LMS J. Comput. Math. 0 (2011) 1–17 © 2011 Authors
doi:10.1112/S146115701000029X

Efficient nonlinear filtering of a singularly perturbed stochastic hybrid system

Jun H. Park, Boris Rozovskii and Richard B. Sowers

Abstract

Our focus in this work is to investigate an efficient state estimation scheme for a singularly perturbed stochastic hybrid system. As stochastic hybrid systems have been used recently in diverse areas, the importance of correct and efficient estimation of such systems cannot be overemphasized. The framework of nonlinear filtering provides a suitable ground for on-line estimation. With the help of intrinsic multiscale properties of a system, we obtain an efficient estimation scheme for a stochastic hybrid system.

1. *Introduction*

The theory of filtering gives a recursive procedure for estimating an evolving signal or state from a noisy observation process. Since the state is usually hidden and evolves according to its own dynamics, the objective is to compute the conditional distribution of the state given noisy observations. Aside from several special cases where the distribution of the state can be described with a finite number of moments or modes (for example, the celebrated Kalman filter [**11**] for linear systems and Beneš [**2**] and Daum [**4**] filters for special nonlinear systems), filtering problems in general deal with infinite-dimensional objects such as stochastic partial differential equations (PDEs) for posterior densities and, thus, require enormous amounts of computation.

In this work, our interest lies in a filtering problem for stochastic hybrid systems. As stochastic hybrid systems have been used in diverse areas to model complex random phenomena which were not captured by state models with either continuous or discrete dynamics alone, their estimation has become an active research field for the last decade (cf. [**9**, **13**, **19**, **20**, **24**] and references therein).

The computation required to solve multidimensional nonlinear filtering problems might be quite intensive. This may hinder the practical implementation of stochastic hybrid systems in time-critical applications such as target tracking, fault detection, volatility estimation in financial markets, etc. However, if the system can be cast into a multiscale setting, significant reduction in the computational complexity may be available.

Singularly perturbed dynamical systems are a natural framework for dealing with multiscale systems [**12**, **14**, **21**]. In several riveting areas including biology [**3**], optimal control [**6**], and finance [**26**], various phenomena have been successfully modeled by singularly perturbed stochastic hybrid systems.

We here consider nonlinear filtering for continuous–discrete state processes given by a pair of fast–slow processes. More specifically, we choose a fast diffusion process with a slow switching process [**24**]. Both the fast and slow processes are coupled so that neither process on its own is Markovian.

More specifically, we will consider a state process $(X_t^\varepsilon, \Theta_t^\varepsilon)$, where ε is a small parameter, X_t^ε is the $\mathbb{R}^d$-valued diffusion process governed by the following stochastic differential equation

Received 17 August 2010; revised 14 March 2011.

2000 Mathematics Subject Classification 60H15 (primary), 35R60, 60G35, 34C29, 70K65 (secondary).

(SDE):

$$dX_t^\varepsilon = -\frac{1}{\varepsilon}\Lambda_{\Theta_t^\varepsilon}(X_t^\varepsilon - \Theta_t^\varepsilon)\,dt + \frac{1}{\sqrt{\varepsilon}}\,dW_t, \tag{1.1}$$

and Θ_t^ε is a (continuous-time) conditionally Markov process taking values in a finite set S. Namely,

$$\mathbb{P}\{\Theta_{t+\Delta}^\varepsilon = \theta_1 \mid \Theta_t^\varepsilon = \theta_2, X_t^\varepsilon = x\} = q_{\theta_1,\theta_2}(x)\Delta + o(\Delta),$$

where

$$q_{\theta_1,\theta_2}(x) \geqslant 0 \quad \text{if } \theta_1 \neq \theta_2,$$
$$q_{\theta,\theta}(x) = -\sum\nolimits_{\substack{\theta' \in S\\ \theta' \neq \theta}} q_{\theta,\theta'}(x), \quad \theta \in S.$$

In the literature, the process Θ_t^ε is often called the parameter process or simply the 'parameter' and we will use this term as well.

We assume that the $q_{\theta,\theta'}$ are all bounded and measurable. We also assume that for each $\theta \in S$, the eigenvalues of the matrix $\Lambda_\theta = (\lambda^\theta_{i,i'})_{1\leqslant i,i'\leqslant d}$ are all strictly positive, and dW in (1.1) is a d-dimensional Brownian motion†.

Note that under these assumptions, the distribution of X_t^ε quickly relaxes to a locally 'invariant' distribution centered at the current value of the parameter process. Parameter process Θ_t^ε evolves on a much slower scale, and its dynamics depend on X^ε. The small parameter ε measures the ratio of slow and fast scales.

We observe a corrupted function of X^ε; that is, the n-dimensional observation process Y^ε is given by

$$dY_t^\varepsilon = h(X_t^\varepsilon, \Theta_t^\varepsilon)\,dt + dV_t \tag{1.2}$$

for some bounded and continuous sensor function h from $\mathbb{R}^d \times S$ to $\mathbb{R}^n$ and where V is a standard n-dimensional Brownian motion. We also assume that $(X_0^\varepsilon, \Theta_0^\varepsilon)$ is independent of the other sources of randomness in our system and that there is a $\rho \in C_0(\mathbb{R}^d \times S)$ such that

$$\mathbb{P}\{X_0^\varepsilon \in A, \Theta_0^\varepsilon \in A'\} = \sum_{\theta \in A'} \int_{x\in A} \rho(x,\theta)\,dx$$

for all $A \in \mathscr{B}(\mathbb{R}^d)$ and $A' \subset S$. Let us also assume that $Y_0^\varepsilon = 0$. Based on the observation process, we want to reconstruct the law of Θ^ε; that is, to compute

$$\mathbb{P}\{\Theta_t^\varepsilon \in A \mid \mathscr{Y}_t^\varepsilon\},$$

where

$$\mathscr{Y}_t^\varepsilon \overset{\text{def}}{=} \sigma\{Y_s^\varepsilon : 0 \leqslant s \leqslant t\}.$$

We want to do this *efficiently*; that is, to find an effective filter which works as the scaling parameter $\varepsilon \searrow 0$.

The standard equations of filtering (which we will develop in a moment) require us to evolve a conditional law for the *full* state; that is, the pair $(X^\varepsilon, \Theta^\varepsilon)$. Since the fast X^ε quickly relaxes to its local invariant measure, it should not have too much information. Our objective is to show that we can track Θ^ε without fully resolving the conditional density of X^ε. Thus, instead of solving an $\mathbb{R}^d \times S$-dimensional Zakai equation, we can effectively solve an approximate Zakai equation whose state space is the finite set S. The resulting equation can be used in place of the original more complex equations to provide qualitatively accurate and computationally feasible descriptions either for simulation and prediction or for real-time control.

† By means of various coordinate changes, we can transform the problem so that dW can have any positive-definite covariance matrix, which may, in fact, depend on Θ_t^ε. Note that we have included Θ_t^ε-dependence in our sensor function in (1.2).

This type of result has been covered in the literature within the framework of homogenization theory; we refer to [16] and the references therein for more detail. Methodologically, this study is similar to [15] and [16]. In the former work, the observation becomes independent of the system in the limit, while the latter has an explicit dependence on the slow variable in the limit. In [15, 16], the fast motion was a fast angular drift. In contrast to these papers, the fast motion (1.1) has both drift and diffusion, so the speed of averaging depends on a spectral gap. In this work, we show that the observation in the limit still has crucial information for estimation even though there is no explicit dependency on the system. This property is quite useful in many practical applications such as molecular motors [22] and rare-event simulations, where the observation could be given in terms of the fast variable. To the authors' knowledge, there is no previous work in this setting.

2. *The Zakai equation*

Our first step is to recall the known framework of nonlinear filtering; that is, the Zakai equation [1, 18, 25]. Let us start with the generator of the fast motion for a fixed value of the parameter. For each $\theta \in S$, define the second-order partial differential operator

$$(\mathscr{L}_\theta f)(x) \stackrel{\text{def}}{=} -\sum\nolimits_{1\leqslant i,j\leqslant d} \lambda^\theta_{i,j}(x_j-\theta_j)\frac{\partial f}{\partial x_i}(x) + \frac{1}{2}\sum_{1\leqslant i\leqslant d}\frac{\partial^2 f}{\partial x_i^2}(x)$$

for $f \in C^\infty(\mathbb{R}^d)$ and $x = (x_1, x_2, \dots, x_d)$. We also define the generator of the parameter process as

$$(\mathscr{Q} f)(x,\theta) \stackrel{\text{def}}{=} \sum_{\theta'\in S} f(\theta')q_{\theta,\theta'}(x)$$

for all $f \in B(\mathbb{R}^d \times S)$, $x \in \mathbb{R}^d$, and $\theta \in S$. These are the generators of the fast and slow motions, and we propagate densities by their adjoints; define

$$(\mathscr{L}^*_\theta f)(x) \stackrel{\text{def}}{=} \sum_{1\leqslant i,j\leqslant d} \lambda^\theta_{i,j}\frac{\partial}{\partial x_i}((x_j-\theta_j)f)(x) + \frac{1}{2}\sum_{1\leqslant i\leqslant d}\frac{\partial^2 f}{\partial x_i^2}(x)$$

for $f \in C^\infty(\mathbb{R}^d)$ and $x = (x_1, x_2 \dots x_d)$ and

$$(\mathscr{Q}^* f)(x,\theta) \stackrel{\text{def}}{=} \sum_{\theta'\in S} f(\theta')q_{\theta',\theta}(x)$$

for all $f \in B(\mathbb{R}^d \times S)$, $x \in \mathbb{R}^d$, and $\theta \in S$.

