University of Toronto
Institute for Aerospace Studies
Technical Report
24 June 2003

Stochastic Algebra for Continuous Variables

T D Barfoot and G M T D'Eleuterio

University of Toronto Institute for Aerospace Studies
4925 Dufferin Street, Toronto, Ontario, Canada, M3H 5T6
<tim.barfoot@utoronto.ca, gabriele.deleuterio@utoronto.ca>

Keywords

stochastic algebra, probability density function, Fourier series, Bayes filtering

Abstract

We report on a method of working with probability density functions over a finite interval using a novel infinitely dimensional algebra. A connection to Fourier analysis is drawn and a method of approximating densities is developed. We also look at the problem of filtering and show how this algebra is helpful in discussing the Kalman and Bayes filters.

1 Introduction

In previous work [2, 3, 4] we have developed a linear algebra for working with probability densities over a finite number of discrete states. In this work we discuss the extension to a finite continuous interval. Let $\boldsymbol{f}(x)$ be a probability density function (PDF) for the random variable, x, over the interval $\big[a, b\big]$. This is a non-negative function that satisfies

$$\int_a^b \boldsymbol{f}(x)\, dx = 1$$

That is, it satisfies the axiom of total probability [15]. Note, the *probability*, $p\left(x \in \big[x_1, x_2\big]\right)$, that x is in some sub-interval, $\big[x_1, x_2\big]$, is given by the integral under the density function in that region:

$$p\left(x \in \big[x_1, x_2\big]\right) = \int_{x_1}^{x_2} \boldsymbol{f}(x)\, dx$$

Let $\boldsymbol{f}(x)$ and $\boldsymbol{g}(x)$ be PDFs over the interval $\big[a, b\big]$. We define the *stochastic addition* of these two densities as

$$\boldsymbol{f}(x) \oplus \boldsymbol{g}(x) \triangleq \frac{\boldsymbol{f}(x)\boldsymbol{g}(x)}{\int_a^b \boldsymbol{f}(x)\boldsymbol{g}(x)\, dx}$$

In other work, stochastic addition has been called *logarithmic opinion pooling* [7] and can be traced back to the *Nash product* [14]. We will use the symbol, $\ominus$, to indicate *stochastic subtraction*. The *scalar multiplication* of density, $\boldsymbol{f}(x)$, by a real scalar, λ, is defined as

$$\lambda \cdot \boldsymbol{f}(x) \triangleq \frac{\boldsymbol{f}(x)^\lambda}{\int_a^b \boldsymbol{f}(x)^\lambda\, dx}$$

Under these definitions, the space of all probability density functions over the finite interval, $\big[a, b\big]$, is a vector space [16] over the field $\mathbb{R}$ under the addition and scalar multiplication defined here. The new zero vector, $\boldsymbol{\omega}(x)$, is the uniform probability density function given by

$$\boldsymbol{\omega}(x) \triangleq \frac{1}{d}$$

where $d \triangleq b - a$. It can be shown that we have an inner product space by using the following definition of the inner product:

$$\langle \boldsymbol{f}(x), \boldsymbol{g}(x) \rangle \triangleq \frac{1}{2d} \int_a^b \int_a^b \ln \frac{\boldsymbol{f}(x)}{\boldsymbol{f}(y)} \ln \frac{\boldsymbol{g}(x)}{\boldsymbol{g}(y)} \, dx \, dy$$

By defining an appropriate basis, we may now express a probability density function as a sum of terms such as:

$$\boldsymbol{f}(x) = \bigoplus_{n=1}^{\infty} \alpha_n \cdot \boldsymbol{\xi}_n(x) \oplus \beta_n \cdot \boldsymbol{\zeta}_n(x)$$

where $\xi_n(x)$ and $\boldsymbol{\zeta}_n(x)$ are basis vectors and α_n and β_n are the coefficients. For example, we might chose our basis vectors to be given by

$$\begin{aligned} \boldsymbol{\xi}_n(x) &= \frac{\exp \phi_n(x)}{\int_a^b \exp \phi_n(x) \, dx} & \boldsymbol{\zeta}_n(x) &= \frac{\exp \psi_n(x)}{\int_a^b \exp \psi_n(x) \, dx} \\ \phi_n(x) &= \sqrt{\tfrac{2}{d}} \cos\left(2\pi n \tfrac{x-a}{d}\right) & \psi_n(x) &= \sqrt{\tfrac{2}{d}} \sin\left(2\pi n \tfrac{x-a}{d}\right) \end{aligned}$$

which form a *complete, orthonormal* basis [9]. Specifically, we have

$$(\forall m, n) \qquad \langle \boldsymbol{\xi}_m(x), \boldsymbol{\xi}_n(x) \rangle = \langle \boldsymbol{\zeta}_m(x), \boldsymbol{\zeta}_n(x) \rangle = \delta_{mn}, \qquad \langle \xi_m(x), \boldsymbol{\zeta}_n(x) \rangle = 0$$

where δ_{mn} is the Kronecker-delta. This follows directly from the definition of the inner product but makes use of the additional knowledge that

$$(\forall n) \qquad \int_a^b \phi_n(x) \, dx = \int_a^b \psi_n(x) \, dx = 0$$

Expanding in terms of this basis is equivalent to doing a Fourier expansion [6] and in fact the coefficients, α_n and β_n, are the usual Fourier coefficients. Note, however, we do not bother including the constant term, α_0, from the canonical Fourier expansion since

$$\boldsymbol{\xi}_0 = \frac{\exp(\cos 0)}{\int_a^b \exp(\cos 0) \, dx} = \frac{1}{d} = \boldsymbol{\omega}(x)$$

which is the zero vector and will not contribute to the expansion. The reason we do not need this term is simply because we are trying to represent probability densities, which must satisfy the axiom of total probability (the integral under the curve is constrained to be 1).

Since we have an orthonormal basis, to determine the expansion coefficients of a particular probability density function, $\boldsymbol{f}(x)$, we must compute the inner product of the density with the appropriate basis vector as follows:

$$\begin{aligned} \alpha_n &= \langle \boldsymbol{f}(x), \boldsymbol{\xi}_n(x) \rangle \\ \beta_n &= \langle \boldsymbol{f}(x), \boldsymbol{\zeta}_n(x) \rangle \end{aligned}$$

which are the lengths of the *projections* of the PDF onto our basis functions. Equivalently, we can simply do a Fourier expansion on the logarithm of our density, $\boldsymbol{f}(x)$, as follows:

$$\begin{aligned} \alpha_n &= \sqrt{\frac{2}{d}} \int_a^b \cos\left(2\pi n \frac{x-a}{d}\right) \ln \boldsymbol{f}(x) \, dx \\ \beta_n &= \sqrt{\frac{2}{d}} \int_a^b \sin\left(2\pi n \frac{x-a}{d}\right) \ln \boldsymbol{f}(x) \, dx \end{aligned}$$

which follows directly from the definition of the inner product. This certainly is of practical importance in that we may now use any standard Fourier series implementation to approximate probability density functions over a finite interval using this framework.

Figure 1 shows an example of approximating a probability density function. The true function is a mixture of truncated Gaussians and a uniform probability density. The figure shows that a very good approximation can be obtained using the Fourier approach outlined here with a small number of terms, $n = 1 \ldots N$.

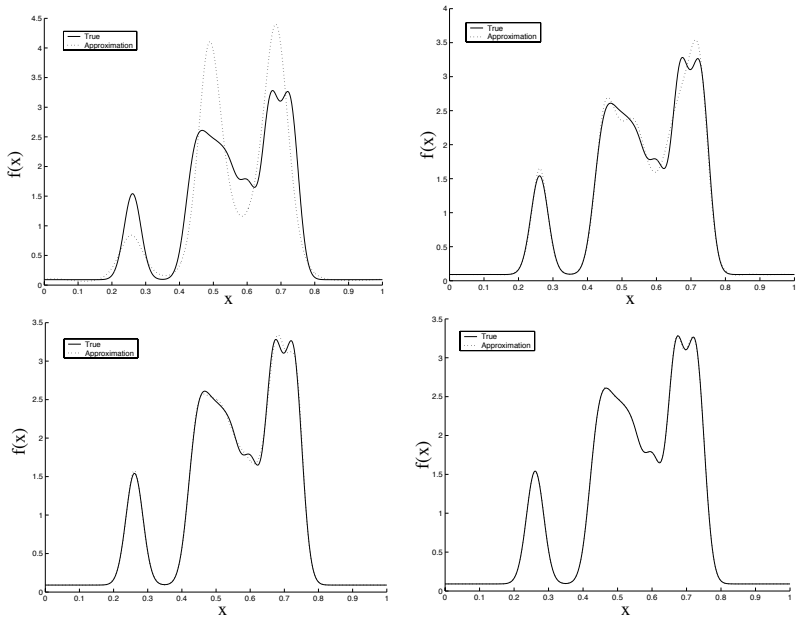

Figure 1: Example of approximating a probability density function using an increasing number of terms in a Fourier expansion ($N = 5, 10, 15, 20$).

2 Inner, Outer and Vector Products

To complete the claim that we have an algebra, we must define the vector product [8]. To do so we must extend the notion of a probability density function to include a second, conditioning variable. Let $\boldsymbol{f}(x|y)$ be a probability density function over $x \in \left[a, b\right]$ conditioned on $y \in \left[r, s\right]$ such that

$$(\forall y) \qquad \int_a^b \boldsymbol{f}(x|y)\, dx = 1$$

This function is directly analogous to a stochastic matrix whose columns (but not rows) sum to one; it is singly but not doubly stochastic.

