SRC TR 85-29

Convergence of Implicit Discretization Schemes for Linear Differential Equations with Application to Filtering

By

M. Piccioni

CONVERGENCE OF IMPLICIT DISCRETIZATION SCHEMES FOR LINEAR DIFFERENTIAL EQUATIONS WITH APPLICATION TO FILTERING

by

M. Piccioni*
Department of Electrical Engineering
University of Maryland
College Park, MD 20742

* On leave for: Dipartimento di Matematica, II Universita di Roma, Via Orazio Raimondo, 00173 Roma, Italy.

This work was supported partially through ONR Grant N00014-84-K-0614, partially through a grant from the Minta Martin Aeronautical Research Fund, College of Engineering, University of Maryland at College Park, and partially through Grant No. 203.01.36 of CNR, Italy.

1. INTRODUCTION

The motivation for the present work arises from the following well-known problem in nonlinear filtering. Let $(\overline{X}_t)$ be a R^d-valued diffusion process with generator A and let (W_t) be an independent R^m-valued standard Brownian motion, both defined on $(\overline{\Omega},\overline{\ },\overline{P})$. Let

$$Y_t = \int_0^t g(\overline{X}_s)ds + W_t, \ t\geq 0, \tag{1.1}$$

where $g:R^d \to R^m$. Compute recursively the conditional expectations $\overline{E}(f(\overline{X}_t)|Y_s, 0\leq s\leq t)$ for some "sufficiently large" class of functions f defined on R^d. Boundedness and smoothness assumptions on the coefficients of A will be given later. We assume from now on that $g_i \in L^\infty(R^d)$, $i=1,\ldots,m$.

A convenient representation for the desired conditional expectations is given by the Kallianpur-Striebel formula [11]. Define on another probability space (Ω,F,P) a diffusion (X_t) with the same distribution of $(\overline{X}_t)$. Then

$$\overline{E}(f(\overline{X}_t)|Y_s, 0\leq s\leq t) = \frac{E[f(X_t)\exp(\int_0^t g^T(X_s)dY_s - \frac{1}{2}\int_0^t |g(X_s)|^2 ds)]}{E[\exp(\int_0^t g^T(X_s)dY_s - \frac{1}{2}\int_0^t |g(X_s)|^2 ds)]}, \tag{1.2}$$

which reduces the problem to the computation of integrals on the paths of (X_t) (the stochastic integral, for each path of (X_t), is a Wiener integral computed on the given path of (Y_t)). By differentiating the numerator a weak stochastic partial differential equation is obtained for a multiple of the conditional probability measure q_t, usually called the Zakai equation [31]

$$dq_t = A^* q_t \, dt + q_t \, g^T dY_t \ . \tag{1.3}$$

Of course, if we want to solve (1.3) recursively "on line" on a digital computer the best we can do is to provide well-behaved discretization algorithms in both space and time.

It is useful to design algorithms which discretize (1.3) but still retain the representation (1.2), merely changing the process (X_t) involved. This could be obtained by replacing the diffusion by a continuous-time finite-state Markov chain with generator A_h, which of course is of finite-difference type (the simplest example is provided by Kushner [17]). But it continues to hold if time is implicitly discretized and the stochastic Trotter product formula [5,2] is used, thereby obtaining the equation

$$(I - \Delta A_h^T) q^{h,\Delta}_{(k+1)\Delta} = \exp(g^T(Y_{(k+1)\Delta} - Y_{k\Delta}) - \tfrac{1}{2}|g|^2\Delta) q^{h,\Delta}_{k\Delta}, \quad k = 0,1,\ldots \tag{1.4}$$

This scheme has been first obtained by Clark [5], by discretizing implicitly time in the "robust" version of (1.3). For us it is more interesting to know that the solution of (1.4) can be written essentially as in (1.2) replacing (X_t) by a discrete-time Markov chain $(X^{h,\Delta}_{k\Delta})$ with transition matrix $(I-\Delta A_h^T)^{-1}$ [20]. This relates the convergence of the approximation scheme (1.4) to the weak convergence of $(X^{h,\Delta}_{k\Delta})$ to (X_t) when $h,\Delta \to 0$ (h is thought as a mesh parameter of the space discretization grid). But this is known to be ensured by the convergence of the discrete semigroup described by the free behaviour of equation (1.4) (i.e., when $Y \equiv 0$) to the semigroup generated by A [14]. The relevant point is that we would like to establish convergence for h and Δ going to zero independently. Our main Theorem 2.4, given in the following section, gives sufficient conditions for this only in term of A_h and A. This is reasonable, in that the matrices A_h are the parameter of the scheme (1.4) and they have to be chosen with the best possible band structure, so as to solve as quickly as possible the equation (1.4), without inverting $I-\Delta A_h^T$ [1]. It turns out that these conditions are slightly stronger than those given by the Trotter-Kato-Kurtz theorem [27, 11, 13] for the convergence of the semigroup generated by A_h to that generated by A, therefore confirming in an abstract setting that the implicit discretization of time allows an independent choice of discretization steps in space and time, respectively [24]. Thus Theorem 2.4 could be of some interest independently of the filtering problem; of course it can be successfully applied to show convergence for different approximate filters which usually do not retain any probabilistic meaning, like those built by Galerkin methods [7].

