Deriving Sensor Models and Non-Linear Filtering for Exponentials of Polynomials

Bradley Tonkes and Alan D. Blair
School of Computer Science and Engineering
University of New South Wales, Australia
btonkes@gmail.com, blair@cse.unsw.edu.au

Abstract

We consider the family of probability density functions given by $e^{-P(x)}$ for elliptic, multivariate polynomials, $P(x)$ in a Bayesian estimation context. Integration of new sensor information with this class is trivial (addition of the polynomial coefficients), and previous work has proposed methods for performing prediction and fusion (covariance intersect). In this paper we consider methods for constructing sensor likelihood models for this class, and demonstrate estimation where the process model is non-linear.

1 Introduction

The Kalman filter solves, in a Bayes-optimal sense, the linear estimation problem for Gaussian uncertainties and is consequently widely used in robotics applications including tracking [Bar-Shalom and Li, 1995], simultaneous localisation and mapping [Smith *et al.*, 1990] and distributed data fusion [Sukkarieh *et al.*, 2003]. Central to the Kalman filter (and many other methods) is the premise that uncertain quantities are Gaussian in nature; that is, their uncertainty can be characterised by their mean and covariance. While the central limit theorem provides a compelling justification for this assumption, non-linear transformations between spaces can produce highly non-Gaussian distributions. For example, a range-only sensor may produce Gaussian noise in terms of the measured range, but the uncertainty in terms of the location of the target in a plane is non-Gaussian. How to efficiently perform non-linear Bayesian estimation remains an open problem.

The efficiency of a particular method is closely tied to the application. This paper considers non-linear estimation in the context of distributed data fusion (DDF). One of the key bottlenecks in DDF is the limited bandwidth available between nodes: it is imperative that nodes can share relatively compact representations of probability densities. Numerous approaches have been made to this problem, each involving different trade-offs in optimality and efficiency. Most recent have been those based on agglomerations of simpler functions such as mixtures of Gaussians [Upcroft *et al.*, 2004] and Parzen density estimators [Ridley *et al.*, 2004]. Closely related are particle filter methods [Gordon *et al.*, 1993] which represent a density as a collection of samples.

The tractability of these methods relies on the fact that operations on densities can be performed efficiently on the components. The downside is that some functions operate pair-wise on the component distributions, leading to an $O(n^2)$ number of components in the resulting distribution, thus requiring resampling [Upcroft *et al.*, 2004].

Other representations have been used for non-linear filtering, although not for DDF including Dirac mixtures [Schrempf *et al.*, 2006] and Fourier densities [Brunn *et al.*, 2006]. Grid-based approaches have also been applied to DDF [Bourgault *et al.*, 2003]. The approach that we consider is the use of a wholly distinct (rather than agglomerated) family of parameterised distributions. The representation we have considered in previous work is that of an exponentiated polynomial, $e^{-P(x)}$ [Blair and Tonkes, 2007; Tonkes and Blair, 2007].

This family is capable of compactly representing complex density functions. For application to Bayesian estimation, three fundamental operations are required for densities: multiplication, convolution and (typically linear) transformation. For our class of densities, multiplication is trivial; convolution with a Gaussian can be efficiently and accurately approximated; and transformation of densities takes place by way of the polynomial (e.g., the class is closed under linear transformation). We have also demonstrated how densities can be conservatively combined in a DDF scenario through an extension of the covariance intersect algorithm.

In this paper we consider two related problems: that of constructing sensor likelihood models and that of non-linear filtering. For the first problem, we propose a geometrically-inspired algorithm for fitting an expo-

nentiated polynomial to an arbitrary density function. The extended Kalman filter attacks the problem of non-linearities by taking a linearisation of the system. We propose, for our density family, instead taking an nth order polynomial expansion, then simplifying the resultant complex density. Finally, we present a contrived demonstration of non-linear Bayesian filtering with this representation.

2 ExPoly Basics

We consider the family of probability density functions of the form $\exp(-P(x))$ where $P(x)$ is a multivariate, elliptic polynomial. (By elliptic we mean that it tends to positive infinity in all directions so that $\exp(-P(x))$ has finite volume.) For convenience, we call this class ExPoly. Since $P(x)$ is a function of arbitrary dimensionality we shall adopt the multi-index notation $[i] = (i_1, ..., i_n)$ so that $x^{[i]} = \prod_{d=1}^{n} x_d^{i_d}$ and

$$P(x) = \sum_{|[i]| \leq l} a_{[i]} x^{[i]},$$

where $|[i]| = \sum_{d=1}^{n} i_d$. Clearly $P(x)$ must be of even order and we shall describe a density as being lth order when no term of $P(x)$ is of mixed order greater than l (i.e., $x_1^a x_2^b$ is an $a + b$ order term). We consider both valid densities, where $\int \exp(-P(x))\ dx = 1$, as well as the denormalised space; normalisation can be achieved by adding $\log \int \exp(-P(x))\ dx$ to the constant term of P, $a_{[0]}$. This class subsumes Gaussian density functions (second-order polynomials) and allows for a much richer variety of forms (see Figure 1).

