THEORY PROBAB. APPL.
Vol. 66, No. 2, pp. 245–262

BACKWARD NONLINEAR SMOOTHING DIFFUSIONS*

B. D. O. ANDERSON†, A. N. BISHOP‡, P. DEL MORAL§, AND C. PALMIER¶

Abstract. We present a backward diffusion flow (i.e., a backward-in-time stochastic differential equation) whose marginal distribution at any (earlier) time is equal to the smoothing distribution when the terminal state (at a later time) is distributed according to the filtering distribution. This is a novel interpretation of the smoothing solution in terms of a nonlinear diffusion (stochastic) flow. This solution contrasts with, and complements, the (backward) deterministic flow of probability distributions (viz. a type of Kushner smoothing equation) studied in a number of prior works. A number of corollaries of our main result are given, including a derivation of the time-reversal of a stochastic differential equation, and an immediate derivation of the classical Rauch–Tung–Striebel smoothing equations in the linear setting.

Key words. nonlinear filtering and smoothing, Kalman–Bucy filter, Rauch–Tung–Striebel smoother, particle filtering and smoothing, diffusion equations, stochastic semigroups, backward stochastic integration, backward Itô–Ventzell formula, time-reversed stochastic differential equations, Zakai and Kushner–Stratonovich equations

DOI. 10.1137/S0040585X97T99037X

1. Introduction. Let $(W_t, V_t) \in (\mathbf{R}^p \times \mathbf{R}^q)$ be a $(p+q)$-dimensional Brownian motion for finite $p, q \geqslant 1$. Consider a signal-observation model $(X_t, Y_t) \in (\mathbf{R}^m \times \mathbf{R}^n)$ given by the Itô stochastic differential equation,

$$\begin{cases} dX_t = a_t(X_t)\, dt + \sigma_t(X_t)\, dW_t, \\ dY_t = b_t(X_t)\, dt + \varsigma_t\, dV_t, \end{cases} \tag{1.1}$$

for some measurable functions $\varsigma_t, a_t(x), \sigma_t(x), b_t(x)$ with appropriate dimensions. We set $Y_0 = 0$ and let X_0 be an initial random variable (r.v.) with absolute moments of any order. We let $\alpha_t(x) := \sigma_t(x)\sigma_t'(x)$ and $\beta_t := \varsigma_t\varsigma_t'$, where A' denotes the transpose of some matrix A.

To avoid unnecessary technical details, we assume $\beta_t \geqslant \varepsilon I$, for some $\varepsilon > 0$, where I denotes the identity matrix. We also assume that the drift and sensor functions $(a_u(x), b_u(x))$, as well as the diffusion matrix $\sigma_u(x)$, are smooth with respect to (u, x) and have uniformly bounded derivatives with respect to x of all order on $(u, x) \in [s, t] \times \mathbf{R}^m$ for any $s \leqslant t$.

These technical conditions ensure that the above stochastic differential equation (1.1) has a global solution (X_t, Y_t) in the sense of Itô. In addition, (X_t, Y_t) as

*Received by the editors November 27, 2019; revised December 10, 2020. The first author was supported by the Australian Research Council (ARC) via grants DP160104500 and DP190100887, and by Data61-CSIRO. The third author was supported in part by the Chair Stress Test, RISK Management and Financial Steering, led by the French Ecole Polytechnique and its Foundation and sponsored by BNP Paribas. Originally published in the Russian journal *Teoriya Veroyatnostei i ee Primeneniya*, 66 (2021), pp. 305–326.

https://doi.org/10.1137/S0040585X97T99037X

†Research School of Electrical, Energy and Material Engineering, Australian National University, Canberra, Australia; Hangzhou Dianzi University, China; and Data61-CSIRO, Canberra, Australia (brian.anderson@anu.edu.au).

‡CSIRO, Sydney, Australia, and University of Technology Sydney (UTS), Sydney, Australia (Adrian.Bishop@uts.edu.au).

§INRIA, Bordeaux Research Center, Talence, France, and CMAP, Polytechnique Palaiseau, 91220 Palaiseau, France (delmoral@cict.fr, pierre.del_moral@inria.fr).

¶Institut de Mathématiques de Bordeaux (IMB), Bordeaux University, Bordeaux, France, and ONERA Palaiseau, 91220 Palaiseau, France (camille.palmi@etu.u-bordeaux.fr).

well as the sensor function $b_t(X_t)$ have absolute moments of any order. The stochastic flow associated with the signal is also smooth with respect to its initial condition, and its derivatives have absolute moments of any order.

The filtering problem then consists of computing the conditional distribution π_t of the random signal states X_t of the signal given the sigma-field $\mathcal{Y}_t = \sigma(Y_s, s \leqslant t)$ generated by the observations. The smoothing problem is to compute the conditional distribution $\pi_{t,s}$ of the random signal states X_s given $\mathcal{Y}_t$, with $t \geqslant s$. With this notation, we have $\pi_{t,t} = \pi_t$.

The filtering and smoothing problems have been studied extensively, and the literature on this topic is too broad to survey in detail here; a survey of this type is beyond the rather narrow scope of our contribution. We point the reader to the general texts [29], [5] for broad coverage of these problems.

We do note some rather seminal early literature in the linear setting [7], [38], [17], [40] and the nonlinear setting [7], [28], [1], [34], [3]. The first work on the smoothing topic is the maximum likelihood solution in [7] in both the linear and nonlinear settings. The study of [38] more formally confirms the linear result in [7] and also provides a simpler formulation for the mean and covariance of the smoothing distribution. In the nonlinear setting, the work of [1], [28] introduces an analogue of a type of Kushner–Stratonovich equation (see [5] for this equation in the filtering context) for the smoothing problem. More specifically, [1], [28] propose a deterministic partial differential equation that describes the flow of the smoothing distribution in terms of a backward flow and the standard filtering distribution, which acts as the boundary condition (the latter follows from the classical Kushner–Stratonovich equation).

In section 2 we state the main contribution of this work. Our main result asserts a backward diffusion flow (i.e., a backward stochastic differential equation), whose marginal distribution at any time $0 \leqslant s \leqslant t$ is equal to the smoothing distribution $\pi_{t,s}$ when the terminal state is distributed according to the filtering distribution π_t.

This is a novel interpretation of the smoothing solution in terms of a nonlinear diffusion (stochastic) flow (in the spirit of McKean–Vlasov-type processes). This solution contrasts with, and complements, say, the (backward) deterministic flow of probability distributions (viz. a type of Kushner smoothing equation) in [1], [28]. We also provide a number of corollaries of our main result in subsection 2.1, including an immediate derivation of the Rauch–Tung–Striebel smoothing equations [38] in the linear setting.

A number of auxiliary contributions are set forth in order to prove our main contribution to the smoothing problem. As is typical (e.g., see [7], [38], [17], [40], [28], [1], [3], [34], [35]), our smoothing solution requires the formulation of a related filtering problem. In section 3 we present a brief review of the Kallianpur–Striebel formula. We then provide a novel and more direct approach to deriving weak versions of the Zakai and Kushner–Stratonovich equations in subsections 3.1 and 3.2, respectively. We also consider the backward versions of these equations in subsection 3.3.

Our approach to the filtering equations in this article combines forward and backward Itô formulas for stochastic transport semigroups with a recent backward version of the Itô–Ventzell formula presented in [14]. This semigroup methodology can be seen as an extension to the Zakai and Kushner–Stratonovich equations, via the forward-backward stochastic analysis of diffusion flows developed in [11], [12], [14], [24], [25].

Our direct semigroup approach to the forward/backward filtering equations in this work contrasts with the classical stochastic partial differential methods and functional analysis in Sobolev spaces; see, e.g., the seminal works of Pardoux [32], [33], [35] as well as Krylov and Rozovskii [20], [21]. Related reverse time diffusions and filtered and smoothed densities are also developed in [2], [3] using discrete time approximation techniques, without a detailed discussion on the existence of these densities. We present a number of auxiliary results in this direction throughout section 3 which are utilized in the proof of our main smoothing result in section 4.

1.1. Some preliminary notation. This subsection presents some notation needed from the onset.

The signal and the observation defined in (1.1) are column vectors. Unless otherwise stated, we use the letters f and g to denote bounded scalar measurable test functions on some measurable space.

We denote by ∇f the column gradient whenever f is a differentiable function on some Euclidian space and denote by $\nabla^2 f$ the Hessian matrix whenever it is twice differentiable.

With $f\colon \mathbf{R}^m \to \mathbf{R}$, we let $\mathrm{div}_{\alpha_t}(f)$ be the α_t-divergence m-column vector operator with jth entry given by the formula

$$\mathrm{div}_{\alpha_t}(f)(x)^j := \sum_{1\leqslant i\leqslant m} \partial_{x_i}\big(\alpha_t^{i,j}(x) f(x)\big).$$

The generator L_t of the signal X_t is also given by the second order differential operator

$$L_t(f)(x) := \nabla f(x)' b_t(x) + \frac{1}{2}\,\mathrm{Tr}\big(\nabla^2 f(x)\,\alpha_t(x)\big)$$

with the trace operator $\mathrm{Tr}(\,\cdot\,)$. Here and throughout, and without further mention, we assume that functions f acted on by a second order differential generator are, in addition, twice differentiable with bounded derivatives.

