|x|}(D\Sigma^{ik}(x) + D\Sigma^{i\ell}(x)).$$

Recall that $D\Sigma(\alpha x) = \alpha^{-1}D\Sigma(x)$; it follows that also $D^2\Sigma(\alpha x) = \alpha^{-2}D^2\Sigma(x)$ for $\alpha > 0$. Using these expressions with $\alpha = |U_{0,r}\nu|$, we get

$$\begin{aligned}\pi_t - \tilde\pi_t &= \int_0^t D\Sigma(\tilde U_{r,t}\pi_r)\tilde U_{r,t}\Delta_\Lambda\pi_r\, dr + \int_0^t D\Sigma(\tilde U_{r,t}\pi_r)\tilde U_{r,t}(H-\tilde H)\pi_r\, dY_r \\ &+ \frac{1}{2}\sum_{k,\ell}\int_0^t \frac{\partial^2\Sigma}{\partial x^k\partial x^\ell}(\tilde U_{r,t}\pi_r)(\tilde U_{r,t}H\pi_r)^k(\tilde U_{r,t}H\pi_r)^\ell\, dr \\ &- \frac{1}{2}\sum_{k,\ell}\int_0^t \frac{\partial^2\Sigma}{\partial x^k\partial x^\ell}(\tilde U_{r,t}\pi_r)(\tilde U_{r,t}\tilde H\pi_r)^k(\tilde U_{r,t}\tilde H\pi_r)^\ell\, dr.\end{aligned}$$

Next we want to express the integrands in terms of $D\tilde\pi_{r,t}(\pi_r)\cdot v$, etc., rather than in terms of $D\Sigma(x)$. Recall that $D\tilde\pi_{r,t}(\pi_r)\cdot v = D\Sigma(\tilde U_{r,t}\pi_r)\tilde U_{r,t}v$ when $v \in T\mathcal{S}^{d-1}$. Similar terms appear in the expression above, but e.g. $\tilde H\pi_r \notin T\mathcal{S}^{d-1}$. To rewrite the expression in the desired form, we use that $D\Sigma(\tilde U_{r,t}\pi_r)\tilde U_{r,t}\pi_r = 0$. Hence

$$D\Sigma(\tilde U_{r,t}\pi_r)\tilde U_{r,t}\tilde H\pi_r = D\Sigma(\tilde U_{r,t}\pi_r)\tilde U_{r,t}(\tilde H - \tilde h^*\pi_r)\pi_r = D\tilde\pi_{r,t}(\pi_r)\cdot(\tilde H - \tilde h^*\pi_r)\pi_r$$

and similarly for the other terms. Note also that

$$\sum_k D^2\Sigma^{ik\ell}(\tilde U_{r,t}\pi_r)(\tilde U_{r,t}\pi_r)^k = -D\Sigma^{i\ell}(\tilde U_{r,t}\pi_r).$$

Substituting this into the expression for $\pi_t - \tilde\pi_t$ and rearranging, we obtain

$$\begin{aligned}\pi_t - \tilde\pi_t &= \int_0^t D\tilde\pi_{r,t}(\pi_r)\cdot\Delta_\Lambda\pi_r\, dr + \int_0^t D\tilde\pi_{r,t}(\pi_r)\cdot\Delta_H(\pi_r)\, dY_r \\ &- \int_0^t D\tilde\pi_{r,t}(\pi_r)\cdot\Big[h^*\pi_r\,(H - h^*\pi_r)\pi_r - \tilde h^*\pi_r\,(\tilde H - \tilde h^*\pi_r)\pi_r\Big]\, dr \\ &+ \frac{1}{2}\sum_{k,\ell}\int_0^t \frac{\partial^2\Sigma}{\partial x^k\partial x^\ell}(\tilde U_{r,t}\pi_r)(\tilde U_{r,t}(H-h^*\pi_r)\pi_r)^k(\tilde U_{r,t}(H-h^*\pi_r)\pi_r)^\ell\, dr \\ &- \frac{1}{2}\sum_{k,\ell}\int_0^t \frac{\partial^2\Sigma}{\partial x^k\partial x^\ell}(\tilde U_{r,t}\pi_r)(\tilde U_{r,t}(\tilde H-\tilde h^*\pi_r)\pi_r)^k(\tilde U_{r,t}(\tilde H-\tilde h^*\pi_r)\pi_r)^\ell\, dr.\end{aligned}$$

It remains to note that we can write

$$(D^2\tilde{\pi}_{s,t}(\mu)\cdot v)^i = \sum_{k,\ell} D^2\Sigma^{ik\ell}(\tilde{U}_{s,t}\mu)(\tilde{U}_{s,t}v)^k(\tilde{U}_{s,t}v)^\ell.$$

The result follows immediately. □

Remark 4.2. We have allowed misspecification of most model parameters of the Wonham filter. One exception is the observation noise intensity: we have not considered observations of the form $dY_t = \mathrm{h}(X_t)\,dt + \sigma\,dB_t$ with $\sigma \neq 1$; in other words, the quadratic variation of Y_t is assumed to be known $[Y,Y]_t = t$. We do not consider this a significant drawback as the quadratic variation can be determined directly from the observation process Y_t. On the other hand, the model parameters ν, Λ, h are "hidden" and would have to be estimated, making these quantities much more prone to modelling errors.

If we allow misspecification of σ, we would have to be careful to specify in which way the filter is implemented: in this case, the normalized solution of the misspecified Zakai equation no longer coincides with the solution of the misspecified Wonham equation. Hence one obtains a different error estimate depending on whether the normalized solution of the misspecified Zakai equation, or the solution of the misspecified Wonham equation, is compared to the exact filter. Both cases can be treated using similar methods, but we have chosen not to pursue this here.

Let $e_t = \pi_t - \tilde{\pi}_t$. We wish to estimate the norm of e_t. Unfortunately, we can no longer use the triangle inequality as in Section 2 due to the presence of the stochastic integral; instead, we choose to calculate $\|e_t\|^2$, which is readily estimated.