Any exponential co-ordinate system naturally determines a dual co-ordinate system in the form of *expectation* parameters [Amari, 1993]. For example, while the Kalman filter operates on the expectation parameters – the mean and the covariance matrix – the information filter form operates on the coefficients of the polynomial underlying the distribution. In the case of Gaussian functions the translation between these co-ordinate systems is relatively straightforward, but with ExPoly functions the relationship is less clear. A Gaussian is uniquely characterised by its mean, $\mu = \mathrm{E}[x]$ and covariance, $\Sigma = \mathrm{E}[(x - \mu)^2]$. Similarly, an lth-order ExPoly density can be uniquely characterised by the first l expectations. Rather than considering the *central* moments of the density we take the moments around zero,

$$M_{[i]} = \mathrm{E}[x^{[i]}] = \int x^{[i]} \exp(-P(x))\ dx.$$

Among all densities of a given mean and covariance, the corresponding Gaussian maximises the entropy. Likewise, for a given set of (algebraic) moments, $M_{[i]}$, the density that maximises the entropy takes the form of an ExPoly with terms given by the specified moments. That is, given moments $M_{[i_1]}, \ldots, M_{[i_n]}$, the density with maximum entropy is of the form $\exp(\theta_1 x^{[i_1]} + \ldots + \theta_n x^{[i_n]})$ [Cover and Thomas, 1991] although determination of θ is non-trivial (see Section 5). Thus, ExPoly densities can be regarded as the most conservative (least informative) choice of representation for a given set of expectation parameters.

3 Linear Filtering

Bayesian filtering is governed by a recursive instantiation of Bayes' law:

$$\mathbf{P}(x_k|Z^k) = \frac{\mathbf{P}(z_k|x_k)\ \mathbf{P}(x_k|Z^{k-1})}{\mathbf{P}(z_k|Z^{k-1})} \tag{1}$$

where z_k and x_k are the observation and estimated state respectively at time t_k, and $Z^k = \{z_k, Z^{k-1}\}$. Consequently, two operations are necessary: prediction (computing $\mathbf{P}(x_k|Z^{k-1})$), and update (multiplying the distributions, normalising if required). Assuming that the state is governed by a Markovian process, its probabilistic evolution is described by the Chapman-Kolmogorov equation,

$$\mathbf{P}(x_k|Z^{k-1}) = \int \mathbf{P}(x_k|x_{k-1})\ \mathbf{P}(x_{k-1}|Z^{k-1}, x_0)\ dx_{k-1}. \tag{2}$$

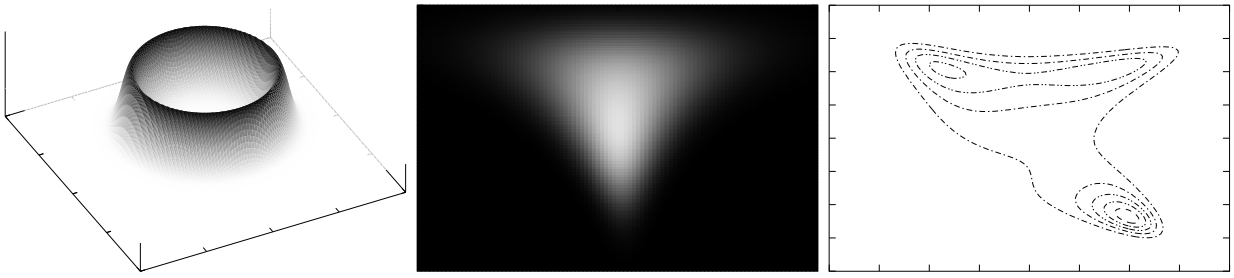
This equation describes a convolution between two density functions, the previous estimate $\mathbf{P}(x_{k-1}|Z^{k-1})$, and the model of the system evolution, $\mathbf{P}(x_k|x_{k-1})$. Assuming a standard linear model of the system, $x_{k+1} = Fx_k + Bu_{k+1} + Gv_k$ where the state estimate, x_k, and noise v_k are Gaussian, the Chapman-Kolmogorov equation describes a convolution of two Gaussian densities which is itself Gaussian.

For ExPoly densities using the same linear model, the Chapman-Kolmogorov equation requires convolution between an ExPoly density and a Gaussian density. ExPoly is not closed under this operation so we have consequently proposed an approximation [Blair and Tonkes, 2007]. The approach to this approximation is to consider the differential rather than the integral form. That is, when one of the densities is Gaussian the solution to the convolution equation is equivalent to the solution of the heat equation described by the Laplacian operator applied to the density over the state estimate.

In the multi-dimensional case the Laplacian operator takes the form

$$\Delta = -\sum_{i,j} D_{ij} \frac{\partial^2}{\partial x_i \partial x_j}$$

for a diffusion matrix, D. The key to the approximation is diagonalising Δ by changing co-ordinates to



(a) (b) (c)

Figure 1: ExPoly density functions of two variables. Density (a) is intended as a range-only sensor model, while (b) is a bearing-only model. Complex, multi-modal densities (c) are easily produced with the maximum modality determined by the order of the polynomial. For these figures, (a) is fourth order and (b) and (c) are sixth order.

