Improved Black-box Variational Inference for High-dimensional Bayesian Inversion involving Black-box Simulators

Dhruv V. Patel
Stanford University
Stanford, CA 94305, USA
dvpatel@stanford.edu

Jonghyun Lee
University of Hawaii at Manoa
Honolulu, HI 96822, USA
jonghyun.harry.lee@hawaii.edu

Matthew W. Farthing
U.S. Army ERDC
Vicksburg, MS 39183, US
matthew.w.farthing@erdc.dren.mil

Tyler Hesser
U.S. Army ERDC
Vicksburg, MS 39183, US
tyler.hesser@usace.army.mil

Peter K. Kitanidis
Stanford University
Stanford, CA 94305, USA
peterk@stanford.edu

Eric F. Darve
Stanford University
Stanford, CA 94305, USA
darve@stanford.edu

Abstract

Black-box forward model simulators are widely used in scientific and engineering domains for their exceptional capability to mimic complex physical systems. However, applying current state-of-the-art gradient-based Bayesian inference techniques like Hamiltonian Monte Carlo or Variational Inference with them becomes infeasible due to the opaque nature of these simulators. We address this challenge by introducing a modular approach that combines black-box variational inference (BBVI) with deep generative priors, making it possible to efficiently and accurately perform high-dimensional Bayesian inversion in these settings. Our method introduces a novel gradient correction term and a sampling strategy for BBVI, which collectively diminish gradient errors by several orders of magnitude across different dimensions, even with minimal batch sizes. Furthermore, integrating our method with Generative Adversarial Network (GAN)-based priors enables the solution of high-dimensional inverse problems. We validate our algorithm's effectiveness on a range of physics-based inverse problems using both simulated and experimental data. In comparison to Markov Chain Monte Carlo (MCMC) methods, our approach consistently delivers superior accuracy and substantial improvements in both statistical and computational efficiency, often by an order of magnitude.

1 Introduction

The ever-increasing computational power of modern hardware, coupled with the expressiveness of advanced programming languages, has facilitated the development of complex, high-fidelity forward model simulators. These simulators excel in accurately modeling multi-scale, multi-physics, and multi-phase phenomena of interest, finding applications across a spectrum of scientific domains, from computational fluid dynamics [1] to geophysics [2, 3] and beyond [4]. These versatile simulators empower us to describe and predict system behavior under diverse circumstances.

37th Conference on Neural Information Processing Systems (NeurIPS 2023), Deep Learning and Inverse Problems Workshop.

While these sophisticated "black-box simulators" deliver high-fidelity solutions, unfortunately, they are ill-suited for performing statistical inference with state-of-the-art gradient-based inference algorithms such as Hamiltonian Monte Carlo (HMC), Langevin Dynamics, or variational inference (VI) due to their black-box nature. The term "black-box simulator" in this context refers to the forward model simulator/computer program (denoted as $\boldsymbol{f}$), capable of taking various problem parameters, $\boldsymbol{x}$, as input (including geometric descriptions of the domain, boundary conditions, initial conditions, and material property distributions within the domain) and producing corresponding solutions, $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$. However, their limitations become evident in their inability to provide gradients of the simulator's output with respect to its inputs, i.e., $\partial \boldsymbol{y}/\partial \boldsymbol{x}$ remains unavailable. This constraint primarily arises from the inherent characteristics of computer programs representing such simulators—typically developed over several years by multiple researchers, often in legacy programming languages, and highly optimized for forward computations (with potential parallel implementations), making it extremely difficult to integrate with modern automatic differentiation libraries or necessitating substantial manual intervention to incorporate gradient functionality.

The inability to perform efficient Bayesian inference with such high-fidelity black-box simulators is a significant barrier to scientific progress. Bayesian inference, especially with high-fidelity models, plays a pivotal role in facilitating precise parameter estimation, conducting uncertainty quantification, informing downstream decision-making, hypothesis testing, model validation, and optimal experimental design across diverse fields in computational science and engineering. Hence, inference algorithms amenable to such black-box simulators can unlock many possibilities in these fields. Furthermore, the parameter to be inferred, $\boldsymbol{x}$, in these problems can often be very high-dimensional (10^3–10^9) due to fine spatio-temporal discretizations needed to capture multiscale processes. Current state-of-the-art inference techniques (both MCMC and VI-based) struggle and often fail to converge in such high-dimensional parameter space—further hampering scientific progress.