In Section 3 we review the already cited results connecting convergence of Markov (Feller-Dynkin) semigroups with weak convergence of their sample paths. With Theorem 2.4 and this type of results, in Section 4 we obtain convergence results for the functionals involved in (1.2), computed averaging on the paths of $(X^{h,\Delta}_{k\Delta})$. For this it is useful to obtain for those functionals a Lipschitz condition in Y, independently of (h,Δ), thereby extending previous results for the Kushner space discretization scheme [20]. This is done by the arguments used in [17], that is integration by parts in (1.2) and some martingale estimates (which require conditions on g, too). In Section 5 this result is shown to imply the robustness of the approximate filters (1.4), in that if they are forced by

$$Y^{\varepsilon}_t = \int_0^t g(\overline{X}^{\varepsilon}_s)ds + W^{\varepsilon}_t,\ t\geq 0 \tag{1.5}$$

where $(\overline{X}^{\varepsilon}_t, W^{\varepsilon}_t)$ converges weakly, as $\varepsilon\to 0$, to $(\overline{X}_t, W_t)$, nonetheless the joint distribution of $(\overline{X}^{\varepsilon}_t)$ and the (h,Δ)-approximate filter computed on the paths of (Y^{ε}_t) converges weakly, as $(h,\Delta,\varepsilon)\to 0$, to the "ideal" one given by (1.2) and (2.2).

The final section deals with two short examples. The first is Kushner's scheme, the other one a variation of that which is intended to show that different choices are possible, depending on the particular structure of the diffusion model. The sufficient conditions are easily checked in any case, but to avoid cumbersome notations we limit ourselves to the case d=1. However, when the dimension of the state space increases, the reasonable choices for the approximating chains increase, surely influencing the speed of computation. Much more work remains to be done on these issues. Anyway boundedness conditions on the coefficient of the diffusion are needed, because the state space is not compact. This suggests that the convergence and robustness results obtained, which holds in general for continuous Feller-Dynkin processes on locally compact state spaces, will be more meaningful for nondegenerate diffusions on compact Riemannian manifolds.

2. THE ABSTRACT CONVERGENCE THEOREM

First of all we recall the Hille-Yosida theorem. Let L by any Banach space.

THEOREM 2.1. A linear operator A on L is the infinitesimal generator of a (strongly continuous) semigroup of operators on L if and only if: D(A) is dense in L and, for $\Delta>0$ $I-\Delta A$ is invertible on the whole L, with the inverse J_Δ which is a contraction.

We remark that Hille's proof [9] is just based on the convergence of the implicit time-discretization scheme (governed by $J_\Delta^{\lfloor t/\Delta \rfloor}$) to the generalized solution of the corresponding Cauchy problem with problem with operator A, when Δ goes to zero. Yosida [30] approximates this solution with the exponentials of the bounded operators $A_\Delta = \Delta^{-1}(J_\Delta - I)$. Of course, because A generates a unique semigroup (which will be called $\{e^{At}\}$) the contraction-valued functions of time $J_\Delta^{\lfloor t/\Delta \rfloor}$ and $e^{A_\Delta t}$ have the same asymptotic behaviour as $\Delta \to 0$. We will utilize a generalization of this result, due to Kurtz [13], to prove Theorem 2.4 by using the more convenient Yosida-type argument.

Let us put ourselves in the setting of [13] which allows to consider convergence of Markov processes defined on different state spaces. Suppose that for each $h>0$, L_h is a Banach space and there exists a bounded map $P_h : L \to L_h$ such that for each $f \in L$ $\lim_{h\to 0} \|P_h f\| = \|f\|$, which in turns imples that $\|P_h\| \leqslant M$ for some $M>0$. On each L_h an infinitesimal generator A_h of some contraction semigroup is specified. For numerical applications L_h will be always finite-dimensional so that A_h will be bounded, but this is not assumed now.

We say that the family of semigroups $\{e^{A_h t}\}$, $h>0$, converges to $\{e^{At}\}$ as $h \to 0$ if, for any $f \in L$ and $T>0$

$$\lim_{h\to 0} \sup_{t\in[0,T]} \|e^{A_h t} P_h f - P_h e^{At} f\| = 0 \tag{2.1}$$

Conditions for (2.1) to hold are given by the Trotter-Kato- Kurtz theorem which is reported below.

THEOREM 2.2. The following are equivalent:

i) $\{e^{A_h t}\}$ converges to $\{e^{At}\}$ for $h\to 0$;

ii) for each $f\in L$

$$\lim_{h\to 0} \|(I-A_h)^{-1}P_h f - P_h(I-A)^{-1}f\|=0 \tag{2.2}$$

iii) for each f in a core S of A there exists $f_h \in D(A_h)$ such that

$$\lim_{h\to 0} \|f_h - P_h f\| \vee \|A_h f_h - P_h Af\| = 0 \tag{2.3}$$

We recall that a core S of the generator A is a linear manifold included in D(A) such that A is the closed extension of $A|_S$. The most used cores are linear manifolds dense in L which are invariant under e^{At}, for $t>0$.

Theorem 2.2 will be a basic tool in the sequel, in that both conditions ii) and iii) will be used to prove Theorem 2.4. The fact that ii) → i) is due to Trotter [27], whereas the converse to Kato [11]. Condition iii) have been introduced by Kurtz [13].

If an implicit discretization scheme is applied to the evolution equation governed by A_h, the discrete contraction semigroup $J^k_{h,\Delta} = (I-\Delta A_h)^{-k}$ is obtained on the space L_h. Our objective is to find sufficient conditions under which $\{J^{[t/\Delta]}_{h,\Delta}\}$ converges to $\{e^{At}\}$ in the same sense of (2.1), as $(h,\Delta)\to 0$. As mentioned before we consider continuous-time semigroups "asymptotically equivalent" to $\{J^{[t/\Delta]}_{h,\Delta}\}$ by means of the following estimate [13]. Define the bounded operator on L_h

$$A_{h,\Delta} = \Delta^{-1}(J_{h,\Delta}-I) \tag{2.4}$$

which is easily shown to generate a contraction semigroup.

