A Split-Step Solution of the Fokker–Planck Equation for the Conditional Density

(Invited Paper)

H. C. Lambert*, F. E. Daum†, and J. L. Weatherwax*
* MIT Lincoln Laboratory
† Raytheon Company

***Abstract*— We have developed and tested a new algorithm, which we call the "wave filter," that solves the nonlinear filtering problem with discrete-time measurements by solving the Fokker–Planck equation for the conditional probability density function using a split-step technique. The wave filter uses fast convolution to compute the effect of process noise at discrete times. Between measurements, the conditional density is propagated by solving a system of ordinary differential equations. Measurement update is carried out via Bayes' rule. We propose the "adjoint method" to reduce the computational complexity of the algorithm by adaptively varying the mesh size.**

I. Introduction

Nonlinear filtering algorithms can be divided into three broad categories. First, there are those filters that compute the sufficient statistics characterizing the conditional probability density function. These filters effectively reduce the problem of solving a partial differential equation (PDE) to that of solving a series of ordinary differential equations (ODEs). This class of filters includes the extended and unscented Kalman filters. Another major class of filters consists of those which approximate the conditional density using Monte Carlo sampling techniques. The particle filter is the quintessence of this class of filters. Finally, there are those filters that compute the conditional density exactly by explicitly solving the Fokker–Planck equation (see e.g. [1]). Note that in general, the computational complexity of both Monte Carlo and exact techniques increases exponentially with the dimension of the state vector, while the computational complexity of filters solving for the sufficient statistics varies polynomially with the dimension of the state vector (see e.g. [2] and [3]).

In this paper, we present a new algorithm, which we refer to as the "wave filter[1]", that solves for the conditional density exactly and hence falls under the third category of nonlinear filters. Our algorithm relies on using a "split-step" technique for solving the Fokker–Planck equation.

We begin by reviewing the nonlinear filtering problem in section II. The split-step solution of the Fokker–Planck equation along with a formulation of the wave filtering algorithm is presented in section III. We give two numerical examples in section IV demonstrating the performance of the wave filter. Finally, in section V, we discuss the possibility of using the "adjoint method" to reduce the computational complexity of the wave filter.

[1]The term "wave filter" is more than a mere wordplay and is in fact reminiscent of the wave-particle duality in quantum mechanics.

II. Problem Statement

In a nutshell, the filtering problem consists of recursively estimating the state of an uncertain variable from noisy measurements. The dynamic behavior of the state is described by the system dynamic model [4]:

$$d\mathbf{x}_t = \mathbf{f}_t(\mathbf{x}_t) + d\mathbf{w}_t, \tag{1}$$

where the process noise vector $\mathbf{w}_t$ is a zero-mean Gaussian white noise process with spectral density matrix Q_t. The mapping of the state space at time index k to the discrete-time measurement space is given by the measurement model [4]:

$$\mathbf{z}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{v}_k, \tag{2}$$

where the measurement noise vector $\mathbf{v}_k$ is a zero-mean Gaussian white noise process with covariance R_k. The dimension of the measurement vector is smaller than or equal to the dimension of the state vector. The spectral density Q_t and covariance R_k of the process and measurement noise vectors are known a-priori and together with the system and measurement functions, $\mathbf{f}_t$ and $\mathbf{h}_k$, respectively, serve as "inputs" to the filtering algorithm.

What we seek is the conditional probability density function, $p(\mathbf{x}_{t_k}|\mathbf{Z}_k)$, of the state variable at time t_k given the sequence of measurements $\mathbf{Z}_k = \{\mathbf{z}_1, \ldots, \mathbf{z}_k\}$. Between measurements, the evolution of the conditional density is prescribed by the Fokker–Planck equation [5]:

$$\frac{\partial p}{\partial t} = -\mathbf{f}_t^{\mathrm{T}} \frac{\partial p}{\partial \mathbf{x}_t} - \mathrm{tr}\left(\frac{\partial \mathbf{f}_t}{\partial \mathbf{x}_t}\right) p + \frac{1}{2}\mathrm{tr}\left(Q_t \frac{\partial^2 p}{\partial \mathbf{x}_t^2}\right) \tag{3}$$

with initial condition $p(\mathbf{x}_{t_{k-1}}|\mathbf{Z}_{k-1})$. We denote the solution as $p(\mathbf{x}_{t_k}|\mathbf{Z}_{k-1})$. This constitutes the "propagation" step of the solution. At the time of measurement arrival, we update the conditional density using Bayes' rule:

$$p(\mathbf{x}_{t_k}|\mathbf{Z}_k) = \frac{p(\mathbf{z}_k|\mathbf{x}_{t_k})p(\mathbf{x}_{t_k}|\mathbf{Z}_{k-1})}{p(\mathbf{z}_k|\mathbf{Z}_{k-1})} \tag{4}$$

where $p(\mathbf{z}_k|\mathbf{x}_{t_k})$ is the likelihood function of the measurement $\mathbf{z}_k$. In the next section, we consider a "split-step" technique for solving the Fokker–Planck equation (3).

III. Solution of the Fokker–Planck Equation

As a foray into solving the Fokker–Planck equation, we examine two limiting cases. First, we consider the case of random walk dynamics with time-invariant process noise:

$$\mathbf{f}_t = \mathbf{0} \quad \text{and} \quad Q_t = Q.$$

The Fokker–Planck equation for this case reduces to

$$\frac{\partial p}{\partial t} = \frac{1}{2}\mathrm{tr}\left(Q\frac{\partial^2 p}{\partial \mathbf{x}_t^2}\right), \tag{5}$$

which we recognize as the diffusion equation with Q interpreted as the diffusion tensor. Taking the Fourier transform with respect to the state vector $\mathbf{x}_t$, we obtain

$$\frac{dp^\dagger}{dt} = -\frac{1}{2}\boldsymbol{\xi}_t^{\mathrm{T}} Q \boldsymbol{\xi}_t p^\dagger, \tag{6}$$

where $p^\dagger$ denotes the Fourier transform of the conditional density, and $\boldsymbol{\xi}_t$ corresponds to the Fourier "wavenumber" variable. Equation (6) is a simple ODE that can be solved in closed form. Taking the inverse Fourier transform of the result, we arrive at a solution for the diffusion equation that can be regarded as the smeared version of the initial density. We conclude that the diffusion equation (5) can be solved using efficient algorithms for computing the convolution integral such as the fast Fourier transform (FFT).

Next, we consider the case when there is no process noise:

$$Q_t = 0.$$

For this case, the Fokker–Planck equation reduces to a convection equation:

$$\frac{\partial p}{\partial t} + \mathbf{f}_t^{\mathrm{T}}\frac{\partial p}{\partial \mathbf{x}_t} = -\mathrm{tr}\left(\frac{\partial \mathbf{f}_t}{\partial \mathbf{x}_t}\right)p, \tag{7}$$

where $\mathbf{f}_t$ can be interpreted as the drift vector, while the left-hand side of the equation denotes a substantive derivative [6]. The PDE (7) can in turn be cast into an equivalent system of ODEs:

$$\frac{d}{dt}\begin{bmatrix}\mathbf{x}_t \\ p\end{bmatrix} = \begin{bmatrix}\mathbf{f}_t \\ -\mathrm{tr}\left(\frac{\partial \mathbf{f}_t}{\partial \mathbf{x}_t}\right)p\end{bmatrix}. \tag{8}$$

We solve equation (8) by first forming a grid for the solution domain. Subsequently, the state vector valid at a particular grid point and the corresponding conditional density are propagated simultaneously using a standard ODE solver such as the 4th-order Runge–Kutta integrator [7].

The results we obtain for these two limiting cases can be combined to solve the Fokker–Planck equation for the general case. This is achieved by mandating the process noise spectral density matrix to be discrete-time; that is, we consider the case when process noise is "injected" into the system at discrete time intervals coinciding with times of measurement. We note that unlike in physics, process noise is an engineering *design parameter*. In physics, we have no control over the behavior of the diffusion tensor. Thus, we let

$$Q_t = \sum_k Q_k \delta(t - t_k), \tag{9}$$

where $\delta(t - t_k)$ is the Dirac delta function. Hence, immediately after the measurement update step, the Fokker–Planck equation reduces to the diffusion equation (5), while between measurements, it reduces to the convection equation (7). We refer to this method for solving the Fokker–Planck equation as the "split-step" technique. Finally, we can always consider a diagonal process noise covariance matrix via an appropriate coordinate transformation and thereby reducing the computational complexity of solving the diffusion equation (5) even further.