The goal of *accurate and efficient high-dimensional Bayesian inversion involving black-box forward model simulators*, therefore, requires designing a novel inference algorithm that can integrate and leverage the black-box nature of such simulators in the inference process while being computationally efficient. To that end, we propose a modular and efficient method that leverages modified black-box variational inference (BBVI) and deep generative priors. BBVI [5] enables extracting valuable information from black-box forward models without requiring its gradient, while GAN-based priors [6, 7] enable performing inference in the low-dimensional latent space. Concretely, we make the following contributions:

- We propose a modular method harnessing modified BBVI and GAN-priors to solve *high-dimensional* Bayesian inverse problems involving *black-box forward model simulators* in a highly efficient manner.
- The traditional BBVI gradient (first proposed in [5]) is sample inefficient. We propose a novel gradient correction term and sampling strategy that reduces gradient error by four to five orders of magnitude across problems of varying dimensionalities, even with very small batch sizes, resulting in significant computational savings.
- We demonstrate the effectiveness of the proposed method on both in-silico and experimental data, where it outperforms traditional MCMC algorithms on both accuracy as well as computational and statistical efficiency for the same compute budget.

2 Method

2.1 Background

Bayesian inference Consider the parameter vector $\boldsymbol{x}$ and solution vector $\boldsymbol{y}$ related by the forward model $\boldsymbol{f}$ such that $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$. The goal of the Bayesian inference is to infer the posterior probability distribution of parameters $\boldsymbol{x}$ given a noisy version of measurement, $\tilde{\boldsymbol{y}} = \boldsymbol{y} + \boldsymbol{\eta}$, i.e., $p(\boldsymbol{x}|\tilde{\boldsymbol{y}})$. It is obtained by invoking the Bayes' rule, $p_{\mathcal{X}}^{\text{post}}(\boldsymbol{x}|\tilde{\boldsymbol{y}}) \propto p_\eta(\tilde{\boldsymbol{y}} - \boldsymbol{f}(\boldsymbol{x})) p_{\mathcal{X}}^{\text{prior}}(\boldsymbol{x})$, where, $p_{\mathcal{X}}^{\text{prior}}(\boldsymbol{x})$ is the prior, and the likelihood term, $p_{\mathcal{Y}}^{\text{like}}(\tilde{\boldsymbol{y}}|\boldsymbol{x}) = p_\eta(\tilde{\boldsymbol{y}} - \boldsymbol{f}(\boldsymbol{x}))$, injects physics into the framework via the direct operator $\boldsymbol{f}$. Statistical inference involves characterizing the posterior distribution by computing the expectation of a statistical quantity of interest (QoI) (defined by $l(\boldsymbol{x})$) with respect to this posterior, $\mathbb{E}_{\boldsymbol{x} \sim p_{\mathcal{X}}^{\text{post}}(\boldsymbol{x}|\tilde{\boldsymbol{y}})}[l(\boldsymbol{x})]$.

Variational inference (VI) The posterior distribution is intractable in general for typical physics-based inverse problems due to the high dimensionality of $\boldsymbol{x}$. Variational inference is an elegant method to approximate this intractable posterior distribution with other tractable variational distributions. Concretely, in VI, we posit a variational family of distributions $q_\lambda(\boldsymbol{x})$ and find the value of variational parameters $\boldsymbol{\lambda}$ such that the variational distribution is close (in Kullback–Leibler (KL) divergence sense) to the posterior distribution, i.e., $\boldsymbol{\lambda}^* = \arg\min_{\boldsymbol{\lambda}} \mathrm{KL}(q_\lambda(\boldsymbol{x}) \| p_{\mathcal{X}}^{\text{post}}(\boldsymbol{x}|\tilde{\boldsymbol{y}}))$. Thus, VI turns an inference problem into an optimization problem. Computing the KL divergence, however, is not possible since it requires the value of the target distribution we are interested in. Hence, instead, an evidence lower bound (ELBO) of the above divergence is computed and maximized to find the optimal variational parameters $\boldsymbol{\lambda}^* = \arg\max_{\boldsymbol{\lambda}} \mathcal{L}$, where $\mathcal{L} = \text{ELBO} = \mathop{\mathbb{E}}_{q_\lambda} [\log p(\boldsymbol{x}, \tilde{\boldsymbol{y}}) - \log q_\lambda(\boldsymbol{x})]$. Inversion examples are illustrated in the result section 3.