THEOREM 2.3. For any $f\in L$, $t>0$ and $\varepsilon>0$

$$\|(J^{[t/\Delta]}_{h,\Delta} - e^{A_{h,\Delta}t})\, P_h f\| \leq 2t\|A_{h,\Delta}P_h f\| \wedge \left(\frac{2\Delta}{\varepsilon^2 t}\|P_h f\| + (\varepsilon t+\Delta)\|A_{h,\Delta}P_h f\|\right). \tag{2.5}$$

We are now ready to prove the promised convergence theorem.

THEOREM 2.4. Let us suppose that for any f in a core S of A, $P_h f \in D(A_h)$ (at least for h sufficiently small) and

$$\lim_{h\to 0} \|A_h P_h f - P_h A f\| = 0 \tag{2.6}$$

Then, for any $f \in L$ and $T>0$

$$\lim_{(h,\Delta)\to 0} \sup_{t\in[0,T]} \|J_{h,\Delta}^{\lfloor t/\Delta \rfloor} P_h f - P_h e^{At} f\| = 0 \tag{2.7}$$

Proof. By Theorem 2.2 it is enough to prove that for any $f \in L$

$$\lim_{(h,\Delta)\to 0} \|(I-A_{h,\Delta})^{-1} P_h f - P_h (I-A)^{-1} f\| = 0. \tag{2.8}$$

Being $\|P_h\| \leq M$ and $\|(I-A_{h,\Delta})^{-1}\| \leq 1$, it is enough to prove (2.8) for $f \in S$. Note that it is possible to write

$$A_{h,\Delta} = A_h (I-\Delta A_h)^{-1} = A_h J_{h,\Delta} \tag{2.9}$$

so that, for any $g \in D(A_h)$

$$(I-A_{h,\Delta})^{-1} g = (I-A_h(I-\Delta A_h)^{-1})^{-1} g = [(I-(\Delta+1)A_h)(I-\Delta A_h)^{-1}]^{-1} g =$$

$$= (I-(\Delta+1)A_h)^{-1}(I-\Delta A_h) g = J_{\Delta+1,h}(g-\Delta A_h g)$$

Therefore, if $f \in S$

$$\|(I-A_{h,\Delta})^{-1} P_h f - P_h (I-A)^{-1} f\| \leq \|\Delta J_{\Delta+1,h} A_h P_h f\| +$$

$$+ \|J_{\Delta+1,h} P_h f - P_h J_1 f\| \leq \Delta \|A_h P_h f\| + \|J_{\Delta+1,h} P_h f - P_h J_1 f\| \tag{2.10}$$

By (2.6) it is clear that for each $f \in S$ there exists $K>0$ such that, for $h>0$

$$\|A_h P_h f\| \leq K \tag{2.11}$$

so that the first term in the r.h.s. of (2.10) goes to zero as $(h,\Delta)\to 0$. The same assumption implies that, for $f\in S$

$$\lim_{(h,\Delta)\to 0} \|(\Delta+1)A_h P_h f - P_h A f\| = 0 \tag{2.12}$$

and by Theorem 2.2 the norm of

$$(I-(\Delta+1)A_h)^{-1}P_h f - P_h(I-A)^{-1}f = J_{\Delta+1,h}P_h f - P_h J_1 f$$

goes to zero as $(h,\Delta)\to 0$. This proves that for any $f\in L$, $T>0$

$$\lim_{(h,\Delta)\to 0}\ \sup_{t\in[0,T]} \|e^{A_{h,\Delta}t}P_h f - P_h e^{At} f\| = 0 \tag{2.13}$$

To get (2.7) it is enough again to consider $f\in S$ in (2.5). Use the fact that $\|P_h\|\leq M$ and, by (2.9) and (2.11), that

$$\|A_{h,\Delta}P_h f\| = \|J_{h,\Delta}A_h P_h f\| \leq \|A_h P_h f\| \leq K,\ f\in S, \tag{2.14}$$

to show that the r.h.s. of (2.7) is uniformly bounded with respect to h and can be made uniformly small for $t\in[0,T]$ with an appropriate choice of ε and taking Δ sufficiently small. So the fact that $\{e^{A_{h,\Delta}t}\}$ has the same asymptotic behaviour as $\{J_{h,\Delta}^{[t/\Delta]}\}$ (as both h and Δ goes to zero) is obtained and (2.7) is finally established.

The condition (2.6) is slightly stronger than the mere convergence of the semigroups $\{e^{A_h t}\}$ to $\{e^{At}\}$ in that in (2.3) the particular choice $f_h = P_h f$ is made. But it is interesting that this condition involves only A_h; it does not require to compute $J_{h,\Delta}$, that is to solve $(I-\Delta A_h)f = g$ for all possible $g\in L_h$, instead that for one g at a time.

3. FELLER-DYNKIN SEMIGROUPS AND WEAK CONVERGENCE

Let us consider a particular class of Markov semigroups. Let E and E_h be complete, separable, locally compact spaces and η_h be a measurable map from E_h to E, for h>0. Let (X_t) and (Z_t^h), h>0 be Markov processes with respect to their own families of σ-algebras, possibly defined on different probability spaces. We suppose that the corresponding algebraic semigroups of operators $\{T_t^h\}$ and $\{T_t\}$ on $L^\infty(E_h)$ and $L^\infty(E)$, respectively, are Feller-Dynkin [6]. If we denote by $\hat{C}(E_h)$ ($\hat{C}(E)$) the Banach space of continuous functions on $E_h(E)$, which go to zero at infinity (when the one-point compactification is done), this means that these semigroups are strongly continuous on $\hat{C}(E_h)(\hat{C}(E))$. Let η_h a continuous map of E_h into E (if E_h is not compact η_h has to map the infinity of E_h into the infinity of E, but this is not usually the case). Those maps induce corresponding bounded linear transformations of C(E) into $\hat{C}(E_h)$ by

$$(P_h f)(x) = f(\eta_h(x)), \qquad x \in E_h \tag{3.1}$$

Finally define $X_t^h = \eta_h(Z_t^h)$, h>0 and observe that this processes have versions with sample path in the Skorohod space $D[0,\infty;E]$ of E-valued cadlag functions [29]. For metrics on this space we refer to [21,15]; a particular case will be discussed in the next section.

