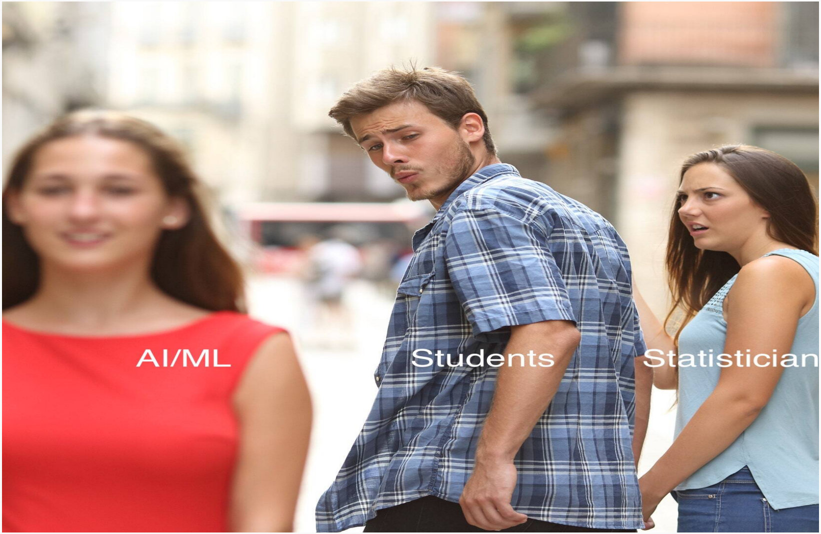
School of Mathematical and Statistical Sciences
Arizona State University

Flexible Bayesian Learning with Q-Exponential Processes beyond Gaussian

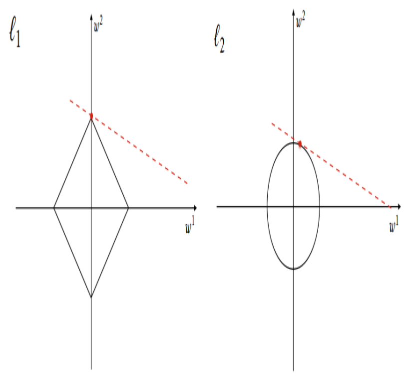
Shiwei Lan*
slan@asu.edu

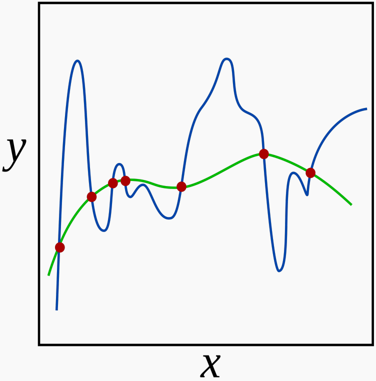
MAR 06, 2026 @ ICERM

AI vs Statistics
Compete or Collaborate?



Regularization





Nonparametric Modeling on Functional Data

Gaussian Process (GP): flexible prior on the space of candidate functions

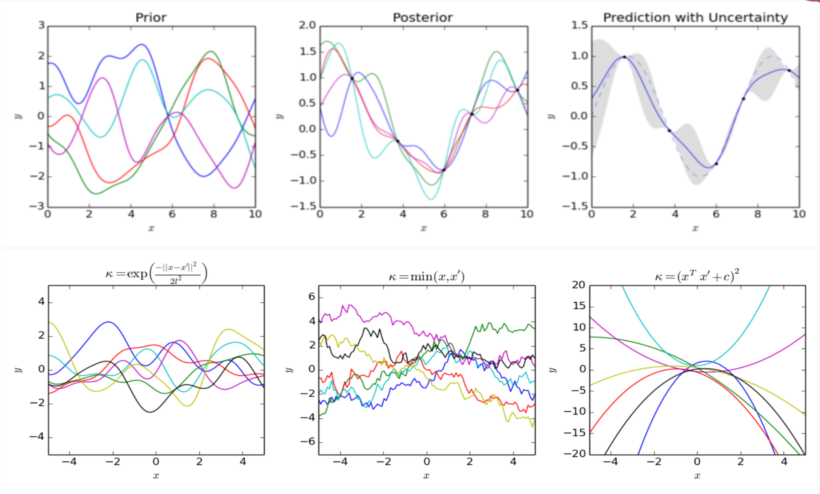


Figure: Top: Gaussian Process Regression; Bottom: GP Priors.

Modeling Derivatives

Is Gaussian process (GP) optimal?

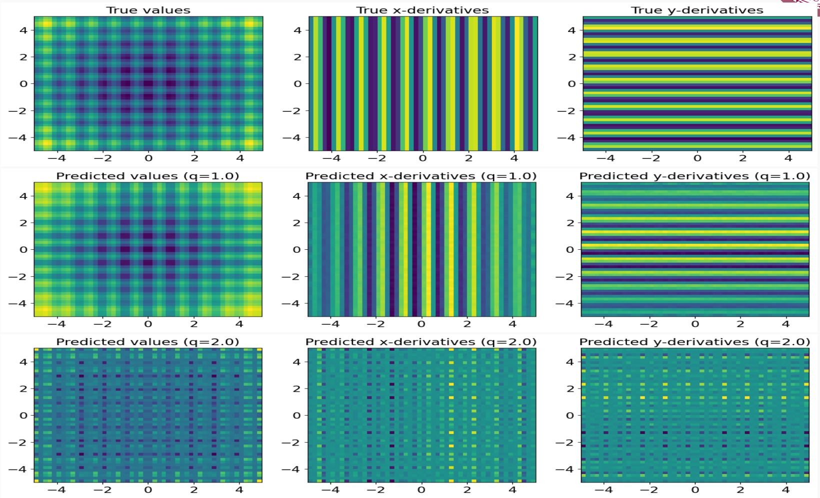
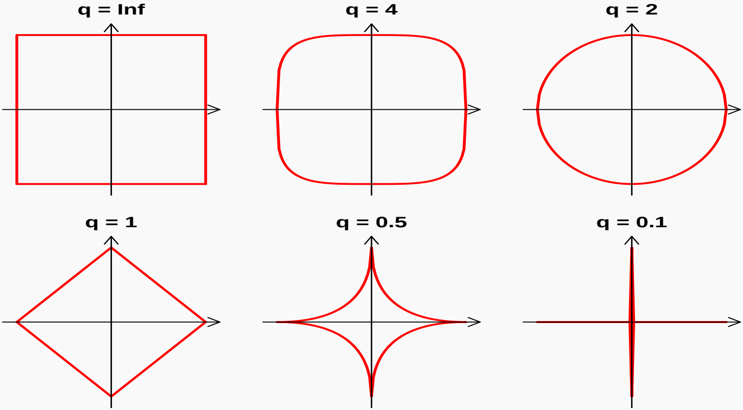
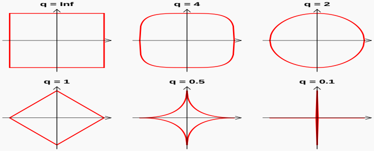


Figure: Contrasting Q-EP ($q = 1.0$, middle) with GP ($q = 2.0$, bottom) against the truth (top) for modeling function values and derivatives in the Rastrigin example.

Regularization
with different degrees



Bayesian Learning via Q-Exponential Process[1]



[1] Shuyi Li, Michael OConnor and Shiwei Lan*, *NIPS*2023

Regularization on Function Spaces

- Regularization is one of the most fundamental topics in optimization, statistics and machine learning.
- To get sparsity in estimating a parameter $u \in \mathbb{R}^d$, an ℓ_q penalty term, $\|u\|_q$, is usually added to the objective function.

Regularization on Function Spaces

- Regularization is one of the most fundamental topics in optimization, statistics and machine learning.
- To get sparsity in estimating a parameter $u \in \mathbb{R}^d$, an ℓ_q penalty term, $\|u\|_q$, is usually added to the objective function.
- What is the probabilistic distribution corresponding to such ℓ_q penalty?

Regularization on Function Spaces

- Regularization is one of the most fundamental topics in optimization, statistics and machine learning.
- To get sparsity in estimating a parameter $u \in \mathbb{R}^d$, an ℓ_q penalty term, $\|u\|_q$, is usually added to the objective function.
- What is the probabilistic distribution corresponding to such ℓ_q penalty?
- What is the *correct* stochastic process corresponding to $\|u\|_q$ when we model functions $u \in L^q$?
- This is important for statistically modeling high-dimensional objects such as images, with penalty to preserve certain properties, e.g. edges in the image.

Regularization on Function Spaces

literature review

- Gaussian process (GP) can be viewed as L_2 regularization on function spaces, sometimes oversmooth [39, 18].
- L_1 penalty based priors include Laplace random field [36, 28, 20] and Besov process (BP) [21, 10, 19, 11].
- Student-t process [42] and elliptical process [1] with heavy tail are proposed as alternatives to GP.
- We propose the *q-exponential process (Q-EP)* based on q-exponential distribution $p(u) \propto \exp\left(-\frac{1}{2}|u|^q\right)$ for random coefficients of BP.
  1. We propose a novel stochastic process Q-EP that includes GP as a special case ($q = 2$).
  2. Q-EP can be regarded as a probabilistic definition of BP with direct configuration of correlation and explicit prediction.
  3. We provide a white-noise representation for Q-EP prior for efficient inference.

Regularization in Image Reconstruction

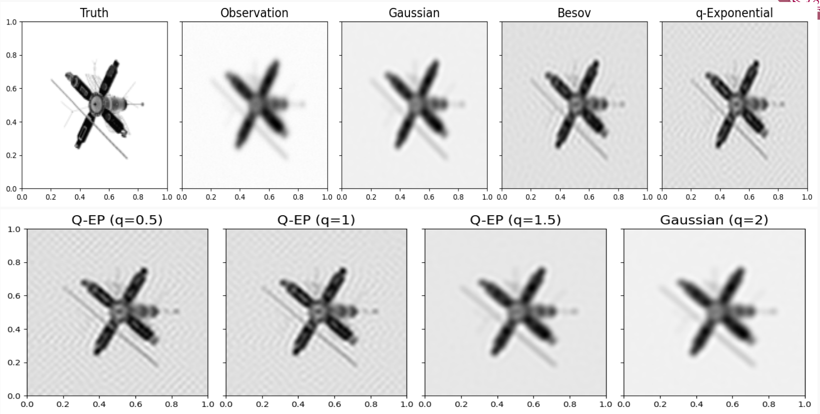


Figure: Image of satellite. Top: true image, blurred observation, and reconstructions by GP, Besov and Q-EP models with relative errors 75.19%, 21.94% and 20.35% respectively. Bottom: MAP estimates by Q-EP with varying q parameters.

