



## King's Research Portal

*Document Version*  
Peer reviewed version

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

*Citation for published version (APA):*  
Armstrong, J., Brigo, D., & Hanzon, B. (Accepted/In press). Optimal Projection Filters with Information Geometry. *Information Geometry*.

### **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](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

# Optimal Projection Filters with Information Geometry

John Armstrong  
Dept. of Mathematics  
King's College London  
john.1.armstrong@kcl.ac.uk

Damiano Brigo  
Dept. of Mathematics  
Imperial College London  
damiano.brigo@imperial.ac.uk

Bernard Hanzon  
Mathematical Science  
University College Cork  
b.hanzon@ucc.ie

First version: Feb. 21, 2022. This version: May 3, 2022

## Abstract

We review the introduction of several types of projection filters. Projection structures coming from information geometry are used to obtain a finite dimensional filter in the form of a stochastic differential equation (SDE), starting from the exact infinite-dimensional stochastic partial differential equation (SPDE) for the optimal filter. We start with the Stratonovich projection filters based on the Hellinger distance as introduced and developed in Brigo, Hanzon and Le Gland (1998, 1999) [19, 20], where the SPDE is put in Stratonovich form before projection, hence the term “Stratonovich projection”. The correction step of the filtering algorithm can be made exact by choosing a suitable exponential family as manifold, there is equivalence with assumed density filters and numerical examples have been studied. Other authors further developed these projection filters and we present a brief literature review. A second type of Stratonovich projection filters was introduced in Armstrong and Brigo (2016) [6] where a direct  $L^2$  metric is used for projection. Projecting on mixtures of densities as a manifold coincides with Galerkin methods. All the above projection filters lack optimality, as the single vector fields of the Stratonovich SPDE are projected optimally but the SPDE solution as a whole is not approximated optimally by the projected SDE solution according to a clear criterion. This led to the optimal projection filters in Armstrong, Brigo and Rossi Ferrucci (2019, 2018) [10, 9], based on the Ito vector and Ito jet projections, where several types of mean square distances between the optimal filter SPDE solution and the sought finite dimensional SDE approximations are minimized, with numerical examples. After reviewing the above developments, we conclude with the remaining challenges.

**Keywords:** Stochastic partial differential equations, Stochastic differential equations, SPDEs projection on a submanifold, Stratonovich projection, Itô-vector projection, Itô-jet projection, Non-linear Filtering, Projection filters, Stratonovich projection filters, Optimal projection filters.

AMS classification codes: 62M20, 93E11, 60G35, 62B10, 58J65, 60H10, 65D18, 58A20

## 1 Introduction and history

### 1.1 Information geometry as the differential geometric approach to statistics

Information Geometry is an informal term to describe the differential geometric approach to statistics, or more precisely to study the differential geometric properties of sets of probability distributions, on which a manifold structure is usually built, leading to so called statistical manifolds. A list of references, far from being comprehensive, describing the evolution of the discipline include C. R. Rao 1945 [48], interpreting the Fisher matrix for a parametric family of distributions as a Riemannian metric on the given finite-dimensional statistical manifold, the dimension being usually related to the number of parameters. It is important for our paper to point out that the Fisher information matrix is related naturally to the Hellinger distance on more general infinite-dimensional spaces of probability measures, a distance based on the  $L^2$  structure of sets of square roots of probability densities. Nand Lal Aggarwal (1974)[1], Shun'ichi Amari (1985) [3], Ole Barndorff-Nielsen (1978) [13] and Giovanni Pistone (Pistone and Sempi 1995 [47]) are other important references that contributed to the development of information geometry and are related to this article, although this list is far from being comprehensive.

### 1.2 Information geometry and filtering dynamics

Our work concerns the application of information geometry to approximation of dynamics of probability distributions, in most cases stemming from the stochastic filtering problem.

To state it in basic terms, in stochastic filtering one observes a random signal perturbed by random noise. The unperturbed random signal cannot be observed but needs to be estimated. For example, the perturbed signal could be the radar reading of the position of a spacecraft, which would not provide the exact position of the spacecraft due to several disturbances (“noise”) in the radar observations. It would then be necessary to estimate the real position of the spacecraft from the noisy radar readings. This is a filtering problem. A filtering algorithm was used in the Apollo 11 mission (Cipra 1993 [24]), the first human landing on the moon. Filtering has also applications in areas such as water level estimation and prediction, submarine navigation, econometrics, target tracking and many others. A good historical book on filtering with an eye to applications is Jazwinski (1970) [36], see also Maybeck (1982) [43], while the mathematical aspects are considered fully in Liptser and Shiryaev (1978) [41]. More recent monographs on filtering are Ahmed (1998) [2] and Bain and Crisan (2009) [12].

The general solution of the filtering problem at a given time is given by the probability density of the unperturbed state of the system at that time, conditional on the perturbed observations up to the given time. When the unobserved signal and the observed signal evolve in continuous time, the filter density follows a stochastic partial differential equation (SPDE). It has been shown that this probability density, the solution of the SPDE, does not evolve in a finite dimensional statistical manifold, except in very special cases. For example, if the dynamics of the unobserved system is linear, the observations are linear, the noises are Gaussian and the initial condition of the unperturbed signal is also Gaussian (or deterministic), then the filter is Gaussian and its density can be characterized by a finite dimensional set of parameters, namely the mean vector and variance-covariance matrix of the resulting Gaussian distribution. This leads to the celebrated Kalman filter. However, this does not happen usually, in the non-linear case, and the filter is infinite dimensional in general, as shown for the cubic sensor example by Hazewinkel, Marcus and Sussmann (1983) [34].

