JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 110, 364–383 (1985)

Discrete Time Galerkin Approximations to the Nonlinear Filtering Solution

JOCELYN FREITAS BENNATON

Departamento Eletricidade, Escola Engenharia de São Carlos Universidade de São Paulo, São Carlos, São Paulo, Brasil

Submitted by Harold Kushner

This paper concerns discrete time Galerkin approximations to the solution of the filtering problem for diffusions. Two families of schemes approximating the unnormalized conditional density, respectively, in an "average" and in a "pathwise" sense, are presented. L^2 error estimates are derived and it is shown that the rate of convergence is linear in the time increment or linear in the modulus of continuity of the sample path.

INTRODUCTION

Filtering problems arise in a variety of practical situations when direct observation of a stochastic process is not possible. The data concerning this so-called signal process is provided by observation on another process which is related to the signal by a dynamic model perturbed by noise. The problem of estimating the conditional probability of the signal given the observed data is called the filtering problem. When the data is coming in continuously, it is worth having a filter which can be revised step by step in order to take into account the new data. This can be done and an interesting way of doing this is by obtaining stochastic differential equations for the filter. However, as the estimate in general depends nonlinearly on the observations, solving these equations it is not an easy task and so numerical approximations are required.

This paper concerns discrete time approximations to the solution of the "standard" nonlinear filtering problem. In this case the signal is a Markov diffusion process and the observation evolves according to a "signal + white noise" model. If some conditions hold it turns out that the dynamic part of the filter can be represented by means of the Zakai formula for the "unnormalized" conditional density of the diffusion. A practical as well as mathematically interesting aspect of the Zakai equation is that it admits a linear "nonstochastic" counterpart. The latter produces a version of the "unnormalized" conditional density which depends smoothly of the obser-

0022-247X/85 $3.00

vation sample paths and so a nice "pathwise' solution of the filtering problem can be obtained.

Both the stochastic and the "nonstochastic" versions of the "unnormalized" conditional density evolve according to parabolic differential forms. Since Galerkin approximation methods have proved to be successful in solving deterministic partial differential equations, it seems natural to extend these methods to stochastic equations. Here, it will be shown that discrete time Galerkin procedures can also be used in order to approximate the filtering solution.

In the next section a brief account of the filtering equations is given. In Section 2, discrete time Galerkin schemes are introduced. They have a "finite element" discretization in the space domain and an implicit Runge–Kutta form in the time domain discretization. In order to achieve a better rate of convergence these schemes must incorporate second order powers of the observation process.

Convergence theorems and estimates for the approximation error of the discrete time Galerkin methods are the main results of this work. They are presented in Sections 3 and 4 and can be resumed as follows: discrete time numerical approximations can be produced in such a way that they (i) converge in an "average" sense with a linear rate in the time increment to the solution of the Zakai equation; (ii) converge in a "pathwise" sense with a linear rate in the modulus of continuity of the sample path to the solution of the "pathwise" version of the Zakai equation. According to previous work (Clark [3, 4]), dedicated to the study of the convergence of numerical approximations to "ordinary" stochastic equations, we believe that these rates are optimal.

Notation

(i) S is a domain in R^n.

(ii) $\mathbb{C}$, L^∞, L^2, denote, respectively, the spaces of continuous, bounded and square integrable real-valued functions defined in S.

(iii) H^m, H_0^m, $m=0, 1,\ldots$, denote, respectively, the Sobolev spaces $W^{m,2}(S)$ and $W_0^{m,2}(S)$ with the usual norm,

$$\|u\|_m = \left(\sum_{|\alpha| \leqslant m} |D_u^\alpha|^2 \right)^{1/2}$$

where $|\cdot| = (\cdot, \cdot)^{1/2}$ is the standard norm in $L^2 = H^0$ and D^α symbolizes the partial derivative of order α_i in x_i, $i = 1,\ldots, n$ with $|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_n$. If $\mathbb{C}^m = \{u \in \mathbb{C}: D^\alpha u \in \mathbb{C}, |\alpha| \leqslant m\}$, we recall that H^m can be defined as the completion of the subset of $\mathbb{C}^m$ consisting of functions u such that

$\|u\|_m < \infty$. If $\mathbb{C}_0^\infty$ is the set of infinitely differentiable functions with compact support in S, H_0^m can be defined as the closure of $\mathbb{C}_0^\infty$ in H^m.

(a) We reserve the symbol $\|\cdot\| \equiv \|\cdot\|_1$ for the norm in H^1 or H_0^1.

(b) The letter K will be used always as an upper bound constant in a great variety of inequalities. Therefore K can take different values in different equations. When it appears this must be translated by "there exists a constant K such that..."

1. *The Filtering Equations*

Although the filtering problem can be formulated for a wide variety of signal–observation pair of processes, this paper will be restricted to the "standard" case. The signal x_t is a Markov diffusion process taking values in $S = R^n$ and the observation y_t, $0 \leqslant t \leqslant T$, is related to the signal by,

$$y_t = \int_0^t h(x_s)\, ds + w_t, \tag{1}$$

where w_t is a standard Brownian motion. Assume the pair (x_t, y_t) satisfying the following set of hypotheses:

(H1) The diffusion x_t is a homogeneous process with differential generator L given by,

$$L = \frac{1}{2} \sum_{i,j=1}^n a_{i,j} \frac{\partial^2}{\partial x_i\, \partial x_j} + \sum_{i=1}^n g_i \frac{\partial}{\partial x_i}.$$

(H2) The diffusion and drift coefficients of x_t are such that, for $i, j = 1,..., n$,

$$a_{i,j} \in \mathbb{C} \cap L^\infty, \qquad g_i, \frac{\partial a_{i,j}}{\partial x_i} \in L^\infty.$$

(H3) The diffusion matrix of x_t is positive definite: there exists $\sigma > 0$ such that,

$$\sum_{i,j=1}^n a_{i,j}(x)\, \xi_i \xi_j \geqslant \sigma \sum_{i=1}^n \xi_i^2$$

for all x, $\xi \in R^n$.

(H4) π, the probability density of x_0 belongs to L^2.

(H5) $h \in \mathbb{C} \cap L^\infty$.

(H6) The Brownian motion w_t and the diffusion x_t are independent stochastic processes in some probability space $(\Omega, \mathscr{A}, P_0)$.

Remark 1. Hypotheses (H1)–(H6) enable us to construct the diffusion x_t by means of the Kolmogorov's evolution equations governing its transition probability density. In particular, the strong ellipticity condition (H3) guarantees a unique solution for these equations (Dynkin, [5]). Hypotheses (H5) and (H6) concerns the filtering problem associate with the pair (x_t, y_t) that is, the problem of calculating the probability density of the signal x_t conditioned by the observation $y_s: 0 \leqslant s \leqslant t$.