Let us also, at this juncture, introduce a normalization operator to simplify notation. It will render any non-negative function over a finite interval a probability density function. If $\boldsymbol{f}(x)$ is such a non-negative function, the *normalization* operator is defined as

$$\downarrow_x \boldsymbol{f}(x) \triangleq \frac{\boldsymbol{f}(x)}{\int_a^b \boldsymbol{f}(x)\, dx}$$

where the subscript, x, indicates we are normalizing with respect to x. We also introduce the *stochastic transpose* operator to be

$$\boldsymbol{f}(x|y)^\dagger \triangleq \downarrow_y \boldsymbol{f}(x|y)$$

such that we can change the variable that is stochastic as follows:

$$\boldsymbol{g}(y|x) = \boldsymbol{f}(x|y)^\dagger$$

We now define the *vector product* as

$$\boldsymbol{f}(x|y) \otimes \boldsymbol{g}(y|z) \triangleq \downarrow_x \exp\langle \boldsymbol{f}(x|y)^\dagger, \boldsymbol{g}(y|z)\rangle$$

The *outer product* is defined as

$$\boldsymbol{f}(x) \rangle\langle \boldsymbol{g}(y) \triangleq \downarrow_x \exp\left(\ln \boldsymbol{f}(x) \ln \boldsymbol{g}(y)\right)$$

where we note that the resultant is of the form

$$\boldsymbol{h}(x|y) = \boldsymbol{f}(x) \; \rangle\langle \; \boldsymbol{g}(y)$$

We also have the usual relation between the inner and outer product:

$$(\boldsymbol{f}(x) \; \rangle\langle \; \boldsymbol{g}(y)) \otimes \boldsymbol{h}(y) \equiv \langle \boldsymbol{g}(y), \boldsymbol{h}(y) \rangle \cdot \boldsymbol{f}(x)$$

3 Functional Calculus

Given a complete basis for the probability density functions (e.g., the Fourier basis functions defined earlier), it is not difficult to generalize the notion of calculus to this algebra and in fact, to linearize a nonlinear operator. For simplicity, we will consider operators that map a probability density function over the interval $[a, b]$ to another such function. We denote such an operator by $\boldsymbol{\phi}(\boldsymbol{f})$. Let $\boldsymbol{b}_n$ denote our complete orthonormal basis vectors where $n = 1 \ldots \infty$. The *stochastic partial derivative* of the i^{th} component of the range of $\boldsymbol{\phi}$ with respect to the j^{th} component of the domain of $\boldsymbol{\phi}$ is then

$$\frac{\eth \boldsymbol{\phi}_{b_i}}{\eth \boldsymbol{f}_{b_j}} \triangleq \lim_{\lambda \to 0} \frac{1}{\lambda} \langle \boldsymbol{b}_i, \boldsymbol{\phi}\left(\boldsymbol{f} \oplus \lambda \cdot \boldsymbol{b}_j\right) \ominus \boldsymbol{\phi}\left(\boldsymbol{f}\right) \rangle$$

The resulting linearized operator (Jacobian) is

$$\boldsymbol{\Phi} = \bigoplus_{i,j} \frac{\eth \boldsymbol{\phi}_{b_i}}{\eth \boldsymbol{f}_{b_j}} \cdot \boldsymbol{b}_i \; \rangle\langle \; \boldsymbol{b}_j$$

which has been expressed in terms of the outer product. Note, that this would have to be evaluated at a particular *operating point* to be used in practice. We may also wish to use only a finite number of basis functions to approximate the linearization process.

As one example, consider the simple case when $\boldsymbol{\phi}(\boldsymbol{f}) = \boldsymbol{A} \otimes \boldsymbol{f}$, which is already a linear operator. In this case the linearization procedure outlined above shows that $\boldsymbol{\Phi} = \boldsymbol{A}$:

$$\begin{aligned}
\boldsymbol{\Phi} &= \bigoplus_{i,j} \langle \boldsymbol{b}_i, \boldsymbol{A} \otimes \boldsymbol{b}_j \rangle \cdot (\boldsymbol{b}_i \; \rangle\langle \; \boldsymbol{b}_j) \\
&= \bigoplus_{i,j} \left\langle \boldsymbol{b}_i, \left(\bigoplus_{k,l} a_{kl} \cdot (\boldsymbol{b}_k \; \rangle\langle \; \boldsymbol{b}_l) \right) \otimes \boldsymbol{b}_j \right\rangle \cdot (\boldsymbol{b}_i \; \rangle\langle \; \boldsymbol{b}_j) \\
&= \bigoplus_{k,l} a_{kl} \cdot \bigoplus_{i,j} \langle \boldsymbol{b}_i, \boldsymbol{b}_k \rangle \langle \boldsymbol{b}_j, \boldsymbol{b}_l \rangle \cdot (\boldsymbol{b}_i \; \rangle\langle \; \boldsymbol{b}_j) \\
&= \bigoplus_{k,l} a_{kl} \cdot (\boldsymbol{b}_k \; \rangle\langle \; \boldsymbol{b}_l) \\
&= \boldsymbol{A}
\end{aligned}$$

As another example, consider the continuous version of a Markov [13] transition matrix:

$$\boldsymbol{\phi}(\boldsymbol{f}) = \int_a^b \boldsymbol{g}(y|x) \boldsymbol{f}(x) \, dx$$

where $\boldsymbol{g}(y|x)$ is the conditional density describing how x and y are dependent. Let us assume we are using the Fourier basis functions described earlier. Some computation reveals that

$$\frac{\eth \boldsymbol{\phi}_{\xi_i}}{\eth \boldsymbol{f}_{\xi_j}} = \frac{2}{d} \int_a^b \frac{1}{\boldsymbol{\phi}(y)} \cos\left(2\pi i \frac{y-a}{d}\right) \int_a^b \cos\left(2\pi j \frac{x-a}{d}\right) \boldsymbol{g}(y|x) \boldsymbol{f}(x) \, dx \, dy$$

which still requires us to evaluate $\boldsymbol{\phi}(y) \triangleq \boldsymbol{\phi}(\boldsymbol{f}(x))$, unfortunately. This expression simplifies to

$$\frac{\eth \boldsymbol{\phi}_{\xi_i}}{\eth \boldsymbol{f}_{\xi_j}} = \frac{2}{d} \int_a^b \cos\left(2\pi i \frac{y-a}{d}\right) \int_a^b \cos\left(2\pi j \frac{x-a}{d}\right) \boldsymbol{g}(y|x) \, dx \, dy$$

if we assume $\boldsymbol{f}(x) = \boldsymbol{\omega}$ and $\boldsymbol{\phi}(\boldsymbol{\omega}) = \boldsymbol{\omega}$, which is the assumption that our zero vector, $\boldsymbol{\omega}$, is the fixed-point of the Markov transition matrix. The expressions are similar for the derivative involving the $\boldsymbol{\zeta}_n$ basis functions (i.e., the sine terms).

We can also speak of the gradient of a functional of a PDF. Let $\phi(\boldsymbol{f})$ map PDFs to real numbers. The *stochastic partial derivative* of ϕ with respect to the i^{th} basis for $\boldsymbol{f}$ is defined as

$$\frac{\eth\phi}{\eth \boldsymbol{f}_{b_i}} \triangleq \lim_{\lambda \to 0} \frac{\phi(\boldsymbol{f} \oplus \lambda \cdot \boldsymbol{b}_i) - \phi(\boldsymbol{f})}{\lambda}$$

The *stochastic gradient* of ϕ with respect to $\boldsymbol{f}$ is then

$$\nabla_f \phi \triangleq \bigoplus_i \frac{\eth\phi}{\eth \boldsymbol{f}_{b_i}} \cdot \boldsymbol{b}_i$$

where again in practice we may only choose to use a finite number of basis functions to approximate the gradient. As an example consider the below functionals and their gradients shown on the right:

$$\begin{aligned} \phi(\boldsymbol{f}) &= \tfrac{1}{2}\langle \boldsymbol{f} \ominus \boldsymbol{g}, \boldsymbol{f} \ominus \boldsymbol{g} \rangle & \nabla_f \phi &= \boldsymbol{f} \ominus \boldsymbol{g} \\ \phi(\boldsymbol{f}) &= \int_a^b \boldsymbol{g}(x) \ln \tfrac{\boldsymbol{g}(x)}{\boldsymbol{f}(x)}\, dx & \nabla_f \phi &= \tfrac{\exp \boldsymbol{f}(x)}{\int_a^b \exp \boldsymbol{f}(x)\, dx} \ominus \tfrac{\exp \boldsymbol{g}(x)}{\int_a^b \exp \boldsymbol{g}(x)\, dx} \end{aligned}$$

The first is simply the inner product of the difference of two PDFs (similar to Euclidean squared distance) while the second is the Kullback-Leibler [12] information theoretic distance between two PDFs. Note that these gradients are exact and hence no approximation is necessary. Also, both measures are positive definite and thus equating either gradient to the zero vector, $\boldsymbol{\omega}$, shows $\boldsymbol{f} = \boldsymbol{g}$ is a minimum. The strength of the algebra here is that no Lagrange multiplier was needed to enforce the axiom of total probability (it is included automatically).

4 Inference

A key issue when working with stochastic equations is inferring one probability density from another using Bayes' rule. In other words, computing the integral

$$\begin{aligned} \boldsymbol{f}(y) &= \int_a^b \boldsymbol{g}(y|x) \boldsymbol{h}(x)\, dx \\ &= \phi(\boldsymbol{h}(x)) \end{aligned}$$

which can be quite expensive to do in the general nonlinear case. In this section we develop a method of approximating this integral by viewing it as an operator, ϕ, taking $\boldsymbol{h}(x)$ to $\boldsymbol{f}(y)$. Our approach will be to linearize the operator about the uniform probability density. We will work with the Fourier basis functions described earlier but point out that the following can use any orthonormal basis.