For a measure μ and test function f of compatible dimension, we write

$$\mu(f) := \int \mu(dx)\, f(x).$$

An integral operator $\mathcal{Q}(x,dz)$ acts on the right on scalar test functions f and on the left on measures μ according to the formulas

$$\mathcal{Q}(f)(x) := \int \mathcal{Q}(x,dz)\, f(z) \quad \text{and} \quad (\mu\mathcal{Q})(dz) := \int \mu(dx)\, \mathcal{Q}(x,dz).$$

We extend this operator to an integral operator on matrix functions $h(x) = (h_{i,j}(x))_{i,j}$ by setting

$$\mathcal{Q}(h)(x)_{i,j} = \mathcal{Q}(h_{i,j})(x).$$

2. Main result. In the further development of this paper we assume, for any $t > 0$, that the conditional distribution π_t has a positive density $p_t := d\pi_t/d\lambda$ with respect to the Lebesgue measure λ on $\mathbf{R}^m$. In addition, $p_u(x)$ and its derivative $\nabla p_u(x)$ are uniformly bounded with respect to $(u,x) \in [s,t] \times \mathbf{R}^m$, for any given $s > 0$, almost surely with respect to the distribution of the observation process. A more detailed discussion on these regularity conditions is provided in subsection 2.2.

The main result of the paper is as follows.

THEOREM 2.1. *For any $t \geqslant u \geqslant s$, we have the transport equation*

$$\pi_{t,s}(dx) = (\pi_{t,u}\,\mathcal{K}_{u,s})(dx) := \int \pi_{t,u}(dz)\,\mathcal{K}_{u,s}(z,dx), \tag{2.1}$$

where $\mathcal{K}_{u,s}$ denotes the Markov semigroup of the backward diffusion flow

$$\begin{aligned} d\mathcal{X}_{u,s}(x) = -\big(&\big(p_s(\mathcal{X}_{u,s}(x))^{-1}\operatorname{div}_{\alpha_s}(p_s)(\mathcal{X}_{u,s}(x)) - a_s(\mathcal{X}_{u,s}(x))\big)\,ds \\ &+ \sigma_s(\mathcal{X}_{u,s}(x))\,d\mathcal{W}_s\big) \end{aligned} \tag{2.2}$$

with the boundary condition $\mathcal{X}_{u,u}(x) = x$, and where $\mathcal{W}_t \in \mathbf{R}^p$ denotes a p-dimensional Brownian motion independent of the observations.

The proof of the above theorem is provided in subsection 4.1. The backward stochastic differential equation (2.2) should be read as shorthand for the backward Itô integration formula

$$\begin{aligned} \mathcal{X}_{t,s}(x) = x &+ \int_s^t \big(p_u(\mathcal{X}_{t,u}(x))^{-1}\operatorname{div}_{\alpha_u}(p_u)(\mathcal{X}_{t,u}(x)) - a_u(\mathcal{X}_{t,u}(x))\big)\,du \\ &+ \int_s^t \sigma_u(\mathcal{X}_{t,u}(x))\,d\mathcal{W}_u \end{aligned} \tag{2.3}$$

with the terminal condition $\mathcal{X}_{t,t}(x) = x$. The rightmost term in the above formula is an Itô backward stochastic integral such that for any terminal time t this process is a square integrable backward martingale with respect to the variable $s \in [0,t]$.

Formally, we may slice the time interval $[s,t]_h := \{u_0, \dots, u_{n-1}\}$ via some time mesh $u_{i+1} = u_i + h$ from $u_0 = s$ to $u_n = t$ and with time step $h > 0$. In this notation, according to the backward equation (2.2) or (2.3), we compute $\mathcal{X}_{t,u-h}(x)$ from $\mathcal{X}_{t,u}(x)$ using the formula

$$\begin{aligned} \mathcal{X}_{t,u-h} - \mathcal{X}_{t,u} \simeq &\big(p_u(\mathcal{X}_{t,u})^{-1}\operatorname{div}_{\alpha_u}(p_u)(\mathcal{X}_{t,u}) - a_u(\mathcal{X}_{t,u})\big)h \\ &+ \sigma_u(\mathcal{X}_{t,u})(\mathcal{W}_u - \mathcal{W}_{u-h}). \end{aligned} \tag{2.4}$$

We provide some comments on the above theorem. By construction, given the observations and for any given $x \in \mathbf{R}^m$ and $t \geqslant s$, the probability $\mathcal{K}_{t,s}(x,dz)$ introduced in (2.1) coincides with the distribution of the random state $\mathcal{X}_{t,s}(x)$. In addition, for any $t \geqslant u \geqslant s$, we have the integral and stochastic semigroup properties

$$\mathcal{K}_{t,s}(x_2,dx_0) := \int \mathcal{K}_{t,u}(x_2,dx_1)\,\mathcal{K}_{u,s}(x_1,dx_0) \tag{2.5}$$

and

$$\mathcal{X}_{t,s} = \mathcal{X}_{u,s} \circ \mathcal{X}_{t,u}, \tag{2.6}$$

where $\mathcal{X}_{u,s} \circ \mathcal{X}_{t,u}$ denotes the composition of the mappings $\mathcal{X}_{u,s}$ and $\mathcal{X}_{t,u}$.

We let $\mathcal{X}_t$ be an r.v. with distribution π_t for some $t \geqslant 0$. According to (2.1) the random state $\mathcal{X}_{t,s}(\mathcal{X}_t)$ of the process (2.2) at any given $s \in [0,t]$ is distributed according to $\pi_{t,s} = \pi_t\mathcal{K}_{t,s}$. In other words, the backward process $\mathcal{X}_{t,s}(\mathcal{X}_t)$ is distributed according to the smoothing distribution $\pi_{t,s}$ for any $s \leqslant t$ whenever the terminal condition $\mathcal{X}_{t,t}(\mathcal{X}_t) = \mathcal{X}_t$ is distributed according to the filtering distribution π_t. In this sense, (2.2) constitutes a backward nonlinear smoothing diffusion. A forward diffusion flow that has a marginal distribution at any time equal to the filtering distribution is considered in [43], [44].

More generally, we have the backward Itô formula

$$df(\mathcal{X}_{t,s}(x)) = -\mathcal{L}_{s,\pi_s}(f)(\mathcal{X}_{t,s}(x))\,ds - \nabla f(\mathcal{X}_{t,s}(x))'\sigma_s(\mathcal{X}_{t,s}(x))\,d\mathcal{W}_s \tag{2.7}$$

with the second order differential operator

$$\mathcal{L}_{s,\pi_s}(f) = \sum_{1\leqslant j\leqslant m}\left(-a_s^j + \frac{1}{p_s}\operatorname{div}_{\alpha_s}(p_s)^j\right)\partial_{x_j}f + \frac{1}{2}\sum_{1\leqslant i,j\leqslant m}\alpha_s^{i,j}\,\partial_{x_ix_j}f. \tag{2.8}$$

Equivalently, we have the backward martingale decomposition

$$\begin{aligned} f(\mathcal{X}_{t,s}(x)) - f(x) - \int_s^t \mathcal{L}_{u,\pi_u}(f)(\mathcal{X}_{t,u}(x))\,du \\ = \int_s^t \nabla f(\mathcal{X}_{t,u}(x))'\sigma_u(\mathcal{X}_{t,u}(x))\,d\mathcal{W}_u. \end{aligned} \tag{2.9}$$

This yields the backward evolution equations

$$\partial_s\mathcal{K}_{t,s}(f)(x) = -\mathcal{K}_{t,s}(\mathcal{L}_{s,\pi_s}(f))(x) \tag{2.10}$$

and

$$\partial_s\pi_{t,s}(f) = -\pi_{t,s}(\mathcal{L}_{s,\pi_s}(f)) \tag{2.11}$$

with the terminal conditions $\mathcal{K}_{t,t}(f) = f$ and $\pi_{t,t} = \pi_t$. Formula (2.11) coincides with the conditional Fokker–Planck equation in [28], which was further developed in [1].

For further discussion on general backward integration of stochastic flows, see [11]; see also the appendix of [4] in the context of nonlinear diffusions, [35] in the context of nonlinear filtering, and [14] on forward-backward perturbation analysis of stochastic flows. Note that there is no issue with adaptation of the backward process in the sense studied in [36] since we rely only on the independent backward Brownian motion in (2.2). The "backward diffusion" in (2.2) is backward in the sense of a time reversed stochastic differential equation as in [2], [15], [31].

2.1. Some corollaries. In this subsection, we present some direct consequences of the above theorem.

Note that when $b_t = 0$ the measure π_t coincides with the distribution of the random state X_t of the signal. In this context, $\mathcal{X}_{t,s}(X_t)$ reduces to the time reversal of the signal starting at $\mathcal{X}_{t,t}(X_t) = X_t$ at the terminal time t. Using Theorem 2.1 we recover the fact that the time reversal process of the signal is itself a Markov diffusion [2], [15], [31]. More precisely, we have the following corollary.

