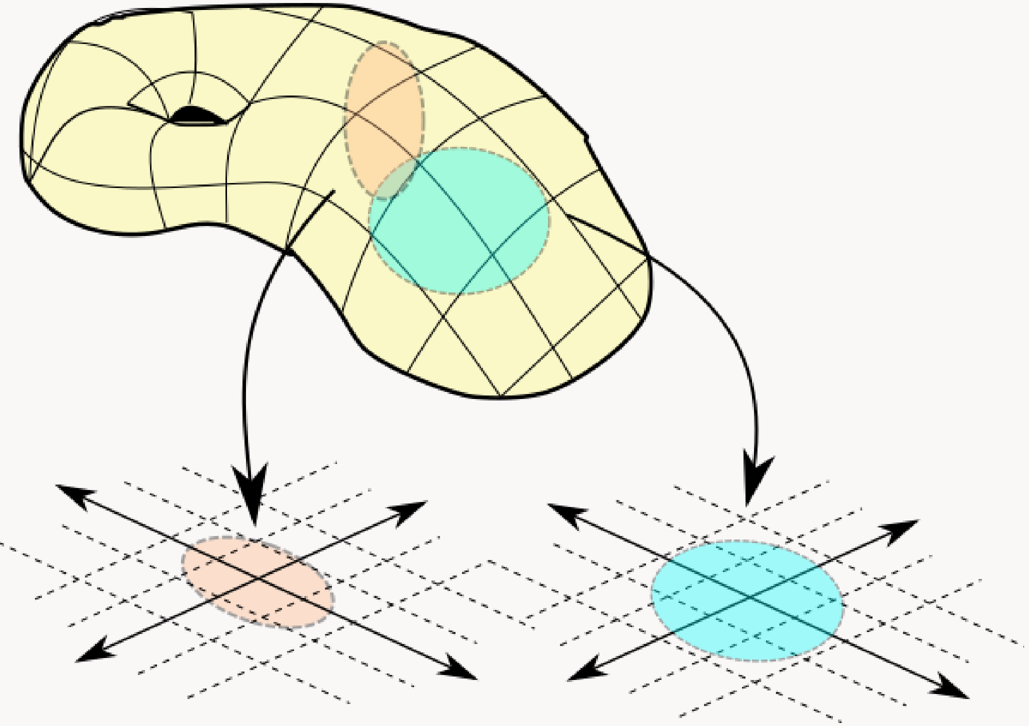


# Continuous Projection Filter

Introduction to Computational SDE

Muhammad F. Emzir
2022

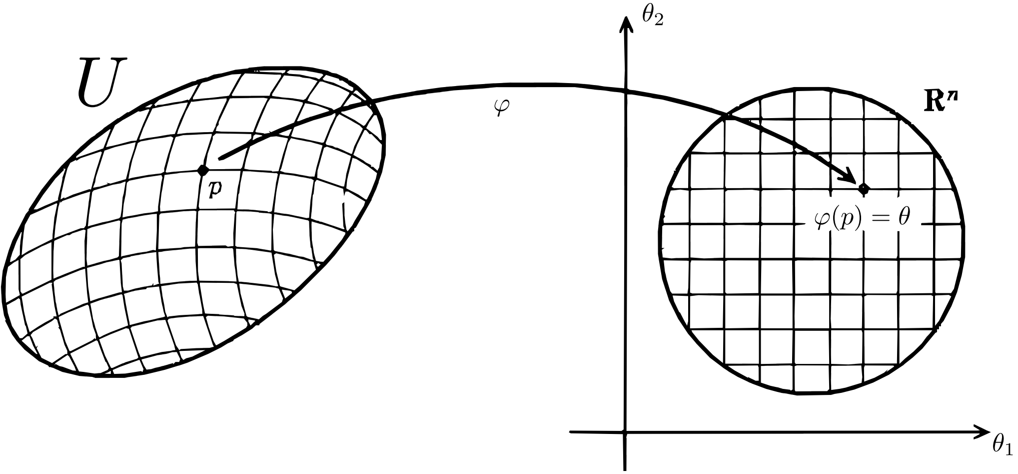
- Biography
- Background
- A Brief Overview of Differential Geometry
- Mathematical Setup
- The Nonlinear Filtering Equation
- The Projection Filtering Equation
- The Difficulties
- Recent Results
- Numerical Examples
- Questions ? 🙋

# Biography

- Muhammad F. Emzir.
- Academic Experience
  - Control and Instrumentation Engineering Dept., KFUPM, 2021- Now
  - Postdoc, Electrical Engineering and Automation Dept. Aalto, 2018-2021.
  - Postdoc, College of Engineering Computing and Computer Science, ANU, 2017-2018.
- Research interest:
  - Stochastic filtering, Nonlinear estimation and control, Quantum Control.