The Zakai equation in our setting is given by

$$\begin{aligned} du^\varepsilon(t,x,\theta) &= \frac{1}{\varepsilon}\mathscr{L}^*_\theta u^\varepsilon(t,x,\theta)\,dt + \mathscr{Q}^* u^\varepsilon(t,x,\theta)\,dt + u^\varepsilon(t,x,\theta)h(x,\theta)^T dY^\varepsilon_t \\ &= \frac{1}{\varepsilon}\mathscr{L}^*_\theta u^\varepsilon(t,x,\theta)\,dt + \mathscr{Q}^* u^\varepsilon(t,x,\theta)\,dt + u^\varepsilon(t,x,\theta)h(x,\theta)^T\,dV_t \\ &\quad + u^\varepsilon(t,x,\theta)h(x,\theta)^T h(X^\varepsilon_t,\Theta^\varepsilon_t)\,dt, \\ u^\varepsilon(0,x,\theta) &= \rho(x,\theta). \end{aligned} \tag{2.1}$$

Under the assumptions of this paper, one could show that for every $\varepsilon > 0$,

$$\mathbb{P}\{X^\varepsilon_t \in A, \Theta^\varepsilon_t \in A' \mid \mathscr{Y}^\varepsilon_t\} = \frac{\sum_{\theta\in A'}\int_{x\in A} u^\varepsilon(t,x,\theta)\,dx}{\sum_{\theta\in S}\int_{x\in\mathbb{R}^d} u^\varepsilon(t,x,\theta)\,dx}, \quad A \in \mathscr{B}(\mathbb{R}^d),\ A' \subset S$$

with probability 1.

Note that in the literature on nonlinear filtering the Zakai equation is usually considered on a new probability space, where Y^ε_t is a Brownian motion. This space changes when the parameter ε changes. This is inconvenient for our purposes. Therefore, we will consider (2.1) on a fixed probability space for all ε.

To see the asymptotic behavior of the solution of (2.1) as $\varepsilon \searrow 0$, we construct the invariant measure of the fast motion. For $\theta \in S$, define

$$B_\theta \overset{\text{def}}{=} \int_{s=0}^{\infty} \exp[-\Lambda_\theta s] \exp[-\Lambda_\theta^T s]\, ds$$

and

$$\mu_\theta(x) \overset{\text{def}}{=} \frac{1}{\sqrt{(2\pi)^d \det B_\theta}} \exp\left[-\frac{1}{2}(x-\theta)' B_\theta^{-1}(x-\theta)\right], \quad x \in \mathbb{R}^d$$

Then

$$\mathscr{L}_\theta^* \mu_\theta(x) = 0 \quad \text{and} \quad \int_{\mathbb{R}^d} \mu_\theta(x)\, dx = 1$$

for all $\theta \in S$. We want to find the effective behavior of the jumps by averaging over the invariant distribution of the fast motion. For each θ_1 and θ_2 in S, define

$$\bar{q}_{\theta_1,\theta_2} \overset{\text{def}}{=} \int_{x\in\mathbb{R}^d} q_{\theta_1,\theta_2}(x)\mu_{\theta_1}(x)\, dx;$$

we still have that $\bar{q}_{\theta_1,\theta_2} \geqslant 0$ if $\theta_1 \neq \theta_2$ and $\bar{q}_{\theta,\theta} = -\sum_{\substack{\theta'\in S\\ \theta'\neq\theta}} \bar{q}_{\theta,\theta'}$ for all $\theta \in S$. Define

$$(\bar{\mathscr{Q}}^* f)(\theta) \overset{\text{def}}{=} \sum_{\theta'\in S} f(\theta')\bar{q}_{\theta',\theta}$$

for all $f \in B(S)$. We also need to average the sensor function and the initial condition; for each $\theta \in S$, define

$$\bar{h}(\theta) \overset{\text{def}}{=} \int_{x\in\mathbb{R}^d} h(x,\theta)\mu_\theta(x)\, dx \quad \text{and} \quad \bar{\rho}(\theta) \overset{\text{def}}{=} \int_{x\in\mathbb{R}^d} \rho(x,\theta)\mu_\theta(x)\, dx. \tag{2.2}$$

The effective Zakai equation for θ^ε should be given by averaging the coefficients of (2.1) with respect to the invariant measure of the fast motion; that is,

$$\begin{gathered} dv^\varepsilon(t,\theta) = \bar{\mathscr{Q}}^* v^\varepsilon(t,\theta)\, dt + v^\varepsilon(t,\theta)\bar{h}(\theta)^T dY_t^\varepsilon, \\ v^\varepsilon(0,\theta) = \bar{\rho}(\theta) \end{gathered} \tag{2.3}$$

(see also Il'in *et al.* [10]). Note that while we can find the effective behavior of the x variable in the coefficients of the Zakai equation (2.1), we cannot really average the observations since they are the inputs to the system. Thus, (2.3) is not a true Zakai equation; this is clear upon writing

$$\begin{aligned} dv^\varepsilon(t,\theta) &= \bar{\mathscr{Q}}^* v^\varepsilon(t,\theta)\, dt + v^\varepsilon(t,\theta)\bar{h}(\theta)^T \{h(X_t^\varepsilon,\Theta_t^\varepsilon)\, dt + dV_t\} \\ &= \bar{\mathscr{Q}}^* v^\varepsilon(t,\theta)\, dt + v^\varepsilon(t,\theta)\bar{h}(\theta)^T \{\bar{h}(\Theta_t^\varepsilon)\, dt + dV_t\} \\ &\quad + v^\varepsilon(t,\theta)\bar{h}(\theta)^T \{h(X_t^\varepsilon,\Theta_t^\varepsilon) - \bar{h}(\Theta_t^\varepsilon)\}\, dt. \end{aligned}$$

The last term captures the deviation from a true Zakai equation.

Let us now collect our thoughts and formulate our results. For each $t > 0$, define

$$\pi_t^\varepsilon(A) \overset{\text{def}}{=} \frac{\sum_{\theta\in A}\int_{x\in\mathbb{R}^d} u^\varepsilon(t,x,\theta)\, dx}{\sum_{\theta\in S}\int_{x\in\mathbb{R}^d} u^\varepsilon(t,x,\theta)\, dx}; \quad A \subset S;$$

then $\pi_t^\varepsilon(A) = \mathbb{P}\{\Theta_t^\varepsilon \in A \mid \mathscr{Y}_t^\varepsilon\}$. Let us also define

$$\bar{\pi}_t^\varepsilon(A) \overset{\text{def}}{=} \frac{\sum_{\theta\in A} v^\varepsilon(t,\theta)}{\sum_{\theta\in S} v^\varepsilon(t,\theta)}; \quad A \subset S.$$

We note that the evolution of v is essentially an $|S|$-dimensional SDE, whereas that of u^ε is a stochastic partial differential equation (SPDE). Thus, $\bar{\pi}_t^\varepsilon$ is a much simpler process to compute.

Note also that the only dependence of $\bar{\pi}^\varepsilon$ on ε is through the observation process Y^ε; the actual *dynamics* are ε-independent. Our main result is that as $\varepsilon \searrow 0$, $\bar{\pi}^\varepsilon$ is a good substitute for π^ε.

THEOREM 2.1. *For each $t > 0$,*

$$\lim_{\varepsilon\to 0} \mathbb{E}[d_{\mathscr{P}(S)}(\pi_t^\varepsilon, \bar{\pi}_t^\varepsilon)] = 0.$$

Here $d_{\mathscr{P}(S)}$ is the Prohorov metric on $\mathscr{P}(S)$.

3. *Asymptotic analysis*

Several preliminary steps will help make the proof of Theorem 2.1 more natural. Firstly, we will consider a slightly more intrinsic formulation of the Zakai equation. Secondly, we will rewrite (2.3) in a way which facilitates comparison to the dynamics of the original problem.