Besov Process and Q-exponential Distribution

- Besov process [21, 10] is proposed to impose L_q regularization as an "edge-preserving" prior ($q = 1$) for images:

$$u(x) = \sum_{\ell=1}^{\infty} \gamma_\ell u_\ell \phi_\ell(x), \quad u_\ell \overset{iid}{\sim} \pi_q(\cdot) \propto \exp\left(-\frac{1}{2}|\cdot|^q\right) \tag{1}$$

- This q-exponential distribution $\pi_q(\cdot)$ is a special case of the *exponential power (EP)* distribution $\mathrm{EP}(\mu, \sigma, q)$ with $\mu = 0$, $\sigma = 1$:

$$p(u|\mu, \sigma, q) = \frac{q}{2^{1+1/q}\sigma\Gamma(1/q)} \exp\left\{-\frac{1}{2}\left|\frac{u-\mu}{\sigma}\right|^q\right\} \tag{2}$$

- When $q = 2$, this is Gaussian distribution.
- When $q = 1$, this becomes Laplace distribution.

Generalization of Q-exponential Distribution

- How can we generalize $\pi_q(\cdot)$ (EP(μ, σ, q)) to a *multivariate distribution* and further to a *stochastic process*?
- By the Kolmogorov' extension theorem [32], one should require
  1. **exchangeability** of the joint distribution, i.e. $p(\boldsymbol{\xi}_{1:J}) = p(\boldsymbol{\xi}_{\tau(1:J)})$ for any finite permutation τ;
  2. **consistency** of marginalization, i.e. $p(\boldsymbol{\xi}_1) = \int p(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) d\boldsymbol{\xi}_2$.
- Gomez [14] provided one possibility of a multivariate EP distribution, denoted as EP$_d(\boldsymbol{\mu}, \mathbf{C}, q)$, with the following density:

$$p(\mathbf{u}|\boldsymbol{\mu}, \mathbf{C}, q) = \frac{q\Gamma(\frac{d}{2})}{2\Gamma(\frac{d}{q})} 2^{-\frac{d}{q}} \pi^{-\frac{d}{2}} |\mathbf{C}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[(\mathbf{u}-\boldsymbol{\mu})^{\mathsf{T}}\mathbf{C}^{-1}(\mathbf{u}-\boldsymbol{\mu})\right]^{\frac{q}{2}}\right\} \tag{3}$$

Generalization of Q-exponential Distribution

marginalization consistency

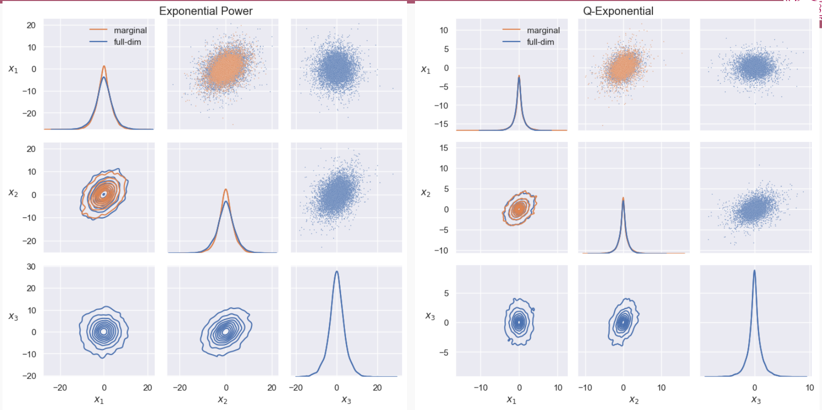


Figure: Inconsistent (Gomez's) EP distribution $\mathrm{EP}_d(\boldsymbol{\mu}, \mathbf{C}, q)$ (left) vs. consistent Q-exponential distribution $\mathrm{q-ED}_d(\boldsymbol{\mu}, \mathbf{C})$ (right). Both can be sampled using $\mathbf{u} \overset{d}{=} \boldsymbol{\mu} + R\mathbf{L}S$ with $R^q \sim \Gamma(\alpha = \frac{d}{q}, \beta = \frac{1}{2})$ and $R^q \sim \Gamma\left(\alpha = \frac{d}{2}, \beta = \frac{1}{2}\right)$ respectively. Note there is significant discrepancy between the marginalization of $\mathrm{EP}_3(\boldsymbol{\mu}, \mathbf{C}, q)$ and $\mathrm{EP}_2(\boldsymbol{\mu}, \mathbf{C}, q)$. However, the marginalization of $\mathrm{q-ED}_3(\boldsymbol{\mu}, \mathbf{C})$ coincides with $\mathrm{q-ED}_2(\boldsymbol{\mu}, \mathbf{C})$. Empirical densities are estimated based on 10000 samples (shown as dots).

Q-Exponential Process
consistent generalization

Definition

A multivariate q-exponential distribution, denoted as $\mathrm{q-ED}_d(\boldsymbol{\mu}, \mathbf{C})$*, has the following density*

$$p(\mathbf{u}|\boldsymbol{\mu}, \mathbf{C}, q) = \frac{q}{2}(2\pi)^{-\frac{d}{2}}|\mathbf{C}|^{-\frac{1}{2}} \boxed{r^{(\frac{q}{2}-1)\frac{d}{2}}} \exp\left\{-\frac{r^{\frac{q}{2}}}{2}\right\}, \quad r(\mathbf{u}) = (\mathbf{u}-\boldsymbol{\mu})^{\mathsf{T}}\mathbf{C}^{-1}(\mathbf{u}-\boldsymbol{\mu}) \tag{4}$$

- If $\mathbf{u} \sim \mathrm{q-ED}_d(0, \mathbf{C})$, then we denote $\mathbf{u}^* = d^{\frac{1}{2}-\frac{1}{q}}\mathbf{u} \sim \mathrm{q-ED}_d^*(0, \mathbf{C})$ following a *scaled* q-exponential distribution.

Definition (Q-EP)

A (centered) q-exponential process $u(x)$ *with kernel* $\mathcal{C}$*,* $\mathrm{q-}\mathcal{EP}(0, \mathcal{C})$*, is a collection of random variables such that any finite set,* $\mathbf{u} = (u(x_1), \cdots u(x_N))$*, follows a scaled multivariate q-exponential distribution, i.e.* $\mathbf{u} \sim \mathrm{q-ED}_N^*(0, \mathbf{C})$.

Stochastic Representation

how to sample?

- As an elliptic contour distribution, the q-exponential has a stochastic representation due to Schoenberg [41].

Theorem

$\mathbf{u} \sim \mathrm{q-ED}_d(\boldsymbol{\mu}, \mathbf{C})$ *if and only if*

$$\mathbf{u} \overset{d}{=} \boldsymbol{\mu} + R\mathbf{L}S \tag{5}$$

where $S \sim \mathrm{Unif}(\mathcal{S}^{d+1})$ *uniformly distributed on the unit-sphere* $\mathcal{S}^{d+1}$, $\mathbf{L}$ *is the Cholesky factor of* $\mathbf{C}$ *such that* $\mathbf{C} = \mathbf{L}\mathbf{L}^\mathsf{T}$, $R \perp S$ *and* $R^q \overset{d}{=} r(\mathbf{u})^{\frac{q}{2}} \sim \chi^2_d$.

- A more direct sampling is easier to work with:

$$\mathbf{u} = T(\mathbf{z}) = \mathbf{L}\mathbf{z}\|\mathbf{z}\|^{\frac{2}{q}-1}, \quad \mathbf{z} = T^{-1}(\mathbf{u}) = \mathbf{L}^{-1}\mathbf{u}\|\mathbf{L}^{-1}\mathbf{u}\|^{\frac{q}{2}-1}, \quad \mathbf{z} \sim \nu_0 \tag{6}$$

Series Representation

equivalent to Besov process but has direct control on the correlation structure through $\mathcal{C}$

- Q-EP and Besov share equivalent series representations.

Theorem (Karhunen-Loéve)

If $u(x) \sim \mathrm{q}\!-\!\mathcal{EP}(0, \mathcal{C})$ with $\mathcal{C}$ having eigen-pairs $\{\lambda_\ell, \phi_\ell(x)\}_{\ell=1}^\infty$ such that $\mathcal{C}\phi_\ell(x) = \phi_\ell(x)\lambda_\ell$, $\|\phi_\ell\|_2 = 1$ for all $\ell \in \mathbb{N}$ and $\sum_{\ell=1}^\infty \lambda_\ell < \infty$, then we have the following series representation for $u(x)$:

$$u(x) = \sum_{\ell=1}^{\infty} u_\ell \phi_\ell(x), \quad u_\ell := \int_D u(x)\phi_\ell(x) \overset{ind}{\sim} \mathrm{q}\!-\!\mathrm{ED}^*(0, \lambda_\ell) \tag{7}$$

where $\mathrm{E}[u_\ell] = 0$ and $\mathrm{Cov}(u_\ell, u_{\ell'}) = \lambda_\ell \delta_{\ell\ell'}$ with Dirac function $\delta_{\ell\ell'} = 1$ if $\ell = \ell'$ and 0 otherwise.