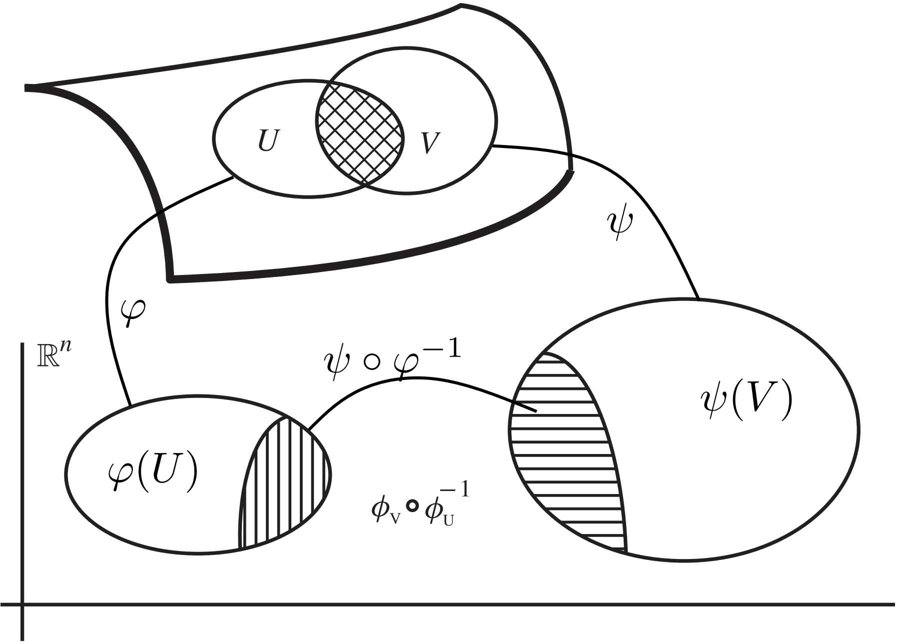
# Background

- Optimal filtering problems involve obtaining the best estimate of unobserved stochastic processes $x_t$ based on a record of noisy observation processes $y_t$.
- It is the equivalent to calculating the evolution of the conditional probability density of the state given the observation up to the present time.
- Under certain regularity assumptions, the evolution of generic nonlinear dynamics with nonlinear observations may be represented by a stochastic partial differential equation (SPDE) known as the Stratonovich-Kushner (SK) equation.
- The SK equation is difficult to solve since it is a nonlinear SPDE with a complex structure.
- Zakai presented an alternative formulation of the SK equation, resulting in a linear SPDE for the conditional density (unnormalized).
- Even though the Zakai equation is linear, it is nonetheless difficult to solve. In the majority of applications, only approximate solutions are available.

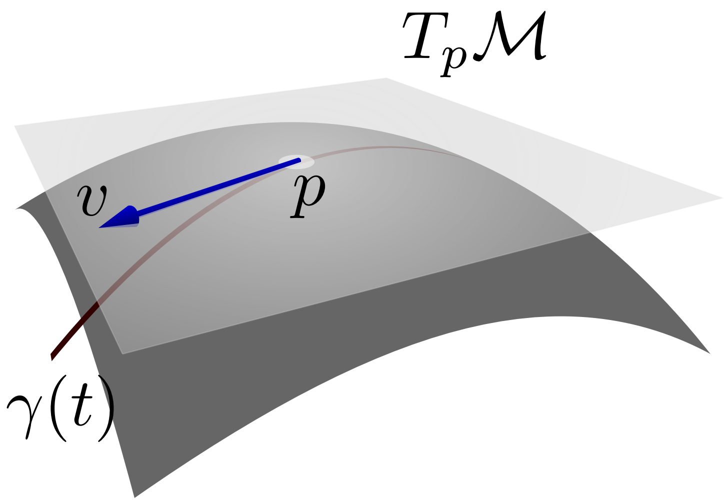



# A Brief Overview of Differential Geometry

- A manifold $\mathcal{M}$ is a set that is locally "similar" to $\mathbb{R}^n$.
- For an open subset $U \subset \mathcal{M}$, we have a bijection $\varphi$ that maps $U$ to an open subset of $\mathbb{R}^n$. $\varphi(p) = [\theta_1(p), \cdots, \theta_n(p)]$ is called coordinate systems for $U$ and $\theta_i : U \to \mathbb{R}$ is called the coordinate function.
- We call $(U, \varphi)$ a chart.




