
Moment Estimates and DeepRitz Methods on Learning Diffusion Systems with Non-gradient Drifts

Fanze Kong*

Department of Applied Mathematics
University of Washington
Seattle, WA 98105, USA
fzkong@uw.edu

Chen-Chih Lai

Department of Mathematics
Columbia University
New York, NY 10027, USA
cclai.math@gmail.com

Yubin Lu

Department of Applied Mathematics
Illinois Institute of Technology
Chicago, IL 60616, USA
ylu117@illinoisitech.edu

Abstract

Conservative-dissipative dynamics are ubiquitous across a variety of complex open systems. We propose a data-driven two-phase method, the Moment-DeepRitz Method, for learning drift decompositions in generalized diffusion systems involving conservative-dissipative dynamics. The method is robust to noisy data, adaptable to rough potentials and oscillatory rotations. We demonstrate its effectiveness through several numerical experiments.

1 Introduction

Nonlinear stochastic dynamics are ubiquitous in complex systems across physics, biology, and engineering. Incorporating the randomness perturbations (e.g. Gaussian noise) in deterministic ordinary differential equations (ODEs), stochastic differential equations (SDEs) provide a powerful mathematical framework for modeling such complex dynamics [1, 2]. By employing simple SDEs with gradient drifts, whose density functions are governed by potential Fokker-Planck equations (typically Langevin equations), a huge volume of research has been done for investigating the equilibrium behaviors of particle systems including the relation between temperature, pressure and entropy at the macroscopic level [3, 4, 5]. However, while investigating the non-equilibrium thermodynamics of particle systems, various complex phenomena occur, including entropy production [6, 7, 8], conservative-dissipative dynamics [9, 10, 11], non-Boltzmann steady states [12, 13], etc. Non-detailed balance is a crucial ingredient in the occurrence of intricate statistical behaviors of collective particles mentioned above and Fokker-Planck equations with the non-gradient structures serve as paradigms to understand such property.

Research on the discovery of physical laws described by ODEs and SDEs from observational data via data-driven approaches has been growing rapidly, see e.g. [14, 15, 16, 17, 18, 19]. The main idea is based on the machine learning framework, for example neural networks are trained to minimize suitable physics-informed loss functions. The loss functions can take various forms, such as PDE-based forms [14, 15, 20], weak forms [21, 22] and variational forms [19, 23, 24, 25]. Due to the fundamental connection between non-detailed balance and non-gradient drift structures, the precise identification of drift decomposition plays a crucial role in the investigation of non-equilibrium

*Corresponding author

thermodynamics. On one hand, the rotational components of the drift provide a way to compute the entropy production of the diffusion process evolving on Riemannian manifolds with non-trivial topology [26]. On the other hand, the computation of irrotational components is significant in the large deviation theory [27]. Motivated by the results of Lu et al. [19], Kong et al. [25] proposed a two-stage data-driven framework for learning the drift decomposition, satisfying pointwise orthogonality, in generalized diffusions by combining the evolution of first moments with the energy dissipation law.

In this paper, our objective is to extend the results in [25] and recover the drift decomposition in Fokker-Planck equations with the more general non-gradient drifts. We remark that, Kong et al. [25] only address the case in which the drift terms satisfy a pointwise orthogonal decomposition, whereas the present work proposes a data-driven framework for tackling general scenarios by combining the first-moment dynamics with DeepRitz method.

In Section 2, we introduce the Moment-DeepRitz method for identifying pseudo-potential and rotation components in diffusion systems. Section 3 presents two-dimensional experiments, and Section 4 summarizes results and open problems.

