Nonlinear Filtering Revisited: a Spectral Approach, II*

Sergey V. Lototsky
Center for Applied Mathematical Sciences
University of Southern California
Los Angeles, CA 90089-1113
sergey@cams-00.usc.edu

Remijigus Mikulevicius
Institute of Mathematics and Informatics
Akademijos 4, Vilnius 2600, Lithuania

Boris L. Rozovskii
Center for Applied Mathematical Sciences
University of Southern California
Los Angeles, CA 90089-1113; rozovski@cams-00.usc.edu

Published in: *Proceedings of the 35th IEEE Conference on Decision and Control*, Kobe, Japan, Dec. 11–13, 1996, v. 4, pp. 4060–4064, Omnipress, Madison, WI.

Abstract

A recursive in time Wiener chaos representation of the optimal nonlinear filter is derived for a continuous time diffusion model with uncorrelated noises. The existing representations are either not recursive or require a prior computation of the unnormalized filtering density, which is time consuming. An algorithm is developed for computing a recursive approximation of the filter, and the error of the approximation is obtained. When the parameters of the model are known in advance, the on-line speed of the algorithm can be increased by performing part of the computations off line.

Key words: multiple Wiener integrals, nonlinear filtering, recursive filter.

*This work was partially supported by ONR Grant #N00014-95-1-0229 and ARO Grant DAAH 04-95-1-0164. The paper is based on [10].

1 Introduction

In a typical filtering model, a non-anticipative functional $f_t(x)$ of the unobserved signal process $X = X(t)$ is estimated from the observations $Y = Y(s),\ s \le t$. The best mean square estimate is known to be the conditional expectation $\mathbf{E}[f_t(X)|Y(s),\ s \le t]$, called the optimal filter. When the observation noise is additive, the Kallianpur-Striebel formula [6, 8] provides the representation of the optimal filter as follows:

$$\mathbf{E}[f_t(X)|Y(s),\ s \le t] = \frac{\phi_t[f]}{\phi_t[1]},$$

where $\phi_t[\cdot]$ is a functional called *the unnormalized optimal filter*. In the particular case $f_t(X) = f(X(t))$, there are two approaches to computing $\phi_t[f]$.

In the first approach [9, 12, 13], the functional $\phi_t[f]$ is expanded in a series of multiple integrals with respect to the observation process. This approach can be used to obtain representations of general functionals, but these representations are not recursive in time. In fact, there is no closed form differential equation satisfied by $\phi_t[f]$.

In the second approach [6, 8, 14] it is proved that under certain regularity assumptions the functional $\phi_t[f]$ can be written as

$$\phi_t[f] = \int f(x)u(t,x)dx \tag{1.1}$$

for some function $u(t,x)$, called *the unnormalized filtering density*. Even though the computation of $u(t,x)$ can be organized recursively in time, and there are many numerical algorithms to do this [1, 2, 3, 5, 10, etc.], these algorithms are time consuming because they involve evaluation of $u(t,x)$ at many spatial points. Moreover, computation of $\phi_t[f]$ using this approach requires subsequent evaluation of the integral (1.1).

The objective of the current work is to develop a recursive in time algorithm for computing $\phi_t[f]$ without computing $u(t,x)$. The analysis is based on the multiple integral representation of the unnormalized filtering density [10, 11, 13] with subsequent Fourier series expansion in the spatial domain. In the proposed algorithm, the computations involving the parameters of the model can be done separately from those involving the observation process. If the parameters of the model are known in advance, this separation can substantially increase the on-line speed of the algorithm.