- Suppose $\psi(p) = [\rho_1(p), \cdots, \rho_n(p)]$ is another coordinate systems for an open set $V \subset \mathcal{M}$ where $U \bigcap V \neq \emptyset$. Then on $U \bigcap V$, $\psi \circ \varphi^{-1}$ and $\varphi \circ \psi^{-1}$ are maps $\mathbb{R}^n \to \mathbb{R}^n$.
- The manifold $\mathcal{M}$ is invariant against the change of coordinate system. Concepts like distance, differentiability, volume, e.t.c, should not depend on the choice of local coordinate systems.
- We said that a topological space $\mathcal{M}$ with an atlas $\{(U_\alpha, \varphi_\alpha)\}$, (such that $\bigcup_\alpha U_\alpha = \mathcal{M}$), is a differentiable manifold if:
  - $\varphi_\alpha$ is bijection from $U_\alpha$ to subset of $\mathbb{R}^n$,
  - for any $\alpha, \beta$ such that $U_\alpha \bigcap U_\beta \neq \emptyset$, $\varphi_\alpha \circ \varphi_\beta^{-1}$ is a diffeomorphism.
- For simplicity, we consider $C^\infty$ diffeomorphism, hence $C^\infty$ (smooth) manifold.




## Tangent Space

- Let $f : \mathcal{M} \to \mathbb{R}$. If for any $p \in \mathcal{M}$, $f \circ \varphi^{-1}$ is $k$-th differentiable, then we say $f \in C^k(\mathcal{M})$, where we write $\left(\frac{\partial}{\partial \theta_i}\right)_p f := \left(\frac{\partial(f \circ \varphi^{-1})}{\partial \theta_i}\right)_{\varphi(p)}$.
- We define a tangent vector $v$ at a point $p$ via curves on $\mathcal{M}$ that passes $p$. It is the velocity of a curve $\gamma(t) \subset \mathcal{M}$ that passes $p$; i.e., $\dot{\gamma}(0) = v$ assuming that $\gamma(0) = p$.
- A tangent vector at a point $p$, $v : f \mapsto v(f) \in \mathbb{R}$ is a linear operator and satisfies Leibniz identity:
  - For $a, b \in \mathbb{R}$, $v(af + bg) = av(f) + bv(g)$.
  - $v(fg) = f(p)v(g) + v(f)g(p)$.

- Set of all tangent vectors at $p$ is called as tangent space at $p$, $T_p\mathcal{M}$, and if $\mathcal{M}$ is $n$ dimensional, then so is $T_p\mathcal{M}$.
- $T_p\mathcal{M}$ is spanned by $\{\frac{\partial}{\partial\theta_i}\}$: this can be seen from $\left(\frac{\partial}{\partial\theta_i}\right)_p \theta_j = \left(\frac{\partial(\theta_j \circ \varphi^{-1})}{\partial\theta_i}\right)_{\varphi(p)} = \delta_{i,j}$.
- Any $v \in T_p\mathcal{M}$ can be expressed as $v(f) = \sum_{i=1}^n v_i \left(\frac{\partial}{\partial\theta_i}\right)_p f$.
- We can define an inner product on $T_p\mathcal{M}$, $\langle v, w\rangle$, where $g_{ij} = \left\langle \left(\frac{\partial}{\partial\theta_i}\right)_p, \left(\frac{\partial}{\partial\theta_j}\right)_p \right\rangle$ is known as the metric tensor.
- In information geometry, $\mathcal{M}$ is related to a set of parametric probability distributions and we normally use single chart $(\mathcal{M}, \varphi)$.
- For more details, please consult Amari & Nagaoka, 2000 and other references in differential geometry.

# Mathematical Setup

- Let $\lambda$ be the Lebesque measure on $\mathbb{R}^{d_x}$, and define $\mathscr{M}$ to be the set of all non-negative, finite measure $\mu$ absolutely continuous w.r.t $\lambda$, whose density is $\lambda$ positive a.e.
- For any density $p$ on $\mathbb{R}^{d_x}$, $\mathbb{E}_p[\cdot]$ denote the expectation w.r.t. $p$.
- Set $\mathscr{H} = \{p = d\mu/d\lambda : \mu \in \mathscr{M}\}$. If $p \in \mathscr{H}$ then $p \in L^1$ and $\sqrt{p} \in L^2$. Hence $\mathscr{R} = \{\sqrt{p} : p \in \mathscr{H}\} \subset L^2$, is a metric space with $d(\sqrt{p}, \sqrt{q}) = \|\sqrt{p} - \sqrt{q}\|$
- There exists $m$-dimensional submanifold $N$ of $L^2$ which is contained in $\mathscr{R}$. Since $N$ is a manifold, $\forall \sqrt{p} \in N \subset \mathscr{R}$, $\exists$ an open neighborhood $S^{1/2}$ of $\sqrt{p}$ in $N$ and diffeomorphism $\varphi : S^{1/2} \to \Theta \subset \mathbb{R}^m$ such that $\varphi^{-1}$ is a differentiable mapping from $\Theta$ to $L^2$; i.e., $\varphi^{-1} : \Theta \to S^{1/2}$, and $\theta \mapsto \sqrt{p(\cdot, \theta)}$.
- We associate $S^{1/2}$ to the set of parametric density $S = \{p(\cdot, \theta) : \theta \in \Theta \subseteq \mathbb{R}^m\}$.