### 1.3 Classic projection filters: Stratonovich–Hellinger projection

Enters information geometry. Can information geometry provide us with a method to approximate the infinite-dimensional filter with a finite-dimensional approximation that is close to the original filter? The idea to apply the Fisher Metric and Hellinger distance to this problem was first sketched in an article of Bernard Hanzon (1987) [32] while he was working at the Technical University of Delft. Hanzon suggested to project the SPDE equation in Stratonovich form for the evolution of the filter density onto a finite dimensional statistical manifold, using the Fisher metric/Hellinger distance. We call this “Stratonovich projection” and it consists in projecting the separate vector fields of the SPDE corresponding to the drift and diffusion part of the Stratonovich version. The projected equation would describe a finite dimensional density evolution, called projection filter, approximating the full filter evolution associated with the optimal filter. The paper was presented to a conference in Lancaster whose proceedings were edited by Christopher T. J. Dodson, in a volume with the almost prophetic title “Geometrization of Statistical Theory”. The following year, on August 22, 1988, Hanzon presented the idea at a seminar in Tokyo University called “The Projection Filter” while visiting Shun'ichi Amari. A few years later, in 1991, Hanzon and a PhD student Ruud Hut also from Technical University of Delft, wrote the paper Hanzon and Hut [31] with new results on the projection filter on Gaussian densities, showing that for the Gaussian family the projection filter coincides with a heuristic-based family of finite dimensional filters, the assumed density filters, previously studied by Harold Kushner (1967) [38], see also [43].

The projection filter idea was formulated precisely, extended and made fully rigorous in subsequent works, during the PhD studies of Damiano Brigo with Bernard Hanzon at the Free University of Amsterdam and with Francois LeGland at IRISA/INRIA, in Rennes, France, in 1993–1996 [16]. In these studies it was shown, among other things, that exponential families played a very particular role in the projection filter, allowing for the correction step of the filtering algorithm to be exact, and also fully generalizing the equivalence to the assumed density filters. The filters were tested numerically on some examples. During his PhD, Brigo also authored other papers on small observation noise for the Gaussian projection filter [15, 17] and on approximations of the Fokker-Planck-Kolmogorov equation, as well as formulations of the filter in discrete time using the Kullback-Leibler information, with application to volatility modeling in finance [18]. The main results on the projection filters were published later in Brigo, Hanzon and LeGland (1998, 1999) [19, 20].

One of the key issues, from the start, was making sure that the given approximated equation for the filter density would stay on the chosen statistical manifold. The Stratonovich projection ensured this, but scholars had been studying the behaviour of stochastic differential equations on manifolds independently of the filtering application above. Among those, we refer to David Elworthy (1988) [27], Michel Emery (1989) [29], and more recently Elton Hsu (2002) [35]. We also notice that Elworthy, Le Jan and Li (2010) would discuss geometric aspects of filtering theory in [28], although their book does not deal with projection filters.

### 1.4 Classic projection filters: Stratonovich–direct $L^2$ projection

Brigo returned to the filtering problem as a side project in 2011 after he moved from a managing director position in the financial industry to a full academic position as Gilbert Chair at the Department of Mathematics of King's College London, earlier in 2010. There, in 2011 he met a new colleague, John Armstrong, a differential geometry PhD from Oxford who had worked on almost Kähler geometry and who also had spent several years in the financial industry and was now turning to a full-time academic career. Brigo explained the filtering problem to Armstrong, who grasped immediately the essential ideas and the mathematics. Brigo had already written a preprint on his new idea of applying the direct  $L^2$  structure without square roots to obtain a new type of projection

filter, showing equivalence with Galerkin-based filters when using mixtures of distributions. Armstrong refined the idea and implemented the filter numerically, studying the cubic sensor problem. This led to a second wave of projection filters based on the direct  $L^2$  metric as opposed to the Hellinger distance. It turned out that, as anticipated in the preprint, while the original Hellinger-based filters worked well with exponential families, being equivalent to assumed density filters, the direct  $L^2$  filters worked best with mixture families, being equivalent to Galerkin-based filters. This research went on in 2011-2013 and was published in Armstrong and Brigo (2016) [6]. By 2012 Brigo had moved to Imperial College. During the review of [6], one of the reviewers asked in which sense, or according to which criterion, the projection filter was providing an optimal approximation of the true filter.

### 1.5 Is the classic projection filter an optimal approximation?

The essence of the problem of optimality of the approximation was based on the way the filtering equation was projected in the projection filter works published until then, mainly [19, 20, 6]. There are two stochastic calculi, Ito and Stratonovich. The two different calculi are suited to different applications, but from a probabilistic point of view the Ito calculus has a more clear interpretation of the stochastic equation coefficients in terms of local mean and local standard deviation, linked to the martingale property. Also, it is believed that even when one works with Stratonovich calculus, under the formalism one can argue that it is still the Ito calculus that “does all the work” (Rogers and Williams (1987) [49], Chapter V.30, p. 184). The problem with Ito calculus is that it violates the chain rule for change of variables. When changing variables, one has to use Ito’s formula, involving a second order term in the transformation.

The true, infinite dimensional filter equation (taking the form of a stochastic partial differential equation, or SPDE) had always been written in Stratonovich form in the previous projection filter works, because in a Stratonovich stochastic equation the two parts describing the drift term and the diffusion coefficient term obey the chain rule under change of variables. This means that they can be interpreted as vector fields and be projected without problems on the tangent space of a submanifold, obtaining vector fields in the submanifolds that would form the approximating finite dimensional stochastic differential equation.

Projecting directly the Ito equation does not work, because the change of variables includes second order terms that do not resemble the behaviour of vector fields. Projection becomes then impossible to perform directly in Ito form. One could re-write the Ito true filter stochastic equation in Stratonovich form, project it, obtain a finite dimensional approximated filter, and transform back this approximate filter equation from Stratonovich to Ito form. But in what sense is this approximation optimal? What criterion does it minimize?