$w = V\Sigma^{-\frac{1}{2}}(x-\mu)$ where V diagonalises $\Sigma^{-\frac{1}{2}}D\Sigma^{-\frac{1}{2}}$; i.e., $V\Sigma^{-\frac{1}{2}}D\Sigma^{-\frac{1}{2}}V^{-1} = \Lambda$ where

$$\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}, \qquad \text{with} \quad 0 < \lambda_1 \leq \ldots \leq \lambda_n$$

is diagonal. The transformed density (in w) can then be diffused according to Λ. The change of co-ordinates involves a straightforward linear transformation of the polynomial, $P(x)$.

For a given transformed ExPoly density, $\exp(-P(w))$ we decompose it into Gaussian (quadratic) and residual terms, $P(w) = Q(w) + R(w)$ and consider the diffusion separately. For the quadratic component, the diffusion can be computed exactly, while the residual requires an approximation that is derived through analogy to the quadratic case. In terms of w, our proposed diffusion process is given by

$$P_t(w) = \frac{1}{2}w^T(I + t\Lambda)^{-1}w + \frac{R(I + t\Lambda)^{-\frac{1}{2}}w}{1 + t\lambda_n^{\eta}} \qquad (3)$$

where t is the diffusion time and η is a free parameter which experimentation suggests should be approximately $2 + \frac{l}{2}$ for an lth order density. Equivalently, we can set t to 1 and pre-multiply D by t (i.e., use $t = 1, D = Gv_k$).

The prediction step, to compute $\mathbf{P}(x_k|Z^{k-1})$ is thus achieved by

1. transforming $\mathbf{P}(x_{k-1}|Z^{k-1})$ to $\mathbf{P}'(x_k|Z^{k-1})$ according to $x_k = Fx_{k-1} + Bu_k$
2. transforming $\mathbf{P}'(x_k|Z^{k-1})$ to $\mathbf{P}'(w_k|Z^{k-1})$ according to $w = V\Sigma^{-\frac{1}{2}}(x-\mu)$ where μ and Σ are the mean and covariance respectively of $\mathbf{P}'(x_k|Z^{k-1})$
3. diffusing $\mathbf{P}'(w_k|Z^{k-1})$ using Eqn. (3) to $\mathbf{P}(w_k|Z^{k-1})$ using $D = Gv$ and computing Λ accordingly
4. transforming $\mathbf{P}(w_k|Z^{k-1})$ by $x = \Sigma^{\frac{1}{2}}V^{-1}w + \mu$ (i.e., the inverse transform of step 2) to $\mathbf{P}(x_k|Z^{k-1})$.

Previous analysis [Blair and Tonkes, 2007] has shown that this approximation is highly accurate for densities that are Gaussian-like in nature and remains quite accurate for the great majority of distributions before breaking down for boundary cases (e.g., extremely peaked multimodal densities).

The other step required under Bayes' law (Eqn 1) — update — involves multiplication of the predicted density, $\mathbf{P}(x_k|Z^{k-1})$, with the sensor likelihood, $\mathbf{P}(z_k|x_k)$, and normalisation. For ExPoly densities, this operation is a trivial addition of the polynomial coefficients, disregarding the normalisation step.

As stated earlier, for ExPoly densities there is no simple relationship between coefficient and expectation parameters. However, for the prediction step knowledge of both the mean and covariance is required. For many operations, given an estimate of these parameters a new one can be derived. Specifically, after step (1), $\mu' = F\mu + Bu_k$ and $\Sigma' = F^{-1}\Sigma F^{T^{-1}}$; and after step (4), $\mu'' \approx \mu$ and $\Sigma'' \approx \Sigma' + Gv$. More difficult is the problem of tracking the mean and covariance through a sensor update (multiplication of densities). For purely Gaussian densities the solution has a closed form, $\Sigma^{-1} = \Sigma_a^{-1} + \Sigma_b^{-1}$. For ExPoly functions, we use the Gaussian solution as an initial estimate then refine it using a Metropolis sampler [Hastings, 1970].

Ensuring efficiency of the sampler is important since many samples are required for accurate estimates of the expectations, making it the most expensive operation required for linear filtering. Experiments suggest that the diffusion approximation – which requires a transformation of the density based on the mean and covariance – is relatively stable with respect to small errors in the estimates. Efficiency concerns also recommend the use of a Horner scheme [Pena, 2000] for evaluation of polynomials.

4 Sensor Fusion

The intended application for ExPoly filtering is in distributed data fusion tasks. The benefit here is that ExPoly densities have a relatively compact representation and can be communicated efficiently between nodes where bandwidth is limited. While it is not the aim of this paper to explore distributed scenarios it is worthwhile reviewing briefly the application of ExPoly to this domain from previous work [Tonkes and Blair, 2007].

Data fusion between independent nodes can occur in two modes: *distributed* data fusion, where the communication network between the nodes is centrally controlled and each node performs a specific role with respect to that architecture; and *decentralised* data fusion where the overall topology of the network is unknown to any of the individual nodes and may in fact change over the lifetime of the system. This difference is significant when considering how information flows through a network.