- Since $(S^{1/2}, \varphi)$ is a chart on $N$, then $\left\{ \left( \frac{\partial}{\partial \theta_i} \right)_{\sqrt{p_\theta}}, \cdots, \left( \frac{\partial}{\partial \theta_m} \right)_{\sqrt{p_\theta}} \right\}$ is the linear independent basis vectors in $L^2$.
- The tangent vector space to $S^{1/2}$ at $\sqrt{p(\cdot, \theta)}$ is given by

$$T_{\sqrt{p(\cdot,\theta)}} S^{1/2} = \mathrm{span} \left\{ \left( \frac{\partial}{\partial \theta_i} \right)_{\sqrt{p_\theta}}, \cdots, \left( \frac{\partial}{\partial \theta_m} \right)_{\sqrt{p_\theta}} \right\}. \quad (1)$$

- We define their inner product

$$g_{ij}(\theta) = \left\langle \left( \frac{\partial}{\partial \theta_i} \right)_{\sqrt{p_\theta}}, \left( \frac{\partial}{\partial \theta_j} \right)_{\sqrt{p_\theta}} \right\rangle = \frac{1}{4} \int \frac{1}{p(x,\theta)} \frac{\partial p(\cdot,\theta)}{\partial \theta_i} \frac{\partial p(\cdot,\theta)}{\partial \theta_j} dx = \frac{1}{4} I_{ij}(\theta)$$

The matrix $g_{ij}$ is the Riemannian metric on $S^{1/2}$ where $I_{ij}$ coincides with the Fisher metric.

- The bases in $(1)$ are generally not orthogonal. Using $(1)$ we can define linear projection $\Pi_\theta : L^2 \to T_{\sqrt{p(\cdot,\theta)}} S^{1/2}$ as below:

$$\Pi_\theta v = \sum_i \left( \sum_j 4I^{ij} \left\langle v, \left(\frac{\partial}{\partial \theta_j}\right)_{\sqrt{p_\theta}} \right\rangle \right) \left(\frac{\partial}{\partial \theta_i}\right)_{\sqrt{p_\theta}}$$

- For a parametric density $p_\theta$, we write $\mathbb{E}_\theta := \mathbb{E}_{p_\theta}$.
- If $\mathbb{E}_\theta[|u|^2] < \infty$ then $v = u\sqrt{p_\theta} \in L^2$, we can explicitly write

$$\Pi_\theta v = \sum_i \left( \sum_j 4I^{ij} \mathbb{E}_\theta \left( u \frac{\partial \log p_\theta}{\partial \theta_j} \right) \right) \left(\frac{\partial}{\partial \theta_i}\right)_{\sqrt{p_\theta}} .$$

See (Brigo, 1999, Lemma 2.1).

## The Exponential Families

- **Definition** : Let $\{c_i\}$ be scalar measurable function on $\mathbb{R}^{d_x}$ such that $\{1, c_1, \ldots, c_m\}$ are linearly independent, and

$$\Theta_o := \left\{\theta \in \mathbb{R}^m : \psi(\theta) = \log \int \exp(c(x)^\top \theta) dx < \infty \right\}$$

has non-empty interior. Then

$$\mathrm{EM}(c) := \{p(\cdot, \theta) = \exp(c(x)^\top \theta - \psi(\theta)), \theta \in \Theta\},$$

is called an exponential family of probability densities, where $\Theta \subset \Theta_o$ open.

- On exponential family, $\eta_i := \mathbb{E}_\theta[c_i] = \frac{\partial \psi(\theta)}{\partial \theta_i}$, $I_{ij} = \frac{\partial^2 \psi(\theta)}{\partial \theta_i \partial \theta_j}$

# The Nonlinear Filtering Equation

- Let $(\Omega, \mathscr{F}, \mathbb{P})$ be standard probability space with filtration $\{\mathscr{F}_t\}$. Consider the following Itô processes for state and observation dynamics:

$$
\begin{aligned}
dx_t =& f(x_t, t)dt + \sigma(x_t, t)dW_t, \quad (2a)\\
dy_t =& h(x_t, t)dt + dV_t. \quad (2b)
\end{aligned}
$$

- The nonlinear filtering problem consists in finding $\pi_t$ the conditional probability distribution of the state $x_t$ given the obserbvation up to time $t$; $\pi_t(g) := \int g(x)\mathbb{P}[dx|\mathscr{Y}_t] = \mathbb{E}[g|\mathscr{Y}_t]$, where $\mathscr{Y}_t := \sigma(y_s, 0 \leq s \leq t)$.

- Under some regularity condition, $\{\pi_t\}$ satisfies SK equation:

$$\pi_t(g) = \pi_0(g) + \int_0^t \pi_s(\mathscr{L}_s g)ds$$
$$+ \sum_{i=1}^{d_y} \int_0^t (\pi_s(h_{s,k} g) - \pi_s(h_{s,k})\pi_s(g)) (dy_{s,k} - \pi_s(h_{s,k})ds). \quad \text{(SK)}$$

where $\mathscr{L}_t$ is the backward diffusion operator.

