

Fast Nonlinear Filter for Continuous-Discrete Time Multiple Models*

Sergey V. Lototsky

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

Chuanxia Rao

Center for Applied Mathematical Sciences
University of Southern California
Los Angeles, CA 90089-1113 USA
crao@cams-00.usc.edu

Boris L. Rozovskii

Center for Applied Mathematical Sciences
University of Southern California
Los Angeles, CA 90089-1113 USA
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. 4071–4076, Omnipress, Madison, WI.

Abstract

A fast algorithm is proposed for computing on line the optimal nonlinear filter in the continuous-discrete time, multiple model setting. Using the finite element approximation on a spatial grid with N points and performing part of the computations off line, the on-line complexity of the algorithm is shown to be $O(N)$ for all dimensions of the state process. The error of the approximation is also studied.

Key words: finite element approximation, multiple model filtering, optimal filter.

*This work was partially supported by ONR Grant #N00014-95-1-0229 and ARO Grant DAAH 04-95-1-0164.

1 Introduction

In the continuous-discrete time filtering model, an unobservable continuous time state process is estimated from the noisy measurements made at discrete time moments. This model seems of special interest from the point of view of applications, because many real life processes evolve in continuous time while the digital devices used to process the measurements require discrete time data. The case of continuous time observations was studied in [9, 8], see also [10].

The desired solution of the filtering problem is an algorithm that provides the best mean square estimate of the given functional of the state process in the form suitable for on-line implementation. In the linear case, such a solution is given by the Kalman filter [6, 7].

It is worth mentioning that the exact solution of the continuous-discrete time filtering problem is known for a wide class of nonlinear models [4, 6] and consists in determining a special random field called the unnormalized filtering density (UFD), but the resulting algorithm requires on-line solution of a partial differential equation (Fokker-Planck equation) and computation of integrals. If the dimension of the state process is greater than three, the real time implementation of this algorithm is practically impossible.

On-line solution of the Fokker-Planck equation can be avoided if the UFD admits a finite dimensional sufficient statistics; in particular, this is the case when the model is linear Gaussian. The first nonlinear example of this sort was discovered by Benesh [2], and a more general class of UFD admitting finite dimensional sufficient statistics was studied by Daum [3]. Unfortunately, for a given nonlinear model, it is usually not clear whether the sufficient statistics exists. The practical algorithms based on this approach use approximations similar to the extended Kalman filter [11], so as a rule the error of such approximations is unknown. Moreover, these algorithms still require on-line solution of an ordinary differential equation and evaluation of integrals.

The objective of the current work is to develop a recursive numerical algorithm for computing the optimal filter in which the on-line part is as simple as possible; in particular, no differential equations are to be solved on-line. The starting point in the derivation is the equation satisfied by the unnormalized filtering density in the general nonlinear model, and the approach is based on the technique known as the parameterization of the UFD [6] using the finite element approximation [5].

In the proposed algorithm (Section 3) both time consuming operations of solving the Fokker-Planck equation and computing the integrals are performed off line, which makes the algorithm suitable for on-line implementation. The on-line complexity of the algorithm is $O(N)$, where N is the number of points in the spatial domain (or the number of functions used in the finite element approximation). Since the result is not the exact optimal filter, the error of the approximation is estimated.

2 The Filtering Problem

Consider the problem of estimating an \mathbb{R}^d -valued state process $X = X(t)$ that evolves in continuous time according to the following diffusion equation

$$dX(t) = b(X(t))dt + \sigma(X(t))dW(t), \quad X(0) = x_0, \quad (2.1)$$

given \mathbb{R}^r -valued measurements $z = z(k)$ made at discrete time moments $t_k = k\Delta$, $k \geq 1$:

$$z(k) = h(X(t_k)) + v(k). \quad (2.2)$$

In the above, $W = (W(t))_{t \geq 0}$ is an \mathbb{R}^{d_1} -valued standard Brownian motion independent of the initial condition x_0 ; functions $b = b(x)$, $\sigma = \sigma(x)$, and $h = h(x)$, $x \in \mathbb{R}^d$, take values in \mathbb{R}^d , $\mathbb{R}^{d \times d_1}$, and \mathbb{R}^r respectively; the sequence $\{v(k)\}_{k \geq 1}$ is independent of the state process and consists of i.i.d. Gaussian random vectors with zero mean and covariance $\mathbf{E}v(k)v(k)^T = \varepsilon I$, $I \in \mathbb{R}^{r \times r}$ is the identity matrix, $\varepsilon > 0$ is a scalar parameter. The underlying probability space $(\Omega, \mathcal{F}, \mathbf{P})$ is assumed to be fixed.

