Efficient Goal-oriented Sequential Bayesian Experimental Design with Applications to Causal and Physical Systems

by

Jiayuan Dong

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Mechanical Engineering and Scientific Computing)
in The University of Michigan
2026

Doctoral Committee:

Associate Professor Xun Huan, Chair
Dr. Tommie Catanach, Sandia National Laboratories
Associate Professor Yang Chen
Associate Professor Daniel Cooper

Jiayuan Dong
jiayuand@umich.edu
ORCID iD: 0009-0004-9111-5555

© Jiayuan Dong 2026

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my advisor, Prof. Xun Huan, for his unwavering support and guidance throughout my academic journey. His expertise and insightful mentorship were instrumental in shaping my research direction, and his constructive feedback greatly refined my work. I am especially grateful for his support during my master's studies in Statistics, where he provided early research opportunities and guided my transition from a statistics background to a Ph.D. program in Mechanical Engineering. Working under his mentorship has been both intellectually rewarding and personally inspiring.

I am also deeply grateful to my thesis committee members, Prof. Cooper, Prof. Chen, and Dr. Catanach, for their time, guidance, and thoughtful feedback. Prof. Cooper's mentorship during my early Bayesian research on material flow analysis helped establish a strong research foundation. Prof. Chen's course in Bayesian statistics sparked my interest in Bayesian methods and influenced my research trajectory. I am particularly thankful to Dr. Catanach for proposing the symmetrized expected information gain formulation, which became a central component of this thesis. I sincerely appreciate the committee's insightful suggestions, which significantly strengthened the quality and clarity of this work.

I would also like to thank my labmates in the UQ-SciML group for their support and collaboration. I am especially grateful to Wanggang Shen for his extensive help with research discussions and code development, which led to a joint paper. I thank Aniket Jivani for his support during conferences, including sharing accommodations and travel moments that made those experiences both enjoyable and memorable, and Thomas Coons for introducing me to power-lifting and offering encouragement beyond academic work. I also appreciate all group members for their valuable discussions and collective support throughout my doctoral studies.

I am thankful to my friends in Michigan and around the world for their companionship and support. I am especially grateful to my roommate, Zheyu Zhang, for his constant help in daily life. I thank Chengyang Huang, Kaiqi Zhu, Jiayi Qiu, Yuting Cao, Haoran Shi, Wanggang Shen, and Sirui Chen for engaging conversations and shared activities such as rock climbing. I also appreciate Haowei Sun, Xinhai Hou, Chenhui Zhao, and Jingzhuo Fang for motivating gym sessions, and Elise Xu, Liang Xu, and Cecilia Xu for memorable skiing trips. I am grateful to Haotian Bai, Yi Guo, Jiazhuo Liang, Yuying Liao, and Junyu Tan for providing an important social buffer during my Ph.D. years, to Yaxuan Hou for her encouragement during my internship applications, and to Tianyi

Xu, Yiyang Lu, and Chaoyu Ding for their care and support. I also thank Zehui Li and Zixuan Shui for their continued remote encouragement, thoughtful conversations, and emotional support, which brought comfort and motivation throughout this journey. I am equally grateful to the many friends not named individually here whose support, kindness, and shared moments have been meaningful throughout this journey.

Finally, I would like to express my deepest gratitude to my family and my beloved dog, Gangan, for their unconditional love and support. My family's encouragement and understanding gave me strength during challenging moments, and Gangan's companionship brought comfort, balance, and joy to my daily life. Their support made this journey not only possible, but meaningful.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ii

LIST OF FIGURES vi

LIST OF TABLES xi

LIST OF APPENDICES xiii

ABSTRACT xiv

CHAPTER

1 Introduction 1

1.1 Background and motivation 1
1.1.1 Bayesian optimal experimental design 1
1.1.2 Sequential Bayesian experimental design 2
1.1.3 Goal-oriented Bayesian experimental design 3
1.1.4 Causal Bayesian experimental design 4
1.1.5 Expected information gain for experimental design 5
1.2 Objectives and outline 6

2 Variational Bayesian Optimal Experimental Design with Normalizing Flows 9

2.1 Problem formulation 10
2.1.1 Bayesian optimal experimental design 10
2.1.2 Variational optimal experimental design 10
2.2 Computational methods 13
2.2.1 Normalizing flows 14
2.3 Numerical experiments 16
2.3.1 Case 1: Low-dimensional nonlinear design 17
2.3.2 Case 2: High-dimensional linear design 19
2.3.3 Case 3: Cathodic electrophoretic deposition (e-coating) 21
2.3.4 Case 4: aphid population 26
2.4 Conclusions 28

3 Variational Sequential Bayesian Optimal Experimental Design 30

3.1 Problem formulation 31
3.1.1 Preliminaries 31

3.1.2 Sequential optimal experimental design (sOED) . . . 32
3.1.3 Experimental design rewards . . . 34
3.1.4 One-point formulation of rewards . . . 35
3.1.5 Variational sequential optimal experimental design (vsOED) . . . 37
3.2 Numerical methods for vsOED . . . 39
3.2.1 Variational gradient . . . 40
3.2.2 Policy gradient . . . 41
3.2.3 Implementation techniques and the overall algorithm . . . 42
3.3 Numerical experiments . . . 44
3.3.1 Demonstration setup . . . 44
3.3.2 Case 1: source location finding . . . 46
3.3.3 Case 2: constant elasticity of substitution . . . 54
3.3.4 Case 3: SIR model for disease spread . . . 55
3.3.5 Case 4: convection-diffusion-reaction . . . 59
3.4 Conclusions . . . 61

4 Goal-Oriented Sequential Bayesian Experimental Design for Causal Learning . . . 65

4.1 Introduction . . . 66
4.2 Preliminaries . . . 67
4.3 Goal-Oriented sequential causal Bayesian experimental design . . . 69
4.3.1 Variational lower bound . . . 69
4.3.2 Variational posteriors and policy network . . . 70
4.4 Numerical experiments . . . 72
4.4.1 Motivating example with fixed graph structure . . . 72
4.4.2 Causal reasoning tasks . . . 73
4.4.3 Causal discovery tasks . . . 75
4.5 Conclusions . . . 76

5 Symmetrized Expected Information Gain for Bayesian Optimal Experimental Design 80

5.1 Problem formulation . . . 81
5.1.1 Symmetric Expected Information Gain . . . 81
5.1.2 Valid measure of information . . . 82
5.2 Numerical experiments . . . 84
5.2.1 Case 1: Linear-Gaussian examples . . . 84
5.2.2 Case 2: Convection-diffusion with building obstacles . . . 86
5.3 Conclusions . . . 88

6 Conclusions and Future Work . . . 90

6.1 Conclusions . . . 90
6.2 Limitations and future work . . . 93

APPENDICES . . . 95

BIBLIOGRAPHY . . . 164

LIST OF FIGURES

FIGURE

2.1 Diagram for the cINN NFs transformation. 16
2.2 Case 1. EIG estimates and sample training convergence plot. 18
2.3 Case 1. Comparison of marginal posteriors obtained from SMC and vOED-NFs at $\xi = 0.2$. 19
2.4 Case 1. Comparison of marginal posteriors obtained from SMC and vOED-NFs at $\xi = 1.0$. 20
2.5 Case 1. EIG under different number of transformations T and optimization sample sizes N_{opt}. 21
2.6 Case 2. Comparison of marginal posteriors obtained from NUTS-HMC and vOED-GG [50] at design where $\xi_j = \mathbf{e}_j$, with KLD denotes the empirical KL divergence from NFs posteriors to MCMC posteriors. 22
2.7 Case 2. Comparison of marginal posteriors obtained from NUTS-HMC and vOED-NFs [50] at design where $\xi_j = \mathbf{e}_j$, with KLD denotes the empirical KL divergence from NFs posteriors to MCMC posteriors. 23
2.8 Case 3. Convergence history of the lower and upper bounds at $d = j_0 = 5$ mA and 10 observations when optimizing ϕ using different N_{opt}. Hi-NMC is the high-quality reference EIG estimate. Here the upper bound estimates appear to be tighter than the lower bound estimates. 24
2.9 Case 3. Marginal distributions $p(y_1|\xi), p(y_2|\xi), p(y_3|\xi)$ at $\xi = j_0 = 5$ mA obtained using MC sampling and vOED-NFs with $N_{\text{opt}} = 10000$. 24
2.10 Case 3. Convergence history of the lower and upper bounds at $\xi = j_0 = 5$ mA and 50 observations when optimizing ϕ using different N_{opt}. Hi-NMC is the high-quality reference EIG estimate. Here the lower bound estimates appear to be tighter than the upper bound estimates. 25
2.11 Case 3. Lower bound estimates for all designs using different methods. Hi-NMC is the high-quality reference EIG estimate. 25
2.12 Case 3. Marginal and pairwise joint posterior distributions obtained using SMC (high-quality reference) and vOED-NFs-LSTM. 26
2.13 Case 4. Comparison of joint posteriors samples from ABC and vOED-NFs at the optimal design. 28

3.1 Case 1a. Expected utility comparisons using policies resulting from different algorithms. The shaded regions represent the standard error. 48

3.2 Case 1a. Examples of GMM and NFs approximate posterior and true posterior for PoIs. The red stars indicate the true data-generating parameter values for these trajectory samples. 48
3.3 Case 1a. Average $\tilde{U}$ over four training replicates for 'OED for QoIs'. The shaded regions represent the standard error. 49
3.4 Case 1a. Examples of policy trajectory for $T = 15$. The contour background plots the true signal strength, and the red stars indicate the true source locations. 49
3.5 Case 1a. Design and posterior comparisons for 'OED for QoIs'. 50
3.6 Case 1b. Average expected utility or $\tilde{U}$ over two training replicates versus design horizon T using policies resulting from the five OED scenarios. 51
3.7 Case 1b. Examples of policy trajectory for $T = 30$. The contour background plots the true signal strength, and the red stars indicate the true source locations. 52
3.8 Case 1b. Examples of approximate posterior and true posterior for model indicator using policy resulting from 'OED for model indicator', for $T = 30$. 53
3.9 Case 2. Expected utility comparisons using policies resulting from different algorithms. The shaded regions represent the standard error. 55
3.10 Case 2. Examples of GMM approximate posterior and true posterior at $T = 10$. The red stars indicate the true data-generating parameter values for these trajectory samples. 56
3.11 Case 3. Average $\tilde{U}$ over four training replicates versus design horizon T. The shaded regions represent the standard error. 58
3.12 Case 3. Examples of infected state trajectory $I(k)$ and corresponding designs (observation times) ξ_t for $T = 10$ with three realizations of (β, ρ) and different ratios $r = \beta/\rho$. 58
3.13 Case 4. Examples of the true and surrogate concentration fields G at $k = 0.2$. 60
3.14 Case 4. Examples of approximate posterior and true posterior for model indicator and PoIs using policy resulting from 'OED for model indicator' and 'OED for PoIs', respectively, for $T = 10$. 62
3.15 Case 4. Examples of policy trajectory for $T = 10$. The background contour plots the true contaminant concentration at the end time $t = 0.2$, and the red stars indicate the true source locations. 63