- The conditional expectation $\pi_t$ corresponds to density $p_t$, which can be written in the Stratonovich form as

$$dp_t = \mathscr{L}_t^* p_t - \frac{1}{2} p_t \left(h_t^\top h_t - \mathbb{E}_{p_t}\left[h_t^\top h_t\right]\right) dt + \sum_{i=1}^{d_y} p_t \left(h_{t,k} - \mathbb{E}_{p_t}\left[h_{t,k}\right]\right) \bullet dy_{t,k}. \quad (3)$$

# The Projection Filtering Equation

- The projection filter for the family $S$ is defined as the projection of $d\sqrt{p_t}$ obtained form (3) onto the tangent space $T_{\sqrt{p(\cdot,\theta)}}S^{1/2}$

$$d\sqrt{p_t} = \frac{1}{2}\sqrt{p_t}\left(\frac{\mathscr{L}_t^* p_t}{p_t} - \frac{1}{2}\left(h_t^\top h_t - \mathbb{E}_{p_t}\left[h_t^\top h_t\right]\right)\right) dt + \frac{1}{2}\sqrt{p_t}\left[h - \mathbb{E}_{p_t}[h_t]\right]^\top \bullet dy_t$$

- By requiring that for any $u \in \{\frac{\mathscr{L}_t^* p_t}{p_t}, h_t^\top h_t, h_{k,t}\}$, $\sup \mathbb{E}_\theta[u^2] < \infty$, the right hand side belongs to $L^2$ for any $p_t = p_\theta$.
- Therefore, we can consider $d\sqrt{p_\theta}$ as a vector in $L^2$, and we can project this vector to $T_{\sqrt{p_\theta}}S^{1/2}$ to obtain $\Pi(d\sqrt{p_\theta}) = \sum_i \frac{\partial \sqrt{p_\theta}}{\partial \theta_i} \bullet d\theta_{i,t}$, where

$$d\theta_t = I(\theta_t)^{-1}\mathbb{E}_{\theta_t}\left[\left(\frac{\mathscr{L}_t^* p_{\theta_t}}{p_{\theta_t}} - \frac{1}{2}h_t^\top h_t\right)\frac{\partial \log p_{\theta_t}}{\partial \theta_t}\right] dt + I(\theta_t)^{-1}\sum_{i=1}^{d_y}\mathbb{E}_{\theta_t}\left[h_{k,t}\frac{\partial \log p_{\theta_t}}{\partial \theta_t}\right] \bullet dy_{k,t}$$

## The Projection Filtering Equation for $\mathrm{EM}(c)$

- For the exponential family $\mathrm{EM}(c)$, the parameter $\theta_t$ of the density $p(\cdot, \theta_t)$ satisfies:

$$d\theta_t = I(\theta_t)^{-1}\mathbb{E}_{\theta_t}\left[\mathscr{L}_t\left[c\right] - \frac{1}{2}h_t^\top h_t\left[c - \eta(\theta_t)\right]\right]dt$$

$$+ I(\theta_t)^{-1}\sum_{k=1}^{d_y}\mathbb{E}_{\theta_t}\left[h_{t,k}\left[c - \eta(\theta_t)\right]\right] \bullet dy_{t,k}.$$

## The Difficulties

- The filter equation for exponential family is model-specific and depends on the natural statistics used. One needs to write down the filtering equation analytically by hand.
- The log partition function and several expected statistics values need to be calculated every step via numerical integration.
- The projection filter numerical implementation relies upon a recursion procedure to compute some expectations and the Fisher information matrix (Brigo, 1995). This recursion is, in general, only feasible for unidimensional problems.

# Recent Results

- Emzir et al., 2022 use the first Chebyshev polynomial expansion for unidimensional problems and sparse-grid integrations for higher dimensional problems to compute $\exp(\psi)$.
- The numerical integration is only required to calculate the log partition function; the expectations and Fisher metric are calculated using auto-differentiations.
- Some assumptions need to be made: state variables and the measurement SDEs have diffusion and drift parts represented by polynomials. Then, choosing natural statistics as monomials simplifies the SDEs for the parameters significantly.
- This configuration permits the automated implementation of projection filters. The derivation of the filter equation by hand is no longer required.
- At each time:
    - Compute $\exp(\psi^N(\theta_t))$ via sparse-grid quadrature and a bijection $\phi : (-1, 1)^{d_x} \rightarrow \mathbb{R}^{d_x}$.
    - Compute $\mathbb{E}_{\theta_t}[c]$ and $I$ via partial differentiation of $\psi^{(N)}(\theta_t)$
    - Compute $d\theta_t$.