In summary, we solve the nonlinear filtering problem with discrete-time measurements in three steps: (1) smear the conditional density valid at time index $k-1$ by convolving it with a Gaussian function with covariance Q_{k-1}; (2) propagate the system of ODEs (8) from time index $k-1$ to time index k for each grid point of the solution domain; and (3) update the conditional density with the measurement received at time index k using Bayes' rule (4). We refer to this algorithm as the "wave filter."

IV. Numerical Examples

A. Spinning Disk

In this and the next section, we consider two simple two-dimensional examples demonstrating the performance of the wave filter with measurement updates. First, we consider the so called "spinning disk" problem [8]. We imagine a flat disk spinning about an axis. An observer observing the spinning disk from an edge measures a feature of the disk that is a nonlinear function of some "metric" property of the disk. Specifically, we consider the state of the disk to be completely specified by the two-dimensional vector:

$$\mathbf{x}_t = \begin{bmatrix}\theta_t & \dot{\theta}_t\end{bmatrix}^{\mathrm{T}}, \tag{10}$$

where θ_t is the angle of a point on the edge of the disk, while $\dot{\theta}$ is the spin rate of the disk. We consider a linear system dynamic model corresponding to a simple "constant-velocity" model:

$$\mathbf{f}_t(\mathbf{x}_t) = \begin{bmatrix}\dot{\theta} & 0\end{bmatrix}^{\mathrm{T}}. \tag{11}$$

On the other hand, for the measurement model, we consider the nonlinear model:

$$h_k = \cos(\theta_{t_k}). \tag{12}$$

Since cosine is an even function, this measurement model gives rise to an ambiguity with regards to the sign of the spin rate. In other words, the posterior density function is expected to be bi-modal.

Figure 1 summarizes our simulation results. Numerical values for the simulation parameters are listed in Table I. Figure 1 shows plots of the conditional density as a function of the two states, angle and spin rate. The first simulation "frame" shown in Figure 1a corresponds to the initial density prior to the acquisition of the measurement. The density is assumed to be uniform. As shown in Figure 1b, the density shows its bi-modal characteristic immediately after the first measurement update. Here, the estimation of spin rate remains poor. After

TABLE I

SPINNING DISK EXAMPLE SIMULATION PARAMETERS

PARAMETERS	VALUE	DIMENSION
Number of grid points per dimension	64	N/A
True initial angle	65	deg
True initial spin rate	2	$\text{deg} \cdot \text{s}^{-1}$
Angle process noise variance	0	$\text{deg}^2 \cdot \text{s}^{-1}$
Spin rate process noise variance	0.1	$\text{deg}^2 \cdot \text{s}^{-3}$
Measurement error std. dev.	0.25	N/A
Sampling time	0.5	s

a few iterations, the estimation of the spin rate improves (Figure 1c). However, the sign of the spin rate remains ambiguous. The last simulation frame shown in Figure 1d demonstrates the convergence of the algorithm. We note how the spacing of the mesh changes over time. Our numerical integration of the convection equation implements the notion of the substantive derivative by explicitly propagating each mesh point according to the underlying system dynamic model. Here, the system dynamic model is linear. This is reflected in the sheared appearance of the mesh after a few iterations of the algorithm.

B. Tracking a Falling Object on Reentry

In this section, we consider a simple example with a nonlinear dynamic model and a linear measurement model (in a way the mirror image of the problem we considered in section IV-A). We consider the problem of tracking a falling object on reentry from beyond the Earth's atmosphere. The state vector is given by

$$\mathbf{x}_t = \begin{bmatrix} r_t & \dot{r}_t \end{bmatrix}^{\mathrm{T}}, \tag{13}$$

where r_t and $\dot{r}_t$ refer to the range and range rate from a sensor immediately below the falling object on a vertical line. A more general scenario would consist of observing the object at an angle and would therefore have the complete position and velocity vectors as the components of the state vector. Furthermore, due to the atmospheric drag, the ballistic coefficient of the object and the air density would have to be known in order to model the drag force in the direction opposite to the object's velocity vector in addition to the Earth's gravitational force. In real applications, the ballistic coefficient is modelled as part of the state since it is in general unknown. For air density, a model is usually assumed with various parameters which depend on the physical characteristics of the local atmosphere. These parameters vary randomly with sensor location and are dependent on meteorological factors. Here, we assume a simple model for the air density, and we assume that the ballistic coefficient of the falling object is known exactly. This scenario is not entirely unrealistic since one can envisage an experiment consisting of dropping say a bullet whose ballistic coefficient is specified by the manufacturer from a high altitude. Here, we follow the reentry example given in section 5.2 of [9].