COROLLARY 2.1 (see [2]). *Assume that $b_t = 0$. For any time horizon $t \geqslant 0$, the process $\mathfrak{X}_s^t := X_{t-s}$ with $s \in [0,t]$ is a Markov process with generator*

$$\mathfrak{L}_s^t(f) = \sum_{1\leqslant j\leqslant m}\left(\frac{1}{p_{t-s}}\operatorname{div}_{\alpha_{t-s}}(p_{t-s})^j - a_{t-s}^j\right)\partial_{x_j}f + \frac{1}{2}\sum_{1\leqslant i,j\leqslant m}\alpha_{t-s}^{i,j}\,\partial_{x_ix_j}f. \tag{2.12}$$

We consider now linear-Gaussian filtering/smoothing models with

$$\begin{aligned} &a_t(x) = A_t x, \quad b_t(x) = B_t x \\ &\text{and homogeneous diffusion matrix} \quad \sigma_t(x) = \Sigma_t \end{aligned} \tag{2.13}$$

for some matrices (A_t, B_t, Σ_t) with appropriate dimensions. Whenever X_0 is a Gaussian r.v. with mean $\widehat{X}_0$ and covariance matrix R_0, the optimal filter π_t is a Gaussian distribution with mean $\widehat{X}_t$ and covariance matrix R_t satisfying the Kalman–Bucy and Riccati equations [6],

$$\begin{cases} d\widehat{X}_t = A_t \widehat{X}_t\, dt + R_t B_t' \beta_t^{-1}(dY_t - B_t \widehat{X}_t\, dt), \\ \partial_t R_t = A_t R_t + R_t A_t' + \alpha_t - R_t B_t' \beta_t^{-1} B_t R_t. \end{cases} \tag{2.14}$$

In this context, we also have

$$-\, p_s(x)^{-1} \operatorname{div}_{\alpha_s}(p_s)(x) = \alpha_s R_s^{-1}(x - \widehat{X}_s). \tag{2.15}$$

This yields the following corollary.

Corollary 2.2. *For linear-Gaussian filtering models* (2.13), *the diffusion flow* $\mathcal{X}_{t,s}(x)$ *satisfies the backward evolution equation*

$$d\mathcal{X}_{t,s}(x) = -\big(\big(-A_s \mathcal{X}_{t,s}(x) - \alpha_s R_s^{-1}(\mathcal{X}_{t,s}(x) - \widehat{X}_s)\big)\, ds + \Sigma_s d\mathcal{W}_s\big) \tag{2.16}$$

with the boundary condition $\mathcal{X}_{t,t}(x) = x$.

Replacing x in (2.16) by an r.v. $\mathcal{X}_t$ with distribution π_t for any $t \geqslant s$, we see that $\mathcal{X}_{t,s}(\mathcal{X}_t)$ has distribution $\pi_{t,s}$. In addition, since the process is linear, the distribution $\pi_{t,s}$ is Gaussian with mean $\widehat{X}_{t,s}$ and covariance matrix $R_{t,s}$. The discrete time version of (2.16) can be found in section 9.9.6 of [13].

Now taking expectations we readily deduce the rather well-known Rauch–Tung–Striebel smoothing equations [38], thereby simplifying the derivation using the innovation techniques and the sophisticated approximation theory developed in [17], [28], [40], or the formal variational approaches and maximum likelihood techniques presented in the pioneering articles [7], [38].

Corollary 2.3 (see [38]). *For any* $t \geqslant s$, *the parameters* $(\widehat{X}_{t,s}, R_{t,s})$ *satisfy the backward evolution equations*

$$\begin{cases} \partial_s \widehat{X}_{t,s} = A_s \widehat{X}_{t,s} + \alpha_s R_s^{-1}(\widehat{X}_{t,s} - \widehat{X}_s), \\ \partial_s R_{t,s} = (A_s + \alpha_s R_s^{-1}) R_{t,s} + R_{t,s}(A_s + \alpha_s R_s^{-1})' - \alpha_s \end{cases} \tag{2.17}$$

with the terminal conditions $(\widehat{X}_{t,t}, R_{t,t}) = (\widehat{X}_t, R_t)$.

2.2. Comments on our regularity conditions. We end this section with some comments on the regularity conditions discussed at the beginning of section 2. These conditions are clearly met for linear-Gaussian filtering models (see, e.g., (2.14) and (2.15)). They are also met for nonlinear models as soon as the signal satisfies a classical controllability-type condition.

Note first that whenever the signal is uniformly elliptic, in the sense that $\alpha_t(x) = \sigma_t(x)\sigma_t'(x) \geqslant \delta\, I$ for some $\delta > 0$, it is well known that X_t has a smooth positive density with respect to the Lebesgue measure on $\mathbf{R}^m$. Nevertheless, in many important applications this ellipticity condition is not satisfied. The parabolic Hörmander condition for time varying models [8], [16] is a weaker condition. For linear-Gaussian filtering problems, this condition reduces to the usual controllability condition. Indeed, if we replace the Brownian motions W_t by some arbitrary smooth control functions, all states are accessible from one to another, as soon as the Lie algebra generated by the controlled vector fields is of full rank. This result is also called the Chow–Rashevskii theorem [9], [39]. Under this Hörmander condition, the Hörmander theorem [16] ensures that the signal states have a smooth density with respect to the Lebesgue measure on $\mathbf{R}^m$. In addition, for any $s < t$ the Markov transition semigroup $P_{s,t}$ of the signal has a smooth positive density $(x,z) \mapsto p_{s,t}(x,z)$ with respect to the Lebesgue measure λ on $\mathbf{R}^m$. In addition, the integral operator $P_{s,t}$ with $s < t$ maps test functions f into bounded smooth functions $P_{s,t}(f)$ given by

$$P_{s,t}(f)(x) = \int P_{s,t}(x,dz)f(z) = \int f(z)p_{s,t}(x,z)\,dz.$$

A natural way to transfer the smoothing properties of $P_{s,t}$ to the optimal filter is to use the equation

$$\pi_t(f) = \pi_0(P_{0,t}(f)) + \int_0^t \pi_s\big(P_{s,t}(f)(b_s - \pi_s(b_s))\big)'\beta_s^{-1}\left(dY_s - \pi_s(b_s)\,ds\right) \tag{2.18}$$

given in Theorem 1.1 of [23]. Using this formula we readily check that for any $t > 0$ the conditional distribution π_t has a positive density p_t on $\mathbf{R}^m$. Whenever $\sigma_t(x)$ and $b_t(x)$ are also bounded, Theorem 3.6 in [30] (see also Theorem 6.3 in [24]) also ensures that p_u is smooth, and for any $k \geqslant 1$, any parameters $h > 0$, and any time horizon $t > 0$ we have

$$\sup_{h\leqslant s\leqslant t}\sup_{x\in\mathbf{R}^m}\left(|p_s(x)| + \|\nabla^k p_s(x)\|\right) < \infty, \tag{2.19}$$

where $\|\cdot\|$ stands for any (equivalent) norm on $\mathbf{R}^m$.

The above estimates are met for linear-Gaussian filtering models. Nevertheless, some caution must be used when considering estimates of the form (2.19). Indeed, most of the literature on stochastic partial differential equations arising in nonlinear filtering, such as the strong formulation of the Zakai and Kushner–Stratonovich equations, assumes that the sensor function is uniformly bounded; see, e.g., [24], [30], [35], [42], [45]. To the best of our knowledge the extension of the estimate (2.19) to more general unbounded sensor functions is still an open and important question.

We also note here that the Kallianpur–Striebel formula [18], [19] is valid as soon as $\beta_u \geqslant \varepsilon I$ for some $\varepsilon > 0$ and that the functions $(a_u(x), b_u(x), \sigma_u(x))$ are smooth with uniformly bounded derivatives with respect to x of all orders on $(u,x) \in [s,t] \times \mathbf{R}^m$ for any $s \leqslant t$. Weaker conditions can also be found in [5] and the recent article [10].

Since X_t has continuous paths, for any continuous function f and any $s \leqslant t$ the random mapping $u \in [s,t] \mapsto f(X_u)$ is a.s. a uniformly bounded function. In addition, $f(X_t)$ is integrable as soon as f has polynomial growth. Up to some classical localization procedure (see, e.g., Chapter 7 in [41]), these rather weak regularity properties also ensure that the integral semigroups that transport (in time) the filtering measures discussed in section 3, as well as their stochastic partial differential evolution equations, are well defined on any test function with polynomial growth.

3. Nonlinear filtering equations. As is well known (see, e.g., [7], [38], [17], [40], [28], [1], [3], [34], [35]), a solution to the smoothing problem will typically make use of the solution of a related filtering problem. Consequently, we need to present and develop some related filtering results to prove our main result, Theorem 2.1. This section is largely self-contained, but it is vital in the proof in section 4 of our main result.

The first part of this section presents the classical Kallianpur–Striebel formula, which acts as a continuous-time version of Bayes law. In subsections 3.1 and 3.2, respectively, we present the Zakai and Kushner–Stratonovich equations for the flow of the conditional filtering distributions (both unnormalized and normalized). These results are rather well known. For further background on these classical ideas, we refer the reader to the pioneering articles by Kallianpur and Striebel [18], [19], Kushner [27], and Zakai [45]. For more recent discussion on these probabilistic models, we refer the reader to [10], [5], [13] and the references therein. In this paper, we present a novel and self-contained derivation based on stochastic transport semigroups and their forward evolution equations.

The solution of the Zakai equation is sometimes termed the unnormalized filter. The semigroup that transports these filtering measures (in time) is discussed in subsection 3.1, and its normalized version is reviewed in subsection 3.2. Subsection 3.3 presents a novel direct approach for deriving the backward evolution of these transport semigroups. Our approach in subsection 3.3 combines the backward Itô formula for stochastic flows with the backward Itô–Ventzell formula presented in [14].

Now we introduce some notation/terminology and briefly present the Kallianpur–Striebel formula and the linear semigroup property of unnormalized measures. Let $X_{s,t}(x)$ be the stochastic flow of the signal on the time interval $[s,t]$ and starting at x at time s. Let $Z_{s,t}(x)$ be the multiplicative functional

$$\ln Z_{s,t}(x) := \int_s^t b_u(X_{s,u}(x))'\beta_u^{-1}\,dY_u - \frac{1}{2}\int_s^t b_u(X_{s,u}(x))'\beta_u^{-1}b_u(X_{s,u}(x))\,du. \tag{3.1}$$

When x is replaced by X_s, we may write $Z_{s,t}$ instead of $Z_{s,t}(X_s)$, and when $s=0$, we may also write Z_t instead of $Z_{0,t}$. With this notation, we have the classical Kallianpur–Striebel formula

$$\pi_t(f) = \frac{\gamma_t(f)}{\gamma_t(1)} \quad \text{with} \quad \gamma_t(f) := \mathbf{E}_0\big(f(X_t)Z_t\big).$$

Here, $\mathbf{E}_0(\,\cdot\,)$ denotes the expectation operator with respect to the signal with a fixed observation process.