2 Learning framework

We present the Moment-DeepRitz method for learning generalized diffusions with non-gradient structures, starting from the stochastic system governed by the following SDE:

$$d\mathbf{X}_t = \mathbf{b}(\mathbf{X}_t)dt + \sigma(\mathbf{X}_t)d\mathbf{W}_t, \quad \mathbf{X}_t = (x_1^t, \dots, x_d^t) \in \mathbb{R}^d, \quad (1)$$

where $\mathbf{b} = (b_1, \dots, b_d) : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a drift term and \mathbf{W}_t denotes a standard d -dimensional Brownian motion with a scalar noise intensity $\sigma = \sigma(\mathbf{X}_t)$. The probability density function $f(\mathbf{x}, t)$ of \mathbf{X}_t evolves according to the Fokker-Planck equation

$$\begin{cases} \partial_t f + \nabla \cdot (\mathbf{b}f) = \frac{1}{2}\Delta(\sigma^2 f), & \mathbf{x} \in \mathbb{R}^d, t > 0, \\ f(\mathbf{x}, 0) = f_0, & \mathbf{x} \in \mathbb{R}^d, \end{cases} \quad (2)$$

where the classical Itô integral formula has been applied to (1) and f_0 denotes the density of initial state \mathbf{X}_0 . The evolution of the first moment $\int_{\mathbb{R}^d} f(\mathbf{x})\mathbf{x} d\mathbf{x}$ is given by (see, e.g., [25])

$$\frac{d}{dt} \int_{\mathbb{R}^d} f x_i d\mathbf{x} = \int_{\mathbb{R}^d} b_i f d\mathbf{x}, \quad \mathbf{x} = (x_1, \dots, x_d). \quad (3)$$

To further extract the potential and rotation components of the drift in generalized diffusions, we assume the drift \mathbf{b} can be decomposed uniquely as

$$\mathbf{b} = -\nabla\psi + \mathbf{R}, \quad (4)$$

where $\nabla \cdot \mathbf{R} = 0$. Then, taking the divergence of both sides in (4) yields

$$-\Delta\psi = \nabla \cdot \mathbf{b}. \quad (5)$$

Since ψ solves the Poisson equation (5), it is a minimizer of the following optimization problem

$$\min_{\psi \in H} I(\psi), \quad I(\psi) := \frac{1}{2} \int_{\Omega} |\nabla\psi|^2 d\mathbf{x} + \int_{\Omega} \nabla\psi \cdot \mathbf{b} d\mathbf{x}, \quad (6)$$

where H is the set of admissible functions.

Combining (3) with (6), we propose a deep learning based method, the Moment-DeepRitz method for identifying the physical laws of generalized diffusions: in the first phase, to learn the general drift \mathbf{b} , one formulates the loss function based on the first-moment dynamics (3); while in the second phase, one follows DeepRitz method in [28] and employs the variational form (6) to construct the loss function and recover the irrotational component.

Given the noise intensity σ^2 , we aim to implement the Moment-DeepRitz method in the continuous data observation setting. Here, we define $\mathbf{b}_{\text{NN}}(x; \theta)$ and $\psi_{\text{NN}}(x; \theta)$ as neural network approximations of the drift and potential components \mathbf{b} and $\psi(x)$, respectively.

We construct the training data as follows. For $j = 1, \dots, M$, Gaussian-type initial states $(f_0)_j(\mathbf{x})$ with mean μ_j^0 and variance σ_0^2 generate numerical solutions $f_j(\mathbf{x}, t)$ of the Fokker-Planck equation

(2). Let Δx_k and Δt be the spatial and temporal mesh sizes, and define observation grids $\delta x_k = \Delta x_k$, $\delta t = m\Delta t$. With uniform mesh points \mathbf{x}_i of spacing $\delta \mathbf{x} = (\delta x_1, \dots, \delta x_d)$, the training dataset is $\{f_j(\mathbf{x}_i, t_1), f_j(\mathbf{x}_i, t_2)\}_{i,j=1}^{N,M}$, where $t_2 = t_1 + \delta t$ and $t_1 \ll 1$ is a short transient time.