4.1 Illustration of goal-oriented versus conventional BOED for causal learning. *Left*: A linear Gaussian SCM with six nodes; the experimental goal is to estimate the causal effect of the intervention $\mathrm{do}(Y_3 = 10)$ on node Y_5 and Y_6. *Middle*: Conventional BOED selects interventions that maximize EIG over all model parameters, resulting in Y_1 and Y_2 being selected as the best. *Right*: Our GO-CBED approach selects interventions by directly maximizing EIG for the specific causal query, leading to a different intervention that prioritizes nodes most relevant to the query, i.e., Y_3 and Y_2. . 67
4.2 Performance comparison of policies trained for $T = 10$, evaluated across different stage lengths. *Left*: Performance on causal query $z = \{Y_5, Y_6 \mid \mathrm{do}(Y_3 = 10)\}$. *Right*: Performance on model parameters $z = \{\theta \setminus \theta_{\mathrm{pa}(3)}\}$. While GO-CBED-$\theta$ achieves higher EIG on the task centering parameters (right), it performs significantly worse than GO-CBED-z on the causal reasoning task (left). Shaded regions represent ± 1 standard error across 4 random seeds. 72

4.3 Performance comparison of policies trained for $T = 10$ on causal queries, using ER and SF graph priors with linear and nonlinear mechanisms (3 interventions per stage, $d = 10$). While GO-CBED-G—which targets structure learning—is a natural baseline, it underperforms on causal queries, particularly in nonlinear settings. In contrast, GO-CBED-z, which directly targets causal QoI, consistently achieves higher EIG, especially under nonlinear mechanisms. Shaded regions represent ±1 standard error over 4 random seeds. 74
4.4 Performance comparison of policies trained for $T = 10$ on *E. coli* gene regulatory networks with nonlinear causal mechanisms ($d = 10$). GO-CBED-z performs comparably to baselines in early stages but exhibits rapid improvement after stage 3, ultimately achieving substantially higher EIG than both GO-CBED-G and Random. These results highlight the value of goal-oriented experimental design in realistic biological settings with complex nonlinear causal mechanisms. Shaded regions represent ±1 standard error across 4 random seeds. 74
4.5 Performance comparison on synthetic SCMs, using ER and SF graph priors with linear and nonlinear mechanisms. Metrics include prior-omitted EIG lower bound $\mathcal{R}_{T;L}$, expected structural Hamming distance $\mathbb{E}$-SHD, and F_1-score. GO-CBED performs better or comparatively in terms of uncertainty reduction ($\mathcal{R}_{T;L}$), structural recovery ($\mathbb{E}$-SHD), and structural accuracy (F_1-score) compared to all baselines. Shaded regions indicate ±1 standard error across 10 random seeds. 78
4.6 GO-CBED outperforms all baselines on semi-synthetic *E. coli* gene regulatory networks ($d = 20$, $T = 10$) with linear mechanisms. Our method achieves: (*Left*) near-zero $\mathcal{R}_{T;L}(\pi_z; \phi)$, (*Middle*) significant lower $\mathbb{E}$-SHD, and (*Right*) superior F_1 score for edge prediction. This demonstrates GO-CBED's ability to efficiently identify the true causal structure in biologically-inspired networks, with the variational posterior tightly concentrated around the ground truth after 10 stages. 79

5.1 Comparison of analytic and Monte Carlo estimators for EIG (left) and sEIG (right) in the linear-Gaussian model with $d = 10$ and $m = 3$. Shaded regions show ±1 standard deviation over 30 independent runs. NMC estimates of EIG suffer from high bias for small inner sample sizes M, whereas the single-loop Monte Carlo estimator of sEIG is unbiased. 85
5.2 Relative variability (SD/mean) of EIG versus sEIG estimators as a function of design strength $\|\xi\|$, calculated over 30 runs. The NMC estimator for EIG ($N = 3000$, $M = 200$ for a total of 600,000 forward runs) exhibits higher relative variability despite using a larger budget, which increases with $\|\xi\|$, while the sEIG estimator ($N = 3000$, $M = 200$ for a total of 13,000 forward runs) remains substantially more stable despite using a much smaller computational budget. 86
5.3 Comparison of EIG landscapes under identical building configurations obtained with different nested Monte Carlo budgets. The top row uses a smaller Monte Carlo budget and exhibits noticeable stochastic variability, while the bottom row uses a substantially larger budget, resulting in smoother utility surfaces and reduced estimator variance. . 87
5.4 Estimated sEIG with $N = M = 10^4$. 88

A.1 Case 1. EIG under different number of transformations T and optimization sample sizes N_{opt}, evaluated using the same samples that it used for performing optimization. . 97
A.2 Case 3. Marginal and pairwise joint posterior distributions obtained using SMC (high-quality reference) and vOED-NFs-LSTM with $N_{\text{opt}} = 80000$. 103

B.1 Case 1a. Training history of average $\tilde{U}$ over four training replicates for $T = 30$. The shaded regions represent standard error. 128
B.2 Case 1b. Training history of average $\tilde{U}$ over four training replicates for 'OED for PoIs' and $T = 30$. The shaded regions represent standard error. 128
B.3 Case 2. Training history of average $\tilde{U}$ over four training replicates for $T = 10$. The shaded regions represent the standard error. 129
B.4 Case 3. Training history of average $\tilde{U}$ over four training replicates for $T = 10$. The shaded regions represent the standard error. 130
B.5 Case 4. Training history of average $\tilde{U}$ over four training replicates for $T = 10$. The shaded regions represent the standard error. 130

C.1 Policy network architecture. The model takes as input a three-dimensional tensor of shape $n \times d \times 2$, where $n = n_{\text{int}} \times T$. It is permutation-invariant along the n-axis and permutation-equivariant along the d-axis. Each of the L layers first applies self-attention across the d-axis, followed by attention across the n-axis, with shared parameters across the non-attended axis. 147
C.2 Comparison between the true posterior predictive distribution $p(z|I_T)$ and the variational approximation $q(z|I_T; \phi)$) for two simulated trajectories. The approximate posterior closely aligns with the true posterior. 149
C.3 Evaluation of policies on ER and SF graphs with both linear and nonlinear causal mechanisms. The $\pi^*_{\boldsymbol{G}}$ demonstrates strong performance in accurately identifying the underlying causal graph across all settings. 151
C.4 Evaluation of policies on *E. coli* graphs with nonlinear causal mechanisms. The $\pi^*_{\boldsymbol{G}}$ demonstrates strong performance in accurately identifying the underlying causal graph. 152
C.5 Evaluation of interventions on node 1 using integers from −5 to 5, with the causal query defined as $z = \{Y_3, Y_4 \mid \text{do}(Y_2 = 2)\}$. 153
C.6 Prior-omitted EIG lower bound estimates, with parenthesis values denoting the inner loop sample size for NMC and training sample size for GO-CBED. Shaded regions represent ±1 standard error across 4 random seeds. 154
C.7 Additional causal reasoning experiments on nonlinear gene regulatory networks. **(a)** *E. coli* network in Section 4.4.2: intervention on node 7 targeting nodes 6 and 8. **(b)** Yeast network: intervention on node 8 targeting nodes 0 and 9. **(c)** *E. coli* network: intervention on node 5 targeting nodes 7 and 9. GO-CBED consistently outperforms baseline methods, with performance gains varying based on structural complexity of the intervention-target relationships. Shaded regions represent ±1 standard error over 4 random seeds. 155
C.8 Evaluation of GO-CBED under a distributional shift in observation noise at deployment for linear causal reasoning tasks. GO-CBED consistently outperforms the random baseline that is using posterior networks pre-trained under the shifted noise, demonstrating the robustness of its jointly optimized policy and posterior networks. . . 156

C.9 Performance comparison on synthetic SCMs, with each method using its originally proposed posterior inference approach. GO-CBED consistently outperforms all baselines across both linear and nonlinear settings, demonstrating the advantage of jointly optimizing policy and posterior networks. Shaded regions represent ±1 standard error across 10 random seeds. 157
C.10 Performance comparison on semi-synthetic gene regulatory (*E. coli* and Yeast) networks, with each method using its originally proposed posterior inference approach. GO-CBED demonstrates strong performance on causal tasks in biologically inspired settings. Shaded regions represent ±1 standard error across 10 random seeds. 158
C.11 Causal discovery performance on nonlinear *E. coli* gene regulatory networks. GO-CBED performs better or comparatively in terms of uncertainty reduction ($\mathcal{R}_{T;L}$), structural recovery ($\mathbb{E}$-SHD), and structural accuracy (F_1-score) compared to all baselines. Shaded regions indicate ±1 standard error across 10 random seeds. 158
C.12 Causal discovery performance on linear Yeast gene regulatory networks. GO-CBED performs better or comparatively in terms of uncertainty reduction ($\mathcal{R}_{T;L}$), structural recovery ($\mathbb{E}$-SHD), and structural accuracy (F_1-score) compared to all baselines. Shaded regions represent ±1 standard error across 10 random seeds. 159
C.13 Causal discovery performance on nonlinear Yeast gene regulatory networks. GO-CBED performs better or comparatively in terms of uncertainty reduction ($\mathcal{R}_{T;L}$), structural recovery ($\mathbb{E}$-SHD), and structural accuracy (F_1-score) compared to all baselines. Shaded regions represent ±1 standard error across 10 random seeds. 159
C.14 Evaluation of GO-CBED under a distributional shift in observation noise at deployment for nonlinear causal reasoning tasks. GO-CBED consistently outperforms the random baseline that is using posterior networks pre-trained under the shifted noise, demonstrating the robustness of its jointly optimized policy and posterior networks. . . 160