The following regularity assumptions are made about the model (2.1), (2.2):

1. *the functions b, σ , and h are infinitely differentiable and bounded with all the derivatives;*
2. *the random vector x_0 has a density $p_0 = p_0(x)$, $x \in \mathbb{R}^d$, so that the function p is infinitely differentiable and decays at infinity with all the derivatives faster than any power of $|x|$.*

Let $f = f(x)$, $x \in \mathbb{R}^d$, be a measurable scalar function such that $\mathbf{E}|f(X(t))|^2 < \infty$ for all $t \geq 0$. Then the filtering problem for (2.1), (2.2) can be stated as follows: find the best mean square estimate of $f(X(t_k))$ given the measurements $z(m)$, $m = 1, \dots, k$. This estimate is called **the optimal filter** and will be denoted by $\hat{f}(k)$. It is a standard fact that

$$\hat{f}(k) = \mathbf{E}(f(X(t_k)) | z(1), \dots, z(k)).$$

For computational purposes, the optimal filter can be characterized as follows.

Denote by T_t the solution operator for the Fokker-Planck equation corresponding to the state process; in other words, $u(t, x) = T_t \varphi(x)$ is the solution of the equation

$$\begin{aligned} \frac{\partial u(t, x)}{\partial t} &= \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2}{\partial x_i \partial x_j} \left(\sum_{l=1}^{d_1} \sigma_{il}(x) \sigma_{jl}(x) u(t, x) \right) - \\ &\quad \sum_{i=1}^d \frac{\partial}{\partial x_i} (b_i(x) u(t, x)), \quad t > 0, \\ u(0, x) &= \varphi(x). \end{aligned} \quad (2.3)$$

Next, define the sequence $p_k(x)$, $x \in \mathbb{R}^d$, $k \geq 0$, by

$$\begin{aligned} p_0(x) &= p(x), \\ p_k(x) &= H_k(x) T_\Delta p_{k-1}(x), \end{aligned}$$

where

$$H_k(x) = \exp \left\{ \frac{1}{\varepsilon^2} \sum_{l=1}^r \left(h_l(x) z_l(k) - \frac{1}{2} h_l^2(x) \right) \right\}.$$

The random field $p_k = p_k(x)$ is called **the unnormalized filtering density**.

Then the optimal filter $\hat{f}(k)$ can be written as follows [6]:

$$\hat{f}(k) = \frac{\int_{\mathbb{R}^d} p_k(x) f(x) dx}{\int_{\mathbb{R}^d} p_k(x) dx}. \quad (2.4)$$

The numerator in (2.4) will be denoted by $\phi_k[f]$. With this notation, (2.4) becomes

$$\hat{f}(k) = \frac{\phi_k[f]}{\phi_k[1]}.$$

3 The Algorithm

Assume that the function f satisfies the following growth condition:

$$|f(x)| \leq C(1 + |x|^\alpha), \quad x \in \mathbb{R}^d, \quad (3.1)$$

for some $\alpha, C > 0$. Then $\mathbf{E}|f(X(t))|^2 < \infty$ for all $t \geq 0$ [7].

To parameterize the function p_k , a finite element approximation is used. Let $\{e_1, \dots, e_n\}$ be a basis in the given finite dimensional set of interpolating functions and $\{x_1, \dots, x_N\}$ the collection of points in \mathbb{R}^d so that $e_i(x_j) = \delta_{ij}$. Assume that the support of each function e_i is a compact set concentrated around the point x_i .

For a continuous function $v = v(x)$ its interpolation $\bar{v} = \bar{v}(x)$ is defined by

$$\bar{v}(x) = \sum_{i=1}^N v(x_i) e_i(x).$$

Consider the following approximation of the unnormalized filtering density:

$$\tilde{p}_k(x) = \overline{H_k(x) T_\Delta \tilde{p}_{k-1}(x)}.$$

For $l, m = 1, \dots, M$ define the numbers

$$Q_{lm} = (T_\Delta e_m)(x_l), \quad \psi_l(k) = \tilde{p}_k(x_l),$$

and

$$f_l = \int_{\mathbb{R}^d} f(x) e_l(x) dx.$$

The coefficients $\psi_l(k)$ can be used to compute the approximations $\tilde{p}_k(x)$ and $\tilde{\phi}_k[f]$ of the UFD and the unnormalized optimal filter according to the following **recursive algorithm**.