Black-box Variational Inference (BBVI) Performing Bayesian inference with VI entails finding the gradient of ELBO and performing gradient-based optimization of $\boldsymbol{\lambda}$. To tackle the challenge of performing VI with non-differentiable simulators, we leverage the black-box variational inference method originally proposed in [5] for latent variable models and adapt it to physics-based inverse problems to allow inference without differentiating through the forward model. Using the BBVI derivation, we can prove that

$$\nabla_{\boldsymbol{\lambda}}^{\text{BBVI}} \mathcal{L} = \mathop{\mathbb{E}}_{\boldsymbol{x} \sim q_\lambda(\boldsymbol{x})} \Big[\nabla_{\boldsymbol{\lambda}} \log q_\lambda(\boldsymbol{x}) \Big\{ \log p_{\mathcal{X}}^{\text{prior}}(\boldsymbol{x}) + \log p_\eta^{\text{like}}(\tilde{\boldsymbol{y}} - \boldsymbol{f}(\boldsymbol{x})) - \log q_\lambda(\boldsymbol{x}) \Big\} \Big] \tag{1}$$

which does not require computing the gradient of $\boldsymbol{f}$. The expectation is approximated by Monte Carlo (MC) sampling. We refer to this gradient as the BBVI gradient. We note that the "log-derivative trick" used in the above gradient derivation appears in different places in ML under various names, such as REINFORCE in reinforcement learning, the likelihood ratio method, or the score function estimator [8–10] with the central idea being approximating the gradient of an expectation as an expectation of a gradient.

2.2 Improved BBVI for High-Dimensional Physics-based Bayesian Inversion

It is well-established that the BBVI gradient estimator of ELBO (shown in (1)) suffers from high variance [10–12]. This necessitates a large number of samples to attain gradients with reasonable accuracy. In the context of high-dimensional physics-based inverse problems, this is problematic since it mandates the solution of a large number of computationally demanding PDEs (forward model) at each optimization iteration, resulting in substantial computational expense. This is one of the reasons for the limited adoption of BBVI for practical PDE-based Bayesian inversion. Here we propose strategies to tackle this issue.

Gradient correction The variance in the BBVI gradient is due to the randomness introduced by the MC sampling. By exploiting the structure of the assumed variational distribution, we can estimate this gradient more accurately by reducing the variance. We do this by first assuming the variational distribution to be multivariate Gaussian, i.e., $q_\lambda(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{I\sigma^2})$ with $\boldsymbol{\lambda} = [\boldsymbol{\mu}, \boldsymbol{\sigma^2}]$ and introduce finite difference gradient correction term (and special sampling strategy complementing this). Specifically, let $l(\boldsymbol{x}) = \log p_{\mathcal{X}}^{\text{prior}}(\boldsymbol{x}) + \log p_\eta^{\text{like}}(\tilde{\boldsymbol{y}} - \boldsymbol{f}(\boldsymbol{x})) - \log q_\lambda(\boldsymbol{x})$. We can rewrite (1) as

$$\nabla_{\boldsymbol{\lambda}}^{\text{BBVI}} \mathcal{L} = \mathop{\mathbb{E}}_{q_\lambda(\boldsymbol{x})} \big[\nabla_{\boldsymbol{\lambda}} \log q_\lambda(\boldsymbol{x}) \{ l(\boldsymbol{x}) - l'_{FD}(\boldsymbol{\lambda})^T (\boldsymbol{x} - \boldsymbol{\lambda}) \} \big] + \mathop{\mathbb{E}}_{q_\lambda(\boldsymbol{x})} \big[\nabla_{\boldsymbol{\lambda}} \log q_\lambda(\boldsymbol{x}) \{ l'_{FD}(\boldsymbol{\lambda})^T (\boldsymbol{x} - \boldsymbol{\lambda}) \} \big] \tag{2}$$

where $l'_{FD}(\boldsymbol{\lambda})$ is a finite difference approximation of the derivative of $l(\boldsymbol{\lambda})$. It is important to note that $l'_{FD}(\boldsymbol{\lambda})$ does not require differentiating through the forward model and can be easily computed using any standard finite difference scheme (such as a central difference). To understand the effect of this correction term, let us focus on $\boldsymbol{\lambda} = \boldsymbol{\mu}$. With this (2) simplifies to

$$\nabla_{\boldsymbol{\mu}}^{\text{BBVI}} \mathcal{L} = \mathop{\mathbb{E}}_{q_\lambda(\boldsymbol{x})} \big[\nabla_{\boldsymbol{\mu}} \log q_\lambda(\boldsymbol{x}) \{ l(\boldsymbol{x}) - l'_{FD}(\boldsymbol{\mu})^T (\boldsymbol{x} - \boldsymbol{\mu}) \} \big] + l'_{FD}(\boldsymbol{\mu}). \tag{3}$$

Here, since the second term is independent of $\boldsymbol{x}$ it has effectively zero variance, and it approximates the gradient of the ELBO quite accurately, whereas the first term acts as an error correction term. To reduce the variance of the total gradient in (3), we can further use the control variate method [5]. We demonstrate in the results section that this overall scheme leads to a reduction in error in approximating the gradient by several orders of magnitude over traditional BBVI gradient evaluation, even with very small batch sizes for a prototypical problem with varying dimensionalities.