2 Representation of the Unnormalized Optimal Filter

Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a complete probability space, on which standard Wiener processes $V = V(t)$ and $W = W(t)$ are given of dimensions d_1 and r. Random processes $X = X(t)$ and $Y = Y(t)$ of dimention d and r are defined by the equations

$$\begin{aligned} X(t) &= X_0 + \int_0^t b(X(s))ds + \int_0^t \sigma(X(s))dV(s), \\ Y(t) &= \int_0^t h(X(s))ds + W(t), \qquad 0 \le t \le T. \end{aligned} \tag{2.1}$$

In applications, $X(t)$ represents the unobserved state process subject to estimation from the observations $Y(s),\ s \le t$. The σ - algebra generated by $Y(s),\ s \le t$, will be denoted by $\mathcal{F}_t^y$.

The following regularity assumptions are made:

(**A1**) The Wiener processes V and W are independent of X_0 and of each other;

(**A2**) The functions $b(x),\ \sigma(x)$, and $h(x)$ are infinitely differentiable and bounded with all the derivatives;

(**A3**) The random variable X_0 has a density $p(x),\ x \in I\!R^d$, so that the function $p = p(x)$ is infinitely differentiable and decays at infinity with all the derivatives faster than any power of $|x|$.

The coefficients $b,\ \sigma,\ h$, the density function p, and the lenght of the time interval T will be referred to as **the parameters** of the filtering model (2.1).

Let $f = f(x)$ be a measurable function such that

$$|f(x)| \le L(1 + |x|^{k_0}) \tag{2.2}$$

for some $k_0 \ge 0$ and $L > 0$. Assumptions (**A2**) and (**A3**) imply that $\mathbf{E}|f(x(t))|^2 < \infty$ for all $t \ge 0$ [8]. Suppose that $T > 0$ is fixed. It is known [6, 8, 14] that the best mean square estimate of $f(X(t))$ given $Y(s),\ s \le t \le T$, is $\hat{f}_t = \mathbf{E}[f(X(t))|\mathcal{F}_t^y]$, and under the above regularity assumptions this estimate can be written as follows:

$$\hat{f}_t = \frac{\phi_t[f]}{\phi_t[1]}.$$

The functional $\phi_t[f]$ is called **the unnormalized optimal filter** and admits a representation

$$\phi_t[f] = \int_{I\!R^d} f(x)u(t,x)dx.$$

The random field $u = u(t,x)$ is called **the unnrmalized filtering density**.

Denote by $P_t\,\varphi(x)$ the solution of the equation

$$\begin{aligned} \frac{\partial v(t,x)}{\partial t} = &\ \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2((\sigma\sigma^{)}_{ij}v(t,x))}{\partial x_i \partial x_j} - \\ &\ \sum_{i=1}^{d} \frac{\partial(b_i(x)v(t,x))}{\partial x_i},\ t > 0; \\ v(0,x) = &\ \varphi(x), \end{aligned}$$

and consider $0 = t_0 < t_1 < \ldots < t_M = T$, a uniform (for simplicity) partition of $[0,T]$ with steps $\Delta = t_i - t_{i-1}$; this partition will be fixed hereafter. The following theorem gives a recursive representation of the unnormalized filtering density at the points of the partition.

Theorem 2.1 *Under assumptions* (**A1**) – (**A3**),

$$\begin{array}{l} u(t_0,x)=p(x), \\ u(t_i,x)=P_t u(t_{i-1},\cdot)(x)+\sum\limits_{k\geq 1}\sum\limits_{l_1,\ldots,l_k=1}^{r}\int_0^{\Delta}\int_0^{s_k}\ldots\int_0^{s_2} \\ P_{t-s_k}h_{l_k}\ldots h_{l_1}P_{s_1}u(t_{i-1},\cdot)(x)dY_{l_1}^{(i)}(s_1)\ldots dY_{l_k}^{(i)}(s_k), \\ \mathrm{P}-\text{a.s.} \end{array} \tag{2.3}$$

for $i=1,\ldots,M$, *where* $Y_l^{(i)}(t)=Y_l(t+t_{i-1})-Y_l(t_{i-1}),\ 0\leq t\leq\Delta$.