LIST OF TABLES

TABLE

2.1 Case 2. Middle column shows the mean of 10 lower bound estimates $\widehat{U}_L$ (each is evaluated with $N = 10000$ samples) at the optimal design found by vOED-GG and vOED-NFs, and the right column shows the mean of 10 high-quality NMC estimates (each with $N_{\text{out}} = 10000$ and $N_{\text{in}} = 10000$) at the optimal designs. The $(\cdot)$ values are one standard deviation estimates for the estimators. 21

2.2 Case 4. The third column shows the optimal design ξ^* found by LB-KLD and vOED-NFs. The fourth column shows the mean of 10 $\widehat{U}_{\text{LB-KLD}}$ estimates with different random seeds (each with 3×10^4 samples for Monte Carlo estimation) at the optimal designs, and $(\cdot)$ indicates the standard deviation originates from different random partitions of the same 30000 samples used for entropy estimation. The last column shows the mean of 10 $\widehat{U}_L$ estimates (each with $N_{\text{opt}} = 2 \times 10^4$ and $N = 10^4$) at the optimal designs but different ϕ^*. Here $(\cdot)$ indicates the standard deviation due to 10 random seeds for different initialization when optimizing ϕ^*, and evaluation is conducted on the same $N = 10^4$ samples. 27

3.1 Properties of different sequential experimental design methods. 45

3.2 Case 1b. PCE estimate of the EIG on model indicator, PoIs, and QoIs, using policies resulting from the five OED scenarios. ± represents standard error. 53

3.3 Case 4. Testing MSE of the surrogate models. 60

A.1 Case 1. Hyperparameters for vOED-G. 96

A.2 Case 1. Hyperparameters for vOED-NFs. 97

A.3 Case 2. Hyperparameters for vOED-GG. 98

A.4 Case 2. Hyperparameters for vOED-NFs. 98

A.5 Case 3. Hyperparameters for the 10-dimensional case, both lower and upper bound estimators. 101

A.6 Case 3. Hyperparameters for vOED-G for the 50-dimensional case, for the lower bound estimator. 101

A.7 Case 3. Hyperparameters for vOED-NFs for the 50-dimensional case, for both lower and upper bound estimators without LSTM summary network. 102

A.8 Case 3. Hyperparameters for vOED-NFs for the 50-dimensional case, for lower bound estimator with LSTM summary network. 103

A.9 Case 4. Hyperparameters for vOED-NFs. 104

B.1 Summary of general symbols and notation used in the paper. 132

B.2 Architecture for the NN-based approximate posteriors of model indicator. 133

B.3 Architecture for the feature net within the GMM net. . . . 133
B.4 Architecture for the weight net within the GMM net. . . . 133
B.5 Architecture for the mean net and standard deviation net within the GMM net. . . . 133
B.6 Architecture for the feature net within the NFs. The first value under the 'Dimension' column is used for the source location problem and the CES problem; the second value is used for the SIR problem. . . . 134
B.7 Architecture for the $\mathtt{s}_1$ and $\mathtt{t}_1$ nets within the NFs. The first value under the 'Dimension' column is used for the source location problem and the CES problem; the second value is used for the SIR problem. . . . 134
B.8 Architecture for the $\mathtt{s}_2$ and $\mathtt{t}_2$ nets within the NFs. The first value under the 'Dimension' column is used for the source location problem and the CES problem; the second value is used for the SIR problem. . . . 134
B.9 Architecture for the actor network. . . . 135
B.10 Architecture for the critic network. . . . 135
B.11 Case 1a. Hyperparameter settings. In the table, 'lr' stands for 'learning rate'. . . . 135
B.12 Case 1a. Case 1a. PCE estimate of the EIG for $T = 30$ using vsOED-G-T with varying number of critic update per iteration. ± represents standard error. . . . 136
B.13 Case 1b. Hyperparameter settings. In the table, 'lr' stands for 'learning rate'. . . . 136
B.14 Case 1a. Expected utility and standard error (±) for four training replicates for 'OED for PoIs', evaluated using PCE, for $T = 30$. . . . 136
B.15 Case 1b. Expected utility and standard error (±) for four training replicates for 'OED for PoIs', evaluated using PCE, for $T = 30$. . . . 137
B.16 Case 2. Hyperparameter settings. In the table, 'lr' stands for 'learning rate'. . . . 137
B.17 Case 2. Expected utility and standard error (±) for four training replicates, evaluated using PCE, for $T = 10$. . . . 137
B.18 Case 3. Hyperparameter settings. In the table, 'lr' stands for 'learning rate'. . . . 138
B.19 Case 3. Expected utility and standard error (±) for four training replicates, evaluated using PCE, for $T = 10$. . . . 138
B.20 Case 4. Hyperparameter settings. In the table, 'lr' stands for 'learning rate'. . . . 139
B.21 Case 4. Prior-omitted $\tilde{U}$ and standard error (±) for four training replicates and $T = 10$. 139
B.22 Case 4. Architecture for the surrogate model of G. . . . 139
B.23 Case 4. Architecture for the surrogate model of φ. . . . 140
B.24 Summary of case-specific symbols and notation used in the numerical experiments. . . . 140

C.1 Hyperparameter settings for the policy network. . . . 147
C.2 Hyperparameter settings for the posterior network in the causal discovery case. . . . 148
C.3 Hyperparameter settings for the posterior network in the causal reasoning case. . . . 149

LIST OF APPENDICES

APPENDIX

A Appendix for Variational BOED with Normalizing Flows 95

B Appendix for vsOED 106

C Appendix for GO-CBED 141

D Appendix for Symmetrized EIG 161

ABSTRACT

Efficient experimentation is central to uncertainty reduction and decision making when data are scarce. Bayesian optimal experimental design (BOED) provides a principled framework for selecting experiments that maximize the expected information gain (EIG), but its practical adoption is often hindered by intractable integrals and the high cost of nested Monte Carlo estimators. This thesis develops efficient, likelihood-free, and scalable computational methods for BOED.

Our approach is motivated by three key bottlenecks. First, estimating and optimizing EIG is prohibitive when likelihoods are expensive or unavailable, especially for partial differential equation-based and simulator-based models. Second, in sequential experimental design, most existing approaches are greedy or myopic, failing to account for future experimental synergy, or rely on high-bias EIG estimators that lead to inefficient policies. Third, in many applications, such as causal learning, the target is a specific Quantity of Interest (QoI), and full parameter uncertainty quantification can be inefficient; experimental designs should therefore focus information on the target QoI.

We develop BOED methods that (i) introduce a variational formulation that defines a tractable lower bound on EIG, (ii) extend this framework to sequential and goal oriented settings, with particular emphasis on causal learning, and (iii) investigate a symmetrized information gain criterion as an alternative to conventional EIG.

In Chapter 2, we advance variational BOED (vOED) by representing posterior distributions with expressive normalizing flows (NFs) and by deriving Monte Carlo estimators for a tight lower bound on EIG. This enables likelihood-free EIG estimation and joint optimization of variational parameters and design variables. Empirically, the flow-based posterior captures non-Gaussian, multimodal structure, yielding lower EIG bias at a fixed forward model budget than prior variational approaches, thereby making BOED more practical for expensive simulators. We demonstrate vOED-NFs on a range of models, including nonlinear benchmark designs, a real-world cathodic electrophoretic deposition (e-coating) application, and an implicit aphid population dynamics model.

In Chapter 3, we develop a non-myopic policy that selects designs across experiment stages to efficiently reduce uncertainty. We further introduce a goal-oriented formulation that targets parameters or QoIs within a unified information-theoretic framework. We validate the vsOED framework on a suite of standard and implicit-likelihood benchmarks, including source location

problems, constant elasticity of substitution (CES) models, SIR dynamics for disease spread, and convection–diffusion source inversion.

In Chapter 4, we instantiate this framework for Causal Bayesian Experimental Design (CBED) by treating interventions as design variables and targeting EIG for specific causal tasks, including causal discovery, edge existence prediction, and causal reasoning about intervention effects. The policy jointly selects a discrete intervention target and a continuous intervention value. Benchmarks include linear structural causal models and nonlinear mechanisms, such as neural networks, evaluated on Erdos–Renyi and scale-free graph families under cost-aware budgets. Performance is assessed using both EIG and downstream metrics, including Wasserstein distance for causal reasoning and structural Hamming distance and F1 for causal discovery.

Finally, in Chapter 5, we propose a novel symmetrized information gain (sEIG) criterion that replaces the conventional directional formulation with a symmetric one. We show that sEIG satisfies the axioms of a valid information measure and admits a simple Monte Carlo estimator requiring only $O(N)$ forward model runs, in contrast to the $O(N^2)$ cost of nested Monte Carlo for conventional EIG. We analyze estimator bias and variance and compare designs optimized under EIG and sEIG.

Overall, variational BOED with normalizing flows tightens EIG lower bound and improves posterior fidelity at fixed computational cost. The sequential extension raises cumulative information gain beyond greedy strategies, and the goal-oriented extension concentrates experimental effort on the stated objective. In causal experimental design, these advances improve causal reasoning and discovery, as reflected by higher information gain and improved downstream metrics. Finally, sEIG offers a promising path toward further efficiency gains relative to conventional EIG.

CHAPTER 1

Introduction

1.1 Background and motivation

1.1.1 Bayesian optimal experimental design

Mathematical models are central for understanding and making predictions in physical systems. Parameters in models are often uncertain, but their uncertainty can be reduced when information is gained from experimental observations. Since experiments are costly in many applications, optimal experimental design (OED) (see, e.g., [123, 30, 8, 128, 126, 75]) can bring substantial resource savings by identifying the experiments that produce the most valuable data.

Bayesian OED (BOED)[1] further incorporates the prior and posterior uncertainty in its objective function. The seminal paper by Lindley [93] proposed to use mutual information (MI) between model parameters and experimental data as the design criterion, which is also called the expected information gain (EIG) in the model parameters. The EIG, however, is generally intractable to compute and has to be approximated numerically.