More in detail, the projection of a vector field always provides the best approximation of the original vector field. But a stochastic equation is given by two terms, the drift and the diffusion part, and if one puts the equation in Stratonovich form, the drift and the diffusion coefficients become described by two vector fields and as such can be projected. As the two vector fields are projected, each projected vector field will be the best approximation of the original vector field, but what does this mean for the solution of the stochastic equation as a whole? The stochastic equation is not just the pair of vector fields. In fact, when the equation is in Ito form, the drift and the diffusion coefficients interact when changing variables or coordinates, involving second order terms in the transformation. The fact that Stratonovich is “less good” probabilistically means that putting together two optimal projections of the coefficients to form a single Stratonovich equation does not provide a solution that is optimal in a probabilistic sense, for example in mean square.

## 1.6 Finding optimal projection filters

Armstrong had previously noticed that an Ito equation behaved exactly as a geometric object he was familiar with, called a 2-jet. Brigo, while helping Armstrong in developing the 2-jet interpretation of stochastic differential equations, started looking at the Schwartz Morphism as studied in Emery (1989) [29] and found it to be very close to the 2-jet approach. The 2-jet interpretation was published in Armstrong and Brigo (2017, 2018) [7, 8], which led next to Armstrong and Brigo investigating how one could project a stochastic differential equation on a sub-manifold in an optimal way. Based on Ito Taylor expansions, two different projections satisfying two different types of optimality were found, the Ito-vector and the Ito-jet projections. The Ito-jet projection is superior in terms of optimality, in that it has a higher order of optimality in a precise sense. These results were presented at ICMS in Edinburgh by Armstrong and Brigo (2016) [5], at a conference co-organized in 2015 again by Dodson, this time with Frank Critchley and Frank Nielsen. The two projections were studied further and some technical problems concerning tubular neighborhoods were solved with the help of the then PhD student Emilio Rossi Ferrucci, leading to the publication Armstrong, Brigo and Rossi Ferrucci (2019) [10], see also Armstrong, Brigo and Rossi Ferrucci (2018) [9], where Rossi Ferrucci helped re-derive the optimal projections through constrained optimizations as opposed to Ito Taylor expansions, and where ambient coordinates are used.

In this last paper [10], information geometry comes back as an application of the now optimal projections both in Hellinger distance and direct  $L^2$  metric, comparing them in a numerical case with the traditional Stratonovich projection of previous works. It turns out that Stratonovich is also optimal for a particular criterion that is, however, not a particularly interesting or natural one, so that the Ito-jet projection filter should be preferred in general.

In this paper we will first present a literature review of projection filtering as done by other authors, following the original papers [19, 20], and then we will explain the basic ideas of the optimal projection filters as compared to the Stratonovich ones. We will finally sketch future problems where information geometry might give a contribution. For the reader's convenience, we summarize the different projection filtering approaches in Table 1.

Metric $\rightarrow$ Projection $\downarrow$	Hellinger	Direct $L^2$
Stratonovich projection	Stratonovich "classic" PF exponential families [19, 20]	Stratonovich "classic" PF mixture families [6]
Ito-vector projection	Optimal Ito-vector PF Gaussian family [10] exponential families?	Optimal Ito-vector PF Gaussian family [10] mixture families?
Ito-jet projection	Optimal Ito-jet PF Gaussian family [10] exponential families?	Optimal Ito-jet PF Gaussian family [10] mixture families?

Table 1: A simplified classification of projection filters (PFs).

## 2 Other works based on the classic projection filters

Our original work on projection filters was further studied and applied to several fields by subsequent authors. Here we mention only a few examples to illustrate the breadth of the possible use of information geometry and dynamics in applications.

Jones and Soatto (2011) [37] briefly mention the projection filter as one of the possible algorithms for on-line estimation in the context of visual-inertial navigation, mapping and localization. Lermusiaux (2006) [40] mentions the projection filter as a possible tool for estimation of uncertainties for ocean dynamics. Kutschireiter, Rast, and Drugowitsch (2022)[39] apply the projection filter to continuous time circular filtering. Projection filters have been applied to quantum systems for example in van Handel and Mabuchi (2005) [52] and in Gao, Zhang and Petersen (2019) [30]. Ma, Zhao, Chen and Chang (2015) [42] apply projection filters to hazard position estimation. Vellekoop and Clark (2006)[53] extend the projection filter theory to deal with changepoint detection. Tronarvand and Särkkä (2019) [51] present a projection filter for systems with discrete time measurement having arbitrary likelihoods. Surace and Pfister (2017) [50] apply the Gaussian projection filter to estimate the parameters of a partially observed diffusion. Harel, Meir and Oppor (2015)[33] apply the assumed density filters, equivalent to the projection filters, to the filtering of optimal point processes with applications to neural encoding. Azimi-Sadjadi and Krishnaprasad (2005)[11] apply projection filter algorithms to navigation. Bröcker and Parlitz (2000)[23] apply projection filter techniques to address noise reduction in chaotic time series. Zhang, Wang, Wu and Xu (2014)[54] apply the Gaussian projection filter as part of their estimation technique to deal with measurements of fiber diameters in melt-blown nonwovens. The projection filter further attracted the attention of the Swedish Defense Research Agency, that summarized and studied it in 2003 in the report [14].

### 3 Optimal projection filters for non-linear filtering via information geometry

We studied the application of the new projections to nonlinear filtering via information geometry in Armstrong, Brigo and Rossi Ferrucci (2019) [10]. Here we summarize the results of that paper, showing how our new projection methods work for stochastic filtering. As explained in the introduction, this enhances optimality of the approximations compared to our previous works in [19], [20] and [6].

Let us first summarize the filtering problem for diffusions. One has a signal  $X$  that evolves according to a SDE, and observes a process  $Y$  which is a function of this signal plus noise.