Deep generative prior While the finite difference correction improves gradient accuracy, its computational cost scales with the dimensionality of the problem. In high-dimensional physics-based inverse problems, where $\boldsymbol{x}$ can be of significant dimension, this can offset the computational gains achieved through the correction term. To mitigate this, we leverage recently proposed GAN-priors [6, 13], which is a sample-based prior method that re-frames the posterior distribution $p_{\mathcal{X}}^{\text{post}}(\boldsymbol{x}|\tilde{\boldsymbol{y}})$ in a lower-dimensional latent space, $\boldsymbol{z}$, of a pre-trained GAN generator. While traditionally GAN-priors have been used with MCMC, one can use them with VI as well by choosing an appropriate variational distribution in the latent space $q_{\lambda}(\boldsymbol{z})$ of the generator and finding the optimal value of $\boldsymbol{\lambda}$ by maximizing the corresponding ELBO. We note a recent study [14] where GAN-priors is used in conjunction with VI in a "white-box" setting. In contrast, our proposed method addresses the challenge with "black-box" setting.

Sampling Strategy Moreover, as the variational distribution is Gaussian, we can re-parameterize its samples using a random normal distribution as $\boldsymbol{z} = \boldsymbol{\xi} \odot \boldsymbol{\sigma} + \boldsymbol{\mu}$ with $\boldsymbol{\xi} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$. The set $\{\boldsymbol{\xi}\}_{i=1}^{N}$ can be obtained through various methods to approximate the expectation in the gradient. Here, we explore three strategies: (i) Normal: drawing N points from a normal distribution, (ii) Symmetric Normal: sampling $N/2$ points from a normal distribution and selecting the remaining $N/2$ points by changing the sign of the first $N/2$ points, (iii) Non-uniform Deterministic: an importance sampling strategy with uniform weights but a non-uniform distribution of points. These points are distributed such that their spacing is inversely proportional to the normal or Gaussian distribution.

3 Results

Bayesian inversion with conjugate priors To understand the effect of additional gradient correction term and sampling scheme on the accuracy of the gradient, we first consider a Bayesian inverse problem involving conjugate priors as used in earlier studies [7]. A detailed description of the problem and the result is provided in Section 5.1.3.

Physics-based inversion with synthetic data (inverse heat conduction) In our study, we direct our attention towards practical physics-based Bayesian inverse problems. Specifically, we tackle the challenging task of inferring the initial condition in the time-dependent heat conduction equation, omitting a source term, and employing a constant conductivity field with a diffusivity of $\kappa = 0.064$ units. Here, the initial condition is prescribed as zero everywhere inside the domain except in a rectangle region, where it varies linearly from left (with the value of 2 units) to right (with the value of 4 units), as can be seen from Figure 1(a) in Section 5.1.1. Given a noisy temperature field at $t = 1$, characterized by a noise variance of 1, our goal is to recover the initial condition. This problem presents a severe ill-posed nature due to information loss during the diffusion process and the presence of substantial measurement noise. We represent the discretized initial and final temperature fields as $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively, on a 32×32 Cartesian grid. Our approach leverages a GAN-prior, trained on a parametric rectangular dataset utilized in prior studies [7, 15], within the 5-dimensional latent space of the GAN. For optimization, we employ the Adam optimizer [16] in conjunction with the Normal Symmetric sampling scheme for BBVI. As a baseline, we compare our method against the random walk Markov Chain Monte Carlo (MCMC) approach, often used with black-box simulators in physics-based inverse problems. We optimally tune the proposal variance for MCMC. Qualitative results, shown in Figure 1(a), indicate slightly superior reconstruction by BBVI. Figure 1(b) presents a quantitative analysis. We solve the inverse problem with a fixed number of forward solves for both methods and assess the error in posterior statistics. Notably, BBVI consistently demonstrates smaller errors and significantly faster convergence rates compared to MCMC, underscoring its superiority in addressing this challenging problem.