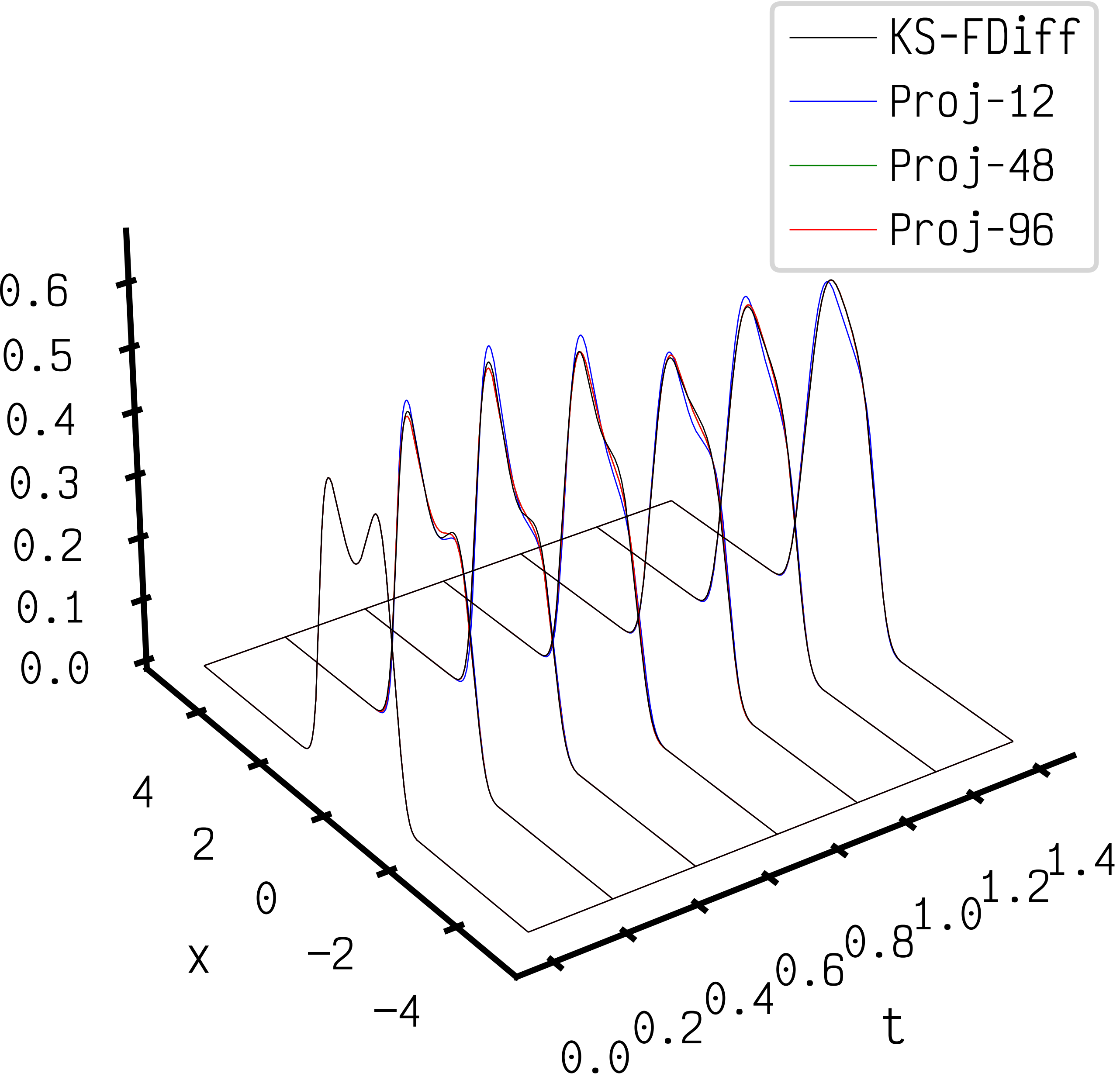
# Numerical Examples

## Unidimensional

- Consider

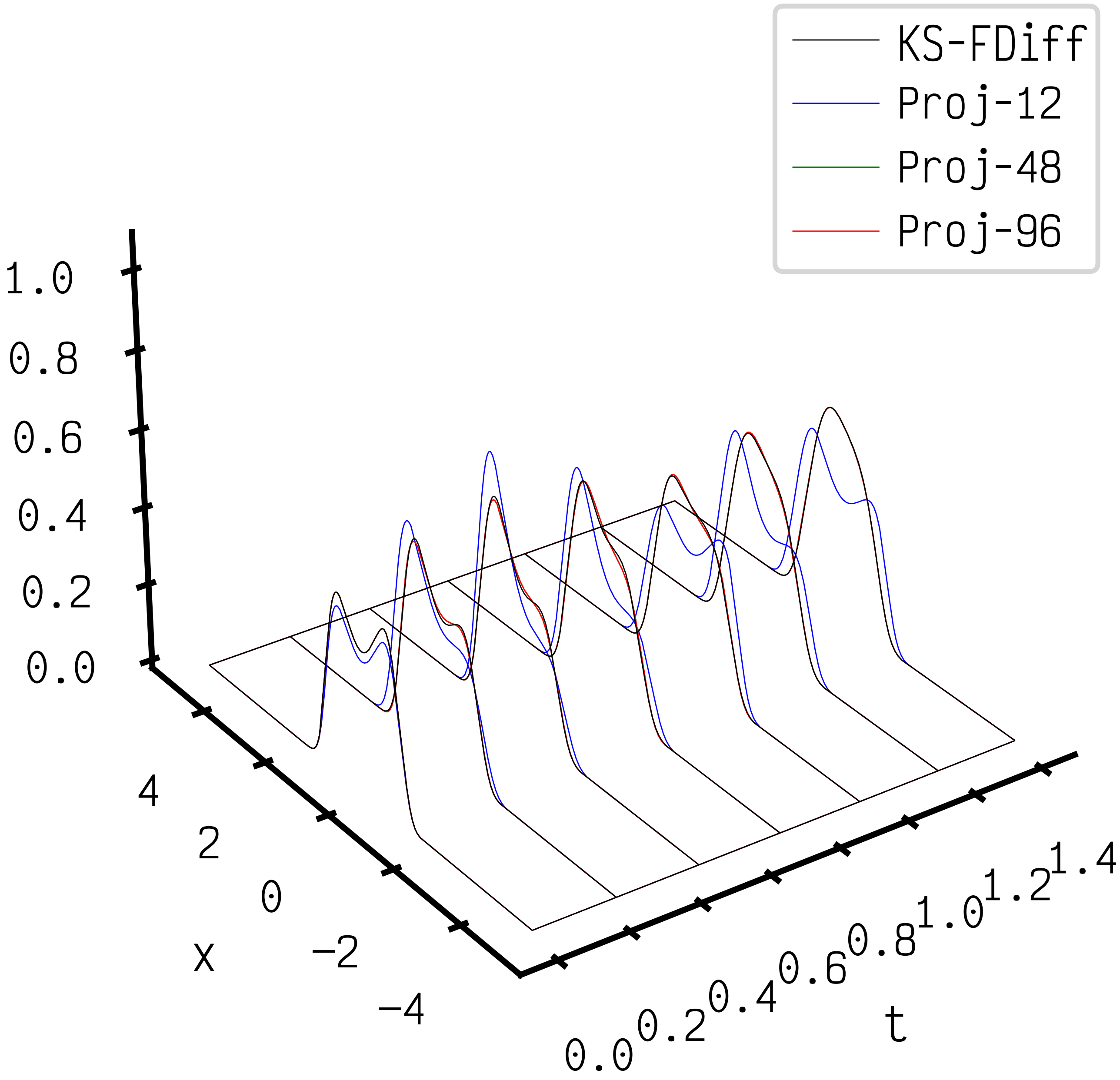
$$
\begin{aligned}
dx_t =& \sigma dW_t, \\
dy_t =& \beta x_t^3 dt + dV_t
\end{aligned}
$$

- We use the exponential manifold with $c_i \in \{x, x^2, x^3, x^4\}$ with $\theta_0 = [0, 1, 0, -1]$.
- We compare two choices of the bijections, the first is $\tanh^{-1}(\tilde{x})$ and the second is $\phi = \frac{\tilde{x}}{1-\tilde{x}^2}$
.




Densities obtained using finite difference scheme, and projection filter using $\tanh^{-1}$.

Densities obtained using finite difference scheme, and projection filter using $\phi = \frac{\tilde{x}}{1-\tilde{x}^2}$.