Let the conditional density, $\boldsymbol{g}(y|x)$, be of the form

$$\begin{aligned} \boldsymbol{g}(y|x) &= \downarrow_y \exp\left(\frac{2}{d} \sum_{i,j} \alpha_{i,j} \cos\left(2\pi i \frac{y-a}{d} + 2\pi j \frac{x-a}{d} \right) + \beta_{i,j} \sin\left(2\pi i \frac{y-a}{d} + 2\pi j \frac{x-a}{d} \right) \right) \\ &= \bigoplus_{i,j} \alpha_{i,j} \cdot \left(\boldsymbol{\xi}_i(y) \rangle\langle \boldsymbol{\xi}_j(x) \ominus \boldsymbol{\zeta}_i(y) \rangle\langle \boldsymbol{\zeta}_j(x) \right) \oplus \beta_{i,j} \cdot \left(\boldsymbol{\xi}_i(y) \rangle\langle \boldsymbol{\zeta}_j(x) \ominus \boldsymbol{\zeta}_i(y) \rangle\langle \boldsymbol{\xi}_j(x) \right) \end{aligned}$$

where the second line shows how it may be expressed in terms of outer products of the Fourier basis functions mentioned earlier.

Inserting $\boldsymbol{g}(y|x)$ into our expression for the Jacobian of the inference (Markov) operator computed previously, and making the assumption that $\alpha_{i,j}$ and $\beta_{i,j}$ are small so that

$$\boldsymbol{g}(y|x) \approx \downarrow_y \left(1 + \frac{2}{d} \sum_{i,j} \alpha_{i,j} \cos\left(2\pi i \frac{y-a}{d} + 2\pi j \frac{x-a}{d} \right) + \beta_{i,j} \sin\left(2\pi i \frac{y-a}{d} + 2\pi j \frac{x-a}{d} \right) \right)$$

we have, by employing $\exp x \approx 1 + x$ for small x, that the linearized operator is given by

$$\boldsymbol{\Phi} = \boldsymbol{g}(y|x)$$

That is, the conditional density is the linearized operator of ϕ (for densities close to the uniform vector). This implies we may do approximate inference by computing

$$\boldsymbol{f}(y) = \boldsymbol{g}(y|x) \otimes \boldsymbol{h}(x)$$

Suppose that

$$\begin{aligned} \boldsymbol{f}(y) &= \bigoplus_i \nu_i \cdot \boldsymbol{\xi}_i(y) \oplus \rho_i \cdot \boldsymbol{\zeta}_i(y) \\ \boldsymbol{h}(x) &= \bigoplus_j \gamma_j \cdot \boldsymbol{\xi}_j(x) \oplus \eta_j \cdot \boldsymbol{\zeta}_j(x) \end{aligned}$$

Then, defining

$$\boldsymbol{\nu} \triangleq \begin{bmatrix} \nu_0 \\ \vdots \\ \nu_N \end{bmatrix}, \quad \boldsymbol{\rho} \triangleq \begin{bmatrix} \rho_0 \\ \vdots \\ \rho_N \end{bmatrix}, \quad \boldsymbol{\gamma} \triangleq \begin{bmatrix} \gamma_0 \\ \vdots \\ \gamma_N \end{bmatrix}, \quad \boldsymbol{\eta} \triangleq \begin{bmatrix} \eta_0 \\ \vdots \\ \eta_N \end{bmatrix}, \quad \boldsymbol{\alpha} \triangleq \begin{bmatrix} \alpha_{0,0} & \cdots & \alpha_{0,N} \\ \vdots & \ddots & \vdots \\ \alpha_{N,0} & \cdots & \alpha_{N,N} \end{bmatrix}, \quad \boldsymbol{\beta} \triangleq \begin{bmatrix} \beta_{0,0} & \cdots & \beta_{0,N} \\ \vdots & \ddots & \vdots \\ \beta_{N,0} & \cdots & \beta_{N,N} \end{bmatrix}$$

we have

$$\begin{bmatrix} \boldsymbol{\nu} \\ \boldsymbol{\rho} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha} & \boldsymbol{\beta} \\ \boldsymbol{\beta} & -\boldsymbol{\alpha} \end{bmatrix} \begin{bmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\eta} \end{bmatrix}$$

as our approximate process of inference in terms of the Fourier coefficients. We note for the determinant of the conditional density matrix that

$$\left| \begin{bmatrix} \boldsymbol{\alpha} & \boldsymbol{\beta} \\ \boldsymbol{\beta} & -\boldsymbol{\alpha} \end{bmatrix} \right|^2 = \left| \begin{bmatrix} \boldsymbol{\alpha} & \boldsymbol{\beta} \\ \boldsymbol{\beta} & -\boldsymbol{\alpha} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} & \boldsymbol{\beta} \\ \boldsymbol{\beta} & -\boldsymbol{\alpha} \end{bmatrix} \right| = \left| \begin{bmatrix} \boldsymbol{\alpha}^2 + \boldsymbol{\beta}^2 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\alpha}^2 + \boldsymbol{\beta}^2 \end{bmatrix} \right| = \left| \boldsymbol{\alpha}^2 + \boldsymbol{\beta}^2 \right|^2$$

so that the product of its $2N$ eigenvalues is given by

$$\left(\prod_{n=1}^{2N} \lambda_n \right)^2 = \left| \boldsymbol{\alpha}^2 + \boldsymbol{\beta}^2 \right|^2$$

where the n^{th} eigenvalue is denoted λ_n.

Recall that we assumed all of the $\alpha_{i,j}$ and $\beta_{i,j}$ to be small. To be safe, a better criterion when using this approximation is that all of the eigenvalues are smaller than 1 in magnitude, which will ensure that information is lost through inference (as it is with the original operator). The special case that $\boldsymbol{\alpha} = \text{diag}\{\alpha_m\}$ and $\boldsymbol{\beta} = \text{diag}\{\beta_m\}$ (i.e., they are diagonal), reveals that

$$(\forall m = 1 \ldots N) \quad \lambda_m^2 = \alpha_m^2 + \beta_m^2$$

where we note that there are actually two sets of eigenvalues whose squares are identical. Thus, the boundary between information being lost and gained through inference is given by $\alpha_m^2 + \beta_m^2 = 1$. Later we will be working with this special case of inference. We will also develop alternate techniques for inferring one density from another, when the assumptions of this section are no longer valid.

5 Joint Densities

We may also denote joint probability densities for N-dimensional continuous variables in our framework as $\boldsymbol{f}(\mathbf{x})$ where $\mathbf{x} = \begin{bmatrix} x_1 & \cdots & x_N \end{bmatrix}^T$ with $x_i \in [a_i,\, b_i]$. We now require

$$\int_{\mathbf{a}}^{\mathbf{b}} \boldsymbol{f}(\mathbf{x})\, d\mathbf{x} = 1$$

where $\mathbf{a} \triangleq \begin{bmatrix} a_1 & \cdots & a_N \end{bmatrix}^T$ and $\mathbf{b} \triangleq \begin{bmatrix} b_1 & \cdots & b_N \end{bmatrix}^T$. In this case we must generalize the definitions of some of the key operations including the inner product:

$$\langle \boldsymbol{f}(\mathbf{x}),\, \boldsymbol{g}(\mathbf{x}) \rangle \triangleq \frac{1}{2|\mathbf{D}|} \int_{\mathbf{a}}^{\mathbf{b}} \int_{\mathbf{a}}^{\mathbf{b}} \ln \frac{\boldsymbol{f}(\mathbf{x})}{\boldsymbol{f}(\mathbf{y})} \ln \frac{\boldsymbol{g}(\mathbf{x})}{\boldsymbol{g}(\mathbf{y})}\, d\mathbf{x}\, d\mathbf{y}$$

where $\mathbf{D} \triangleq \text{diag}\{d_i\}$ and $d_i \triangleq b_i - a_i$.

We may still use the Fourier series expansion for a joint density but must generalize the basis functions to be

$$\begin{aligned} \boldsymbol{\xi}_{\mathbf{n}}(\mathbf{x}) &= \frac{\exp \phi_{\mathbf{n}}(\mathbf{x})}{\int_{\mathbf{a}}^{\mathbf{b}} \exp \phi_{\mathbf{n}}(\mathbf{x})\, d\mathbf{x}} & \boldsymbol{\zeta}_{\mathbf{n}}(\mathbf{x}) &= \frac{\exp \psi_{\mathbf{n}}(\mathbf{x})}{\int_{\mathbf{a}}^{\mathbf{b}} \exp \psi_{\mathbf{n}}(\mathbf{x})\, d\mathbf{x}} \\ \phi_{\mathbf{n}}(\mathbf{x}) &= \sqrt{\tfrac{2^N}{|\mathbf{D}|}} \cos\left(2\pi \mathbf{n}^T \mathbf{D}^{-1}(\mathbf{x}-\mathbf{a})\right) & \psi_{\mathbf{n}}(\mathbf{x}) &= \sqrt{\tfrac{2^N}{|\mathbf{D}|}} \sin\left(2\pi \mathbf{n}^T \mathbf{D}^{-1}(\mathbf{x}-\mathbf{a})\right) \end{aligned}$$

where $\mathbf{n} \triangleq \begin{bmatrix} n_1 & \cdots n_N \end{bmatrix}^T$. The series expansion is thus given as

$$\boldsymbol{f}(\mathbf{x}) = \bigoplus_{\mathbf{n}}^{\infty} \alpha_{\mathbf{n}} \cdot \boldsymbol{\xi}_{\mathbf{n}}(\mathbf{x}) \oplus \beta_{\mathbf{n}} \cdot \boldsymbol{\zeta}_{\mathbf{n}}(\mathbf{x})$$

where we note that the upper limit, ∞, is meant to imply summations over all possible combinations of the n_i with each variable independently marching towards infinity (it is an N-dimensional summation). The Fourier coefficients are given as before by

$$\begin{aligned} \alpha_{\mathbf{n}} &= \langle \boldsymbol{f}(\mathbf{x}), \boldsymbol{\xi}_{\mathbf{n}}(\mathbf{x}) \rangle \\ \beta_{\mathbf{n}} &= \langle \boldsymbol{f}(\mathbf{x}), \boldsymbol{\zeta}_{\mathbf{n}}(\mathbf{x}) \rangle \end{aligned}$$