The transport semigroup of the unnormalized measures γ_t is given, for any $s \leqslant t$, by the formula

$$\gamma_t = \gamma_s Q_{s,t} \quad \text{with} \quad Q_{s,t}(f)(x) := \mathbf{E}_0\big(f(X_{s,t}(x))Z_{s,t}(x)\big). \tag{3.2}$$

To check this claim observe that

$$\begin{aligned} Z_t = Z_s\,Z_{s,t} \quad \Longrightarrow \quad \mathbf{E}_0\big(f(X_t)Z_t\big) &= \mathbf{E}_0\big(Z_s\mathbf{E}_0(f(X_t)Z_{s,t} \mid X_s)\big) \\ &= \mathbf{E}_0\big(Z_sQ_{s,t}(f)(X_s)\big). \end{aligned}$$

Now, for any $s \leqslant u \leqslant t$, we have

$$\begin{aligned} Q_{s,t}(f)(X_s) &= \mathbf{E}_0\big(f(X_t)Z_{s,t} \mid X_s\big) = \mathbf{E}_0\big(Z_{s,u}\mathbf{E}\big(f(X_t)Z_{u,t} \mid X_u\big) \mid X_s\big) \\ &= \mathbf{E}_0\big(Z_{s,u}Q_{u,t}(f)(X_u) \mid X_s\big) = Q_{s,u}(Q_{u,t}(f))(X_u). \end{aligned}$$

This yields the integral semigroup formula

$$Q_{s,t}(x_0, dx_2) = (Q_{s,u}Q_{u,t})(x_0, dx_2) := \int Q_{s,u}(x_0, dx_1)\, Q_{s,u}(x_1, dx_2).$$

In a more compact form, the semigroup property takes the form

$$Q_{s,t} = Q_{s,u}Q_{u,t} \quad \text{with} \quad Q_{t,t} = I, \quad \text{where } I \text{ denotes the identity operator.}$$

3.1. Unnormalized stochastic semigroups. Consider the stochastic transport semigroups $\mathbf{P}_{s,t}$ and $\mathbf{Q}_{s,t}$ defined by the composition of functions

$$\mathbf{P}_{s,t}(f)(x) := (f \circ X_{s,t})(x) \quad \text{and} \quad \mathbf{Q}_{s,t}(f)(x) := \mathbf{P}_{s,t}(f)(x) Z_{s,t}(x).$$

Using the semigroup properties of the stochastic flow $X_{s,t}(x)$ for any $s \leqslant u \leqslant t$, we check that

$$\mathbf{P}_{s,t}(f)(x) = (f \circ X_{s,t})(x) = (f \circ X_{u,t})(X_{s,u}(x)) = \mathbf{P}_{s,u}(\mathbf{P}_{u,t}(f))(x).$$

Similarly, we have

$$\mathbf{Q}_{s,t}(f)(x) = Z_{s,u}(x)\big(Z_{u,t}(X_{s,u}(x))(f \circ X_{s,t})(X_{s,u}(x))\big) = \mathbf{Q}_{s,u}(\mathbf{Q}_{u,t}(f))(x).$$

In the more compact form we have the semigroup properties

$$\mathbf{P}_{s,t} = \mathbf{P}_{s,u} \circ \mathbf{P}_{u,t} \quad \text{and} \quad \mathbf{Q}_{s,t} = \mathbf{Q}_{s,u} \circ \mathbf{Q}_{u,t} \quad \text{with} \quad \mathbf{P}_{t,t} = I = \mathbf{Q}_{t,t}.$$

We also observe that

$$P_{s,t}(f)(x) := \mathbf{E}_0\big(\mathbf{P}_{s,t}(f)(x)\big) \quad \text{and} \quad Q_{s,t}(f)(x) := \mathbf{E}_0\big(\mathbf{Q}_{s,t}(f)(x)\big).$$

The forward evolution equations of the above semigroups are described in the following proposition.

Proposition 3.1. *For any $t \geqslant s$, we have the forward stochastic evolution equation*

$$d\mathbf{Q}_{s,t}(f) = \mathbf{Q}_{s,t}(L_t(f))\, dt + \mathbf{Q}_{s,t}(f\, b_t')\beta_t^{-1}\, dY_t + \mathbf{Q}_{s,t}((\nabla f)'\sigma_t)\, dW_t, \tag{3.3}$$

with the initial condition $\mathbf{Q}_{s,s}(f) = f$, when $t = s$. In particular, we have the forward equation

$$dQ_{s,t}(f) = Q_{s,t}(L_t(f))\, dt + Q_{s,t}(f\, b_t')\beta_t^{-1}\, dY_t, \tag{3.4}$$

with the initial condition $Q_{s,s}(f) = f$, when $t = s$.

Proof. Assume that the sensor function $b_u(x)$ is uniformly bounded on $[s,t] \times \mathbf{R}^m$ for any $s \leqslant t$. Then the random process $(X_{s,u}(x), Z_{s,u}(x))$ also has uniformly bounded absolute moments of any order on any compact interval $[s,t]$ for any time parameters $s \leqslant t$. In this context, we use the Itô formula to check that

$$dZ_{s,t}(x) = Z_{s,t}(x) b_t(X_{s,t}(x))'\beta_t^{-1}\, dY_t$$

as well as that

$$d\mathbf{P}_{s,t}(f)(x) = \mathbf{P}_{s,t}(L_t(f))(x)\, dt + \mathbf{P}_{s,t}(\nabla f'\sigma_t)(x)\, dW_t.$$

An integration by parts yields

$$\begin{aligned} d\mathbf{Q}_{s,t}(f)(x) &= Z_{s,t}(x)\, d\mathbf{P}_{s,t}(f)(x) + \mathbf{P}_{s,t}(f)(x)\, dZ_{s,t}(x) \\ &= L_t(f)(X_{s,t}(x))Z_{s,t}(x)\, dt + Z_{s,t}(x) f(X_{s,t}(x)) b_t(X_{s,t}(x))' \beta_t^{-1}\, dY_t \\ &\quad + Z_{s,t}(x)\nabla f(X_{s,t}(x))'\sigma_t(X_{s,t}(x))\, dW_t. \end{aligned}$$

By classical localization principles of Itô integrals (see, for instance, Chapter 7 in [41]), the above result is also true for unbounded sensor functions. This completes the proof of (3.3). Taking the expectations, we conclude that

$$d\mathbf{E}_0(\mathbf{Q}_{s,t}(f)(x)) = \mathbf{E}_0\big(\mathbf{Q}_{s,t}(L_t(f))(x)\big)\, dt + \mathbf{E}_0\big(\mathbf{Q}_{s,t}(f b_t')(x)\big)\beta_t^{-1}\, dY_t.$$

This completes the proof of (3.4). The proof of the proposition is completed.

Combining (3.2) with Fubini's theorem, we readily check the weak form of the Zakai equation given by the formula

$$d\gamma_t(f) = \gamma_t(L_t(f))\, dt + \gamma_t(f b_t')\beta_t^{-1}\, dY_t. \tag{3.5}$$

Arguing as in (2.18), we transfer the smoothing properties of $P_{s,t}$ to $Q_{s,t}$ using the perturbation formulas given, for any $s < t$, by

$$Q_{s,t}(f) = P_{s,t}(f) + \int_s^t Q_{s,u}(P_{u,t}(f) b_u')\beta_u^{-1}\, dY_u.$$

Arguing as in [45], the above formula shows that, for any $s < t$, the integral operator $Q_{s,t}(x_0, dx_1)$ has a density $x_1 \mapsto q_{s,t}(x_0, x_1)$ with respect to the Lebesgue measure on $\mathbf{R}^m$ given by the integral equation

$$q_{s,t}(x_0,x_1) = p_{s,t}(x_0,x_1) + \int_s^t \left[\int q_{s,u}(x_0,z) p_{u,t}(z,x_1) b_u'(z)\, dz\right]\beta_u^{-1}\, dY_u. \tag{3.6}$$

3.2. Normalized stochastic semigroups. Let $\overline{Z}_{s,t}(x)$ be the multiplicative functional defined as $Z_{s,t}(x)$ by replacing in (3.1) the function b_u and the observation increment dY_u by the centered function $\overline{b}_u$ and the innovation increment $d\overline{Y}_u$, respectively, defined by the formulas

$$\overline{b}_u := b_u - \pi_u(b_u) \quad \text{and} \quad d\overline{Y}_u := dY_u - \pi_u(b_u)\, du.$$

Under our assumptions, the random process $\pi_t(b_t)$ is a.s. square integrable on any compact time interval so that the innovation process is well defined. Choosing $f = 1$ in (3.5), we check that

$$\ln \gamma_t(1) = \int_0^t \pi_u(b_u)'\beta_u^{-1}\, dY_u - \frac{1}{2}\int_0^t \pi_u(b_u)'\beta_u^{-1}\pi_u(b_u)\, du.$$

Observe that

$$\pi_s Q_{s,t}(1) = \frac{\gamma_t(1)}{\gamma_s(1)} = \exp\left(\int_s^t \pi_u(b_u)'\beta_u^{-1}\, dY_u - \frac{1}{2}\int_s^t \pi_u(b_u)'\beta_u^{-1}\pi_u(b_u)\, du\right).$$

We also consider the normalized stochastic semigroup

$$\overline{\mathbf{Q}}_{s,t}(f)(x) := (f\circ X_{s,t})(x)\overline{Z}_{s,t}(x) = \mathbf{P}_{s,t}(f)(x)\overline{Z}_{s,t}(x).$$

Arguing as above, for any $s \leqslant u \leqslant t$ we check that

$$\overline{\mathbf{Q}}_{s,t} = \overline{\mathbf{Q}}_{s,u} \circ \overline{\mathbf{Q}}_{u,t} \quad \text{and} \quad \overline{Z}_{s,t}(x) = Z_{s,t}(x)/\pi_s Q_{s,t}(1).$$

Consider the semigroup

$$\begin{aligned}\overline{Q}_{s,t}(f)(x) &:= \mathbf{E}_0\big(\overline{\mathbf{Q}}_{s,t}(f)(x)\big) = \mathbf{E}_0\big(f(X_{s,t}(x))\overline{Z}_{s,t}(x)\big) \\ &= Q_{s,t}(f)(x)/\pi_s Q_{s,t}(1).\end{aligned}$$

In this notation, using the same arguments as in the proof of Proposition 3.1, we have the following forward evolution equations.