We now present the Moment-DeepRitz method (see Algorithm 1 in Appendix 1):

Phase 1: First-moment estimates. This phase is identical to Stage 1 of [25]. The objective is to solve the minimization problem

$$\theta_{\mathbf{b}}^* = \underset{\theta}{\operatorname{argmin}} L_{\mathbf{b}}^{\text{dyn}}(\theta). \quad (7)$$

where $L_{\mathbf{b}}^{\text{dyn}}(\theta) = \sum_{k=1}^d L_{b_k}^{\text{dyn}}(\theta)$ in which

$$L_{b_k}^{\text{dyn}}(\theta) = \sum_{j=1}^M \left\| \frac{(\mu_j)_k(t_2) - (\mu_j)_k(t_1)}{t_2 - t_1} - |\delta \mathbf{x}| \sum_{i=1}^N (b_k)_{\text{NN}}(\mathbf{x}_i; \theta) f_j(\mathbf{x}_i, t_1) \right\|^2, \quad k = 1, \dots, d. \quad (8)$$

where $(\mu_j)_k(t) = |\delta \mathbf{x}| \sum_{i=1}^N (x_i)_k f_j(\mathbf{x}_i, t)$, $\mathbf{x}_i = ((x_i)_1, \dots, (x_i)_d)$, denotes the approximation of the centroid of density function f solving (2) for $j = 1, \dots, M$, and $k = 1, \dots, d$.

Phase 2: DeepRitz method. Approximating (6) by a Riemann sum yields the loss function

$$I_{\psi}^{\text{dyn}}(\theta) = |\delta \mathbf{x}| \sum_{i=1}^N \left[\frac{1}{2} |\nabla \psi_{\text{NN}}(\mathbf{x}_i; \theta)|^2 + \nabla \psi_{\text{NN}}(\mathbf{x}_i; \theta) \cdot \mathbf{b}^*(\mathbf{x}_i) \right], \quad \mathbf{b}^*(\mathbf{x}_i) := \mathbf{b}_{\text{NN}}(\mathbf{x}_i; \theta_{\mathbf{b}}^*) \quad (9)$$

where $\theta_{\mathbf{b}}^*$ is given by (7) in Phase 1. Learning ψ reduces to the optimization problem:

$$\theta_{\psi}^* = \underset{\theta}{\operatorname{argmin}} I_{\psi}^{\text{dyn}}(\theta). \quad (10)$$

It is worth mentioning the key differences between our proposed method and those in [27, 25]. The approach in [27] requires estimating the velocity of individual particles. This can yield highly accurate results for learning both the potential and rotational components when a sufficient number of particles are available and the data is minimally affected by noise. In contrast, our method only requires computing the temporal change in ensemble data. Specifically, we estimate the time derivative of the density rather than tracking individual particle velocities. Compared with [25], our method offers two notable advantages. First, it does not require pointwise orthogonality between the gradient of the potential and the rotational component of the drift coefficient. Second, it remains effective even under small or vanishing noise intensity σ , thanks to the use of DeepRitz in the second stage, which circumvents the nonconvex optimization problem encountered when learning the potential function in [25].

3 Numerical examples

We present representative two-dimensional experiments ($d = 2$). Since the Fokker-Planck equation (2) is posed on \mathbb{R}^2 , we approximate its solution on the domain $[-4, 4] \times [-4, 4]$. In Phase 1, observation data are generated from M initial distributions $\mathcal{N}(\boldsymbol{\mu}^0, \sigma_0^2 I_2)$, where $\boldsymbol{\mu}^0$ is uniformly sampled from $[-2, 2] \times [-2, 2]$, $\sigma_0^2 = 0.01$, and I_2 denotes the 2×2 identity matrix.

The training dataset $\{f_j(x_i, y_i, t_1), f_j(x_i, y_i, t_2)\}_{i,j=1}^{N,M}$ is generated by numerically solving the Fokker-Planck equation with spatial resolution $\Delta x = \Delta y = 0.1$ and time step $\Delta t = 0.0001$. Unless otherwise specified, we construct $M = 40$ distinct initial distributions and extract solution snapshots at $t_1 = 0.015$ and $t_2 = 0.016$ (i.e., $m = 10$) for training.

To test the robustness of our method to noisy data, the training samples are perturbed with Gaussian noise of level 0.1, obtained by convolving the numerical solution f with the distribution $\mathcal{N}(\mathbf{0}, 0.1 I_2)$.

For training in both phases, we employ a fully connected neural network with four layers: two hidden layers, each containing 50 neurons. The activation function is **Tanh()**, and the parameters are optimized using the Adam algorithm [29] with a learning rate of 1×10^{-4} . In Phase 1, the network is trained with a batch size of 5. Unless otherwise specified, both phases are trained for 10,000 epochs.