Lemma 4.3. *The filtering error can be estimated by*

$$\begin{aligned}\mathbf{E_P}\|e_t\|^2 \le \int_0^t \mathbf{E_P}|D\tilde{\pi}_{r,t}(\pi_r)\cdot\Delta_\Lambda\pi_r|\,dr + K\int_0^t \mathbf{E_P}|D\tilde{\pi}_{r,t}(\pi_r)\cdot\Delta_H(\pi_r)|\,dr \\ + \int_0^t \mathbf{E_P}|D\tilde{\pi}_{r,t}(\pi_r)\cdot(h^*\pi_r\,(H-h^*\pi_r)\pi_r - \tilde{h}^*\pi_r\,(\tilde{H}-\tilde{h}^*\pi_r)\pi_r)|\,dr \\ + \frac{1}{2}\int_0^t \mathbf{E_P}|D^2\tilde{\pi}_{r,t}(\pi_r)\cdot(H-h^*\pi_r)\pi_r - D^2\tilde{\pi}_{r,t}(\pi_r)\cdot(\tilde{H}-\tilde{h}^*\pi_r)\pi_r|\,dr,\end{aligned}$$

where $K = 2\max_k|h^k| + \max_k|\tilde{h}^k|$.

Proof. We wish to calculate $\mathbf{E_P}\|e_t\|^2 = \mathbf{E_P}e_t^*e_t$. Using Prop. 4.1, we obtain

$$\begin{aligned}\mathbf{E_P}\|e_t\|^2 = \int_0^t \mathbf{E_P}\,e_t^*D\tilde{\pi}_{r,t}(\pi_r)\cdot\Delta_\Lambda\pi_r\,dr \\ + \mathbf{E_P}\left[e_t^*\int_0^t D\tilde{\pi}_{r,t}(\pi_r)\cdot\Delta_H(\pi_r)\,dY_r\right] \\ - \int_0^t \mathbf{E_P}\,e_t^*D\tilde{\pi}_{r,t}(\pi_r)\cdot\left[h^*\pi_r\,(H-h^*\pi_r)\pi_r - \tilde{h}^*\pi_r\,(\tilde{H}-\tilde{h}^*\pi_r)\pi_r\right]dr \\ + \frac{1}{2}\int_0^t \mathbf{E_P}\,e_t^*\left[D^2\tilde{\pi}_{r,t}(\pi_r)\cdot(H-h^*\pi_r)\pi_r - D^2\tilde{\pi}_{r,t}(\pi_r)\cdot(\tilde{H}-\tilde{h}^*\pi_r)\pi_r\right]dr.\end{aligned}$$

The chief difficulty is the stochastic integral term. Using Eq. (2.2), we can write

$$\mathbf{E_P}\left[e_t^*\int_0^t D\tilde{\pi}_{r,t}(\pi_r)\cdot\Delta_H(\pi_r)\,dY_r\right] = \mathbf{E_Q}\left[|U_{0,t}\nu|\,e_t^*\int_0^t D\tilde{\pi}_{r,t}(\pi_r)\cdot\Delta_H(\pi_r)\,dY_r\right].$$

We would like to apply Eq. (A.1) to evaluate this expression. First, we must establish that the integrand is in $\operatorname{Dom}\boldsymbol{\delta}$; this does not follow directly from Proposition 4.1, as the anticipative Itô rule which was used to obtain that result can yield integrands which are only in $\mathbb{L}^{1,2}_{\rm loc}$. We can verify directly, however, that the integrand in this case is indeed in $\operatorname{Dom}\boldsymbol{\delta}$, see Lemma B.3. Next, we must establish that $|U_{0,t}\nu|\,e_t^i$ is in $\mathbb{D}^{1,2}$ for every i. Note that $|U_{0,t}\nu| = \sum_i (U_{0,t}\nu)^i$, so $|U_{0,t}\nu|$ is in $\mathbb{D}^\infty$. Moreover, we establish in Lemma B.4 that $e_t\in\mathbb{D}^{1,2}$ and that $\mathbf{D}_r e_t$ is a bounded random variable for every t. Hence it follows from Proposition A.1 that $|U_{0,t}\nu|\,e_t^i\in\mathbb{D}^{1,2}$. Consequently we can apply Eq. (A.1), and we obtain

$$\begin{aligned}
&\mathbf{E_Q}\left[|U_{0,t}\nu|\,e_t^*\int_0^t D\tilde\pi_{r,t}(\pi_r)\cdot\Delta_H(\pi_r)\,dY_r\right] \\
&\qquad= \int_0^t \mathbf{E_Q}\left[(|U_{0,t}\nu|\,\mathbf{D}_r e_t^* + \mathbf{D}_r|U_{0,t}\nu|\,e_t^*)D\tilde\pi_{r,t}(\pi_r)\cdot\Delta_H(\pi_r)\right]dr \\
&\qquad= \int_0^t \mathbf{E_Q}\left[|U_{0,t}\nu|\,(\mathbf{D}_r\pi_t - \mathbf{D}_r\tilde\pi_t)^* D\tilde\pi_{r,t}(\pi_r)\cdot\Delta_H(\pi_r)\right]dr \\
&\qquad\qquad + \int_0^t \mathbf{E_Q}\left[\sum_i (U_{r,t}HU_{0,r}\nu)^i\,e_t^* D\tilde\pi_{r,t}(\pi_r)\cdot\Delta_H(\pi_r)\right]dr.
\end{aligned}$$

Now note that $|e_t^i|\le 1$, and that by Lemma B.4

$$|(\mathbf{D}_r\pi_t - \mathbf{D}_r\tilde\pi_t)^i| \le |(\mathbf{D}_r\pi_t)^i| + |(\mathbf{D}_r\tilde\pi_t)^i| \le \max_k |h^k| + \max_k |\tilde h^k|.$$

Furthermore we can estimate

$$\left|\frac{\sum_i (U_{r,t}HU_{0,r}\nu)^i}{|U_{0,t}\nu|}\right| \le \frac{1}{|U_{0,t}\nu|}\sum_{i,j,k} U_{r,t}^{ij}\,|h^j|\,U_{0,r}^{jk}\nu^k \le \max_k |h^k|,$$

where we have used a.s. nonnegativity of the matrix elements of $U_{0,r}$ and $U_{r,t}$ (this must be the case, as e.g. $U_{r,t}\mu$ has nonnegative entries for any vector μ with nonnegative entries). Hence we obtain, using the triangle inequality,

$$\begin{aligned}
&\mathbf{E_Q}\left[|U_{0,t}\nu|\,e_t^*\int_0^t D\tilde\pi_{r,t}(\pi_r)\cdot\Delta_H(\pi_r)\,dY_r\right] \\
&\qquad\le (2\max_k|h^k| + \max_k|\tilde h^k|)\int_0^t \mathbf{E_Q}|U_{0,t}\nu|\;|D\tilde\pi_{r,t}(\pi_r)\cdot\Delta_H(\pi_r)|\,dr.
\end{aligned}$$

The result follows after straightforward manipulations. □

Unlike in the case $\tilde h = h$, we now have to deal also with second derivatives of the filter with respect to its initial condition. These can be estimated much in the same way as we dealt with the first derivatives.