PROPOSITION 3.2. *For any given time horizon s and for any $t \geqslant s$, we have the forward stochastic evolution equation*

$$d\overline{\mathbf{Q}}_{s,t}(f) = \overline{\mathbf{Q}}_{s,t}(L_t(f))\,dt + \overline{\mathbf{Q}}_{s,t}(f\overline{b}_t')\beta_t^{-1}\,d\overline{Y}_t + \overline{\mathbf{Q}}_{s,t}((\nabla f)'\sigma_t)\,dW_t,$$

with the initial condition $\overline{Q}_{s,s}(f) = f$, when $t = s$. In particular, we have the forward equation

$$d\overline{Q}_{s,t}(f) = \overline{Q}_{s,t}(L_t(f))\,dt + \overline{Q}_{s,t}(f\overline{b}_t')\beta_t^{-1}\,d\overline{Y}_t,$$

with the initial condition $\overline{Q}_{s,s}(f) = f$, when $t = s$.

The above proposition yields the weak form of the Kushner–Stratonovich equation defined by

$$d\pi_t(f) = \pi_t(L_t(f))\,dt + \pi_t(f\overline{b}_t)'\beta_t^{-1}\,d\overline{Y}_t. \tag{3.7}$$

Formally, using the same notation as in (3.11), we have the forward approximation equation

$$\pi_{u+h}(f) \simeq \pi_u(f) + \pi_u(L_u(f))h + \pi_u(f\overline{b}_u)'\beta_u^{-1}(\overline{Y}_{u+h} - \overline{Y}_u). \tag{3.8}$$

3.3. Backward evolution equations. This subsection is concerned with the backward evolution equation associated with the unnormalized semigroup $\mathbf{Q}_{s,t}$ and its normalized version. The main result of this subsection is the following theorem.

THEOREM 3.1. *For any twice differentiable function f with bounded derivatives and for any $s \leqslant t$, we have the backward evolution equation*

$$\begin{aligned}d\mathbf{Q}_{s,t}(f)(x) = &-\left(\nabla\mathbf{Q}_{s,t}(f)(x)'a_s(x) + \frac{1}{2}\operatorname{Tr}\big(\nabla^2\mathbf{Q}_{s,t}(f)(x)\alpha_s(x)\big)\right)ds \\ &- \mathbf{Q}_{s,t}(f)(x)b_s(x)'\beta_s^{-1}\,dY_s - \nabla\mathbf{Q}_{s,t}(f)(x)'\sigma_s(x)\,dW_s,\end{aligned} \tag{3.9}$$

with the terminal condition $\mathbf{Q}_{t,t}(f) = f$, when $s = t$. In particular, we have the backward equation

$$dQ_{s,t}(f) = -\big(L_s(Q_{s,t}(f))\,ds + Q_{s,t}(f)b_s'\beta_s^{-1}\,dY_s\big), \tag{3.10}$$

with the terminal condition $Q_{t,t}(f) = f$, when $s = t$.

Proof. We use a direct approach combining the backward filtering calculus developed in [24], [42] based on the backward Itô calculus developed in [11], [12], [22], [26]; see also the more recent article [14] and the references therein.

Consider the stochastic flow $\chi_{s,t}(\overline{x})$ starting at

$$\chi_{s,s}(\overline{x}) = \overline{x} := \begin{pmatrix} x \\ z \end{pmatrix} \in (\mathbf{R}^m \times \mathbf{R})$$

on the time interval $[s, \infty[$ and given, for any $t \geqslant s$, by

$$\chi_{s,t}(\overline{x}) := \begin{pmatrix} X_{s,t}(x) \\ Z_{s,t}(x)z \end{pmatrix} \in (\mathbf{R}^m \times \mathbf{R}).$$

We set

$$\mathcal{B}_t(\overline{x}) := \begin{pmatrix} a_t(x) \\ 0 \end{pmatrix}, \quad \mathcal{U}_t := \begin{pmatrix} W_t \\ Y_t \end{pmatrix},$$
$$\Lambda_t(\overline{x}) := \begin{pmatrix} \sigma_t(x) & 0 \\ 0 & z b_t(x)'\beta_t^{-1} \end{pmatrix}, \quad \mathcal{A}_t(\overline{x}) := \Lambda_t(\overline{x})\Lambda_t(\overline{x})'.$$

Assume that the sensor function $b_u(x)$ is uniformly bounded on $[s,t] \times \mathbf{R}^m$ for any $s \leqslant t$. Then, the process $(Z_{s,u}(x), \chi_{s,u}(\overline{x}))$ has continuous partial derivatives and also has uniformly bounded absolute moments of any order on $([s,t] \times \mathbf{R}^m)$ for any $s \leqslant t$. In this situation, we have the forward stochastic evolution equation

$$d\chi_{s,t}(\overline{x}) = \mathcal{B}_t(\chi_{s,t}(\overline{x}))\, dt + \Lambda_t(\chi_{s,t}(\overline{x}))\, d\mathcal{U}_t.$$

For any twice differentiable function F on $(\mathbf{R}^m \times \mathbf{R})$ with bounded derivatives we also have the backward equation

$$\begin{aligned} d(F \circ \chi_{s,t})(\overline{x}) = &-\left(\nabla(F \circ \chi_{s,t})(\overline{x})'\mathcal{B}_s(\overline{x}) + \frac{1}{2}\operatorname{Tr}\big(\nabla^2(F \circ \chi_{s,t})(\overline{x})\mathcal{A}_s(\overline{x})\big)\right) ds \\ &- \nabla(F \circ \chi_{s,t})(\overline{x})'\Lambda_s(\overline{x})\, d\mathcal{U}_s. \end{aligned}$$

A proof of the above formula can be found in [11], [12]; see also [14]. Choosing the function $F(\overline{x}) = f(x)z$, for some twice differentiable function f on $\mathbf{R}^m$ with bounded derivatives and letting $z = 1$ we check that

$$\begin{aligned} d(f(X_{s,t}(x))Z_{s,t}(x)) = &-\Big(\nabla\big(f(X_{s,t}(x))Z_{s,t}(x)\big)'a_s(x) \\ &+ \frac{1}{2}\operatorname{Tr}\big(\nabla^2\big(f(X_{s,t}(x))Z_{s,t}(x)\big)\alpha_s(x)\big)\Big)\, ds \\ &- \big(f(X_{s,t}(x))Z_{s,t}(x)\big)b_s(x)'\beta_s^{-1}\, dY_s - \nabla\big(f(X_{s,t}(x))Z_{s,t}(x)\big)'\sigma_s(x)\, dW_s. \end{aligned}$$

This completes the proof of (3.9). By localization arguments, the above result is also true for unbounded sensor functions. Integrating the flow of the signal we obtain (3.10). The theorem is proved.

We can also check (3.10) by considering a discrete time interval $[s,t]_h := \{t_0, \dots, t_{n-1}\}$ associated with some refining time mesh $t_{i+1} = t_i + h$ from $t_0 = s$ to $t_n = t$ for some time step $h > 0$. By (3.4), for any $u \in [s,t]_h$, we compute $Q_{u,t}(f)$ from $Q_{u+h,t}(f)$ using the backward equation

$$\begin{aligned} Q_{u,t}(f) &= Q_{u+h,t}(f) + (Q_{u,u+h} - I)(Q_{u+h,t}(f)) \\ &\simeq Q_{u+h,t}(f) + L_u(Q_{u+h,t}(f))h + Q_{u+h,t}(f)b_u'\beta_u^{-1}(Y_{u+h} - Y_u). \end{aligned} \tag{3.11}$$

For null sensor functions the evolution equation (3.9) coincides with the backward Itô formula discussed in [11], [12], [14], [24], [25].