Given two independent estimates of a state, $\mathbf{P}(x|Z_a)$ and $\mathbf{P}(x|Z_b)$, a combined estimate can be made through simple multiplication. However, if the estimates are not independent then a simple combination will result in a skewed estimate. In a distributed scenario the topology of the network can be controlled so that each node can determine how to factor out the information which it itself has provided. This effect is achieved through the use of a *channel filter* which keeps track of the mutual information between a node and its neighbours, and removes it accordingly:

$$\mathbf{P}(x|Z) = \frac{\mathbf{P}(x|Z_a)\mathbf{P}(x|Z_b)}{\mathbf{P}(x|Z_a \cap Z_b)}.$$

For ExPoly densities this operation is straightforward since division and multiplication involve subtraction and addition of polynomial coefficients respectively.

The case of decentralised data fusion is more interesting. Here it is impossible to maintain the shared information between two nodes since information can be passed between them via an unknown, indirect route. A solution for Gaussian densities is the covariance intersect (CI) algorithm [Julier and Uhlmann, 1997b]. The aim of CI is to find the most informative density that is consistent with all possible cross-correlations of the two source densities. That is, regardless of the common information used to produce the two densities, CI finds a density that represents a consistent combination. Formally, CI combines two densities with a parameter, $\omega \in [0, 1]$, such that

$$\Sigma^{-1} = \omega\Sigma_a^{-1} + (1-\omega)\Sigma_b^{-1},$$
$$\Sigma^{-1}\mu = \omega\Sigma_a^{-1}\mu_a + (1-\omega)\Sigma_b^{-1}\mu_b$$

where ω is chosen to optimise some property of Σ such as minimisation of its trace.

In terms of polynomial coefficients, CI is remarkably simple. In this space, CI becomes

$$\exp(-Q(x)) = \exp(-(\omega Q_a(x) + (1-\omega)Q_b(x))).$$

This form of CI suggests an immediate extension to ExPoly densities which has an information theoretic justification [Hurley, 2002].

5 Deriving Sensor Models

The task of building sensor likelihood models, $P(z|x)$, is an arduous one involving collection of large amounts of sensor data and ground truth measurements. For Gaussian models it is relatively straightforward to construct a density which captures the measured data, but for ExPoly densities it is more complex. There are two possible needs for non-Gaussian models: (i) the sensor inherently exhibits a non-Gaussian pattern of errors, (ii) the sensor is subject to Gaussian errors, but the measurement itself is a non-linear function of the state. For the Kalman filter the first problem is typically solved by conservatively fitting a Gaussian model to the data. The second problem is solved by linearising the observation function.

For ExPoly densities, the issue of a non-linear relationship between state and observation can be handled analogously to the Kalman filter. However, instead of *linearising* the system (which maintains the errors in the form of a Gaussian) we can instead consider a *polynomial* expansion of the system. That is, where the Kalman filter limits the Taylor series expansion of the function to first order, ExPoly densities can use an expansion of any order (although with increasing computational cost). We shall discuss this idea further in the next section in terms of non-linear state transitions.

For the other problem, that of fitting an ExPoly distribution to intrinsically non-linear sensor data, the typical approach is to consider the moments. The problem of determining a density (or distribution) from its moments has a long history and is an intensively studied problem, at least for univariate functions [Landau, 1987].

Our initial approach to the problem for multivariate ExPoly densities was to use the Fisher information matrix to provide a locally approximate map between coefficients and expectations. If the points $M = M_{[i]}$ in expectation parameters corresponds to $a = a_{[i]}$ in coefficient parameters, then the transformation between the two co-ordinate systems can be locally approximated by the linear map:

$$M + \delta M \;\mapsto\; a + \mathrm{G}^{-1}\delta M \tag{4}$$

where G^{-1} is the inverse of the Fisher information matrix G. The Fisher metric for an ExPoly density,

$\exp(-P(x))$, is given by

$$\begin{aligned} g(\tfrac{\partial}{\partial a_{[i]}}, \tfrac{\partial}{\partial a_{[j]}}) &= \int \tfrac{\partial P(x)}{\partial a_{[i]}} \tfrac{\partial P(x)}{\partial a_{[j]}} e^{-P(x)} dx \\ &= \int x^{[i]} x^{[j]} e^{-P(x)} dx = M_{[i+j]}. \end{aligned}$$

In principle this technique defines an iterative procedure for finding a set of coefficients that define a distribution having the given expectation parameters. In practice we find that the process is numerically unstable, leading to non-elliptic polynomials. It is also computationally very expensive, requiring evaluation of many moments for each iteration.