To begin, let us define

$$\tilde{u}^\varepsilon(t,x,\theta) \overset{\text{def}}{=} \frac{u^\varepsilon(t,x,\theta)}{\mu_\theta(x)}; \quad t \geqslant 0, x \in \mathbb{R}^d, \theta \in \mathbb{R}^d;$$

then

$$\int_{x\in\mathbb{R}^d} u^\varepsilon(t,x,\theta) f(x)\, dx = \int_{x\in\mathbb{R}^d} \tilde{u}^\varepsilon(t,x,\theta) f(x)\mu_\theta(x)\, dx, \quad t \geqslant 0, \theta \in S$$

for all $f \in C_c^\infty(\mathbb{R}^d)$ and

$$\pi_t^\varepsilon(A) \overset{\text{def}}{=} \frac{\sum_{\theta\in A}\int_{x\in\mathbb{R}^d} \tilde{u}^\varepsilon(t,x,\theta)\mu_\theta(x)\, dx}{\sum_{\theta\in S}\int_{x\in\mathbb{R}^d} \tilde{u}^\varepsilon(t,x,\theta)\mu_\theta(x)\, dx} \quad A \subset S.$$

In other words, let us make our reference measure the invariant measure of the fast motion. If we define

$$\tilde{\mathscr{L}}_\theta^* f \overset{\text{def}}{=} \frac{1}{\mu_\theta}(\mathscr{L}_\theta^*(f\mu_\theta))$$

for $f \in C^\infty(\mathbb{R}^d)$ and $x = (x_1, x_2, \ldots, x_d)$ and

$$(\tilde{\mathscr{Q}}^* f)(x,\theta) \overset{\text{def}}{=} \frac{1}{\mu_\theta(x)} \sum_{\theta'\in S} \mu_{\theta'}(x) f(\theta') q_{\theta',\theta}(x)$$

for all $f \in B(\mathbb{R}^d \times S)$, $x \in \mathbb{R}^d$, and $\theta \in S$, we then have that

$$\begin{aligned} d\tilde{u}^\varepsilon(t,x,\theta) &= \frac{1}{\varepsilon}\tilde{\mathscr{L}}_\theta^*\tilde{u}^\varepsilon(t,x,\theta)\, dt + \tilde{\mathscr{Q}}^*\tilde{u}^\varepsilon(t,x,\theta)\, dt + \tilde{u}^\varepsilon(t,x,\theta)h(x)^T\, dY_t^\varepsilon \\ &= \frac{1}{\varepsilon}\tilde{\mathscr{L}}_\theta^*\tilde{u}^\varepsilon(t,x,\theta)\, dt + \tilde{\mathscr{Q}}^*\tilde{u}^\varepsilon(t,x,\theta)\, dt + \tilde{u}^\varepsilon(t,x,\theta)h(x)^T h(X_t^\varepsilon,\Theta_t^\varepsilon)\, dt \\ &\quad + \tilde{u}^\varepsilon(t,x,\theta)h(x)^T\, dV_t, \\ \tilde{u}^\varepsilon(0,x,\theta) &= \frac{\rho(x,\theta)}{\mu_\theta(x)}. \end{aligned} \tag{3.1}$$

We next observe that v^ε satisfies a similar SPDE. Since μ_θ is the invariant measure for the generator $\mathscr{L}_\theta$, $\mathscr{L}_\theta^*\mu_\theta \equiv 0$. Defining $\mathbf{1}: \mathbb{R}^d \to \mathbb{R}$ as $\mathbf{1}: \mathbb{R}^d \mapsto 1$, we thus have that $\tilde{\mathscr{L}}_\theta^*\mathbf{1} \equiv 0$. Hence, the function $(t,x,\theta) \mapsto v^\varepsilon(t,\theta) = v^\varepsilon(t,\theta)\mathbf{1}(x)$ from $\mathbb{R}_+ \times \mathbb{R}^d \times S$ satisfies

$$\begin{gathered} dv^\varepsilon(t,\theta) = \frac{1}{\varepsilon}(\tilde{\mathscr{L}}_\theta^* v^\varepsilon)(t,x,\theta)\, dt + \bar{\mathscr{Q}}^* v^\varepsilon(t,\theta)\, dt + v^\varepsilon(t,\theta)\bar{h}(\theta)^T\, dY_t^\varepsilon, \\ v^\varepsilon(0,\theta) = \bar{\rho}(\theta). \end{gathered}$$

We also note, of course, that

$$\bar{\pi}^{\varepsilon}_t(A) = \frac{\sum_{\theta\in A}\int_{x\in\mathbb{R}^d} v^{\varepsilon}(t,\theta)\mu_\theta(x)\,dx}{\sum_{\theta\in S}\int_{x\in\mathbb{R}^d} v^{\varepsilon}(t,\theta)\mu_\theta(x)\,dx}; \quad A\subset S.$$

Our immediate goal is then the following lemma.

LEMMA 3.1. *For each $t>0$, we have that*

$$\lim_{\varepsilon\searrow 0}\sum_{\theta\in S}\mathbb{E}\left[\int_{x\in\mathbb{R}^d}|\tilde{u}^{\varepsilon}(t,x,\theta)-v^{\varepsilon}(t,\theta)|^2\mu_\theta(x)\,dx\right]=0.$$

This will be a crucial step towards the proof of Theorem 2.1; see Section 4.

The value of the linear dynamics of (1.1) is that they allow us to get explicit rates at which the fast motion achieves its stationary behavior (if the dynamics of X^ε had nonlinearities, one could in general get only abstract bounds on the rate of convergence to a stationary distribution). To formalize the notation surrounding this, fix $\theta\in S$. For each $x\in\mathbb{R}^d$, define

$$\tilde{X}^{\theta,x}_t \overset{\text{def}}{=} \theta+\exp[-\Lambda_\theta t]x+\int_{s=0}^t \exp[-\Lambda_\theta(t-s)]\,dW_s, \quad t>0.$$

Then $\tilde{X}^{\theta,x}$ satisfies the SDE

$$\begin{aligned} d\tilde{X}^{\theta,x}_t &= -\Lambda_\theta(\tilde{X}^{\theta,x}_t-\theta)\,dt+dW_t,\\ \tilde{X}^{\theta,x}_0 &= x; \end{aligned} \tag{3.2}$$

this is a Markov process with generator $\mathscr{L}_\theta$. For $t>0$ and $f\in B(\mathbb{R}^d)$, define

$$(P^\theta_t f)(x) \overset{\text{def}}{=} \mathbb{E}[f(\tilde{X}^{\theta,x}_t)].$$

This is the semigroup on $B(\mathbb{R}^d)$ generated by $\mathscr{L}_\theta$ (and of course $\lim_{t\searrow 0}P^\theta_t f=f$ pointwise). We can write a kernel representation for P^θ_t. For every $t>0$, define

$$B_\theta(t) \overset{\text{def}}{=} \int_{s=0}^t \exp[-\Lambda_\theta(t-s)]\exp[-\Lambda_\theta^T(t-s)]\,ds = \int_{s=0}^t \exp[-\Lambda_\theta s]\exp[-\Lambda^T_\theta s]\,ds.$$

Of course, $\lim_{t\to\infty}B_\theta(t)=B_\theta$. Define

$$\begin{aligned} p^\theta_x(t,z) &\overset{\text{def}}{=} (2\pi)^{-d/2}(\det B_\theta(t))^{-1/2}\\ &\quad\times\exp[-\tfrac{1}{2}(z-(\theta+\exp[-\Lambda_\theta t]x))^T B^{-1}_\theta(t)(z-(\theta+\exp[-\Lambda_\theta t]x))] \end{aligned}$$

for all $t>0$ and x and z in $\mathbb{R}^d$. Then

$$(P^\theta_t f)(x)=\int_{z\in\mathbb{R}^d}p^\theta_x(t,z)f(z)\,dz$$

for all $t>0$, $x\in\mathbb{R}^d$, and $f\in B(\mathbb{R}^d)$. For each $f\in C^\infty_c(\mathbb{R}^d)$, let us next define

$$(S^\theta_t f)(x) \overset{\text{def}}{=} \frac{1}{\mu_\theta(x)}\int_{z\in\mathbb{R}^d} f(z)p^\theta_z(t,x)\mu_\theta(z)\,dz$$

for all $t>0$ and $x\in\mathbb{R}^d$, so that

$$\int_{x\in\mathbb{R}^d}(S^\theta_t f)(x)g(x)\mu_\theta(x)\,dx=\int_{z\in\mathbb{R}^d}f(z)\mu_\theta(z)(P^\theta_t g)(z)\,dz \tag{3.3}$$

for all f and g in $C^\infty_c(\mathbb{R}^d)$ and all $t>0$. In fact, we should think of S^θ_t as the adjoint of P^θ_t. For all f and g in $C_c(\mathbb{R}^d)$, define

$$\langle f,g\rangle_\theta \overset{\text{def}}{=} \int_{x\in\mathbb{R}^d}f(x)g(x)\mu_\theta(x)\,dx$$

and define $\|f\|_\theta \overset{\text{def}}{=} \sqrt{\langle f, f\rangle_\theta}$ for all $f \in C_c(\mathbb{R}^d)$; this is a norm on $C_c(\mathbb{R}^d)$. Define $L^2(\mu_\theta)$ as the closure of $C_c(\mathbb{R}^d)$ with respect to this norm; then $\langle \cdot, \cdot\rangle_\theta$ can uniquely be extended to an inner product on $L^2(\theta)$. Then (3.3) can be written as

$$\langle S^\theta_t f, g\rangle_\theta = \langle f, P^\theta_t g\rangle_\theta$$

for all f and g in $C_c(\mathbb{R}^d)$ and all $t > 0$.