Lemma 4.4. *Let $\tilde\lambda_{ij}>0$ $\forall i\ne j$ and $\mu\in\mathcal{S}^{d-1}$, $v,w\in T\mathcal{S}^{d-1}$. Then a.s.*

$$\begin{aligned}
&|D^2\tilde\pi_{s,t}(\mu)\cdot v - D^2\tilde\pi_{s,t}(\mu)\cdot w| \\
&\qquad\le 2\sum_k \frac{|v^k+w^k|}{\mu^k}\sum_j \frac{|v^j-w^j|}{\mu^j}\,\exp\left(-2(t-s)\min_{p,q\ne p}\sqrt{\tilde\lambda_{pq}\tilde\lambda_{qp}}\right).
\end{aligned}$$

Moreover, the result still holds if μ, v, w are $\mathscr{F}_s^Y$-measurable random variables with values a.s. in $\mathcal{S}^{d-1}$ and $T\mathcal{S}^{d-1}$, respectively.

Proof. Proceeding as in the proof of Proposition 3.3, we can calculate directly the second derivative of (3.1):

$$(D^2\pi_t(\mu)\cdot v)^i = -2\,(D\pi_t(\mu)\cdot v)^i\,\frac{\sum_j (v^j/\nu^j)\,\mathbf{P}(X_0=a_j|\mathscr{F}_t^Y)}{\sum_j (\mu^j/\nu^j)\,\mathbf{P}(X_0=a_j|\mathscr{F}_t^Y)}.$$

Setting $\mu=\nu$ and using the triangle inequality, we obtain

$$|D^2\pi_t(\nu)\cdot v - D^2\pi_t(\nu)\cdot w| \le 2\sum_{i,j}\frac{|v^j(D\pi_t(\nu)\cdot v)^i - w^j(D\pi_t(\nu)\cdot w)^i|}{\nu^j}.$$

Another application of the triangle inequality and using Proposition 3.3 gives

$$\begin{aligned}&|D^2\pi_t(\nu)\cdot v - D^2\pi_t(\nu)\cdot w|\\ &\quad\le \sum_k \frac{|v^k+w^k|}{\nu^k}\,|D\pi_t(\nu)\cdot(v-w)| + \sum_k\frac{|v^k-w^k|}{\nu^k}\,|D\pi_t(\nu)\cdot(v+w)|\\ &\qquad\qquad\le 2\sum_k\frac{|v^k+w^k|}{\nu^k}\sum_j\frac{|v^j-w^j|}{\nu^j}\,\exp\left(-2t\min_{p,q\ne p}\sqrt{\lambda_{pq}\lambda_{qp}}\right).\end{aligned}$$

We can now repeat the arguments of Corollary 3.4 to establish that the result still holds if we replace $\pi_{0,t}$ by $\tilde\pi_{s,t}$, λ_{pq} by $\tilde\lambda_{pq}$, and ν, v, w by $\mathscr{F}_s^Y$-measurable random variables μ, v, w. This completes the proof. □

We are now ready to complete the proof of Theorem 1.1.

Proof of Theorem 1.1. Set $\beta = 2\min_{p,q\ne p}(\tilde\lambda_{pq}\tilde\lambda_{qp})^{1/2}$. Let us collect all the neccessary estimates. First, we have

$$\int_0^t \mathbf{E_P}|D\tilde\pi_{r,t}(\pi_r)\cdot\Delta_\Lambda\pi_r|\,dr \le \beta^{-1}\sup_{s\ge0}\mathbf{E_P}(1/\min_k\pi_s^k)\,|\Lambda^*-\tilde\Lambda^*|,$$

as we showed in Section 3. Next, we obtain

$$\int_0^t \mathbf{E_P}|D\tilde\pi_{r,t}(\pi_r)\cdot\Delta_H(\pi_r)|\,dr \le \beta^{-1}\sup_{\pi\in\mathcal{S}^{d-1}}\sum_k|h^k-\tilde h^k+\tilde h^*\pi-h^*\pi|$$

using Corollary 3.4. Using the triangle inequality, we can estimate this by

$$\int_0^t \mathbf{E_P}|D\tilde\pi_{r,t}(\pi_r)\cdot\Delta_H(\pi_r)|\,dr \le (d+1)\beta^{-1}|h-\tilde h|.$$

Next, we estimate using Corollary 3.4

$$\begin{aligned}&\int_0^t \mathbf{E_P}|D\tilde\pi_{r,t}(\pi_r)\cdot(h^*\pi_r\,(H-h^*\pi_r)\pi_r - \tilde h^*\pi_r\,(\tilde H-\tilde h^*\pi_r)\pi_r)|\,dr\\ &\qquad\le\beta^{-1}\sup_{\pi\in\mathcal{S}^{d-1}}\sum_k|h^*\pi\,(h^k-h^*\pi)-\tilde h^*\pi\,(\tilde h^k-\tilde h^*\pi)|\\ &\qquad\qquad\le\beta^{-1}\left((d+1)\max_k|h^k| + d\max_{k,\ell}|\tilde h^k-\tilde h^\ell|\right)|h-\tilde h|,\end{aligned}$$

where we have used the estimate

$$\begin{aligned}
&\sum_k |h^*\pi\,(h^k - h^*\pi) - \tilde h^*\pi\,(\tilde h^k - \tilde h^*\pi)| \\
&\qquad \le |h^*\pi| \sum_k |h^k - \tilde h^k + \tilde h^*\pi - h^*\pi| + |h^*\pi - \tilde h^*\pi| \sum_k |\tilde h^k - \tilde h^*\pi| \\
&\qquad\qquad \le (d+1)\max_k |h^k|\ |h-\tilde h| + |h-\tilde h| \sum_k |\tilde h^k - \tilde h^*\pi| \\
&\qquad\qquad\qquad \le \left((d+1)\max_k |h^k| + d\max_{k,\ell} |\tilde h^k - \tilde h^\ell|\right)|h-\tilde h|.
\end{aligned}$$