Similar to what happens in the "unconditional" case, we can again make use of evolution equations in order to describe the law governing the conditioned diffusion x_t. The steps leading to these equations will now be briefly described. Let z_t be the stochastic process,

$$z_t = \exp\left(\int_0^t h(x_s)\, dy_s - \tfrac{1}{2}\int_0^t h^2(x_s)\, ds\right).$$

It follows from our hypotheses that the relation $dP = (z_T)^{-1}\, dP_0$ defines a new probability measure on $(\Omega, \mathscr{A})$. Besides, under measure P: *i*) the distributions of x_t are the ame as under P_0; (ii) the observation y_t becomes a standard Brownian motion; (iii) x_t and y_t are independent stochastic processes. This result was first introduced by Girsanov and its proof can be found in Liptser–Shiryaev [7].

The important consequence of the Girsanov transformation is that, since x_t and y_t are independent under P, we are allowed to "integrate out" x_t in order to obtain the conditional expectation. Following this idea, we arrive at the Zakai equation

$$du_t = L^* u_t\, dt + h u_t\, dy_t, \qquad (u_o = \pi), \tag{2}$$

where L^*, the adjoint of L is the Fokker–Planck operator. It can be shown (Pardoux [11]) that, under hypotheses (H1)–(H6) the equation above has a unique solution which satisfies, for all $f \in L^\infty \cap L^2$,

$$E(f(x_t)\, z_t / \mathscr{Y}_t) = \int_{R^n} f(x)\, u_t(x)\, dx, \tag{3}$$

where $\mathscr{Y}_t = \sigma - \{y_s : s \leqslant t\}$ and $E(\cdot)$ is the expectation relatively to P. Recalling standard formula relating conditional expectations of equivalent probability measures we can write,

$$E_0(f(x_t) \mid \mathscr{Y}_t) = \frac{E(f(x_t)\, z_t \mid \mathscr{Y}_t)}{E(z_t \mid \mathscr{Y}_t)} \qquad \text{(a.s.)}. \tag{4}$$

Therefore, up to a normalizing factor, the stochastic evolution equation (2) describes the law governing the density of the diffusion x_t conditioned by the observation y_t. If $h=0$, Eq. (2) reduces, as expected, to the Fokker–Planck equation governing the "unconditional density of x_t. From (2) using Ito's calculus, we can also obtain a dynamic representation for the conditional density of x_t. This is the Kushner–Stratonovich equation (Kushner [6]) which is quite similar to (2) but unfortunately has nonlinear terms in the stochastic integral.

According to what has been suggested, among others, by Clark [3], a nice "pathwise" solution for the filtering problem may also be found. Consider the following evolution equation, defined for all (continuous) sample paths $y(\cdot)$ of the observation process:

$$\dot{v}(t)=\exp(-hy(t))\{L^*-\tfrac{1}{2}h^2\}\exp(hy(t))\,v(t). \tag{5}$$

Using basically Ito's rule of transformation it can be shown that, given the solution of Eq. (2), the stochastic process v_t defined by $v_t=\exp(-hy_t)\,u_t$ has its sample paths described by Eqs. (5). It follows from (3) and (4) that the expression

$$\frac{\exp(hy(t))\,v(t)}{\int_{R^n}\exp(hy(t))\,v(t)\,dx}$$

is a version of the conditional density of x_t. So Eq. (5) can also be used as a mean of describing the dynamic part of the filter.

We conclude that, in order to calculate the probability density of x_t given an observed data $y(s)$, $0\leqslant s\leqslant t$, we can choose between: (i) solving the stochastic evolution equation (2); (ii) solving the ordinary evolution equation (5) parametrized by the observation sample path. Although they represent two complementary aspects of the same task, both these alternatives are identical as means of solving the filtering problem. We believe that this idea will become clear in this work.

We have introduced the formulas for the filtering problem where the diffusion process occurs in R^n. However, one can also consider the diffusion taking place in a bounded domain of R^n. Two main situations can be distinguished: (i) the diffusion is absorbed by the boundaries; (ii) the diffusion is reflected by a nonelastic boundary. Both cases have been studied by Pardoux [10] and the resulting filtering equations correspond, respectively, to Dirichlet and Neumann boundary conditions attached to Eq. (2) or (5).

In what follows we shall be concerned only with absorbed diffusions. If we write x_t^0 for a diffusion in the R^n, the diffusion absorbed by the boundary Γ of an open domain $S\subset R^n$ can be expressed by $x_t=x^0_{t\wedge\tau}$, where $\tau=\inf\{t\geqslant 0: x(t)\notin S\}$. In order to evaluate the conditional law of x_t one has to consider an "inner" density in S and a "superficial" density on Γ. To avoid the introduction of new formulas and hypotheses we can consider the function f in (3) as being zero for all $x\in\Gamma$. So, only the "inner" part of the density must be taken into account. The law governing this density will be given by Eq. (2) under Dirichlet boundary conditions, that is,

$$\begin{aligned} u_o(x) &= \pi, \qquad x \in S, \\ u_t(x) &= 0, \qquad (t, x) \in (0, T] \times \Gamma \end{aligned} \tag{6}$$

Remark 2. The stochastic equation (2) constitutes the "natural," dynamic representation of the filtering solution. However, we are giving to both equations, (2) and (5), the same status as means of calculating the filtering solution. This point of view is based on two interesting properties of the "pathwise" solution. First, the Fokker–Planck operator and the one in (5) differ only by a zeroth order term depending smoothly on the observation sample paths. In other words, the dynamic part of the law governing the conditional density of x_t can be made similar to that governing its "unconditional" density, with the same diffusion coefficients. Thus, the task of calculating the filtering solution is basically the same of calculating the density of x_t. Second, it is not difficult to show that the solution of (5) depends continuously on the sample paths of the observation process (Bennaton [1]). This fact is auspicious from the numerical analysis point of view. We can work with approximations of the observation sample path without the risk of "explosions" in the set of the solution paths. In other words, representation (5) is robust.