LEMMA 3.2. *Fix $f \in C^\infty_c(\mathbb{R}^d)$. For $t > 0$ and $x \in \mathbb{R}^d$, define $w(t, x) \overset{\text{def}}{=} (S^\theta_t f)(x)$. Then w satisfies the PDE*

$$\begin{aligned}\frac{\partial w}{\partial t}(t, x) &= \tilde{\mathscr{L}}^*_\theta w(t, x) \quad t > 0, x \in \mathbb{R}^d,\\ w(0, \cdot) &= f.\end{aligned}$$

Proof. The proof is fairly standard, but, for the sake of completeness, we will outline it. Fix $g \in C^\infty_c(\mathbb{R}^d)$ and set

$$\xi_t \overset{\text{def}}{=} \int_{x\in\mathbb{R}^d} w(t, x)g(x)\mu_\theta(x)\, dx = \langle f, P^\theta_t g\rangle_\theta$$

for all $t > 0$. Clearly,

$$\lim_{t\searrow 0} \xi_t = \int_{x\in\mathbb{R}^d} f(x)g(x)\mu_\theta(x)\, dx.$$

Secondly,

$$\begin{aligned}\dot{\xi}_t &= \langle f, (P^\theta_t(\mathscr{L}_\theta g))\rangle_\theta\\ &= \langle S^\theta_t f, \mathscr{L}_\theta g\rangle_\theta\\ &= \int_{x\in\mathbb{R}^d} w(t, x)\mu_\theta(x)(\mathscr{L}_\theta g)(x)\, dx\\ &= \int_{x\in\mathbb{R}^d} (\tilde{\mathscr{L}}^*_\theta w)(t, x)g(x)\mu_\theta(x)\, dx.\end{aligned}$$

Collecting things together, we get the result. □

An important result which will form the basis for our averaging estimates is the following.

LEMMA 3.3. *There are a $K > 0$ and a $\nu > 0$ such that, for all $t > 0$,*

$$\|P^\theta_t f - \langle f, \mathbf{1}\rangle_\theta \mathbf{1}\|_\theta \le Ke^{-\nu t}\|f\|_\theta$$

for all $f \in L^2(\theta)$ and all $\theta \in S$.

Proof. This is Proposition 4.3 of [7]. □

This gives us our central averaging estimate (the proof of which is also fairly standard).

PROPOSITION 3.4. *There are a $K > 0$ and a $\nu > 0$ such that*

$$\|S^\theta_t f\|_\theta \le Ke^{-\nu t}\|f\|_\theta$$

for all $\theta \in S$, $t > 0$, and $f \in C_c(\mathbb{R}^d)$ such that $\langle f, \mathbf{1}\rangle_\theta = 0$.

Proof. Fix $g \in C^\infty_c(\mathbb{R}^d)$. Then, since $\langle f, \mathbf{1}\rangle_\theta = 0$,

$$\langle S^\theta_t f, g\rangle_\theta = \langle f, P^\theta_t g\rangle_\theta - \langle f, P^\theta_t g - \langle g, \mathbf{1}\rangle_\theta \mathbf{1}\rangle_\theta,$$

so

$$|\langle S_t^\theta f, g\rangle_\theta| \le Ke^{-\nu t}\|f\|_\theta \|g\|_\theta.$$

Take $g = S_t^\theta f$, and the claim follows. □

To proceed, we make the decomposition

$$u^\varepsilon(t,x,\theta) = v^\varepsilon(t,\theta) + \Phi^\varepsilon(t,x,\theta) + R^\varepsilon(t,x,\theta),$$

where

$$\begin{aligned}
d\Phi^\varepsilon(t,x,\theta) &= \frac{1}{\varepsilon}\tilde{\mathscr{L}}_\theta^*\Phi^\varepsilon(t,x,\theta)\,dt + \{(\tilde{\mathscr{Q}}^* v)(t,x,\theta) - (\bar{\mathscr{Q}}^* v)(t,\theta)\}\,dt \\
&\quad + v^\varepsilon(t,\theta)\{h(x,\theta) - \bar{h}(\theta)\}h(X_t^\varepsilon, \Theta_t^\varepsilon)^T\,dt \\
&\quad + v^\varepsilon(t,\theta)\{h(x,\theta) - \bar{h}(\theta)\}^T\,dV_t, \\
\Phi^\varepsilon(0,x) &= \rho(x,\theta) - \bar{\rho}(\theta), \\
dR^\varepsilon(t,x,\theta) &= \frac{1}{\varepsilon}\tilde{\mathscr{L}}_\theta^* R^\varepsilon(t,x,\theta)\,dt + \tilde{\mathscr{Q}}^* R^\varepsilon(t,x,\theta)\,dt \\
&\quad + R^\varepsilon(t,x,\theta)h(x,\theta)h(X_t^\varepsilon,\Theta_t^\varepsilon)^T\,dt \\
&\quad + R^\varepsilon(t,x,\theta)h(x,\theta)^T\,dV_t + \tilde{\mathscr{Q}}^*\Phi^\varepsilon(t,x,\theta)\,dt \\
&\quad + \Phi^\varepsilon(t,x,\theta)h(x,\theta)h(X_t^\varepsilon,\Theta_t^\varepsilon)^T\,dt \\
&\quad + \Phi^\varepsilon(t,x,\theta)h(x,\theta)^T\,dV_t, \\
R^\varepsilon(0,x) &= 0.
\end{aligned} \tag{3.4}$$

We want to show that Φ^ε is small since it reflects an averaging correction. We will then use standard SPDE estimates to show that R^ε, which is driven by Φ^ε, is also small.

Let us further split Φ^ε into several parts, writing $\Phi^\varepsilon = \sum_{i=1}^4 \Phi_i^\varepsilon$, where

$$\begin{gathered}
\frac{\partial\Phi_1^\varepsilon}{\partial t}(t,x,\theta) = \frac{1}{\varepsilon}\tilde{\mathscr{L}}_\theta^*\Phi_1^\varepsilon(t,x,\theta), \\
\Phi_1^\varepsilon(0,x,\theta) = \rho(x,\theta) - \bar{\rho}(\theta), \\
\frac{\partial\Phi_2^\varepsilon}{\partial t}(t,x,\theta) = \frac{1}{\varepsilon}\tilde{\mathscr{L}}_\theta^*\Phi_2^\varepsilon(t,x,\theta) + \{(\tilde{\mathscr{Q}}^* v)(t,x,\theta) - (\bar{\mathscr{Q}}^* v)(t,\theta)\}, \\
\Phi_2^\varepsilon(0,x,\theta) = 0, \\
\frac{\partial\Phi_3^\varepsilon}{\partial t}(t,x,\theta) = \frac{1}{\varepsilon}\tilde{\mathscr{L}}_\theta^*\Phi_3^\varepsilon(t,x,\theta) + v^\varepsilon(t,\theta)\{h(x,\theta) - \bar{h}(\theta)\}h(X_t^\varepsilon,\Theta_t^\varepsilon)^T, \\
\Phi_3^\varepsilon(0,x,\theta) = 0, \\
d\Phi_4^\varepsilon(t,x,\theta) = \frac{1}{\varepsilon}\tilde{\mathscr{L}}_\theta^*\Phi_4^\varepsilon(t,x,\theta) + v^\varepsilon(t,\theta)\{h(x,\theta) - \bar{h}(\theta)\}^T\,dV_t, \\
\Phi_4^\varepsilon(0,x,\theta) = 0.
\end{gathered} \tag{3.5}$$

Let us start to bound the various terms.

LEMMA 3.5. *For each $t > 0$, we have that*

$$\lim_{\varepsilon\searrow 0}\max_{\theta\in S}\mathbb{E}\left[\int_{x\in\mathbb{R}^d}|\Phi_1^\varepsilon(t,x,\theta)|^2\mu_\theta(x)\,dx\right] = 0.$$

Proof. For convenience, define

$$\check{\rho}(x,\theta) \stackrel{\text{def}}{=} \rho(x,\theta) - \bar{\rho}(\theta).$$

By definition, $\langle \check{\rho}(\cdot,\theta), \mathbf{1}\rangle_\theta = 0$. We then have that

$$\Phi_1^\varepsilon(t,x,\theta) = (S^\theta_{t/\varepsilon}\check{\rho}(\cdot,\theta))(x),$$

so

$$\|\Phi^\varepsilon(t,\cdot)\|_\theta \le K e^{-\nu t/\varepsilon}\|\check{\rho}(\cdot,\theta)\|_\theta.$$

This gives the claimed result. □

Before proceeding, we need some uniform bounds on v^ε.