It is quite clear that the convergence of $\{T_t^h\}$ to $\{T_t\}$ as $h \to 0$ in the sense of (2.1) relates the expectation of functionals of the corresponding processes at each instant of time. The following important theorem, due to Kurtz [14] involves the whole sample paths.

THEOREM 3.1. If $\{T_t^h\}$ converges to $\{T_t\}$ and X_0^h converges weakly to X_0, then (X_t^h) converges weakly to (X_t), as $h \to 0$, as a $D[0,\infty;E]$-valued random variable.

The previous theorem refers only to continuous-time Markov processes, but it can be easily extended. The argument used in Section 2 has in fact a stochastic interpretation. In fact, let A_h be the infinitesimal generator of $\{T_t^h\}$ on $C(E_h)$ and Q_h its transition function: then, for $\Delta>0$, $f\varepsilon\hat{C}(E_h)$

$$(I-\Delta A_h)^{-1}f(x) = \int_{E_h} f(y)\left[\Delta^{-1}\int_0^\infty e^{-\Delta^{-1}t}Q_h(t,x,dy)dt\right], \qquad x\varepsilon E_h \tag{3.2}$$

which shows that a discrete-time Markov process $(Z_{k\Delta}^{h,\Delta}, k=0,1,\ldots)$ can be built such that for each $f\varepsilon\hat{C}(E_h)$

$$E(f(Z_{(k+1)\Delta}^{h,\Delta})|Z_{k\Delta}^{h,\Delta} = x) = (I-\Delta A_h)^{-1}f(x) = J_{h,\Delta}f(x). \tag{3.3}$$

Now it is quite easy to show that $A_{h,\Delta}$ defined in (2.4) is the infinitesimal generator of a Feller-Dynkin process $(\tilde{Z}_t^{h,\Delta})$, which can be obtained with a random time change which turns the intervals between the jumps of $(Z_{k\Delta}^{h,\Delta})$ to be i.i.d. exponential variables with mean Δ. Moreover the distance between the processes $\tilde{X}_t^{h,\Delta} = \eta_h(\tilde{Z}_t^{h,\Delta})$ and $X_{[t/\Delta]\Delta}^{h,\Delta} = \eta_h(Z_{[t/\Delta]\Delta}^{h,\Delta})$ in the $D[0,\infty;E]$-metric goes to zero in probability as (h,Δ) goes to zero [15]. This allows to modify the previous theorem in the following way.

THEOREM 3.2. If $\{T_t^h\}$ converges to $\{T_t\}$ and $X_0^{h,\Delta}$ converges weakly to X_0, then $(X_{[t/\Delta]\Delta}^{h,\Delta})$ converges weakly to (X_t), as a $D[0,\infty;E]$-valued random variable.

4. APPLICATION TO FILTERING

We return to the problem stated in Introduction by identifying the "copy" of the state process in (1.2) with the Feller-Dynkin one of the last section. Moreover we have now to suppose that this process is continuous (so that, at least locally, it is a diffusion [8]). We need to introduce an extension of its infinitesimal generator, called the full generator [15], which is a possibly multivalued operator $\tilde{A} \subset L^\infty(E) \times L^\infty(E)$ such that

$$(g,h) \in \tilde{A} \quad g(X_t) - \int_0^t h(X_s)ds$$

is a martingale w.r.t. the increasing family of σ-algebras generated by (X_t).

Suppose now that each component of g in (1.1) is bounded, uniformly continuous and belongs to $D(\tilde{A})$, and the products $g_i g_j$, $i,j=1,\dots,m$, too. We let $\tilde{A}g = (\tilde{A}g_1,\dots,\tilde{A}g_m)$ where $\tilde{A}g_i$ stands for any element of the $\tilde{A}$-image of g_i, $i=1,\dots,m$. By integrating by parts inside the expectations of the Kallianpur-Striebel formula (1.2) this can be expressed for any path $y \in C_0[0,\infty;R^m]$ (continuous functions which starts from zero) of (Y_t) through the "robust" version [5]

$$\overline{E}(f(\overline{X}_t)|Y_s=y(s),\ 0\leq s\leq t) = \frac{E[f(X_t)\exp(y^T(t)g(X_t) - \int_0^t y^T(s)dg(X_s) - \frac{1}{2}\int_0^t |g(X_s)|^2 ds)]}{E[\exp(y^T(t)g(X_t) - \int_0^t y^T(s)dg(X_s) - \frac{1}{2}\int_0^t |g(X_s)|^2 ds)]} \tag{4.1}$$

for any f bounded and measurable.