1. *Off line (before the measurements are available), compute Q_{lm} , f_l , and $\psi_l(0) = p_0(x_l)$, $l, m = 1, \dots, N$.*
2. *On line, step k (as the measurements become available): compute*

$$\psi_m(k) = H_k(x_m) \sum_{l=1}^N Q_{ml} \psi_l(k-1), \quad m = 1, \dots, N, \quad (3.2)$$

then compute

$$\tilde{p}_k(x) = \sum_{l=1}^N \psi_l(k) e_l(x), \quad (3.3)$$

$$\tilde{\phi}_k[f] = \sum_{l=1}^N \psi_l(k) f_l,$$

and

$$\tilde{f}_k = \frac{\tilde{\phi}_k[f]}{\tilde{\phi}_k[1]}. \quad (3.4)$$

According to formula (3.2), the matrix Q can be viewed as **the filter kernel**.

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

- (1) The time consuming operations of solving the partial differential equation (2.3) 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 (3.4) can be used to compute an approximation to $\hat{f}(k)$ (e.g. conditional moments) without computing $\tilde{p}_k(x)$ as an intermediate step;
- (4) Only the coefficients ψ_l must be computed at every time step while the approximate filter \tilde{f}_k and UFD $\tilde{p}_k(x)$ 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 on each step of the algorithm is $O(N)$, where N is the number of the spatial points. The reason is that the number of operations to update each of the coefficients ψ_m according to (3.2) is effectively independent of N . Indeed, since $T_\Delta e_l$ is a solution of a parabolic equation with the initial condition e_l and each function e_l is compactly supported around the point x_l , the values of $(T_\Delta e_l)(x_m)$ will be close to zero when the distance $|x_l - x_m|$ is sufficiently large. In other words, the the matrix $Q = (Q_{ml})$ is *numerically* banded meaning that in actual computations Q is replaced by a banded matrix with the number of nonzero entries in each column not exceeding some fixed number L ; the number L is independent of N . The specific value of L depends on the coefficients of the state equation and the basis functions, and can be controlled by the suitable choice of Δ . The number of on-line operations, measured in flops, on each step of the algorithm is then equal to

$$K_{op} = (3r + 2L + 3)N.$$

The approximation error can be controlled by increasing the dimension of the approximation. The specific error bounds depend on the basis functions e_l . For simplicity, consider the one dimensional model ($d = r = 1$) and the linear interpolation on a uniform grid so that the functions $e_l(x)$ are given by

$$e_l(x) = \left(1 - \frac{|x - x_l|}{x_{l+1} - x_l}\right) I(x_{l-1} \leq x \leq x_{l+1}).$$

Denote by \mathbf{C}_b^2 the space of twice continuously differentiable functions bounded with both derivatives; $\|f\|_{\mathbf{C}_b^2} := \sup_{x \in \mathbb{R}} |f(x)| + \sup_{x \in \mathbb{R}} |f'(x)| + \sup_{x \in \mathbb{R}} |f''(x)|$. Also denote by $L_2(w)$, $w \in \mathbb{R}$, the space of measurable functions $f = f(x)$ for which $\|f\|_{L_2(w)}^2 := \int_{\mathbb{R}} |f(x)|^2 (1+x^2)^w dx < \infty$; similarly, $\mathbf{H}^1(w) = \{f : \|f\|_{\mathbf{H}^1(w)}^2 := \|f\|_{L_2(w)}^2 + \|f'\|_{L_2(w)}^2 < \infty\}$.

Theorem. *Assume that the spatial grid $x_1 < \dots < x_N$ is uniform ($x_{i+1} - x_i = \delta x$), $[-R, R] \subset [x_1, x_N]$ for some $R > 0$, and the linear interpolation is used. Then for every $w > 0$ there is a constant $C(w)$ depending only on w and possibly the coefficients of the state equation, and a constant $C(h)$ depending only on the ratio $h(x)/\varepsilon$ so that*

$$\sup_{x \in \mathbb{R}} \mathbf{E}|p_k(x) - \tilde{p}_k(x)| \leq \exp(k(C(w)\Delta + C(h)) \times \left(\frac{(\delta x)^2}{8} \|p_0\|_{\mathbf{C}_b^2} + \frac{C(w)}{(1+R^2)^{w/2}} \|p_0\|_{\mathbf{H}^1(w)} \right).$$