Proof. Under different assumptions that are typically less restrictive than (**A1**–**A3**) the various forms of the representation (2.3) were established by Kunita [7], Mikulevicius and Rozovskii [12], and Ocone [13]. □

To simplify the further presentation, the following notations are introduced. For an $\mathcal{F}_{t_{i-1}}^{y}$ - measurable function $g=g(x,\omega)$ and $0\leq t\leq\Delta$,

$$\begin{array}{l} F_0^{(i)}(t,g)(x):=P_t g(x), \\ F_k^{(i)}(t,g)(x):=\sum\limits_{l=1}^{r}\int_0^t P_{t-s}h_l F_{k-1}^{(i)}(s,g)dY_l^{(i)}(s),\ k\geq 1. \end{array} \tag{2.4}$$

With these notations, (2.3) becomes

$$u(t_i,x)=\sum_{k\geq 0}F_k^{(i)}(\Delta,u(t_{i-1},\cdot))(x),\ i=1,\ldots,M. \tag{2.5}$$

It can be shown [14] that the regularity assumptions (**A1**) and (**A2**) imply

$$u(t_i,\cdot)\in L_2(I\!R^d),\quad i=1,\ldots,M\quad(\mathrm{P}-\text{a.s.})$$

Consequently, if $\{e_n\}_{n\geq 0}$ is an orthonormal basis in $L_2(I\!R^d)$, then

$$u(t_i,\cdot)=\sum_{n\geq 0}\psi_n(i)e_n\quad(\mathrm{P}-\text{a.s.}), \tag{2.6}$$

where $\psi_n(i)=(u(t_i,\cdot),e_n)_0$ and $(\cdot,\cdot)_0$ is the inner product in $L_2(I\!R^d)$. Substitution of (2.6) into (2.5) yields

$$\begin{array}{c} \psi_n(i)=\sum\limits_{k\geq 0}\Big(\sum\limits_{l\geq 0}(F_k^{(i)}(\Delta_i,e_l),e_n)_0\psi_l(i-1)\Big), \\ i=1,\ldots,M. \end{array} \tag{2.7}$$

Together with the initial relation

$$\psi_n(0)=(p,e_n)_0$$

equation (2.7) can be used for recursive computation of the coefficients $\psi_n(i)$.

Clearly, for computational purposes both infinite sums in (2.7) must be truncated. To study the effects of the truncation it is necessary to specify the basis $\{e_n\}$.

There are two main reasons why the desired filtering algorithm cannot be described using a general basis $\{e_n\}$. One reason is that, to estimate the error due to the truncation of the infinite sum in (2.7), it is necessary to know the rate of decay of the coefficients $\psi_n(i)$ as $n \to \infty$, and this rate, in general, depends on the basis $\{e_n\}$. The other reason is that, once the UFD is represented in the form (2.6) – exact or truncated, it is natural to write the unnormalized optimal filter $\phi_t[f]$ as

$$\phi_t[f] = \sum_{n \geq 0} \psi_n(i) f_n,$$

where

$$f_n = \int_{I\!R^d} f(x) e_n(x) dx.$$

On the other hand, if the only restriction on the function f is the growth condition (2.2), then the last integral need not be defined for a general basis functions e_n.

Below, the approximations of the UFD and of the optimal filter will be constructed using the Hermite basis in $L_2(I\!R^d)$. The basis is defined as follows. Let $\Gamma = \{\gamma : \gamma = (\gamma_1, \ldots, \gamma_2), \gamma_i \in \mathbf{Z}_+\}$ be the set of d dimensional multi-indices; $|\gamma| = \sum_{i=1}^d \gamma_i$. To define the ordering of Γ, set $\gamma < \tau$ if $|\gamma| < |\tau|$ or if $|\gamma| = |\tau|,\ \gamma_i = \tau_i,\ i < j < d,\ \gamma_j < \tau_j$. For $\gamma \in \Gamma$ define

$$e_\gamma(x_1, \ldots, x_d) = \prod_{i=1}^d e_{\gamma_i}(x_i), \tag{2.8}$$

where, for $n \in \mathbf{Z}_+,\ t \in I\!R$,

$$e_n(t) = \frac{1}{\sqrt{2^n \pi^{1/2} n!}} e^{t^2/2} \frac{d^n}{dt^n} e^{-t^2}.$$