- If we factor $\sqrt{\lambda_\ell}$ out of u_ℓ, we have the following expansion for Q-EP more comparable to (19) for Besov:

$$u(x) = \sum_{\ell=1}^{\infty} \sqrt{\lambda_\ell} u_\ell \phi_\ell(x), \quad u_\ell \overset{iid}{\sim} \mathrm{q}\!-\!\mathrm{ED}(0, 1) \propto \pi_q(\cdot) \tag{8}$$

- Let $L(\cdot; 0, \Sigma)$ be the likelihood model, and μ_0 be the prior.

$$\begin{aligned} y &= u(x) + \varepsilon, \quad \varepsilon \sim L(\cdot; 0, \Sigma) \\ u &\sim \mu_0(du) \end{aligned} \tag{9}$$

- **Conjugate case**: $\mu_0 = \mathrm{q}-\mathcal{EP}(0, \mathcal{C})$ and $L(\cdot; 0, \Sigma) = \mathrm{q}-\mathrm{ED}(\mathbf{0}, \Sigma)$

Theorem (Posterior Prediction)

Given covariates $\mathbf{x} = \{x_i\}_{i=1}^N$ *and observations* $\mathbf{y} = \{y_i\}_{i=1}^N$ *following* $\mathrm{q}-\mathrm{ED}$ *in the model* (9) *with* $\mathrm{q}-\mathcal{EP}$ *prior for the same* $q > 0$*, we have the following posterior predictive distribution for* $u(x_*)$ *at (a) new point(s)* x_**:*

$$u(x_*)|\mathbf{y}, \mathbf{x}, x_* \sim \mathrm{q}-\mathrm{ED}(\boldsymbol{\mu}^*, \mathbf{C}^*), \ \boldsymbol{\mu}^* = \mathbf{C}_*^{\mathrm{T}}(\mathbf{C} + \Sigma)^{-1}\mathbf{y}, \ \mathbf{C}^* = \mathbf{C}_{**} - \mathbf{C}_*^{\mathrm{T}}(\mathbf{C} + \Sigma)^{-1}\mathbf{C}_* \tag{10}$$

where $\mathbf{C} = \mathcal{C}(\mathbf{x}, \mathbf{x})$*,* $\mathbf{C}_* = \mathcal{C}(\mathbf{x}, x_*)$*, and* $\mathbf{C}_{**} = \mathcal{C}(x_*, x_*)$*.*

- **Non-conjugate case**: posterior sampling by dimension-independent MCMC algorithms [8, 5, 2, 3, 4] with the pushforward $\mu_0 = T^{\#}\nu_0$:

$$\mathbf{u} = T(\mathbf{z}) = \mathbf{L}\mathbf{z}\|\mathbf{z}\|^{\frac{2}{q}-1}, \quad \mathbf{z} = T^{-1}(\mathbf{u}) = \mathbf{L}^{-1}\mathbf{u}\|\mathbf{L}^{-1}\mathbf{u}\|^{\frac{q}{2}-1}, \quad \mathbf{z} \sim \nu_0$$

Computed Tomography Imaging

preserving the edges

$$\mathbf{y} = \mathbf{A}\mathbf{u} + \varepsilon, \quad \varepsilon \sim \mathcal{N}(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}), \quad \mathbf{y} \in \mathbb{R}^{n_\theta n_s}, \quad \mathbf{A} \in \mathbb{R}^{n_\theta n_s \times n^2}, \quad \mathbf{u} \in \mathbb{R}^{n^2}$$

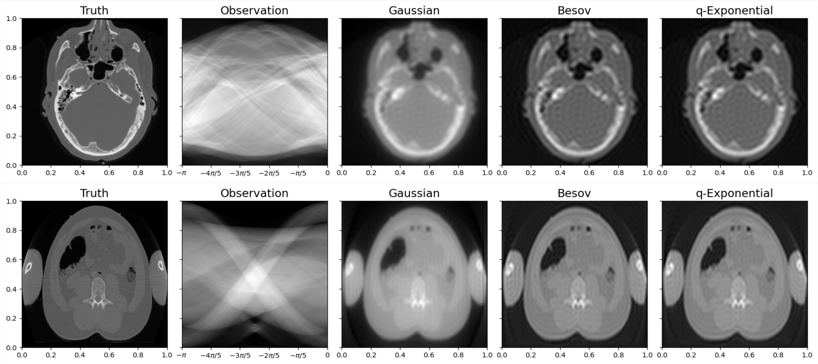
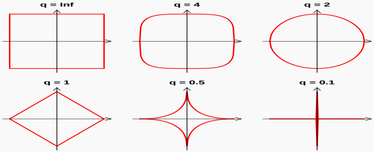


Figure: CT of human head (upper) and torso (lower): true image, observation (sinogram), and MAP estimates by GP, Besov and Q-EP models with relative errors 29.99%, 22.41% and 22.24% (for head) and 26.11%, 21.77% and 21.53% (for torso) respectively.

Table: Posterior estimates of Shepp–Logan phantom by GP, Besov and Q-EP prior models: relative error, RLE $:= \|\hat{u} - u^\dagger\|/\|u^\dagger\|$, of MAP ($\hat{u} = u^*$) and posterior mean ($\hat{u} = \overline{u}$) respectively, log-likelihood (LL), peak signal-to-noise ratio (PSNR) [13], structured similarity index (SSIM) [48], Haar wavelet-based perceptual similarity index (HaarPSI) [40]. Numbers in the bracket are standard deviations obtained repeating the experiments for 10 times with different random seeds.

	MAP			Posterior Mean		
	GP	Besov	Q-EP	GP	Besov	Q-EP
RLE	0.6810	0.7027	**0.4087**	0.4917(6.16e-7)	0.4894(3.53e-5)	**0.4890**(4.79e-5)
LL	-1.55e+6	-1.54e+6	-1.57e+5	-5.21e+5(8.47)	-4.80e+5(196.34)	-4.56e+5(307.97)
PSNR	15.5531	15.2806	**19.9887**	18.3826(1.09e-5)	18.4226(6.27e-4)	**18.4303**(8.51e-4)
SSIM	0.4028	0.3703	**0.5967**	**0.5561**(3.92e-7)	0.5535(2.38e-4)	0.5403(5.26e-4)
HaarPSI	0.0961	0.0870	**0.3105**	0.3126(1.52e-8)	**0.3126**(3.36e-4)	0.3122(3.06e-4)

Solving & Learning Differential Equations with Variational Q-Exponential Processes[2]



[2] Guangting Yu and Shiwei Lan*, *NIPS*2025

Modeling Derivatives

Is Gaussian process (GP) optimal?

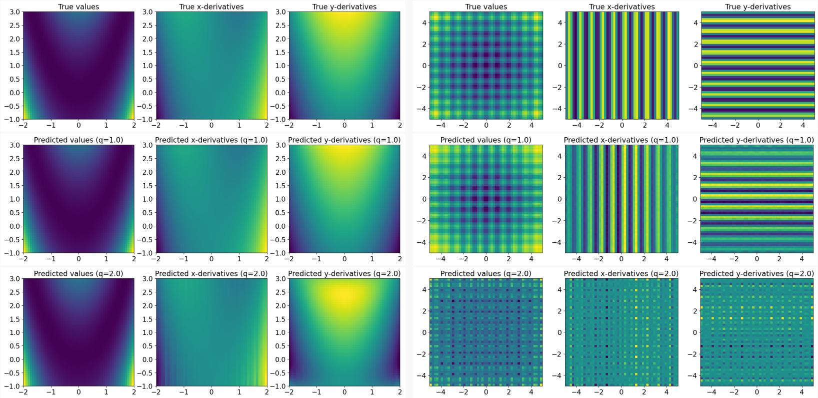


Figure: Contrasting Q-EP ($q = 1.0$, middle row) with GP ($q = 2.0$, bottom row) against the truth (top row) for modeling function values and derivatives of Rosenbrock (left) and Rastrigin (right).

Physics-Informed Machine Learning

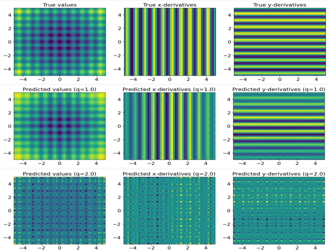
- Solving PDEs is of fundamental importance in science and technology. There is increasing interest and effort to automate this process.
- There are two main thrusts: neural network (NN)-based algorithms and Gaussian process (GP)-based probabilistic solvers.
  - NN methods include PINN [37, 49], deep Ritz [12], deep Galerkin [43], FNO [26], DeepONet [27], NIO [30], etc.
  - They require large samples but lack convergence guarantee or uncertainty quantification (UQ)
  - GP is introduced to solve and learn ODEs [44, 38, 6, 16] and PDEs [33, 34, 7, 29, 15] with theoretic guarantee [35] and UQ [17].
  - GP tends to be over-smooth and lacks edge-preserving property [22, 10].
- We advocate the recently proposed q-exponential process (Q-EP) [24] as a superior probabilistic method for solving and learning PDEs.

Modeling with Derivative Information

- Let $u \sim \text{q}-\mathcal{EP}(0, \mathcal{C})$. Denote the function and its derivatives by $\tilde{u} = (u, \frac{\partial}{\partial \mathbf{x}} u, \cdots, \frac{\partial^k}{\partial \mathbf{x}^k} u)$ up to order k.
- $\tilde{u} \sim \text{q}-\mathcal{EP}(0, \tilde{\mathcal{C}})$ is also a Q-EP if $\mathcal{C}$, e.g. `matern52`, is differentiable.
- Heuristically, the superiority of Q-EP in modeling derivatives over GP comes from its improved ability to handle inhomogeneity.

Table: The structure of kernel $\tilde{\mathcal{C}}$ with derivatives.

$\text{Cov}(\cdot,\cdot)$	$u(\mathbf{x}')$	$\frac{\partial}{\partial \mathbf{x}'} u(\mathbf{x}')$	$\frac{\partial^2}{\partial (\mathbf{x}')^2} u(\mathbf{x}')$
$u(\mathbf{x})$	$\mathcal{C}(\mathbf{x}, \mathbf{x}')$	$\frac{\partial}{\partial \mathbf{x}'} \mathcal{C}(\mathbf{x}, \mathbf{x}')$	$\frac{\partial^2}{\partial (\mathbf{x}')^2} \mathcal{C}(\mathbf{x}, \mathbf{x}')$
$\frac{\partial}{\partial \mathbf{x}} u(\mathbf{x})$	$\frac{\partial}{\partial \mathbf{x}} \mathcal{C}(\mathbf{x}, \mathbf{x}')$	$\frac{\partial^2}{\partial \mathbf{x} \partial \mathbf{x}'} \mathcal{C}(\mathbf{x}, \mathbf{x}')$	$\frac{\partial^3}{\partial \mathbf{x} \partial (\mathbf{x}')^2} \mathcal{C}(\mathbf{x}, \mathbf{x}')$
$\frac{\partial^2}{\partial \mathbf{x}^2} u(\mathbf{x})$	$\frac{\partial^2}{\partial \mathbf{x}^2} \mathcal{C}(\mathbf{x}, \mathbf{x}')$	$\frac{\partial^3}{\partial \mathbf{x}^2 \partial \mathbf{x}'} \mathcal{C}(\mathbf{x}, \mathbf{x}')$	$\frac{\partial^4}{\partial \mathbf{x}^2 \partial (\mathbf{x}')^2} \mathcal{C}(\mathbf{x}, \mathbf{x}')$



Function Space for Solving PDE
with variational Q-EP

- To prepare for solving PDEs, we define $\|\cdot\|_{s,q}$ for u [21, 10]:

$$\|u(\cdot)\|_{s,q} = \left(\sum_{\ell=1}^{\infty} \ell^{\tau_q(s)q} |u_\ell|^q \right)^{\frac{1}{q}}, \tau_q(s) = \frac{s}{d} + \frac{1}{2} - \frac{1}{q}. \tag{11}$$

- Consider the Banach space $B^{s,q}(\Omega) := \{u : \Omega \to \mathbb{R} \mid \|u(\cdot)\|_{s,q} < \infty\}$.