2. *Numerical Schemes*

We start by introducing the family of approximating subspaces V_d^p, $d > 0$, $p = 1, 2, \ldots$, which will appear all along this paper. They are of "finite element" type: V_d^p are finite dimensional subspaces of $H^p \cap H_0^1$ that satisfy the following *approximation property*:

If $u \in H^q \cap H_0^1$, $0 \leqslant q \leqslant p+1$, there exists $u^* \in V_d^p$ such that, $\|u - u^*\|_j \leqslant K d^{(q-j)} \|u\|_q$, where $0 \leqslant j \leqslant p$, $j \leqslant q$ and K is a constant independent of d. (7)

Remark 3. The family V_d^p of approximating subspaces has been largely employed in order to approximate the solution of elliptic or parabolic equations. In general, the parameter d is related in some fashion to the maximum diameter of the "elements" partitioning S (see, e.g., Strang–Fix [12]). Subspaces of H_0^1 have been introduced because we are concerned with the solution of Eqs. (2) and (5) in a bounded domain with Dirichlet boundary conditions. A similar class of approximating subspaces can be constructed if Neumann boundary conditions are imposed on these equations.

Now, consider Eq. (2). Associate with the Fokker–Planck operator L^* we can define for all $(u, v) \in H_0^1 \times H_0^1$ the bilinear form,

$$a(u, v) = \frac{1}{2} \int_S \sum_{i,j=1}^{n} a_{i,j} \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_j} dx + \frac{1}{2} \int_S \sum_{i,j=1}^{n} \left(\frac{\partial v}{\partial x_i} a_{i,j} \right) u \frac{\partial v}{\partial x_j} dx - \int_S \sum_{i=1}^{n} g_i u \frac{\partial v}{\partial x_i} dx. \tag{8}$$

Bilinear forms like $a(\cdot, \cdot)$ always appear in connection to unbounded differential operators. They have some interesting properties. In our case, from hypotheses (H1), (H2), (H3), it follows that, for all $u, v \in H_0^1$ $a(\cdot, \cdot)$ satisfies,

$$a(u, \sigma) + (L^* u, v) = 0 \qquad (\text{if } L^* u \in L^2). \tag{9}$$

$a(\cdot, \cdot)$ is bounded,

$$|a(u, v)| \leqslant K \|u\| \, \|v\|. \tag{10}$$

$a(\cdot, \cdot)$ is coercive: for some $\lambda \in R$, $\sigma > 0$,

$$a(u, u) + \lambda |u|^2 \geqslant \sigma \|u\|^2. \tag{11}$$

A Galerkin approximation of the solution of the stochastic evolution equation (2) can be determined by solving the following finite dimensional stochastic equation in V_d^p,

$$dU_t + A U_t \, dt = H U_t \, dy_t, \tag{12}$$

where A, H are linear operators in V_d^p defined by,

$$(Au, v) = a(u, v), \tag{13}$$

$$(Hu, v) = (hu, v), \tag{14}$$

for all $u, v \in V_d^p$.

Any discretization in time domain of Eq. (12) can be viewed as an approximate method for solving the stochastic evolution equation (2). However, one must take into account that the eigenvalues of the operator A are unbounded. Most of the classical numerical methods for solving ordinary differential equations require, for reasons of stability, these eigenvalues to be bounded. Such methods will provide conditionally stable schemes. In our case, we must restrict ourselves to the methods which do not require the boundedness of the eigenvalues of A, that is, to methods that provide unconditionally stable schemes. One class of such methods are the implicit Runge–Kutta methods introduced by Butcher [2]. On the other hand, it is a very-well-known fact that, schemes incorporating terms with powers in the noise increment can produce a faster rate of convergence in time domain. (Clark [4], McShane, [8]). Hence, in order to take full advantage of the discretization in time domain and guarantee unconditional stability of the schemes, we shall use here a modified first order Runge–Kutta method given by the following numerical scheme:

$$\begin{aligned} &U_{k+1} - U_k + \Delta A U_{k+1} + \tfrac{1}{2}\Delta Q U_k \\ &\quad = \Delta_y H U_k + \tfrac{1}{2}\Delta_y^2 Q U_k, \qquad k = 0, 1, \dots, N-1, \end{aligned} \tag{15}$$

where $\{0 = t_0 < t_1 < \cdots < t_N = T\}$ is a uniform partition of the interval $[0, T]$ with increment $\Delta = T/N$, $\Delta_y = y(t_{k+1}) - y(t_k)$ is the noise increment and Q is a linear operator in V_d^p defined by,

$$(Qu, v) = (h^2 u, v), \tag{16}$$

for all $u, v \in V_d^p$.

We make the following assumption:

(H7) The coercivity condition (11) holds strongly, that is, for some $\sigma > 0$ and for all $u \in H_0^\perp$,

$$a(u, u) \geqslant \sigma \|u\|^2.$$

Remark 4. This hypothesis is usual and it can be made without loss of generality since it is equivalent to the addition of a potential term $-\lambda$ in the Fokker–Planck operator. Such a hypothesis would not be necessary if, instead of (2), we were to consider the equation governing the process $\exp(-\lambda t)\, u_t$.

From hypothesis (H7) it follows that $a(\cdot, \cdot)$ is a nom in H_0^1. In other words, the operator A has only positive eigenvalues and thus, Eq. (15) represents a well defined unconditionally stable scheme for the stochastic differential equation (12).

Now, let us consider the evolution equation (5). The bilinear form associated with its differential operator can be expressed by,

$$\begin{aligned} &a(\exp(hy(t))\, u, \exp(-hy(t))\, v) + \tfrac{1}{2}(h^2 u, v) \\ &\quad = a(u, v) - y(t)\, a_1(u, v) - \tfrac{1}{2} y^2(t)\, a_2(u, v) + \tfrac{1}{2}(h^2 u, v), \end{aligned} \tag{17}$$

where,

$$a_1(u), v) = a(u, hv) - a(hu, v), \tag{18}$$

$$a_2(u, v) = 2a(hu, hv) - a(u, h^2 v) - a(h^2 u, v). \tag{19}$$

In order to have this form defined for all $u, v \in H_0^1$ we assume,

(H8) $\partial h/\partial x_i, \partial h/\partial x_i \partial x_j \in L^\infty$.