LEMMA 3.6. *For each $t > 0$, we have that*

$$\sup_{\substack{\varepsilon\in(0,1)\\ \theta\in S\\ 0\le s\le t}} \mathbb{E}[|v^\varepsilon(s,\theta)|^2] < \infty.$$

Proof. Define

$$V^\varepsilon(t) \overset{\text{def}}{=} \sum_{\theta\in S}(v^\varepsilon(t,\theta))^2.$$

We have that

$$dV^\varepsilon(t) = 2\sum_{\theta,\theta'\in S} v^\varepsilon(t,\theta)\bar{q}_{\theta',\theta}v^\varepsilon(t,\theta')\,dt + 2\sum_{\theta\in S}\bar{h}(\theta)(v^\varepsilon(t,\theta))^2\,dY^\varepsilon_t + \sum_{\theta\in S}(\bar{h}(\theta)v^\varepsilon(t,\theta))^2\,dt.$$

~~Define now~~

$$Q \overset{\text{def}}{=} \sup\left\{\sum_{\theta,\theta'\in S}\bar{q}_{\theta,\theta'}f(\theta)f(\theta') : f\in B(S), \sum_{\theta\in S} f^2(\theta) = 1\right\};$$

then $Q < \infty$. Thus,

$$\mathbb{E}[V^\varepsilon(t)] \le \sum_{\theta\in S}\bar{\rho}(\theta)^2 + \{2Q + 8\|h\|_B\}K\int_{s=0}^t \mathbb{E}[V^\varepsilon(s)]\,ds.$$

Gronwall's inequality then implies the claim. □

LEMMA 3.7. *For each $t > 0$, we have that*

$$\lim_{\varepsilon\searrow 0}\sup_{\theta\in S}\mathbb{E}\left[\int_{x\in\mathbb{R}^d}|\Phi_2^\varepsilon(t,x,\theta)|^2\mu_\theta(x)\,dx\right] = 0.$$

Proof. Let us start by writing

$$(\tilde{\mathscr{Q}}^* v^\varepsilon)(t,x,\theta) - (\bar{\mathscr{Q}}^* v^\varepsilon)(t,\theta) = \sum_{\theta'\in S}\check{q}_{\theta',\theta}(x)v^\varepsilon(t,\theta'),$$

where

$$\check{q}_{\theta',\theta}(x) \overset{\text{def}}{=} \frac{\mu_{\theta'}(x)}{\mu_\theta(x)}q_{\theta',\theta}(x) - \bar{q}_{\theta',\theta}$$

for all θ and θ' in S and all $x\in\mathbb{R}^d$. Note that $\langle\check{q}_{\theta',\theta},\mathbf{1}\rangle_\theta = 0$ for all θ and θ' in S. We then have that

$$\Phi_2^\varepsilon(t,x,\theta) = \sum_{\theta'\in S}\int_{s=0}^t (S^\theta_{(t-s)/\varepsilon}\check{q}_{\theta',\theta})(x)v^\varepsilon(s,\theta')\,ds.$$

Thus,

$$\|\Phi_2^\varepsilon(t,\cdot,\theta)\|_\theta \le K \sum_{\theta'\in S} \int_{s=0}^t e^{-\nu(t-s)/\varepsilon}\|\check{q}_{\theta',\theta}\|_\theta |v^\varepsilon(s,\theta')|\, ds$$
$$\le K\varepsilon \sum_{\theta'\in S} \|\check{q}_{\theta,\theta'}\| \int_{s=0}^{t/\varepsilon} e^{-\nu s}|v^\varepsilon(t-s\varepsilon,\theta')|\, ds.$$

The claim follows. □

Let us next define the function

$$\check{h}(x,\theta) \stackrel{\text{def}}{=} h(x,\theta) - \bar{h}(\theta); \quad x\in\mathbb{R}^d, \theta\in S;$$

from (2.2), we have that $\langle \check{h}_j(\cdot,\theta), \mathbf{1}\rangle_\theta = 0$ for all $j\in\{1,2,\dots,n\}$.

The bound on Φ_3^ε follows from arguments similar to those of Lemma 3.7.

LEMMA 3.8. *For each $t>0$, we have that*

$$\lim_{\varepsilon\searrow 0} \max_{\theta\in S} \mathbb{E}\left[\int_{x\in\mathbb{R}^d} |\Phi_3^\varepsilon(t,x,\theta)|^2\mu_\theta(x)\right] dx = 0.$$

Proof. We have that

$$\Phi_3^\varepsilon(t,x,\theta) = \sum_{j=1}^n \int_{s=0}^t (S^\theta_{(t-s)/\varepsilon}\check{h}_j)(x) h_j(X_s^\varepsilon,\Theta_s^\varepsilon) v^\varepsilon(s,\theta)\, ds.$$

Thus,

$$\|\Phi_3^\varepsilon(t,\cdot,\theta)\|_\theta \le \sum_{j=1}^n \int_{s=0}^t \|S^\theta_{(t-s)/\varepsilon}\check{h}_j\||h_j(X_s^\varepsilon,\Theta_s^\varepsilon)||v^\varepsilon(s,\theta)|\, ds$$
$$\le K \sum_{j=1}^n \int_{s=0}^t e^{-\nu(t-s)/\varepsilon}\|\check{h}_j\|_\theta |h_j(X_s^\varepsilon,\Theta_s^\varepsilon)||v^\varepsilon(s,\theta)|\, ds$$
$$\le K\varepsilon \sum_{j=1}^n \|\check{h}_j\|_\theta \int_{s=0}^t e^{-\nu s}|h_j(X_{t-s\varepsilon}^\varepsilon,\Theta_{t-s\varepsilon}^\varepsilon)||v^\varepsilon(t-s\varepsilon,\theta)|\, ds.$$

This gives us the result. □

The bound on Φ_4^ε follows from similar arguments once we use Ito's isometry.

LEMMA 3.9. *For each $t>0$, we have that*

$$\lim_{\varepsilon\searrow 0} \sup_{\theta\in S} \mathbb{E}\left[\int_{x\in\mathbb{R}^d} |\Phi_4^\varepsilon(t,x,\theta)|^2\mu_\theta(x)\, dx\right] = 0.$$

Proof. We have that

$$\Phi_4^\varepsilon(t,x,\theta) = \sum_{j=1}^n \int_{s=0}^t (S^\theta_{(t-s)/\varepsilon}\check{h}_j)(x) v^\varepsilon(s,\theta)\, dV_s^j.$$

The Ito isometry thus gives us that

$$\mathbb{E}[\|\Phi_4^\varepsilon(t,\cdot,\theta)\|_\theta^2] \le \sum_{j=1}^n \int_{s=0}^t \|S^\theta_{(t-s)/\varepsilon}\check{h}_j\|^2 \mathbb{E}[|v^\varepsilon(s,\theta)|^2]\, ds$$

$$
\begin{aligned}
&\le K^2 \sum_{j=1}^{n} \int_{s=0}^{t} e^{-2\nu(t-s)/\varepsilon} \|\check{h}_j\|_\theta \mathbb{E}[|v^\varepsilon(s,\theta)|^2]\, ds \\
&\le K\varepsilon \sum_{j=1}^{n} \|\check{h}_j\|_\theta^2 \int_{s=0}^{t/\varepsilon} e^{-2\nu s} \mathbb{E}[|v^\varepsilon(t-s\varepsilon,\theta)|^2]\, ds.
\end{aligned}
$$

The result follows. □

Summarizing, we have that Φ^ε is small.