The filtering problem consists in estimating the signal  $X$  given the present and past observations  $Y$ . If  $t$  is the current time, the solution of the filtering problem is the probability density of the state  $X_t$  conditional on the observations from time 0 to time  $t$ , call this density  $p_t$ . The density  $p_t$  follows the Kushner-Stratonovich (or alternatively the Zakai) stochastic partial differential equation (SPDE) that, under some technical assumptions, can be seen as a stochastic differential equation in the infinite dimensional  $L^2$  space of square roots of densities (Hellinger metric) or of densities themselves (direct  $L^2$  metric).

The process we wish to approximate on a low dimensional manifold is  $p_t$ , evolving in the  $L^2$  infinite dimensional space, while the submanifold  $M$  where we seek approximation is a finite dimensional family of probability densities parametrized by  $\theta$ , acting as coordinates:  $M = \{p(\cdot, \theta), \theta \in \Theta \subset \mathbb{R}^n\}$ . We aim at finding a SDE for  $\theta$  such that  $p(\cdot, \theta_t)$  approximates the optimal filter  $p_t(\cdot)$  in an optimal way.

### 3.1 The Kushner Stratonovich equation

We suppose that the state  $X_t \in \mathbb{R}^m$  of a system evolves according to the Itô stochastic differential equation:

$$dX_t = f(X_t, t) dt + \sigma(X_t, t) dW_t$$

where  $f$  and  $\sigma$  are smooth  $\mathbb{R}^m$  valued functions and  $W_t$  is a Brownian motion. One typically adds growth conditions to ensure a global existence and uniqueness result for the signal equation, see for example [6] and references therein for the details.

We suppose that an associated process, the observation process,  $Y_t \in \mathbb{R}^d$  evolves according to the equation:

$$dY_t = b(X_t, t) dt + dV_t$$

where  $b$  is a smooth  $\mathbb{R}^d$  valued function and  $V_t$  is a Brownian motion independent of  $W_t$ . Note that the filtering problem is often formulated with an additional constant in terms of the observation noise. For simplicity we have assumed that the system is scaled so that this can be omitted.

The filtering problem is to compute the conditional distribution of  $X_t$  given a prior distribution for  $X_0$  and the values of  $Y$  for all times up to and including  $t$ .

Subject to various bounds on the growth of the coefficients of this equation, the assumption that the distribution has a density  $p_t$  and suitable bounds on the growth of  $p_t$  one can show that  $p_t$  satisfies the Kushner–Stratonovich SPDE:

$$dp_t = \mathcal{L}_t^* p_t dt + p_t [b(\cdot, t) - E_{p_t}(b(\cdot, t))]^T [dY_t - E_{p_t}(b(\cdot, t)) dt] \quad (1)$$

where  $E_p$  denotes the expectation with respect to the density  $p$ ,

$$E_p[\psi] = \int \psi(x) p(x) dx, \quad E_p[\phi(\cdot, t)] = \int \phi(x, t) p(x) dx,$$

and the forward diffusion operator  $\mathcal{L}_t^*$  is defined by:

$$\mathcal{L}_t^* \phi = -\frac{\partial}{\partial x^i} [f_i(x, t) \phi] + \frac{1}{2} \frac{\partial^2}{\partial x^i \partial x^j} [a_{ij}(x, t) \phi] \quad (2)$$

where  $a = \sigma \sigma^T$ . Note that we are using the Einstein summation convention in this expression.

In the event that the coefficient functions  $f$  and  $b$  are all linear and  $\sigma$  is a deterministic function of time one can show that so long as the prior distribution for  $X$  is Gaussian, or deterministic, the density  $p$  will be Gaussian at all subsequent times. This allows one to reduce the infinite dimensional equation (1) to a finite dimensional stochastic differential equation for the mean and covariance matrix of this normal distribution. This finite dimensional problem solution is known as the Kalman filter.

For more general coefficient functions, however, equation (1) cannot be reduced to a finite dimensional problem [34]. Instead one might seek approximate solutions of (1) that belong to some given statistical family of densities. This is a very general setup and includes, for example, approximating the density using piecewise linear functions to derive a finite difference approximation or approximating the density with Hermite polynomials to derive a spectral method. Other examples include exponential families (considered in [20, 19]) and mixture families (considered in [4, 6]).

Our projection theory tells us how one can find good approximations on a given statistical family with respect to a given metric on the space of distributions. We illustrate this by writing down the Itô-vector and Itô-jet projection of (1) for the  $L^2$  and Hellinger metrics onto a general manifold<sup>1</sup>.

<sup>1</sup>Note that it is also possible to consider projecting the Zakai equation. However, as explained in [6], one expects that projecting the Kushner–Stratonovich equation will lead to smaller error terms in direct metric, whereas the projected equations are the same in Hellinger metric. See [6] for a discussion



A good part of the classic filtering literature focuses on the very specific case of seeking approximate solutions using Gaussian distributions. The idea of approximating the solution to the filtering problem using a Gaussian distribution has been considered by numerous authors who have derived variously, the extended Kalman filter [46], assumed density filters [38] and Stratonovich projection filters [19]. Some of these are related, for example the assumed density filters and Stratonovich projection filters in Hellinger metrics for Gaussian (and more generally exponential) families coincide [20]. Using our projection methods, we have been able to derive projection filters which outperform all these other filters also in the specific Gaussian case (assuming performance is measured over small time intervals using the appropriate Hilbert space metric).

We will be using  $L^2$  geometry here. More generally, for the the geometry of approximations to the infinite dimensional filtering problems based on  $L^2$  or Orlicz charts we refer for example to [19, 20, 6, 10, 21, 22, 44, 45].