As one can see, the principal part of the bilinear form (17) (the part containing derivatives in both of u and v) is identical to that of the form (8). The time varying parameter $y(t)$ only effects the secondary part of (17). The component in $y(t)$ has at most only one derivative in u or v and the component in $y^2(t)$ has no derivatives. From hypothesis (H2), (H5), and (H8) it follows that $a_1(\cdot, \cdot)$ and $a_2(\cdot, \cdot)$ are also bounded:

$$|a_1(u, v)| \leqslant K \|u\| \, |v| \quad \text{or} \quad |a_1(u, v)| \leqslant K \, |u| \, \|v\|, \tag{20}$$

$$|a_2(u, v)| \leqslant K \, |u| \, |v|, \tag{21}$$

for all $u, v \in H_0^1$. A Galerkin approximation for the evolution equation (5) can be given by the following ordinary differential equation in the subspace V_d^p:

$$\dot{V}(t) + AV(t) + \tfrac{1}{2}QV(t) = y(t)\, FV(t) + \tfrac{1}{2} y^2(t)\, GV(t), \tag{22}$$

where F, G are linear operators in V_d^p defined by,

$$(Fu, v) = a_1(u, v), \tag{23}$$

$$(Gu, v) = a_2(u, v), \tag{24}$$

for all $u, v \in V_d^p$.

Equation (22) suggests to us the use of the following first order, implicit Runge–Kutta scheme in order to approximate the solution of the evolution equation (5),

$$\begin{aligned} &V_{k+1} - V_k + \Delta A V_{k+1} + \tfrac{1}{2}\Delta Q V_k \\ &\quad = y(t_k)\, \Delta F V_k + \tfrac{1}{2} y^2(t_k)\, \Delta G V_k, \qquad k = 0, 1, ..., N-1, \end{aligned} \tag{25}$$

with the partition of the interval $[0, T]$ defined as before.

Remark 5. A general first order unconditionally stable Runge–Kutta scheme would have in the left side of (15) the term $\Delta A(\rho U_{k+1} + (1-\rho)\, U_k)$ with $0 < \rho \leqslant 1$. Here, for simplicity, we have chosen $\rho = 1$. It is a very-well-known fact that best rates of convergence in time domain can be reached with $\rho = 1/2$. In our case, as the functions we have are highly oscillatory, the rates of convergence will be the same for all ρ.

3. *Convergence to the "Unnormalized" Density*

Denote by R the Ritz projection of H_0^1 onto V_d^p with respect to the norm $a(\cdot, \cdot)$, that is, if $u \in H_0^1$, then $Ru \in V_d^p$ and

$$a(u - Ru, v) = a(R^- u, v) = 0 \tag{26}$$

for all $v \in V_d^p$, where $R^- = I - R$, $I =$ identity. We assume the Ritz projection R satisfying the following hypothesis:

(H9) For all $u \in H^{q+1} \cap H_0^1$, $q = 0, 1, \ldots$, we have,

$$|R^- u| \leqslant Kd \, \|R^- u\|,$$

where K is a constant independent of d.

Remark 6. This is another usual hypothesis employed in association with finite element methods. It can be shown that the above inequality holds depending on the regularity of the bilinear form $a(\cdot, \cdot)$ or, in the other words, on the smoothness of the boundary of the domain S (see Nitsche [9]).

From (10), (26), and hypothesis (H7) we have,

$$\begin{aligned} \sigma \, \|R^- u\|^2 &\leqslant a(R^- u, u - Ru) = a(R^- u, u - v) \\ &\leqslant K \, \|R^- u\| \; \|u - v\|, \end{aligned}$$

for all $v \in V_d^p$. Thus; using hypothesis (H9) and the approximation property (7) we deduce for all $u \in H^{q+1} \cap H_0^1$, $q = 0, 1, \ldots, p$, the following inequality:

$$|R^- u| \leqslant Kd \, \|R^- u\| \leqslant Kd^{q+1} \, \|u\|_{q+1}. \tag{27}$$

We also assume that the solution of Eq. (2) with boundary conditions (6) satisfies:

(H10) π, u_t and $L^* u_t \in H_0^1$ for $t \in (0, T]$.

The purpose of this section is the estimation of the error of approximating the solution of Eqs. (2) and (6) by means of the numerical scheme (15). That is, the error,

$$u_k - U_k = e_k + R^- u_k, \qquad k = 0, 1, \ldots, N, \tag{28}$$

where, hereafter, the index k will denote the value of time varying functions at instant t_k and $e_k = Ru_k - U_k$.

Let us concentrate on e_k: the Ritz projection of the approximation error.

From (15), after eliminating U_k and U_{k+1} we obtain,

$$e_{k+1} - e_k + \Delta A e_{k+1} = (\Delta_y H + \tfrac{1}{2}(\Delta_y^2 - \Delta)\, Q)(e_k - Ru_k) + R(u_{k+1} - u_k) + \Delta A R u_{k+1}. \tag{29}$$

We have seen in Section 1 that, under the probability measure P the observation process y_t becomes a Brownian motion. So, let $\mathscr{F}_t$, $t \in [0, T]$, be a nonanticipating family of σ-algebras relative to y_t. Taking the conditional expectation $E(\cdot \mid \mathscr{F}_k)$ of both sides in (29) and using $E(\Delta_y \mid \mathscr{F}_k) = 0$, $E(\Delta_y^2 \mid \mathscr{F}_k) = \Delta$, we have,

$$\hat{e} - e_k + \Delta A \hat{e} = R(\hat{u}_{k+1} - u_k) + \Delta A R \hat{u}_{k+1}, \tag{30}$$

where the symbol $\hat{}$ denotes $E(\cdot \mid \mathscr{F}_k)$ and $\hat{e} = \hat{e}_{k+1}$. Now, defining $\tilde{e} = e_{k+1} - \hat{e}$ and $\tilde{u}_t = u_t - \hat{u}_t$ for $t \in (t_k, t_{k+1}]$, it comes from (29) and (30) that,

$$\tilde{e} + \Delta A \tilde{e} = (\Delta_y H + \tfrac{1}{2}(\Delta_y^2 - \Delta)\, Q)(e_k - Ru_k) + R\tilde{u}_{k+1} + \Delta A R \tilde{u}_{k+1}. \tag{31}$$

The Ritz projection, e_{k+1}, has been separated into two parcels, $\hat{e}$ and $\tilde{e}$, where $\tilde{e}$ represents the error at t_{k+1} when the new data y_{k+1} is taken into account. The critical point in the estimation of the approximation error is the evaluation of $\tilde{e}$. But before we proceed in this direction, Eq. (31) must be written in a more convenient form. Using (8), (13), (14), (16), (26) and manipulating (31) we obtain the following identity:

$$(\tilde{e}, v) + \Delta a(\tilde{e}, v) = (\varphi e_k, v) + (\gamma, v) - (R^- \tilde{u}_{k+1}, v) + (\varphi R^- u_k, v), \tag{32}$$

for all $v \in V_d^p$ where, for simplicity, we denote,

$$\varphi = \Delta_y h + \tfrac{1}{2}(\Delta_y^2 - \Delta)\, h^2, \tag{33}$$

$$\gamma = (I - \Delta L^*)\, \tilde{u}_{k+1} - \varphi u_k. \tag{34}$$

The auxiliary variables φ and γ deserve a special treatment. First, we introduce the following lemma concerning Ito calculus:

LEMMA 1. *For u_t and y_t defined as before we have*

$$(y_t - y_s)^2 = (t - s) + 2 \int_s^t (y_\tau - y_s)\, dy_\tau, \tag{35}$$

$$\int_s^t \int_s^\tau u_\xi \, d\xi \, dy_\tau = \int_s^t (y_t - y_\tau)\, u_\tau \, d\tau. \tag{36}$$

Proof. Equation (35) is a basic identity concerning Brownian motions. It follows from Ito's rule of transformations applied to the process $(y_t - y_s)^2$. Equation (36) follows from Ito's rule applied to the process $W_t = y_t \cdot \int_s^t u_\tau \, d\tau$: W_t can be expressed by,

$$dW_t = y_t u_t \, dt + \left(\int_s^t u_t \, d\tau \right) dy_t.$$

Then,

$$y_t \cdot \int_s^t u_\tau \, d\tau = \int_s^t y_\tau u_\tau \, d\tau + \int_s^t \int_s^\tau u_s \, ds \, dy_\tau,$$

and we conclude the proof. ∎

Now, using (35), (36), and the integral form of Eq. (2) in order to evaluate increments of the kind $(u_s - u_k)$ we can write,

$$\begin{aligned}
\gamma &= (I - \Delta L^*) \tilde{u}_{k+1} - \int_k^{k+1} \left\{ hu_k \, dy_s + \int_k^s h^2 u_k \, dy_\tau \right\} dy_s \\
&= \int_k^{k+1} \left\{ \int_{k+1}^s (L^*)^2 u_\tau \, d\tau + L^*(\hat{u}_{k+1} - \hat{u}_s) + \int_{k+1}^s L^* h u_\tau \, dy_\tau \right\} ds \\
&\quad + \int_k^{k+1} \left\{ h(u_s - u_k) - \int_k^s h^2 u_k \, dy_\tau \right\} dy_s \\
&= \int_k^{k+1} \left\{ \int_{k+1}^s (L^*)^2 u_\tau \, d\tau + L^*(\hat{u}_{k+1} - \hat{u}_s) + \int_{k+1}^s L^* h u_\tau \, dy_\tau \right. \\
&\quad + \int_s^{k+1} h L^* u_\tau \, dy_\tau - \int_s^{k+1} \int_s^\tau h (L^*)^2 u_\xi \, d\xi \, dy_\tau \\
&\quad \left. - \int_s^{k+1} \int_s^\tau h L^* h u_\xi \, dy_\xi \, dy_\tau \right\} ds \\
&\quad + \int_k^{k+1} \int_k^s \left\{ \int_k^\tau h^2 L^* u_\xi \, d\xi + \int_k^\tau h^3 u_\xi \, dy_\xi \right\} dy_\tau \, dy_s. \qquad (37)
\end{aligned}$$

From (37) and using (9) and (18) we can express (γ, v) for all $v \in V_a^p$ as follows:

$$\begin{aligned}
(\gamma, v) = \int_k^{k+1} &\left\{ \int_s^{k+1} a(L^* u_\tau, v) \, d\tau - a(\hat{u}_{k+1} - \hat{u}_s, v) \right. \\
&- \int_s^{k+1} a_1(u_\tau, v) \, dy_\tau + \int_s^{k+1} \int_s^\tau a(L^* u_\xi, hv) \, d\xi \, dy_\tau
\end{aligned}$$

$$+ \int_s^{k+1} \int_s^{\tau} a(hu_\xi, hv)\, dy_\xi\, dy_\tau \Big\}\, ds$$

$$- \int_k^{k+1} \int_k^{s} \left\{ \int_k^{\tau} a(u_\xi, h^2 v)\, d\xi - \int_k^{\tau} (u_\xi, h^3 v)\, dy_\xi \right\} dy_\tau\, dy_s. \quad (38)$$

Let us return to the evaluation of $\tilde{e}$. In order to estimate $E\,|\tilde{e}|^2$, the ideal is to take the expectation of both sides in (32) with $v = \tilde{e}$ as a test element. We start by estimating terms in the right side of (32):

$$\begin{aligned} E(\varphi e_k, \tilde{e}) &\leqslant K\,\Delta E\,|e_k|^2 + \tfrac{1}{8}E\,|\tilde{e}|^2, \\ E(\gamma, \tilde{e}) &\leqslant K\,\Delta^3(\sup E\,\|L^* u_t\|^2 + \sup E\,\|u_t\|^2) \\ &\quad + \Delta\sigma\, E\,\|\tilde{e}\|^2 + \tfrac{1}{8}E\,|\tilde{e}|^2, \qquad (39) \\ E(R^- \tilde{u}_{k+1}, \tilde{e}) &\leqslant K\Delta(\Delta \sup E\,|R^- L^* u_t|^2 + \sup E\,|R^- h u_t|^2 + \tfrac{1}{8}E\,|\tilde{e}|^2, \\ E(\varphi R^- u_k, \tilde{e}) &\leqslant K\Delta \sup E\,|R^- u_t|^2 + \tfrac{1}{8}E|\tilde{e}|^2, \end{aligned}$$

where $\sup(\cdot)$ is evaluated over $(0, T)$. The auxiliary equations leading to the results above are the following: (i) Eqs. (33) and (38); (ii) inequalities (10) and (20); (iii) standard propeties of Brownian motions concerning the estimation of increments of the kind $(y_t - y_s)$; (iv) Schwartz's inequality; (v) Cauchy,'s inequality, that is, $2xy \leqslant \varepsilon^{-1}x^2 + \varepsilon y^2$ for $x, y \in R$ and $\varepsilon > 0$, where the coefficient ε has been selected in a convenient way in order to produce terms $\frac{1}{8}E\,/\tilde{e}|^2$ in the right side of all estimates. Now, setting estimates (39) into Eq. (32) (with $E(\cdot)$ applied in both sides and $v = \tilde{e}$), using (H7) and inequality (27) (with $q = 0$ because hypothesis (H10) guarantees only $L^* u_t \in H_0^1$) we get,

$$\begin{aligned} E\,|\tilde{e}|^2 \leqslant K\Delta(E\,|e_k|^2 &+ \Delta(\Delta + d^2) \sup E\,\|L^* u_t\|^2 \\ &+ (\Delta^2 + d^2) \sup E\,\|u_t\|^2). \qquad (40) \end{aligned}$$

Equation (40) gives us an estimate for $\tilde{e}$. We need now an estimate for $\hat{e}$ in order to obtain the estimation of the Ritz projection of the approximation error, that is, $e_{k+1} = \hat{e} + \tilde{e}$.