Physics-based inversion with real-world experimental data (hydraulic tomography) In this study, we validate our method using experimental data. Specifically, we solve the hydraulic tomography problem of inferring the permeability field from sparse measurements of a pressure field. This is an important problem in the geophysics and petroleum engineering communities. The forward model is defined by Darcy flow. The details of the experimental setup, data, and forward model are provided in [17]. We use a GAN-prior trained using a parametric dataset of different rectangular fields randomly located inside the domain, mimicking the sandbox structure with sands with different permeabilities.

We perform inference using the proposed BBVI method and compare it against random walk MCMC, and the reconstruction results are provided in Figure 2 in Section 5.1.2. As can be observed qualitatively from the figure, BBVI performs significantly better than MCMC. A quantitative comparison of both methods is provided in Table 1 in Section 5.1.2.

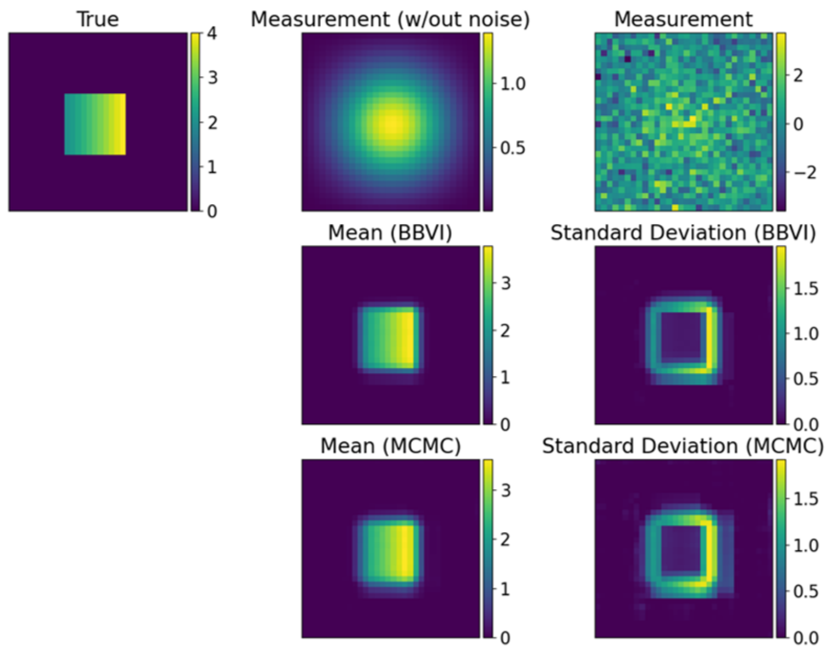
4 Conclusion

In this study, we have developed a modular framework that integrates black-box variational inference (BBVI) with deep generative priors to effectively address high-dimensional Bayesian inversion in the context of black-box forward model simulators. Our method, validated across practical use-cases, consistently outperforms traditional Markov Chain Monte Carlo (MCMC) techniques in both accuracy and convergence speed. By overcoming the inherent challenges of high-dimensional spaces and black-box models, our approach paves the way for advancements in Bayesian inference across diverse scientific fields.

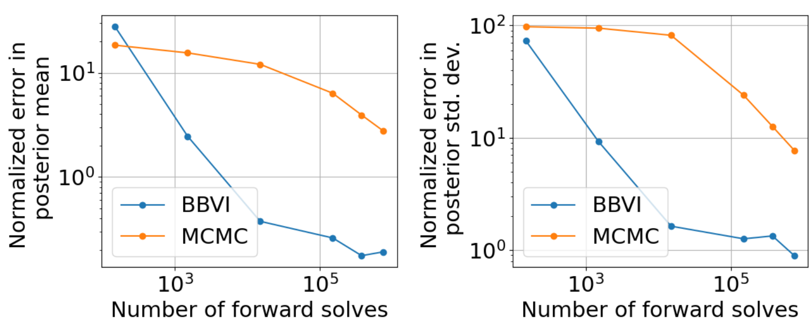
5 Supplementary Material

5.1 Figures and Additional Results

5.1.1 Physics-based inversion with synthetic data (inverse heat conduction)



(a)



(b)

Figure 1: (a) Initial condition inversion. *Top row*: true inferred field $\boldsymbol{x}$, final temperature $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$, measured temperature field $\tilde{\boldsymbol{y}} = \boldsymbol{y} + \boldsymbol{\eta}$. *Second and third row*: posterior mean and standard deviation obtained using BBVI and MCMC, respectively. (b) Normalized error in posterior statistics as a function of the number of forward solves for BBVI and MCMC.