In fact, by assumption $g(X_t)$ is a R^m-valued semi-martingale whose decomposition is given by

$$g(X_t) = g(X_0) + \int_0^t (\tilde{A}g)(X_s)ds + M_t \tag{4.2}$$

where M_t is a continuous square-integrable martingale having the matrix-valued increasing process [10]

$$\langle M\rangle_t^{i,j} = \int_0^t [(\tilde{A}g_i g_j)(X_s) - g_j(X_s)(\tilde{A}g_i)(X_s) - g_i(X_s)(\tilde{A}g_j)(X_s)]ds \tag{4.3}$$

Being $\langle M\rangle_t$ locally bounded, $R_t = \int_0^t y^T(s)dM_s$ and $\exp(R_t - \frac{1}{2}\langle R\rangle_t)$ are martingales [23], from which the boundedness of the denominator is easily obtained for each $y \in C_0[0,\infty;R^m]$. By Riesz theorem this implies that for each y there exist finite measures $\mu_t(y)$, $t \geq 0$, such that

$$\overline{E}(f(\overline{X}_t)|Y_s = y(s),\ 0 \leq s \leq t) = \frac{\langle f, \mu_t(y)\rangle}{\langle 1, \mu_t(y)\rangle},\ t \geq 0 \tag{4.4}$$

where $\langle f,\mu\rangle = \int_E f(x)\mu(dx)$ and

$$\Pi_t(y)(dx) = \frac{\mu_t(y)(dx)}{\int_E \mu_t(y)(dz)} \tag{4.5}$$

is a regular conditional probability measure.

Now let us suppose that, for $h>0$, A_h is the infinitesimal generator of a continuous-time Markov chain with finite state space $E_h = \{1,2,\dots,N_h\}$. Let η_h associate with each state i a point x_i^h in E and for each function on E let $P_h f$ be the N_h-vector of the evaluations of f at points $\{x_1^h,\dots,x_{N_h}^h\}$, which is always considered with the sup norm. We are allowed to identify E_h with $\eta_h(E_h)$ and the values of $P_h f$ with those of f in the sequel. For $\Delta>0$ and $y \in C_0[0,\infty;R^m]$ let us consider the implicit time-discretization equation (1.4), which can be rewritten as

$$q_{(k+1)\Delta}^{h,\Delta} = J_{h,\Delta}^T\ B_{k\Delta}^{\Delta}\ q_{k\Delta}^{h,\Delta},\ k = 0,1,\dots, \tag{4.6}$$

$B_{k\Delta}^{\Delta}$ being a N_h-th order diagonal matrix, whose i-th diagonal element is $\exp(g^T(x_i^h)(y((k+1)\Delta)-y(k\Delta)) - \frac{1}{2}\Delta|g(x_i^h)|^2))$.

Let us consider on some probability space the discrete-time Markov chain

$(X_{k\Delta}^{h,\Delta}$, $k=0,1,\dots)$ with initial probability vector q_0^h (it is supposed $1^T q_0^h = 1$) and transition matrix $J_{h,\Delta}$.

THEOREM 4.1. The solution of equation (4.6) can be expressed in the following way: for any $f \in R^{N_h}$

$$f^T q_{k\Delta}^{h,\Delta} = E\lfloor f(X_{k\Delta}^{h,\Delta})\exp(\sum_{\ell=0}^{k-1} \lfloor g^T(X_{\ell\Delta}^{h,\Delta})(y((\ell+1)\Delta) - y(\ell\Delta)) - \frac{1}{2} |g(X_{\ell\Delta}^{h,\Delta})|^2 \Delta \rfloor)\rfloor. \quad (4.7)$$

Proof. It requires only a substitution of (4.7) into (4.6) which yields

$$f^T J_{h,\Delta}^T B_{k\Delta}^{\Delta} q_{k\Delta}^{h,\Delta} = (B_{k\Delta}^{\Delta} J_{h,\Delta} f)^T q_{k\Delta}^{h,\Delta} =$$

$$= E\lfloor E(f(X_{(k+1)\Delta}^{h,\Delta}) | X_{k\Delta}^{h,\Delta})\exp(\sum_{\ell=0}^{k} \lfloor g^T(X_{\ell\Delta}^{h,\Delta})(y((\ell+1)\Delta) - y(\ell\Delta)) - \frac{1}{2} |g(X_{\ell\Delta}^{h,\Delta})|^2 \Delta \rfloor)\rfloor$$

which is equal to $f^T q_{(k+1)\Delta}^{h,\Delta}$ by the Markov property of $(X_{k\Delta}^{h,\Delta})$ and the projective property of conditional expectations.

It can be easily shown that (4.7) gives the numerator of a Kallianpur-Striebel type formula for an estimation problem in discrete time.

We can extend the function $q_{k\Delta}^{h,\Delta}$ to continuous time by

$$f^T q_t^{h,\Delta}(y) = E\lfloor f(X_{\lfloor t/\Delta \rfloor \Delta}^{h,\Delta})\exp(\int_0^t g^T(X_{\lfloor s/\Delta \rfloor \Delta}^{h,\Delta})\dot{y}(s)ds - \frac{1}{2}\int_0^t |g(X_{\lfloor s/\Delta \rfloor \Delta}^{h,\Delta})|^2 ds)\rfloor \quad (4.8)$$

if y is in $C_0^1\lfloor 0,\infty;E\rfloor$. A similar expression holds in this case for μ_t

$$\langle f,\mu_t(y)\rangle = E\lfloor f(X_t)\exp(\int_0^t g(X_s)\dot{y}(s)ds - \frac{1}{2}\int_0^t |g(X_s)|^2 ds)\rfloor \quad (4.9)$$

Now let $p_t^{h,\Delta}(y) = q_t^{h,\Delta}(y)/(1^T q_t^{h,\Delta}(y))$. The following theorem states the relevant consequence of Theorem 2.4 for our problem.