### 3.2 Stratonovich projections

The Stratonovich projection filters have been abundantly studied in [19, 20] in Hellinger metric, and in [6] in direct metric, see also references in Section 2 for the Hellinger case. Here we briefly summarize them. To shorten notation, we will omit time dependence when obvious from the context, so  $p = p_t$ ,  $b = b(\cdot, t)$ , and so on. For this method, the optimal filter SPDE is given by putting the optimal filter equation (1) in Stratonovich form, obtaining

$$dp = \mathcal{L}^* p dt - \frac{1}{2} p [ |b|^2 - E_p\{|b|^2\} ] dt + p [b - E_p\{b\}]^T \circ dY . \quad (3)$$

For convenience, let us rewrite this as

$$dp = A dt + B \circ dY . \quad (4)$$

This is a Stratonovich SPDE. The P in SPDE and the particular type of SPDE we have imply that, in general, the solution  $p$  will not belong to any finite dimensional family of densities  $M = \{p(\cdot, \theta), \theta \in \Theta \subset \mathbb{R}^n\}$ . As every computer implementation is inherently finite dimensional, we need a way to get  $p_t$  approximated through a finite dimensional density  $p(\cdot, \theta_t)$  for all times  $t$ .

Now, with the Stratonovich filter SPDE equation above, one can do something very simple. Since the Stratonovich SPDE satisfies the chain rule,  $A$  and  $B$  behave like two vector fields in a suitable function space. So the equation is characterized by a “ $dt$ ” vector field  $A$  and a “ $dY_t$ ” vector field  $B$ . These are two separate vector fields and for the time being we are content with dealing with them separately, but as we will discuss later this is not a choice without consequences. Dealing with  $A$  and  $B$  separately, one can project them on the tangent space of  $M = \{p(\cdot, \theta), \theta \in \Theta \subset \mathbb{R}^n\}$  (direct metric) or of their square roots (Hellinger metric) obtaining, in the direct metric case for example,

$$dp(\cdot, \theta_t) = \Pi_{p(\cdot, \theta_t)}[A] dt + \Pi_{p(\cdot, \theta_t)}[B] \circ dY \quad (5)$$

where  $\Pi$  is the tangent space projection at the denoted point for the manifold  $M$ . Applying the chain rule gives immediately a finite dimensional SDE for  $d\theta_t$  from the above equation, where the coefficients are known and where the SDE can be implemented easily in a finite dimensional setting, giving a finite dimensional filter.

This is basically the  $L^2$  direct metric or Hellinger projection filter in a nutshell, it has been studied and implemented in [19, 20, 6] and by a number of subsequent authors, as summarized in Section 2.

We conclude the summary of the Stratonovich projections by saying that they do satisfy an optimality criterion, although it is a criterion that is somewhat unnatural and not helpful. It requires to run an artificial filter in negative time and to include it into the criterion to be minimized. This is summarized in Table 2.

### 3.3 Itô-vector projections

Let us go back to our exact filter equation in Stratonovich form (4):

$$dp = A dt + B \circ dY .$$

Now in the Stratonovich projection filters we projected separately the vector fields  $A$  and  $B$  obtaining a projected equation. By nature, the projection is the best (optimal) approximation for  $A$  and  $B$  separately on the chosen manifold tangent space. However, does this translate into an optimality of the solution  $p(\cdot, \theta_{t+\delta t})$  as an approximation of the exact  $p_{t+\delta t}$  for say small  $\delta t$ , given that we had the optimal filter up to  $t$  and now we wish to approximate the next step  $\delta t$  (in most cases  $t = 0$  as we will seek an optimal approximation from time 0)? In other terms, is there a norm  $\|\cdot\|$  for which we can say that in some sense

$$\theta_{t+\delta t} \approx \operatorname{argmin}_{\theta} \|p_{t+\delta t} - p(\cdot, \theta_{t+\delta t})\| \quad (6)$$

so long as  $\delta t$  is small? This is a very legitimate question, and it comes from the fact that the two vector fields of a SDE or SPDE ( $A$  and  $B$  in our example) interact in a very specific way in determining the solution. If we agree it is the Ito solution we are considering primarily (Rogers and Williams (1987) [49], Chapter V.30, p. 184), note that to transform a Stratonovich SDE into an Ito one with the same solution, the drift  $A$  is modified by terms involving partial derivatives of the term  $B$ . In the Ito form, therefore, there is no neat separation into two vector fields. Not just that, but the behaviour of the solution of the SDE or SPDE *as a whole* is more than the behaviour of the two separate vector fields  $A$  and  $B$ . This is why the optimality of the separate projections of  $A$  and  $B$  does not guarantee any optimality of the type sought in (6). Consider then (1) and write it as

$$dp = C dt + B dY_t.$$

Again, this is a SPDE, this time in Ito form, and has an infinite dimensional solution in general.

The Ito-vector projection sets out to approach the problem starting from a criterion like (6). It does not resort to a Stratonovich version of the Kushner-Stratonovich equation but keeps the original Ito version.

Let us choose a norm for the space of densities,  $\|\cdot\|$  which might be the direct metric or the Hellinger metric.

Given the diffusion term in the approximating equation minimizing (but not zeroing) the  $\delta t$  term of the expansion for the mean square difference  $E_t[\|p_{t+\delta t} - p(\cdot, \theta_{t+\delta t})\|^2]$ , we find the drift term that minimizes the  $(\delta t)^2$  term of the same difference while holding the earlier diffusion term fixed. Note that the  $\delta t$  order term is minimized, not zeroed, so that we do not attain  $(\delta t)^2$  convergence.

As a bonus, we also minimize the order 1 Taylor expansion (in  $t$ ) of the norm of the expectation of the difference between the optimal filter  $p_{t+\delta t}$  and  $p(\cdot, \theta_{t+\delta t})$ , namely  $\|E[p_{t+\delta t} - p(\cdot, \theta_{t+\delta t})]\|$ .

To achieve  $(\delta t)^2$  convergence, rather than  $\delta t$  convergence, we will need the Ito-jet projection.

Finally, the expectation  $E_t$  is necessary because one should not forget that  $p$  and  $p(\cdot, \theta)$  are random objects. The randomness of  $p$ , in particular, comes from  $Y$  and the random  $\theta$  is supposed to capture it optimally.

