



King's Research Portal

DOI:

[10.1007/978-3-642-40020-9_76](https://doi.org/10.1007/978-3-642-40020-9_76)

Document Version

Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Armstrong, J., & Brigo, D. (2013). Stochastic filtering by projection: The example of the quadratic sensor. In F. Nielsen, & F. Barbaresco (Eds.), *Geometric Science of Information: First International Conference, GSI 2013, Paris, France, August 28-30, 2013: proceedings* (N/A ed., Vol. N/A, pp. 685-692). (Lecture Notes in Computer Science; Vol. 8085). Springer. https://doi.org/10.1007/978-3-642-40020-9_76

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Stochastic Filtering by Projection: the Example of the Quadratic Sensor

John Armstrong¹ and Damiano Brigo²

¹ Dept. of Mathematics, King's College London, Strand, London, WC2R 2LS

john.1.armstrong@kcl.ac.uk

² Dept. of Mathematics, Imperial College, London

Abstract. The “projection method” is an approach to finding numerical approximations to the optimal filter for non linear stochastic filtering problems. One uses a Hilbert space structure on a space of probability densities to project the infinite dimensional stochastic differential equation given by the filtering problem onto a finite dimensional manifold inside the space of densities. This reduces the problem to finite dimensional stochastic differential equation.

Previously, the projection method has only been considered for the Hilbert space structure associated with the Hellinger metric. We show through the numerical example of the quadratic sensor that the approach also works well when one projects using the direct L^2 metric.

Previous implementations of projection methods have been limited to solving a single problem. We indicate how one can build a computational framework for applying the projection method more generally.

1 The Filtering Problem in Continuous Time

The state of a system X evolves over time according to some stochastic process. We cannot observe the state of directly instead we make an imperfect measurement Y which is also perturbed stochastically. Thus we assume that we have processes X and Y related by stochastic differential equations of the following form:

$$\begin{aligned}dX_t &= f_t(X_t) dt + \sigma_t(X_t) dW_t, \quad X_0, \\dY_t &= b_t(X_t) dt + dV_t, \quad Y_0 = 0.\end{aligned}\tag{1}$$

In this paper we will assume that X and Y are processes taking values in \mathbb{R} and that V and W are two independent Wiener processes.

Using the observations Y one cannot hope to determine the state X . Instead one hopes that given a prior probability distribution π_0 for X one might be able to compute the probability distribution π_t for X at subsequent times. This is called the filtering problem, see Bain & Crisan [5] for an introduction to filtering.

If one imposes certain asymptotic bounds and regularity conditions on the coefficients in the above equations, it turns out that there is a well defined probability measure π_t . Moreover the evolution of π_t over time is determined

by a stochastic partial differential equation called the Kushner–Stratonovich equation. For a given test function ϕ we can write:

$$\pi_t(\phi) = \pi_0(\phi) + \int_0^t \pi_s(\mathcal{L}_s \phi) ds + \int_0^t [\pi_s(b_s \phi) - \pi_s(b_s) \pi_s(\phi)] [dY_s - \pi_s(b_s) ds] . \quad (2)$$

The *backward diffusion operator* \mathcal{L} is defined by:

$$\mathcal{L}_t = f_t \frac{\partial}{\partial x} + \frac{1}{2} a_t \frac{\partial^2}{\partial x^2} .$$

We suppose that the measure π_t is determined by a probability density. A formal calculation then gives the following Itô equation for the evolution of p :

$$dp_t = \mathcal{L}_t^* p_t dt + p_t [b_t - E_{p_t}\{b_t\}] [dY_t - E_{p_t}\{b_t\} dt] .$$

Here \mathcal{L}^* is the formal adjoint of \mathcal{L} – the so-called forward diffusion operator.

As we shall explain shortly we will need a version of the above equation written in Stratonovich form. With a little calculation one can show that the Stratonovich version of our equation is:

$$dp_t = \mathcal{L}_t^* p_t dt - \frac{1}{2} p_t [|b_t|^2 - E_{p_t}\{|b_t|^2\}] dt + p_t [b_t - E_{p_t}\{b_t\}] \circ dY_t^k .$$

If the coefficients are linear and the prior distribution is normal, this equation can be solved analytically to give the so-called Kalman Filter. This Kalman filter reduces the problem to a two dimensional SDE for the mean and variance of the distribution. However, in general, as was shown in [6], one cannot reduce this problem to a finite dimensional one.

2 The Projection Method

The projection method can be understood abstractly as an approach to solving a differential equation on a Riemannian manifold M . Given a vector field \mathcal{X} defined on M , we wish to find the trajectory of a particle p as it flows along \mathcal{X} . We attempt to approximate this trajectory by choosing a submanifold Σ of M and using the Riemannian metric on M to project \mathcal{X} onto the tangent space of Σ . This gives rise to a vector field \mathcal{X}' on Σ . The hope is that the trajectories of \mathcal{X}' will be a good approximation for the trajectories of \mathcal{X} . The distance-minimizing properties of projection will ensure that infinitesimally, this is the best achievable approximation using a vector field on Σ .