Next we estimate using Lemma 4.4

$$\begin{aligned}
&\frac{1}{2}\int_0^t \mathbf{E_P}|D^2\tilde\pi_{r,t}(\pi_r)\cdot(H-h^*\pi_r)\pi_r - D^2\tilde\pi_{r,t}(\pi_r)\cdot(\tilde H - \tilde h^*\pi_r)\pi_r|\,dr \\
&\qquad \le \beta^{-1} \sup_{\pi\in\mathcal{S}^{d-1}} \sum_k |h^k - h^*\pi + \tilde h^k - \tilde h^*\pi| \sum_j |h^j - \tilde h^j + \tilde h^*\pi - h^*\pi| \\
&\qquad\qquad \le d(d+1)\beta^{-1}\left(\max_{k,\ell}|h^k - h^\ell| + \max_{k,\ell}|\tilde h^k - \tilde h^\ell|\right)|h-\tilde h|.
\end{aligned}$$

We have now estimated all the terms in Lemma 4.3, and hence we have bounded $\mathbf{E_P}\|e_t\|^2 = \mathbf{E_P}\|\pi_t(\nu) - \tilde\pi_t(\nu)\|^2$. It remains to allow for misspecified initial conditions. To this end, we estimate

$$\|\pi_t(\nu) - \tilde\pi_t(\mu)\|^2 \le \|e_t\|^2 + \|\tilde\pi_t(\nu) - \tilde\pi_t(\mu)\|\,(\|\tilde\pi_t(\nu) - \tilde\pi_t(\mu)\| + 2\|\pi_t(\nu) - \tilde\pi_t(\nu)\|).$$

Hence we obtain using the equivalence of finite-dimensional norms $\|x\| \le K_{21}\,|x|$

$$\|\pi_t(\nu) - \tilde\pi_t(\mu)\|^2 \le \|e_t\|^2 + 6K_{21}\,|\tilde\pi_t(\nu) - \tilde\pi_t(\mu)|$$

where we have used that the simplex is contained in the $(d-1)$-dimensional unit sphere, so $\|\mu_1 - \mu_2\| \le 2\ \forall \mu_1,\mu_2 \in \Delta^{d-1}$. The statement of the Theorem now follows directly from Lemma 4.3, Proposition 3.5, and the estimates above. □

Appendix A. Anticipative stochastic calculus

The goal of this appendix is to recall briefly the main results of the Malliavin calculus, Skorokhod integrals and anticipative stochastic calculus that are needed in the proofs. In our application of the theory we wish to deal with functionals of the observation process $(Y_t)_{t\in[0,T]}$, where T is some finite time (usually we will calculate integrals from 0 to t, so we can choose any $T > t$). Recall from Section 2 that Y is an $\mathscr{F}_t^Y$-Wiener process under the measure $\mathbf{Q}$; it will thus be convenient to work always under $\mathbf{Q}$, as this puts us directly in the framework used e.g. in [13]. As the theory described below is defined $\mathbf{Q}$-a.s. and as $\mathbf{P} \sim \mathbf{Q}$, the corresponding properties under $\mathbf{P}$ are unambiguously obtained by using Eq. (2.2). We will presume this setup whenver the theory described here is applied.

A smooth random variable F is one of the form $f(Y(h_1),\ldots,Y(h_n))$, where $Y(h)$ denotes the Wiener integral of the deterministic function $h \in L^2([0,T])$ with respect to Y and f is a smooth function which is of polynomial growth together with all its derivatives. For smooth F the Malliavin derivative $\mathsf{D}F$ is defined by

$$\mathsf{D}_t F = \sum_{i=1}^n \frac{\partial f}{\partial x^i}(Y(h_1),\ldots,Y(h_n))\,h_i(t).$$

The Malliavin derivative D can be shown [13, page 26] to be closeable as an operator from $L^p(\Omega, \mathscr{F}_T^Y, \mathbf{Q})$ to $L^p(\Omega, \mathscr{F}_T^Y, \mathbf{Q}; L^2([0,T]))$ for any $p \geq 1$, and we denote the domain of D in $L^p(\Omega)$ by $\mathbb{D}^{1,p}$ (for notational convenience we will drop the measure $\mathbf{Q}$ and σ-algebra $\mathscr{F}_T^Y$ throughout this section, where it is understood that $L^p(\Omega)$ denotes $L^p(\Omega, \mathscr{F}_T^Y, \mathbf{Q})$, etc.). In fact, $\mathbb{D}^{1,p}$ is simply the closure of the set of smooth random variables in $L^p(\Omega)$ with respect to the norm

$$\|F\|_{1,p} = \left[\mathbf{E_Q}|F|^p + \mathbf{E_Q}\|\mathsf{D}F\|^p_{L^2([0,T])}\right]^{1/p}.$$

More generally, we consider iterated derivatives $\mathsf{D}^k F \in L^p(\Omega; L^2([0,T]^k))$ defined by $\mathsf{D}^k_{t_1,\ldots,t_k} F = \mathsf{D}_{t_1} \cdots \mathsf{D}_{t_k} F$. The domain of D^k in $L^p(\Omega)$ is denoted by $\mathbb{D}^{k,p}$, and coincides with the closure in $L^p(\Omega)$ of the smooth random variables with respect to the norm

$$\|F\|_{k,p} = \left[\mathbf{E_Q}|F|^p + \sum_{j=1}^k \mathbf{E_Q}\|\mathsf{D}^j F\|^p_{L^2([0,T]^j)}\right]^{1/p}.$$

The local property of the Malliavin derivative allows us to localize these domains [13, page 44–45]. For $F \in L^2(\Omega)$, suppse there exists a sequence $(\Omega_n, F_n)_{n\geq 1}$ with $\Omega_n \in \mathscr{F}_T^Y$ and $F_n \in \mathbb{D}^{k,p}$, such that $\Omega_n \nearrow \Omega$ a.s. and $F = F_n$ a.s. on Ω_n. Then $(\Omega_n, F_n)_{n\geq 1}$ localizes F in $\mathbb{D}^{k,p}$, and we define $\mathsf{D}F = \mathsf{D}F_n$ on Ω_n. The space of random variables that can be localized in $\mathbb{D}^{k,p}$ is denoted by $\mathbb{D}^{k,p}_{\mathrm{loc}}$.

The first result we will need is a chain rule for the Malliavin derivative.