### 3.4 Itô-jet projection

The Ito jet projection uses the notion of metric projection.

The metric projection of a general density  $p$  in  $L^2$  onto the manifold  $M$  is the closest point on  $M$  to  $p$  and is denoted by  $\pi(p)$ . This is not a vector projection, it is a projection of a point onto the submanifold  $M$ . Given that the metric projection is, according to the chosen metric, the best we can ever do in approximating  $p$  on  $M$ , as it is the closest point on  $M$  to  $p$ , we can try to find a projection filter that gets as close as possible to the metric projection. In other term, our criterion has changed to

$$\theta_{t+\delta t} \approx \operatorname{argmin}_{\theta} \|\pi(p_{t+\delta t}) - p(\cdot, \theta_{t+\delta t})\| \quad (7)$$

The Ito jet projection satisfies the following optimality criterion: it zeros the  $\delta t$  term and minimizes the  $(\delta t)^2$  term of the Taylor expansion of the mean square of the distance in  $L^2$  or  $M$  between  $\pi(p_{t+\delta t})$  and  $p(\cdot, \theta_{t+\delta t})$ . This is the most optimal projection we derived and it converges with order  $(\delta t)^2$ , as opposed to the  $(\delta t)^1$  of the Ito vector projection.

Again, in real applications we won't have the optimal filter at time  $t$  so we will start our approximation directly at time  $t = 0$ . This is reflected in the summary table 2 where  $t = 0$  and we call  $\delta t$  with the name  $t$ , assuming it is small.

### 3.5 Comparison of filters

In [10] we compare the different projection filters with each other in a case of cubic sensor perturbing a linear system (where, without perturbation, the Kalman filter would work well). In other words, the state equation is trivial,  $dX = dW$ , while the observation function is  $b(x) = x + \varepsilon x^3$ . For small  $\varepsilon$ , this will be close to a linear system and the extended Kalman filter and other Gaussian filters are supposed to perform well. We make the comparison in [10], comparing the different projection filters with the extended Kalman filter and with the Ito assumed density filter (ADF) with assumed Gaussian density. We refer to the paper for the full details.

In [10] we compare first the direct  $L^2$  residuals for the various methods. The Itô-vector projection in the direct  $L^2$  metric results in the lowest residuals over short time horizons. The Stratonovich projection comes a close second. Over medium time horizons, the Itô-jet projection out performs the Itô-vector projection. The projection methods out-performed all other methods like extended Kalman filter or assumed density filters.

Second, in [10] we compared the Hellinger residuals for different filters, where projection filters are w.r.t. the Hellinger metric. This second analysis indicates that the Itô ADF and the Itô-jet projection are almost indistinguishable in their performance, and we explain why in [10]. Over the short term, the Itô-vector projection gives the best results. Over medium term, the Itô-jet projection and the Itô ADF give the best results.

We also note that in previous works such as [19, 20, 6] where we only studied the Stratonovich projection filter, filtering problems for systems like the cubic and quadratic sensors were studied. For such systems, the optimal filter density would often turn out to be bimodal and a projection filter based on a manifold consisting of mixtures of two Gaussians or of exponential families with fourth order polynomial exponents would track the optimal filter well, while approximated filters such as the Extended Kalman filter, Gaussian Assumed Density filters and even particle filters with the same number of parameters as the projection filters would fare worse than the projection filters in terms of  $L^2$ , Hellinger or Lévy-Prokhorov norms of errors.

We may conclude that information geometry based filters contributed in a relevant way to finite dimensional approximations of the optimal filter.

## 4 Conclusions and further work

The notion of projecting a vector field onto a manifold is unambiguous. By contrast, there are multiple distinct generalizations of this notion to SDEs, as summarized in Table 2.

The two Itô projections we recalled in this review can both be derived from minimization arguments. However, the Itô-jet projection has some clear advantages.

- The Itô-jet projection is the best approximation to the metric projection of the true solution and leads to a mean-squared error of order  $O(t^2)$ . By contrast, the Itô-vector projection only tracks the true solution with an accuracy of  $O(t)$  for the mean-square error.
- The Itô-jet projection gives a more intuitive answer than the Itô-vector projection for the low dimensional example of the cross-diffusion considered in [10].
- The Itô-jet projection gives better numerical results in the longer term than the Itô-vector projection in our application to filtering.
- The Itô-jet projection has an elegant definition when written in terms of 2-jets, which is described in [10].

The Stratonovich projection satisfies an ad hoc minimization that is less appealing than the ones of the Itô projections, since it requires a deterministic anchor point at time 0 and negative time copies of the processes. The Itô-jet and Itô-vector projection arguments allow one to derive new Gaussian approximations to non-linear filters, and new exponential and mixture filters more generally, although the more general cases have not been explored in [10]. Some of the possibilities with different projections, metrics and manifolds are shown in Table 1. This could be investigated in further work to complete the table. In the Gaussian case we do explore in [10] applying the methods summarized in this review, unlike previous Gaussian approximations to non-linear filters, the projection approximations are derived by fully explicit minimization arguments rather than heuristic arguments. Thus, the notion of projecting an SDE onto a manifold, coupled with information geometry, is able to give new results even for this well-worn topic of approximate Gaussian nonlinear filtering.

A further important investigation line could be in deriving approximations based on approximating bases that are not made of densities or their square roots. Working with densities has the advantage of allowing information geometry to act clearly, but at the same time puts strong constraints on the approximating bases. As a simple example, one might wish to use “mixtures” of Hermite polynomials, which are not densities, as a basis for the approximation. One might wish to investigate to what extent it is possible to use non-density bases while retaining an information geometric approach.