Assumption

Suppose $\lambda = \{\lambda_\ell\}_{\ell=1}^{\infty}$ are eigenvalues of HS operator $T_{\tilde{\mathcal{C}}}$ for the kernel $\tilde{\mathcal{C}}$. We assume

$$\lambda \in \ell^{\frac{q}{2}}, \quad i.e. \quad \|\lambda\|_{\frac{q}{2}}^{\frac{q}{2}} = \sum_{\ell=1}^{\infty} \lambda_\ell^{\frac{q}{2}} < \infty.$$

Proposition

If $\tilde{u}(\cdot) \sim \mathrm{q}-\mathcal{EP}(0, \tilde{\mathcal{C}})$ satisfying Assumption 1, then $\tilde{u}(\cdot) \in L^q_{\mathbb{P}}(\mathbb{R}^\infty, L^q(\Omega)) := \{\tilde{u} : \Omega \times \mathbb{R}^\infty \to \mathbb{R} | \mathbb{E}(\|\tilde{u}\|_q^q) < \infty\}$.

- Consider the general PDE defined on a bounded domain $\Omega \subset \mathbb{R}^d$:

$$\begin{aligned} \mathcal{D}(u)(\mathbf{x}) &= f(\mathbf{x}), \quad \mathbf{x} \in \Omega, \\ \mathcal{B}(u)(\mathbf{x}) &= g(\mathbf{x}), \quad \mathbf{x} \in \partial\Omega. \end{aligned} \tag{12}$$

where $\mathcal{D} : B^{s,q}(\Omega) \to L^q(\Omega)$, $\mathcal{B} : B^{s,q}(\partial\Omega) \to L^q(\partial\Omega)$, $f \in L^q(\Omega)$, and $g \in L^q(\partial\Omega)$. Note that $\tilde{u}(\mathbf{x}) \in \mathbb{R}^D$ with $D = 1 + kd$.

- Let $\overline{\Omega} = \Omega \cup \partial\Omega$, $\mathcal{P} = (\mathcal{D}, \mathcal{B}) : B^{s,q}(\overline{\Omega}) \to L^q(\overline{\Omega})$, $h = (f, g) \in L^q(\overline{\Omega})$.

Assumption

There exists a differentiable function $P : \mathbb{R}^D \to \mathbb{R}$ such that $\mathcal{P}(u)(\mathbf{x}) = P(\tilde{u}(\mathbf{x}))$. And there is a constant $C > 0$ such that $\|\nabla P\| \leq C$.

- Let $\mathbf{X} = \{\mathbf{x}_n\}_{n=1}^N$ be a set of collocation points. Then $\mathcal{P}(u)(\mathbf{X})$ becomes a nonlinear function of $\tilde{\mathbf{U}} := \tilde{u}(\mathbf{X})$, i.e. $P(\tilde{\mathbf{U}}) = \mathcal{P}(u)(\mathbf{X})$. Let $\mathbf{h} = h(\mathbf{X})$.
- The numeric solver seeks to obtain $\tilde{\mathbf{U}}$ based on $P(\tilde{\mathbf{U}}) = \mathbf{h}$.

Bayesian Model of PDE Solutions
with variational Q-EP

- Model $\tilde{u}(\mathbf{X}) \sim \text{q}-\text{ED}(\tilde{\mathbf{u}}, \mathbf{S})$ with some variational distribution. To define likelihood, we *propagate this distribution* by linearizing P.

$$
\begin{aligned}
P(\tilde{u}(\mathbf{X})) &\approx P(\tilde{\mathbf{u}}_0) + \nabla P(\tilde{\mathbf{u}}_0)(\tilde{u}(\mathbf{X}) - \tilde{\mathbf{u}}_0) \sim \text{q}-\text{ED}(\mathbf{m}, \mathbf{\Gamma}), \\
\mathbf{m} &= P(\tilde{\mathbf{u}}_0) + \nabla P(\tilde{\mathbf{u}}_0)(\tilde{\mathbf{u}} - \tilde{\mathbf{u}}_0), \quad \mathbf{\Gamma} = \nabla P(\tilde{\mathbf{u}}_0)\mathbf{S}\nabla P(\tilde{\mathbf{u}}_0)^{\mathsf{T}} + \delta \mathbf{I}_N.
\end{aligned} \tag{13}
$$

where the reference point $\tilde{\mathbf{u}}_0$ can be chosen as $\tilde{\mathbf{u}}_{n-1}$ from the previous training epoch or simply $\tilde{\mathbf{u}}$.

- Let $\mathbf{Y} = P(\tilde{u}(\mathbf{X}))$ and $\mathbf{h} = h(\mathbf{X})$. We have the solution to (12) probabilistically as the posterior $p(\tilde{u}(\mathbf{X})|\mathbf{Y})$ of the Bayesian model

$$
\begin{aligned}
\mathbf{Y}|\tilde{u}(\mathbf{X}), \mathbf{h} &\sim \text{q}-\text{ED}_N(\mathbf{h}, \mathbf{\Gamma}), \\
\tilde{u} &\sim \text{q}-\mathcal{EP}(0, \tilde{\mathcal{C}}).
\end{aligned} \tag{14}
$$

- We approximate the posterior to (14) using sparse variational Bayes which introduces M inducing points to bring the complexity from $\mathcal{O}(N^3)$ to $\mathcal{O}(NM^2)$ [46, 31].

Theorem (Posterior Contraction)

Let $u \sim \text{q}-\mathcal{EP}(0,\mathcal{C})$ in $L^q(\Omega)$ with $\mathcal{C}$ satisfying above assumptions. If the true solution to (12) $u^\dagger \in B^{s^\dagger,q^\dagger}(\Omega)$ *with* $s^\dagger > s' + \left(\frac{d}{q^\dagger} - \frac{d}{q}\right)_+$, $s' = \frac{d}{q} - \frac{d}{2}$, *and* $q^\dagger, q \in [1,2]$, *then the posterior of* (14) *contracts to* $u^\dagger$ *at the optimal rate* $\varepsilon_n^\dagger = n^{-\frac{1}{2+\frac{d}{s^\dagger - s'}}}$ *whenever* $q \leq q^\dagger$.

- When $q^\dagger \geq 1$, setting $q=1$ guarantees the fastest convergence.
- The numerical experiments support the optimal choice of $q=1$.

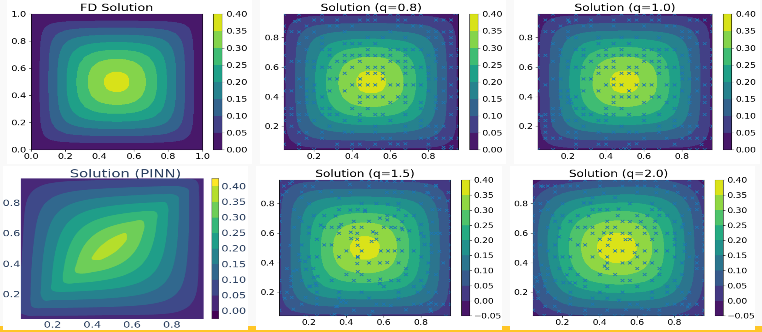
Eikonal Equation

- Let $f \equiv 1$ and $\varepsilon = 0.1$ in Eikonal equation:

$$\begin{aligned} |\nabla u(\mathbf{x})|^2 - \varepsilon \Delta u(\mathbf{x}) &= f(\mathbf{x})^2, \quad \mathbf{x} \in \Omega, \\ u(\mathbf{x}) &= 0, \quad \mathbf{x} \in \partial\Omega. \end{aligned} \tag{15}$$

- Let $D(u, d_1 u, d_2 u, d_1^2 u, d_2^2 u) = (d_1 u)^2 + (d_2 u)^2 - \varepsilon(d_1^2 u + d_2^2 u)$. Then

$$\mathbf{Y} = P(\tilde{u}(\mathbf{X})) = \begin{bmatrix} D(u(\mathbf{X}_d), \frac{\partial}{\partial x_1} u(\mathbf{X}_d), \frac{\partial}{\partial x_2} u(\mathbf{X}_d), \frac{\partial^2}{\partial x_1^2} u(\mathbf{X}_d), \frac{\partial^2}{\partial x_2^2} u(\mathbf{X}_d)) \\ u(\mathbf{X}_b) \end{bmatrix}.$$



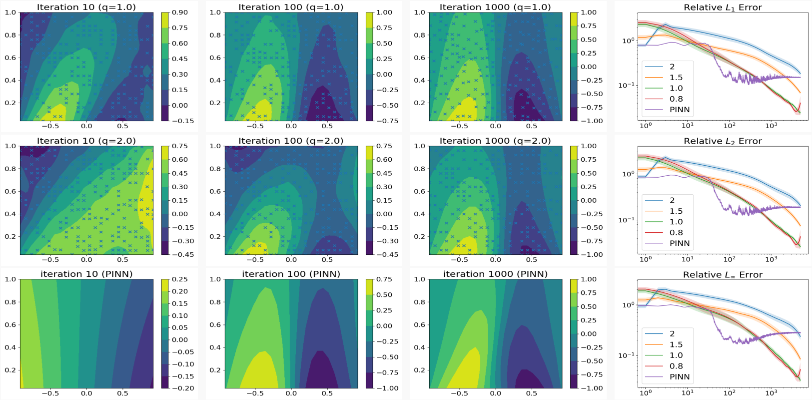
Eikonal Equation

Table: Comparing accuracy of Q-EP ($q = 1$) against GP ($q = 2$) solvers with various kernels for Eikonal equation (15) in terms of mean absolute error (MAE), mean squared error (MSE), and relative errors in L_1 norm (RLE-1), L_2 norm (RLE-2), and L_∞ norm (RLE-∞) respectively. Result in each cell are averaged over 10 experiments with different random seeds; values after $\pm$ are standard deviations of these repeated experiments.