It is known [14] that under the regularity assumptions (**A2**) and (**A3**) the UFD $u(t, x)$ is infinitely differentiable as a function of x and decays at infinity with all the derivatives faster than any power of $|x|$. Consequently, using the results from [4], it can be shown that for every $\nu > 0$ there is a constant $C(\nu)$ depending only on ν and the parameters of the model so that for all $i = 0, \ldots, M$

$$\mathbf{E}|\psi_\gamma(i)| \leq \frac{C(\nu)}{|\gamma|^\nu}.$$

Also, since each function $e_\gamma = e_\gamma(x)$ decays at infinity as $\exp(-C(\gamma)|x|^2)$ for some positive number $C(\gamma)$, the coefficients

$$f_\gamma = \int_{I\!R^d} f(x) e_\gamma(x) dx$$

can be defined for every function f satisfying the condition (2.2).

3 Recursive Approximation of the Unnormalized Optimal Filter

It was already mentioned that the infinite sums in (2.7) must be approximated by truncating the number of terms if the formula is to be used for practical computations. Multiple integrals in (2.4) must also be approximated. The effects of these approximations are studied below.

Given a positive integer κ, define the set

$$\Gamma_\kappa = \{\gamma \in \Gamma : |\gamma| \le \kappa\}$$

and then the random variables $\psi_{\gamma,\kappa}(i)$, $\gamma \in \Gamma_\kappa$, $i = 0, \dots, M$, by

$$\begin{aligned}
&\psi_{\gamma,\kappa}(0) = (p, e_\gamma)_0, \\
&\psi_{\gamma,\kappa}(i) = \sum_{\tau \in \Gamma_\kappa} ((P_\Delta e_\tau, e_g m)_0 + \\
&\sum_{l=1}^{r} (P_\Delta h_l e_\tau, e_\gamma)_0 [Y_l(t_i) - Y_l(t_{i-1})] + \\
&\frac{1}{2} \sum_{l=1}^{r} (P_\Delta h_l^2 e_\tau, e_\gamma)_0 [(Y_l(t_i) - Y_l(t_{i-1}))^2 - \Delta] + \\
&\sum_{l=1}^{r} \sum_{m=1}^{l-1} (P_\Delta h_l h_m e_\tau, e_\gamma)_0 (Y_l(t_i) - Y_l(t_{i-1})) \times \\
&(Y_m(t_i) - Y_m(t_{i-1}))) \, \psi_{\tau,\kappa}(i-1), \ i = 1, \dots, M.
\end{aligned} \tag{3.1}$$

Then the corresponding approximations to $u(t_i, x)$ and $\phi_{t_i}[f]$ are

$$\begin{aligned}
u_\kappa(t_i, x) &= \sum_{\gamma \in \Gamma_\kappa} \psi_{\gamma,\kappa}(i) e_\gamma(x), \\
\phi_{t_i,\kappa}[f] &= \sum_{\gamma \in \Gamma_\kappa} \psi_{\gamma,\kappa}(i) f_\gamma.
\end{aligned} \tag{3.2}$$

The errors of these approximations are given in the following theorem. The main steps in deriving (3.1) and (3.2) and the idea of the proof are given in the Appendix. Below, $\|\cdot\|_0$ denotes the norm in the space $L_2(I\!R^d)$.