One strategy entails directly approximating the EIG. For example, Ryan [129] introduced a nested Monte Carlo (NMC) estimator for the EIG, but it is computationally expensive with the number of likelihood evaluations scaling as the product of sample sizes of the nested Monte Carlo loops. Advanced numerics have been developed to accelerate NMC computations through surrogate modeling [77, 76], reusing samples across nested loops [77], and 'Gaussianizing' the posterior via Laplace approximations [96, 164, 162, 13]. Another estimator similar to the NMC is the prior contrastive estimator (PCE) [51]. A key difference is that PCE reuses each outer loop sample once in the inner loop, resulting the estimator mean to be negatively biased (i.e., a lower bound of the EIG). The above techniques, however, all require the ability to evaluate the likelihood function

[1] In the remainder of this thesis, we use OED and BOED interchangeably to refer to Bayesian experimental design approaches, without intending to distinguish them from non-Bayesian approaches unless explicitly stated.

and cannot accommodate implicit or intractable likelihood settings that often arise from stochastic models. A lower bound estimator (from directly approximating the EIG) that is able to handle intractable likelihood is the LB-KLD [5], which can be derived from Shannon's entropy power inequality.

Another strategy is to adopt a variational bound for the EIG [120] and tighten the bound by optimizing its variational parameters. (Note that PCE and LB-KLD do not have variational parameters for tightening the bounds, and therefore are not considered to be variational bounds.) We call these variational OED (vOED) approaches. For example, the tractable unnormalized Barber–Agakov (TUBA) lower bound [120] incorporates the tuning of a 'critic' function, and the Nguyen–Wainwright–Jordan (NWJ) bound [111], also known as the mutual information neural estimation f-divergence (MINE-f) bound [14], can be shown to be a special case of TUBA and has been used by Kleinegesse and Gutmann [86] in the OED context. Another lower bound, the information noise-contrastive estimation (InfoNCE) bound [150], takes a similar form as PCE, but replaces the likelihood with an exponentiated critic function. More recently, Ivanova *et al.* [80] extended the NWJ and InfoNCE bounds to the sequential OED setting. All of these bounds can accommodate implicit likelihoods.

A particular type of variational bounds, known as the Barber–Agakov (BA) lower bound [12], emerges from approximating the posterior density function directly. A potential advantage of the BA bound is its ability to enable both density evaluation and sampling from the resulting approximate posteriors. This is in contrast to several other lower bound choices, such as the NWJ bound that allows density evaluation but not sampling, and the InfoNCE bound that does not offer density or sampling. Foster *et al.* [50] were the first to use the BA bound in the OED setting, but remained with relatively simple variational distributions such as Gaussian, Bernoulli, and Gamma. The approximations thus can deteriorate when the true posterior distributions depart from these forms. Improving upon their work, Foster *et al.* [51] then introduced the adaptive contrastive estimation (ACE) lower bound based on the PCE but used a variational biasing distribution for the inner loop sampling. However, ACE required explicit likelihood.

1.1.2 Sequential Bayesian experimental design

Sequential experimental design involves planning multiple experiments conducted in sequence, where the results of earlier experiments can inform the design of subsequent ones—i.e., it involves 'feedback'. A straightforward approach is to design one experiment (or a subset) at a time: plan the next experiment, perform it, use the resulting data to update the model and its uncertainty, and repeat. This approach is known as greedy or myopic design since it focuses on the immediate next experiment without considering future ones; i.e., it lacks 'lookahead'. Greedy design is

advantageous for its simplicity and flexibility in accommodating an unknown total number of experiments, and it has been studied extensively [20, 44, 28, 136, 45, 46, 82, 66, 85].

Recent works follow a sequential experimental design formulation (see, e.g., [106], [155, VII.G], [74, Chapter 3]) that incorporates both (i) lookahead, where each design is selected while accounting for all remaining experiments, and (ii) feedback, where designs are given by adaptive policies that determine the next experiment based on the current state of information. This formulation, referred to as sequential OED (sOED) in [74, 78, 132], is general but computationally challenging. Previous approaches [27, 57, 122, 21, 33, 109, 158, 107, 143] have largely been limited to discrete settings or have not employed a Bayesian framework with information-theoretic design criteria. Only recently have more efficient computational techniques—often leveraging reinforcement learning (RL)—been developed to tackle sOED in more general settings [74, 78, 132, 49, 80, 18].

Huan and Marzouk [74, 78] presented sOED as a finite-horizon Markov decision process (MDP) and proposed an approximate dynamic programming algorithm to solve it. However, their approach relies on an implicit policy representation, where each policy evaluation involves solving a stochastic optimization problem, making it computationally expensive for online applications. To address this, Shen and Huan [134] introduced policy-gradient sOED (PG-sOED), an actor-critic method that trains an explicit parameterized policy network, significantly reducing computational costs. Around the same time, Foster *et al.* [49] proposed Deep Adaptive Design (DAD), which trains a policy network to maximize a lower bound estimator of the EIG, known as the sequential PCE, using its gradients. Unlike PG-sOED, DAD does not rely on an actor-critic framework but requires access to the derivatives of the observation model (i.e., parameter-to-observable map). Blau *et al.* [18] further proposed learning a stochastic policy to maximize the sequential PCE via randomized ensembled double Q-learning [31], which sidestepped the need for model derivatives. Building upon DAD, Ivanova *et al.* [80] developed implicit DAD (iDAD) to handle problems with implicit likelihood, where the likelihood is intractable but data samples can still be generated. iDAD constructs variational lower bounds of the EIG and tightens these bounds by simultaneously optimizing the policy and variational parameters, requiring only sample-based updates. However, similar to DAD, iDAD also requires model derivatives.

EIG bounds [120] (see Sec. 1.1.1) also have proven to be effective for sOED, as demonstrated by iDAD and related works. Notably, all these bounds support implicit likelihoods.

1.1.3 Goal-oriented Bayesian experimental design

The above OED efforts primarily focused on maximizing the EIG of unknown model parameters. However, the ultimate goals of the experiments often extend beyond parameter inference to predicting other goal quantities that depend on the model parameters but differ from the observation

quantities. For example, while observations such as temperature and pressure facilitate the inference of a weather model's parameters, the goal might not be to learn these parameters, but to forecast future precipitation using the resulting parameter distributions.

A goal-oriented OED framework therefore directly targets the EIG for such goal quantities. This concept builds on classical I-, V-, and G-optimal designs that seek to minimize predictive variance in linear regression models [8]. Attia *et al.* [9] demonstrated gradient-based optimization techniques to solve linear Bayesian D_A- and L-optimal design problems, while Wu *et al.* [163] developed scalable offline-online decompositions and low-rank approximations for high-dimensional linear settings. The theoretical formulation for nonlinear goal-oriented OED was presented in [15]. Computationally, Butler *et al.* [25] proposed OED for prediction (OED4P), a non-Bayesian approach based on matching push-forward distributions to observed data distributions [24, 26]. Later, Smith *et al.* [16] introduced expected predictive information gain (EPIG), a nested Monte Carlo estimator for predictive EIG under a greedy sequential (i.e., active learning) setting, but requiring the goal quantities to be the same as observation quantities except at new designs. More recently, Zhong et al. [166] combined Markov chain Monte Carlo with kernel density estimation to estimate the probability densities of goal quantities directly. These studies highlight the challenge of intractable predictive densities for the goal quantities, motivating the potential effectiveness of likelihood-free EIG bounds. Kleinegesse and Gutmann [87] explored such bounds for various goal-oriented OED settings. However, existing goal-oriented efforts all focused on batch (i.e., non-sequential) or greedy OED, and integrating goal-oriented objectives into sOED has yet to be developed.

1.1.4 Causal Bayesian experimental design

Causal OED studies how to choose interventions that are maximally informative about a structural causal model (SCM), which provides a mathematical framework for representing causal relationships via a directed acyclic graph (DAG). Early work in causal OED demonstrated the utility of active interventions for efficiently uncovering causal graph structures, moving beyond passive observational learning [108, 148, 32, 110, 154, 140]. Subsequent research expanded to learning full SCMs, including the selection of both intervention targets and values [147, 146]. More recent work has highlighted the importance of tailoring experiments to specific causal queries [149]. However, many of these approaches remain myopic—focusing on single-step gains—or are oriented toward global fidelity rather than user-specific causal objectives.

In parallel, the broader OED literature has seen growing interest in goal-orientation design, where experiments are optimized for their utility to downstream tasks [9, 161, 166, 29]. These methods have shown substantial benefits in predictive settings, particularly with complex nonlinear models. However, they generally do not address the unique challenges of causal inference, including

the interventional nature of learning and the structural constraints of SCMs.

Recognizing the limitations of greedy approaches, non-myopic sequential OED seeks to optimize entire experimental trajectories rather than one step at a time. Approaches based on amortized policy learning and RL [49, 18, 134, 132] have shown promise in this area. Yet in the causal setting, non-myopic strategies often focus solely on structure learning [4, 56] and do not integrate flexible, user-defined causal goals into the long-term optimization framework.

1.1.5 Expected information gain for experimental design

While EIG is the conventional measure of information in OED, it is not the only way to quantify information in an experiment. Ginebra [61] formalized a minimal set of axiomatic requirements for valid information measures: (i) the measure must be a well-defined real number; (ii) it should return zero for a completely non-informative experiment, i.e., when the data are independent of the parameters; and (iii) it must respect the sufficiency ordering of experiments, meaning that if experiment A is statistically more informative (in the Blackwell sense) than experiment B, then A's information measure must be larger. Lindley's EIG satisfies these conditions [93], confirming its soundness, but these axioms also permit alternative criteria.

We propose one such alternative, the symmetric expected information gain (sEIG), which captures the information between parameters and data in a bidirectional manner, rather than focusing only on the one-way reduction of parameter uncertainty. Importantly, sEIG also satisfies Ginebra's axiomatic requirements, ensuring its theoretical validity as a measure of information. Whereas EIG emphasizes how much the posterior contracts relative to the prior, sEIG emphasizes the alignment between the prior and posterior distributions—a perspective that is distinct but equally principled. This flexibility underscores that EIG is not uniquely valid; other measures like sEIG may provide additional robustness and practical advantages.