Choosing $f = 1$ in (3.10) we recover the backward evolution of the likelihood function presented in [3], [34] (see formula (5.9) in [3] and equation (3.15) in [34]). Arguing as in (3.6), using (3.10) we check the perturbation formulas given, for any $s < t$, by

$$Q_{s,t}(f) = P_{s,t}(f) + \int_s^t P_{s,u}(Q_{u,t}(f)b_u')\beta_u^{-1}\,dY_u.$$

So, for any $s<t$, the integral operator $Q_{s,t}(x_0, dx_1)$ has a density $(x_0, x_1) \mapsto q_{s,t}(x_0, x_1)$ given by (3.6) and the integral formula

$$q_{s,t}(x_0, x_1) = p_{s,t}(x_0, x_1) + \int_s^t \left[\int p_{s,u}(x_0, z) q_{u,t}(z, x_1) b_u'(z)\,dz\right] \beta_u^{-1}\,dY_u. \tag{3.12}$$

Using the same arguments as in the proof of Theorem 3.1 we also have the following backward evolution equation.

PROPOSITION 3.3. *For any twice differentiable function f with bounded derivatives and for any $s \leqslant t$ we also have the backward equation*

$$\begin{aligned} d\overline{\mathbf{Q}}_{s,t}(f)(x) = &-\left(\nabla\overline{\mathbf{Q}}_{s,t}(f)(x)' a_s(x) + \frac{1}{2}\operatorname{Tr}\bigl(\nabla^2\overline{\mathbf{Q}}_{s,t}(f)(x)\alpha_s(x)\bigr)\right) ds \\ &- \overline{\mathbf{Q}}_{s,t}(f)(x)\overline{b}_s(x)'\beta_s^{-1}\,d\overline{Y}_s - \nabla\overline{\mathbf{Q}}_{s,t}(f)(x)'\sigma_s(x)\,dW_s \end{aligned}$$

with the terminal condition $\overline{\mathbf{Q}}_{t,t}(f) = f$. In particular, we have the backward equation

$$d\overline{Q}_{s,t}(f) = -\bigl(L_s(\overline{Q}_{s,t}(f))\,ds + \overline{Q}_{s,t}(f)\overline{b}_s'\beta_s^{-1}\,d\overline{Y}_s\bigr) \tag{3.13}$$

with the terminal condition $\overline{Q}_{t,t}(f) = f$.

Using the same notation as in (3.11), we also have the approximating backward equation

$$\overline{Q}_{u,t}(f) \simeq \overline{Q}_{u+h,t}(f) + L_u(\overline{Q}_{u+h,t}(f))h + \overline{Q}_{u+h,t}(f)\overline{b}_u'\beta_u^{-1}(\overline{Y}_{u+h} - \overline{Y}_u). \tag{3.14}$$

4. Smoothing semigroups and proof of the main result. This section is concerned with forward-backward evolution equations for the conditional smoothing distribution and with the proof of our main result.

Let $\mathcal{K}_{t,s}$ be the backward integral operator defined by

$$\mathcal{K}_{t,s}(f)(x) := \int \pi_s(dz)\,\frac{d\overline{Q}_{s,t}(z,\cdot)}{d\pi_t}(x)\,f(z). \tag{4.1}$$

For any $s \leqslant u \leqslant t$, we have the backward semigroup property

$$\mathcal{K}_{t,s} = \mathcal{K}_{t,u}\,\mathcal{K}_{u,s}, \tag{4.2}$$

which follows via

$$\begin{aligned} (\mathcal{K}_{t,u}\,\mathcal{K}_{u,s})(f)(x) &= \int \pi_s(dx_0)\overline{Q}_{s,u}(x_0, dx_1)\,\frac{d\overline{Q}_{u,t}(x_1,\cdot)}{d\pi_t}(x)\,f(x_0) \\ &= \int \pi_s(dx_0)\frac{d\overline{Q}_{s,t}(x_0,\cdot)}{d\pi_t}(x)\,f(x_0) = \mathcal{K}_{t,s}(f)(x), \end{aligned}$$

and where we exploit the semigroup properties of the operators $\overline{Q}_{s,t}$.

Also observe that, for any $t > s > 0$, the integral operator $\mathcal{K}_{t,s}(x_1, dx_0)$ has a density $(x_1, x_0) \mapsto k_{s,t}(x_1, x_0)$ with respect to the Lebesgue measure on $\mathbf{R}^m$ given by $k_{t,s}(x_1, x_0) := p_s(x_0)\overline{q}_{s,t}(x_0, x_1)/p_t(x_1)$ with $\overline{q}_{s,t}(x_0, x_1) = q_{s,t}(x_0, x_1)/\pi_s(Q_{s,t}(1))$. The function $q_{s,t}$ denotes the density of the integral operator $Q_{s,t}$ discussed in (3.6) and (3.12).

Now, for any pair of functions (f, g) we readily check the duality formula

$$\pi_s(f\overline{Q}_{s,t}(g)) = \pi_t(\mathcal{K}_{t,s}(f)g). \tag{4.3}$$

The following technical result is key in the proof of Theorem 2.1.

LEMMA 4.1. *For any time parameter $s \leqslant t$, we have the forward-backward differential equation*

$$\partial_s\big(\pi_s(f\overline{Q}_{s,t}(g))\big) = -\pi_s\big(\overline{Q}_{s,t}(g)\mathcal{L}_{s,\pi_s}(f)\big) \tag{4.4}$$

with the second order differential operator

$$\mathcal{L}_{s,\pi_s}(f) := -L_s(f) + \frac{1}{p_s}\sum_{1\leqslant i,j\leqslant m} \partial_{x_i}(p_s\alpha_s^{i,j}\partial_{x_j}f).$$

Proof. Observe that (4.4) does not involve the derivatives of the function g. Thus, up to a smooth-mollifier-type approximation of the function g, it suffices to check (4.4) for any pair of bounded and twice differentiable functions f, g with bounded differentials. Arguing as in the proof of Proposition 3.1 and Theorem 3.1, it suffices to prove the result for uniformly bounded sensor functions $b_u(x)$ on $[s, t] \times \mathbf{R}^m$ for any $s \leqslant t$.

In this situation, combining the Kushner–Stratonovich equation (3.7) and the backward equation (3.13), it is straightforward to check that the forward-backward evolution equation,

$$\partial_s\big(\pi_s(f\overline{Q}_{s,t}(g))\big) = \pi_s\big(L_s(f\overline{Q}_{s,t}(g)) - fL_s(\overline{Q}_{s,t}(g))\big), \tag{4.5}$$

follows for any $s \leqslant t$. The above equation can be proved using the backward Itô–Ventzell formula in [14]. We use the same notation as in the proof of Theorem 3.1. Let $\overline{Z}_{s,t}(x)$ be the multiplicative functional defined as $Z_{s,t}(x)$ by replacing the function b_u and the observation Itô-increment dY_u by the centered function $\overline{b}_u$ and the innovation increment $d\overline{Y}_u$.

Consider the backward random field $F_{s,t}$ with the terminal condition $F_{t,t}(\overline{x}) = f(x)g(x)z$ defined by the formula

$$F_{s,t}(\overline{x}) := f(x)\overline{\mathbf{Q}}_{s,t}(g)(x)z \quad \text{and we set} \quad \overline{\chi}_s := \begin{pmatrix} X_s \\ \overline{Z}_s \end{pmatrix} \in (\mathbf{R}^m \times \mathbf{R}).$$

In this notation, we have

$$\mathbf{E}_0(F_{s,t}(\overline{\chi}_s)) = \mathbf{E}_0\big(f(X_s)\overline{Z}_s\mathbf{E}_0\big(\overline{\mathbf{Q}}_{s,t}(g)(X_s) \mid (X_s, Z_s)\big)\big) = \pi_s(f\overline{Q}_{s,t}(g)).$$

Observe that $F_{s,t}(\overline{x}) = f(x)(F \circ \overline{\chi}_{s,t})(\overline{x})$ with the function $F(\overline{x}) := g(x)z$ and the stochastic flow

$$\overline{\chi}_{s,t}(x, z) := \begin{pmatrix} X_{s,t}(x) \\ \overline{Z}_{s,t}(x)z \end{pmatrix}.$$

Following the proof of Theorem 3.1, we check that

$$dF_{s,t}(\overline{x}) = f(x)\,d(F \circ \overline{\chi}_{s,t})(\overline{x}) = -\big(\mathcal{G}_{s,t}(\overline{x})\,ds + \mathcal{H}_{s,t}(\overline{x})\,d\mathcal{U}_s\big)$$

with the drift function

$$\mathcal{G}_{s,t}(\overline{x}) := f(x)z\left(\nabla\overline{\mathbf{Q}}_{s,t}(g)(x)'a_s(x) + \frac{1}{2}\operatorname{Tr}\big(\nabla^2\overline{\mathbf{Q}}_{s,t}(g)(x)'\alpha_s(x)\big)\right)$$

and the diffusion term

$$\mathcal{H}_{s,t}(\overline{x})\,d\mathcal{U}_s := f(x)z\big(\nabla\overline{\mathbf{Q}}_{s,t}(g)(x)'\sigma_s(x)\,dW_s + \overline{\mathbf{Q}}_{s,t}(g)(x)b_s(x)'\beta_s^{-1}\,dY_s\big).$$

Applying the backward Itô–Ventzell formula [14] we check that

$$dF_{s,t}(\overline{\chi}_s) = (dF_{s,t})(\overline{\chi}_s) + \nabla F_{s,t}(\overline{\chi}_s)'\,d\chi_s + \frac{1}{2}\operatorname{Tr}\big(\nabla^2 F_{s,t}(\chi_s)'\mathcal{A}_t(\overline{\chi}_s)\big)\,ds,$$

from which we conclude that

$$\begin{aligned}
dF_{s,t}(\overline{\chi}_s) = {}& \overline{Z}_s\big(\nabla\big(\overline{\mathbf{Q}}_{s,t}(g)(x)f(x)\big)'\big|_{x=X_s} \\
&\qquad - f(X_s)\overline{Z}_s\nabla\overline{\mathbf{Q}}_{s,t}(g)(X_s)'\big)\sigma_s(X_s)\,dW_s \\
& - f(X_s)\overline{Z}_s\left(\nabla\overline{\mathbf{Q}}_{s,t}(g)(X_s)'a_s(X_s)\,ds + \frac{1}{2}\operatorname{Tr}\big(\nabla^2\overline{\mathbf{Q}}_{s,t}(g)(X_s)\alpha_s(X_s)\big)\right)ds \\
& + \overline{Z}_s\left(\nabla\big(\overline{\mathbf{Q}}_{s,t}(g)(x)f(x)\big)'\big|_{x=X_s}a_s(X_s)ds\right. \\
&\qquad\left. + \frac{1}{2}\operatorname{Tr}\big(\nabla^2\big(\overline{\mathbf{Q}}_{s,t}(g)(x)f(x)\big)'\big|_{x=X_s}\alpha_s(X_s)\big)\right)ds.
\end{aligned}$$

We complete the proof of (4.5) by simple integration.