An alternative approach is to consider the moments on a bounded subspace, typically, $[0,1]^n$. One heuristic method in this vein assumes that the moments come from an ExPoly function and then solves the problem directly [Borwein and Huang, 1995]. Given the constraints,

$$M_{[k]} = \int_0^1 x^{[k]} \exp(\sum_{[i]} a_{[i]} x^{[k]})$$

for $|[k]| \leq 2l$, finding the coefficients, $a_{[i]}$ reduces, for univariate functions, to solving $b = Br$ for r, where

$$B = \begin{bmatrix} 1 & M_1 & M_2 & \dots & M_l \\ 1 & M_2 & M_3 & \dots & M_{l+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & M_{n+1} & M_{n+2} & \dots & M_{2l} \end{bmatrix}, \; b = \begin{bmatrix} M_1 \\ 2M_2 \\ \vdots \\ (l+1)M_l \end{bmatrix},$$

$$r = [r_0 \; r_1 \; \dots \; r_l]^T$$

with $r_0 = \exp(\sum a_i)$ and $r_k = k a_k$ for $k > 0$. The multivariate case is somewhat more involved, requiring the solution of n sets of linear equations for a density of n variables.

Should the moments indeed come from an lth order ExPoly density, then this procedure is guaranteed to reconstruct it. When the moments do not come from an ExPoly density it is important that the function be effectively confined to $[0,1]^n$. If the moments derive from a density which is non-trivially supported outside this range then the procedure provides no guarantees on the extrapolated moments and typically performs poorly.

Our second approach has been to consider the problem as a purely geometrical one. Rather than using the moments as the starting point, we assume that we are given a function, $f(x)$, that we can evaluate for arbitrary x. This conceptualisation suggests fitting $-\log f(x)$ with a polynomial, $P(x)$, in a least-squares sense. In this logarithmic space, small values of $f(x)$ are overweighted so we instead consider a weighted-least-squares fit, using a weighting of $f(x)$. The problem now reduces to solving a linear equation, $Ga = m$ for a, with

$$G_{[\bar{i}],[\bar{j}]} = \sum_{m=1}^{N} f(x) x_m^{[i]+[j]}, \text{and}$$

$$m_{[\bar{i}]} = \sum_{m=1}^{N} -f(x) x_m^{[i]} \log f(x)$$

where $[\bar{i}]$ denotes an arbitrary total ordering of the multi-indices, and $x_m, m = 0, \dots, N$ are samples drawn uniformly from the support of $f(x)$. Alternatively, we can draw samples from the distribution, $f(x)$, using a Metropolis sampler, then weight them equally using the sampling frequency as a substitute for the weight. Simulations suggest than uniform sampling is marginally better. Again, if f is an lth order ExPoly, then this method will reproduce it exactly.

Note that, assuming f is an ExPoly density, G is effectively its Fisher information matrix since its entries reflect the moments of f (consider integration instead of sampling). Furthermore, the second method involves a similar quantity in the form of the matrix, B. There are clearly some deep parallels between all three methods but we are unsure of their nature.

Unlike the moment-based approaches, where $M_{[0]}$ is an explicit constraint, the geometrical solution provides no guarantees on the volume of the fitted density. In fact, the geometrical approach sometimes finds the best fit to be one which is non-elliptic. This error is particularly apparent when $f(x)$ vanishes very quickly so that, where the density is almost zero, its weighting ($f(x)$) is so small that samples from that region are effectively ignored.

One solution to this problem is to fit a polynomial of degree $l+2$ where the $(l+2)$th order terms are given by $\sum_{d=1}^{n} \epsilon x_d^{l+2}$ (i.e., coefficients for terms of order $(l+2)$ are ϵ when the exponent is of the form $(0, \dots, l+2, \dots, 0)$). This constraint ensures that the polynomial is elliptic. The linear problem then becomes $Ga + e = m$ where e describes the regularising components given by ϵ. An alternative approach is to introduce a minimum value on the size of the weights.

6 Non-Linear Filtering

Given the preceding sections, we are now in a position to consider non-linear filtering with ExPoly densities. Sources of non-linearity arise in two places: (i) transforming between sensor and state spaces; and (ii) in the evolution of the state. The extended Kalman filter (EKF) linearises these transformations (about the expected value) making it once again a linear problem. An alternative approach is taken by the unscented Kalman filter (UKF) [Julier and Uhlmann, 1997a] which applies the transformation to specifically chosen points ('Sigma points'), reconstructing the Gaussian by the transformed

values. For ExPoly densities it is unclear how to perform the same style of computation as the UKF, but it is obvious that the techniques of the EKF are immediately applicable. We also propose an alternative solution.

Non-linearities in the sensor model have already been addressed in a previous section. Non-linearities in the state propagation can be treated similarly. Given a state model, $x_{k+1} = f(x_k, u_{k+1}) + Gv_{k+1}$, we can take a Taylor series expansion around the expected state, $\mathrm{E}[x_k]$, to nth order, resulting in a polynomial state transition. With an lth order density over the state, $\exp(-P(x))$, this polynomialisation produces an nlth order ExPoly density, $\exp(-P'(x))$. Since the state transition model is applied recursively, it is obvious that the order of the state density will grow exponentially. It will therefore become necessary to periodically 'downsample' the density to lower order. This simplification can be achieved with our geometric solution outlined in section 5.