Since ψ is defined up to an additive constant, we shift ψ_{NN} so its mean matches that of ψ , i.e., $\psi_{\text{NN}}(\mathbf{x}) \mapsto \psi_{\text{NN}}(\mathbf{x}) + \frac{1}{N} \sum_{i=1}^N (\psi(\mathbf{x}_i) - \psi_{\text{NN}}(\mathbf{x}_i))$.

To assess the accuracy of our method at each phase, we compute the relative root mean square error (rRMSE) of the drift \mathbf{b} as follows: $\text{rRMSE}_{\mathbf{b}} = \sqrt{\sum_{i=1}^N |\mathbf{b}(\mathbf{x}_i) - \mathbf{b}_{\text{NN}}(\mathbf{x}_i; \theta_{\mathbf{b}}^*)|^2} / \sqrt{\sum_{i=1}^N |\mathbf{b}(\mathbf{x}_i)|^2}$. The rRMSE of ψ (rRMSE_{ψ}) and \mathbf{R} ($\text{rRMSE}_{\mathbf{R}}$) are defined in a similar fashion.

Double-well Potential. Consider the potential $\psi(x, y) = \frac{1}{4}(x^2 - 1)^2 + \frac{1}{2}y^2$ and the rotation $\mathbf{R}(x, y) = [y, -x]^\top$, with the noise intensity $\sigma^2 = 2, 10^{-3}$, and 0.

Using the Moment-DeepRitz method, we recover both the drift vector field and the associated pseudo-potential, which together yield the corresponding rotation field. The relative root mean square errors are reported in Table 1a, and the first three rows of Figure 2 compare the learned results with the ground truth.

Table 1: Relative root mean square errors across different examples.

(a) Learning of the double-well potential under different noise intensities.

σ^2	$\text{rRMSE}_{\mathbf{b}}$	rRMSE_{ψ}	$\text{rRMSE}_{\mathbf{R}}$
2	2.83e-02	7.27e-02	9.03e-02
10^{-3}	2.58e-02	8.07e-02	9.10e-02
0	2.50e-02	1.09e-01	1.09e-01

(b) Learning of different potentials and rotations.

Example	$\text{rRMSE}_{\mathbf{b}}$	rRMSE_{ψ}	$\text{rRMSE}_{\mathbf{R}}$
QW ψ	1.37e-02	2.41e-01	7.07e-02
Osc. \mathbf{R}	6.44e-03	1.84e-01	2.17e-01
Rough ψ	2.64e-02	1.17e-01	1.33e-01

Quadruple-well (QW) Potential. Consider the potential $\psi(x, y) = \frac{1}{8}(x^2 - 1)^2 + \frac{1}{8}(y^2 - 1)^2$ and the rotation $\mathbf{R}(x, y) = [y, -x]^\top$, with the noise intensity $\sigma^2 = 2$. We choose a larger number of datasets $M = 80$ and increase the number of epochs to 100,000.

Oscillatory Rotation (Osc. \mathbf{R}). Consider the potential $\psi(x, y) = \frac{1}{2}(x^2 + y^2)$ and the rotation $\mathbf{R}(x, y) = [\cos(y), -\sin(x)]^\top$, with the noise intensity $\sigma^2 = 2$. We choose a larger number of datasets $M = 80$ and increase the number of epochs to 50,000.

For the two examples above, the relative root mean square errors are reported in the first two rows in Table 1b, and the first two rows in Figure 1 compare the learned results with the ground truth.