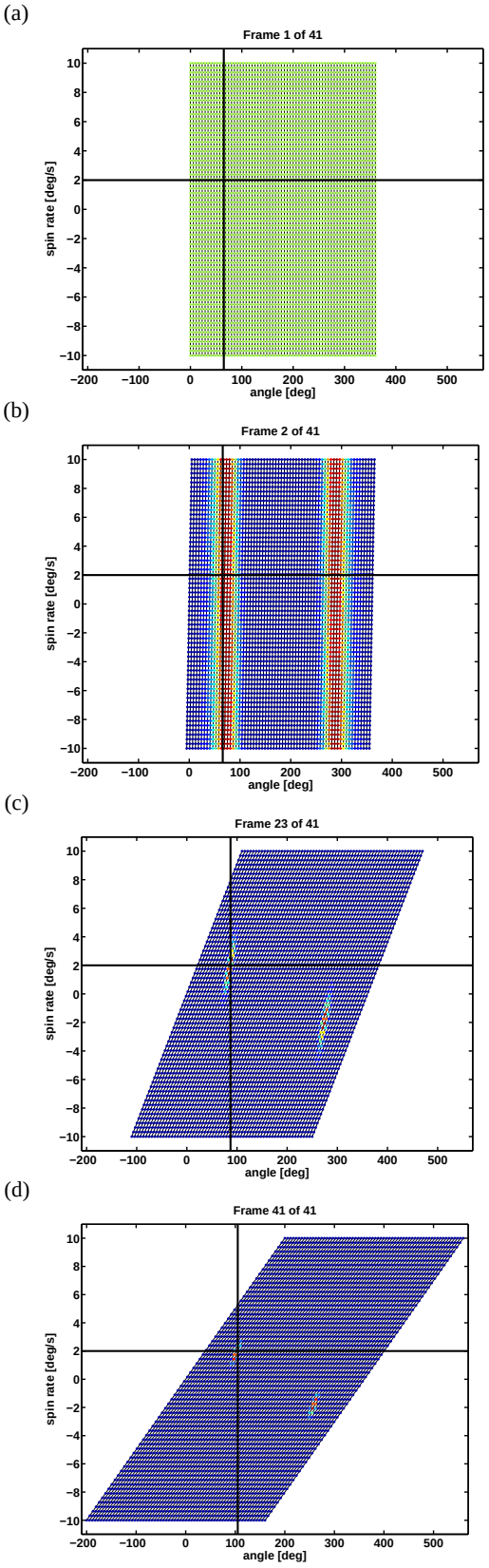

Fig. 1. Unnormalized conditional density for the spinning disk problem: (a) initial uniform density with uniform mesh, (b) density after one measurement update, (c) density after twenty-two measurement updates, and (d) density after forty measurement updates. The true state is shown as the cross section of the two thick lines. cross-section of the two thick lines.

TABLE II

REENTRY EXAMPLE SIMULATION PARAMETERS

PARAMETERS	VALUE	DIMENSION
Number of grid points per dimension	32	N/A
True initial range	6.1	km
True initial range rate	3048	$\mathrm{m} \cdot \mathrm{s}^{-1}$
Range process noise variance	0	$\mathrm{m}^2 \cdot \mathrm{s}^{-1}$
Range rate process noise variance	0.01	$\mathrm{m}^2 \cdot \mathrm{s}^{-3}$
Measurement error std. dev.	1	km
Sampling time	1	s

The system dynamic model is given by

$$\mathbf{f}_t(\mathbf{x}_t) = \begin{bmatrix} -\dot{r}_t \\ -\frac{\rho(r_t) g \dot{r}_t^2}{2\beta} + g \end{bmatrix} \tag{14}$$

where $g = 9.81\ \mathrm{m} \cdot \mathrm{s}^{-2}$ is the gravitational acceleration, and $\beta = 19161\ \mathrm{kg} \cdot \mathrm{m}^{-1} \cdot \mathrm{s}^{-2}$ is the ballistic coefficient of the falling object, which, for simplicity we are assuming is known exactly. The air density is given by

$$\rho(r_t) = \gamma e^{-\eta r_t} \tag{15}$$

where $\gamma = 1.754\ \mathrm{kg} \cdot \mathrm{m}^{-3}$ and $\eta = 1.49 \times 10^{-4}\ \mathrm{m}^{-1}$. From equation (14), it can be shown that the trace of the system Jacobian matrix is given by

$$\mathrm{tr}\left(\frac{\partial \mathbf{f}_t}{\partial \mathbf{x}_t}\right) = \frac{\partial f_{1,t}}{\partial r_t} + \frac{\partial f_{2,t}}{\partial \dot{r}_t} = -\frac{\rho(r_t) g \dot{r}_t}{\beta} \tag{16}$$