The procedure for prediction, i.e., computing $\mathbf{P}(x_k|Z^{k_1})$ now becomes

1. transforming $\mathbf{P}(x_{k-1}|Z^{k-1})$ to $\mathbf{P}'(x_k|Z^{k-1})$ according to $(x_i)_k = p(x_{k-1})$ with p being the nth order polynomial expansion of the state transition function, f, around $\mathrm{E}[x_{k-1}]$ with respect to x_i.
2. transforming $\mathbf{P}'(x_k|Z^{k-1})$ to $\mathbf{P}'(w_k|Z^{k-1})$ according to $w = V\Sigma^{-\frac{1}{2}}(x - \mu)$ where μ and Σ are the mean and covariance respectively of $\mathbf{P}'(x_k|Z^{k-1})$
3. diffusing $\mathbf{P}'(w_k|Z^{k-1})$ using Eqn. (3) to $\mathbf{P}(w_k|Z^{k-1})$ using $D = Gv$ and computing Λ accordingly
4. transforming $\mathbf{P}(w_k|Z^{k-1})$ by $x = \Sigma^{\frac{1}{2}}V^{-1}w + \mu$ to $\mathbf{P}(x_k|Z^{k-1})$.

In this case tracking the expectation parameters through step (1) becomes more difficult than in the linear case. Indeed, since relatively accurate estimates of these quantities are required for the diffusion approximation in step (2), it is necessary to re-estimate them numerically after step (1). Again we use linear and Gaussian assumptions to provide an initial configuration for a Metropolis sampler. Since step (4) is only a linear transformation, the estimate update remains the same as for the corresponding step for the linear filter.

Since polynomial transformations rapidly increase the order of densities it will be prudent to downsample them frequently. An obvious choice is to simplify the function after the prediction (after step 4). Another alternative would be to simplify after step (3) since mean and covariance need to be re-estimated at that point anyway.

There remains an additional potential source of non-linearity in the state transition: that of error in the control. This noise source is generally assumed to be additive with the state noise and incorporated into a single Gaussian error term. If this is not the case, then the Chapman-Kolmogorov equation (Eqn 2) describes a convolution between an ExPoly and a non-Gaussian density. Since our approximation technique relies on treating the problem as a Laplacian, it breaks down for non-Gaussian functions. Consequently, for ExPoly filters, these noise sources must be treated as linearised Gaussian noise in the process model.

6.1 Non-Linear Example

We consider the problem of tracking a target moving according to a tricycle motion model

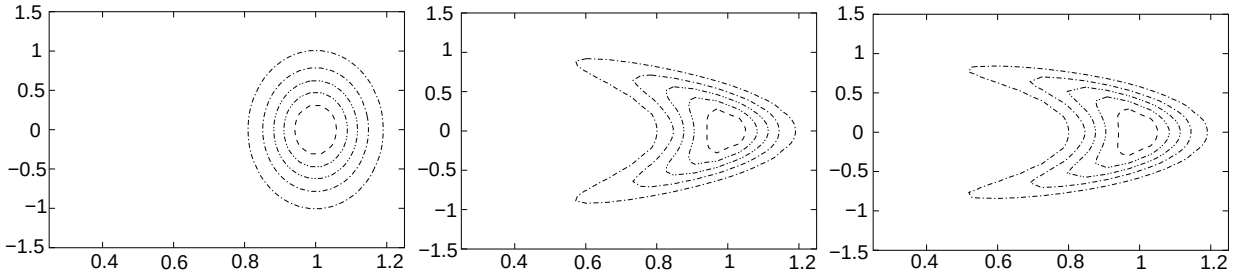
$$\begin{aligned} x_{k+1} &= x_k + v_{k+1}\Delta_T \cos(\phi_k + \psi_{k+1}) + q_x \\ y_{k+1} &= y_k + v_{k+1}\Delta_T \sin(\phi_k + \psi_{k+1}) + q_y \\ \phi_{k+1} &= \phi_k + \Delta_T \frac{v_{k+1}}{B} \sin\psi_k + q_\phi \end{aligned}$$

where velocity, v, and steering angle ψ are control variables, B is the distance between the rear wheels, and Δ_T is the time between updates. The state is described in terms of position, (x, y), and bearing, ϕ. We make a simplifying assumption that each state variable experiences independent Gaussian noise, q_x, q_y and q_θ which have variances Q_x, Q_y and Q_θ respectively.

This motion model is well-known for producing 'banana' densities which aren't adequately expressed by Gaussian models. Figure 2 shows densities that result from first (linear), second and third order polynomial approximations of the state transition given a Gaussian prior. It is obvious that as the accuracy of the approximation increases, new parts of the space provide non-trivial support to the density.

To highlight the advantages of our non-linear approach we consider a highly contrived situation where $Q_x = Q_y = Q_\theta = 0$, $B = 0.1$ and $\phi = 0.0$ and $\Delta_T = v = 1.0$ for all steps. While we could have used an ExPoly range-only likelihood, for simplicity we assume the sensor gives an estimate of the position (but not the heading) with Gaussian uncertainty having standard deviation in each direction of 0.5. The initial state of the target is $(0, 0, \frac{\pi}{4})$ and the initial estimate has mean $(0, 0, 0)$ and a diagonal covariance with terms $(0.01, 0.01, \frac{\pi}{12})$ (i.e., the target heading is unlikely given the prior).