Consider Eq. (30). Using (9), (13), and (2) in order to evaluate increments of the kind $(\hat{u}_s - u_k)$ we find,

$$\begin{aligned} &(\hat{e} - e_k, v) + \Delta a(\hat{e}, v) \\ &\quad = \Delta a(\hat{u}_{k+1}, v) - \int_k^{k+1} a(\hat{u}_s, v)\, ds - (R^-(\hat{u}_{k+1} - u_k), v) \\ &\quad = \int_k^{k+1} \left\{ \int_s^{k+1} a\left(\frac{d}{dt}\hat{u}_\tau, v\right) d\tau - \left(R^- \frac{d}{dt} u_s, v\right) \right\} ds. \qquad (41) \end{aligned}$$

for all $v \in V_d^p$. Making $v = \hat{e}$ as a test element in (41), taking the expectation $E(\cdot)$ of both sides, using Schwartz and Cauchy's inequalities and considering inequality (10) and hypothesis (H7), we have,

$$\begin{aligned}
\tfrac{1}{2}E\,|\hat{e}|^2 - \tfrac{1}{2}E\,|e_k|^2 + \Delta\sigma E\,\|\hat{e}\|^2 &\leqslant E(\hat{e} - e_k, e_k) \\
&\quad + E\,\Delta a(\hat{e}, e_k) = E\int_k^{k+1}\int_k^s a\left(\frac{d}{dt}\hat{u}_\tau, v\right) d\tau\, ds \\
-E\int_k^{k+1}\left(R^- \frac{d}{dt}u_s, v\right) ds &\leqslant K\Delta\left(\Delta^2 \sup E\left\|\frac{d}{dt}\hat{u}_t\right\|^2\right. \\
&\quad \left. + \sup E\left|R^- \frac{d}{dt}\hat{u}_t\right|^2\right) + \Delta\sigma\, E\,\|\hat{e}\|^2 + \tfrac{1}{2}\Delta E\,|\hat{e}|^2. \qquad (42)
\end{aligned}$$

Recalling that $E\,\|(d/dt)\,\hat{u}_t\|^2 \leqslant E\,\|L^* u_t\|^2$ and using inequality (27) as before we get from (42) the following estimate for $\hat{e}$,

$$E\,|\hat{e}|^2 \leqslant (1/1 - \Delta)\,E\,|e_k|^2 + K\Delta(\Delta^2 + d^2)\sup E\,\|L^* u_t\|^2, \qquad (43)$$

where, for consistency, we have made the hypothesis:

(H11) $\Delta \leqslant \delta < 1$.

We are now able to estimate the approximation error. Recalling that $e_{k+1} = \hat{e} + \tilde{e}$, using estimates (40) and (43) we can write,

$$\begin{aligned}
E\,|e_{k+1}|^2 = E\,|\hat{e}|^2 + E\,|\tilde{e}|^2 &\leqslant (1 + \Delta\beta)\,E\,|e_k|^2 \\
&\quad + K\Delta(\Delta^2 + d^2)(\sup E\,\|L^* u_t\|^2 + \sup E\,\|u_t\|^2) \\
&= (1 + \Delta\beta)\,E\,|e_k|^2 + K\Delta(\Delta^2 + d^2), \qquad k = 0, 1, \ldots, N-1, \qquad (44)
\end{aligned}$$

where, here, $K \equiv K(\sup E\,\|L^* u_t\|^2 + \sup E\,\|u_t\|^2)$, that is, we have incorporated in the upper bound constant K estimates concerning the process u_t. The symbol β represents a upper bound for $K + 1/(1 - \Delta)$ and so we have $1 + \Delta\beta \leqslant \exp(\Delta\beta)$. Substituting $E\,|e_k|^2$, $E\,|e_{k-1}|^2, \ldots, E\,|e_0|^2$ into equation (44) it follows that,

$$E\,|e_{k+1}|^2 \leqslant \exp\left(\sum_0^{N-1} \Delta\beta\right)\left\{E\,|e_0|^2 + \sum_0^{N-1} K\Delta(\Delta^2 + d^2)\right\}, \qquad k = 0, 1, \ldots, N-1. \qquad (45)$$

Hence, considering Eq. (28) and using (27) and (45) we can finally deduce the result expressed in the following theorem:

THEOREM 1. *Assuming hypotheses* (H1)–(H11), *the error of* *approximating the solution of Eq.* (2) *with boundary conditions* (6) *by means of the numerical scheme* (15) *with initial condition* $U_0 = \pi$ *is bounded by*,

$$\sup(E\,|u_k - U_k|^2)^{1/2} \leqslant K(\Delta + d), \tag{46}$$

where K *is a constant independent of* d *and* Δ.

Theorem 1 tell us that the approximation method (15) converges to the (Dirichlet) solution of the stochastic evolution equation (2) in the average sense expressed in the left of (46). According to inequality (27), the rate of convergence in the space domain can be increased up to $q+1$ depending on the regularity of u_t and L^*u_t. Here, by hypothesis (H10), we have assumed $q=0$. However, the rate of convergence in the time increment cannot be improved even if another numerical method is employed. We are led to this conclusion by the fact that linear rate is the maximum possible rate of convergence for all numerical procedures that depend on the values of the noise only at the dividing points of the time domain partition. This fact has been shown by Clark [4] for finite dimensional stochastic differential equations and so, with respect to the rate of convergence in time domain, scheme (15) is optimal.

Remark 7. About the hypotheses leading to estimate (46): Hypotheses (H2)–(H6) guarantee the existence of the filtering solution. Hypothesis (H7) complements (H3) and, according to Remark 4, it is only made in order to simplify the proof of Theorem 1. Hypothesis (H9) guarantees the convergence in the space domain; this hypothesis or any other yielding estimate (27) is necessary when finite element subspaces of the type V_d^p are used. Hypotheses (H8) and (H10) concern the regularity conditions which are necessary for the derivation of (46); if they are weakened, convergence may still occur, but the rate of convergence in time domain will be of a lower order. Finally, hypothesis (H11) is also made in order to simplify the proof. The restriction $\Delta < 1$ must be interpreted as "for sufficient small Δ." The unitary bound for Δ is a consequence of the particular choice we have made for the varying parameter in Cauchys inequality.