Theorem 3.1 *If assumptions* (**A1**) – (**A3**) *and* (2.2) *hold and the basis* $\{e_n\}$ *is chosen according to* (2.8) *then*

$$\max_{1 \le i \le M} \mathbf{E} \|u_\kappa(t_i, \cdot) - u(t_i, \cdot)\|_0 \le C\Delta + \frac{C(\nu)}{\kappa^\nu \Delta}, \tag{3.3}$$

$$\max_{1 \le i \le M} \mathbf{E} |\phi_{t_i,\kappa}[f] - \phi_{t_i}[f]| \le C\Delta + \frac{C(\nu)}{\kappa^\nu \Delta}. \tag{3.4}$$

Remark. The constants in (3.3) and (3.4) are determined by the parameters of the model, i.e. the bounds on the functions b, σ, h, and p and their derivatives and by the length T of the time interval. The constants in (3.4) also depend on L and k_0 from (2.2).

The error bounds in (3.3) and (3.4) involve two asymptotic parameters: Δ (the size of the partition of the time interval) and κ (which determines the number of the spatial basis functions). With the appropriate choice of these parameters, the errors can be made arbitrarily small.

The following is the description of the algorithm based on the formulas (3.1) and (3.2).

1. Off line (before the observations are available):

— *for* $l, m = 1, \dots, r$ *and* $\gamma, \tau \in \Gamma_\kappa$, *compute* $(P_\Delta e_\tau, e_g m)_0$, $(P_\Delta h_l e_\tau, e_g m)_0$, $(P_\Delta h_l^2 e_\tau, e_g m)_0$, $(P_\Delta h_l h_m e_\tau, e_g m)_0$, $f_\gamma = \int_{I\!R^d} f(x) e_\gamma(x) dx$, *and* $\psi_{\gamma,\kappa}(0)$.

— *set* $u_\kappa(t_0, x) = \sum_{\gamma \in \Gamma_\kappa} \psi_{\gamma,\kappa}(0) e_\gamma(x)$ *and* $\phi_{t_0,\kappa}[f] = \sum_{\gamma \in \Gamma_\kappa} \psi_{\gamma,\kappa}(0) f_{\lfloor}\gamma$.

2. On line, k-th step (as the measurements become available): compute

$$\psi_{\gamma,\kappa}(i) = \sum_{\tau \in \Gamma_\kappa} Q_{\gamma\tau}(Y^{(i)}) \psi_(i-1), \quad \gamma \in \Gamma_\kappa, \tag{3.5}$$

where

$$\begin{aligned} &Q_{\gamma\tau}(Y^{(i)}) = (P_\Delta e_\tau, e_\gamma)_0 + \\ &\sum_{l=1}^{r} (P_\Delta h_l e_\tau, e_\gamma)_0 [Y_l(t_i) - Y_l(t_{i-1})] + \\ &\frac{1}{2} \sum_{l=1}^{r} (P_\Delta h_l^2 e_\tau, e_\gamma)_0 [(Y_l(t_i) - Y_l(t_{i-1}))^2 - \Delta] + \\ &\sum_{l=1}^{r} \sum_{m=1}^{l-1} (P_\Delta h_l h_m e_\tau, e_\gamma)_0 (Y_l(t_i) - Y_l(t_{i-1})) \times \\ &(Y_m(t_i) - Y_m(t_{i-1})), \end{aligned}$$

then compute

$$u_\kappa(t_i, x) = \sum_{\gamma \in \Gamma_\kappa} \psi_{\gamma,\kappa}(i) e_\gamma(x), \tag{3.6}$$

$$\phi_{t_i,\kappa}[f] = \sum_{\gamma \in \Gamma_\kappa} \psi_{\gamma,\kappa}(i) f_\gamma,$$

and

$$\hat{f}_{t_i,\kappa} = \frac{\phi_{t_i,\kappa}[f]}{\phi_{t_i,\kappa}[1]}. \tag{3.7}$$

We would like to remark the following features of the algorithm:

(1) The time consuming operations of solving the partial differential equations and computing integrals are performed off line;

(2) The overall amount of the off-line computations does not depend on the number of the on-line time steps;

(3) Formula (3.7) can be used to compute an approximation to $\hat{f}_{t_i}$ (e.g. conditional moments) without the time consuming computations of $u_\kappa(t_i, x)$ and the related integrals;

(4) Only the Fourier coefficients $\psi_{\gamma,\kappa}$ must be computed at every time step while the approximate filter $\hat{f}_{t_i,\kappa}$ and/or UFD $u_\kappa(t_i, x)$ can be computed as needed, e.g. at the final time moment.