Lemma 3.10. *For each $t > 0$, we have that*

$$\lim_{\varepsilon \searrow 0} \sup_{\theta \in S} \mathbb{E}\left[\iint_{x\in\mathbb{R}^d} |\Phi^\varepsilon(t,x,\theta)|\, dx\, dt\right] = 0.$$

Proof. Collect Lemmas 3.5, 3.7, 3.8, and 3.9. □

3.1. *Proof of Lemma 3.1*

By standard SPDE methods [18], we have that

$$
\begin{aligned}
\sum_{\theta\in S} \mathbb{E}[\|R^\varepsilon(t,\cdot,\theta)\|_\theta^2] \le{}& \frac{2}{\varepsilon} \sum_{\theta\in S} \int_{s=0}^{t} \mathbb{E}[\langle \mathscr{L}_\theta R^\varepsilon(s,\cdot,\theta), R^\varepsilon(s,\cdot,\theta)\rangle_\theta]\, ds \\
&+ 2 \sum_{\theta\in S} \int_{s=0}^{t} \mathbb{E}[\langle \mathscr{Q} R^\varepsilon(s,\cdot,\theta), R^\varepsilon(s,\cdot,\theta)\rangle_\theta]\, ds \\
&+ 2 \sum_{\substack{\theta\in S\\ 1\le j\le n}} \int_{s=0}^{t} \mathbb{E}[\langle R^\varepsilon(s,\cdot,\theta), R^\varepsilon(s,\cdot,\theta) h_j(\cdot,\theta)\rangle_\theta h_j(X_s^\varepsilon,\Theta_s^\varepsilon)]\, ds \\
&+ 2 \sum_{\theta'\in S} \int_{s=0}^{t} \mathbb{E}[\langle \mathscr{Q} R^\varepsilon(s,\cdot,\theta), \Phi^\varepsilon(s,\cdot,\theta)\rangle_\theta]\, ds \\
&+ 2 \sum_{\substack{\theta\in S\\ 1\le j\le n}} \int_{s=0}^{t} \mathbb{E}[\langle R^\varepsilon(s,\cdot,\theta), \Phi^\varepsilon(s,\cdot,\theta) h_j(\cdot,\theta)\rangle_\theta h_j(X_s^\varepsilon,\Theta_s^\varepsilon)]\, ds \\
&+ \sum_{1\le j\le n} \int_{s=0}^{t} \mathbb{E}\left[\left\|\sum_{\theta\in S}\{R^\varepsilon(s,\cdot,\theta) + \Phi^\varepsilon(s,\cdot,\theta)\} h_j(\cdot,\theta)\right\|_\theta^2\right] ds.
\end{aligned}
$$

The bound on the $1/\varepsilon$ term is standard. For any $f \in C^\infty(\mathbb{R}^d)$, we have that

$$\mathscr{L}_\theta f^2 = 2f\mathscr{L}_\theta f + \sum_{1\le i\le d} \left(\frac{\partial f}{\partial x_i}\right)^2 \geqslant 2f\mathscr{L}_\theta f$$

and, hence, since μ_θ is an invariant distribution,

$$\langle \mathscr{L}_\theta f, f\rangle_\theta = \int_{x\in\mathbb{R}^d} f(x)\mathscr{L}_\theta f(x)\mu_\theta(x)\, dx \le \frac{1}{2}\int_{x\in\mathbb{R}^d} (\mathscr{L}_\theta f^2)(x)\mu_\theta(x)\, dx = 0.$$

Let us also define

$$Q \stackrel{\text{def}}{=} \sup\left\{\sum_{\theta,\theta'\in S} q_{\theta,\theta'}(x) f(\theta) g(\theta') : x\in\mathbb{R}^d,\ f,g\in B(S),\ \sum_{\theta\in S} f^2(\theta) = \sum_{\theta\in S} g^2(\theta) = 1\right\};$$

then $Q < \infty$. Hence,

$$\sum_{\theta \in S} \mathbb{E}[\|R^\varepsilon(t,\cdot,\theta)\|_\theta^2]$$

$$\leq 2Q \sum_{\theta' \in S} \int_{s=0}^t \mathbb{E}[\|R^\varepsilon(s,\cdot,\theta)\|_\theta^2]\,ds$$

$$+ 2\left\{\sum_{1\le j\le n} \|h_j\|_B^2\right\} \sum_{\theta \in S} \int_{s=0}^t \mathbb{E}[\|R^\varepsilon(s,\cdot,\theta)\|_\theta^2]\,ds$$

$$+ Q\left\{\sum_{\theta \in S} \int_{s=0}^t \mathbb{E}[\|R^\varepsilon(s,\cdot,\theta)\|_\theta^2]\,ds + \sum_{\theta \in S} \int_{s=0}^t \mathbb{E}[\|\Phi^\varepsilon(s,\cdot,\theta)\|_\theta^2]\,ds\right\}$$

$$+ \left\{\sum_{1\le j\le n} \|h_j\|_B^2\right\}\left\{\sum_{\theta \in S} \int_{s=0}^t \mathbb{E}[\|R^\varepsilon(s,\cdot,\theta)\|_\theta^2]\,ds + \sum_{\theta \in S} \int_{s=0}^t \mathbb{E}[\|\Phi^\varepsilon(s,\cdot,\theta)\|_\theta^2]\,ds\right\}$$

$$+ 2\left\{\sum_{1\le j\le n} \|h_j\|_B^2\right\} \int_{s=0}^t \left\{\sum_{\theta \in S} \int_{s=0}^t \mathbb{E}[\|R^\varepsilon(s,\cdot,\theta)\|_\theta^2]\,ds + \sum_{\theta \in S} \int_{s=0}^t \mathbb{E}[\|\Phi^\varepsilon(s,\cdot,\theta)\|_\theta^2]\,ds\right\}.$$

Apply Gronwall's inequality and use Lemma 3.10 to bound R^ε. Combining things together, the claim follows.

4. *Proof of Theorem 2.1*

We finally want to return to our analysis of π_t^ε. We want to use Lemma 3.1 to show that π_t^ε and $\bar{\pi}_t^\varepsilon$ are close. To start, define

$$\bar{V}^\varepsilon(t) \stackrel{\text{def}}{=} \sum_{\theta \in S} v^\varepsilon(t,\theta).$$

From standard calculations, we have that $v^\varepsilon(t,\theta) \geqslant 0$ for all $t > 0$ and $\theta \in S$.

LEMMA 4.1. *For all $t \geqslant 0$, $\varepsilon \in (0,1)$, and $L > 0$,*

$$d_{\mathscr{P}(S)}(\pi_t^\varepsilon, \bar{\pi}_t^\varepsilon) \leq \frac{2}{V^\varepsilon(t)} \sum_{\theta \in S} \sqrt{\int_{x\in\mathbb{R}^d} |\tilde{u}^\varepsilon(t,x,\theta) - v^\varepsilon(t,\theta)|^2 \mu_\theta(x)\,dx} \tag{4.1}$$

if $\bar{V}^\varepsilon(t) > 0$.

Proof. For each $t \geqslant 0$ and $\varepsilon > 0$, define the random σ-finite measures $\pi_t^{\circ,\varepsilon}$ and $\bar{\pi}_t^{\circ,\varepsilon}$ on $(\mathbb{R}^d \times S, \mathscr{B}(\mathbb{R}^d \times S))$ as

$$\pi_t^{\circ,\varepsilon}(A) \stackrel{\text{def}}{=} \sum_{\theta \in A} \int_{x\in\mathbb{R}^d} \tilde{u}^\varepsilon(t,x,\theta)\mu_\theta(x)\,dx,$$

$$\bar{\pi}_t^{\circ,\varepsilon}(A) \stackrel{\text{def}}{=} \sum_{\theta \in A} v^\varepsilon(t,\theta) = \sum_{\theta \in A} \int_{x\in\mathbb{R}^d} v^\varepsilon(t,\theta)\mu_\theta(x)dx$$

for all $A \subset S$. Then

$$\pi_t^\varepsilon(A) - \bar{\pi}_t^\varepsilon(A) = \frac{\pi_t^{\circ,\varepsilon}(A)}{\pi_t^{\circ,\varepsilon}(S)} - \frac{\bar{\pi}_t^{\circ,\varepsilon}(A)}{\bar{\pi}_t^{\circ,\varepsilon}(S)}$$
$$= \frac{\pi_t^{\circ,\varepsilon}(A)}{\pi_t^{\circ,\varepsilon}(S)\bar{\pi}_t^{\circ,\varepsilon}(S)}\{\bar{\pi}_t^{\circ,\varepsilon}(S) - \pi_t^{\circ,\varepsilon}(S)\} + \frac{1}{\bar{\pi}_t^{\circ,\varepsilon}(S)}\{\bar{\pi}_t^{\circ,\varepsilon}(A) - \pi_t^{\circ,\varepsilon}(A)\}.$$

For any $A' \subset S$,

$$\begin{aligned}|\pi_t^{\circ,\varepsilon}(A') - \bar{\pi}_t^{\circ,\varepsilon}(A')| &= \left|\sum_{\theta \in A'} \int_{x\in\mathbb{R}^d} \{\tilde{u}^\varepsilon(t,x,\theta) - v^\varepsilon(t,\theta)\}\mu_\theta(x)\,dx\right| \\ &\le \sum_{\theta\in S}\left|\int_{x\in\mathbb{R}^d} \{\tilde{u}^\varepsilon(t,x,\theta) - v^\varepsilon(t,\theta)\}\mu_\theta(x)\,dx\right| \\ &\le \sum_{\theta\in S}\sqrt{\int_{x\in\mathbb{R}^d} |\tilde{u}^\varepsilon(t,x,\theta) - v^\varepsilon(t,\theta)|^2\mu_\theta(x)\,dx}.\end{aligned}$$

Of course, $\bar{\pi}_t^{\circ,\varepsilon}(S) = \bar{V}^\varepsilon(t)$. The claim follows. □