Proposition A.1. *Let $\varphi : \mathbb{R}^m \to \mathbb{R}$ be C^1 and $F = (F^1, \ldots, F^m)$ be a random vector with components in $\mathbb{D}^{1,2}$. Then $\varphi(F) \in \mathbb{D}^{1,2}_{\mathrm{loc}}$ and*

$$\mathsf{D}\varphi(F) = \sum_{i=1}^m \frac{\partial \varphi}{\partial x^i}(F)\, \mathsf{D}F^i.$$

If $\varphi(F) \in L^2(\Omega)$ and $\mathsf{D}\varphi(F) \in L^2(\Omega \times [0,T])$, then $\varphi(F) \in \mathbb{D}^{1,2}$. These results still hold if F a.s. takes values in an open domain $V \subset \mathbb{R}^m$ and φ is $C^1(V)$.

The first (local) statement can be found in [14, Prop. 2.9]; the second statement can be proved in the same way as [16, Lemma A.1], and the proofs are easily adapted to the case where F a.s. takes values in some domain.

A useful class of random variables is $\mathbb{D}^\infty = \cap_{p\geq 1} \cap_{k\geq 1} \mathbb{D}^{k,p}$. Then $\mathsf{D}_t F \in \mathbb{D}^\infty$ for any $F \in \mathbb{D}^\infty$, and the chain rule extends as follows [13, page 62].

Proposition A.2. *Let $\varphi : \mathbb{R}^m \to \mathbb{R}$ be a smooth function which is of polynomial growth together with all its derivatives, and let $F = (F^1, \ldots, F^m)$ be a random vector with components in $\mathbb{D}^\infty$. Then $\varphi(F) \in \mathbb{D}^\infty$ and the usual chain rule holds. This implies that $\mathbb{D}^\infty$ is an algebra, i.e. $FG \in \mathbb{D}^\infty$ for $F, G \in \mathbb{D}^\infty$.*

We will also need the following property [13, page 32].

Lemma A.3. *For a Borel set $A \subset [0,T]$, denote by $\mathscr{F}_A^Y$ the σ-algebra generated by the random variables $\{Y(I_B) : B \subset A \text{ Borel}\}$. Let $F \in \mathbb{D}^{1,2}$ be $\mathscr{F}_A^Y$-measurable. Then $\mathsf{D}_t F = 0$ a.e. in $\Omega \times ([0,T]\backslash A)$.*

It is useful to be able to calculate explicitly the Malliavin derivative of the solution of a stochastic differential equation. Consider

$$dx_t = f(x_t)\, dt + \sigma(x_t)\, dY_t, \qquad x_0 \in \mathbb{R}^m,$$

where $f(x)$ and $\sigma(x)$ are smooth functions of x with bounded derivatives of all orders. It is well known that such equations generate a smooth stochastic flow of diffeomorphisms $x_t = \xi_t(x)$ [9]. We now have the following result.

Proposition A.4. *All components of x_t belong to $\mathbb{D}^\infty$ for every $t \in [0,T]$. We have $\mathbf{D}_r x_t = D\xi_t(x_0) D\xi_r(x_0)^{-1}\sigma(x_r)$ a.e. $r < t$, where $(D\xi_t(x))^{ij} = \partial\xi_t^i(x)/\partial x^j$ is the Jacobian matrix of the flow, and $\mathbf{D}_r x_t = 0$ a.e. $r > t$.*

The first statement is given in [13, Theorem 2.2.2, page 105], the second on [13, Eq. (2.38), page 109]. $\mathbf{D}_r x_t = 0$ a.e. $r > t$ follows from adaptedness.

We now consider the Malliavin derivative as a closed operator from $L^2(\Omega)$ to $L^2(\Omega \times [0,T])$ with domain $\mathbb{D}^{1,2}$. Its Hilbert space adjoint $\boldsymbol{\delta} = \mathbf{D}^*$ is well defined in the usual sense as a closed operator from $L^2(\Omega\times[0,T])$ to $L^2(\Omega)$, and we denote its domain by $\mathrm{Dom}\,\boldsymbol{\delta}$. The operator $\boldsymbol{\delta}$ is called the Skorokhod integral, and coincides with the Itô integral on the subspace $L^2_a(\Omega \times [0,T]) \subset L^2(\Omega \times [0,T])$ of adapted square integrable processes [13, Prop. 1.3.4, page 41].

Lemma A.5. *$L^2_a(\Omega \times [0,T]) \subset \mathrm{Dom}\,\boldsymbol{\delta}$, and $\boldsymbol{\delta}|_{L^2_a}$ coincides with the Itô integral*

$$\boldsymbol{\delta}(u) = \int_0^T u_t\, dY_t, \qquad \forall u \in L^2_a(\Omega \times [0,T]).$$

The Skorokhod integral is thus an extension of the Itô integral to a class of possibly anticipative integrands. To emphasize this point we will use the same notation for Skorokhod integrals as for Itô integrals, i.e. we will write

$$\boldsymbol{\delta}(uI_{[s,t]}) = \int_s^t u_r\, dY_r, \qquad uI_{[s,t]} \in \mathrm{Dom}\,\boldsymbol{\delta}.$$

The Skorokhod integral has the following properties. First, its expectation vanishes $\mathbf{E_Q}\boldsymbol{\delta}(u) = 0$ if $u \in \mathrm{Dom}\,\boldsymbol{\delta}$. Second, by its definition as the adjoint of $\mathbf{D}$ we have

$$\mathbf{E_Q}(F\boldsymbol{\delta}(u)) = \mathbf{E_Q}\left[\int_0^T (\mathbf{D}_t F) u_t\, dt\right] \tag{A.1}$$

if $u \in \mathrm{Dom}\,\boldsymbol{\delta}$, $F \in \mathbb{D}^{1,2}$. We will also use the following result, the proof of which proceeds in exactly the same way as its one-dimensional counterpart [13, page 40].

Lemma A.6. *If u is an n-vector of processes in $\mathrm{Dom}\,\boldsymbol{\delta}$ and F is an $m \times n$-matrix of random variables in $\mathbb{D}^{1,2}$ such that $\mathbf{E_Q}\int_0^T \|Fu_t\|^2\,dt < \infty$, then*

$$\int_0^T Fu_t\, dY_t = F\int_0^T u_t\, dY_t - \int_0^T (\mathbf{D}_t F) u_t\, dt$$

in the sense that $Fu \in \mathrm{Dom}\,\boldsymbol{\delta}$ iff the right hand side of this expression is in $L^2(\Omega)$.