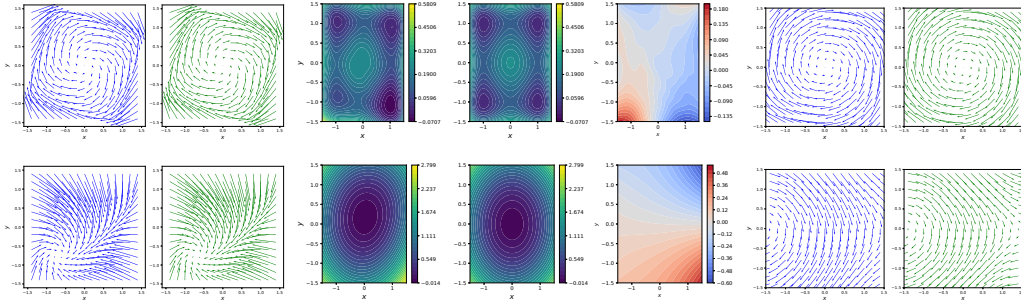


Figure 1: From left to right: $\mathbf{b}_{\text{NN}}/\mathbf{b}$ (quiver plots), $\psi_{\text{NN}}/\psi/\psi_{\text{NN}} - \psi$ (heatmaps), and $\mathbf{R}_{\text{NN}}/\mathbf{R}$ (quiver plots). **Rows:** (1) Quadruple-well potential, (2) Oscillatory rotation.

Rough Potential. To demonstrate the robustness of our method to rough, oscillatory potentials, we consider $\psi(\mathbf{x}) = \frac{1}{4}(x^2 - 1)^2 + \frac{1}{2}y^2 + \epsilon^4 \sin(\frac{2\pi x}{\epsilon}) \sin(\frac{2\pi y}{\epsilon})$ with a parameter $\epsilon = 0.2$. Here ϵ controls the oscillations: smaller values produce stronger, higher-frequency fluctuations, which grow more pronounced in higher-order derivatives and thus are harder to approximate accurately. The noise intensity is set to be $\sigma^2 = 2$.

The relative root mean square errors are reported in the last row of Table 1b. The last row of Figure 2 compare the learned results with the ground truth.

The experimental results summarized in Table 1 and shown in Figures 1 and 2 demonstrate the robustness of the Moment-DeepRitz method under several challenging scenarios. First, the method remains stable in the vanishing noise limit ($\sigma \rightarrow 0$), where the pseudo-potential converges to the quasi-potential. Second, it performs reliably on noisy datasets, tolerating perturbations up to a noise

level of 0.1. Finally, it is effective for complex potentials ψ and rotation fields \mathbf{R} even without imposing the pointwise orthogonality condition between $\nabla\psi$ and \mathbf{R} .

4 Conclusions

A data-driven approach has been proposed for learning the governing physical laws in complex stochastic dynamics described by the Fokker-Planck equations with non-gradient drifts. Combining the first-moment dynamics with a variational formulation, we have developed the Moment-DeepRitz method to learn both the pseudo-potential and rotational components in non-equilibrium diffusions. Compared to the results shown in [25], our framework is applicable to the case of non-pointwise orthogonal drift decomposition. Another advantage of our method lies in its robustness to noisy data and varying noise intensity shown in Section 3. In particular, we validate the effectiveness through a series of challenging numerical experiments involving quadruple-well potentials, rough potentials and oscillatory rotations. We remark that the numerical results for the cases $\sigma^2 = 10^{-3}$ and $\sigma = 0$ in Table 1a demonstrate that our framework offers an alternative approach of computing the quasi-potential, parallel to the data-driven method used in [27]. Numerous open problems remain and a promising research direction in the near future is to extend the data-driven learning framework to the case of time-dependent drifts.

Acknowledgments and Disclosure of Funding

F. Kong thanks Professor Hong Qian for critical discussions. Y. Lu is partially supported by DOE DE-SC0222766. All the code used for making data and figures in this paper is available at <https://github.com/lai-chen-chih/Moment-DeepRitz>.