To take the final step, we recall the integration by parts formula

$$L_t(fg) = fL_t(g) + gL_t(f) + \Gamma_{L_t}(f,g)$$

with the carré-du-champ (also known as the square field) operator Γ_{L_t} associated with the generator L_t defined by $\Gamma_{L_t}(f,g) := (\nabla f)'\alpha_t\nabla g$. Combining (4.5) with the above formula we check that

$$\partial_s\big(\pi_s\big(f\overline{Q}_{s,t}(g)\big)\big) = \pi_s\big(L_s(f)\overline{Q}_{s,t}(g)\big) + \pi_s\big(\Gamma_{L_s}(\overline{Q}_{s,t}(g),f)\big).$$

On the other hand, by integration by parts we have

$$\pi_s\big(\Gamma_{L_s}(\overline{Q}_{s,t}(g),f)\big) = -\sum_{i,j}\int p_s(x)\overline{Q}_{s,t}(g)(x)\frac{1}{p_s(x)}\,\partial_{x_i}\big(p_s(x)\alpha_t^{i,j}\partial_{x_j}f(x)\big)\,dx.$$

This completes the proof of the lemma.

Another approach for finding (4.5) is to use, for any $u \in [s,t]_h$, the decomposition

$$\begin{aligned}
&\pi_{u+h}\big(f\overline{Q}_{u+h,t}(g)\big) - \pi_u\big(f\overline{Q}_{u,t}(g)\big) \\
&\qquad = \pi_u\big(f(\overline{Q}_{u+h,t} - \overline{Q}_{u,t})(g)\big) + (\pi_{u+h} - \pi_u)\big(f\overline{Q}_{u+h,t}(g)\big).
\end{aligned} \tag{4.6}$$

Note that π_u depends on the observations $(Y_s - Y_0)$ from $s = 0$ up to time $s = u$, while the increment $\overline{Q}_{u,t}$ is computed backward in time and depends only on the observations $(Y_s - Y_u)$ from $s > u$ up to $s = t$. Conversely, π_{u+h} depends on the observations $(Y_s - Y_0)$ from $s = 0$ up to time $s = u + h$, while $\overline{Q}_{u+h,t}$ is computed backward in time and depends only on the observations $(Y_s - Y_{u+h})$ from $s > u + h$ up to time $s = t$.

Following the two-sided stochastic integration calculus developed by Pardoux and Protter in [37] (see also [14] for extended versions to interpolating stochastic flows), combining the forward equation (3.8) with the backward equation (3.14), when $h \simeq 0$ we can check the approximation

$$\sum_{u\in[s,t]_h} \big\{\pi_{u+h}\big(f\overline{Q}_{u+h,t}(g)\big) - \pi_u\big(f\overline{Q}_{u,t}(g)\big) - \pi_u\big(L_u(f\overline{Q}_{u+h,t}(g)) - fL_u(\overline{Q}_{u+h,t}(g))\big)h\big\} \simeq 0.$$

4.1. Proof of Theorem 2.1. With the definition in (4.1) we have

$$\pi_{t,s}(dx) = (\pi_t\mathcal{K}_{t,s})(dx) = \pi_s(dx)\,\overline{Q}_{s,t}(1)(x). \tag{4.7}$$

The formulation of the conditional distribution $\pi_{t,s}$ of X_s given $\mathcal{Y}_t$ in (4.7) is rather well known; see, e.g., Theorem 3.7 and Corollary 3.8 in [35], as well as equation (3.9) in [3]. The proof of this formula is a direct consequence of (4.1). With (4.2) we have

$$\pi_t\mathcal{K}_{t,s} = \pi_{t,u}\mathcal{K}_{u,s} = \pi_{t,s}.$$

Thus with $\mathcal{K}_{t,s}$ as defined in (4.1) we immediately have the transport equation (2.1).

It remains to show that this integral operator (as defined in (4.1)) is also the Markov transition kernel of the backward diffusion flow in (2.2). The rest of the proof of Theorem 2.1 is a consequence of the duality formula (4.3) and Lemma 4.1.

Rewritten in a slightly different form, the duality formula (4.3) reads

$$\mathbf{E}\big(f(X_s)g(X_t) \mid \mathcal{Y}_t\big) = \mathbf{E}\big(\mathcal{K}_{t,s}(f)(X_t)g(X_t) \mid \mathcal{Y}_t\big).$$

This implies that

$$\mathcal{K}_{t,s}(f)(X_t) = \mathbf{E}\big(f(X_s) \mid X_t, \mathcal{Y}_t\big).$$

Finally, combining (4.4) with the duality formula (4.3) we have

$$\pi_t\big(g\partial_s\mathcal{K}_{t,s}(f)\big) = -\pi_t\big(g\mathcal{K}_{t,s}(\mathcal{L}_{s,\pi_s}(f))\big).$$

Since the above formula is valid for any test function g, and since π_t has a bounded positive density, we check the backward Kolmogorov equation

$$\partial_s\mathcal{K}_{t,s}(f)(x) = -\mathcal{K}_{t,s}(\mathcal{L}_{s,\pi_s}(f))(x), \tag{4.8}$$

with the terminal condition $\mathcal{K}_{t,t}(f) = f$, when $s = t$, for a.e. $x \in \mathbf{R}^m$ (and a.s. with respect to the law of the observation process from the origin up to the time t). Since both terms in (4.8) are continuous, the above equality holds for any $x \in \mathbf{R}^m$ a.s.

We now complete the proof by showing that the integral operator $\mathcal{K}_{t,s}(x, dz)$ (defined in (4.1)) does indeed coincide with the transition kernel associated with the flow $\mathcal{X}_{t,s}(x)$ in (2.2). First, observe that (4.8) coincides with the backward Kolmogorov equation (2.11) associated with the transition semigroup of the stochastic flow $\mathcal{X}_{t,s}(x)$. Denote this transition semigroup by $\overline{\mathcal{K}}_{t,s}(x, dz)$ temporarily.

By the semigroup properties of $\overline{\mathcal{K}}_{t,s}$, for any $s \leqslant u \leqslant t$ and any smooth function f, we have

$$\partial_u\overline{\mathcal{K}}_{t,s}(f) = 0 = \partial_u\big(\overline{\mathcal{K}}_{t,u}(\overline{\mathcal{K}}_{u,s}(f))\big) = -\overline{\mathcal{K}}_{t,u}(\mathcal{L}_{u,\pi_u}(\overline{\mathcal{K}}_{u,s}f)) + \overline{\mathcal{K}}_{t,u}(\partial_u\overline{\mathcal{K}}_{u,s}(f)).$$

Choosing $u = t$ we obtain the forward equation $\partial_t \overline{\mathcal{K}}_{t,s}(f) = \mathcal{L}_{t,\pi_t}(\overline{\mathcal{K}}_{t,s}(f))$. Arguing as above, this implies that

$$\partial_u\big(\mathcal{K}_{t,u}(\overline{\mathcal{K}}_{u,s}(f))\big) = -\mathcal{K}_{t,u}(\mathcal{L}_{u,\pi_u}(\overline{\mathcal{K}}_{u,s}f)) + \mathcal{K}_{t,u}(\mathcal{L}_{u,\pi_u}(\overline{\mathcal{K}}_{u,s}(f))) = 0.$$

Integrating over the interval $[s,t]$, we check that $\mathcal{K}_{t,s} = \overline{\mathcal{K}}_{t,s}$. This completes the proof of Theorem 2.1.

REFERENCES

[1] B. D. O. Anderson, *Fixed interval smoothing for nonlinear continuous time systems*, Information and Control, 20 (1972), pp. 294–300, https://doi.org/10.1016/S0019-9958(72)90451-2.

[2] B. D. O. Anderson, *Reverse-time diffusion equation models*, Stochastic Process. Appl., 12 (1982), pp. 313–326, https://doi.org/10.1016/0304-4149(82)90051-5.

[3] B. D. O. Anderson and I. B. Rhodes, *Smoothing algorithms for nonlinear finite-dimensional systems*, Stochastics, 9 (1983), pp. 139–165, https://doi.org/10.1080/17442508308833251.

[4] M. Arnaudon and P. Del Moral, *A variational approach to nonlinear and interacting diffusions*, Stoch. Anal. Appl., 37 (2019), pp. 717–748, https://doi.org/10.1080/07362994.2019.1609985.

[5] A. Bain and D. Crisan, *Fundamentals of Stochastic Filtering*, Stoch. Model. Appl. Probab. 60, Springer, New York, 2009, https://doi.org/10.1007/978-0-387-76896-0.