THEOREM 4.2. Let us suppose that the convergence condition (2.6) holds and

p_0^h converges weakly to the law of X_0, as $h \to 0$. Then, for each $y \in C_0^1[0,\infty;R^m]$, $T>0$ and f bounded and uniformly continuous

$$\lim_{(h,\Delta)\to 0} \sup_{t\in[0,T]} |f^T p_t^{h,\Delta}(y) - \langle f, \mu_t(y)\rangle| = 0 \tag{4.10}$$

Proof. For each y as above define the function $\varphi_1 : D[0,\infty;E] \to D[0,\infty]$ as

$$\varphi_1(x)(t) = \exp\left(\int_0^t g^T(x(s))\dot{y}(s)ds - \frac{1}{2}\int_0^t |g(x(s))|^2 ds\right) \tag{4.11}$$

and observe that for each $T>0$ there exist two real constants $\underline{K}$ and $\overline{K}$ such that

$$\underline{K} \leq \log \varphi_1(x)(t) \leq \overline{K}, \qquad x \in D[0,\infty;E],\ t\in[0,T] \tag{4.12}$$

Therefore, if $\varphi_f(x)(t) = f(x(t))\varphi_1(x)(t)$, then

$$\sup_{t\in[0,T]} \left| \frac{f^T q_t^{h,\Delta}(y)}{1^T q^{h,\Delta}(y)} - \frac{\langle f, \mu_t(y)\rangle}{\langle 1, \mu_t(y)\rangle} \right| \leq \left(\inf_{t\in[0,T]} 1^T q_t^{h,\Delta}(y)\right)^{-1} \left(\sup_{t\in[0,T]} |f^T q_t^{h,\Delta}(y) - \langle f, \mu_t(y)\rangle|\right) +$$

$$+ \left(\inf_{t\in[0,T]} \langle 1, \mu_t(y)\rangle \inf_{t\in[0,T]} 1^T q_t^{h,\Delta}(y)\right)^{-1} \left(\sup_{t\in[0,T]} |\langle f, \mu_t(y)\rangle| \sup_{t\in[0,T]} |1^T q_t^{h,\Delta}(y) - \langle 1, \mu_t(y)\rangle|\right) \leq$$

$$\leq e^{-\underline{K}} E(\|\varphi_f(X^{h,\Delta}) - \varphi_f(X)\|_{\infty,T}) + e^{-2\underline{K}} e^{\overline{K}} \|f\| E(\|\varphi_1(X^{h,\Delta}) - \phi_1(X)\|_{\infty,T}) \tag{4.13}$$

where $\|\cdot\|_{\infty,T}$ stands for the sup norm on $[0,T]$. Note that we have placed $(X_t^{h,\Delta})$, $h>0, \Delta>0$ and (X_t) on the same probability space: by the weak convergence assured by Theorems 2.4 and 3.2 this can be done even assuring that $(X_t^{h,\Delta}(\omega))$ converges to $(X_t(\omega))$ in $D[0,\infty;E]$ for each ω [4]. Being the paths of (X_t) continuous this implies uniform convergence on each compact set. But f and g are uniformly continuous, so that the two terms under the expectation sign in (4.13) converge to zero for each ω. By bounded convergence theorem, the proof is accomplished.

The successive step will be to extend the convergence in (4.10) to all possible are reduced to the corresponding ones for $\psi^{h,\Delta}_{\phi_i}(y)$ and $\psi_{\phi_i}(y)$, $i = 1,2,\ldots$.

This allows to solve the following "robustness" problem. Let us consider continuous processes $(\overline{X}^{\varepsilon}_t, W^{\varepsilon}_t)$, $\varepsilon > 0$, considered as $C[0,\infty;E] \times C_0[0,\infty;R^m]$ - valued random variables, which are a family of "physical" state and noise models depending on some parameter, converging to the "ideal" diffusion plus white noise model of the Introduction $(\overline{X}_t, W_t)$ as this parameter degenerates. The typical situations to have in mind are carefully reviewed in [16]. Note that the output map defined in (1.1) is defined on each sample path (X,W) of the state and noise processes, yielding a continuous map

$$y: C[0,\infty;E] \times C_0[0,\infty;R^m] \to C_0[0,\infty;R^m] \tag{5.6}$$

with all the spaces endowed with the metric of uniform convergence on compact intervals.

The approximate filter (5.5) is applied to the "physical" output process $Y^{\varepsilon} = y(\overline{X}^{\varepsilon}, W^{\varepsilon})$. The following result extends the similar one proved by Kushner [19] for one particular chain in continuous time, in the meantime giving a more direct proof in that unnormalized conditional probabilities are not used.

THEOREM 5.1. Let us suppose that $(\overline{X}^{\varepsilon}_t, W^{\varepsilon}_t)$ converges weakly to $(\overline{X}_t, W_t)$ as $\varepsilon \to 0$, where $(\overline{X}_t)$ is a continuous E-valued Feller-Dynkin process and (W_t) an independent R^m-valued standard Brownian motion. Then, under the hypotheses of Corollary 4.1, the process $(\overline{X}^{\varepsilon}_t, W^{\varepsilon}_t, p^{h,\Delta}_t(Y^{\varepsilon}))$ converges weakly to $(\overline{X}_t, W_t, \Pi_t(Y))$ as $(\varepsilon,h,\Delta) \to 0$, considered as $C[0,\infty;E] \times C_0[0,\infty;R^m] \times D[0,\infty;\mathcal{P}(E)]$ - valued random variables.

