School of Mathematical and Statistical Sciences
Arizona State University

Solving and Learning Partial Differential Equations with Variational Q-Exponential Processes [a]

Guangting Yu
guangtin@asu.edu

Shiwei Lan*
slan@asu.edu

DEC 2-7, 2025 NIPS @ San Diego

[a] https://openreview.net/pdf?id=WjhS0EpJH7

Modeling Derivatives

Is Gaussian process (GP) optimal?

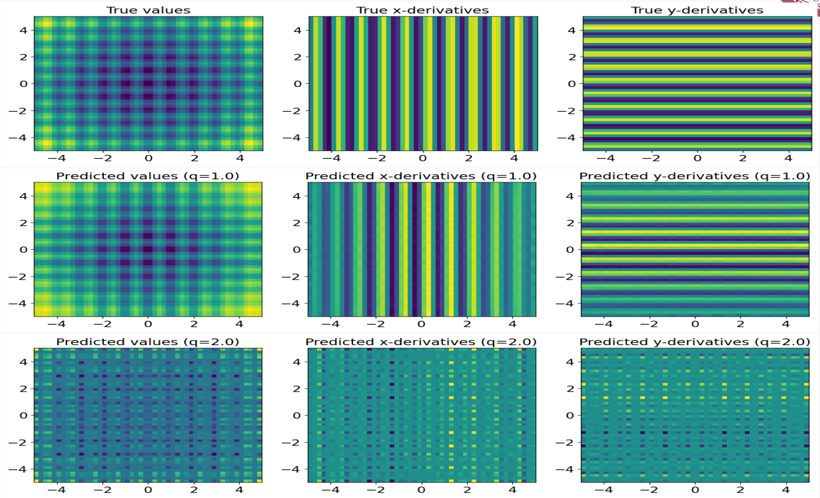


Figure: Contrasting Q-EP ($q = 1.0$, middle) with GP ($q = 2.0$, bottom) against the truth (top) for modeling function values and derivatives in the Rastrigin example.

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 [18, 23], deep Ritz [4], deep Galerkin [20], FNO [10], DeepONet [11], NIO [13], etc.
  - They require large samples but lack convergence guarantee or uncertainty quantification (UQ)
  - GP is introduced to solve and learn ODEs [21, 19, 1, 6] and PDEs [15, 16, 2, 12, 5] with theoretic guarantee [17] and UQ [7].
  - GP tends to be over-smooth and lacks edge-preserving property [8, 3].
- We advocate the recently proposed q-exponential process (Q-EP) [9] as a superior probabilistic method for solving and learning PDEs.

Q-Exponential Process

principled regularization over function spaces

Definition

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

$$p(\mathbf{u}|\boldsymbol{\mu}, \mathbf{C}, q) = \frac{q}{2}(2\pi)^{-\frac{N}{2}}|\mathbf{C}|^{-\frac{1}{2}} \boxed{r^{(\frac{q}{2}-1)\frac{N}{2}}} \exp\left\{-\frac{r^{\frac{q}{2}}}{2}\right\},$$

$$r(\mathbf{u}) = (\mathbf{u}-\boldsymbol{\mu})^{\mathsf{T}}\mathbf{C}^{-1}(\mathbf{u}-\boldsymbol{\mu}).$$

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.* $\mathbf{u} \sim \mathrm{q-ED}_N^*(0, \mathbf{C})$.

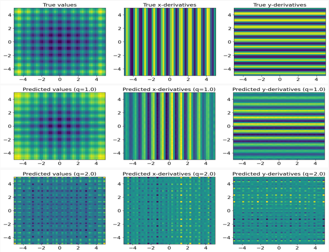
- If $q = 2$, $\mathrm{q-}\mathcal{EP}(0, \mathcal{C})$ reduces to $\mathcal{GP}(0, \mathcal{C})$. When $q \in (0, 2)$, Q-EP imposes stronger regularization than GP.

Modeling with Derivative Information

- Let $u \sim \mathrm{q}{-}\mathcal{EP}(0, \mathcal{C})$. 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.
- $\tilde{u} \sim \mathrm{q}{-}\mathcal{EP}(0, \tilde{\mathcal{C}})$ 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.

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



Bayesian Model of PDE Solutions

with variational Q-EP

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

$$\begin{aligned} \mathcal{D}(u)(\mathbf{x}) &= f(\mathbf{x}), \quad \mathbf{x} \in \Omega, \\ \mathcal{B}(u)(\mathbf{x}) &= g(\mathbf{x}), \quad \mathbf{x} \in \partial\Omega. \end{aligned} \tag{1}$$

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 $g \in L^q(\partial\Omega)$.