Model (q)	kernel	MAE	MSE	RLE-1	RLE-2	RL-∞
1.0	Matern	**1.68e-3** $\pm$ 5.41e-4	**4.42e-6** $\pm$ 2.26e-6	**0.0106** $\pm$ 0.0034	**0.0110** $\pm$ 0.0030	**0.0321** $\pm$ 0.0071
2.0 (Gaussian)	Matern	9.64e-3 $\pm$ 1.33e-3	1.30e-4 $\pm$ 3.61e-5	0.0610 $\pm$ 0.0084	0.0612 $\pm$ 0.0087	0.1009 $\pm$ 0.0124
1.0	rbf	**4.39e-3** $\pm$ 1.05e-3	**3.91e-5** $\pm$ 2.52e-5	**0.0278** $\pm$ 0.0066	**0.0327** $\pm$ 0.0107	**0.0648** $\pm$ 0.0324
2.0 (Gaussian)	rbf	1.55e-2 $\pm$ 4.97e-4	3.07e-4 $\pm$ 2.87e-5	0.0982 $\pm$ 0.0031	0.0949 $\pm$ 0.0044	0.1236 $\pm$ 0.0048
1.0	rq	**1.86e-3** $\pm$ 5.40e-4	**5.51e-6** $\pm$ 3.44e-6	**0.0118** $\pm$ 0.0034	**0.0122** $\pm$ 0.0038	**0.0168** $\pm$ 0.0064
2.0 (Gaussian)	rq	3.07-3 $\pm$ 1.70e-3	1.89e-5 $\pm$ 2.83e-5	0.0194 $\pm$ 0.0107	0.0200 $\pm$ 0.0134	0.0325 $\pm$ 0.0242

Burgers' Equation

$$\frac{\partial}{\partial t}u + u\frac{\partial}{\partial x}u - \nu\frac{\partial^2}{\partial x^2}u = 0, \quad (x,t) \in (-1,1)\times(0,1],$$
$$u(x,0) = -\sin(\pi x), \quad x \in (-1,1), \qquad (16)$$
$$u(-1,t) = u(1,t) = 0, \quad t \in (0,1].$$



Burgers' Equation

Table: Comparing accuracy of various solvers for Burgers' equation (16) in terms of mean absolute error (MAE), mean squared error (MSE), and relative errors in L_1 norm (RLE-1), L_2 norm (RLE-2), and L_∞ norm (RLE-∞) respectively. Result in each cell are averaged over 10 experiments with different random seeds; values after $\pm$ are standard deviations of these repeated experiments.

Model (q)	MAE	MSE	RLE-1	RLE-2	RL-∞
PINN	5.81e-2 $\pm$ 1.03e-4	7.12e-3 $\pm$ 3.96e-4	0.1508 $\pm$ 0.0017	0.1896 $\pm$ 0.0052	0.2842 $\pm$ 0.0071
B-PINN	2.94e-2 $\pm$ 1.57e-2	1.67e-3 $\pm$ 1.58e-3	0.0785 $\pm$ 0.0420	0.0833 $\pm$ 0.0409	0.1327 $\pm$ 0.0446
0.5	7.77e-2 $\pm$ 7.09e-2	3.50e-2 $\pm$ 7.15e-2	0.2018 $\pm$ 0.3751	0.2056 $\pm$ 0.3764	0.2177 $\pm$ 0.3838
0.8	1.56e-2 $\pm$ 4.96e-3	1.59e-3 $\pm$ 5.25e-3	0.0405 $\pm$ 0.0780	0.0434 $\pm$ 0.0804	0.0529 $\pm$ 0.0924
1.0	**9.33e-3** $\pm$ 1.76e-4	**1.77e-4** $\pm$ 2.16e-4	**0.0242** $\pm$ 0.0128	**0.0266** $\pm$ 0.0140	**0.0324** $\pm$ 0.0145
1.2	1.87e-2 $\pm$ 3.23e-4	5.92e-4 $\pm$ 4.06e-4	0.0485 $\pm$ 0.0151	0.0522 $\pm$ 0.0168	0.0560 $\pm$ 0.0181
1.5	2.64e-2 $\pm$ 8.58e-4	1.31e-3 $\pm$ 1.23e-3	0.0684 $\pm$ 0.0262	0.0754 $\pm$ 0.0316	0.0854 $\pm$ 0.0430
2.0(Gaussian)	7.13e-2 $\pm$ 8.68e-3	1.08e-2 $\pm$ 1.28e-2	0.1848 $\pm$ 0.0908	0.2068 $\pm$ 0.1118	0.2376 $\pm$ 0.1456
2.5	2.26e-1 $\pm$ 3.20e-1	2.49e-1 $\pm$ 6.04e-1	0.5879 $\pm$ 0.8296	0.6669 $\pm$ 0.9576	0.7558 $\pm$ 1.0521

Bayesian Inverse Problems

same variational framework

- Let the PDE (12) contain a quantity of interest, $a(\mathbf{x})$. The task of the inverse problem is to find a true solution, $a^\dagger$, with proper UQ based on observations $\mathcal{O}(u)(\mathbf{X}) = u(\mathbf{X}_o)$, obtained with observation operator $\mathcal{O}$.
- Denote $\tilde{a} = (a, \frac{\partial}{\partial \mathbf{x}} a, \cdots, \frac{\partial^{k'}}{\partial \mathbf{x}^{k'}} a)$ to the order $k' \leq k$. Assume a nonlinear function P such that $P(\tilde{u}(\mathbf{X}), \tilde{a}(\mathbf{X})) = \mathcal{P}(u, a)(\mathbf{X})$. Let $\tilde{\mathbf{Y}} = [P(\tilde{u}(\mathbf{X}), \tilde{a}(\mathbf{X})), \mathcal{O}(u)(\mathbf{X})]$, and $\tilde{\mathbf{h}} = [\mathbf{h}, u(\mathbf{X}_o)]$.
- Model $\tilde{a}(\mathbf{X})$ using variational distribution $q(\tilde{a}(\mathbf{X})) \sim \mathrm{q-ED}(\tilde{\mathbf{a}}, \mathbf{S}_a)$. The final Bayesian model for the inverse problem becomes

$$\tilde{\mathbf{Y}} | \tilde{u}(\mathbf{X}), \tilde{a}(\mathbf{X}), \tilde{\mathbf{h}} \sim \mathrm{q-ED}_{\tilde{N}}(\tilde{\mathbf{h}}, \tilde{\boldsymbol{\Gamma}}), \quad \tilde{\boldsymbol{\Gamma}} = \nabla \tilde{P}(\tilde{\mathbf{u}}_0, \tilde{\mathbf{a}}_0) \begin{bmatrix} \mathbf{S}_u & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_a \end{bmatrix} \nabla \tilde{P}(\tilde{\mathbf{u}}_0, \tilde{\mathbf{a}}_0)^{\mathsf{T}} + \delta \mathbf{I}_{\tilde{N}},$$
$$\tilde{u} \sim \mathrm{q-}\mathcal{EP}(0, \tilde{\mathcal{C}}_u), \quad \tilde{a} \sim \mathrm{q-}\mathcal{EP}(0, \tilde{\mathcal{C}}_a). \tag{17}$$

- We obtain the variational solution of $\tilde{u}(\mathbf{X}) | \tilde{\mathbf{Y}}$ as a byproduct.

Inverse Darcy Flow

$$
\begin{aligned}
-\mathrm{div}(\exp(a)\nabla u)(\mathbf{x}) &= f(\mathbf{x}), \quad \mathbf{x} \in \Omega := [0,1]^2, \\
u(\mathbf{x}) &= 0, \quad \mathbf{x} \in \partial\Omega.
\end{aligned} \tag{18}
$$

where the true coefficient such that
$\exp(a^\dagger(\mathbf{x})) = \exp(\sin(2\pi x_1) + \sin(2\pi x_2)) + \exp(-\sin(2\pi x_1) - \sin(2\pi x_2))$ [7].

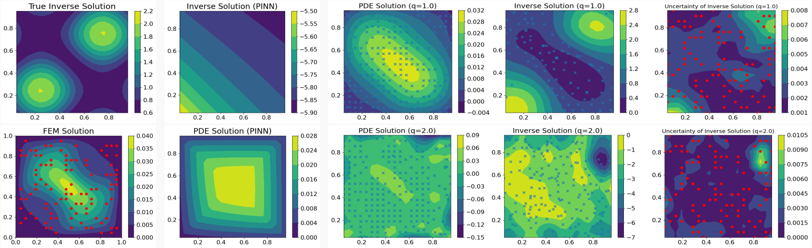
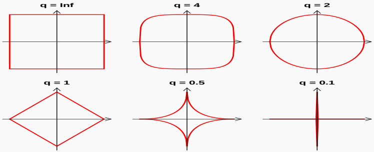


Figure: Solving inverse Darcy flow (18) using PINN (second column), Q-EP (right three in upper row: $q = 1.0$), and GP (right three in lower row: $q = 2.0$) respectively. Upper left: true inverse solution $a^\dagger$; lower left: fine-resolution finite element solution $u^\dagger$ to (18) with $a^\dagger$. Blue crosses are learned inducing points, and red dots indicate locations of observations.