As it is difficult to obtain general statements for integrands in $\mathrm{Dom}\,\boldsymbol{\delta}$, it is useful to single out restricted classes of integrands that are easier to deal with. To this end, define the space $\mathbb{L}^{1,2} = L^2([0,T];\mathbb{D}^{1,2})$, i.e. the space of processes u such that $u_t \in \mathbb{D}^{1,2}$ and such that the norm

$$\|u\|_{1,2} = \left[\|u\|^2_{L^2(\Omega\times[0,T])} + \|\mathbf{D}u\|^2_{L^2(\Omega\times[0,T]^2)}\right]^{1/2}$$

is finite. Similarly, we define $\mathbb{L}^{k,p} = L^p([0,T];\mathbb{D}^{k,p})$ for $k \geq 1$, $p \geq 2$. Then $\mathbb{L}^{k,p} \subset \mathbb{L}^{1,2} \subset \mathrm{Dom}\,\boldsymbol{\delta}$ [13, page 38]. Moreover, the Skorokhod integral satisfies the local property on $\mathbb{L}^{1,2}$, so that the domains $\mathbb{L}^{k,p}$ can be localized to $\mathbb{L}^{k,p}_{\mathrm{loc}}$ in the same way as we localized $\mathbb{D}^{k,p}$ to $\mathbb{D}^{k,p}_{\mathrm{loc}}$ [13, page 43–45].

We finish this appendix with a statement of the Itô change of variables formula for Skorokhod integral processes. Various versions of the formula can be found in [14, 15, 13]. The extension to processes that a.s. take values in some domain is straightforward through localization.

Proposition A.7. *Consider an m-dimensional process of the form*

$$x_t = x_0 + \int_0^t v_s\, ds + \int_0^t u_s\, dY_s,$$

where we assume that x_t has a continuous version and $x_0 \in (\mathbb{D}^{1,4}_{\rm loc})^m$, $v \in (\mathbb{L}^{1,4}_{\rm loc})^m$, and $u \in (\mathbb{L}^{2,4}_{\rm loc})^m$. Let $\varphi : \mathbb{R}^m \to \mathbb{R}$ be a C^2 function. Then

$$\varphi(x_t) = \varphi(x_0) + \int_0^t D\varphi(x_s) v_s\, ds + \int_0^t D\varphi(x_s) u_s\, dY_s + \frac{1}{2}\int_0^t (D^2\varphi(x_s)\boldsymbol{\nabla}_s x_s, u_s)\, ds,$$

where we write $\boldsymbol{\nabla}_s x_s = \lim_{\varepsilon\searrow 0} \mathbf{D}_s(x_{s+\varepsilon} + x_{s-\varepsilon})$, $D\varphi(x_s)u_s = \sum_i (\partial\varphi/\partial x^i)(x_s)u^i_s$, $(D^2\varphi(x_s)\boldsymbol{\nabla}_s x_s, u_s) = \sum_{ij}(\partial^2\varphi/\partial x^i\partial x^j)(x_s)u^i_s\boldsymbol{\nabla}_s x^j_s$. The result still holds if x_s a.s. takes values in an open domain $V \subset \mathbb{R}^m$ $\forall s \in [0,t]$ and φ is $C^2(V)$.

Appendix B. Some technical results

Lemma B.1. *The following equality holds:*

$$\tilde U_{0,t}\int_0^s \tilde U^{-1}_{0,r}(H-\tilde H)U_{0,r}\nu\, dY_r = \int_0^s \tilde U_{r,t}(H-\tilde H)U_{0,r}\nu\, dY_r + \int_0^s \tilde U_{r,t}\tilde H(H-\tilde H)U_{0,r}\nu\, dr.$$

The integral on the left is an Itô integral, on the right a Skorokhod integral.

Proof. We have already established in the proof of Proposition 4.1 that the matrix elements of $\tilde U_{0,t}$ are in $\mathbb{D}^\infty \subset \mathbb{D}^{1,2}$. Moreover,

$$\begin{aligned}\mathbf{E}_{\mathbf{Q}}\|\tilde U_{r,t}(H-\tilde H)U_{0,r}\nu\|^2 &\le \|H-\tilde H\|^2\,\mathbf{E}_{\mathbf{Q}}(\|\tilde U_{r,t}\|^2\,\|U_{0,r}\|^2) \\ &\le \|H-\tilde H\|^2\sqrt{\mathbf{E}_{\mathbf{Q}}\|\tilde U_{r,t}\|^4\,\mathbf{E}_{\mathbf{Q}}\|U_{0,r}\|^4} \\ &\le C_4^4\,\|H-\tilde H\|^2\sqrt{\mathbf{E}_{\mathbf{Q}}|||\tilde U_{r,t}|||_4^4\,\mathbf{E}_{\mathbf{Q}}|||U_{0,r}|||_4^4},\end{aligned}$$

where we have used the Cauchy-Schwarz inequality and $\|\nu\| \le 1$ for $\nu \in \mathcal{S}^{d-1}$. Here $|||U|||_p = (\sum_{ij} U^p_{ij})^{1/p}$ is the elementwise p-norm of U, $\|U\|$ is the usual matrix 2-norm, and C_p matches the norms $\|U\| \le C_p|||U|||_p$ (recall that all norms on a finite-dimensional space are equivalent). As $U_{0,r}, \tilde U_{r,t}$ are solutions of linear stochastic differential equations, standard estimates give for any integer $p \ge 2$

$$\mathbf{E}_{\mathbf{Q}}\left(\sup_{0\le r\le t}|||\tilde U_{r,t}|||_p^p\right) \le D_1(p) < \infty, \quad \mathbf{E}_{\mathbf{Q}}\left(\sup_{0\le r\le t}|||U_{0,r}|||_p^p\right) \le D_2(p) < \infty,$$

and we obtain

$$\int_0^s \mathbf{E}_{\mathbf{Q}}\|\tilde U_{r,t}(H-\tilde H)U_{0,r}\nu\|^2\, dr \le s\sup_{0\le r\le s}\mathbf{E}_{\mathbf{Q}}\|\tilde U_{r,t}(H-\tilde H)U_{0,r}\nu\|^2 < \infty.$$

Hence we can apply Lemma A.6 to obtain the result. By a similar calculation we can establish that the right hand side of the expression in Lemma A.6 for our case is square integrable, so that the Skorokhod integral is well defined. □

Lemma B.2. *The anticipating Itô rule with $\Sigma(x) = x/|x|$ can be applied to*

$$\tilde U_{s,t}U_{0,s}\nu = \tilde U_{0,t}\nu + \int_0^s \tilde U_{r,t}(\Lambda^* - \tilde\Lambda^*)U_{0,r}\nu\,dr + \int_0^s \tilde U_{r,t}(H-\tilde H)U_{0,r}\nu\,dY_r.$$

Proof. Clearly the Skorokhod integral term has a.s. continuous sample paths, as both $\tilde U_{s,t}U_{0,s}\nu$ and the time integrals do; moreover, $\tilde U_{0,t}\nu \in (\mathbb{D}^\infty)^d$. In order to be able to apply Proposition A.7, it remains to check the technical conditions $v_r = \tilde U_{r,t}(\Lambda^* - \tilde\Lambda^*)U_{0,r}\nu \in (\mathbb{L}^{1,4})^d$, $u_r = \tilde U_{r,t}(H-\tilde H)U_{0,r}\nu \in (\mathbb{L}^{2,4})^d$.