- Let $\overline{\Omega} = \Omega \cup \partial\Omega$, $\mathcal{P} = (\mathcal{D}, \mathcal{B}) : B^{s,q}(\overline{\Omega}) \to L^q(\overline{\Omega})$, $h = (f, g) \in L^q(\overline{\Omega})$.
- Assume $P : \mathbb{R}^D \to \mathbb{R}$ such that $\mathcal{P}(u)(\mathbf{x}) = P(\tilde{u}(\mathbf{x}))$ with $\|\nabla P\| \le C$.
- Model $\tilde{u}(\mathbf{X}) \sim \mathrm{q-ED}(\tilde{\mathbf{u}}, \mathbf{S})$ with some variational distribution. To define likelihood, we propagate this distribution by linearizing P.
- Let $\mathbf{Y} = P(\tilde{u}(\mathbf{X}))$ and $\mathbf{h} = h(\mathbf{X})$. We have the solution to (1) probabilistically as the posterior of the Bayesian model

$$\begin{aligned} \mathbf{Y}|\tilde{u}(\mathbf{X}), \mathbf{h} &\sim \mathrm{q-ED}_N(\mathbf{h}, \boldsymbol{\Gamma}), \\ \tilde{u} &\sim \mathrm{q-}\mathcal{EP}(0, \tilde{\mathcal{C}}). \end{aligned} \tag{2}$$

- We approximate the posterior to (2) using sparse variational Bayes [22, 14].

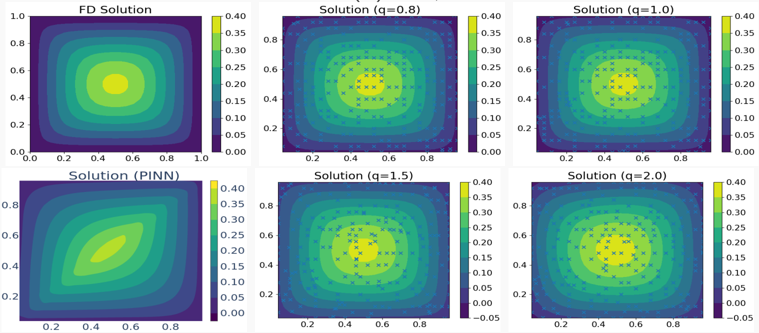
Theorem (Posterior Contraction)

Let $u \sim \mathrm{q}\!-\!\mathcal{EP}(0, \mathcal{C})$ with $\mathcal{C}$ in $L^q(\Omega)$ satisfying certain assumptions. If the true solution to (1) $u^\dagger \in B^{s^\dagger, q^\dagger}(\Omega)$ 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 (2) 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 $q = 1$.

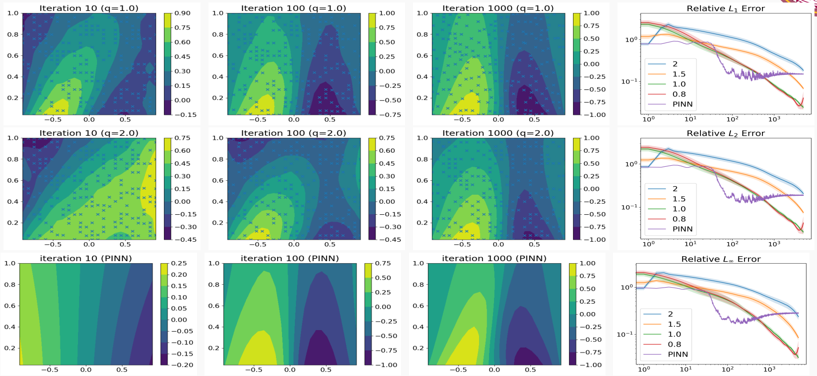
Eikonal Equation

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



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



Model (q)	MAE	MSE	RLE-1	RLE-2	RL-∞
PINN	5.81e-2 ± 1.03e-4	7.12e-3 ± 3.96e-4	0.1508 ± 0.0017	0.1896 ± 0.0052	0.2842 ± 0.0071
B-PINN	2.94e-2 ± 1.57e-2	1.67e-3 ± 1.58e-3	0.0785 ± 0.0420	0.0833 ± 0.0409	0.1327 ± 0.0446
0.5	7.77e-2 ± 7.09e-2	3.50e-2 ± 7.15e-2	0.2018 ± 0.3751	0.2056 ± 0.3764	0.2177 ± 0.3838
1.0	**9.33e-3** ± 1.76e-4	**1.77e-4** ± 2.16e-4	**0.0242** ± 0.0128	**0.0266** ± 0.0140	**0.0324** ± 0.0145
1.5	2.64e-2 ± 8.58e-4	1.31e-3 ± 1.23e-3	0.0684 ± 0.0262	0.0754 ± 0.0316	0.0854 ± 0.0430
2.0(Gaussian)	7.13e-2 ± 8.68e-3	1.08e-2 ± 1.28e-2	0.1848 ± 0.0908	0.2068 ± 0.1118	0.2376 ± 0.1456
2.5	2.26e-1 ± 3.20e-1	2.49e-1 ± 6.04e-1	0.5879 ± 0.8296	0.6669 ± 0.9576	0.7558 ± 1.0521

Conclusion

- In this paper, we propose a novel probabilistic PDE solver based on q-exponential process (Q-EP).
- We theoretically justify and numerically demonstrate why Q-EP with $q = 1$ is the preferable choice over GP with $q = 2$.
- In future, we will extend this work to diffusion probabilistic models with q-exponential noise.

https://github.com/lanzithinking/Diff_QEP

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

References I

[1] 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.

[2] 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.

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

[4] 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.

[5] 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.

[6] 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.

[7] 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.

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

References II

[9] 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.

[10] 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.

[11] 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.

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

[13] 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.

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

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

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

References III

[17] 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.

[18] 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.

[19] 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.

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

[21] 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.

[22] 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.

[23] 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