From a computational perspective, the motivation for considering sEIG stems from the prohibitive cost of estimating conventional EIG. Exact computation of EIG is typically intractable for nonlinear or simulator-based models, and numerical approximations usually rely on the NMC estimator. In NMC, an outer loop samples potential data realizations, and for each realization, an inner loop estimates its marginal likelihood, resulting in a computational cost that grows as the product of the loop samples ($O(M \times N)$ forward model runs). Such scaling quickly becomes impractical when the forward model itself is expensive, such as models based on differential equations. Moreover, the NMC estimator suffers from high error under small sample sizes, making reliable estimation challenging in high-dimensional or resource-constrained settings.

sEIG mitigates these difficulties by eliminating the need for nested sampling. Owing to its symmetric formulation, sEIG admits a single-loop Monte Carlo estimator with linear computational

dependence ($O(N)$). This makes it a natural candidate to explore as an alternative to conventional EIG, particularly in settings where forward model runs are expensive or likelihoods are implicit.

1.2 Objectives and outline

The objectives of this thesis are formed around addressing several key research gaps.

Objective 1 (Reduce the Bias in BA Bound Estimator)**.**

One important research gap in for EIG estimation—particularly those base on variational bounds—is the BA bound estimator bias (even asymptotically under infinite samples) in estimating the EIG when approximating the posterior with over-simplified variational family such as mean-field Gaussian. We propose to reduce this bias by improving the accuracy of posterior approximation through enriching the space of variational distributions using normalizing flows (NFs) [39, 127, 115, 88] in vOED, we call this approach **vOED-NFs**. We achieve this while maintaining the ability to handle implicit likelihoods. The key novelty and contributions are as follows.

- We present the use of NFs for estimating the EIG lower bound in vOED, and illustrate its sampling efficiency (lower bias) compared to previous approaches.
- We show the ability of NFs trained for vOED to achieve good posterior approximations, including non-Gaussian and multi-modal distributions.
- We validate vOED-NFs against established reference methods, and demonstrate vOED-NFs on a partial differential equation (PDE)-governed application of cathodic electrophoretic deposition.
- We illustrate vOED-NFs in handling an implicit likelihood case with stochastic modeling of aphid population.

Objective 2 (Unify Sequential and Goal-Oriented OED)**.**

Despite recent progress, there does not exist a unified framework that (i) is goal-oriented, optimizing information about user-specified quantities rather than global model uncertainty; (ii) is sequential and non-myopic, planning entire intervention trajectories; and (iii) is computationally viable for expensive and possibly implicit likelihood models, with flexible posterior approximations and amortized policies for real-time decisions.

We fill this gap by introducing variational sequential optimal experimental design (vsOED), a method to solve the sOED problem under an actor-critic RL framework powered by policy gradient

and variational BA bound to the EIG. While vsOED shares some foundational elements with the previous sOED approach in [132]—such as the finite-horizon MDP formulation and actor-critic policy gradient solution strategy—it introduces several key differences. These distinctions are reflected in the main novelty and contributions of this work, summarized below.

- We formulate vsOED that accommodates nuisance parameters, implicit likelihoods, and multiple candidate models, while offering flexible design criterion that can target designs for model discrimination, parameter inference, goal-oriented prediction, or weighted combinations thereof.
- We develop techniques for solving vsOED, highlighted by estimating the actor-critic policy gradient and approximating posteriors using Gaussian mixture models and NFs.
- We illustrate the effectiveness of vsOED, particularly under limited sample budgets, through comparisons with existing algorithms across diverse numerical experiments.

Objective 3 (Enable Sequential and Goal-Oriented OED for Causal Learning)**.**

Most causal OED methods optimize EIG for full-model recovery (structure and mechanisms), even when the scientific task is a specific causal query (interventional or counterfactual). Moreover, existing causal design strategies are predominantly greedy or one-step lookahead, ignoring synergies across experiments and often yielding suboptimal long-horizon policies. In causal settings, design should also contend with (i) mixed discrete–continuous decisions (jointly choosing targets and intervention values) and (ii) high-dimensional, relational observations.

To address these limitations, we implement Goal-Oriented Causal Bayesian Experimental Design (GO-CBED), a sequential, non-myopic framework that:

- Directly targets user-defined causal queries by maximizing the BA bound on the EIG for the desired quantity of interest (QoI) [120, 12];
- Plans non-myopically over entire intervention sequences using a learned RL policy to capture cross-experiment complementarities;
- Enables real-time intervention selection via an amortized transformer-based policy trained offline for fast deployment.

Objective 4 (Develop Symmetric Expected Information Gain for Bayesian OED)**.**

While EIG is theoretically sound and widely used, symmetric expected information gain (sEIG) is a valid alternative that satisfies Ginebra's axioms for information measures and admits a single-loop ($O(N)$) Monte Carlo estimator. Yet, its role in OED remains underexplored. In particular, there is no

systematic assessment of (i) estimator properties (bias/variance, sample efficiency) of sEIG versus NMC EIG across likelihood-free and simulator-based settings; (ii) design quality and downstream task performance under sEIG-optimized designs as compared to EIG; and (iii) integration of sEIG within modern OED pipelines, including sequential (non-myopic) and goal-oriented formulations, and causal design (mixed discrete-continuous interventions). This objective fills these gaps by developing sEIG-based estimators and optimization routines, establishing empirical comparisons to EIG on PDE-governed models, and evaluating their impact on both classical parameter-centric utilities and task-specific goal metrics.

Overall, this thesis develops a unified, information-theoretic program for Bayesian OED—progressing from variational OED (vOED) to sequential, non-myopic OED (vsOED) that are both capable of handling goal-oriented design, and culminating in a causal instantiation. In parallel, it also explores sEIG as an alternative criterion to conventional EIG, motivated by its distinct interpretation and computational efficiency.

The dissertation is organized as follows: Chapter 2 introduces **vOED-NFs**, a variational OED method that reduces BA-bound bias by enriching posterior families with NFs while accommodating implicit likelihoods, addressing Objective 1. Chapter 3 presents **vsOED**, a sequential BOED framework that maximizes a variational EIG bound with an amortized actor-critic policy and supports OED for model discrimination, parameter inference, and goal-oriented designs under nuisance parameters and implicit likelihoods. This unified treatment of vsOED is evaluated against multiple state-of-the-art baselines and addresses Objective 2. Chapter 4 instantiates the framework for causal learning via **GO-CBED**: sequential, non-myopic, goal-oriented causal design. It handles mixed discrete-continuous intervention decisions, and evaluates on synthetic SCMs and semi-synthetic gene regulatory networks, addressing Objective 3. Chapter 5 introduces **sEIG** as an alternative measure of information with a more computationally efficient estimator and a distinct interpretation from conventional EIG, while still remaining a valid information measure. It compares this approach with the traditional NMC estimator for EIG, focusing on computational efficiency and addressing Objective 4.

CHAPTER 2

Variational Bayesian Optimal Experimental Design with Normalizing Flows

Direct optimization of the expected information gain (EIG) is often computationally prohibitive due to its nested expectation structure. Variational optimal experimental design (vOED) circumvents this difficulty by optimizing a tractable lower bound of the EIG, replacing repeated posterior inference with an amortized variational approximation. This reformulation substantially improves sampling efficiency and enables scalable design optimization.

Research gap and contributions. Despite the computational advantages of vOED, the resulting bound estimators are generally biased, with the magnitude of the bias depending on the expressiveness of the chosen variational family. This bias limits the accuracy of EIG estimation and, consequently, the quality of the optimized designs. The central research question addressed in this chapter is whether more expressive variational approximations can substantially reduce this bias while preserving the scalability and amortization benefits of the vOED framework. To answer this question, this chapter introduces a normalizing-flows-based vOED framework (vOED-NFs) that enriches the variational posterior family, yielding tighter EIG lower bounds and improved posterior approximations. The key contributions include (i) the integration of normalizing flows into the variational EIG estimation framework, (ii) empirical demonstration of reduced bias and improved sampling efficiency relative to existing methods, and (iii) validation of the proposed approach on both implicit-likelihood models and PDE-governed applications.

The contents of this chapter are published in [40].

2.1 Problem formulation

2.1.1 Bayesian optimal experimental design

We adopt the following notation: upper case for random variable, lower case for realization, and subscript in a probability density function (PDF) is generally omitted but retained in some cases for clarification or emphasis; for example, $p_X(X = x) = p(x)$ denotes the PDF of random vector X evaluated at value $X = x$.

When an experiment is conducted under design $\xi \in \mathcal{A} \subseteq \mathbb{R}^{n_\xi}$ and yields an observation $Y = y \in \mathbb{R}^{n_y}$, the PDF for the model parameters $\Theta \in \mathbb{R}^{n_\theta}$ is updated via Bayes' rule by

$$p(\theta|y, \xi) = \frac{p(\theta)\, p(y|\theta, \xi)}{p(y|\xi)}, \tag{2.1}$$

where $p(\theta)$ is the prior PDF (we assume the prior does not depend on ξ), $p(y|\theta, \xi)$ is the data likelihood, $p(y|\xi)$ is the marginal likelihood (model evidence), and $p(\theta|y, \xi)$ is the posterior PDF. The likelihood can be computed based on an underlying observation model, for example,

$$Y = G(\Theta, \xi) + \mathcal{E}, \tag{2.2}$$

where $G : \mathbb{R}^{n_\theta} \times \mathbb{R}^{n_\xi} \to \mathbb{R}^{n_y}$ is the forward model (e.g., from solving a governing system of PDEs) and $\mathcal{E}$ is the random noise associated with the observation model. Each likelihood evaluation thus entails computing $p(y|\theta, \xi) = p_{\mathcal{E}}(y - G(\theta, \xi))$, which requires a forward model solve.