Spatiotemporal Besov Priors for Bayesian Inverse Problems[3]



[3] Shiwei Lan*, Mirjeta Pasha, Shuyi Li, Weining Shen (2025), JASA

Spatiotemporal Generalization of Besov Process

- We generalize BP to the spatiotemporal domain by replacing the random coefficients with random functions following q$-\mathcal{EP}$:

$$u(x) = \sum_{\ell=1}^{\infty} \gamma_\ell \xi_\ell \phi_\ell(x), \quad \xi_\ell \overset{\text{i.i.d.}}{\sim} \pi_q(\cdot) \propto \exp\left(-\frac{1}{2}|\cdot|^q\right), \forall \ell \in \mathbb{N}. \quad (19)$$

$\downarrow \qquad \downarrow \qquad \downarrow$

$$u(\mathbf{z}) = \sum_{\ell=1}^{\infty} \gamma_\ell \xi_\ell(t) \phi_\ell(\mathbf{x}),\ \xi_\ell(\cdot) \overset{\text{i.i.d.}}{\sim} \text{q}-\mathcal{EP}(0, \mathcal{C}),\ \forall \ell \in \mathbb{N}. \quad (20)$$

where $\gamma_\ell = \kappa \ell^{-\tau_q(s)}$ and $\tau_q(s) = \frac{s}{d} + \frac{1}{2} - \frac{1}{q}$.

- Let $\|u\|_{s,q} = \left(\sum_{\ell=1}^{\infty} \ell^{\tau_q(s)q} \|u_\ell(\cdot)\|_q^q\right)^{\frac{1}{q}}$ and $\mathcal{Z} = \mathcal{X} \times \mathcal{T}$. Consider Banach space $B^{s,q}(\mathcal{Z}) = \{u : \mathcal{Z} \to \mathbb{R} \mid \|u\|_{s,q} < \infty\}$.

Spatiotemporal Besov Process (STBP)

- STBP $\mathcal{STBP}(\mathcal{C}, B^{s,q}(\mathcal{Z}))$ generalizes BP to capture the spatial inhomogeneity while modeling temporal correlations via a kernel $\mathcal{C}$.
- We also have the Karhunen-Loéve theorem for an STBP $u(\cdot)$ represented in the spatial ($\{\phi_\ell\}_{\ell=1}^\infty$) and temporal ($\{\psi_{\ell'}\}_{\ell'=1}^\infty$) bases.

Theorem (Karhunen-Loéve)

If $u \sim \mathcal{STBP}(\mathcal{C}, B^{s,q}(\mathcal{Z}))$ as in (20) with a trace-class HS operator $T_\mathcal{C}$ having eigen-pairs $\{\lambda_\ell, \psi_\ell(\cdot)\}_{\ell=1}^\infty$, then we have

$$u(\mathbf{z}) = \sum_{\ell=1}^\infty \sum_{\ell'=1}^\infty u_{\ell\ell'} \phi_\ell(\mathbf{x}) \psi_{\ell'}(t), \;\; u_{\ell\ell'} := \int_\mathcal{T} u_\ell(t) \psi_{\ell'}(t) dt \sim \text{q-ED}(0, \gamma_\ell^2 \lambda_{\ell'}) \,.$$

Moreover, the spatiotemporal covariance of STBP bears a separable structure, i.e.

$$\mathrm{Cov}(u(\mathbf{z}), u(\mathbf{z}')) = \sum_{\ell=1}^\infty \gamma_\ell^2 \phi_\ell(\mathbf{x}) \phi_\ell(\mathbf{x}') \mathcal{C}(t, t') \,.$$

Dynamic Tomography Reconstruction
linear inverse problem

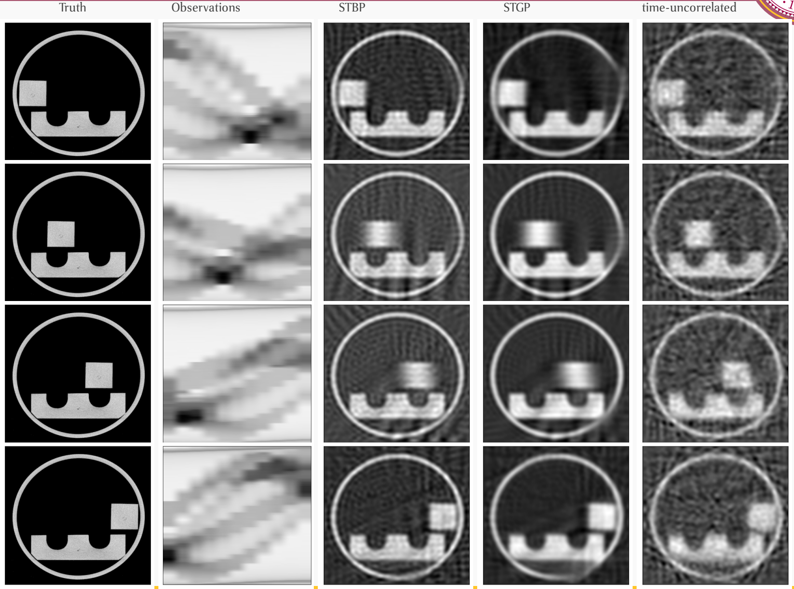


Figure: MAP reconstruction of dynamic STEMPO tomography in the whitened space.

- Consider the following 2-d Navier-Stokes equation (NSE) for a viscous, incompressible fluid in vorticity form on $\mathbb{T}^2 = (0,1)^2$:

$$\begin{aligned} \partial_t w(x,t) + u(x,t)\cdot\nabla w(x,t) &= \nu\Delta w(x,t) + f(x), && x\in(0,1)^2, t\in(0,T], \\ \nabla\cdot u(x,t) &= 0, && x\in(0,1)^2, t\in[0,T], \\ w(x,0) &= w_0(x), && x\in(0,1)^2. \end{aligned}$$

  where $u \in C([0,T];H^r(\mathbb{T}^2;\mathbb{R}^2))$ is the velocity field, $w = \nabla\times u$ is the vorticity, $w_0 \in L^2(\mathbb{T}^2;\mathbb{R})$ is the initial vorticity, $\nu\in\mathbb{R}_+$ is the viscosity coefficient, and $f\in L^2(\mathbb{T}^2;\mathbb{R})$ is the forcing function.

- We build an emulator based on the Fourier operator neural network (FNO) [25] that maps the vorticity up to time $T_0 = 10$ to the vorticity up to some later time $T > 10$:

$$\mathcal{G} : C([0,T_0];H^r(\mathbb{T}^2;\mathbb{R}^2)) \to C((T_0,T];H^r(\mathbb{T}^2;\mathbb{R}^2)),$$
$$w\big|_{(0,1)^2\times[0,10]} \mapsto w\big|_{(0,1)^2\times(10,T]}.$$

- We choose the viscosity $\nu = 1\mathrm{e}-3$ and set $T - T_0 = 30$.
- We initialize the vorticity w_0 with a (star-convex) polygon.
- We then observe data of vorticity $w|_{(0,1)^2\times(10,40]}(\mathbf{X}, t_j)$ with $t_j \in (T_0, T]$ for $j = 0, \cdots, 29$ based on the true initial inputs $w^\dagger|_{(0,1)^2\times[0,10]}$, with Gaussian noise contamination, i.e.,

$$\mathbf{y}_j = \mathcal{G}(w^\dagger|_{(0,1)^2\times[0,10]}(\mathbf{X}, t_j)) + \boldsymbol{\eta}_j, \quad \boldsymbol{\eta}_j \sim N(0, \Gamma_{\text{noise}}),$$

  where Γ_{noise} is empirically estimated.
- Instead of the solution of the initial condition w_0 alone, we are interested in the inverse solution of vorticity for an initial period, i.e., $w|_{(0,1)^2\times[0,10]}$.

Navier-Stokes Inverse Problem

nonlinear inverse problem

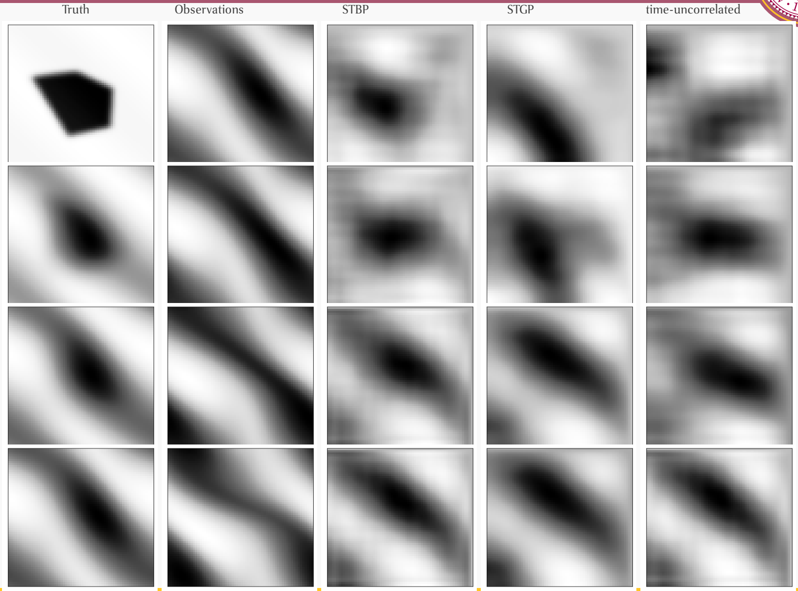
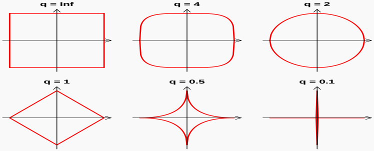


Figure: MAP inverse solutions of Navier-Stokes equation in the whitened space.

Deep Q-Exponential Processes[4]



[4] Zhi Chang, Chukwudi Obite, Shuang Zhou, Shiwei Lan*, *AABI* 2025

Shallow Q-EP Model

- Let $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_Q]_{N\times Q}$, $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_D]_{N\times D}$, $f := (f_1, \cdots, f_D)$, and $\mathbf{F} = f(\mathbf{X})_{N\times D}$, . With m.i.u. Q-EP priors imposed on $f := (f_1, \cdots, f_D)$, we consider the multivariate regression problem:

$$\begin{aligned} \text{likelihood :} \quad & \mathrm{vec}(\mathbf{Y})|\mathbf{F} \sim \mathrm{q-ED}_{ND}(\mathrm{vec}(\mathbf{F}), \mathbf{I}_D \otimes \Sigma), \\ \text{prior on latent function :} \quad & f \sim \mathrm{q-}\mathcal{EP}(0, \mathcal{C}, \mathbf{I}_D). \end{aligned} \tag{21}$$

- Marginalizing f yields the following Bayesian shallow Q-EP model:

$$\begin{aligned} \text{marginal likelihood :} \quad & \mathrm{vec}(\mathbf{Y})|\mathbf{X} \sim \mathrm{q-ED}(\mathbf{0}, \mathbf{I}_D \otimes \mathbf{K}_\mathbf{X}), \\ \text{prior on input/latent variable :} \quad & \mathrm{vec}(\mathbf{X}) \sim \mathrm{q-ED}(\mathbf{0}, \mathbf{I}_{NQ}). \end{aligned} \tag{22}$$

- If $\mathbf{X}$ is latent variable, this is a latent variable model (LVM) [23, 47].
- It can be regarded as a *non-Gaussian, nonlinear, probabilistic PCA* [45].