The approach becomes interesting when one considers an infinite dimensional Hilbert manifold M . One is now using a finite dimensional ordinary differential equation (ODE) to approximate an infinite dimensional equation. If we take M to be a function space on which we wish to solve a partial differential equation (PDE), we have a possible approach for approximating the solution to PDE's

with finite dimensional ODE's. Indeed many standard approaches to the numerical solution of PDE's can be re-interpreted geometrically this way.

One can extend the approach to stochastic differential equations. The only additional complexity is that one needs to use Stratonovich differential equations in order to invariantly define stochastic vector fields on a manifold.

Thus we will attempt to numerically solve the filtering problem by mapping the space of probability distributions into a Hilbert manifold and then projecting onto a finite dimensional submanifold. In fact the Hilbert manifolds we use will simply be Hilbert spaces.

3 Choice of Hilbert Space Structure

There are two obvious ways of embedding the state of our system as lying in a Hilbert space. One can consider \sqrt{p} which lies inside $L^2(\mathbb{R})$ or one can assume that p is itself square integrable and so lies inside $L^2(\mathbb{R})$. These two approaches give two different metrics on the space of probability distributions. The former yields the Hellinger metric, the latter we will call the direct L^2 metric.

Since there are no assumptions on the integrability of p , the Hellinger metric immediately seems more attractive from a theoretical standpoint. It has other advantages: its definition can be extended to probability measures; its definition is invariant under re-parameterizations of \mathbb{R} . These properties explain why the Hellinger metric is the most popular choice when considering the differential geometry of probability distributions.

The direct L^2 metric is only defined on square integrable distributions and is not invariant under re-parameterizations. However, it has one distinct advantage over the Hellinger metric: it is defined in terms of p rather than \sqrt{p} . Since the metric is bilinear in p , using the L^2 metric gives more convenient formulae for mixture distributions than does the Hellinger metric. These simpler formulae have a practical consequence: when we come to consider numerical implementations of the projection method we will find that numerical integration is normally necessary to apply the projection method in the Hellinger metric, but the corresponding integrals for the direct metric can sometimes be performed analytically. This should ultimately translate into faster and more scalable computer algorithms.

We should remark that the space of probability distributions is not a submanifold of $L^2(\mathbb{R})$. Fortunately we can view the stochastic PDE we wish to solve as an equation on the whole of $L^2(\mathbb{R})$ and so avoid the thorny question of defining a manifold structure on the space of probability measures.

4 Choice of Submanifold

We will consider the following submanifolds of our Hilbert spaces:

Definition 1. *The polynomial exponential family of degree m consists of densities of the form:*

$$p(x) = \exp(a_m x^m + a_{m-1} x^{m-1} + \dots + a_0)$$

We require that m is even and a_m is negative in order for p to be integrable. a_0 is determined from the other coefficients by requiring that p integrates to 1. Thus this defines an m -dimensional submanifold of our Hilbert space.

Definition 2. A mixture of m normal distributions is a distribution of the form

$$p(x) = \sum_{i=1}^m c_i \exp\left(-\frac{(x - \mu_i)^2}{2\sigma_i^2}\right)$$

where $c_i > 0$. We can consider c_1 to be determined by the normalization condition. Thus the mixtures of m normal distributions give rise to a $3m - 1$ dimensional family.

The motivation for considering these particular submanifolds is that, even in low dimensions, they allow us to reproduce many of the qualitative phenomena seen in the filtering problem. In particular we can produce highly skewed distributions and multi modal distributions. Many other possible choices of submanifold are worth consideration. For example by considering spaces of functions defined piecewise on a grid one might hope to reinterpret finite difference methods in terms of projection.

5 The Projected Equation

Let M be an m dimensional submanifold of L^2 parameterized by $\theta = (\theta_1, \theta_2, \dots, \theta_m)$. Define $v_i = \frac{\partial p}{\partial \theta^i}$ so that $\{v_1, v_2, \dots, v_m\}$ gives a basis for the tangent space of M at a point θ .

The direct L^2 metric induces a Riemannian metric h_{ij} on M . By projecting both sides of the Stratonovich equation for the evolution of p given above, we can obtain a stochastic differential for the evolution of the parameter θ .