We consider both linear and quadratic approximations of the process model for five time steps. The linear approximation mirrors an EKF since both initial estimate and sensor likelihood are Gaussian. The quadratic approximation is simplified to sixth order in each prediction step. Figure 3 shows the estimate after prediction and observation for (a) the linear approximation and (b) the quadratic approximation. This example clearly shows the linear version converges poorly due to the inaccurate initial estimate of the heading, while the quadratic approximation recovers far more quickly. Figure 2 shows



(a) Linear (b) Quadratic (c) Cubic

Figure 2: (a) First, (b) second and (c) third order approximations of the tricycle model state transition, applied to a Gaussian prior, and marginalised over the bearing, ϕ. For these figures, $v = 1$, $\Delta_t = 1$, $B = 0.1$, $\psi = 0$ and the prior density has mean 0 and diagonal covariance with terms $(0.01, 0.01, 0.274)$.

the likely source of the benefit: the quadratic approximation better supports an unlikely heading.

7 Discussion and Conclusion

In this paper we have considered two problems: fitting an ExPoly given certain constraints, and non-linear filtering with ExPoly densities. Fitting ExPoly densities was considered in two (successful) ways: fitting a given set of moments, and fitting a sample of points. Although the motivations behind these approaches differ, the end results are quite similar, solution of a set of linear equations with a matrix involving (effectively) the moments up to twice the order of the ExPoly density. Preliminary simulations suggest that the geometrical approach outperforms the moment-based approach for higher-order densities. The geometrical approach also has the advantage of not requiring transformation down to $[0, 1]^n$.

Non-linear filtering with ExPoly can be treated identically to the EKF (via linearisation). A more accurate solution can be achieved by considering a polynomial (Taylor series) expansion around the expectation. While ExPoly is closed under polynomial transformation, it is not true for ExPoly densities of a fixed order. Since iteratively applying polynomial transformations will exponentially increase the order of the ExPoly, it is important to downsample the density to lower order. Either method for fitting densities could be used, but we have chosen to use the geometrical solution since it bypasses a linear transformation (to $[0, 1]^n$) and appears to work better.

Relative to the EKF, estimation with ExPoly densities is computationally expensive. There are two parts in particular that impose a high burden: estimation of the moments up to $2l$ for downsampling the density, and estimation of the mean and covariance for the diffusion (prediction step). These steps may be combined, but filtering still requires evaluation of the polynomial many times. Some strategies exist for improving polynomial evaluation [Roy and Minocha, 1991; Pena, 2000] but these are likely to provide only a small benefit. The well-studied convergence properties of the Metropolis sampler provide the most important guide to the efficiency of filtering. Speed of convergence is highly dependent on the sampled density: for unimodal densities samplers tend to 'mix' well and converge rapidly, while samplers take much longer to traverse sparse peaks in a multimodal density (these can occur frequently in, for example, range-only scenarios). A better solution lies in being able to approximate the moments directly as the filter progresses, and a solution for the univariate case has been proposed [Rauh and Hanebeck, 2005]. Since the number of coefficients is $\binom{n+l}{l}$, ExPoly is practically limited to a handful of variables of moderate order.

We have presented ExPoly as a candidate for representing non-Gaussian probability densities with specific application to DDF. All operations that are necessary for supporting (non-linear) DDF have been proposed in this and previous papers, although some aspects lack formal proof. We propose that ExPoly can be added to the set of available non-Gaussian representations, to be used under appropriate circumstances.

References

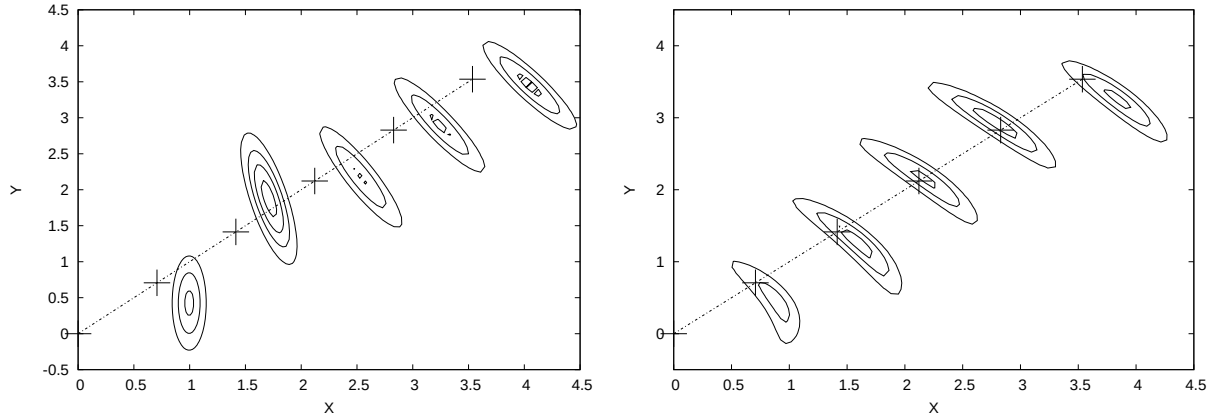
[Amari, 1993] Shun'ichi Amari. *Methods of Information Geometry*. Oxford University Press, 1993.