It follows that the solution of the convection equation for this problem can be obtained from the following system of nonlinear ODEs:

$$\frac{d}{dt}\begin{bmatrix} \mathbf{x}_t \\ p \end{bmatrix} = \begin{bmatrix} \mathbf{f}_t(\mathbf{x}_t) \\ \frac{\rho(r_t) g \dot{r}_t p}{\beta} \end{bmatrix}, \tag{17}$$

which we solve numerically using a standard ODE solver such as the 4th-order Runge–Kutta integrator. Finally, the linear measurement model is given by

$$h_k = r_{t_k}. \tag{18}$$

The parameters of the simulation are listed in Table II. The results are shown in Figure 2. As in the previous example, it is evident from the Figures that the spacing of the mesh changes over time. Again this is a direct consequence of our numerical integration of the convection equation. Note that for this example, the nonlinear system dynamic model results in a highly deformed mesh distribution after a few iterations of the algorithm.

V. ADAPTIVE MESHING

The algorithm presented in section III suffers from the so called curse of dimensionality; that is, the computational complexity of solving the Fokker–Planck equation increases exponentially with the dimension of the state vector [3]. As can be seen in Figure 1, the value of the conditional density is negligible for many grid points. It is expected that by adaptively changing the mesh size over the solution domain,

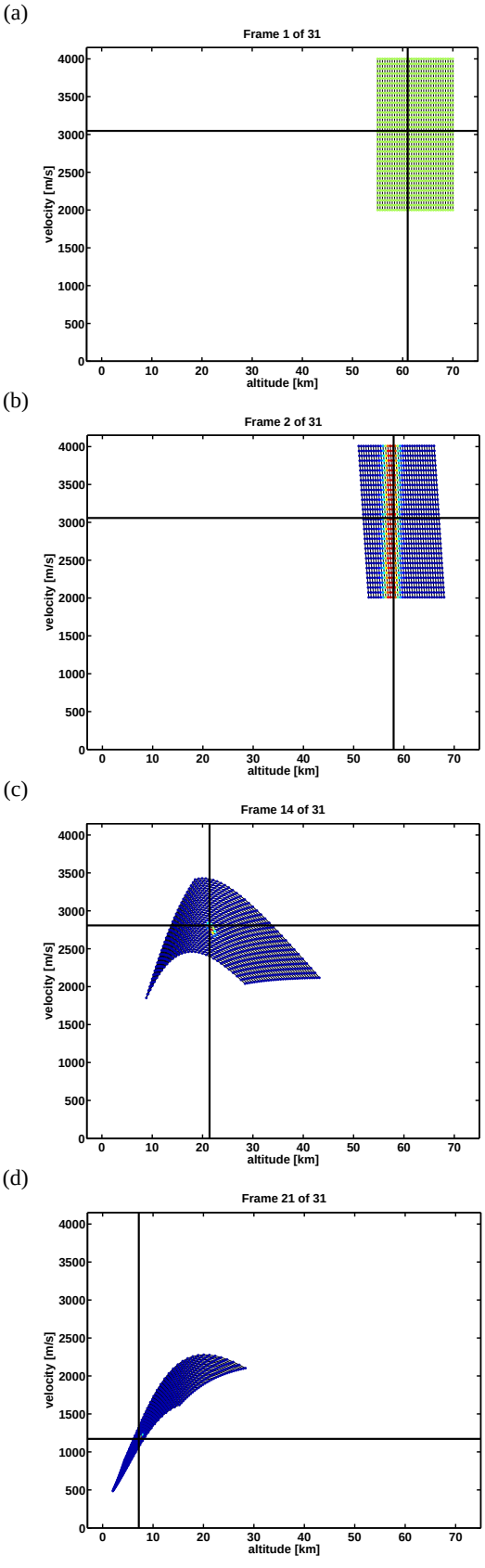

Fig. 2. Unnormalized conditional density for the reentry problem: (a) initial uniform density with uniform mesh, (b) density after one measurement update, (c) density after thirteen measurement updates, and (d) density after twenty measurement updates. The true state is shown as the cross section of the two thick lines. cross-section of the two thick lines.

it is possible to reduce the computational complexity of the algorithm by concentrating on regions of high probability density. Many useful adaptive techniques have been developed for the numerical solution of PDEs (see e.g. [10]). One possible technique suggested by Daum and Krichman [11] is the adjoint method [12].