A useful simplification occurs in the case that the joint density is independent in the x_i. We will show this in the context of $N = 2$ but the concept generalizes. Suppose we have a two-dimensional density, $\boldsymbol{f}(\mathbf{x}) = \boldsymbol{f}(x_1, x_2)$, which is of the form

$$\boldsymbol{f}(x_1, x_2) = \boldsymbol{g}(x_1)\boldsymbol{h}(x_2)$$

where $\boldsymbol{g}(x_1)$ and $\boldsymbol{h}(x_2)$ are one-dimensional densities. This product is somewhat reminiscent of a *dyadic* and it is not hard to show is indeed a valid probability density function. To compute the Fourier coefficients, we use the logarithmic form mentioned previously to reveal

$$\begin{aligned} \alpha_{n_1,n_2} &= \sqrt{\frac{4}{d_1 d_2}} \int_{a_1}^{b_1} \int_{a_2}^{b_2} \ln \boldsymbol{f}(x_1, x_2) \cos\left(2\pi n_1 \frac{x_1 - a_1}{d_1} + 2\pi n_2 \frac{x_2 - a_2}{d_2}\right) dx_1\, dx_2 \\ &= \begin{cases} \sqrt{\frac{2}{d_1}} \int_{a_1}^{b_1} \ln \boldsymbol{g}(x_1)\, dx_1 + \sqrt{\frac{2}{d_2}} \int_{a_2}^{b_2} \ln \boldsymbol{h}(x_2)\, dx_2 & \text{if } n_1 = 0, n_2 = 0 \\ \sqrt{\frac{2}{d_1}} \int_{a_1}^{b_1} \ln \boldsymbol{g}(x_1) \cos\left(2\pi n_1 \frac{x_1 - a_1}{d_1}\right) dx_1 & \text{if } n_1 > 0, n_2 = 0 \\ \sqrt{\frac{2}{d_2}} \int_{a_2}^{b_2} \ln \boldsymbol{h}(x_2) \cos\left(2\pi n_2 \frac{x_2 - a_2}{d_2}\right) dx_2 & \text{if } n_1 = 0, n_2 > 0 \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} \alpha_{n_1=0} + \alpha_{n_2=0} & \text{if } n_1 = 0, n_2 = 0 \\ \alpha_{n_1} & \text{if } n_1 > 0, n_2 = 0 \\ \alpha_{n_2} & \text{if } n_1 = 0, n_2 > 0 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

where α_{n_1} are the coefficients for $\boldsymbol{g}(x_1)$ alone and α_{n_2} are the coefficients for $\boldsymbol{h}(x_2)$ alone. As before, we will not need the coefficients for $n_1 = 0$ and $n_2 = 0$ since $\boldsymbol{\xi}_0 = \boldsymbol{\omega}$. Similarly we find

$$\beta_{n_1,n_2} = \begin{cases} \beta_{n_1} & \text{if } n_1 > 0, n_2 = 0 \\ \beta_{n_2} & \text{if } n_1 = 0, n_2 > 0 \\ 0 & \text{otherwise} \end{cases}$$

where β_{n_1} are the coefficients for $\boldsymbol{g}(x_1)$ alone and β_{n_2} are the coefficients for $\boldsymbol{h}(x_2)$ alone. Using these simplifications, the expansion for $\boldsymbol{f}(x_1, x_2)$ thus becomes

$$\begin{aligned} \boldsymbol{f}(x_1, x_2) &= \left(\bigoplus_{n_1=1}^{\infty} \alpha_{n_1} \cdot \boldsymbol{\xi}_{n_1}(x_1) \oplus \beta_{n_1} \cdot \boldsymbol{\zeta}_{n_1}(x_1) \right) \oplus \left(\bigoplus_{n_2=1}^{\infty} \alpha_{n_2} \cdot \boldsymbol{\xi}_{n_2}(x_2) \oplus \beta_{n_2} \cdot \boldsymbol{\zeta}_{n_2}(x_2) \right) \\ &= \boldsymbol{g}(x_1) \oplus \boldsymbol{h}(x_2) \end{aligned}$$

which has the nice property of separation into a summation of terms. To see how this works directly, we can use the definition of stochastic addition, generalized to two dimensions:

$$\begin{aligned} \boldsymbol{g}(x_1) \oplus \boldsymbol{h}(x_2) &= \frac{\boldsymbol{g}(x_1)\boldsymbol{h}(x_2)}{\int_{a_1}^{b_1} \int_{a_2}^{b_2} \boldsymbol{g}(x_1)\boldsymbol{h}(x_2)\, dx_1\, dx_2} \\ &= \frac{\boldsymbol{g}(x_1)\boldsymbol{h}(x_2)}{\left(\int_{a_1}^{b_1} \boldsymbol{g}(x_1)\, dx_1 \right) \left(\int_{a_2}^{b_2} \boldsymbol{h}(x_2)\, dx_2 \right)} \\ &= \boldsymbol{g}(x_1)\boldsymbol{h}(x_2) \end{aligned}$$

where we have used the axiom of total probability. This reveals the true nature of stochastic addition: it is an expression of statistical independence (a.k.a., linearity).

As a final note, to compute the density over a subset of the dimensions of a joint density we integrate out the undesired variables as follows:

$$\boldsymbol{g}(x_1) = \int_{a_2}^{b_2} \boldsymbol{f}(x_1, x_2)\, dx_2$$

which implies the following identity for joint densities wherein all of the x_i are independent:

$$\boldsymbol{f}(\mathbf{x}) = \bigoplus_{i=1}^{N} \int_{a_{j\neq i}}^{b_{j\neq i}} \boldsymbol{f}(\mathbf{x})\, dx_{j\neq i}$$

where the integration in the i^{th} term occurs over all variables, $j = 1 \ldots N$, excluding i.

6 Filtering

Filtering can be thought of as combining estimates from different sources, with differing degrees of certainty. In other words, it is the process of combining multiple PDFs into a single overall PDF. The Kalman filter [11] allows this to be done in an optimal way when each of the estimates is a Gaussian. It is, however, a simplifying case of the general problem of combining PDFs, which is handled quite nicely by the Bayes filter [5, 1, 10]. If we let $\boldsymbol{f}_i(\mathbf{x})$ represent estimates from several sources, the Bayes filter suggests that we combine them into a single estimate, $\boldsymbol{f}(\mathbf{x})$, through

$$\boldsymbol{f}(\mathbf{x}) = \bigoplus_i \boldsymbol{f}_i(\mathbf{x})$$

which is simply vector addition in our framework.

It turns out that if each of the $\boldsymbol{f}_i(\mathbf{x})$ is a Gaussian, the Bayes filter exactly reproduces the Kalman filter[1]. To see this, let each of the $\boldsymbol{f}_i(\mathbf{x})$ be of the form

$$\boldsymbol{f}_i(\mathbf{x}) = \downarrow_{\mathbf{x}} \exp\left(-\frac{1}{2} (\mathbf{x}_i - \bar{\mathbf{x}}_i)^T \mathbf{C}_i^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}_i) \right)$$

[1]This requires that we let the integration limits be $[a, b] = [-\infty, \infty]$ in the definition of stochastic addition, which is allowable. However, the extension to infinite intervals becomes problematic for the inner product.

where $\bar{\mathbf{x}}_i$ and $\mathbf{C}_i$ are the mean vector and covariance matrix, respectively, of the i^{th} individual estimate, $\boldsymbol{f}_i(\mathbf{x})$. Carrying out the stochastic addition reveals

$$\begin{aligned} \mathbf{C}^{-1} &= \sum_i \mathbf{C}_i^{-1} \\ \bar{\mathbf{x}} &= \mathbf{C} \sum_i \mathbf{C}_i^{-1} \bar{\mathbf{x}}_i \end{aligned}$$

where $\bar{\mathbf{x}}$ and $\mathbf{C}$ are the mean vector and covariance matrix of the overall Gaussian estimate, $\boldsymbol{f}(\mathbf{x})$. These equations are the so-called *inverse covariance* updates for the Kalman filter, which do not require the explicit calculation of the usual Kalman gain matrix.

In our framework, the $\mathcal{L}_2$-norm of a PDF is given as

$$\|\boldsymbol{f}\|_{\mathcal{L}_2} \triangleq \sqrt{\langle \boldsymbol{f}, \boldsymbol{f} \rangle}$$

and represents how far away that vector is from the uniform PDF (i.e., the zero vector). In other words, it is a measure of the amount of information in a PDF, not unlike Shannon's information metric [17]. For example, the $\mathcal{L}_2$-norm of a truncated Gaussian is inversely proportional to the variance. To see this, let

$$\boldsymbol{f}(x) = \frac{\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)}{\int_a^b \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)\, dx}$$

where μ and σ^2 are the mean and variance, respectively. Then we have

$$\begin{aligned} \|\boldsymbol{f}\|_{\mathcal{L}_2} &= \left(\frac{1}{2d} \int_a^b \int_a^b \left(\ln \frac{\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)}{\exp\left(-\frac{1}{2\sigma^2}(y-\mu)^2\right)} \right)^2 dx\, dy \right)^{\frac{1}{2}} \\ &= \frac{1}{\sigma^2} \left(\frac{1}{4d} \int_a^b \int_a^b \left((x-\mu)^2 - (y-\mu)^2 \right)^2 dx\, dy \right)^{\frac{1}{2}} \\ &\propto \frac{1}{\sigma^2} \end{aligned}$$

This indicates that a PDF with a greater variance will have less information and vice versa, which has an intuitive appeal: the variance affects only the "length" of the vector representing a truncated Gaussian. Note, the $\mathcal{L}_2$-norm also depends on the location of the mean with respect to the truncated interval. In multiple dimensions, the $\mathcal{L}_2$-norm is inversely proportional to the determinant of the covariance matrix.