The above development might potentially ease another fundamental problem that remains to this day: controlling the long term error of the projection filter compared to the optimal filter. This is a very difficult problem in general. Again in an information geometry setting, when the unobserved signal process  $X$  is a finite-state Markov chain, Cohen and Fausti (2023) [25] derive results on a well-controlled error, based on ergodic theory and symplectic structures. This result builds on their previous work [26]. The theory needs to be extended to the diffusion setting we have been using here, but this is a promising result in controlling the long term error between the optimal filter and the projection filter.

Projection	Optimality property
Itô-vector	<p>(i) Minimizes order 1 Taylor expansion (in <math>t</math>) of norm of the expectation of the difference between <math>p_t</math> &amp; <math>p(\cdot, \theta_t)</math>, namely <math>\ \mathbb{E}[p_t - p(\cdot, \theta_t)]\ </math></p> <p>(ii) Given the diffusion term in the approximating equation minimizing (but not zeroing) the <math>t</math> term of the expansion for the mean square difference <math>\mathbb{E}[\ p_t - p(\cdot, \theta_t)\ ^2]</math>, finds the drift term that minimizes the <math>t^2</math> term while holding that diffusion term fixed. Order <math>O(t)</math> convergence.</p>
Itô-jet	<p>Zeroes <math>t</math> term and minimizes <math>t^2</math> term of Taylor expansion of the mean square of the distance in <math>L^2</math> or <math>M</math> between <math>\pi(p_t)</math> and <math>p(\cdot, \theta_t)</math>. Order <math>O(t^2)</math> convergence.</p>
Stratonovich	<p>Similar to Itô vector but for the Taylor series of the “time-symmetric” mean square difference between <math>p</math> and its lower dimensional approximation <math>p(\cdot, \theta_t)</math>:</p> $\frac{1}{2} (\mathbb{E}[\ p_{-t} - p(\cdot, \theta_{-t})\ ^2] + \mathbb{E}[\ p_t - p(\cdot, \theta_t)\ ^2])$ <p>where negative time processes are defined ad hoc by propagating a second input Brownian motion backward in time.</p>

Table 2: Projections and the associated optimality criteria

## Statement on lack of conflict of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

## References

- [1] Nand Lal Aggarwal. Sur l'information de Fisher. In J. Kampé de Fériet and C. F. Picard, editors, *Théories de l'Information*, pages 111–117. Springer, Berlin Heidelberg, 1974.
- [2] Nasir Uddin Ahmed. *Linear and Nonlinear Filtering for Scientists and Engineers*. World Scientific, Singapore, 1998.
- [3] Shun'ichi Amari. *Differential Geometrical Methods in Statistics*, volume 28 of *Lecture Notes in Statistics*. Springer Verlag, 1985.
- [4] John Armstrong and Damiano Brigo. Stochastic filtering via L2 projection on mixture manifolds with computer algorithms and numerical examples. *arXiv preprint arXiv:1303.6236*, 2013.
- [5] John Armstrong and Damiano Brigo. Extrinsic projection of Itô SDEs on submanifolds with applications to non-linear filtering. In: *Nielsen, F., Critchley, F., & Dodson, K. (Eds), Computational Information Geometry for Image and Signal Processing*, Springer Verlag, 2016.
- [6] John Armstrong and Damiano Brigo. Nonlinear filtering via stochastic PDE projection on mixture manifolds in  $L^2$  direct metric. *Mathematics of Control, Signals and Systems*, 28(1):1–33, 2016.
- [7] John Armstrong and Damiano Brigo. Ito stochastic differential equations as 2-jets. In *Geometric Science of Information, Proceedings of the 2017 Conference, Paris, France*, pages 543–551. Springer, 2017.
- [8] John Armstrong and Damiano Brigo. Intrinsic stochastic differential equations as jets. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 474, 2018.
- [9] John Armstrong, Damiano Brigo, and Emilio Rossi Ferrucci. Projection of SDEs onto submanifolds. *arXiv preprint arXiv:1810.03923*, to appear in “*Information Geometry*”, 2023 special issue on half a century of information geometry, 2018.
- [10] John Armstrong, Damiano Brigo, and Emilio Rossi Ferrucci. Optimal approximation of sdes on submanifolds: the ito-vector and ito-jet projections. *Proceedings of the London Mathematical Society*, 119(1):176–213, 2019.
- [11] Babak Azimi-Sadjadi and P.S. Krishnaprasad. Approximate nonlinear filtering and its application in navigation. *Automatica*, 41(6):945–956, 2005.
- [12] Alan Bain and Dan Crisan. *Fundamentals of stochastic filtering*, volume 3. Springer, 2009.
- [13] Ole Barndorff-Nielsen. *Information and Exponential Families*. John Wiley & Sons, New York, 1978.
- [14] F. Berezantsev, Johan Hamberg, and John W.C. Robinson. Geometric aspects of nonlinear filtering. *Tech. rep., Swedish Defence Research Agency*, 2003.