References

- [1] Evans, L. (2012) *An introduction to stochastic differential equations*. Vol. 82, American Mathematical Society.
- [2] Särkkä, S. and Solin, A. (2019) *Applied stochastic differential equations*. Vol. 10, Cambridge University Press.
- [3] Ken, S. (1998) Langevin equation and thermodynamics. *Prog. Theor. Phys. Suppl.* **130**:17–27.
- [4] Hatano, T. and Sasa, S.-I. (2001) Steady-state thermodynamics of Langevin systems. *Phys. Rev. Lett.* **86**(16):3463.
- [5] Quigley, D. and Probert, M. (2004) Langevin dynamics in constant pressure extended systems. *J. Chem. Phys.* **120**(24):11432–11441.
- [6] Esposito, M. (2012) Stochastic thermodynamics under coarse graining. *Phys. Rev. E* **84**(4):041125.
- [7] Crooks, G. (1999) Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E* **60**(3):2721.
- [8] Qian, M.-P., Qian, M. and Gong, G. (1991) The reversibility and the entropy production of Markov processes. *Contemp. Math* **118**:255–261.
- [9] Kurchan, J. (1998) Fluctuation theorem for stochastic dynamics. *J. Phys. A: Math. Gen.* **31**(16):3719.
- [10] Marconi, U., Puglisi, A., Rondoni, L. and Vulpiani, A. (2008) Fluctuation–dissipation: response theory in statistical physics. *Phys. Rep.* **461**(4–6):111–195.
- [11] Esposito, M. and Van den Broeck, C. (2010) Three detailed fluctuation theorems. *Phys. Rev. Lett.* **104**(9):090601.
- [12] Dorfman, J. (1999) *An introduction to chaos in nonequilibrium statistical mechanics*. No. 14, Cambridge university press.
- [13] Jiang, D.-Q. and Jiang, D.-H. (2004) *Mathematical theory of nonequilibrium steady states: on the frontier of probability and dynamical systems*. Springer Science & Business Media.
- [14] Brunton, S., Proctor, J. and Kutz, J. (2016) Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl. Acad. Sci.* **113**(15):3932–3937.
- [15] Raissi, M., Perdikaris, P. and Karniadakis, G. Em. (2019) Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **378**:686–707.

- [16] Messenger, D.A. and Bortz, D.M. (2021) Weak SINDy for partial differential equations. *J. Comput. Phys.* **443**:110525.
- [17] Huang, S., He, Z. and Reina, C. (2022) Variational Onsager Neural Networks (VONNs): A thermodynamics-based variational learning strategy for non-equilibrium PDEs. *J. Mech. Phys. Solids* **163**:104856.
- [18] Kharazmi, E., Zhang, Z. and Karniadakis, G. Em. (2021) hp-VPINNs: Variational physics-informed neural networks with domain decomposition. *Comput. Methods Appl. Mech. Eng.* **374**:113547.
- [19] Lu, Y., Li, X., Liu, C., Tang, Q. and Wang Y. (2024) Learning Generalized Diffusions using an Energetic Variational Approach. *arXiv preprint arXiv:2412.04480*.
- [20] Chen, X., Yang, L., Duan, J., Karniadakis, G. Em. (2021) Solving Inverse Stochastic Problems from Discrete Particle Observations Using the Fokker–Planck Equation and Physics-Informed Neural Networks. *SIAM J. Sci. Comput.* **43**(3):B811–B830.
- [21] Gao, Y., Lang, Q. and Lu, F. (2024) Self-test loss functions for learning weak-form operators and gradient flows. *arXiv preprint arXiv:2412.03506*.
- [22] Zang, Y., Bao, G., Ye, X. and Zhou, H. (2020) Weak adversarial networks for high-dimensional partial differential equations. *J. Comput. Phys.* **411**:109409.
- [23] Lee, K., Trask, N. and Stinis, P. (2021) Machine learning structure preserving brackets for forecasting irreversible processes. *Advances in Neural Information Processing Systems*.
- [24] Huang, S., He, Z., Dirr N., Zimmer, J. and Reina, C. (2024) Statistical-Physics-Informed Neural Networks (Stat-PINNs): A machine learning strategy for coarse-graining dissipative dynamics. *J. Mech. Phys. Solids*:105908.
- [25] Kong, F., Lai, C.-C. and Lu, Y. (2025) Moment Estimate and Variational Approach for Learning Generalized Diffusion with Non-gradient Structures. *arXiv preprint arXiv:2508.01854*.
- [26] Qian, M. and Wang, Z. (1999) The entropy production of diffusion processes on manifolds and its circulation decompositions. *Commun. Math. Phys.* **206**:429–445.
- [27] Lin, B., Li, Q. and Ren W. (2022) A data driven method for computing quasipotentials. *Math. Sci. Mach. Learn.*:652–670.
- [28] E, W. and Yu, B. (2018) The deep Ritz method: a deep learning-based numerical algorithm for solving variational problems. *Commun. Math. Stat.* **6**(1):1–12.
- [29] Kingma, D. P. and Ba, J. (2015) Adam: A Method for Stochastic Optimization. *3rd International Conference on Learning Representations*.