Proof. For $\varepsilon > 0$, $h > 0$, $\Delta > 0$, define the functions

$$\lambda^{\varepsilon,h,\Delta}: C[0,\infty;E] \times C_0[0,\infty;R^m] \to C[0,\infty;E] \times C_0[0,\infty;R^m] \times D[0,\infty;(E)]$$

$$\lambda^{\varepsilon,h,\Delta}(\overline{x},w) = (\overline{x},w,p^{h,\Delta}(y(\overline{x},w))) \tag{5.7}$$

and χ, with the same domain and range space, defined by

$$\chi(\overline{x},w) = (\overline{x},w,\Pi(y(\overline{x},w))). \tag{5.8}$$

Let $\overline{X}^{\varepsilon,h,\Delta} = \overline{X}^{\varepsilon}$ and $W^{\varepsilon,h,\Delta} = W^{\varepsilon}$. By the remarks following Corollary 4.1, $\chi^{\varepsilon,h,\Delta}$ converges to χ uniformly on compact sets. Being χ continuous, it suffices to to apply Theorem 5.5 in [3] to get the desired result. $\square$

A comprehensive discussion of the meaning of weak convergence-type results like Theorem 5.1 is given in [14]. However, again, the important thing is to note that the way ε,h,Δ approach zero cannot destroy convergence. We believe that those results could be of particular importance for sequential decision problems on partially observed diffusions [1].

6. TWO EXAMPLES

Let us first consider the chain proposed by Kushner, which is obtained by suitably modifying a simple difference scheme applied to the generator of a diffusion [16]. We limit ourselves to the one-dimensional case in which such a scheme always works.

Let $C_b^k(\hat{C}^k)$ the space of k-times continuously differentiable function on R, with all those derivatives bounded (which go to zero at infinity). Let $(\overline{X}_t)$ be the solution of the martingale problem with full generator

$$(\tilde{A}f)(x) = \frac{1}{2}\, a(x) \frac{\partial^2 f}{\partial x}(x) + b(x) \frac{\partial f}{\partial x}(x) \quad , \ f \in C_b^2 \tag{6.1}$$

where it must be supposed that $a(x) \geq \lambda > 0$ for $x \in R$ and $a(\cdot)$ and $b(\cdot)$ are bounded and Hölder, in order to have a well-posed martingale problem [28] and the restriction A of $\tilde{A}$ to $C_b^2 \cap \hat{C}$ to be extendible to a generator of a Feller-Dynkin semigroup on $\hat{C}$ [6]. But we need also to use $\hat{C}^2$ as a core, and for $\hat{C}^2$ to be invariant under e^{At}, $a(\cdot)$ and $b(\cdot)$ have to be also in C_b^2. In this case, the parabolic equation $(\partial/\partial t - A)f = 0$ can be differentiated twice w.r.t. the space variable [25].

For each $h > 0$, let us consider a finite grid G_h of equispaced points of distance h, which tends to cover the whole line as $h \to 0$, and define a Markov chain on G_h by the following non-zero intensities: for $x \in G_h$

$$\begin{aligned} a^h(x,x-h) &= \frac{1}{2h^2}\, a(x) + \frac{1}{h}\, b^-(x) \\ a^h(x,x) &= -\frac{1}{2h^2}\, a(x) - \frac{1}{h}\, |b(x)| \\ a^h(x,x+h) &= \frac{1}{2h^2}\, a(x) + \frac{1}{h}\, b^+(x) \end{aligned} \tag{6.2}$$

except for the first and the last point of the grid, which are made absorbent. Let η_h be the inclusion of G_h into R, and let P_h be defined as in Section 4.

It is clear that condition (2.6) is satisfied once we show that for $f \varepsilon C^2$

$$\sup_{x \varepsilon R} \left| \frac{1}{2} a(x) \frac{\partial^2 f}{\partial x^2} + b(x) \frac{\partial f}{\partial x} - \left(\frac{1}{2h^2} a(x) + \frac{1}{h} b^-(x)\right) f(x-h) + \left(\frac{1}{2h^2} a(x) + \frac{1}{h} |b(x)|\right) f(x) + \right.$$

$$\left. - \left(\frac{1}{2h^2} a(x) + \frac{1}{h} b^+(x)\right) f(x+h) \right| \leq \delta(h) \qquad (6.3)$$

where $\delta(h)$ goes to zero as $h \to 0$. The behavior at the boundary is controlled by the boundedness assumptions on a and b and the fact that $f \varepsilon \hat{C}^2$. The expression of the r.h.s. of (6.3) can be rewritten as

$$\frac{a(x)}{2h^2} (f(x+h) + f(x-h) - 2f(x) - f''(x)h^2)) + \frac{b^+(x)}{h} (f(x+h) - f(x) - f'(x)h) +$$

$$- \frac{b^-(x)}{h} (f(x-h) - f(x) - f'(x)h)$$

which clearly shows uniform convergence (f" is in fact uniformly continuous). Moreover, for any $g \varepsilon C_b^2$, the boundedness condition (4.16) is verified, and Corollary 4.1 and Theorem 5.1 can be applied.

Such method can be extended to the case R^d, $d > 1$, with additional assumptions on the coefficients [16]. The verification of conditions (2.7) and (4.9) is still straightforward. It is clear that the method could take into account boundary conditions, too.