To simplify the result, we introduce the following notation:

$$\begin{aligned} \gamma_t^0(p) &:= \frac{1}{2} [|b_t|^2 - E_p\{|b_t|^2\}] p, \\ \gamma_t^1(p) &:= [b_t - E_p\{b_t\}] p. \end{aligned} \tag{3}$$

One can then show that the projected equation for θ is equivalent to the stochastic differential equation:

$$d\theta^i = \sum_{j=1}^m h^{ij} \{ \langle p(\theta), \mathcal{L}v_j \rangle dt - \langle \gamma^0(p(\theta)), v_j \rangle dt + \langle \gamma^1(p(\theta)), v_j \rangle \circ dY \}. \tag{4}$$

Here $\langle \cdot, \cdot \rangle$ denotes the direct L^2 inner product.

One can similarly derive an equation for θ using the Hellinger metric. See [2] for the details.

6 Numerical Implementation

Being an ordinary stochastic differential equation, our equation for the evolution of θ can be approximated numerically using standard techniques. One must be a little careful as it is a Stratonovich equation and so one cannot use the simple Euler scheme. In our implementations we used the Stratonovich–Heun scheme described in [4].

The difficulty in putting this idea into practice is the complexity of equation (4). Recall that the v_j are defined in terms of partial derivatives of p and that the inner product $\langle \cdot, \cdot \rangle$ is defined in terms of integration over \mathbb{R} .

In our implementation we have addressed this complexity by introducing two object oriented software abstractions: an interface `FunctionRing` and an interface `Submanifold`.

The role of the `FunctionRing` is to perform computations such as the multiplication, differentiation and integration of elements of the ring. If one restricts to a class of functions such as products of polynomials and Gaussians, this is reasonably easy to implement. In this particular case, one can even perform the integrals analytically.

The role of the `Submanifold` is to compute the tangent vectors at a point θ as elements of the `FunctionRing`.

Given implementations of these two interfaces, one can then compute the coefficients of dt and dY in equation 4. One can then use the Stratonovich–Heun scheme to approximate the evolution of θ .

Notice that the precise behaviour of the `Submanifold` interface depends not only on the submanifold selected but on the choice of parameterization. To ensure the best numerical results, one should choose a parameterization with as large a domain as possible. For the sake of brevity we omit the details of the parameterizations used in our implementations.

The projection method has been implemented previously (see [2] and [3]) but only for the special case of the cubic sensor and only for projection onto the polynomial exponential family using the Hellinger metric. One problem found in this case is that the corresponding integrals can only be performed numerically. One expects that this would lead to performance problems when extending the approach to higher dimensions. By contrast, if one approximates the coefficients of the problem using Taylor series, L^2 projection onto normal mixtures can be performed using analytic integrals.

7 Numerical Results

In [1] we examine the performance of these approaches against a variety of different problems. In this paper we will simply consider the quadratic sensor problem:

$$\begin{aligned}dX_t &= dW_t \\dY_t &= X^2 + dV_t .\end{aligned}$$

To run a simulation for this problem we also need to choose a prior distribution. We have taken this to be given by: $p(x) = \exp(0.25 - x^2 + x^3 - 0.25x^4)$

We have then compared the numerical results obtained using the following approaches:

- Projection using the Hellinger metric onto the degree 4 polynomial exponential family. We will label this P1 in graphs.
- Projection using the L^2 metric onto a mixture of two normal distributions (labelled P2). Since the prior distribution is not of this form, we chose an initial value for our parameter θ by numerically minimizing the L^2 between the true prior distribution and $p(\theta(0))$.
- The extended Kalman filter (labelled EK). This is described in [5]. It is derived, in essence, by linearizing the filtering problem and then applying the Kalman filter. Thus it approximates the true distribution using a normal distribution.
- A finite difference method with a very fine grid. This is assumed to be extremely close to the true solution and so provides a performance benchmark and so is labelled as Exact in our graphs.

An advantage of considering the quadratic sensor problem is that its behaviour can be understood heuristically quite easily. Since the sensor equation contains only an X^2 term, our measurements give us no information about the sign of X , they only tell us its magnitude. Thus we expect that once the state moves close to the origin, the probability distribution will become nearly symmetrical and remain symmetrical thereafter. When the state moves away from the origin, one expects the distribution to be reasonably well approximated by two normal distributions whose standard deviations decrease as X increases.

This behaviour can be seen in Fig. 1. We have not shown the results from using Hellinger projection to reduce visual clutter, but they are qualitatively similar to those obtained using L^2 projection.

To give a more objective view of the performance of the filters we have plotted the L^2 norm of the distance between our numerical results and the “exact” result obtained using a fine grid. We have termed this the L^2 residual and have plotted these residuals against time in Fig. 2.