[6] A. N. Bishop and P. Del Moral, *On the stability of Kalman–Bucy diffusion processes*, SIAM J. Control Optim., 55 (2017), pp. 4015–4047, https://doi.org/10.1137/16M1102707; preprint version available at https://arxiv.org/abs/1610.04686.

[7] A. E. Bryson and M. Frazier, *Smoothing for linear and nonlinear dynamic systems*, in Proceedings of the Optimum System Synthesis Conference (ASD, 1962), Tech. rep. ASD-TDR-63-119, Aeronautical Systems Div., Wright-Patterson AFB, Ohio, 1963, pp. 353–364.

[8] P. Cattiaux and L. Mesnager, *Hypoelliptic non-homogeneous diffusions*, Probab. Theory Related Fields, 123 (2002), pp. 453–483, https://doi.org/10.1007/s004400100194.

[9] W.-L. Chow, *Über Systeme von linearen partiellen Differentialgleichungen erster Ordnung*, Math. Ann., 117 (1939), pp. 98–105, https://doi.org/10.1007/BF01450011.

[10] T. Cass, M. Clark, and D. Crisan, *The filtering equations revisited*, in Stochastic Analysis and Applications 2014, Springer Proc. Math. Stat. 100, Springer, Cham, 2014, pp. 129–162, https://doi.org/10.1007/978-3-319-11292-3_5.

[11] G. Da Prato, J.-L. Menaldi, and L. Tubaro, *Some results of backward Itô formula*, Stoch. Anal. Appl., 25 (2007), pp. 679–703, https://doi.org/10.1080/07362990701283045.

[12] G. Da Prato, *Some remarks about backward Itô formula and applications*, Stoch. Anal. Appl., 16 (1998), pp. 993–1003, https://doi.org/10.1080/07362999808809576.

[13] P. Del Moral and S. Penev, *Stochastic Processes: From Applications to Theory*, Chapman & Hall/CRC Texts Stat. Sci. Ser., CRC Press, Boca Raton, FL, 2017.

[14] P. Del Moral and S. S. Singh, *A Forward-Backward Stochastic Analysis of Diffusion Flows*, preprint, 2019, https://arxiv.org/abs/1906.09145v1.

[15] U. G. Haussmann and E. Pardoux, *Time reversal of diffusions*, Ann. Probab., 14 (1986), pp. 1188–1205, https://doi.org/10.1214/aop/1176992362.

[16] L. Hörmander, *Hypoelliptic second order differential equations*, Acta Math., 119 (1967), pp. 147–171, https://doi.org/10.1007/BF02392081.

[17] T. Kailath and P. Frost, *An innovations approach to least-squares estimation: Part* II*: Linear smoothing in additive white noise*, IEEE Trans. Automat. Control, AC-13 (1968), pp. 655–660, https://doi.org/10.1109/TAC.1968.1099019.

[18] G. Kallianpur and C. Striebel, *Estimation of stochastic systems: Arbitrary system process with additive white noise observation errors*, Ann. Math. Statist., 39 (1968), pp. 785–801, https://doi.org/10.1214/aoms/1177698311.

[19] G. Kallianpur and C. Striebel, *Stochastic differential equations occurring in the estimation of continuous parameter stochastic processes*, Theory Probab. Appl., 14 (1969), pp. 567–594, https://doi.org/10.1137/1114076.

[20] N. V. Krylov and B. L. Rozovskii, *On the Cauchy problem for linear stochastic partial differential equations*, Math. USSR-Izv., 11 (1977), pp. 1267–1284, https://doi.org/10.1070/IM1977v011n06ABEH001768.

[21] N. V. Krylov and B. L. Rozovskii, *On conditional distributions of diffusion processes*, Math. USSR-Izv., 12 (1978), pp. 336–356, https://doi.org/10.1070/IM1978v012n02ABEH001857.
[22] N. V. Krylov and B. L. Rozovskiĭ, *On the first integrals and Liouville equations for diffusion processes*, in Stochastic Differential Systems (Visegrád, 1980), Lect. Notes Control Inf. Sci. 36, Springer, Berlin, 1981, pp. 117–125, https://doi.org/10.1007/BFb0006415.
[23] H. Kunita, *Asymptotic behavior of the nonlinear filtering errors of Markov processes*, J. Multivariate Anal., 1 (1971), pp. 365–393, https://doi.org/10.1016/0047-259X(71)90015-7.
[24] H. Kunita, *Stochastic partial differential equations connected with non-linear filtering*, in Nonlinear Filtering and Stochastic Control, (Cortona, 1981), Lecture Notes in Math. 972, Springer, Berlin, 1982, pp. 100–169, https://doi.org/10.1007/BFb0064861.
[25] H. Kunita, *On backward stochastic differential equations*, Stochastics, 6 (1982), pp. 293–313, https://doi.org/10.1080/17442508208833209.
[26] H. Kunita, *First order stochastic partial differential equations*, in Stochastic Analysis (Katata/Kyoto, 1982), North-Holland Math. Library 32, North-Holland, Amsterdam, 1984, pp. 249–269, https://doi.org/10.1016/S0924-6509(08)70396-9.
[27] H. J. Kushner, *On the differential equations satisfied by conditional probability densities of Markov processes, with applications*, J. SIAM Control Ser. A, 2 (1964), pp. 106–119, https://doi.org/10.1137/0302009.
[28] C. T. Leondes, J. B. Peller, and E. B. Stear, *Nonlinear smoothing theory*, IEEE Trans. Syst. Sci. Cybern., 6 (1970), pp. 63–71, https://doi.org/10.1109/TSSC.1970.300330.
[29] J. S. Meditch, *A survey of data smoothing for linear and nonlinear dynamic systems*, Automatica J. IFAC, 9 (1973), pp. 151–162 (in French), https://doi.org/10.1016/0005-1098(73)90070-8.
[30] D. Michel, *Régularité des lois conditionnelles en théorie du filtrage non-linéaire et calcul des variations stochastique*, J. Funct. Anal., 41 (1981), pp. 8–36, https://doi.org/10.1016/0022-1236(81)90059-8.
[31] A. Millet, D. Nualart, and M. Sanz, *Integration by parts and time reversal for diffusion processes*, Ann. Probab., 17 (1989), pp. 208–238, https://doi.org/10.1214/aop/1176991505.
[32] E. Pardoux, *Equations aux Dérivées Partielles Stochastiques Non Linéaires Monotones*, Ph.D. thesis, Univ. Paris XI, Orsay, 1975.
[33] E. Pardoux, *Stochastic partial differential equations and filtering of diffusion processes*, Stochastics, 3 (1979), pp. 127–167, https://doi.org/10.1080/17442507908833142.
[34] E. Pardoux, *Non-linear filtering, prediction and smoothing*, in Stochastic Systems: The Mathematics of Filtering and Identification and Applications (Les Arcs, 1980), Nato Sci. Ser. C 78, Reidel, Dordrecht, 1981, pp. 529–557, https://doi.org/10.1007/978-94-009-8546-9.
[35] E. Pardoux, *Équations du filtrage non linéaire de la prédiction et du lissage*, Stochastics, 6 (1982), pp. 193–231, https://doi.org/10.1080/17442508208833204.
[36] E. Pardoux and S. G. Peng, *Adapted solution of a backward stochastic differential equation*, Systems Control Lett., 14 (1990), pp. 55–61, https://doi.org/10.1016/0167-6911(90)90082-6.
[37] E. Pardoux and P. Protter, *A two-sided stochastic integral and its calculus*, Probab. Theory Related Fields, 76 (1987), pp. 15–49, https://doi.org/10.1007/BF00390274.
[38] H. E. Rauch, F. Tung, and C. T. Striebel, *Maximum likelihood estimates of linear dynamic systems*, AIAA J., 3 (1965), pp. 1445–1450, https://doi.org/10.2514/3.3166.
[39] K. Rashevskii, *On joining any two points of a nonholonomic space by an admissible line*, Uchen. Zap. Pedag. Inst. Liebkhecht Ser. Fiz.–Mat., 3 (1938), pp. 83–94 (in Russian).
[40] M. Rutkowski, *A simple proof for the Kalman–Bucy smoothed estimate formula*, Statist. Probab. Lett., 17 (1993), pp. 377–385, https://doi.org/10.1016/0167-7152(93)90258-K.
[41] J. M. Steele, *Stochastic Calculus and Financial Applications*, corr. reprint of the 1st ed., Appl. Math. (N.Y.) 45, Springer-Verlag, New York, 2012, https://doi.org/10.1007/978-1-4684-9305-4.
[42] A. Yu. Veretennikov, *On backward filtering equations for SDE systems (direct approach)*, in Stochastic Partial Differential Equations (Edinburgh, 1994), London Math. Soc. Lecture Note Ser. 216, Cambridge Univ. Press, Cambridge, 1995, pp. 304–311, https://doi.org/10.1017/CBO9780511526213.019.
[43] T. Yang, P. G. Mehta, and S. P. Meyn, *Feedback particle filter*, IEEE Trans. Automat. Control, 58 (2013), pp. 2465–2480, https://doi.org/10.1109/TAC.2013.2258825.
[44] T. Yang, R. S. Laugesen, P. G. Mehta, and S. P. Meyn, *Multivariable feedback particle filter*, Automatica J. IFAC, 71 (2016), pp. 10–23, https://doi.org/10.1016/j.automatica.2016.04.019.
[45] M. Zakai, *On the optimal filtering of diffusion processes*, Z. Wahrsch. Verw. Gebiete, 11 (1969), pp. 230–243, https://doi.org/10.1007/BF00536382.