[Bar-Shalom and Li, 1995] Y. Bar-Shalom and X. Li. *Multisensor target tracking: Principles and techniques*. YBT Publishing, 1995.

[Blair and Tonkes, 2007] A. D. Blair and B. Tonkes. Geometry and approximate diffusion for exponentials of polynomials. Submitted, 2007.

[Borwein and Huang, 1995] J. Borwein and W. Huang. A fast heuristic algorithm for polynomial moment problems with Boltzmann-Shannnon entropy. *SIAM Journal of Optimization*, 5(1):68–99, 1995.

[Bourgault *et al.*, 2003] F. Bourgault, T. Furukawa, and H. F. Durrant-Whyte. Coordinated decentralized search for a lost target in a Bayesian world. In *Proc. of the 2003*



(a) (b)

Figure 3: ExPoly filters for a tricycle model target. (a) Shows the estimate of the linear filter after each iteration given the observation shown, while (b) shows the same data for the quadratic approximation. Densities are shown over the estimated position of the target, marginalising the heading.

IEEE/RSJ Intl Conf. on Intelligent Robots and Systems, pages 48–53, 2003.

[Brunn *et al.*, 2006] D. Brunn, F. Sawo, and U. D. Hanebeck. Nonlinear multidimensional Bayesian estimation with fourier densities. In *IEEE Conf. on Decision and Control*, 2006.

[Cover and Thomas, 1991] T. A. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley and Sons, 1991.

[Gordon *et al.*, 1993] N. Gordon, D. Salmond, and A. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *IEE Proceedings-F*, volume 140, pages 107–113, 1993.

[Hastings, 1970] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1), 1970.

[Hurley, 2002] M. B. Hurley. An information theoretic justification for covariance intersection and its generalisation. In *Proc. of the Int'l Symposium on Information Fusion*, 2002.

[Julier and Uhlmann, 1997a] S. J. Julier and J. K. Uhlmann. A new extension of the Kalman filter to nonlinear systems. In *The Proc. of AeroSense: The 11th Int'l Symposium on Aerospace/Defense Sensing, Simulation and Controls*, 1997.

[Julier and Uhlmann, 1997b] S. J. Julier and J. K. Uhlmann. A non-divergent estimation algorithm in the presence of unknown correlations. In *Proc. of the 1997 American Control Conf.*, volume 4, pages 2369–2373, 1997.

[Landau, 1987] H. J. Landau, editor. *Moments in Mathematics*, volume 37 of *Proc. of Symposia in Applied Mathematics*, Providence, RI, USA, 1987. American Mathematics Society.

[Pena, 2000] J. M. Pena. On the multivariate Horner scheme. *SIAM Journal on Numerical Analysis*, 37(4):1186–1197, 2000.

[Rauh and Hanebeck, 2005] A. Rauh and U. D. Hanebeck. Moment-based prediction stepp for nonlinear discrete-time dynamic systems using exponential densities. In *Proc. of the 44th IEEE Conf. on Decision and Control*, Sevilla, Spain, 2005.

[Ridley *et al.*, 2004] M.F. Ridley, B. Upcroft, L. Ong, S. Kumar, and S. Sukkarieh. Decentralised data fusion with parzen density estimates. In *Proc. of the Int'l Conf. on Intelligent Sensors, Sensor Networks and Information Processing*, pages 161–166, 2004.

[Roy and Minocha, 1991] S. C. Dutta Roy and S Minocha. A note on ëfficient evaluation of polynomials and exponentials of polynomials for equispaced arguments*:* *IEEE Transactions on Signal Processing*, 39(11):2554–2556, 1991.

[Schrempf *et al.*, 2006] O. C. Schrempf, D. Brunn, and U. D. Hanebeck. Density approximation based on Dirac mixtures with regard to nonlinear estimation and filtering. In *IEEE Conf. on Decision and Control*, 2006.

[Smith *et al.*, 1990] R. Smith, M. Self, and P. Cheeseman. Estimating uncertain spatial relationships in robotics. In I.J. Cox and G.T. Wilfong, editors, *Autonomous Robot Vehicles*, pages 167–193. Springer-Verlag, Berlin, 1990.

[Sukkarieh *et al.*, 2003] S. Sukkarieh, E. Nettleton, J.-H. Kim, M. Ridley, A Goktogan, and H. F. Durrant-Whyte. The ANSER project: Data fusion across multiple uninhabited air vehicles. *International Journal of Robotics Research*, 22(7/8):505–540, 2003.

[Tonkes and Blair, 2007] B. Tonkes and A. D. Blair. Decentralised data fusion with exponentials of polynomials. In To appear: *Proc. of the 2007 IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems*, 2007.

[Upcroft *et al.*, 2004] B. Upcroft, S. Kumar, M. Ridley, L. Ong, and H. F. Durrant-Whyte. Fast reparameterisation of Gaussian mixture models for robotics applications. In *Australasian Conf. on Robotics and Automation 2004*, 2004.