As $\mathbb{D}^\infty$ is an algebra, u_t and v_t take values in $\mathbb{D}^\infty$. Moreover, we can establish exactly as in the proof of Lemma B.1 that u and v are in $L^4(\Omega\times[0,t])$. To complete the proof we must establish that

$$\sum_i \int_0^t \mathbf{E_Q}\left[\int_0^t (\mathbf{D}_s u_r^i)^2\,ds\right]^2 dr < \infty, \quad \sum_i \int_0^t \mathbf{E_Q}\left[\int_0^t (\mathbf{D}_s v_r^i)^2\,ds\right]^2 dr < \infty,$$

thus ensuring that $u, v \in (\mathbb{L}^{1,4})^d$, and

$$\sum_i \int_0^t \mathbf{E_Q}\left[\int_0^t\int_0^t (\mathbf{D}_\sigma\mathbf{D}_s u_r^i)^2\,ds\,d\sigma\right]^2 dr < \infty$$

which ensures that $u \in (\mathbb{L}^{2,4})^d$. Using the Cauchy-Schwarz inequality we have

$$\sum_i \int_0^t \mathbf{E_Q}\left[\int_0^t (\mathbf{D}_s u_r^i)^2\,ds\right]^2 dr \\ \le t\int_0^t\int_0^t \mathbf{E_Q}\|\mathbf{D}_s u_r\|_4^4\,ds\,dr \le t^3 \sup_{0\le r,s\le t}\mathbf{E_Q}\|\mathbf{D}_s u_r\|_4^4,$$

and similarly for v. Moreover, we obtain

$$\sum_i \int_0^t \mathbf{E_Q}\left[\int_0^t\int_0^t (\mathbf{D}_\sigma\mathbf{D}_s u_r^i)^2\,ds\,d\sigma\right]^2 dr \le t^5 \sup_{0\le r,s,\sigma\le t}\mathbf{E_Q}\|\mathbf{D}_\sigma\mathbf{D}_s u_r\|_4^4.$$

But using the chain rule Proposition A.2 we can easily establish that

$$\mathbf{D}_s u_r = \begin{cases} \tilde U_{r,t}(H-\tilde H)U_{s,r}HU_{0,s}\nu & \text{a.e. } 0<s<r<t, \\ \tilde U_{s,t}\tilde H\tilde U_{r,s}(H-\tilde H)U_{0,r}\nu & \text{a.e. } 0<r<s<t, \end{cases}$$

and similarly

$$\mathbf{D}_\sigma\mathbf{D}_s u_r = \begin{cases} \tilde U_{r,t}(H-\tilde H)U_{s,r}HU_{\sigma,s}HU_{0,\sigma}\nu & \text{a.e. } 0<\sigma<s<r<t, \\ \tilde U_{r,t}(H-\tilde H)U_{\sigma,r}HU_{s,\sigma}HU_{0,s}\nu & \text{a.e. } 0<s<\sigma<r<t, \\ \tilde U_{\sigma,t}\tilde H\tilde U_{r,\sigma}(H-\tilde H)U_{s,r}HU_{0,s}\nu & \text{a.e. } 0<s<r<\sigma<t, \\ \tilde U_{s,t}\tilde H\tilde U_{r,s}(H-\tilde H)U_{\sigma,r}HU_{0,\sigma}\nu & \text{a.e. } 0<\sigma<r<s<t, \\ \tilde U_{s,t}\tilde H\tilde U_{\sigma,s}\tilde H\tilde U_{r,\sigma}(H-\tilde H)U_{0,r}\nu & \text{a.e. } 0<r<\sigma<s<t, \\ \tilde U_{\sigma,t}\tilde H\tilde U_{s,\sigma}\tilde H\tilde U_{r,s}(H-\tilde H)U_{0,r}\nu & \text{a.e. } 0<r<s<\sigma<t. \end{cases}$$

The desired estimates now follow as in the proof of Lemma B.1. □

Lemma B.3. *The Skorokhod integrand obtained by applying the anticipative Itô formula as in Lemma B.2 is in* $\mathrm{Dom}\,\boldsymbol{\delta}$*.*

Proof. We use the notation $\rho_r = U_{0,r}\nu$. The Skorokhod integral in question is

$$\int_0^s D\Sigma(\tilde{U}_{r,t}\rho_r)\tilde{U}_{r,t}(H-\tilde{H})\rho_r\,dY_r = \int_0^s f_r\,dY_r.$$

To establish $f \in \operatorname{Dom}\boldsymbol{\delta}$, it suffices to show that $f \in \mathbb{L}^{1,2}$. We begin by showing

$$|D\Sigma(\tilde{U}_{r,t}\rho_r)\tilde{U}_{r,t}(H-\tilde{H})\rho_r| = \sum_i \left|\sum_{j,k}\frac{\delta^{ij}-\Sigma^i(\tilde{U}_{r,t}\rho_r)}{|\tilde{U}_{r,t}\rho_r|}\tilde{U}^{jk}_{r,t}(h^k-\tilde{h}^k)\rho^k_r\right| \le$$

$$\frac{1}{|\tilde{U}_{r,t}\rho_r|}\sum_{i,j,k}\tilde{U}^{jk}_{r,t}\,|h^k-\tilde{h}^k|\,\rho^k_r \le \frac{\max_k|h^k-\tilde{h}^k|}{|\tilde{U}_{r,t}\rho_r|}\sum_{i,j,k}\tilde{U}^{jk}_{r,t}\rho^k_r = d\,\max_k|h^k-\tilde{h}^k|,$$

where we have used the triangle inequality, $|\delta^{ij}-\Sigma^i(x)| \le 1$ for any $x \in \mathbb{R}^d_{++}$, and the fact that $U_{r,t}$ and ρ_r have nonnegative entries a.s. Hence f_r is a bounded process. Similarly, we will show that $\mathbf{D}_s f_r$ is a bounded process. Note that f_r is a smooth function on $\mathbb{R}^d_{++}$ of positive random variables in $\mathbb{D}^\infty$; hence we can apply the chain rule Proposition A.1. This gives