5.1.2 Physics-based inversion with experimental data (hydraulic tomography)

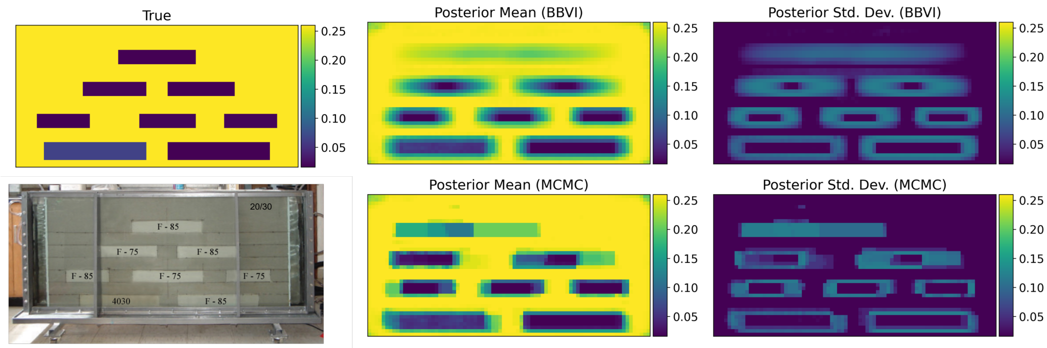


Figure 2: Hydraulic tomography using experimental data. *First column*: (top) true permeability field, (bottom) photograph of the sandbox used during the experiment. *Second column*: posterior mean obtained using BBVI and MCMC method. *Third column*: posterior standard deviation obtained using BBVI and MCMC method.

Table 1: Quantitative comparison of MCMC and BBVI

For a fixed number of forward solves

QoI	MCMC	**BBVI**
Error ↓	31.58%	**26.06%**
Coverage (95% CI) ↑	0.760	**0.903**
Time (in seconds) ↓	108.25	**11.07**

For fixed CPU wall-clock time

QoI	MCMC	**BBVI**
Error ↓	39.81%	**26.09%**
Coverage (95% CI) ↑	0.274	**0.900**
No. of forward solves ↑	3000	**72000**

5.1.3 Bayesian inversion with conjugate priors

In this study, we consider a Bayesian inverse problem with Gaussian prior and Gaussian likelihood resulting in analytically tractable Gaussian posterior. We approximate this Gaussian posterior with a variational distribution whose mean and variance vectors are optimized. We consider problems of varying dimensionalities. The detailed results are presented in Figure 3.

5.1.4 1D Pedagogical toy problem

We examine an additional pedagogical example that illustrates the application of BBVI in a 1D setting for visualization purposes. In this example, the prior and likelihood distributions are both Gaussian, while the forward model $f(x) = 0.2x^3 \sin(x)$ is non-linear, leading to a non-convex posterior distribution. Specifically, we choose the prior and the likelihood distribution as $N(0, 1)$, and set the observation as $\tilde{y} = 2$. To compute the true posterior statistics, we utilize a Monte Carlo estimate with 1 million samples. For comparison, we consider a baseline method that utilizes a quadratic approximation approach. This method iteratively approximates the target density with a local quadratic function, aiming to capture the posterior characteristics.

Figure 4 shows the comparison of BBVI with quadratic approximation and MCMC for the non-linear forward model. Here, the term quadratic approximation means fitting the log of the posterior with three points (with the central point approximating the mean and two side points are one standard deviation away from the central point). The figure showcases both qualitative (first column) and quantitative (next three columns) comparisons of the performance of these methods. From the results, it is evident that BBVI outperforms the quadratic approximation and MCMC methods in this simple 1D toy problem with multi-modal posterior, exhibiting superior performance. Further, quadratic approximation exhibits mode-seeking behavior, whereas BBVI tries to approximate the whole distribution (as also indicated by their KL divergence). It is worth noting that while we have employed a Gaussian distribution as our variational family in this example (and hence have selected quadratic approximation as the corresponding baseline), it is possible to select more expressive