Proof of Theorem 2.1. From Lemma 3.1, it suffices to show that

$$\sup_{\varepsilon\in(0,1)} \mathbb{E}\left[\frac{1}{(\bar{V}^\varepsilon(t))^2}\right] < \infty. \tag{4.2}$$

In fact, we have that

$$d\bar{V}^\varepsilon(t) = \left\{\sum_{\theta\in S} \bar{h}(\theta) v^\varepsilon(t,\theta)\right\} dY_t^\varepsilon.$$

Of course, $\bar{V}^\varepsilon(0) = 1$. For each $n\in\mathbb{N}$, define

$$\tau_n \overset{\text{def}}{=} \inf\left\{t \geqslant 0 : \bar{V}^\varepsilon(t) \le \frac{1}{n}\right\}.$$

Define $\tau \overset{\text{def}}{=} \lim_{n\to\infty}\tau_n = \inf\{t\geqslant 0 : \bar{V}^\varepsilon(t) = 0\}$. For $t\in[0,\tau)$, define

$$a(t) \overset{\text{def}}{=} \frac{\sum_{\theta\in S}\bar{h}(\theta)v^\varepsilon(t,\theta)}{\sum_{\theta\in S} v^\varepsilon(t,\theta)};$$

since the $v^\varepsilon(t,\theta)$ are non-negative, we have that

$$\|a(t)\|_{\mathbb{R}^n} \le \sup_{\theta\in S}\|\bar{h}(\theta)\|_{\mathbb{R}^n}. \tag{4.3}$$

For every $n\in\mathbb{N}$, we have that

$$\bar{V}^\varepsilon(t\wedge\tau_n) = \exp\left[\int_{s=0}^{t\wedge\tau_n} a(s)^T h(X_s^\varepsilon,\Theta_s^\varepsilon)\,ds + \int_{s=0}^{t\wedge\tau_n} a(s)^T\,dV_s - \frac{1}{2}\int_{s=0}^{t\wedge\tau_n}\|a(s)\|^2\,ds\right].$$

Letting $n\to\infty$, we get that

$$\bar{V}^\varepsilon(t\wedge\tau) = \exp\left[\int_{s=0}^{t\wedge\tau} a(s)^T h(X_s^\varepsilon,\Theta_s^\varepsilon)\,ds + \int_{s=0}^{t\wedge\tau} a(s)^T\,dV_s - \frac{1}{2}\int_{s=0}^{t\wedge\tau}\|a(s)\|^2\,ds\right].$$

Since the exponential term is finite thanks to (4.3), we must have that $\tau > t$. Thus,

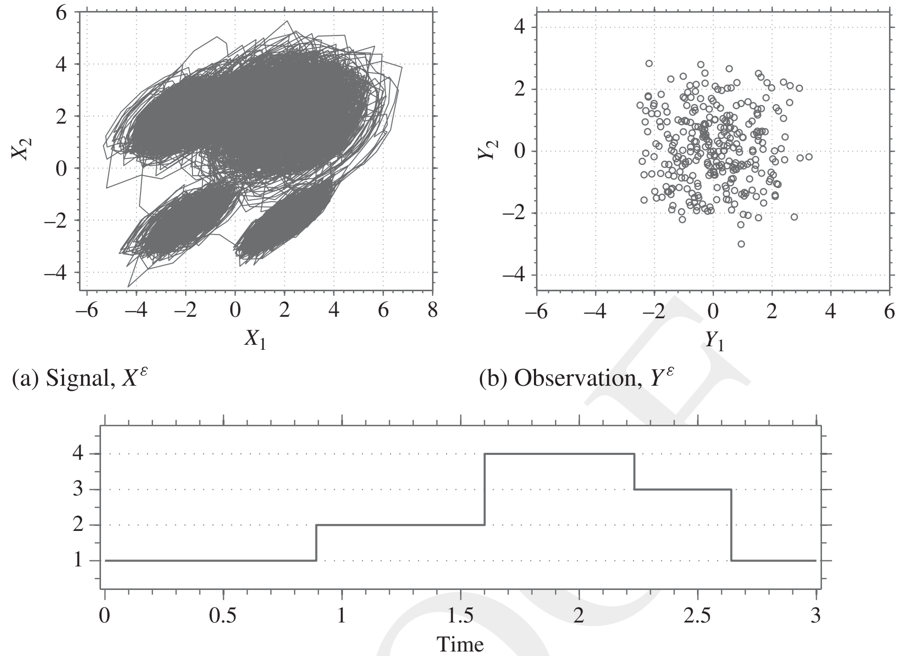
$$\frac{1}{(\bar{V}^\varepsilon(t))^2} = \exp\left[-2\int_{s=0}^{t\wedge\tau} a(s)^T h(X_s^\varepsilon,\Theta_s^\varepsilon)\,ds - 2\int_{s=0}^{t\wedge\tau} a(s)^T\,dV_s + \int_{s=0}^{t\wedge\tau}\|a(s)\|^2\,ds\right].$$

This implies (4.2), completing the proof. □

5. *A numerical example*

We consider a simple system in a continuous–discrete set-up; the fast variable $X^\varepsilon\in\mathbb{R}^2$ is given as a continuous process

$$dX_t^\varepsilon = -\frac{1}{\varepsilon}\Lambda_{\Theta_t^\varepsilon}(X_t^\varepsilon - \Theta_t^\varepsilon)\,dt + \frac{1}{\sqrt{\varepsilon}}\,dW_t, \quad X_0^\varepsilon = \xi,$$



(a) Signal, X^ε (b) Observation, Y^ε

(c) 1 : (2, 2), 2 : (–2, 2), 3 : (– 2, –2), 4 : (2, –2)

FIGURE 1. *Signal and observation.*

while the slow variable Θ^ε is a jump process with a finite state space $S = \{(2,2), (-2,2), (-2,-2), (2,-2)\}$. The matrix Λ_θ is given for each $\theta \in S$ as

$$\Lambda_{(2,2)} \stackrel{\text{def}}{=} \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}, \qquad \Lambda_{(-2,2)} \stackrel{\text{def}}{=} \begin{pmatrix} 2 & -2 \\ 2 & 2 \end{pmatrix},$$
$$\Lambda_{(-2,-2)} \stackrel{\text{def}}{=} \begin{pmatrix} 3 & -2 \\ 2 & 3 \end{pmatrix}, \quad \Lambda_{(2,-2)} \stackrel{\text{def}}{=} \begin{pmatrix} 4 & -2 \\ 2 & 4 \end{pmatrix}$$

and the generator of Θ^ε is defined as

$$Q(x) \stackrel{\text{def}}{=} \begin{pmatrix} -\|x\|^2 & \|x\|^2 & 0 & 0 \\ \|x\|^2 & -2\|x\|^2 & 0 & \|x\|^2 \\ \|x\|^2 & 0 & -2\|x\|^2 & \|x\|^2 \\ 0 & 0 & \|x\|^2 & -\|x\|^2 \end{pmatrix} \times 10^{-5},$$

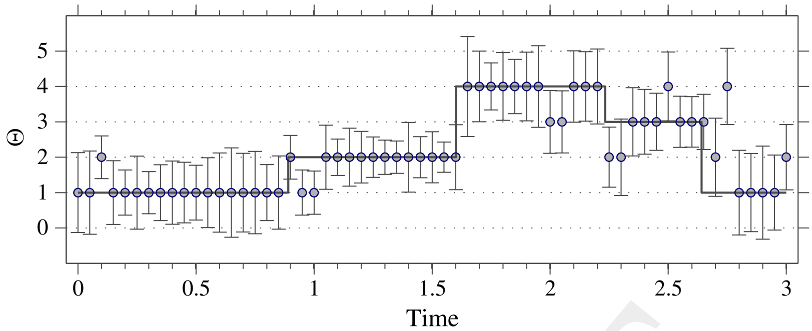
where $\|x\| \stackrel{\text{def}}{=} \sqrt{x_1^2 + x_2^2}$.

In this example, observations are made at equally spaced discrete points as follows:

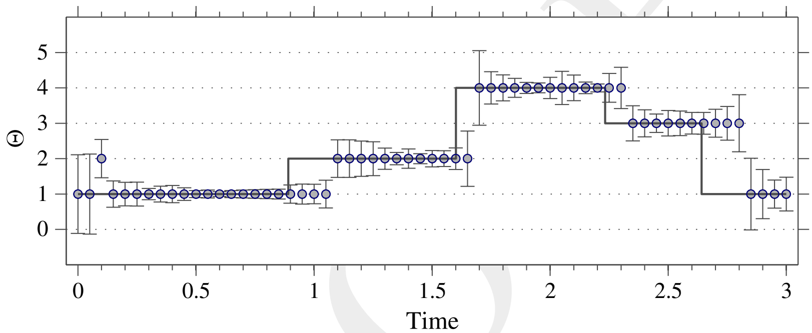
$$Y^\varepsilon_{t_k} = \sin X^\varepsilon_{t_k} + B_{t_k},$$

where B_{t_k} is a standard Gaussian white noise sequence.