We adopt the EIG on Θ to reflect the expected utility of performing an experiment. Mathematically, the EIG is the expected Kullback–Leibler (KL) divergence from the parameter prior to the posterior, and can be interpreted as the expected parameter uncertainty reduction due to the experiment:

$$U(\xi) = \mathbb{E}_{Y|\xi}[D_{\mathrm{KL}}(\, p_{\Theta|Y,\xi} \,||\, p_\Theta \,)] = \iint p(\theta, y|\xi) \log[\frac{p(\theta|y, \xi)}{p(\theta)}] d\theta\, dy. \tag{2.3}$$

The Bayesian OED problem then entails solving for the optimal design:

$$\xi^* = \underset{\xi \in \mathcal{A}}{\operatorname{argmax}}\, U(\xi). \tag{2.4}$$

2.1.2 Variational optimal experimental design

We summarize below the variational OED (vOED) approach introduced by Foster *et al.* [50]. vOED replaces the posterior or marginal likelihood with an approximate PDF $q(\cdot)$ parameterized by ϕ

$\in \mathbb{R}^{n_\phi}$. When approximating the posterior, the expected utility from Eqn. (2.3) becomes

$$U_L(\xi;\phi) = \mathbb{E}_{\Theta,Y|\xi}[\log \frac{q(\Theta|Y;\phi)}{p(\Theta)}] = \iint p(\theta, y|\xi) \log[\frac{q(\theta|y;\phi)}{p(\theta)}] d\theta\, dy. \tag{2.5}$$

Note that the outer expectation is still taken with respect to the true distribution $p(\theta, y|\xi)$. $U_L(\xi;\phi)$ is known as the Barber–Agakov bound [12]; it is a lower bound for $U(\xi)$:

$$U(\xi) - U_L(\xi;\phi) = \mathbb{E}_{Y|\xi}[D_{\mathrm{KL}}(\, p_{\Theta|Y,\xi} \,||\, q_{\Theta|Y;\phi}\,)] \geq 0. \tag{2.6}$$

The bound is tight if and only if $q(\theta|y;\phi) = p(\theta|y,\xi)$ for all θ and y. The best (tightest) lower bound at a given ξ is then

$$U_L(\xi;\phi^*) = \max_\phi U_L(\xi;\phi). \tag{2.7}$$

We note that the ordering of posterior approximation quality for $q(\theta|y;\phi)$ as measured by $U_L(\xi;\phi)$ is preserved when measured by the expected KL divergence in the form of moment projection as used in expectation propagation [103] (i.e., the 'reverse' KL divergence compared to the information projection form used in variational inference). Hence, the approximate posterior resulting from a tighter lower bound is also closer to the true posterior in this expected KL sense. We make this precise in the following proposition.

Proposition 2.1.1. *Consider probability densities $q(\theta|y;\phi_1)$ and $q(\theta|y;\phi_2)$ formed at variational parameter values ϕ_1 and ϕ_2, respectively, both as approximations to the true posterior density $p(\theta|y,\xi)$. Then,*

$$U_L(\xi;\phi_1) \leq U_L(\xi;\phi_2) \tag{2.8}$$

if and only if

$$\mathbb{E}_{Y|\xi}\left[D_{\mathrm{KL}}(p_{\Theta|y,\xi} \,||\, q_{\Theta|y;\phi_1})\right] \geq \mathbb{E}_{Y|\xi}\left[D_{\mathrm{KL}}(p_{\Theta|y,\xi} \,||\, q_{\Theta|y;\phi_2})\right]. \tag{2.9}$$

A proof of the proposition is provided in Eqn. (A.1).

With $U_L(\xi;\phi^*)$ being the best variational approximation to the true EIG at ξ, the vOED problem looks for the design that maximizes the tightest bound by simultaneously optimizing for ξ and ϕ:

$$\xi^*, \phi^* = \operatorname*{argmax}_{\xi\in\mathcal{A},\phi} U_L(\xi;\phi). \tag{2.10}$$

When $U_L(\xi;\phi)$ is differentiated with respect to ξ and ϕ, their gradients are

$$\nabla_\phi U_L(\xi;\phi) = \mathbb{E}_\Theta\{\mathbb{E}_{Y|\Theta,\xi}[\nabla_\phi \log q(\Theta|Y;\phi)]\}, \tag{2.11}$$

$$\begin{aligned}\nabla_\xi U_L(\xi;\phi) &= \mathbb{E}_\Theta\{\nabla_\xi \mathbb{E}_{Y|\Theta,\xi}[\log q(\Theta|Y;\phi)]\} = \mathbb{E}_\Theta\left\{\int [\nabla_\xi p(y|\theta,\xi)] \log q(\theta|y;\phi) dy\right\} \\ &= \mathbb{E}_\Theta\{\mathbb{E}_{Y|\Theta,\xi}[\log q(\Theta|Y;\phi)\nabla_\xi \log p(Y|\Theta,\xi)]\},\end{aligned} \tag{2.12}$$

where the last equality uses the identity $\nabla_\xi \log p(y|\theta,\xi) = \frac{\nabla_\xi p(y|\theta,\xi)}{p(y|\theta,\xi)}$. Access to these gradient formulas allows the adoption of gradient-based optimization algorithms to solve Eqn. (2.10).

Alternatively, one can approximate the marginal likelihood instead of the posterior as introduced in [50], and the expected utility becomes

$$U_U(\xi;\phi) = \mathbb{E}_{\Theta,Y|\xi}\left[\log \frac{p(Y|\Theta,\xi)}{q(Y;\phi)}\right] = \iint p(\theta,y|\xi) \log\left[\frac{p(y|\theta,\xi)}{q(y;\phi)}\right] d\theta\, dy, \tag{2.13}$$

where the posterior-to-prior density ratio has been replaced with likelihood-to-marginal-likelihood density ratio via Bayes' rule in Eqn. (2.1). Again, the outer expectation remains with respect to the true distribution $p(\theta,y|\xi)$. U_U forms an upper bound for $U(\xi)$:

$$U_U(\xi;\phi) - U(\xi) = D_{\mathrm{KL}}(\, p_{Y|\xi} \,||\, q_{Y;\phi}\,) \geq 0. \tag{2.14}$$

The bound is tight if and only if $q(y;\phi) = p(y|\xi)$ for all y. The best (tightest) upper bound at a given ξ is then

$$U_U(\xi;\phi^*) = \min_\phi U_U(\xi;\phi). \tag{2.15}$$

The corresponding vOED problem using $U_U(\xi;\phi)$ involves solving for

$$\xi^* = \operatorname*{argmax}_{\xi\in\mathcal{A}} \min_\phi U_U(\xi;\phi). \tag{2.16}$$

When $U_U(\xi;\phi)$ is differentiate with respect to ϕ, its gradient is

$$\nabla_\phi U_U(\xi;\phi) = -\mathbb{E}_{Y|\xi}[\nabla_\phi \log q(Y;\phi)]. \tag{2.17}$$

The use of $U_U(\xi;\phi)$ can offer advantages when $n_y \ll n_\theta$ so that the variational density approximation is applied to a lower-dimensional space. However, the use of $U_U(\xi;\phi)$ involves solving a more difficult maximin saddle-point problem in Eqn. (2.16), and its two optimizations cannot be combined like in Eqn. (2.10). In this work, we will primarily focus on the lower bound $U_L(\xi;\phi)$

and its corresponding vOED problem in Eqn. (2.10).

2.2 Computational methods

The expected utility in Eqn. (2.3) is generally intractable and needs be estimated numerically. One widely used approach is the NMC estimator [129]:

$$U(\xi) \approx \widehat{U}_{\text{NMC}}(\xi) := \frac{1}{N_{\text{out}}} \sum_{i=1}^{N_{\text{out}}} \left\{ \log p(y^{(i)}|\theta^{(i)}, \xi) - \log[\frac{1}{N_{\text{in}}} \sum_{j=1}^{N_{\text{in}}} p(y^{(i)}|\theta^{(i,j)}, \xi)] \right\}, \tag{2.18}$$

where samples $\theta^{(i)} \sim p(\theta)$, $y^{(i)} \sim p(y|\theta^{(i)}, \xi)$, and $\theta^{(i,j)} \sim p(\theta)$ are drawn from the prior and likelihood. Calculating the NMC estimator $\widehat{U}_{\text{NMC}}$ requires evaluating the likelihood $p(y|\theta, \xi)$ and cannot directly handle implicit likelihoods settings. $\widehat{U}_{\text{NMC}}$ is a biased estimator to U under any finite inner-loop sample size N_{in}, but it is asymptotically unbiased as $N_{\text{in}} \to \infty$. Properties of the NMC estimator has been studied extensively [129, 47, 13, 125], and we will use it for reference comparison in this work when applicable.

For vOED, the lower and upper bounds in Eqn. (2.5) and (2.13) require numerical approximations for their expectation operators, for example through standard Monte Carlo (MC) estimation:

$$U_L(\xi; \phi) \approx \widehat{U}_L(\xi; \phi) := \frac{1}{N} \sum_{i=1}^{N} \left\{ \log q(\theta^{(i)}|y^{(i)}; \phi) - \log p(\theta^{(i)}) \right\}, \tag{2.19}$$

$$U_U(\xi; \phi) \approx \widehat{U}_U(\xi; \phi) := \frac{1}{N} \sum_{i=1}^{N} \left\{ \log p(y^{(i)}|\theta^{(i)}, \xi) - \log q(y^{(i)}; \phi) \right\}, \tag{2.20}$$

where $\theta^{(i)} \sim p(\theta)$ and $y^{(i)} \sim p(y|\theta^{(i)}, \xi)$. In particular, calculating Eqn. (2.19) only requires sampling the likelihood without any likelihood evaluations; it is therefore suitable to handle implicit likelihoods. Note that both $\widehat{U}_L$ and $\widehat{U}_U$ are not bounds themselves, but are (unbiased) estimators of the bounds U_L and U_U, respectively. With respect to the true EIG U, then, $\widehat{U}_L$ and $\widehat{U}_U$ are biased estimators whose biases are, respectively, negative and positive.

We can also form MC estimators to the lower bound gradients in Eqn. (2.11) and (2.12), yielding

$$\widehat{\nabla_\phi U_L}(\xi; \phi) := \frac{1}{N_{\text{batch}}} \sum_{i=1}^{N_{\text{batch}}} \nabla_\phi \log q(\theta^{(i)}|y^{(i)}; \phi), \tag{2.21}$$

$$\widehat{\nabla_\xi U_L}(\xi; \phi) := \frac{1}{N_{\text{batch}}} \sum_{i=1}^{N_{\text{batch}}} \log q_\phi(\theta^{(i)}|y^{(i)}, \xi) \nabla_\xi \log p(y^{(i)}|\theta^{(i)}, \xi). \tag{2.22}$$

Similarly, the upper bound gradient in Eqn. (2.17) yields

$$\widehat{\nabla_\phi U_U}(\xi;\phi) := -\frac{1}{N_{\text{batch}}}\sum_{i=1}^{N_{\text{batch}}} \nabla_\phi \log q(y^{(i)};\phi). \tag{2.23}$$

In this work, we employ stochastic gradient ascent (SGA) for optimization. In order to control the total number of forward model runs, instead of sampling $(\theta^{(i)}, y^{(i)})$ from their true distributions every instance, we first generate a pool of N_{opt} sample pairs at each ξ, and then subdivide them into mini-batches with size N_{batch} for the SGA update iterations.