(5) The on-line part of the algorithm can be easily parallelized.

The number of on-line operations (in flops) to update the coefficients $\psi(i)$ is about

$$K_{op} = (r^2 + 3r + 1)|\Gamma_\kappa|^2,$$

where $|\Gamma_\kappa| = (\kappa + d)!/(\kappa! d!)$ is the number of elements in the set Γ_κ; for $\kappa = 10$ and $d = 6$, $|\Gamma_\kappa = 8008$. We believe that, due to the high spatial regularity of the unnormalized filtering density, relatively small values of κ will suffice to achieve the desired accuracy so that the above number of operations is not too large and can be performed on line by a sufficiently powerful computer even for high dimensional state processes.

4 Conclusion

An optimal nonlinear filtering algorithm is suggested for a continuous time diffusion model with independent noise. The algorithm is based on the multiple Wiener integral expansion of the unnormalized filtering density. Recursive approximations of the unnormalized filtering density and of the optimal filter are computed independently of each other using a finite collection of (approximate) Fourier coefficients with respect to the Hermite basis. The coefficients are updated using the increments of the observation process. The on-line speed of the algorithm is increased by shifting off line the time consuming operations of solving partial differential equations and computing integrals. The approximation error can be controlled by increasing the number of the coefficients and decreasing the size of the time step.

References

[1] A. Budhiraja and G. Kallianpur. Approximations to the Solution of the Zakai Equations Using Multiple Wiener and Stratonovich Integral Expansions. *Stochastics and Stochastics Reports*, 56:271–315, 1996.

[2] R. J. Elliott and R. Glowinski. Approximations to Solutions of the Zakai Filtering Equation. *Stochastic Analysis and Applications*, 7(2):145–168, 1989.

[3] P. Florchinger and F. LeGland. Time Discretization of the Zakai Equation for Diffusion Processes Observed in Correlated Noise. *Stochastics and Stochastics Reports*, 35(4):233–256, 1991.

[4] D. Gottlieb and S. A. Orszag. *Numerical Analysis of Spectral Methods: Theory and Applications*. CBMS-NSF Regional Conference, Series in Applied Mathematics, Vol.26, 1977.

[5] Kazufumi Ito. Approximation of the Zakai Equation for Nonlinear Filtering. *SIAM Journal on Control and Optimization*, 34(2):620–634, 1996.

[6] G. Kallianpur. *Stochastic Filtering Theory*. Springer, New York, 1980.

[7] H. Kunita. Cauchy Problem for Stochastic Partial Differential Equations Arising in Nonlinear Filtering Theory. *System & Control Letters*, 1(1):37–41, 1981.

[8] R. Sh. Liptser and A. N. Shiryayev. *Statistics of Random Processes*. Springer, New York, 1992.

[9] J. T.-H. Lo and S.-K. Ng. Optimal Orthogonal Expansion For Estimation I: Signal in White Gaussian Noise. In R.S. Bucy and J.M.F Moura, editors, *Nonlinear Stochastic Problems*, pages 291–309, D. Reidel Publ. Company, 1983.

[10] S. Lototsky, R. Mikulevicius, and B. L. Rozovskii. Nonlinear Filtering Revisited: A Spectral Approach. *SIAM Journal on Control and Optimization*, 1997. To appear.

[11] R. Mikulevicius and B. L. Rozovskii. Separation of Observations and Parameters in Nonlinear Filtering. In *Proceedings of the 32nd IEEE Conference on Decision and Control, Part 2, San Antonio, 1993*, pages 1564–1569. IEEE Control Systems Society, 1993.