We have compared our results with the extended Kalman filter as this is a commonly used algorithm that also approximates solutions to the filtering problem using a low dimensional family of distributions. One might also wish to compare the projection algorithms with particle filter methods since these give some of the most effective numerical approaches to the problem currently known. The difficulty is knowing what would constitute a fair comparison. Particle filters are a Monte Carlo approach requiring one to generate a large number of “particles” - these particles are Dirac masses which when combined give an approximation to the exact filter in the weak topology. Since particle filters require many particles, so it not too surprising that numerical examples show that low dimensional particle filters do not perform as well as projection methods of the same dimension. Details of such comparisons are given in [1].

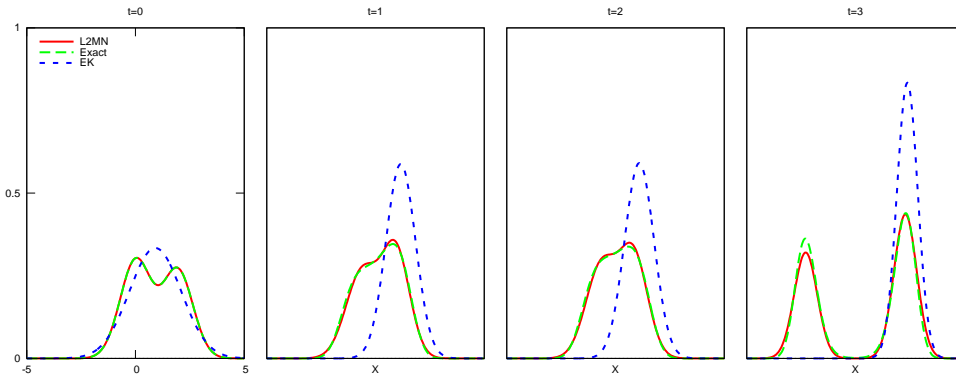


Fig. 1. Evolution of the probability density over time

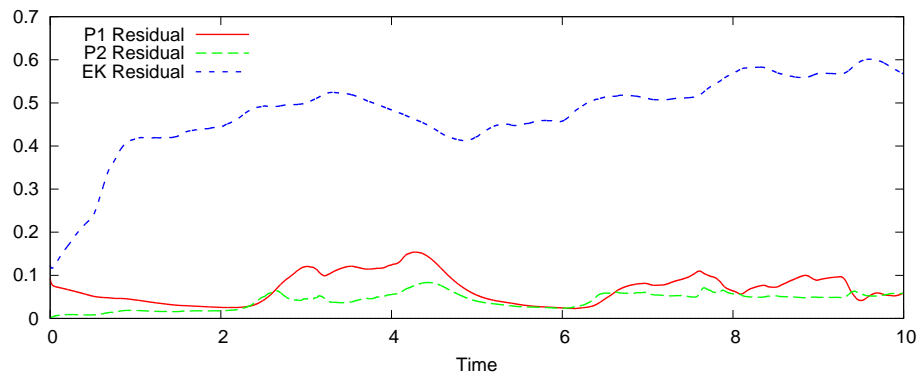


Fig. 2. L^2 residuals for the quadratic sensor

8 Conclusions

L^2 projection and Hellinger projection both give rise to numerical methods for solving the filtering problem. Both methods allow one to find surprisingly accurate approximations to non-linear problems by solving only low dimensional stochastic differential equations. In the example of the quadratic sensor, we get good results by considering manifolds of dimensions as low as 4 and 5.

While the Hellinger metric has certain theoretical advantages over the direct L^2 metric, the L^2 metric can sometimes lead to simpler formulae. In particular if the filtering equation has polynomial coefficients projection using the direct L^2 metric can be achieved by evaluating integrals analytically. By contrast, numerical integration is required for projection using the Hellinger metric onto the polynomial exponential family of degree greater than 2.

References

1. Armstrong, J, Brigo D, Stochastic filtering via L^2 projection on mixture manifolds with computer algorithms and numerical examples. Forthcoming.
2. Brigo, D, Hanzon, B, LeGland, F, A differential geometric approach to nonlinear filtering: The projection filter, IEEE T AUTOMAT CONTR, 1998, Vol: 43, Pages: 247 – 252
3. Brigo, D, Hanzon, B, Le Gland, F, Approximate nonlinear filtering by projection on exponential manifolds of densities, BERNOULLI, 1999, Vol: 5, Pages: 495 – 534
4. Burrage, K., Burrage, P. M., and Tian, T. Numerical methods for strong solutions of stochastic differential equations: an overview, Proc. R. Soc. Lond. A 460 (2004), 373402.
5. Bain, A., and Crisan, D. (2010). Fundamentals of Stochastic Filtering. Springer-Verlag, Heidelberg.
6. M. Hazewinkel, S.I.Marcus, and H.J. Sussmann, Nonexistence of finite dimensional filters for conditional statistics of the cubic sensor problem, *Systems and Control Letters* **3** (1983) 331–340.