As we are working with a vector space, we may now use the $\mathcal{L}_2$-norm to make the following statement about the Bayes filter:

$$\|\boldsymbol{f}\|_{\mathcal{L}_2} \leq \sum_i \|\boldsymbol{f}_i\|_{\mathcal{L}_2}$$

which follows directly from the triangle inequality for inner product spaces [8]. This states that the information in the combined estimator can never exceed the sum of that in the individual estimators. Also, when the individual estimators are not in perfect agreement (i.e., they are not parallel vectors), the information in the overall estimator is reduced. Two truncated Gaussians, for example, are parallel when they have the same mean but, possibly, different variances. For PDFs that are not Gaussian, it is difficult to speak of the Bayes filter being optimal in the sense that the Kalman filter is optimal because variance is not necessarily an appropriate measure to optimize for multi-modal PDFs.

7 Time Dependence

In this section we begin to develop a technique to tackle estimation problems relevant to, for example, mobile robotics. We will develop the equations for one dimension. Let $x \in [a, b]$ represent the position of a mobile robot. Our estimate of the robot's position is represented by the density, $\boldsymbol{f}(x)$. Here we must bring in the concept of time, as the robot will be allowed to move. To do this we generalize our Fourier basis functions to be of the form

$$\begin{aligned} \boldsymbol{\xi}_n(x|v,t) &= \frac{\exp \phi_n(x|v,t)}{\int_a^b \exp \phi_n(x|v,t)\, dx} & \boldsymbol{\zeta}_n(x|v,t) &= \frac{\exp \psi_n(x|v,t)}{\int_a^b \exp \psi_n(x|v,t)\, dx} \\ \phi_n(x|v,t) &= \sqrt{\tfrac{2}{d}} \cos\left(\tfrac{2\pi n}{d}\,(x-a-r)\right) & \psi_n(x|v,t) &= \sqrt{\tfrac{2}{d}} \sin\left(\tfrac{2\pi n}{d}\,(x-a-r)\right) \end{aligned}$$

where

$$r = \int_0^t v(\tau)\, d\tau$$

and $v(t)$ is the instantaneous velocity of the robot and t is the current time. These basis functions are, naturally, the solution form of the classical wave equation.

It can be shown that these time-dependent basis functions may be built from outer products of the time-independent basis functions as follows:

$$\begin{aligned} \boldsymbol{\xi}_n(x|v,t) &= \boldsymbol{\xi}_n(x) \,\rangle\langle\, \boldsymbol{\xi}_n(v,t) \oplus \boldsymbol{\zeta}_n(x) \,\rangle\langle\, \boldsymbol{\zeta}_n(v,t) \\ \boldsymbol{\zeta}_n(x|v,t) &= \boldsymbol{\zeta}_n(x) \,\rangle\langle\, \boldsymbol{\xi}_n(v,t) \ominus \boldsymbol{\xi}_n(x) \,\rangle\langle\, \boldsymbol{\zeta}_n(v,t) \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{\xi}_n(v,t) &= \downarrow_{v,t} \exp\left(\cos\frac{2\pi n r}{d}\right) \\ \boldsymbol{\zeta}_n(v,t) &= \downarrow_{v,t} \exp\left(\sin\frac{2\pi n r}{d}\right) \end{aligned}$$

which follows from the trigonometric summation of angles identities. The outer product formulation becomes quite useful if we are concerned about how well we can determine the current time. In some cases we may not know exactly what time it is due errors in clock readings. Then, to determine the robot's location at some future time, we would require the use of the Markov transition function

$$\boldsymbol{f}(x) = \int \boldsymbol{f}(x|v,t)\boldsymbol{g}(v,t)\, dv\, dt$$

where $\boldsymbol{g}(v,t)$ is a PDF over the velocity and time variables. We will be assuming this density to be a impulse-pdf function, located at the correct velocity-time. We will relax this assumption in a later section.

Two forms for the equations of a propagating PDF (with no change in shape) are given by

$$\begin{aligned} \boldsymbol{f}(x|v,t) &= \bigoplus_n \alpha_n[t]\cdot\boldsymbol{\xi}_n(x) \oplus \beta_n[t]\cdot\boldsymbol{\zeta}_n(x) \\ &= \bigoplus_n \alpha_n[0]\cdot\boldsymbol{\xi}_n(x|v,t) \oplus \beta_n[0]\cdot\boldsymbol{\zeta}_n(x|v,t) \end{aligned}$$

where the first has the velocity/time dependence in the Fourier coefficients while the second has it in the basis functions. To relate the above two forms we compute

$$\begin{aligned} \alpha_n[t] &= \langle \boldsymbol{\xi}_n(x), \bigoplus_m \alpha_m\cdot\boldsymbol{\xi}_m(x|v,t) \oplus \beta_m\cdot\boldsymbol{\zeta}_m(x|v,t)\rangle \\ \beta_n[t] &= \langle \boldsymbol{\zeta}_n(x), \bigoplus_m \alpha_m\cdot\boldsymbol{\xi}_m(x|v,t) \oplus \beta_m\cdot\boldsymbol{\zeta}_m(x|v,t)\rangle \end{aligned}$$

from whence it follows that

$$\begin{aligned}\alpha_n[t+h] &= \alpha_n[t]\cos\left(\frac{2\pi nq}{d}\right) - \beta_n[t]\sin\left(\frac{2\pi nq}{d}\right)\\ \beta_n[t+h] &= \alpha_n[t]\sin\left(\frac{2\pi nq}{d}\right) + \beta_n[t]\cos\left(\frac{2\pi nq}{d}\right)\end{aligned}$$

where $q = \int_t^{t+h} v(\tau)\, d\tau$. which allows us to propagate the n^{th} set of coefficients forward in time, given knowledge of them at a previous time.

This is a perfect, time-reversible propagation, however, which is not realistic for robotics applications. This came about through our assumption of perfect knowledge of the velocity and time. One way to avoid this is to go back to the Markov transition and assume something other than an impulse for $\boldsymbol{g}(v,t)$. We could, for example, use the approximate inference method outlined earlier. Another way to accommodate the growth of uncertainty as the robot moves, is to install a gain which is determined by how far the robot moves. There are several possibilities for the real gain, $k \in [0,1]$, but we might choose

$$k \triangleq e^{-\frac{1}{2}\hat{r}\,\|\boldsymbol{f}(x|v,t)\|}$$

where $\hat{r} = \int_0^t |v(\tau)|\, d\tau$, $\kappa \geq 0$ is a real, and the $\mathcal{L}_2$-norm of $\boldsymbol{f}$ can be gleaned from Parseval's relation:

$$\|\boldsymbol{f}(x|v,t)\|^2 = \sum_n \alpha_n^2[t] + \beta_n^2[t]$$

This gain has the appeal that our PDF takes on the form of a damped waveform. We have deviated somewhat from the classic diffusion model which has v^2 rather than $|v|$ for the damping coefficient. The reason for this deviation is empirical; growth of odometry errors, for example, are often given proportional to the distance the robot has travelled. In the limit of small steps, k behaves as

$$k \triangleq \sqrt{\frac{1}{1+\kappa\hat{r}\,\|\boldsymbol{f}(x|v,t)\|}}$$

The reasoning behind this choice for a gain derives from the desire to have additive variance, σ^2, when $\boldsymbol{f}(x|t)$ is Gaussian:

$$\sigma^2[t+h] = \sigma^2[t] + \kappa \int_t^{t+h} |v|\, d\tau$$

where recall that in a previous section we reasoned that $\|\boldsymbol{f}\| \propto 1/\sigma^2$ for a Gaussian PDF. Note, $\kappa = 0$ implies no loss of information as the PDF propagates. When $\kappa > 0$ information will be monotonically lost.

This part of our model accounts for all generic sensors that rely on the integration of some quantity forward in time (e.g., encoders, inertial sensors). It allows us to propagate an old estimate of the robot's position forward in time as the robot is moving to arrive at

$$\boldsymbol{f}(x|v,t+h) = k \cdot \bigoplus_n \cos\left(\frac{2\pi nq}{d}\right) \cdot \boldsymbol{f}_n(x|v,t) \oplus \sin\left(\frac{2\pi nq}{d}\right) \cdot \boldsymbol{f}_n(x|v,t)^\perp$$

where

$$\begin{aligned}\boldsymbol{f}_n(x|v,t) &\triangleq \alpha_n[t]\cdot\boldsymbol{\xi}_n(x) \oplus \beta_n[t]\cdot\boldsymbol{\zeta}_n(x)\\ \boldsymbol{f}_n(x|v,t)^\perp &\triangleq \alpha_n[t]\cdot\boldsymbol{\zeta}_n(x) \ominus \beta_n[t]\cdot\boldsymbol{\xi}_n(x)\end{aligned}$$

and we have the useful relations

$$\begin{aligned}\langle \boldsymbol{f}_n(x|v,t), \boldsymbol{f}_n(x|v,t)^\perp \rangle &= 0\\ \boldsymbol{\xi}_n(x|v,t)^\perp &= \boldsymbol{\zeta}_n(x|v,t)\\ \boldsymbol{\zeta}_n(x|v,t)^\perp &= \ominus\, \boldsymbol{\xi}_n(x|v,t)\end{aligned}$$

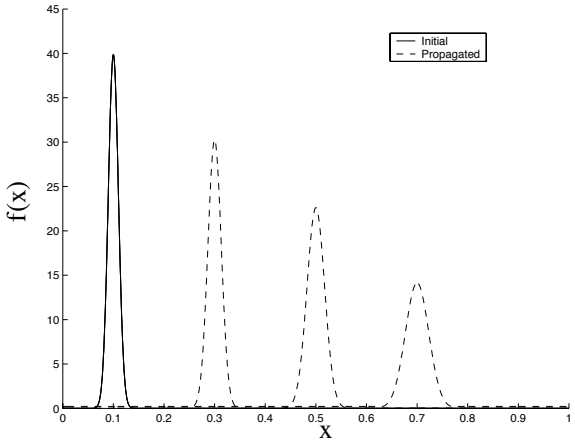

Figure 2: Example of an initial Gaussian PDF that propagates to the right while degrading.