If in addition the function f satisfies the growth condition (3.1), then

$$\begin{aligned} \mathbf{E}|\phi_k[f] - \tilde{\phi}_k[f]| &\leq \\ \exp(k(C(w)\Delta + C(h)) \times &\left(CR^{\alpha+3/2} \frac{(\delta x)^2}{8} \|p_0\|_{\mathbf{C}_b^2} + \right. \\ &\left. \frac{C(w)\|f\|_{L_2(-\alpha-2)}}{(1+R^2)^{w/2}} \|p_0\|_{L_2(w+\alpha+2)} \right). \end{aligned}$$

Proof. It is well known that in the case of linear interpolation

$$\sup_{x \in [-R, R]} |f(x) - \bar{f}(x)| \leq \frac{(\delta x)^2}{8} \max_{x \in [-R, R]} |f''(x)| \leq \frac{(\delta x)^2}{8} \|f\|_{\mathbf{C}_b^2}.$$

Another well known inequality is

$$\sup_{x \notin [-R, R]} |f(x)| \leq \frac{C(w)}{(1+R^2)^{1/2}} \|f\|_{\mathbf{H}^1(w)}.$$

It can be shown using the definition of p_k and the regularity assumptions that

$$\mathbf{E}\|p_k\|_{\mathbf{X}} \leq e^{k(C(w)\Delta + C(h))} \|p_0\|_{\mathbf{X}},$$

where \mathbf{X} is either \mathbf{C}_b^2 or $\mathbf{H}^1(w)$. After that the first statement of the theorem follows from the discrete time version of the Gronwall lemma.

The second statement of the theorem follows from the first and the growth condition (3.1). \square

The theorem implies that for every k the approximation error converges to zero in the limit $\lim_{R \rightarrow \infty} \lim_{\delta x \rightarrow 0}$.

Remark. If $\varepsilon^{-1}\|h\|_{\mathbf{C}_b^2} \leq 1$, then $C(h) = 8\varepsilon^{-2}\|h\|_{\mathbf{C}_b^2}^2$, and if $\varepsilon^{-1}\|h\|_{\mathbf{C}_b^0} \geq 1$, then $C(h) = 5\varepsilon^{-2}\|h\|_{\mathbf{C}_b^2}^2 + 7\varepsilon^{-4}\|h\|_{\mathbf{C}_b^1}^4$.

4 Multiple Model Filtering

The optimal nonlinear filter can also be used even if the parameters of the state equation are not known to the observer. To construct the filter in this situation, the multiple model approach can be used. According to [1], the basic assumption of the approach is that at each moment the state process can be described by one of the given number of models with switching between the models governed by a homogeneous Markov process. For example, an aircraft can be in one of few possible flight modes (constant velocity, constant acceleration, coordinated turn, etc.), and while the current mode may be not known to the observer, the dynamical equations corresponding to each mode are available.

Assume that each of the possible models of the state process is of the form (2.1), i.e. is a nonlinear diffusion with drift $b(x, i)$ and diffusion coefficient $\sigma(x, i)$, with $i \in \{1, \dots, M\}$ corresponding to the particular model. If $\theta = \theta_t$ is a homogeneous Markov process with values in the set $\{1, \dots, M\}$, initial distribution $\pi_0 \in \mathbb{R}^M$, and the jump intensity matrix $\Lambda \in \mathbb{R}^{M \times M}$, then the state process $X = X(t)$ can be modeled by

$$dX(t) = b(X(t), \theta_t)dt + \sigma(X(t), \theta_t)dW(t). \quad (4.1)$$

It is also assumed that the measurements

$$z(k) = h(X(t_k), \theta_{t_k}) + v(k) \quad (4.2)$$

with the sequence $v = v(k)$ as in (2.2), and that the process θ is independent of v , W , and $X(0)$. As before, the functions $b(\cdot, i)$, $\sigma(\cdot, i)$, and $h(\cdot, i)$ are infinitely differentiable and bounded with all the derivatives for every $i = 1, \dots, M$, and the initial density p_0 of $X(0)$ is infinitely differentiable and decays at infinity faster than any power of $|x|$.

The choice of the initial distribution π_0 and of the jump intensity matrix Λ is made during the overall system design. For example, in the case of an aircraft tracking both π_0 and Λ can be estimated by analyzing the real flight data. A general discussion of the question is given in [1].