Foster *et al.* [50] initially proposed parameterizing $q(\,\cdot\,;\phi)$ using relatively simple forms of distributions such as Gaussian, Bernoulli, Beta, and simple transformations of them. Below we present the use of NFs to parameterize $q(\,\cdot\,;\phi)$, with the potential to capture a richer space of distributions and in turn to achieve tighter U_L and U_U bounds.

2.2.1 Normalizing flows

An NF is an invertible mapping from a target random variable $X \sim p_X(x)$ to a standard normal random variable $Z \sim p_Z(z) = \mathcal{N}(\mathbf{0}, \mathbb{I})$ of the same dimension: $Z = f(X)$ and $X = g(Z)$ where $g := f^{-1}$. In terms of transport maps, f is known as the pushforward map and g is the pullback map [101]. In practice, we approximate f via a mapping parameterized with ϕ, which produces an approximate transformation $\widetilde{Z} = f(X;\phi)$ and its inverse $g(\,\cdot\,;\phi) := f(\,\cdot\,;\phi)^{-1}$ produces $X = g(\widetilde{Z};\phi)$. If acting on the exact standard normal Z, then $\widetilde{X} = g(Z;\phi)$ and also $Z = f(\widetilde{X};\phi)$.

In general, the approximate mappings used in NFs are often structured as compositions of successive simple invertible mappings: $f(\widetilde{X};\phi) = f_n \circ f_{n-1} \circ \ldots \circ f_1(\widetilde{X}) = f_n(f_{n-1}(\ldots(f_1(\widetilde{X}))\ldots))$ and $g(Z;\phi) = g_1 \circ \ldots g_{n-1} \circ g_n(Z) = g_1(g_2(\ldots(g_n(Z))\ldots))$ with $g_i = f_i^{-1}$ and $n \geq 1$. Note that all intermediate mappings f_i and g_i depend on ϕ, but we omit their subscripts to simplify notation. The log density of $\widetilde{X}$ can be tracked via the change-of-variable formula:

$$\log q_{\widetilde{X}}(\widetilde{X} = x;\phi) = \log p_Z(f_n \circ f_{n-1} \circ \ldots \circ f_1(\widetilde{X} = x)) + \sum_{i=1}^{n} \log \left| \det \frac{\partial f_i \circ f_{i-1} \circ \ldots f_1(\widetilde{X})}{\partial \widetilde{X}} \right|_{\widetilde{X}=x}, \tag{2.24}$$

where $\frac{\partial f_i(\widetilde{X})}{\partial \widetilde{X}}$is the Jacobian of f_i. By applying successive transformations on Z, the PDF of the resulting variable can be highly expressive [127, 141] and effective for multi-modal, skewed, or other non-standard distribution shapes.

In the context of vOED lower bound in Eqn. (2.5), we use NF-based $q(\theta|y;\phi)$ to approximate $p(\theta|y,\xi)$. Among a range of choices for the architecture of invertible mappings [115, 88], we

adopt the coupling layers [39] as a special type of invertible neural network (INN) [91, 124] for its efficient density evaluations and sampling in both forward (f) and inverse (g) directions. Several papers have shown that composing coupling layers can create flexible flows [84, 121, 6, 91], and recent work by Draxler *et al.* [43] shows that coupling layers form a distributional universal approximator. The basic form of the coupling layer that completes one full transformation starts by partitioning the standard normal random vector $Z = [Z_1, Z_2]^\top$ into two parts of approximately equal dimension—that is, $Z_1 \in \mathbb{R}^{n_{\theta_1}}$, $Z_2 \in \mathbb{R}^{n_{\theta_2}}$, and $n_{\theta_1} + n_{\theta_2} = n_\theta$—and then composes together $n = 2$ transformations that transform one part at a time. This maps Z to an approximate target random vector $\widetilde{\Theta}$, i.e., $g(Z; \phi) = g_1 \circ g_2(Z) = \widetilde{\Theta}$, and is defined as:

$$g_2(Z) = \begin{bmatrix} \widetilde{\Theta}_1 = [Z_1 - t_2(Z_2)] \odot \exp[-s_2(Z_2)] \\ Z_2 \end{bmatrix}, \tag{2.25}$$

$$g_1(g_2(Z)) = \begin{bmatrix} \widetilde{\Theta}_1 \\ \widetilde{\Theta}_2 = [Z_2 - t_1(\widetilde{\Theta}_1))] \odot \exp(-s_1(\widetilde{\Theta}_1)) \end{bmatrix}, \tag{2.26}$$

where $\odot$ denotes element-wise product, and $s_1, t_1 : \mathbb{R}^{n_{\theta_1}} \to \mathbb{R}^{n_{\theta_2}}$ and $s_2, t_2 : \mathbb{R}^{n_{\theta_2}} \to \mathbb{R}^{n_{\theta_1}}$ are arbitrary functions. The parameterizations of these functions make up ϕ; for instance, if these functions are represented by neural networks, then ϕ encompasses all neural networks' weight and bias parameters.

The inverse of $g(\,\cdot\,; \phi)$, which is $f(\widetilde{\Theta}; \phi) = f_2 \circ f_1(\widetilde{\Theta}) = Z$, similarly involves partitioning $\widetilde{\Theta} = [\widetilde{\Theta}_1, \widetilde{\Theta}_2]^\top$ and can be shown to be:

$$f_1(\widetilde{\Theta}) = \begin{bmatrix} \widetilde{\Theta}_1 \\ Z_2 = \widetilde{\Theta}_2 \odot \exp(s_1(\widetilde{\Theta}_1)) + t_1(\widetilde{\Theta}_1) \end{bmatrix}, \tag{2.27}$$

$$f_2(f_1(\widetilde{\Theta})) = \begin{bmatrix} Z_1 = \widetilde{\Theta}_1 \odot \exp(s_2(Z_2)) + t_2(Z_2) \\ Z_2 \end{bmatrix}. \tag{2.28}$$

The Jacobians of f_1 and f_2 are triangular matrices:

$$\frac{\partial f_1(\widetilde{\Theta})}{\partial \widetilde{\Theta}} = \begin{bmatrix} \mathbb{I}_{n_{\theta_1}} & \mathbf{0} \\ \frac{\partial Z_2}{\partial \widetilde{\Theta}_1} & \operatorname{diag}(\exp(s_1(\widetilde{\Theta}_1))) \end{bmatrix}, \qquad \frac{\partial f_2(f_1(\widetilde{\Theta}))}{\partial (f_1(\widetilde{\Theta}))} = \begin{bmatrix} \operatorname{diag}(\exp(s_2(Z_2)) & \frac{\partial Z_1}{\partial Z_2} \\ \mathbf{0} & \mathbb{I}_{n_{\theta_2}} \end{bmatrix}, \tag{2.29}$$

and have respective determinants $\exp(\sum_{j=1}^{n_{\theta_2}} s_1(\widetilde{\Theta}_1)_j)$ and $\exp(\sum_{j=1}^{n_{\theta_1}} s_2(Z_2)_j)$.

Multiple such complete transformations can be composed together for greater expressiveness. For example, adopting NFs with $T = 3$ sets of complete transformations entails $f(\widetilde{\Theta}; \phi) = (f_2 \circ f_1)^{T3} \circ (f_2 \circ f_1)^{T2} \circ (f_2 \circ f_1)^{T1}(\widetilde{\Theta})$ and $g(Z; \phi) = (g_1 \circ g_2)^{T3} \circ (g_1 \circ g_2)^{T2} \circ (g_1 \circ g_2)^{T1}(Z)$.

To incorporate the y-dependence into the approximate posteriors $q(\theta|y; \phi)$, the s and t functions

are designed to additionally take y as input, leading to a form of conditional INNs (cINNs) [114]. When the dimension of y is large, a summary network [124] that is common to all the s and t functions, effectively functioning as an encoder, can be employed to compress y into a lower-dimensional summary statistic (i.e., latent variables) y'. The parameters of the summary network then become part of ϕ. The summary network can be particularly beneficial to contain the growth of parameters when s and t functions are represented by neural networks. Figure 2.1 illustrates the overall cINN transformation for vOED lower bound.

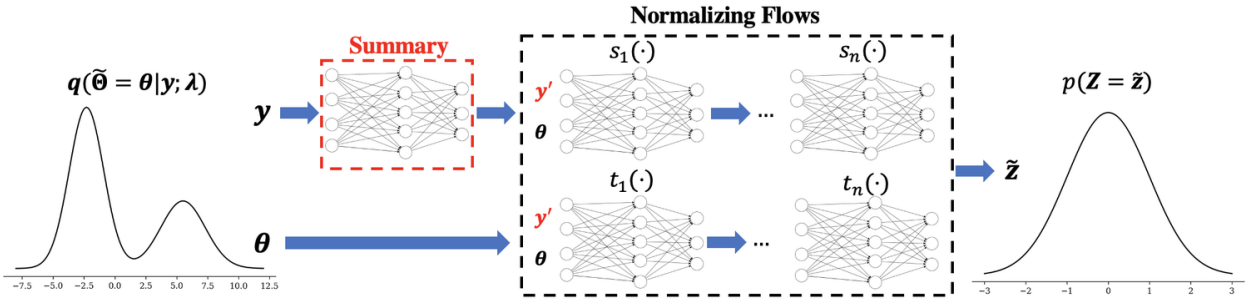


Figure 2.1: Diagram for the cINN NFs transformation.

In the case of vOED upper bound in Eqn. (2.13), NF-based $q(y;\phi)$ can be similarly used to approximate $p(y)$. The transformation $g(Z;\phi)$ then mirrors Eqn. (2.25) and (2.26), while $f(Y;\phi)$ mirrors Eqn. (2.27) and (2.28).