The use of the *orthogonal complement operator*, $^\perp$, avoids the need to use complex numbers in our framework, which we feel is desirable. It is really just translating the PDF by one-quarter wavelength, $d/4$; the quadruple composition of this operator is the identity operator. Figure 2 shows an example of a degrading Gaussian PDF that is propagating to the right.

In terms of the Fourier coefficients, this can be written in matrix form as

$$\mathbf{f}_n[t+h] = k\mathbf{C}_n(h)\mathbf{f}_n[t]$$

where

$$\mathbf{f}_n[t] \triangleq \begin{bmatrix}\alpha_n[t]\\ \beta_n[t]\end{bmatrix} \qquad \mathbf{C}_n(h) \triangleq \begin{bmatrix}\cos\left(\frac{2\pi nq}{d}\right) & -\sin\left(\frac{2\pi nq}{d}\right)\\ \sin\left(\frac{2\pi nq}{d}\right) & \cos\left(\frac{2\pi nq}{d}\right)\end{bmatrix}$$

If we only keep M terms in the Fourier expansion and define

$$\mathbf{f}[t] \triangleq \begin{bmatrix}\mathbf{f}_1[t]\\ \vdots\\ \mathbf{f}_M[t]\end{bmatrix} \qquad \mathbf{C}(h) \triangleq \begin{bmatrix}\mathbf{C}_1(h) & \cdots & \mathbf{0}\\ \vdots & \ddots & \vdots\\ \mathbf{0} & \cdots & \mathbf{C}_M(h)\end{bmatrix}$$

then the combined update equation is

$$\mathbf{f}[t+h] = k\mathbf{C}(h)\mathbf{f}[t]$$

which is nonlinear due to the gain, k, which depends on

$$\|\boldsymbol{f}(x|v,t)\|^2 \approx \mathbf{f}[t]^T\mathbf{f}[t]$$

where the approximation symbol indicates that we have kept only M terms in the Fourier expansion.

Note, it may make more sense to let $k \in [0,1]$ be given by

$$\triangleq e^{-\kappa\hat{r}}$$

whence we have

$$\begin{aligned}\mathbf{f}[t+h+s] &= e^{-\kappa|v|(h+s)}\mathbf{C}(h+s)\mathbf{f}[t]\\ &= \left(e^{-\kappa|v|h}\mathbf{C}(h)\right)\left(e^{-\kappa|v|s}\mathbf{C}(s)\right)\mathbf{f}[t]\end{aligned}$$

when the velocity is constant. This is reasonable to expect as we would like our choice of integration time-step to not affect the behaviour of the equations.

The matrix, $\mathbf{C}(h)$, is time-varying, but it should be noted that it preserves length (it is a rotation matrix and hence $|\mathbf{C}| = 1$); the rotation angle depends on how far the robot actually moved, h. With $k \in [0,1]$,

we are always shortening our vector (losing information). We should also point out that we have made no approximations up to this point except to keep only M basis vectors when approximating the PDF representing the robot's position. Specifically, we have not linearized anything up to this point. This is partly because we have been working only in a single dimension, x.

8 Time and Joint Densities

To extend the time-dependent case to higher dimensions, we again must generalize our basis functions to be

$$\begin{aligned}
\boldsymbol{\xi}_{\mathbf{n}}(\mathbf{x}|\mathbf{v},t) &= \frac{\exp \phi_{\mathbf{n}}(\mathbf{x}|\mathbf{v},t)}{\int_{\mathbf{a}}^{\mathbf{b}} \exp \phi_{\mathbf{n}}(\mathbf{x}|\mathbf{v},t)\, d\mathbf{x}} & \boldsymbol{\zeta}_{\mathbf{n}}(\mathbf{x}|\mathbf{v},t) &= \frac{\exp \psi_{\mathbf{n}}(\mathbf{x}|\mathbf{v},t)}{\int_{\mathbf{a}}^{\mathbf{b}} \exp \psi_{\mathbf{n}}(\mathbf{x}|\mathbf{v},t)\, d\mathbf{x}} \\
\phi_{\mathbf{n}}(\mathbf{x}|\mathbf{v},t) &= \sqrt{\tfrac{2^N}{|\mathbf{D}|}} \cos\left(2\pi \mathbf{n}^T \mathbf{D}^{-1}(\mathbf{x}-\mathbf{a}-\mathbf{r})\right) & \psi_{\mathbf{n}}(\mathbf{x}|\mathbf{v},t) &= \sqrt{\tfrac{2^N}{|\mathbf{D}|}} \sin\left(2\pi \mathbf{n}^T \mathbf{D}^{-1}(\mathbf{x}-\mathbf{a}-\mathbf{r})\right)
\end{aligned}$$

where

$$\mathbf{r} = \int_0^t \mathbf{v}(\tau)\, d\tau$$

and $\mathbf{v}(t) \triangleq \begin{bmatrix} v_1 & \cdots & v_N \end{bmatrix}^T$ is the instantaneous velocity and t is time. As in the one-dimensional case, we find

$$\begin{aligned}
\alpha_{\mathbf{n}}[t+h] &= \alpha_{\mathbf{n}}[t] \cos\left(2\pi \mathbf{n}^T \mathbf{D}^{-1}\mathbf{q}\right) - \beta_{\mathbf{n}}[t] \sin\left(2\pi \mathbf{n}^T \mathbf{D}^{-1}\mathbf{q}\right) \\
\beta_{\mathbf{n}}[t+h] &= \alpha_{\mathbf{n}}[t] \sin\left(2\pi \mathbf{n}^T \mathbf{D}^{-1}\mathbf{q}\right) + \beta_{\mathbf{n}}[t] \cos\left(2\pi \mathbf{n}^T \mathbf{D}^{-1}\mathbf{q}\right)
\end{aligned}$$

where $\mathbf{q} = \int_t^{t+h} \mathbf{v}(\tau)\, d\tau$. This is the expression for the propagation of PDFs without degradation included. We now would like to include a diffusion-like process but it might be directional. That is, information will be lost at different rates, in different directions. Our gain will thus be of the form

$$k_{\mathbf{n}} \triangleq \exp\left(-\frac{\mathbf{n}^T \mathbf{K} \hat{\mathbf{r}}}{\mathbf{n}^T \mathbf{n}}\right)$$

where $\hat{\mathbf{r}} = \int_0^t |\mathbf{v}|\, d\tau$ and $\mathbf{K}$ is an appropriately sized positive semi-definite real square matrix. Since $\mathbf{n}$ is really a vector pointing in the direction of travel of the $\mathbf{n}^{th}$ mode, this has the effect of degrading modes which are parallel to $\mathbf{K}\hat{\mathbf{r}}$. Letting

$$\mathbf{f}_{\mathbf{n}}[t] \triangleq \begin{bmatrix} \alpha_{\mathbf{n}}[t] \\ \beta_{\mathbf{n}}[t] \end{bmatrix} \qquad \mathbf{C}_{\mathbf{n}}(h) \triangleq \begin{bmatrix} \cos\left(2\pi \mathbf{n}^T \mathbf{D}^{-1}\mathbf{q}\right) & -\sin\left(2\pi \mathbf{n}^T \mathbf{D}^{-1}\mathbf{q}\right) \\ \sin\left(2\pi \mathbf{n}^T \mathbf{D}^{-1}\mathbf{q}\right) & \cos\left(2\pi \mathbf{n}^T \mathbf{D}^{-1}\mathbf{q}\right) \end{bmatrix}$$

Then, in terms of the Fourier coefficients, our update can be written in matrix form as

$$\mathbf{f}_{\mathbf{n}}[t+h] = k_{\mathbf{n}} \mathbf{C}_{\mathbf{n}}(h) \mathbf{f}_{\mathbf{n}}[t]$$

so that, again, each mode can be updated completely independently (unless we were to once again include the factor of $\|\boldsymbol{f}(\mathbf{x}|\mathbf{v},t)\|$ in the gain, $k_{\mathbf{n}}$).

9 Gaussian Inference

In the previous two sections, we outlined a method of propagating a PDF waveform including a degradation of information over distance travelled. The resulting model can be used to represent the degradation of information, given the knowledge of the quantity, r (distance travelled). This is really an approximate model based on empirical knowledge. In this section we attempt to provide a better justification for using this type of model. We will be assuming that r is Gaussian with mean, μ, and variance, σ^2. We represent its density using

$$\boldsymbol{h}(r) = \downarrow_r \exp\left(-\frac{(r-\mu)^2}{2\sigma^2}\right)$$

We then seek to compute

$$\boldsymbol{f}(x) = \int \boldsymbol{g}(x|r)\boldsymbol{h}(r)\,dr$$

where

$$\boldsymbol{g}(x|r) = \downarrow_x \exp\sqrt{\frac{2}{d}}\sum_n \left(\alpha_n \cos\left(2\pi n\frac{x-a-r}{d}\right) + \beta_n \sin\left(2\pi n\frac{x-a-r}{d}\right)\right)$$

We then seek to linearize each of the sines and cosines (with respect to r) about the mean of the Gaussian such that

$$\begin{aligned}
\cos\left(2\pi n\frac{x-a-r}{d}\right) &\approx c_n + \frac{2\pi n}{d}s_n(r-\mu) \\
\sin\left(2\pi n\frac{x-a-r}{d}\right) &\approx s_n - \frac{2\pi n}{d}c_n(r-\mu)
\end{aligned}$$

where

$$c_n \triangleq \cos\left(2\pi n\frac{x-a-\mu}{d}\right) \qquad s_n \triangleq \sin\left(2\pi n\frac{x-a-\mu}{d}\right)$$