From the point of view of the general theory, the model (4.1), (4.2) is equivalent to the original model (2.1), (2.2). Indeed, the unobservable component is the pair of processes (X, θ) with the generating operator

$$\mathcal{L} = \begin{pmatrix} \mathcal{L}_1 & & 0 \\ & \ddots & \\ 0 & & \mathcal{L}_M \end{pmatrix} + \Lambda,$$

where

$$\begin{aligned} \mathcal{L}_k g(x) = & \frac{1}{2} \sum_{i,j=1}^d (\sigma(x, k)\sigma^T(x, k))_{ij} \frac{\partial^2 g(x)}{\partial x_i \partial x_j} + \\ & \sum_{i=1}^d b(x, k) \frac{\partial g(x)}{\partial x_i} \end{aligned}$$

and σ^T is the transpose of the matrix σ . Denote by $\Gamma_t \mathbf{F}(x)$ the solution of the equation

$$\frac{\partial \mathbf{v}(t, x)}{\partial t} = \mathcal{L}^* \mathbf{v}(t, x), \quad t > 0, \quad \mathbf{v}(0, x) = \mathbf{F}(x),$$

in which $\mathbf{v}(t, x)$ is a vector¹ in \mathbb{R}^M and \mathcal{L}^* is the formal adjoint of \mathcal{L} .

Define the vectors $\mathbf{p}(k, x)$, $k \geq 0$, $x \in \mathbb{R}^d$, by

$$\begin{aligned}\mathbf{p}(k, x) &= \mathcal{H}(k, x)\Gamma_{\Delta}\mathbf{p}(k-1, \cdot)(x), \quad k \geq 1, \\ \mathbf{p}(0, x) &= p_0(x)\pi_0,\end{aligned}\tag{4.3}$$

where $\mathcal{H}(k, x)$ is a diagonal matrix with entries

$$\exp\left\{\frac{1}{\varepsilon^2}\sum_{l=1}^r\left(h_l(x, i)z_l(k) - \frac{1}{2}h_l^2(x, i)\right)\right\} \quad i = 1, \dots, M.$$

For a function f satisfying the growth condition (3.1) define

$$\phi_k[f, i] = \int_{\mathbb{R}^d} f(x)[\mathbf{p}(k, x)]_i dx.$$

Then

$$\mathbf{E}f(X(t_k))|z(1), \dots, z(k) = \frac{\sum_{i=1}^M \phi_k[f, i]}{\sum_{i=1}^M \phi_k[1, i]}$$

and

$$\mathbf{P}(\theta_{t_k} = i|z(1), \dots, z(k)) = \frac{\phi_k[1, i]}{\sum_{i=1}^M \phi_k[1, i]}.$$

Recursive approximations of $\mathbf{p}(k, x)$ and $\phi_k[f, i]$ can be computed using the same approach as in the previous section. The vector $\mathbf{p}(k, x)$ is approximated by

$$\tilde{\mathbf{p}}(k, x) = \sum_{l=1}^N \sum_{i=1}^M \psi_l^i(k) e_l(x) \mathbf{u}_i,$$

where e_1, \dots, e_N are the interpolating functions and $\{\mathbf{u}_1, \dots, \mathbf{u}_M\}$ is the standard unit basis in \mathbb{R}^M (i.e. $[\mathbf{u}_i]_j = \delta_{ij}$). The coefficients ψ_l^i are updated using the pre-computed filter kernel $Q = (Q_{lm}^{ij})$, $l, m = 1, \dots, N$, $i, j = 1, \dots, M$, and the measurements $z(k)$. The following is the description of the corresponding algorithm.

1. *Off line (before the observations are available): compute the components of the kernel*

$$Q_{lm}^{ij} = [\Gamma_{\Delta}(e_m(\cdot)\mathbf{u}_j)]_i(x_l),$$

the coefficients

$$f_l = \int_{\mathbb{R}^d} f(x)e_l(x) dx$$

and

$$\psi_l^i(0) = p_0(x_l)[\pi_0]_i;$$

set

$$\tilde{\mathbf{p}}(0, x) = \sum_{l=1}^N \sum_{i=1}^M \psi_l^i(0) e_l(x) \mathbf{u}_i;$$

¹All vectors are column vectors; the components of a vector v are denoted by either v_i or $[v]_i$.