Remark 8. One cannot say that the proof of Theorem 1 is difficult but, surely, one can complain about how complicated the proof is. This is so because the proof involves, at the same time, numerical analysis technique and facts concerning Ito calculus. We have tried to separate both of these aspects. Ito calculus is crucial in the evaluation of the term $\tilde{e}$ and, in order to have things a bit explicit, Lemma 1 has been introduced. For instance, thanks to identity (36), one can obtain the term containing $a_1(\cdot, \cdot)$ in (38). As a consequence a linear rate of convergence in time domain could be deduced. With respect to the auxiliary estimates (39) the technique employed can be summarized as follows: (i) each differential in the time variable contributes with a Δ in the estimate; (ii) each differential in the noise, when squared, contributes with a Δ in the estimate; (iii) Cauchy's inequality is used in order to present the estimate in a convenient form.

4. *Convergence to the "Pathwise" Solution*

Here, the objective is the estimation of the error of approximating by means of scheme (25) the solution of Eq. (5) under the same Dirichlet boundary conditions used before, that is, those expressed in (6). The approximation error is now given by,

$$v_k - V_k = e_k + R^- v_k, \qquad k = 0, 1, ..., N, \tag{47}$$

where, as an abuse, we have used the same symbol used before for the Ritz projection of the error. So, in this section $e_k = Rv_k - V_k$ with R defined as in (26).

Eliminating V_k, V_{k+1} in Eq. (25) we can write a recursive formula for e_k, that is,

$$\begin{aligned} &e_{k+1} - e_k + \Delta A e_{k+1} + (\tfrac{1}{2}\Delta Q - y_k\,\Delta F - \tfrac{1}{2}y_k^2\,\Delta G)\,e_k \\ &\quad = R(v_{k+1} - v_k) + \Delta A R v_{k+1} + (\tfrac{1}{2}\Delta Q - y_k\,\Delta F - \tfrac{1}{2}y_k^2\,\Delta G)\,Rv_k. \end{aligned} \tag{48}$$

Using (13), (16), (23), (24) and rearranging terms we obtain from (46) the following equation:

$$\begin{aligned} (e_{k+1} - e_k,\ v) + \Delta a(e_{k+1}, v) = {} & -\tfrac{1}{2}\Delta(h^2 e_k, v) \\ & + y_k \Delta a_1(e_k, v) + \tfrac{1}{2}y_k^2\,\Delta a_2(e_k, v) \\ & + (v_{k+1} - v_k, v) + \Delta a(v_{k+1}, v) + \tfrac{1}{2}\Delta(h^2 v_k, v) \\ & - y_k\,\Delta a_1(v_k, v) - \tfrac{1}{2}y_k^2\,\Delta a_2(v_k, v) \\ & - (R^-(v_{k-1} - v_k), v) - \tfrac{1}{2}\Delta(h^2 R^- v_k, v) \\ & + y_k\,\Delta a_1(R^- v_k, v) + \tfrac{1}{2}y_k^2\,\Delta a_2(R^- v_k, v), \end{aligned} \tag{49}$$

for all $v \in V_d^p$. Now, define the auxiliary bilinear form,

$$\begin{aligned} \psi(u, v) = {} & (u, v) - \tfrac{1}{2}\Delta(h^2 u, v) + y_k\,\Delta a_1(u, v) \\ & + \tfrac{1}{2}y_k^2\,\Delta a_2(u, v), \end{aligned} \tag{50}$$

for all $u, v \in H_0^1$. Setting $v = e_{k+1}$ as a test element in (49); using (H7), (20), (21), and (50); considering Schwartz's inequality and Cauchy's inequality conveniently parametrized, it follows that

$$\tfrac{1}{2}|e_{k+1}|^2 - \tfrac{1}{2}|e_k|^2 + \Delta\sigma \, \|e_{k+1}\|^2 \leqslant \tfrac{1}{2}K\Delta \, |e_k|^2 + \tfrac{1}{3}\Delta\sigma \, \|e_{k+1}\|^2$$
$$+ \tfrac{1}{6}\Delta \, |e_{k+1}|^2 + |(v_{k+1}, e_{k+1}) + \Delta a(v_{k+1}, e_{k+1})$$
$$- \psi(v_k, e_{k+1})| + |(R^- v_{k+1}, e_{k+1}) - \psi(R^- v_k, e_{k+1})|. \tag{51}$$

where, here, the upper bound constant K also incorporates bounds of the function $y(\cdot)$. Before we proceed, it is necessary to evaluate the remaining terms in inequality (51). First, recall that the bilinear form defined in (17) suggests the following weak form for Eq. (5):

$$(\dot{v}(t), v) + a(v(t), v) - y(t)\, a_1(v(t), v)$$
$$- \tfrac{1}{2} y^2(t)\, a_2(v(t), v) + \tfrac{1}{2}(h^2 v(t), v) = 0. \tag{52}$$

for all $v \in H_0^1$. If $v(\cdot)$ solves (5) with boundary conditions (6) it follows from (H5), (H8), and (H10) that: (i) $v(\cdot)$ also solves (52); (ii) $v(t)$ and $\dot{v}(t) \in H_0^1$, $t \in (0, T]$.

Let us return to the evaluation of the terms in (51). Taking into account (52) we can write,

$$(v_{k+1}, v) + \Delta a(v_{k+1}, v) - \psi(v_k, v) = \int_k^{k+1} \{(\dot{v}(s) - \dot{v}_k, v)$$
$$+ a(v_{k+1} - v_k, v)\}\, ds = \int_k^{k+1} \left\{ \int_s^{k+1} a(\dot{v}(\tau), v)\, d\tau \right.$$
$$+ y(s) \int_k^s a_1(\dot{v}(\tau), v)\, d\tau + \tfrac{1}{2} y^2(s) \int_k^s a_2(\dot{v}(\tau), v)\, d\tau$$
$$+ (y(s) - y_k)\, a_1(v_k, v) + \tfrac{1}{2}(y^2(s) - y_k^2)\, a_2(v_k, v)$$
$$\left. - \tfrac{1}{2} \int_k^s (h^2 \dot{v}(\tau), v)\, d\tau \right\} ds, \tag{53}$$

for all $v \in V_d^n$. Starting with (53), considering inequalities (10), (20), and (21), using Schwartz and Cauchy's inequalities conveniently we obtain,