A Algorithm for the Moment-DeepRitz Method

Here we present the algorithm for the Moment-DeepRitz method described in Section 2.

Algorithm 1 Learning generalized diffusions with non-gradient drifts via Moment-DeepRitz method

- Choose the training data as $\{(f_j(\mathbf{x}_i, t_1), f_j(\mathbf{x}_i, t_2))\}_{i,j=1}^{N,M}$, where $f_j(\mathbf{x}_i, t)$ are probability density functions and t_1, t_2 are two time steps.
 - Phase 1: Minimize the loss function (8) via the deep learning based method, then find the “best” parameters of the neural networks to reconstruct \mathbf{b}_{NN} for learning the drift \mathbf{b} .
 - Phase 2: Minimize the loss function I_ψ^{dyn} given by (9) and find the “best” parameters of the neural networks to reconstruct ψ_{NN} for learning the irrotational component ψ .
-

B Results of learning double-well potential under different noise intensities

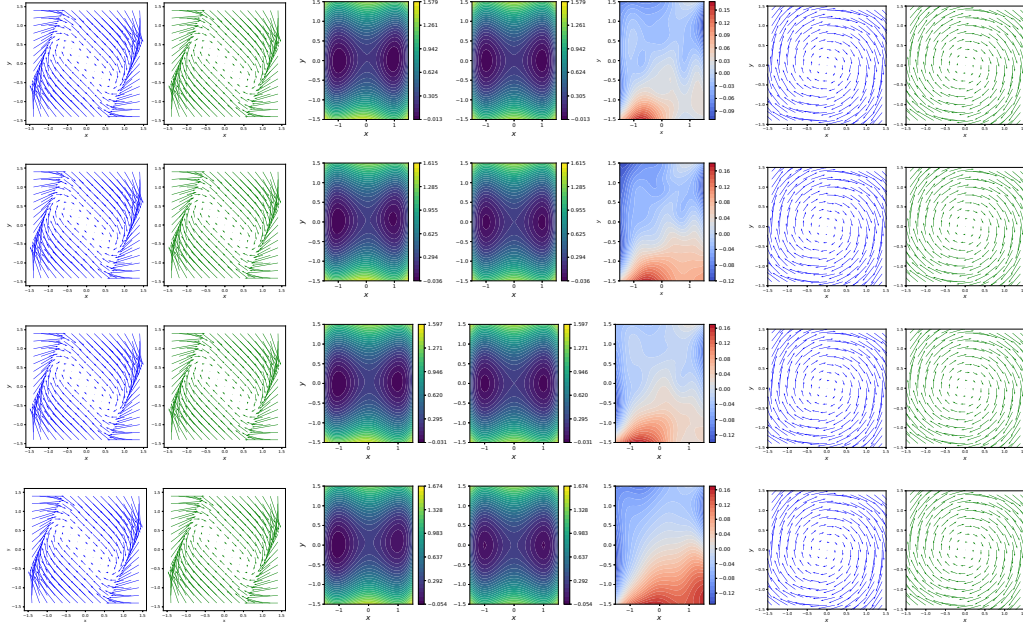


Figure 2: From left to right: $\mathbf{b}_{\text{NN}}/\mathbf{b}$ (quiver plots), $\psi_{\text{NN}}/\psi/\psi_{\text{NN}} - \psi$ (heatmaps), and $\mathbf{R}_{\text{NN}}/\mathbf{R}$ (quiver plots). **Rows:** (1) $\sigma^2 = 2$, (2) $\sigma^2 = 0.001$, (3) $\sigma^2 = 0$, (4) $\sigma^2 = 2$ with the rough potential.