Note, we could have linearized $\boldsymbol{g}(x|r)$ directly as

$$\boldsymbol{g}(x|r) \approx \boldsymbol{g}(x|\mu) \oplus (r-\mu)\cdot\boldsymbol{g}'(x|\mu)$$

where

$$\boldsymbol{g}'(x|\mu) = \lim_{h\to 0}\frac{1}{h}\cdot\left(\boldsymbol{g}(x|r+h) \ominus \boldsymbol{g}(x|r)\right)\Big|_{r=\mu} = \downarrow_x \exp\sqrt{\frac{2}{d}}\sum_n \frac{2\pi n}{d}\left(\alpha_n s_n - \beta_n c_n\right)$$

Using these approximations, we have for the desired integral that

$$\begin{aligned}
\boldsymbol{f}(x) &= \int \boldsymbol{g}(x|r)\boldsymbol{h}(r)\,dr \\
&\approx \downarrow_x \int \exp\left(-\frac{(r-\mu)^2}{2\sigma^2} + \sqrt{\frac{2}{d}}\sum_n \alpha_n\left(c_n + \frac{2\pi n}{d}s_n(r-\mu)\right) + \beta_n\left(s_n - \frac{2\pi n}{d}c_n(r-\mu)\right)\right)dr \\
&= \exp\sqrt{\frac{2}{d}}\sum_n\left(\alpha_n c_n + \beta_n s_n\right) \oplus \exp\frac{\sigma^2}{d}\left(\sum_n \frac{2\pi n}{d}\left(\alpha_n s_n - \beta_n c_n\right)\right)^2 \oplus \int \exp\left(-\frac{(r-\hat{\mu})^2}{2\sigma^2}\right)dr \\
&= \boldsymbol{g}(x|\mu) \oplus \frac{1}{2}\sigma^2\cdot\boldsymbol{g}'^2(x|\mu)
\end{aligned}$$

where

$$\begin{aligned}
\hat{\mu} &\triangleq \mu + \sigma^2\sqrt{\frac{2}{d}}\sum_n \frac{2\pi n}{d}\left(\alpha_n s_n - \beta_n c_n\right) \\
\boldsymbol{g}'^2(x|\mu) &\triangleq \downarrow_x \exp\left(\sqrt{\frac{2}{d}}\sum_n \frac{2\pi n}{d}\left(\alpha_n s_n - \beta_n c_n\right)\right)^2
\end{aligned}$$

The third term above was dropped since

$$\int_{-\infty}^{\infty} \exp\left(-\frac{(r-\hat{\mu})^2}{2\sigma^2}\right) dr = \sqrt{2\pi\sigma^2}$$

which is a constant with respect to x and thus the term becomes $\boldsymbol{\omega}$, the zero vector. We also see that the integration has yielded a nonlinear term, $\boldsymbol{g}'^2(x|\mu)$. This nonlinear term is somewhat problematic in that we do not have a simple matrix update for the Fourier coefficients. However, we may proceed by decomposing this term over the basis functions once again such that

$$\begin{aligned} \hat{\nu}_n &\triangleq \langle \boldsymbol{\xi}_n, \boldsymbol{g}'^2(x|\mu)\rangle \\ \hat{\rho}_n &\triangleq \langle \boldsymbol{\zeta}_n, \boldsymbol{g}'^2(x|\mu)\rangle \end{aligned}$$

Expanding the cosine coefficient we find

$$\begin{aligned} \hat{\nu}_n &= \sqrt{\frac{2}{d}}\int_a^b \cos\left(2\pi m\frac{x-a}{d}\right)\left(\sqrt{\frac{2}{d}}\sum_m \frac{2\pi m}{d}\left(\alpha_m s_m - \beta_m c_m\right)\right)^2 \\ &= \frac{2}{d}\sqrt{\frac{2}{d}}\int_a^b \cos\left(2\pi m\frac{x-a}{d}\right)\left(\sum_m \gamma_m \cos\left(2\pi m\frac{x-a}{d}\right) + \eta_m \sin\left(2\pi m\frac{x-a}{d}\right)\right)^2 \end{aligned}$$

where

$$\begin{aligned} \gamma_m &\triangleq -\frac{2\pi m}{d}\left(\alpha_m \sin\left(2\pi m\frac{\mu}{d}\right) + \beta_m \cos\left(2\pi m\frac{\mu}{d}\right)\right) \\ \eta_m &\triangleq \frac{2\pi m}{d}\left(\alpha_m \cos\left(2\pi m\frac{\mu}{d}\right) - \beta_m \sin\left(2\pi m\frac{\mu}{d}\right)\right) \end{aligned}$$

Computing this integral we find

$$\begin{aligned} \hat{\nu}_n &= \begin{cases} \frac{1}{4}\sqrt{\frac{2}{d}}\left(\gamma_{n/2}^2 - \eta_{n/2}^2 + 2\gamma_n\gamma_{2n} + 2\eta_n\eta_{2n}\right) & \text{if } n \text{ even} \\ \frac{1}{4}\sqrt{\frac{2}{d}}\left(2\gamma_n\gamma_{2n} + 2\eta_n\eta_{2n}\right) & \text{if } n \text{ odd} \end{cases} \\ \hat{\rho}_n &= \begin{cases} \frac{1}{4}\sqrt{\frac{2}{d}}\left(2\gamma_{n/2}\eta_{n/2} - 2\eta_n\gamma_{2n} + 2\gamma_n\eta_{2n}\right) & \text{if } n \text{ even} \\ \frac{1}{4}\sqrt{\frac{2}{d}}\left(-2\eta_n\gamma_{2n} + 2\gamma_n\eta_{2n}\right) & \text{if } n \text{ odd} \end{cases} \end{aligned}$$

Letting

$$\boldsymbol{f}(x) = \bigoplus_n \nu_n \cdot \boldsymbol{\xi}_n(x) \oplus \rho_n \cdot \boldsymbol{\zeta}_n(x)$$

we find that

$$\begin{aligned} \nu_n &= \alpha_n \cos\left(2\pi n\frac{\mu}{d}\right) - \beta_n \sin\left(2\pi n\frac{\mu}{d}\right) + \frac{1}{2}\sigma^2\hat{\nu}_n \\ \rho_n &= \alpha_n \sin\left(2\pi n\frac{\mu}{d}\right) + \beta_n \cos\left(2\pi n\frac{\mu}{d}\right) + \frac{1}{2}\sigma^2\hat{\rho}_n \end{aligned}$$

which is very similar to a previous set of equations, with the new hatted (nonlinear) terms on the right. Notice that as the variance, σ^2, tends to zero, we recover our perfect wave propagation from before where the distance the waveform travels is the mean of the Gaussian, μ. We should also point out that the Fourier coefficients are now coupled in the update equations through the hatted quantities.

We could also consider keeping an additional term in the Taylor expansion of $\boldsymbol{g}(x|r)$ such that

$$\boldsymbol{g}(x|r) \approx \boldsymbol{g}(x|\mu) \oplus (r-\mu)\cdot\boldsymbol{g}'(x|\mu) \oplus \frac{1}{2!}(r-\mu)^2\cdot\boldsymbol{g}''(x|\mu)$$

where

$$\boldsymbol{g}''(x|\mu) = \lim_{h\to 0} \frac{1}{h} \cdot \left(\boldsymbol{g}'(x|r+h) \ominus \boldsymbol{g}'(x|r)\right)\Big|_{r=\mu} = \downarrow_x \exp -\sqrt{\frac{2}{d}} \sum_n \left(\frac{2\pi n}{d}\right)^2 (\alpha_n c_n + \beta_n s_n)$$

In this case we have for $\boldsymbol{f}(x)$ that

$$\boldsymbol{f}(x) = \boldsymbol{g}(x|\mu) \oplus \frac{1}{2}\hat{\sigma}^2 \cdot \boldsymbol{g}'^2(x|\mu) \oplus \int \exp\left(-\frac{(r-\hat{\mu})^2}{2\hat{\sigma}^2}\right) dr$$

where

$$\begin{aligned} \hat{\sigma}^2 &\triangleq \left(\frac{1}{\sigma^2} + \sqrt{\frac{2}{d}} \sum_n \left(\frac{2\pi n}{d}\right)^2 (\alpha_n c_n + \beta_n s_n)\right)^{-1} \\ \hat{\mu} &\triangleq \mu + \hat{\sigma}^2 \sqrt{\frac{2}{d}} \sum_n \frac{2\pi n}{d} (\alpha_n s_n - \beta_n c_n) \end{aligned}$$

Keeping up to terms of order σ^2 and σ^4, we have

$$\boldsymbol{f}(x) = \boldsymbol{g}(x|\mu) \oplus \frac{1}{2}\sigma^2 \cdot \left(\boldsymbol{g}'^2(x|\mu) \oplus \boldsymbol{g}''(x|\mu)\right) \oplus \frac{1}{4}\sigma^4 \left(2 \cdot \boldsymbol{g}'^2(x|\mu) \cdot \boldsymbol{g}'' \ominus \boldsymbol{g}''^2(x\mu)\right)$$

where

$$\begin{aligned} \boldsymbol{g}'^2(x|\mu) \cdot \boldsymbol{g}''(x|\mu) &\triangleq \downarrow_x \exp\left(\sqrt{\frac{2}{d}} \sum_n \frac{2\pi n}{d} (\alpha_n s_n - \beta_n c_n)\right)^2 \left(-\sqrt{\frac{2}{d}} \sum_n \left(\frac{2\pi n}{d}\right)^2 (\alpha_n c_n + \beta_n s_n)\right) \\ \boldsymbol{g}''^2(x|\mu) &\triangleq \downarrow_x \exp\left(-\sqrt{\frac{2}{d}} \sum_n \left(\frac{2\pi n}{d}\right)^2 (\alpha_n c_n + \beta_n s_n)\right)^2 \end{aligned}$$

We see that additional nonlinear terms have appeared and thus these would have to be decomposed onto the basis functions in order to come up with a Fourier coefficient update as before.