$$|(v_{k+1}, e_{k+1}) + \Delta a(v_{k+1}, e_{k+1}) - \psi(v_k, e_{k+1})| \leqslant \tfrac{1}{3}\Delta\sigma \, \|e_{k+1}\|^2$$
$$+ \tfrac{1}{6}\Delta \, |e_{k+1}|^2 + K\Delta(\Delta^2 \sup \|\dot{v}(t)\|^2$$
$$+ |\Delta^y|^2 \sup \|v(t)\|^2), \tag{54}$$

where $|\Delta^y|$ denotes the modulus of continuity of the function $y(\cdot)$, that is,

$$|\Delta^y| = \sup\{ |y(t) - y(s)| : |t - s| \leqslant \Delta \}.$$

Using the same technique as above and recalling inequality (27) (with $q=0$) we can also obtain,

$$|(R^- v_{k+1}, e_{k+1}) - \psi(R^- v_k, e_{k+1})| \leqslant \tfrac{1}{3}\Delta\sigma \, \|e_{k+1}\|^2 + \tfrac{1}{6}\Delta \, |e_{k+1}|^2 + K\,\Delta d^2(\sup \|\dot{v}(t)\|^2 + \sup \|v(t)\|^2). \tag{55}$$

Now, substituting inequalities (54) and (55) in (51), rearranging terms and considering hypothesis (H11), we have

$$|e_{k+1}|^2 \leqslant (1+\Delta\beta)\,|e_k|^2 + K\Delta(\Delta^2 + d^2 + |\Delta^y|^2), \qquad k=0, 1,..., N-1, \tag{56}$$

where, here, the constant K also incorporates estimates concerning the function $v(\cdot)$ and the symbol β represents an upper bound for $(K+1)/(1-\Delta)$.

Eliminating $|e_k|^2$, $|e_{k-1}|^2$,..., $|e_0|^2$, from Eq. (56) it follows that,

$$|e_{k+1}|^2 \leqslant \exp\left(\sum_0^{N-1} \Delta\beta\right)\left\{|e_0|^2 + \sum_0^{N-1} K\Delta(\Delta^2 + d^2 + |\Delta^y|^2\right\}, \qquad k=0, 1,..., N-1. \tag{57}$$

Thus, considering Eq. (47) and using (27) and (57) we can finally obtain an estimate for the approximation error. We register this result in the following theorem:

THEOREM 2. *Assuming hypothesis* (H1)–(H11) *the error of approximating the solution of Eq.* (5) *with boundary conditions* (6) *by means of the numerical scheme* (25) *with initial condition* $V_0 = R\pi$ *is bounded by,*

$$\sup |v_k - V_k| \leqslant K(\Delta + d + |\Delta^y|), \tag{58}$$

where K *is a constant independent of* d, Δ *and* $|\Delta^y|$.

Inequality (58) shows that the approximation method (25) converges to the (Dirichlet) solution of the evolution equation (5) in a pathwise sense. As before, the rate of convergence in the space domain discretization can be improved up to $q+1$ depending on the regularity of the solution of (5). On the other hand, the rate of convergence in time domain depends, in a linear fashion, on the modulus of continuity of the sample paths of the observation process. This order of convergence is optimal and cannot be improved whatever the numerical scheme employed. This is so, because the function $y(\cdot)$ is continuous but, in general, highly oscillatory. In particular, if we select sample paths that satisfy a uniform Hölder condition, $y(\cdot) \in \{y(\cdot): |\Delta^y| \leqslant K\Delta^\alpha, 0<\alpha<1\}$, then the convergence will be uniform and the rate of convergence in the time increment will have the value of the Hölder coefficient. Clark [3] has shown that the pathwise solution of the filtering problem for Markov chains admits a discrete approximation that converges with a rate depending, in a linear fashion, on the modulus of continuity of the sample paths. Here, we have extended Clark's result to filtering problems for diffusion processes.

Remark 9. As the set of hypotheses leading to Theorems 1 and 2 are the same they deserve the same comments as those in Remark 7. In particular, as before, if the regularity conditions (H8) (H10) are weakened the numerical method may converge but with a lower rate of convergence in the time domain.

Conclusion

Numerical approximations to the nonlinear filtering solution is still an unexplored area of research for the most part. Here, convergence properties of two discrete time numerical schemes are studied. Under the hypotheses we have made these schemes converge with an optimal rate to the equations governing the filter for a diffusion absorbed by the boundaries of a given domain. If the hypotheses are weakened or if the quadratic term in the noise is removed from theses schemes, convergence may occur but with a slow rate. Using the same analytical procedure, extensions of these results can be obtained for nonhomogeneous signal (and observation) processes and also for diffusions reflected by the boundaries. The choice between the two schemes presented is certainly a matter of convenience but an analysis of the computational aspects of their solution should be done.

Acknowledgment

This work is an extract from the author's Ph.D. thesis at Imperial College, University of London, January 1980.

References

1. J. F. Bennaton, "Galerkin Procedures for Stochastic Partial Differential Equations," Ph.D. thesis, Imperial College, London, 1980.
2. J. C. Butcher, Implicit Runge–Kutta processes, *Math. Comp.* **18** (1964), 50–64.
3. J. M. C. Clark, The design of robust approximations to the stochastic differential equations of non-linear filtering, *in* "Communication Systems and Random Process Theory" (J. K. Skwirzynski, Ed.), NATO, Adv. Study Inst. Series Sijithoff & Noordhoff, Alphen aan de Rijn, 1978.

4. J. M. C. Clark, and R. J. Cameron, The maximum rate of convergence of discrete approximation for stochastic differential equations, *in* "Proc. Intern. Symposium on Stoch. Diff. Equations, Vilnius, 1978."
5. E. B. Dynkin, "Markov Processes," Springer-Verlag Berlin, 1965.
6. H. J. Kushner, Dynamical equations for optimal non-linear filtering, *J. Differential Equations* **3** (1967), 179–190.
7. R. S. Liptser and A. N. Shiryaev, "Statistics of Random Processes," Springer-Verlag, New York, 1977.
8. E. J. McShane, "Stochastic Calculus and Stochastic Models," Academic Press, New York, 1974.
9. J. Nitsche, Ein Kriterium für die quasi-optimalital des Ritzschen verfahrend, *Numer. Math.* **11** (1968), 346–348.
10. E. Pardoux, Filtrage de diffusions avec conditions frontières in Proc. Journées de Statistique dans les Processus Stochastiques, *in* "Lecture Notes in Math., Vol. 636," Springer, New York, 1978.
11. E. Pardoux, Stochastic partial differential equations and filtering of diffusion processes, *Stochastics* **3**, No. 2 (1979), 127–167.
12. G. Strang and G. Fix, "An Analysis of the Finite Element Method," Prentice–Hall, New Jersey, 1973.