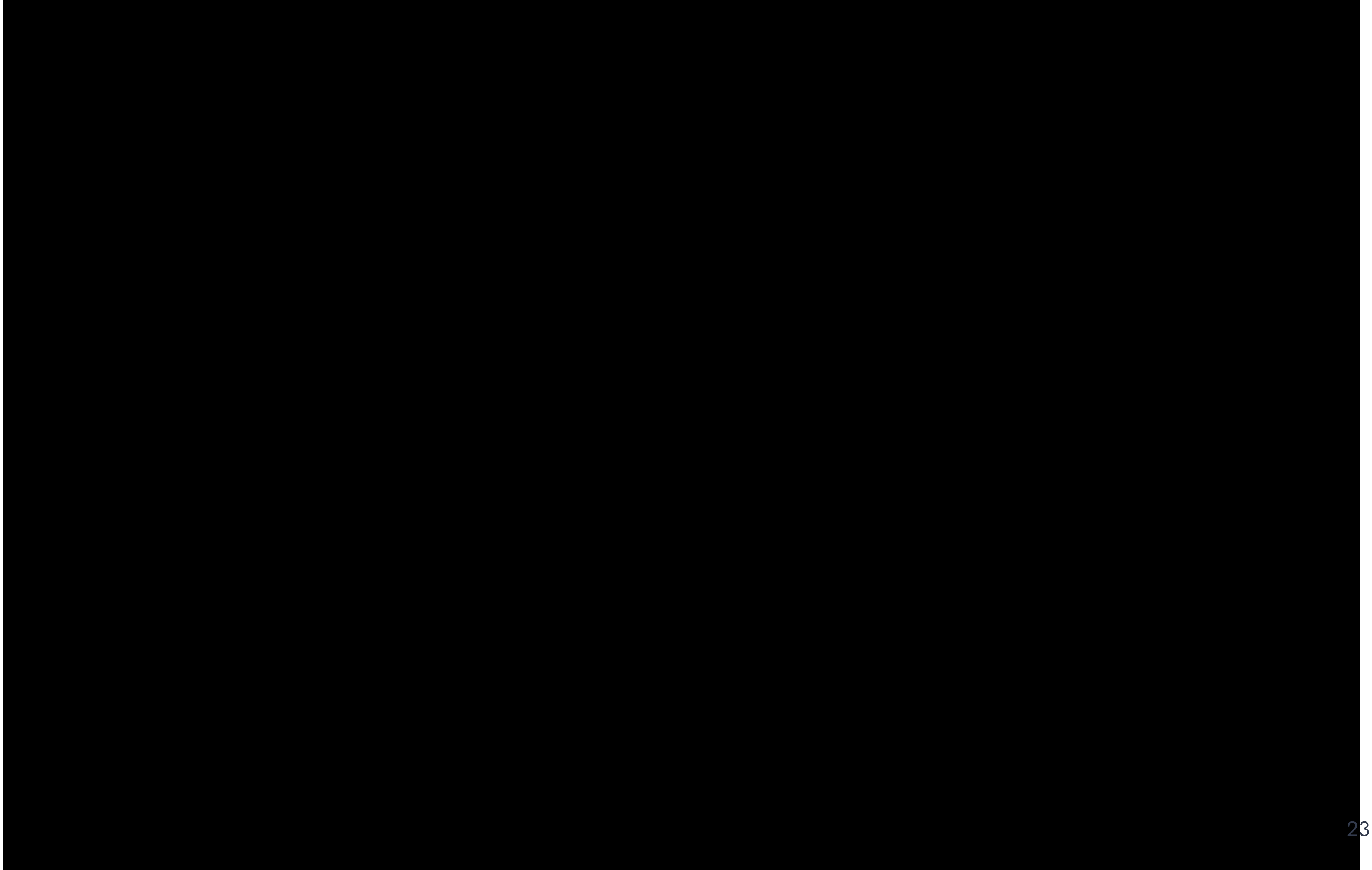


## Multidimensional

- The dynamic model considered is the Van-der-Pol oscillator:

$$
\begin{aligned}
d\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} &= \begin{bmatrix} \kappa x_{1,t} + x_{2,t} \\ \mu(1 - x_{1,t}^2)x_{2,t} - x_{1,t} + \kappa x_{2,t} \end{bmatrix} dt + \begin{bmatrix} 0 \\ \sigma_w \end{bmatrix} dW_t, \\
dY &= x_{1,t} dt + \sigma_v dV_t.
\end{aligned}
$$

- $\mu = 0.3, \kappa = 1$, and $\sigma_v = \sigma_w = 1$. We also set $dt = 2.5 \times 10^{-4}$. We use sparse-grid integration where we set the level equals to 4.
- We compare with the result with empirical density from the particle filter with $1.6 \times 10^6$ samples.



# Questions ? 🙋

## References


- Aubin, T., 2001. **A course in differential geometry**, Graduate studies in mathematics. American Mathematical Society, Providence, R.I.
- Frankel, T., 2012. **The geometry of physics: an introduction**, 3rd ed. ed. Cambridge University Press, Cambridge ; New York.
- Lee, J.M., 2013. **Introduction to smooth manifolds**, 2nd ed. ed, Graduate texts in mathematics. Springer, New York ; London.
- Amari, S., Nagaoka, H., 2000. **Methods of information geometry**, Translations of mathematical monographs. American Mathematical Society, Providence, RI.
- Brigo, D., Hanzon, B., Gland, F.L., 1999. *Approximate nonlinear filtering by projection on exponential manifolds of densities*. **Bernoulli** 5, 495.
- Emzir, M.F., Zhao, Z., Särkkä, S., 2023. *Multidimensional projection filters via automatic differentiation and sparse-grid integration*. **Signal Processing** 204, 108832.
- Emzir, M.F., Zhao, Z., *Sparse-Grid-AutoDiff-Projection-Filters*, GitHub Repository, https://github.com/puat133/Sparse-Grid-Autodiff-Projection-Filters