$$(\mathbf{D}_s f_r)^i = \begin{cases} \sum_{jk} D^2\Sigma^{ijk}(\tilde{U}_{r,t}\rho_r)(\tilde{U}_{r,t}(H-\tilde{H})\rho_r)^j(\tilde{U}_{r,t}U_{s,r}H\rho_s)^k \\ \qquad +\sum_j D\Sigma^{ij}(\tilde{U}_{r,t}\rho_r)(\tilde{U}_{r,t}(H-\tilde{H})U_{s,r}H\rho_s)^j \qquad \text{a.e. } s<r, \\ \sum_{jk} D^2\Sigma^{ijk}(\tilde{U}_{r,t}\rho_r)(\tilde{U}_{r,t}(H-\tilde{H})\rho_r)^j(\tilde{U}_{s,t}\tilde{H}\tilde{U}_{r,s}\rho_r)^k \\ \qquad +\sum_j D\Sigma^{ij}(\tilde{U}_{r,t}\rho_r)(\tilde{U}_{s,t}\tilde{H}\tilde{U}_{r,s}(H-\tilde{H})\rho_r)^j \qquad \text{a.e. } s>r. \end{cases}$$

Proceeding exactly as before, we find that $\mathbf{D}f \in L^\infty(\Omega\times[0,t]^2)$. But then by Proposition A.1 we can conclude that $\mathbf{D}_s f_r \in \mathbb{D}^{1,2}$ for a.e. $(s,t)\in[0,t]^2$, and in particular $f \in \mathbb{L}^{1,2}$. Hence the proof is complete. □

Lemma B.4. *$\mathbf{D}_r\pi_s = D\pi_{r,s}(\pi_r)\cdot(H-h^*\pi_r)\pi_r$ a.e. $r<s$, $\mathbf{D}_r\pi_s = 0$ a.e. $r>s$. Moreover $|(\mathbf{D}_r\pi_s)^i| \le \max_k|h^k|$ for every i. The equivalent results hold for $\mathbf{D}_r\tilde{\pi}_s$. In particular, this implies that π_s and $\tilde{\pi}_s$ are in $\mathbb{D}^{1,2}$.*

Proof. The case $r>s$ is immediate from adaptedness of π_s. For $r<s$, apply the chain rule to $\pi_s = \Sigma(U_{0,s}\nu) \in \mathbb{D}^{1,2}_{\mathrm{loc}}$. Boundedness of the resulting expression follows e.g. as in the proof of Lemma B.3, and hence it follows that $\pi_s \in \mathbb{D}^{1,2}$. □

References

[1] Baxendale, P., Chigansky, P., and Liptser, R. Asymptotic stability of the Wonham filter: ergodic and nonergodic signals. *SIAM J. Control Optim. 43*, 2 (2004), 643–669 (electronic).

[2] Bhatt, A. G., Kallianpur, G., and Karandikar, R. L. Uniqueness and robustness of solution of measure-valued equations of nonlinear filtering. *Ann. Probab. 23*, 4 (1995), 1895–1938.

[3] Bhatt, A. G., Kallianpur, G., and Karandikar, R. L. Robustness of the nonlinear filter. *Stochastic Process. Appl. 81*, 2 (1999), 247–254.

[4] Brigo, D., Hanzon, B., and Le Gland, F. Approximate nonlinear filtering by projection on exponential manifolds of densities. *Bernoulli 5*, 3 (1999), 495–534.

[5] Budhiraja, A., and Kushner, H. J. Robustness of nonlinear filters over the infinite time interval. *SIAM J. Control Optim. 36*, 5 (1998), 1618–1637 (electronic).

[6] Budhiraja, A., and Kushner, H. J. Approximation and limit results for nonlinear filters over an infinite time interval. *SIAM J. Control Optim. 37*, 6 (1999), 1946–1979 (electronic).

[7] Elliott, R. J., Aggoun, L., and Moore, J. B. *Hidden Markov models*, vol. 29 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 1995.

[8] Guo, X., and Yin, G. The Wonham filter with random parameters: rate of convergence and error bounds. *IEEE Trans. Automat. Control 51*, 3 (2006), 460–464.

[9] Kunita, H. Stochastic differential equations and stochastic flows of diffeomorphisms. In *École d'été de probabilités de Saint-Flour, XII—1982*, vol. 1097 of *Lecture Notes in Math.* Springer, Berlin, 1984, pp. 143–303.

[10] Le Gland, F., and Oudjane, N. A robustification approach to stability and to uniform particle approximation of nonlinear filters: the example of pseudo-mixing signals. *Stochastic Process. Appl. 106*, 2 (2003), 279–316.

[11] Le Gland, F., and Oudjane, N. Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filters. *Ann. Appl. Probab. 14*, 1 (2004), 144–187.

[12] Liptser, R. S., and Shiryaev, A. N. *Statistics of random processes. I*, vol. 5 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 2001.

[13] Nualart, D. *The Malliavin calculus and related topics*. Probability and its Applications (New York). Springer-Verlag, New York, 1995.

[14] Nualart, D., and Pardoux, É. Stochastic calculus with anticipating integrands. *Probab. Theory Related Fields 78*, 4 (1988), 535–581.

[15] Ocone, D., and Pardoux, É. A generalized Itô-Ventzell formula. Application to a class of anticipating stochastic differential equations. *Ann. Inst. H. Poincaré Probab. Statist. 25*, 1 (1989), 39–71.

[16] Ocone, D. L., and Karatzas, I. A generalized Clark representation formula, with application to optimal portfolios. *Stochastics Stochastics Rep. 34*, 3-4 (1991), 187–220.

[17] Protter, P. E. *Stochastic integration and differential equations*, vol. 21 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 2004.

[18] Wonham, W. M. Some applications of stochastic differential equations to optimal nonlinear filtering. *J. Soc. Indust. Appl. Math. Ser. A Control 2* (1965), 347–369 (1965).

Pavel Chigansky, Department of Mathematics, The Weizmann Institute of Science, Rehovot 76100, Israel

E-mail address: pavel.chigansky@weizmann.ac.il

Ramon van Handel, Physical Measurement and Control 266-33, California Institute of Technology, Pasadena, CA 91125, USA

E-mail address: ramon@its.caltech.edu