Figure 1(a) and (b) show typical plots for the fast process, X^ε_t, of the above multiscale hybrid system and the observation process Y^ε_t, respectively. Figure 1(c) shows the evolution of the slow process, Θ^ε, in time, where the original state S is mapped into $\{1, 2, 3, 4\}$. To show the validity and efficiency of the homogenized filter, we applied the particle filter (PF) [**5**] and the homogenized hybrid particle filter (HHPF) [**8, 17**] algorithms for a comparison. Figure 2(a) and (b) show maximum a posteriori (MAP) estimates with error bars representing one standard deviation, where 400 particles are used.



(a) PF (Thick line: Θ^ε, Circle: $\hat{\Theta}^{\mathrm{PF}}$)



(b) HHPF (Thick line: Θ^ε, Circle: $\hat{\Theta}^{\mathrm{HHPF}}$)

FIGURE 2. *MAP and standard deviation.*

We also compare the errors for PF and HHPF in Table 1. The errors are obtained from a 0–1 error estimate given by

$$\mathbb{E}\mathbf{1}_{\Theta^\varepsilon_{t_k}\neq\widehat{\Theta}_k^{\{\cdot\}}} \approx \frac{1}{T}\sum_{l=1}^{T}\mathbf{1}_{\Theta^\varepsilon_{t_k}\neq\widehat{\Theta}_k^{\{\cdot\}}},$$

where $\widehat{\Theta}_k^{\mathrm{PF}}$ and $\widehat{\Theta}_k^{\mathrm{HHPF}}$ are MAP estimates at a discretized point k obtained respectively from PF and HHPF algorithms and $\Theta^\varepsilon_{t_k}$ represents the value of Θ^ε at k. The values in Table 1 are based on 50 Monte Carlo simulations. The mean times taken for these simulations with Intel Xeon 5540 (2.53 GHz) quad-core Nehalem processors are given in the parentheses (the unit is 10^3 seconds). While the errors for both algorithms are comparable, the time taken for HHPF is much less that that of PF.

Acknowledgements. The third author would like to thank the Departments of Mathematics and Statistics of Stanford University for their hospitality in the Spring of 2010 during

TABLE 1. *Errors of PF and HHPF.*

Algorithm/N	50	100	200	400	800
PF	0.2246 (0.26)	0.2061 (0.52)	0.2026 (1.07)	0.1972 (2.23)	0.1937 (4.86)
HHPF	0.2607 (0.007)	0.2187 (0.02)	0.2036 (0.07)	0.2010 (0.23)	0.1970 (0.89)

a sabbatical stay. The first author acknowledges the support from OSD/AFOSR grant 9550-05-1-0613. The second author acknowledges the support from NSF grant DMS 0604863, ARO grant W911NF-07-1-0044, and OSD/AFOSR grant 9550-05-1-0613. The third author acknowledges the support from AFOSR grant FA9550-08-1-0206 and NSF grant DMS 0604249 during the preparation of this work.

References

1. A. Bain and D. Crisan, *Fundamentals of stochastic filtering*, Stochastic Modelling and Applied Probability 60 (Springer, New York, 2009).

2. V. E. Beneš, 'Exact finite-dimensional filters for certain diffusions with nonlinear drift', *Stochastics* 5 (1981) 65–92.

3. A. Crudu, A. Debussche and O. Radulescu, 'Hybrid stochastic simplifications for multiscale gene networks', *BMC Syst. Biol.* 3 (2009) no. 89,.

4. F. Daum, 'Exact finite-dimensional nonlinear filters', *IEEE Trans. Automat. Control* 31 (1986) no. 7, 616–622.

5. A. Doucet, 'On sequential simulation-based methods for Bayesian filtering', *Technical report* (Department of Engineering, University of Cambridge, Cambridge, UK, 1998).

6. J. Filar, V. Gaitsgory and A. Haurie, 'Control of singularly perturbed hybrid stochastic systems', *IEEE Trans. Automat. Control* 46 (2001) no. 2, 179–190.

7. M. Fuhrman, 'Hypercontractivity properties of nonsymmetric Ornstein–Uhlenbeck semigroups in Hilbert spaces', *Stoch. Anal. Appl.* 16 (1998) no. 2, 241–260.

8. D. Givon, P. Stinis and J. Weare, 'Variance reduction for particle filters of systems with time scale separation', *IEEE Trans. Signal Process.* 57 (2009) no. 2, 424–435.

9. I. Hwang, H. Balakrishnan and C. Tomlin, 'State estimation for hybrid systems: applications to aircraft tracking', *IEE Proc. – Control Theory Appl.* 153 (2006) no. 5, 556–566.

10. A. M. Il'in, R. Z. Khasminskii and G. Yin, 'Singularly perturbed switching diffusions: rapid switchings and fast diffusions', *J. Optim. Theory Appl.* 102 (1999) no. 3, 555–591.

11. R. E. Kalman, 'A new approach to linear filtering and prediction problems', *J. Basic Eng.* 82 (1960) no. 1, 35–45.

12. P. Kokotović, H. Khalil and J. O'Reilly, *Singular perturbation methods in control: analysis and design* (Society for Industrial Mathematics, 1999).

13. M. Miller, U. Grenander, J. O'Sullivan and D. Snyder, 'Automatic target recognition organized via jump-diffusion algorithms', *IEEE Trans. Image Process.* 6 (1997) no. 1, 157–174.

14. G. C. Papanicolaou, 'Asymptotic analysis of stochastic equations', *Stud. Probab. Theory* (1978) 111–179.

15. J. H. Park, N. S. Namachchivaya and R. B. Sowers, 'A problem in stochastic averaging of nonlinear filters', *Stoch. Dyn.* 8 (2008) no. 3, 543–560.

16. J. H. Park, R. B. Sowers and N. S. Namachchivaya, 'Dimensional reduction in nonlinear filtering', *Nonlinearity* 23 (2010) 305–324.

17. J. H. Park, N. S. Namachchivaya and H. C. Yeong, Particle filters in a multiscale environment: homogenized hybrid particle filter (HHPF). Under review.

18. B. L. Rozovskii, *Stochastic evolution systems*, Mathematics and its Applications (Soviet Series) 35 (Kluwer Academic, Dordrecht, 1990) Linear theory and applications to nonlinear filtering, translated from the Russian by A. Yarkho.

19. B. Rozovskii, R. Blazek and A. Petrov, *Interactive banks of Bayesian matched filters*, In SPIE Proceedings (Volume 4048): Signal and Data Processing of Small Targets (Orlando, FL, 2000) (ed. O. E. Drummond; SPIE (The International Society for Optical Engineering), Bellingham, WA, 2000).

20. D. D. Sworder and J. Boyd, *Estimation problems in hybrid systems* (Cambridge University Press, Cambridge, UK, 1999).

21. F. Verhulst, *Methods and applications of singular perturbations: boundary layers and multiple timescale dynamics* (Springer, Berlin, 2005).

22. H. Wang, 'Mathematical theory of molecular motors and a new approach for uncovering motor mechanisms', *IEE Proc. – Nanobiotechnology* 150 (2003) no. 3, 127–133.

23. E. Wong and J. B. Thomas, 'On polynomial expansions of second-order distributions', *J. Soc. Ind. Appl. Math.* 10 (1962) no. 3, 507–516.

24. G. Yin and C. Zhu, *Hybrid switching diffusions: properties and applications*, Stochastic Modelling and Applied Probability 63 (Springer, Berlin, 2010).

25. M. Zakai, 'On the optimal filtering of diffusion processes', *Z. Wahrscheinlichkeitstheorie verw. Geb.* 11 (1969) 230–243.

26. Q. Zhang and G. Yin, 'Nearly-optimal asset allocation in hybrid stock investment models', *J. Optim. Theory Appl.* 121 (2004) no. 2, 419–444.

Jun H. Park
182 George Street
Providence, RI 02912
USA

jun_park@brown.edu

Boris Rozovskii
182 George Street
Providence, RI 02912
USA

boris_rozovsky@brown.edu

Richard B. Sowers
1409 W. Green Street
Urbana, IL 61801
USA

r-sowers@illinois.edu

AUTHOR QUERIES

Q1 (page 16)
Please check that the journal title given in Ref. [3] is OK as set. Also, please give the page range.

Q2 (page 16)
Please check added organization details in Ref. [5].

Q3 (page 16)
Please confirm the publisher name in Ref. [12]. Maybe "Society for Industrial and Applied Mathematics"?

Q4 (page 16)
Please give the volume number in Ref. [14].

Q5 (page 16)
Please update Ref. [17], if possible.

Q6 (page 16)
Reference [23] is not cited in the text. Please cite or remove from the list.