Shallow Q-EP Model

Bayesian Regularization on Latent Representation [a]

[a]Chukwudi Obite, Zhi Chang, Keyan Wu, Shiwei Lan*, *ICLR*2025

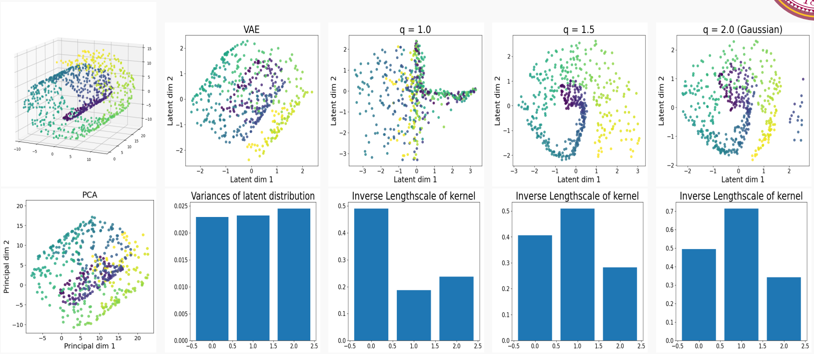


Figure: Latent representation of Swiss roll dataset. Upper left: 3d cloud of 1000 points; lower left: PCA in 2d principal space; right 4 columns: 2d latent representations (upper row) by VAE and QEP-LVMs with $q = 1.0, 1.5$ and 2.0 (GP-LVM) showing a regularization effect via the parameter q, and the corresponding variances of latent distribution (VAE) and inverse length-scales γ ordered on the x-axes (lower row). Colors are used to aid visualization but not for training.

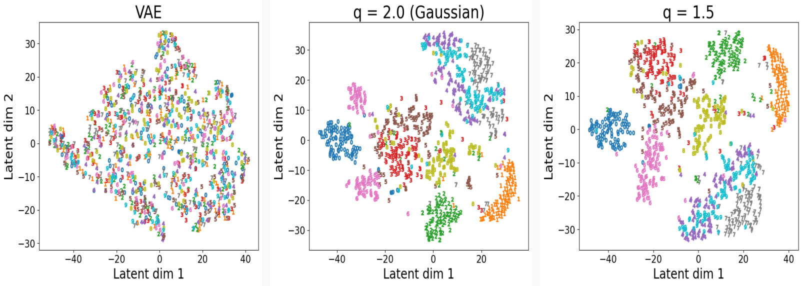


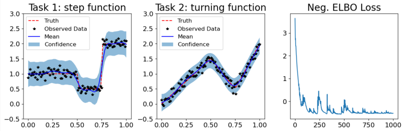
Figure: Latent representations of MNIST database by VAE (left), GP-LVM (middle), QEP-LVM with $q = 1.5$ (right). For the convenience of visualization, 10-dimensional latent spaces learned by these algorithms are projected to 2-d subspace by t-SNE respectively.

Deep Q-EP Model

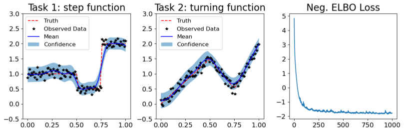
- Similarly as building deep GP with GP-LVMs [9], we consider a hierarchy of L shallow Q-EP layers (22) as follows:

$$
\begin{aligned}
y_{nd} &= f_d^0(\mathbf{x}_n^1) + \varepsilon_{nd}^0, \quad d = 1, \cdots, D_0, \quad \mathbf{x}_n^1 \in \mathbb{R}^{D_1}, \\
x_{nd}^1 &= f_d^1(\mathbf{x}_n^2) + \varepsilon_{nd}^1, \quad d = 1, \cdots, D_1, \quad \mathbf{x}_n^2 \in \mathbb{R}^{D_2}, \\
\vdots & \qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\
x_{nd}^{L-1} &= f_d^{L-1}(\mathbf{z}_n) + \varepsilon_{nd}^{L-1}, \quad d = 1, \cdots, D_{L-1}, \quad \mathbf{z}_n \in \mathbb{R}^{D_L},
\end{aligned}
$$

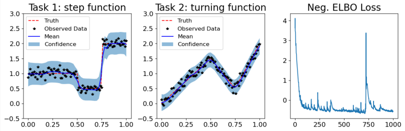
  where $\boldsymbol{\varepsilon}^\ell \sim \text{q}-\text{ED}(\mathbf{0}, \Gamma^\ell)$, $f^\ell \sim \text{q}-\mathcal{EP}(0, k^\ell, I_{D_\ell})$ for $\ell = 0, \cdots, L-1$.
- The evidence lower bound (ELBO) is complicated but tractable formula can be obtained with the help of Jensen's inequality.
- The deep Q-EP indeed has enhanced expressiveness and flexibility of regularization.



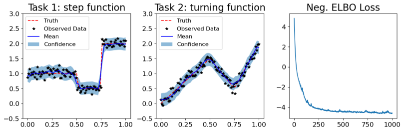
(a) Shallow GP regression.



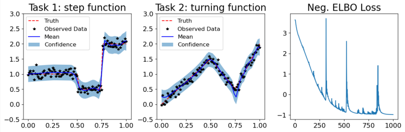
(b) Shallow Q-EP regression.



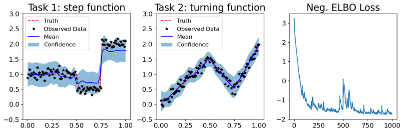
(c) Deep GP regression.



(d) Deep Q-EP regression.



(e) DKL-GP regression.



(f) DSPP regression.

Deep Q-EP Model

Classification

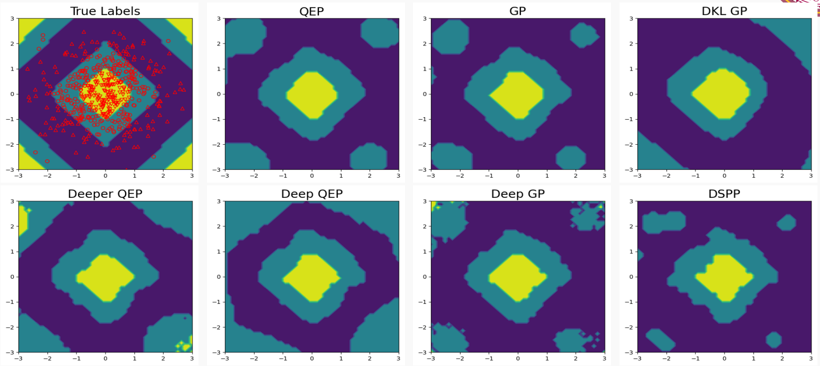


Figure: Comparing shallow (1-layer), deep (2-layer) and deeper (3-layer) Q-EPs with GP, deep GP, DKL-GP and DSPP on a classification problem defined on annular rhombus. Circles, upper and lower triangles label three classes in the training data. Classification accuracy on testing data are 81.04% (GP), 82.2% (Deep GP), 76.4% (DKL-GP), 78.88% (DSPP), 83.4% (Q-EP), 85.64% (Deep Q-EP) and 87.2% (Deeper Q-EP) respectively.

Conclusion

- We propose q-exponential process (Q-EP) as a prior on L^q functions with a flexible parameter $q > 0$ to control the degree of regularization.
- Q-EP includes GP as a special case ($q = 2$) but can impose sharper regularization for $0 < q < 2$.
- Diff Q-EP is novel probabilistic PDE solver based on Q-EP with $q = 1$ preferable to GP with $q = 2$ for a good reason.
- Spatiotemporal Besov process (STBP) can effectively capture spatial data inhomogeneity while modeling temporal correlations.
- Deep Q-EP enhances the expressiveness of (shallow) Q-EP for modeling complex data.
- In future, we plan to study diffusion process with Q-EP noise.

github.com/lanzithinking/Q-EXP

https://github.com/lanzithinking/Diff_QEP

QPyTorch

https://github.com/lanzithinking/QePyTorch

Q$^{ⓔ}$PyTorch

Run Test Suite passing | docs passing | license MIT

python 3.10+ | conda | conda-forge v0.2 | pypi v0.2

Q$^{ⓔ}$PyTorch is a Python package for Q-exponential process (QEP) implemented using PyTorch and built upon GPyTorch. Q$^{ⓔ}$PyTorch is designed to facilitate creating scalable, flexible, and modular QPE models.

Different from GPyTorch for Gaussian process (GP) models, Q$^{ⓔ}$PyTorch focuses on QEP, which generalizes GP by allowing flexible regularization on function spaces through a parameter $q > 0$ and embraces GP as a special case with $q = 2$. QEP is proven to be superior than GP in modeling inhomogeneous objects with abrupt changes or sharp contrast for $q < 2$ [Li et al (2023)]. Inherited from GPyTorch, Q$^{ⓔ}$PyTorch has an efficient and scalable implementation by taking advantage of numerical linear algebra library LinearOperator and improved GPU utilization.

Tutorials, Examples, and Documentation

See **documentation** on how to construct various QEP models in Q$^{ⓔ}$PyTorch.

Installation

Requirements:

- Python >= 3.10
- PyTorch >= 2.0
- GPyTorch >= 1.14

Stable Version

Install Q$^{ⓔ}$PyTorch using pip or conda:

https://lanzithinking.github.io/QePyTorch/

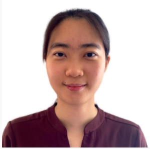
```
pip install qpytorch
conda install qpytorch
```

Acknowledgments

Special Thanks to My Collaborators!