We note that for a majority of applications, instead of the conditional density, we are mainly interested in the solution of expectation integrals of the form:

$$J = \langle \psi | p \rangle = \int \psi(\mathbf{x}) p(\mathbf{x}|\mathbf{Z}_k)\, d\mathbf{x}. \tag{19}$$

The adjoint method seeks the solution that minimizes the error in estimating the functional J. If we regard the differential equation (written in operator notation)

$$Lp = \phi \tag{20}$$

as the "prime" problem, then we can define a "dual" problem described by

$$L^* q = \psi, \tag{21}$$

where L^* is the adjoint of the L operator. From the definition of the adjoint operator, it follows that the expectation integral (19) can be equivalently expressed in terms of q:

$$J = \langle \psi | p \rangle = \langle L^* q | p \rangle = \langle q | L p \rangle = \langle q | \phi \rangle \,. \tag{22}$$

It can be shown that the error δJ in estimating the expectation integral (19) satisfies the bound [13]:

$$\|\delta J\| \leq \left\| L^{-1} \right\| \|L p_h - \phi\| \, \|L^* q_h - \psi\| \,, \tag{23}$$

where p_h and q_h refer to the approximate solutions of (20) and (21), respectively. By adaptively changing the mesh size such that the error bound (23) is minimized, it can be shown that the computational complexity of solving the Fokker–Planck equation can be significantly reduced [12]. We defer to a future study the application of the adjoint method to the wave filter.

VI. Summary and Conclusion

We have shown that the Fokker–Planck equation for the nonlinear filtering problem with discrete-time measurements can be solved efficiently using a split-step technique. However, the computational complexity of the algorithm still increases exponentially with the dimension of the state vector. We recommend the adjoint method discussed in section V as a means to adaptively changing the mesh size and thereby reducing the computational complexity of the algorithm.

Acknowledgements

We wish to thank Dr. K. P. Dunn and Dr. C. B. Chang of MIT Lincoln Laboratory for helpful discussions.

References

[1] K. Kastella, "Finite difference methods for nonlinear filtering and automatic target recoginition," in *Multitarget/Multisensor Tracking*, Y. Bar-Shalom and W. D. Blair, Eds. Artech House, Inc., 2000, vol. III.

[2] F. E. Daum and J. Huang, "The curse of dimensionality and particle filters," in *Proceedings of IEEE Aerospace Conference*, March 2004.

[3] F. E. Daum, "Nonlinear filters: beyond the Kalman filter," *IEEE AES Systems Magazine*, August 2005.

[4] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. New York: Academy Press, 1970.

[5] F. E. Daum, "New exact nonlinear filters," in *Bayesian Analysis of Time Series and Dynamic Models*, J. C. Spall, Ed. New York: Marcel Dekker, Inc., 1988.

[6] D. J. Tritton, *Physical Fluid Dynamics*, 2nd ed. Oxford Sciences Publications, 1988.

[7] J. D. Hoffman, *Numerical Methods for Engineers and Scientists*. McGraw Hill, 1992.

[8] A. C. Schmidt, "Dynamic Bayesian networks for the classification of spinning disks," Master's thesis, Massachusetts Institute of Technology, 2004.

[9] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Boston: Artech House, 2004.

[10] K. Eriksson, D. Estep, P. Hansbo, and C. Johnson, "Introduction to adaptive methods for differential equations," in *Acta Numerica*. Cambridge University Press, 1995, pp. 105–158.

[11] F. E. Daum and M. Krichman, "Meshfree adjoint methods for nonlinear filtering," in *Proccedings of IEEE Aerospace Conference*, Big Sky, Montana, March 2006.

[12] W. Bangerth and R. Rannacher, *Adaptive finite element methods for differential equations*. Birkhauser Inc., 2003.

[13] N. Pierce and M. B. Giles, "Adjoint and defect error bounding and correction for functional estimates," *Journal of Computational Phyiscs*, 2004.