[12] R. Mikulevicius and B. L. Rozovskii. Fourier-Hermite Expansion for Nonlinear Filtering. In A. A. Novikov, editor, *Festschrift in honor of A. N. Shiryayev*, 1995.

[13] D. Ocone. Multiple Integral Expansions for Nonlinear Filtering. *Stochastics*, 10:1–30, 1983.

[14] B. L. Rozovskii. *Stochastic Evolution Systems*. Kluwer Academic Publishers, 1990.

Appendix

Below, the main steps of the proof of Theorem 3.1 are presented. These steps also illustrate how formulas (3.1) and (3.2) were obtained.

Step 1. Define

$$u^1(t_0,x) := p(x),$$
$$u^1(t_i,x) := \sum_{k=0}^{2} F_k^{(i)}(\Delta, u(t_{i-1},\cdot))(x).$$

It is proved in [10, Theorem 2.4], that

$$\max_{0\le i\le M} \mathbf{E}\|u(t_i,\cdot) - u^1(t_i,\cdot)\|_0 \le C\Delta. \tag{A.1}$$

Step 2. Define

$$\begin{aligned}
&\bar{F}_0^{(i)}(\Delta, g)(x) := P_\Delta g(x),\\
&\bar{F}_1^{(i)}(\Delta, g)(x) :=\\
&\qquad \sum_{l=1}^{r} P_\Delta(h_l g)(x)[Y_l(t_i) - Y_l(t_{i-1})],\\
&\bar{F}_2^{(i)}(\Delta, g)(x) :=\\
&\quad \frac{1}{2}\sum_{l=1}^{r} P_\Delta(h_l^2 g)(x)[(Y_l(t_i) - Y_l(t_{i-1}))^2 - \Delta] +\\
&\sum_{l=1}^{r}\sum_{m=1}^{l-1} P_\Delta(h_l h_m g)(x)(Y_l(t_i) - Y_l(t_{i-1}))\times\\
&\qquad\qquad (Y_m(t_i) - Y_m(t_{i-1}))
\end{aligned}$$

and then by induction

$$\begin{aligned}
&\bar{u}^1(t_0, x) := p(x),\\
&\bar{u}^1(t_i, x) = \sum_{k=0}^{2} \bar{F}_k^{(i)}(\Delta, \bar{u}^1(t_{i-1}, \cdot))(x).
\end{aligned}$$

It can be shown that

$$\max_{0\le i\le M} \mathbf{E}\|\bar{u}^1(t_i, \cdot) - u^1(t_i, \cdot)\|_0 \le C\Delta. \tag{A.2}$$

Step 3. The same arguments as in the proof of Theorem 2.6 in [10] show that

$$\mathbf{E}\|\bar{u}^1(t_i, \cdot) - u_\kappa(t_i, \cdot)\|_0 \le \frac{C(\nu)}{\kappa^\nu \Delta}. \tag{A.3}$$

Combining (A.1), (A.2), (A.3), and the triangle inequality results in (3.3).

2. Inequality (3.4) follows from (3.3) and the Cauchy inequality. To deal with the technical difficulty that, in general, $f \notin L_2(\mathbb{R}^d)$, the following spaces are introduced [14, Sec. 4.3]: for $w \in \mathbb{R}$, $L_2(\mathbb{R}^d, w) = \{\varphi : \int_{\mathbb{R}} \varphi^2(x)(1 + |x|^2)^w dx < \infty\}$. The growth condition (2.2) then implies that $f \in L_2(-w)$ for sufficiently large positive w, while the regularity assumptions and the choice of the basis $\{e_n\}$ imply that both the UFD and its approximation belong to $L_2(\mathbb{R}^d, w)$ for all positive w. The idea is then to establish the analog of inequality (3.3) with the norm of the space $L_2(\mathbb{R}, w)$ for sufficiently large positive w and then to use the generalized Cauchy – Schwartz inequality.