- [15] Damiano Brigo. On the nice behaviour of the gaussian projection filter with small observation noise. *Systems & Control Letters*, 26(5):363–370, 1995.
- [16] Damiano Brigo. *Filtering by Projection on the Manifold of Exponential Densities*. Free University of Amsterdam, PhD dissertation, 1996.
- [17] Damiano Brigo. New results on the gaussian projection filter with small observation noise. *Systems & Control Letters*, 28(5):273–279, 1996.
- [18] Damiano Brigo and Bernard Hanzon. On some filtering problems arising in mathematical finance. *Insurance: Mathematics and Economics*, 22(1):53–64, 1998.
- [19] Damiano Brigo, Bernard Hanzon, and François LeGland. A differential geometric approach to nonlinear filtering: the projection filter. *IEEE Transactions on Automatic Control*, 43(2):247–252, 1998.
- [20] Damiano Brigo, Bernard Hanzon, and François LeGland. Approximate nonlinear filtering by projection on exponential manifolds of densities. *Bernoulli*, 5(3):495–534, 1999.
- [21] Damiano Brigo and Giovanni Pistone. Dimensionality reduction for measure-valued evolution equations in statistical manifolds. In *Proceedings of the conference on Computational Information Geometry for Image and Signal Processing*. Springer, 2016.
- [22] Damiano Brigo and Giovanni Pistone. Optimal approximations of the Fokker-Planck-Kolmogorov equation: projection, maximum likelihood eigenfunctions and Galerkin methods. *arXiv preprint arXiv:1603.04348*, 2016.
- [23] Jochen Bröcker and Ulrich Parlitz. Noise reduction and filtering of chaotic time series. *Proc. NOLTA 2000*, 2000.
- [24] Barry Cipra. Engineers look to Kalman Filtering for Guidance. *SIAM News*, 26, 1993.
- [25] Sam Cohen and Eliana Fausti. Stability and approximation of projection filters. *Seminar presented at Imperial College London, Department of Mathematics, April 18, 2023*.
- [26] Sam Cohen and Eliana Fausti. Exponential contractions and robustness for approximate Wonham filters. *arXiv preprint*, 2023.
- [27] David Elworthy. Geometric aspects of diffusions on manifolds. In *École d’Été de Probabilités de Saint-Flour XV–XVII, 1985–87*, pages 277–425. Springer, 1988.
- [28] David Elworthy, Yves Le Jan, and Xue-Mei Li. *The Geometry of Filtering*. Frontiers in Mathematics. Birkhauser, 2010.
- [29] Michel Emery. *Stochastic calculus in manifolds*. Springer-Verlag, Heidelberg, 1989.
- [30] Qing Gao, Guofeng Zhang, and Ian R. Petersen. An exponential quantum projection filter for open quantum systems. *Automatica*, 99:59–68, 2019.
- [31] B. Hanzon and R. Hut. New results on the Projection Filter. In *Proceedings of the European Control Conference, Grenoble, France*, pages 623–628. ECC91, 1991.
- [32] Bernard Hanzon. A differential-geometric approach to approximate nonlinear filtering. In C.T.J. Dodson, editor, *Geometrization of Statistical Theory*, pages 219–223. ULMD Publications, University of Lancaster, 1987.

- [33] Yuval Harel, Ron Meir, and Manfred Opper. A tractable approximation to optimal point process filtering: Application to neural encoding. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.
- [34] Michiel Hazewinkel, Steven I. Marcus, and H.J. Sussmann. Nonexistence of finite-dimensional filters for conditional statistics of the cubic sensor problem. *Systems & control letters*, 3(6):331–340, 1983.
- [35] Elton P. Hsu. *Stochastic Analysis on Manifolds*. Contemporary Mathematics. American Mathematical Society, 2002.
- [36] Andrew H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, New York, 1970.
- [37] Eagle S. Jones and Stefano Soatto. Visual-inertial navigation, mapping and localization: A scalable real-time causal approach. *The International Journal of Robotics Research*, 30(4):407–430, 2011.
- [38] Harold J. Kushner. Approximations to optimal nonlinear filters. *Automatic Control, IEEE Transactions on*, 12(5):546–556, 1967.
- [39] Anna Kutschireiter, Luke Rast, and Jan Drugowitsch. Projection filtering with observed state increments with applications in continuous-time circular filtering. *IEEE Transactions on Signal Processing*, 70, 2022.
- [40] Pierre F.J. Lermusiaux. Uncertainty estimation and prediction for interdisciplinary ocean dynamics. *Journal of Computational Physics*, 217(1):176–199, 2006.
- [41] R.S Liptser and A.N. Shiriyayev. *Statistics of Random Processes I, General Theory*. Springer Verlag, Berlin, 1978.
- [42] Yan Ma, Yu-xin Zhao, Li-juan Chen, and Chang Shuai. Hazard position estimation of sudden accidents based on projection filter. *Systems Engineering - Theory & Practice*, 35(3):651, 2015.
- [43] Peter S. Maybeck. *Stochastic models, estimation, and control*, volume 3. Academic press, 1982.
- [44] Nigel J. Newton. An infinite-dimensional statistical manifold modelled on Hilbert space. *J. Funct. Anal.*, 263(6):1661–1681, 2012.
- [45] Nigel J. Newton. Information geometric nonlinear filtering. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.*, 2(18):1550014, 24, 2015.
- [46] Étienne Pardoux. Filtrage non linéaire et équations aux dérivées partielles stochastiques associées. In *Ecole d’Eté de Probabilités de Saint-Flour XIX 1989*, pages 68–163. Springer, 1991.
- [47] Giovanni Pistone and Carlo Sempì. An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one. *Ann. Statist.*, 23(5):1543–1561, 10 1995.
- [48] C. R. Rao. Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.*, pages 81–91, 1945.



- [49] L. C. G. Rogers and David Williams. *Diffusions, Markov processes, and martingales. Vol. 2: Ito calculus*. Cambridge University Press, 1987.
- [50] Simone Surace and Jean-Pascal Pfister. Online maximum likelihood estimation of the parameters of partially observed diffusion processes. *IEEE Transactions on Automatic Control*, 64, 09 2017.
- [51] Filip Tronarp and Simo Särkkä. Updates in bayesian filtering by continuous projections on a manifold of densities. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5032–5036, 2019.
- [52] Ramon van Handel and Hideo Mabuchi. Quantum projection filter for a highly nonlinear model in cavity QED. *Journal of Optics B: Quantum and Semiclassical Optics*, 7, 2005.
- [53] M. H. Vellekoop and J. M. C. Clark. A nonlinear filtering approach to changepoint detection problems: Direct and differential-geometric methods. *SIAM Review*, 48(2):329–356, 2006.
- [54] Xian Miao Zhang, Rong Wu Wang, Hai Bo Wu, and Bugao Xu. Automated measurements of fiber diameters in melt-blown nonwovens. *Journal of Industrial Textiles*, 43(4):593–605, 2014.