Weining Shen (UCI)

Mirjeta Parsha (VTech)



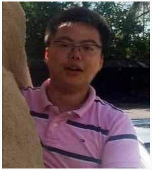
Shuang Zhou (ASU)

Shuyi Li (ASU, former student)

Chukwudi Obite (ASU, current student)

Zhi Chang (ASU, current student)

Keyan Wu (ASU, current student)



Guangting Yu (ASU, former student)

References I

[1] Maria Bånkestad, Jens Sjölund, Jalil Taghia, and Thomas Schön. The elliptical processes: a family of fat-tailed stochastic processes. 03 2020.

[2] A. Beskos, F. J. Pinski, J. M. Sanz-Serna, and A. M. Stuart. Hybrid Monte-Carlo on Hilbert spaces. *Stochastic Processes and their Applications*, 121:2201–2230, 2011.

[3] Alexandros Beskos. A stable manifold MCMC method for high dimensions. *Statistics & Probability Letters*, 90:46–52, 2014.

[4] Alexandros Beskos, Mark Girolami, Shiwei Lan, Patrick E. Farrell, and Andrew M. Stuart. Geometric MCMC for infinite-dimensional inverse problems. *Journal of Computational Physics*, 335(Supplement C):327 – 351, 2017.

[5] Alexandros Beskos, Gareth Roberts, Andrew Stuart, and Jochen Voss. MCMC methods for diffusion bridges. *Stochastics and Dynamics*, 8(03):319–350, 2008.

[6] Ben Calderhead, Mark Girolami, and Neil Lawrence. Accelerating bayesian inference over nonlinear differential equations with gaussian processes. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, *Advances in Neural Information Processing Systems*, volume 21. Curran Associates, Inc., 2008.

[7] Yifan Chen, Bamdad Hosseini, Houman Owhadi, and Andrew M. Stuart. Solving and learning nonlinear pdes with gaussian processes. *Journal of Computational Physics*, 447:110668, 2021.

[8] Simon L Cotter, Gareth O Roberts, AM Stuart, and David White. MCMC methods for functions: modifying old algorithms to make them faster. *Statistical Science*, 28(3):424–446, 2013.

[9] Andreas Damianou and Neil D. Lawrence. Deep Gaussian processes. In Carlos M. Carvalho and Pradeep Ravikumar, editors, *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, volume 31 of *Proceedings of Machine Learning Research*, pages 207–215, Scottsdale, Arizona, USA, 29 Apr–01 May 2013. PMLR.

References II

[10] Masoumeh Dashti, Stephen Harris, and Andrew Stuart. Besov priors for bayesian inverse problems. *Inverse Problems and Imaging*, 6(2):183–200, may 2012.

[11] Masoumeh Dashti and Andrew M. Stuart. *The Bayesian Approach to Inverse Problems*, pages 311–428. Springer International Publishing, Cham, 2017.

[12] Weinan E and Bing Yu. The deep ritz method: A deep learning-based numerical algorithm for solving variational problems. *Communications in Mathematics and Statistics*, 6(1):1–12, Mar 2018.

[13] Osama S. Faragallah, Heba El-Hoseny, Walid El-Shafai, Wael Abd El-Rahman, Hala S. El-Sayed, El-Sayed M. El-Rabaie, Fathi E. Abd El-Samie, and Gamal G. N. Geweid. A comprehensive survey analysis for present solutions of medical image fusion and future directions. *IEEE Access*, 9:11358–11371, 2021.

[14] E. Gómez, M.A. Gomez-Viilegas, and J.M. Marín. A multivariate generalization of the power exponential family of distributions. *Communications in Statistics - Theory and Methods*, 27(3):589–600, jan 1998.

[15] Oliver Hamelijnck, Arno Solin, and Theodoros Damoulas. Physics-informed variational state-space gaussian processes. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

[16] Markus Heinonen, Cagatay Yildiz, Henrik Mannerström, Jukka Intosalmi, and Harri Lähdesmäki. Learning unknown ODE models with Gaussian processes. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 1959–1968. PMLR, 10–15 Jul 2018.

[17] Philipp Hennig, Michael A. Osborne, and Mark Girolami. Probabilistic numerics and uncertainty in computations. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2179):20150142, July 2015.

References III

[18] A. P. Dawid J. M. Bernardo, J. O. Berger and A. F. M. Smith. Regression and classification using gaussian process priors. *Bayesian Statistics*, 6:475–501, 1998.

[19] Junxiong Jia, Jigen Peng, and Jinghuai Gao. Bayesian approach to inverse problems for functions with a variable-index besov prior. *Inverse Problems*, 32(8):085006, 2016.

[20] Tomasz J. Kozubowski, Krzysztof Podgórski, and Igor Rychlik. Multivariate generalized laplace distribution and related random fields. *Journal of Multivariate Analysis*, 113:59–72, 2013. Special Issue on Multivariate Distribution Theory in Memory of Samuel Kotz.

[21] Matti Lassas, Eero Saksman, and Samuli Siltanen. Discretization-invariant bayesian inversion and besov space priors. *Inverse Problems and Imaging*, 3(1):87–122, 2009.

[22] Matti Lassas and Samuli Siltanen. Can one use total variation prior for edge-preserving Bayesian inversion? *Inverse Problems*, 20(5):1537, 2004.

[23] Neil Lawrence. Gaussian process latent variable models for visualisation of high dimensional data. In S. Thrun, L. Saul, and B. Schölkopf, editors, *Advances in Neural Information Processing Systems*, volume 16. MIT Press, 2003.

[24] Shuyi Li, Michael O′ Connor, and Shiwei Lan. Bayesian learning via q-exponential process. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Proceedings of the 37th Conference on Neural Information Processing Systems*, volume 36, pages 72867–72887. Curran Associates, Inc., 2023.

[25] Yikuan Li, Shishir Rao, Abdelaali Hassaine, Rema Ramakrishnan, Dexter Canoy, Gholamreza Salimi-Khorshidi, Mohammad Mamouei, Thomas Lukasiewicz, and Kazem Rahimi. Deep bayesian gaussian processes for uncertainty estimation in electronic health records. *Scientific Reports*, 11(1), October 2021.

References IV

[26] Zongyi Li, Nikola Borislavov Kovachki, Kamyar Azizzadenesheli, Burigede liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021.

[27] Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, March 2021.

[28] Felix Lucka. Fast markov chain monte carlo sampling for sparse bayesian inference in high-dimensional inverse problems using l1-type priors. *Inverse Problems*, 28(12):125012, nov 2012.

[29] Rui Meng and Xianjin Yang. Sparse gaussian processes for solving nonlinear pdes. *Journal of Computational Physics*, 490:112340, 2023.

[30] Roberto Molinaro, Yunan Yang, Björn Engquist, and Siddhartha Mishra. Neural inverse operators for solving PDE inverse problems. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 25105–25139. PMLR, 23–29 Jul 2023.

[31] Chukwudi Paul Obite, Zhi Chang, Keyan Wu, and Shiwei Lan. Bayesian regularization of latent representation. In *The Thirteenth International Conference on Learning Representations*, Singapore, April 2025. ICLR.

[32] Bernt Øksendal. *Stochastic Differential Equations*. Springer Berlin Heidelberg, 2003.

[33] Houman Owhadi. Bayesian numerical homogenization. *Multiscale Modeling and Simulation*, 13(3):812–828, January 2015.

References V

[34] Houman Owhadi. Multigrid with rough coefficients and multiresolution operator decomposition from hierarchical information games. *SIAM Review*, 59(1):99–149, January 2017.

[35] Houman Owhadi and Clint Scovel. *Operator-Adapted Wavelets, Fast Solvers, and Numerical Homogenization: From a Game Theoretic Approach to Numerical Approximation and Algorithm Design*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2019.

[36] Krzysztof Podgórski and Jörg Wegener. Estimation for stochastic models driven by laplace motion. *Communications in Statistics - Theory and Methods*, 40(18):3281–3302, sep 2011.

[37] M. Raissi, P. Perdikaris, and G.E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

[38] J. O. Ramsay, G. Hooker, D. Campbell, and J. Cao. Parameter estimation for differential equations: a generalized smoothing approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(5):741–796, 2007.

[39] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2005.

[40] Rafael Reisenhofer, Sebastian Bosse, Gitta Kutyniok, and Thomas Wiegand. A haar wavelet-based perceptual similarity index for image quality assessment. *Signal Processing: Image Communication*, 61:33–43, 2018.

[41] I. J. Schoenberg. Metric spaces and completely monotone functions. *Annals of Mathematics*, 39:811–841, 1938.

References VI

[42] Amar Shah, Andrew Wilson, and Zoubin Ghahramani. Student-t Processes as Alternatives to Gaussian Processes. In Samuel Kaski and Jukka Corander, editors, *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics*, volume 33 of *Proceedings of Machine Learning Research*, pages 877–885, Reykjavik, Iceland, 22–25 Apr 2014. PMLR.

[43] Justin Sirignano and Konstantinos Spiliopoulos. Dgm: A deep learning algorithm for solving partial differential equations. *Journal of Computational Physics*, 375:1339–1364, 2018.

[44] John Skilling. Bayesian solution of ordinary differential equations. In C. Ray Smith, Gary J. Erickson, and Paul O. Neudorfer, editors, *Maximum Entropy and Bayesian Methods: Seattle, 1991*, pages 23–37. Springer Netherlands, Dordrecht, 1992.

[45] Michael E. Tipping and Christopher M. Bishop. Probabilistic Principal Component Analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 61(3):611–622, 09 1999.

[46] Michalis Titsias. Variational learning of inducing variables in sparse gaussian processes. In David van Dyk and Max Welling, editors, *Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics*, volume 5 of *Proceedings of Machine Learning Research*, pages 567–574, Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA, 16–18 Apr 2009. PMLR.

[47] Michalis Titsias and Neil D. Lawrence. Bayesian gaussian process latent variable model. In Yee Whye Teh and Mike Titterington, editors, *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, volume 9 of *Proceedings of Machine Learning Research*, pages 844–851, Chia Laguna Resort, Sardinia, Italy, 13–15 May 2010. PMLR.

[48] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[49] Liu Yang, Xuhui Meng, and George Em Karniadakis. B-pinns: Bayesian physics-informed neural networks for forward and inverse pde problems with noisy data. *Journal of Computational Physics*, 425:109913, 2021.

Thank you !

https://math.la.asu.edu/∼slan