2. On line, step k : compute

$$\psi_l^i(k) = \mathcal{H}(k, x_l) \sum_{m=1}^N \sum_{j=1}^M Q_{lm}^{ij} \psi_m^j(k-1),$$

then, if necessary, compute

$$\begin{aligned} \tilde{\mathbf{p}}(k, x) &= \sum_{l=1}^N \sum_{i=1}^M \psi_l^i(k) e_l(x) \mathbf{u}_i, \\ \tilde{\phi}_k[f, i] &= \sum_{l=1}^N \psi_l^i(k) f_l, \\ \tilde{f}_k &= \frac{\sum_{i=1}^M \tilde{\phi}_k[f, i]}{\sum_{i=1}^M \tilde{\phi}_k[1, i]}, \end{aligned} \tag{4.4}$$

and

$$\tilde{\theta}_k(i) = \frac{\tilde{\phi}_k[1, i]}{\sum_{i=1}^M \tilde{\phi}_k[1, i]}$$

(the probability that the current model is i).

Formula (4.4) gives only one possible way of estimating the state. Once the coefficients $\psi(k)$ are known, other estimates can be constructed, for example,

$$\hat{X}_k = \arg \max_x [\tilde{\mathbf{p}}(k, x)]_{\hat{\theta}_k},$$

where

$$\hat{\theta}_k = \arg \max_i \phi_k[1, i].$$

Using the same arguments as in the previous section, it can be shown that the number of on-line operations (in flops) to update the coefficients ψ is

$$K_{op} = (3r + 2L + 3)M^2N.$$

The operator Γ_Δ can be approximated by the Trotter formula as follows. For $t > 0$ denote by $T_t^{(k)}g(x)$, $k = 1, \dots, M$, the solution of the equation

$$\frac{\partial v(t, x)}{\partial x} = \mathcal{L}_k^* v(t, x), \quad v(0, x) = g(x),$$

and let

$$T_t = \begin{pmatrix} T_t^{(1)} & & 0 \\ & \ddots & \\ 0 & & T_t^{(M)} \end{pmatrix}.$$

If $0 = t_0 < t_1 \dots < t_\lambda = \Delta$ is a partition of the interval $[0, \Delta]$, then Γ_Δ is approximated by

$$\exp((t_\lambda - t_{\lambda-1})\Lambda^T) T_{t_\lambda - t_{\lambda-1}} \cdots \exp((t_1 - t_0)\Lambda^T) T_{t_1 - t_0}.$$

References

- [1] Y. Bar-Shalom and X.-R. Li. *Estimation and Tracking: Principles, Techniques, and Software*. Artech House, Boston, 1993.
- [2] V. E. Benesh. Exact Finite-dimensional Filters for Certain Diffusions with Nonlinear Drift. *Stochastics*, 5:65–92, 1981.
- [3] F. E. Daum. New Exact Nonlinear Filters. In J. C. Spall, editor, *Bayesian Analysis of Time Series and Dynamic Models*, pages 199–226, Marcel Dekker, Inc., New York, 1988.
- [4] R. J. Elliott, L. Aggoun, and J. B. Moore. *Hidden Markov Models*. Springer, New York, 1995.
- [5] A. Germani and M. Piccioni. Semi-discretization of Stochastic Partial Differential Equations on \mathbf{R}^d by a Finite Element Technique. *Stochastics*, 23(2):131–148, 1988.
- [6] A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, New York, 1970.
- [7] R. Sh. Liptser and A. N. Shiriyayev. *Statistics of Random Processes*. Springer, New York, 1977.
- [8] S. Lototsky, R. Mikulevicius, and B. L. Rozovskii. Nonlinear Filtering Revisited: A Spectral Approach. *SIAM Journal on Control and Optimization*, 1997. (To appear).
- [9] S. Lototsky and B. L. Rozovskii. Recursive Multiple Wiener Integral Expansion for Nonlinear Filtering of Diffusion Processes. In N. Gretskey, editor, *Festschrift in honor of M. M. Rao*, Marcel Dekker, Inc., New York, 1996. (To appear).
- [10] D. Ocone. Multiple Integral Expansions for Nonlinear Filtering. *Stochastics*, 10:1–30, 1983.
- [11] G. C. Schmidt. Designing Nonlinear Filters Based on Daum’s Theory. *Journal of Guidance, Control, and Dynamics*, 16(2):371–376, 1993.