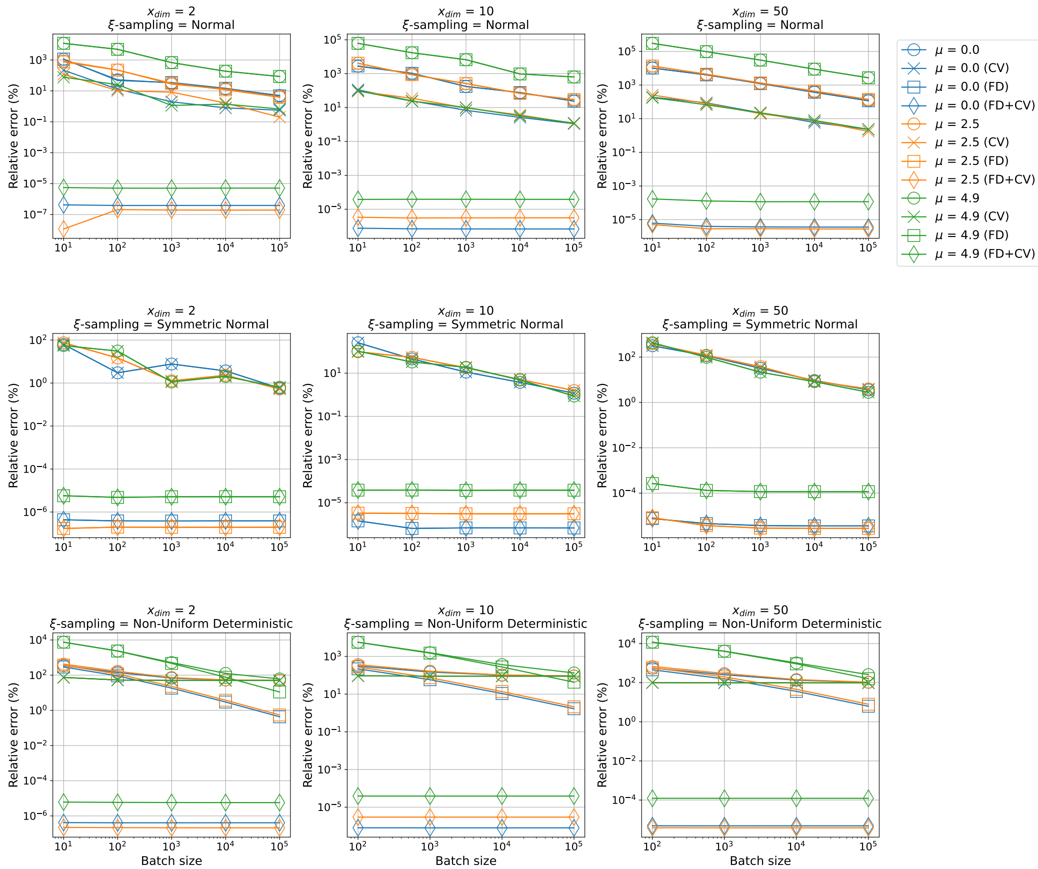


Figure 3: Relative error in gradient as a function of batch size. Each row represents a different sampling scheme, and each column represents the problem dimension. Different colors in each subplot indicate different points at which the gradient is evaluated, and different markers represent different gradients used.

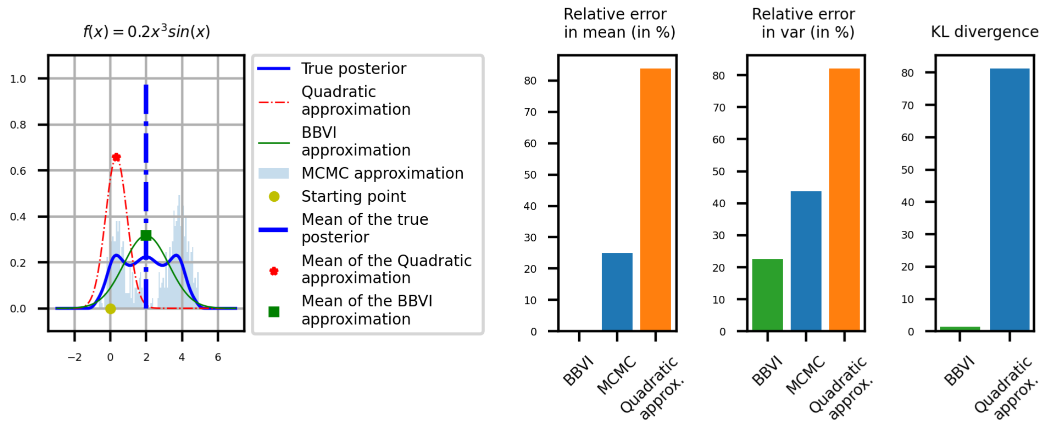


Figure 4: 1D pedagogical example: comparison of BBVI with quadratic approximation and MCMC for $f(x) = 0.2x^3 \sin(x)$. *Second, third, and fourth columns*: relative error in posterior mean, variance, and KL divergence between target posterior distribution and approximated distribution, respectively.

distributions that can closely capture the characteristics of the target density resulting in even more improvement in quantitative metrics.