If we drop the σ^4 terms then the Fourier updates are

$$\begin{aligned} \nu_n &= \left(1 - \frac{1}{2}\sigma^2 \left(\frac{2\pi n}{d}\right)^2\right) \left(\alpha_n \cos\left(2\pi n \frac{\mu}{d}\right) - \beta_n \sin\left(2\pi n \frac{\mu}{d}\right)\right) + \frac{1}{2}\sigma^2 \hat{\nu}_n \\ \rho_n &= \left(1 - \frac{1}{2}\sigma^2 \left(\frac{2\pi n}{d}\right)^2\right) \left(\alpha_n \sin\left(2\pi n \frac{\mu}{d}\right) + \beta_n \cos\left(2\pi n \frac{\mu}{d}\right)\right) + \frac{1}{2}\sigma^2 \hat{\rho}_n \end{aligned}$$

which differ from the previous update equations by our inclusion of the term involving $\boldsymbol{g}''(x|\mu)$. Figure 3 shows an example of using this update. The input or initial pdf is a Gaussian and which should propagate to the right and degrade. We see that the approximations begin to break down as the variance is increased too far.

The major drawback of this technique is that more and more Fourier modes will become active with multiple applications of the inference. Really we would like something that keeps the number of active Fourier modes constant.

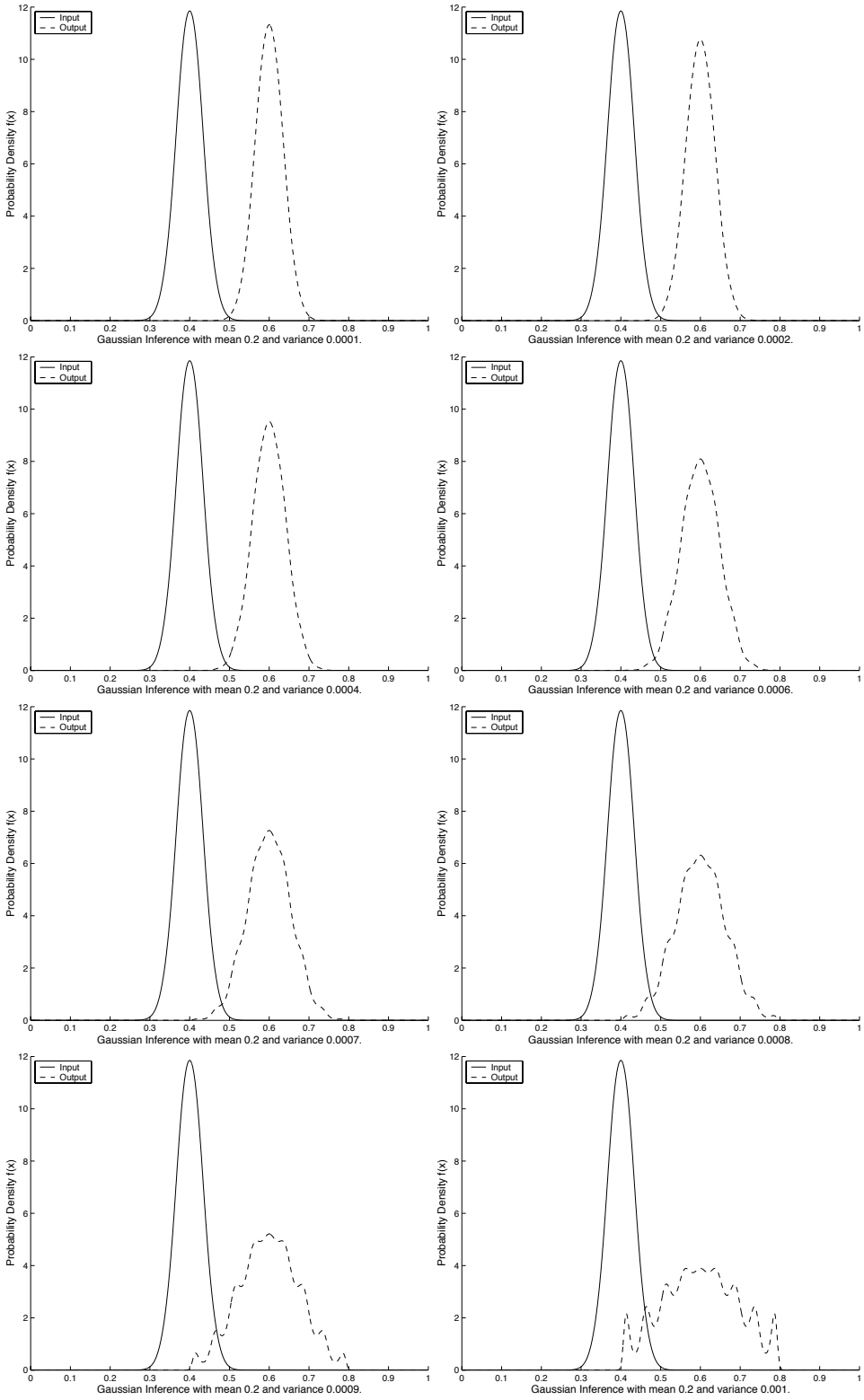

Figure 3: Approximate Gaussian inference with increasing variance.

10 An Example

In this section we consider the specific case that we are working with a mobile robot with a nonholonomic constraint whose state is given by $\mathbf{x} = \begin{bmatrix} x & y & \theta \end{bmatrix}^T$ and whose continuous time kinematics are given by

$$\begin{aligned} \dot{x} &= v\cos\theta \\ \dot{y} &= v\sin\theta \\ \dot{\theta} &= \omega \end{aligned}$$

where v and ω are the translational and rotational speeds, respectively. The robot's change in position, $\mathbf{r}_a$ in the absolute reference frame, $\mathcal{F}_a$, will be given by

$$\mathbf{r}_a = \mathbf{C}_{ab}\mathbf{r}_b = \mathbf{v}h$$

where $\mathbf{r}_b$ is the position change in a body-fixed frame, $\mathcal{F}_b$. The body-fixed frame has the x-coordinate in the direction of travel and the z-coordinate straight up, perpendicular to the xy-plane. The rotation matrix from $\mathcal{F}_b$ to $\mathcal{F}_a$ is denoted $\mathbf{C}_{ab}(\theta)$ and is given by

$$\mathbf{C}_{ab}(\theta) \triangleq \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Integrating the kinematics equations over a short interval reveals

$$\mathbf{r}_b = \begin{bmatrix} v\omega^{-1}\sin\omega h \\ v\omega^{-1}(1-\cos\omega h) \\ \omega h \end{bmatrix}$$

where we have assumed v and ω to be constants in order to perform the integration (reasonable for small h). Defining $\mathbf{K}_b = \mathrm{diag}\{k_x, k_y, k_\theta\}$, we construct the damping coefficient, $k_{\mathbf{n}}$, as

$$k_{\mathbf{n}} = \exp\left(-\frac{|\mathbf{n}^T\mathbf{C}_{ab}(\theta)\mathbf{K}_b\mathbf{r}_b|}{\mathbf{n}^T\mathbf{n}}\right)$$

where we note that

$$\mathbf{n}^T\mathbf{C}_{ab}(\theta)\mathbf{K}_b\mathbf{r}_b = \mathbf{n}^T\mathbf{K}_a\mathbf{r}_a$$

with $\mathbf{K}_a = \mathbf{C}_{ab}(\theta)\mathbf{K}_b\mathbf{C}_{ab}^T(\theta)$.

To use this in practice, we assume that v, ω, and t can be measured from, for example, odometry and a clock. This allows us to determine $\mathbf{r}_b$. The determination of $\mathbf{C}_{ab}(\theta)$ requires the knowledge of θ, which we do not have, directly. Instead we must use an estimate. This brings up the next issue, which is what to do in general when we do not have perfect knowledge of $\mathbf{v}$.

References

[1] K J Åstrom. Optimal control of markov processes with incomplete state information. *Journal of Mathematical Analysis and Applications*, 10:174–205, 1965.

[2] T D Barfoot. *Stochastic Decentralized Systems*. PhD thesis, Institute for Aerospace Studies, University of Toronto, 2002.

[3] T D Barfoot and G M T D'Eleuterio. An algebra for the analysis of stochastic systems. Submitted to *Linear Algebra and Its Applications*, 2002.

[4] T D Barfoot and G M T D'Eleuterio. An algebra for the control of stochastic systems: Exercises in linear algebra. Cambridge, UK, July 14-18 2002. The Fifth International Conference on Dynamics and Control of Structures in Space.

[5] Thomas Bayes. Essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 1764.

[6] J B J Fourier. *Théorie Analytique de la Chaleur*. 1822.

[7] Christian Genest and James V Zidek. Combining probability distributions: A critique and an annotated bibliography. *Statistical Science*, 1(1):114–148, 1986.

[8] Werner Greub. *Linear Algebra*. Springer-Verlag, New York, 1974. Graduate Texts in Mathematics, 23.

[9] J R Higgins. *Completeness and Basis Properties of Sets of Special Functions*. Cambridge University Press, 1977.

[10] A M Jazwinsky. *Stochastic Processes and Filtering Theory*. Academic, New York, 1970.

[11] R E Kalman. A new approach to linear filtering and prediction problems. *Trans. ASME, Journal of Basic Engineering*, 82:35–45, 1960.

[12] S Kullback. *Information Theory and Statistics*. Wiley, New York, 1959.

[13] Andrei Andreevich Markov. Essai d'une recherche statistique sur le texte du roman 'eugene onegin' illustrant la liaison des epreuve en chain. *Izvestia Imperatorskoi Akademii Nauk (Bulletin de l'Academie Imperiale des Sciences de St-Petersbourg)*, 7:153–162, 1913.

[14] John F Nash. The bargaining problem. *Econometrica*, 18:155–162, 1950.

[15] Athanasios Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill Book Companay, New York, 1965.

[16] Giuseppe Peano. *Geometrical Calculus*. 1888.

[17] C E Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, July 1948.