Example 2. This rather artificial example serves only as a sample to show that reasonable alternatives to the previous space discretization scheme exist, even in dimension one. Of course, this is much more true in higher dimensions, given that the complexity of the topology of a grid increases. Suppose that $a \equiv 1$ in (6.1), and write $b = -\partial V/\partial x$. Usually, it will be easier to compute the "potential" V than its derivative so that it makes sense to define the following approximating chain, holding fixed the grid G_h as before:

$$\bar{a}^h(x,x-h) = \begin{cases} \frac{1}{h^2}\left(\frac{1}{2} + V(x) - V(x-h)\right), \text{ if } V(x-h) \leqslant \min\{V(x),\ V(x+h)\} \\ \frac{1}{2h^2} \text{ otherwise,} \end{cases}$$

$$\bar{a}^h(x,x+h) = \begin{cases} \frac{1}{h^2}\left(\frac{1}{2} + V(x) - V(x+h)\right), \text{ if } V(x+h) < \min\{V(x),\ V(x-h)\} \\ \frac{1}{2h^2} \text{ otherwise} \end{cases} \qquad (6.4)$$

letting $a^h(x,x) = -(a^h(x,x-h) + a^h(x,x+h))$ and the other terms to be zero (including the boundary ones). Condition (2.6) is reduced to checking

$$\sup_{x \in R} \left| \frac{1}{h^2} \lfloor (V(x) - V(x \pm h))f(x \pm h) - (V(x) - V(x \pm h))f(x) \rfloor + \frac{\partial V}{\partial x} f(x) \right| \leqslant \delta(h)$$

for each $f \in \hat{C}^2$, where $\delta(h)$ goes to zero as $h \to 0$. This is because, when the V-terms in (6.4) repeatedly disappear around x, as $h \to 0$, it is necessarily $(\partial V/\partial x)(x) = 0$. This allows to prove the boundedness condition for any $g \in C_b^2$, so the convergence property of the filtering algorithm derived from (6.4) is the same as in the previous example.

ACKNOWLEDGEMENTS

The author wishes to thank Professor J. S. Baras for his valuable guidance during the realization of the present work.

REFERENCES

1. J. S. Baras, A. La Vigna, "Expert systems and VLSI architectures for real-time non-Gaussian detectors and filters", in C. Byrnes and A. Lindquist eds., Proceedings of MTNS-85, North-Holland, to appear.

2. Ya. I. Belopol'skaya, Z. I. Nagolkina, "On a class of stochastic partial differential equations," Th. of Prob. and its Appl., 27, 592-599, 1982.

3. P. Billingsley, Convergence of Probability Measures, John Wiley, New York, 1968.

4. P. Billingsley, Weak Convergence of Measures: Applications in Probability, SIAM, Philadelphia, 1971.

5. J. M. C. Clark, "The design of robust approximations to the stochastic differential equations of nonlinear filtering," in J. K. Skwirzynski ed., Communication Systems and Random Process Theory, Sijthoff and Noordhoff, Aalpen aan den Rijn, 1978.

6. E. B. Dynkin, Markov Processes - I, Springer-Verlag, Berlin, 1965.

7. A. Germani, M. Piccioni, "A Galerkin approximation for the Zakai equation," in P. Thoft-Christensen, ed., Systems Modelling and Optimization, Springer-Verlag, Berlin, 1984.

8. I. I. Gihman, A. V. Skorohod, The Theory of Stochastic Processes, III, Springer-Verlag, New York, 1979.

9. E. Hille, Functional Analysis and Semigroups, AMS, New York, 1948.

10. J. Jacod, Calcul Stochastique et Problémes de Martingales, Springer-Verlag, Berlin, 1977.

11. G. Kallianpur, C. Striebel, "Estimation of stochastic systems: arbitrary system process with additive white noise observation errors," Ann. Math. Stat., 39, 785-801, 1968.

12. T. Kato, Perturbation Theory for Linear Operators, Springer-Verlag, Berlin, 1976.

13. T. G. Kurtz, "Extensions of Trotter's operator semigroup approximation theorems," J. Funct. Anal., 3, 354-375, 1969.

14. T. G. Kurtz, "Semigroups of conditional shifts and approximations of Markov processes," Ann. Prob., 4, 618-642, 1975.

15. T. G. Kurtz, Approximation of Population Processes, SIAM, Philadelphia, 1981.

16. H. J. Kushner, Probability Methods for Approximation in Stochastic Control and for Elliptic Equations, Academic Press, New York, 1977.

17. H. J. Kushner, "A robust discrete state approximation to the optimal nonlinear filter for a diffusion," Stochastics, 3, 75-83, 1979.

18. H. J. Kushner, Approximation and Weak Convergence Methods for Random Processes, MIT, Cambridge, 1984.

19. H. J. Kushner, H. Huang, "Approximate and limit results for nonlinear filters with wide bandwidth observation noise," Report LCDS 84-36, 1984.

20. F. Legland, "Estimation de paramètres dans les processus stochastiques en observation incomplète," Thèse, Université Paris IX, 1981.

21. T. Lindvall, "Weak convergence of probability measures and random functions in the function space $D[0,\infty)$," J. Appl. Prob., 10, 109-121, 1973.

22. P.-A. Meyer, Martingales and Stochastic Integrals I, Springer-Verlag, Berlin, 1978.

23. M. Metivier, Semimartingales, W. de Gruyter, Berlin, 1982.

24. R. D. Richtmyer, K. W. Morton, Difference Methods for Initial-Value Problems, Interscience, New York, 1967.

25. B. Stewart, "Generation of analytic semigroups by strongly elliptic operators," Trans. Amer. Math. Soc., 199, 141-162, 1974.

26. D. W. Stroock, S. R. S. Varadhan, Multidimensional Diffusion Processes, Springer-Verlag, Berlin, 1979.

27. H. Trotter, "Approximation of semigroups of operators," Pacific J. Math., 8, 887-919, 1958.

28. H. Trotter, "On the product of semigroups of operators," Proc. Amer. Math. Soc., 10, 545-551, 1959.

29. D. Williams, Diffusions, Markov Processes and Martingales, Vol. 1, John Wiley, Chichester, 1979.

30. K. Yosida, Functional Analysis, Springer-Verlag, Berlin, 1968.

31. M. Zakai, "On the optimal filtering of diffusion processes," Z. Wahr. verw. Geb., 11, 230-243, 1969.