References

[1] Cameron W. Smith, Eroma Abeysinghe, Suresh Marru, and Kenneth E. Jansen. Phasta science gateway for high performance computational fluid dynamics. In *Proceedings of the Practice and Experience on Advanced Research Computing*, PEARC '18, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450364461. doi: 10.1145/3219104.3229243. URL `https://doi.org/10.1145/3219104.3229243`.

[2] Arlen W Harbaugh. *MODFLOW-2005, the US Geological Survey modular ground-water model: the ground-water flow process*, volume 6. US Department of the Interior, US Geological Survey Reston, VA, USA, 2005.

[3] Glenn E. Hammond, Peter C. Lichtner, and Richard T. Mills. Evaluating the performance of parallel subsurface simulators: An illustrative example with pflotran. *Water Resources Research*, 50:208–228, 2014. doi: 10.1002/2012WR013483.

[4] David E Keyes, Lois C McInnes, Carol Woodward, William Gropp, Eric Myra, Michael Pernice, John Bell, Jed Brown, Alain Clo, Jeffrey Connors, Emil Constantinescu, Don Estep, Kate Evans, Charbel Farhat, Ammar Hakim, Glenn Hammond, Glen Hansen, Judith Hill, Tobin Isaac, Xiangmin Jiao, Kirk Jordan, Dinesh Kaushik, Efthimios Kaxiras, Alice Koniges, Kihwan Lee, Aaron Lott, Qiming Lu, John Magerlein, Reed Maxwell, Michael McCourt, Miriam Mehl, Roger Pawlowski, Amanda P Randles, Daniel Reynolds, Beatrice Rivière, Ulrich Rüde, Tim Scheibe, John Shadid, Brendan Sheehan, Mark Shephard, Andrew Siegel, Barry Smith, Xianzhu Tang, Cian Wilson, and Barbara Wohlmuth. Multiphysics simulations: Challenges and opportunities. *The International Journal of High Performance Computing Applications*, 27(1): 4–83, 2013. doi: 10.1177/1094342012468181.

[5] Rajesh Ranganath, Sean Gerrish, and David Blei. Black Box Variational Inference. 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 814–822, Reykjavik, Iceland, 22–25 Apr 2014. PMLR. URL `https://proceedings.mlr.press/v33/ranganath14.html`.

[6] Eric Laloy, Romain Hérault, Diederik Jacques, and Niklas Linde. Training-image based geostatistical inversion using a spatial generative adversarial neural network. *Water Resources Research*, 54(1):381–406, Jan 2018. ISSN 1944-7973. doi: 10.1002/2017wr022148. URL `http://dx.doi.org/10.1002/2017WR022148`.

[7] Dhruv V Patel, Deep Ray, and Assad A Oberai. Solution of physics-based bayesian inverse problems with deep generative priors. *Computer Methods in Applied Mechanics and Engineering*, 400:115428, 2022.

[8] Ronald J Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8:229–256, 1992.

[9] Peter W Glynn. Likelihood ratio gradient estimation for stochastic systems. *Communications of the ACM*, 33(10):75–84, 1990.

[10] Shakir Mohamed, Mihaela Rosca, Michael Figurnov, and Andriy Mnih. Monte carlo gradient estimation in machine learning. *The Journal of Machine Learning Research*, 21(1):5183–5244, 2020.

[11] Francesco Locatello, Gideon Dresdner, Rajiv Khanna, Isabel Valera, and Gunnar Rätsch. Boosting black box variational inference. *Advances in Neural Information Processing Systems*, 31, 2018.

[12] Rajesh Ranganath, Dustin Tran, and David Blei. Hierarchical variational models. In *International conference on machine learning*, pages 324–333. PMLR, 2016.

[13] Dhruv V. Patel and Assad A. Oberai. Gan-based priors for quantifying uncertainty in supervised learning. *SIAM/ASA Journal on Uncertainty Quantification*, 9(3):1314–1343, 2021. doi: 10.1137/20M1354210. URL `https://doi.org/10.1137/20M1354210`.

[14] Agnimitra Dasgupta, Dhruv V Patel, Deep Ray, Erik A Johnson, and Assad A Oberai. A dimension-reduced variational approach for solving physics-based inverse problems using generative adversarial network priors and normalizing flows. *arXiv preprint arXiv:2310.04690*, 2023.

[15] Dhruv V Patel, Deep Ray, Harisankar Ramaswamy, and Assad Oberai. Bayesian inference in physics-driven problems with adversarial priors. In *NeurIPS 2020 Workshop on Deep Learning and Inverse Problems*, 2020.

[16] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[17] X Liu and PK Kitanidis. Large-scale inverse modeling with an application in hydraulic tomography. *Water Resources Research*, 47(2), 2011.