In summary, the final approach to determine the optimal design entails maximizing the EIG via its lower bound:

$$\xi^*, \phi^* = \underset{\xi\in\mathcal{A},\phi}{\operatorname{argmax}}\, \widehat{U}_L = \underset{\xi,\phi}{\operatorname{argmax}}\, \frac{1}{N}\sum_{i=1}^{N}\left\{\log q(\theta^{(i)}|y^{(i)};\phi) - \log p(\theta^{(i)})\right\}, \tag{2.30}$$

where $\theta^{(i)} \sim p(\theta)$, and $y^{(i)} \sim p(y^{(i)}|\theta^{(i)},\xi)$, and ϕ represents all the neural network weight and bias parameters in the cINN and summary network. For results that show upper bound estimates, they are based on solving the problem:

$$\xi^* = \underset{\xi\in\mathcal{A}}{\operatorname{argmax}}\, \min_{\phi}\, \widehat{U}_U = \underset{\xi\in\mathcal{A}}{\operatorname{argmax}}\, \min_{\phi}\, \frac{1}{N}\sum_{i=1}^{N}\left\{\log p(y^{(i)}|\theta^{(i)},\xi) - \log q(y^{(i)};\phi)\right\}. \tag{2.31}$$

2.3 Numerical experiments

We demonstrate vOED-NFs across four design cases involving models of varying complexity: (1) a nonlinear model with 3 parameters and 1 design variable; (2) a linear model with 21 parameters

and 400-dimensional design vector; (3) a PDE model for cathodic electrophoretic deposition; and (4) a stochastic model of aphid population with an implicit likelihood.

2.3.1 Case 1: Low-dimensional nonlinear design

Consider a nonlinear observation model

$$y = G(\theta, \xi) + \epsilon \qquad \text{where} \qquad G(\theta, \xi) = \theta_1^3 \xi^2 + \theta_2 e^{-|0.2-\xi|} + \sqrt{\theta_3^2 \xi \times 2}, \tag{2.32}$$

with independent prior on parameters $\theta_1 \sim \mathcal{N}(0.5, 0.3^2), \theta_2 \sim \mathcal{N}(0.3, 0.7^2), \theta_3 \sim \mathcal{N}(0.5, 0.8^2)$; $\xi \in [0, 1]$; and measurement noise ϵ following a Gaussian mixture distribution:

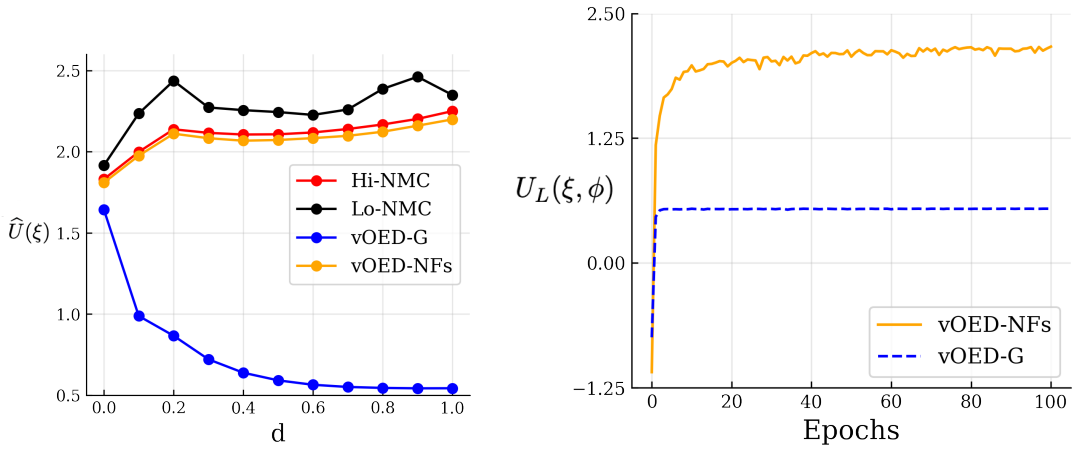
$$p(\epsilon) = \frac{1}{2\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(\epsilon-\mu_1)^2}{2\sigma_0^2}\right) + \frac{1}{2\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(\epsilon-\mu_2)^2}{2\sigma_0^2}\right), \tag{2.33}$$

where $\mu_1 = 0.1, \mu_2 = -0.1, \sigma = 0.05$.

Figure 2.2a presents the EIG estimates across the discretized design space using (a) (Hi-NMC) a high-quality $\widehat{U}_{\text{NMC}}$ with $N_{\text{out}} = 2 \times 10^4$ and $N_{\text{in}} = 2 \times 10^4$ samples, equating to 4×10^8 forward model evaluations total, (b) (Lo-NMC) a low-quality $\widehat{U}_{\text{NMC}}$ with $N_{\text{out}} = 200$ and $N_{\text{in}} = 200$ samples, equating to 4×10^4 forward model evaluations total, (c) (Re-NMC) a NMC estimator with $N_{\text{out}} = N_{\text{in}} = 2 \times 10^4$, where the inner loop reuses the outer loop samples, equating to 2×10^4 forward model evaluations total, (d) (vOED-G) EIG lower bound $\widehat{U}_L$ using Gaussian distributions optimized with $N_{\text{opt}} = 2 \times 10^4$ samples, and (e) (vOED-NFs) EIG lower bound $\widehat{U}_L$ using NFs optimized with $N_{\text{opt}} = 2 \times 10^4$ samples. Furthermore, the lower bound estimate $\widehat{U}_L$ at each d is evaluated using $N = 1 \times 10^4$ samples. For (d) and (e), the total number of forward model evaluation is therefore $2 \times 10^4 + 1 \times 10^4 = 3 \times 10^4$. vOED-NFs adopts $T = 5$ sets of complete transformations of the cINN described in Sec. 2.2.1. Additional training details and hyperparameter choices for (d) and (e) can be found in Appendix A.3.1.

In Fig. 2.2a, Hi-NMC is designated as the reference solution. In comparison to this reference, Lo-NMC exhibits moderate bias while vOED-NFs performs extremely well. Re-NMC yields near-identical results as Hi-NMC; while faster, it still requires an explicit likelihood and does not provide approximate posterior distributions, just like the other NMC methods. vOED-G, due to its much simplistic Gaussian variational distributions, deviates significantly from the expected utility trend and misidentifies the optimal design. However, the training cost for vOED-NFs is also higher than vOED-G, as illustrated through the training convergence history for $\xi = 1.0$ in Fig. 2.2b. Here vOED-G shows a very rapid convergence to ϕ^*, while vOED-NFs takes longer to stabilize and experiences greater fluctuation. This timing difference induced by optimization is negligible when

the overall computations are dominated by the forward model runs, but can become significant when the forward model is inexpensive. Additional information regarding the time required for EIG estimation is given in Appendix A.7.



(a) EIG estimates with different estimators

(b) Training convergence at $\xi = 1$

Figure 2.2: Case 1. EIG estimates and sample training convergence plot.

To further analyze the ability of vOED-NFs in approximating the true posteriors, we compare the resulting variational posteriors to a high-quality reference posterior obtained using sequential Monte Carlo (SMC) [42] at two designs: $\xi = 0.2$ in Fig. 2.3 and $\xi = 1.0$ in Fig. 2.4. Each column plots the marginal posteriors of the three parameters $q(\theta_1|y, \xi; \phi^*)$, $q(\theta_2|y, \xi; \phi^*)$, $q(\theta_3|y, \xi; \phi^*)$, while each row corresponds to a different sample of observed y. The vOED-NFs posteriors approximate the SMC reference posteriors very well, even with multi-modal posteriors as seen in Fig. 2.4 for θ_3.

Figure 2.5 investigates the impact of the number of transformations and optimization sample size N_{opt} in vOED-NFs. The plots show the EIG lower bound estimates $\widehat{U}_L$ when adopting $T = \{1, 3, 5, 7\}$ sets of complete cINN transformations, and using $N_{\text{opt}} = \{5 \times 10^3, 10^4, 5 \times 10^4\}$. The EIG lower bound estimates are still evaluated with $N = 10^4$ samples. In general, employing 3 transformations is adequate to identify the optimal design $\xi^* = 1$. However, when N_{opt} is small, adding more transformations does not provide significant improvement. When N_{opt} is increased, the value from adding transformations is more notable and stabilized around $T = 5$; this is consistent with previous work [124] that reported composing more than 5 transformations provides only slight additional benefits. This diminishing return is likely due to the much larger number of NFs parameters when a high number of transformations is used, making them more prone to overfitting. Additional testing and discussions for the $N_{\text{opt}} = 5 \times 10^3$ case can be found in Appendix A.3.1.

2.3.2 Case 2: High-dimensional linear design

This next example showcases the joint optimization of ξ and ϕ and in a higher dimensional setting. The setup follows the example in [51], with an observation model:

$$y_j = \xi_j^\top \mathbf{w} + \epsilon_j, \qquad j = 1, \ldots, n, \tag{2.34}$$

where $y_j \in \mathbb{R}$ is the jth observation, $\xi_j \in \mathbb{R}^p$ is its corresponding design vector, $\epsilon_j \sim \mathcal{N}(0, \sigma^2)$ is a Gaussian noise random variable, and $\theta = \{\mathbf{w}, \sigma\}$ encompasses the p-dimensional regression coefficient vector $\mathbf{w}$ and noise standard deviation σ. We set $p = 20$ and $n = 20$, hence the parameter dimension is 20 and the total design dimension is 400. Independent priors are adopted for $\mathbf{w}_j \sim \text{Laplace}(0, 1)$ and $\sigma \sim \text{Exp}(1)$. Lastly, a design constraint $||\xi_j||_1 = 1$ is imposed to reflect a resource budget; this is implemented through re-normalization of ξ after each SGA update in order to map it back to the feasible region. During the joint optimization of ξ and ϕ, the total number of forward model runs (i.e., the total number of $(\theta^{(i)}, y^{(i)})$ pairs across all design epochs) is fixed to 10^6. This is the same computational budget used by [51] that adopted Gaussian and Gamma variational distributions for $\mathbf{w}$ and σ, respectively; we refer to the setup in [51] as vOED-GG. Further details on the computational setup can be found in A.4.1.

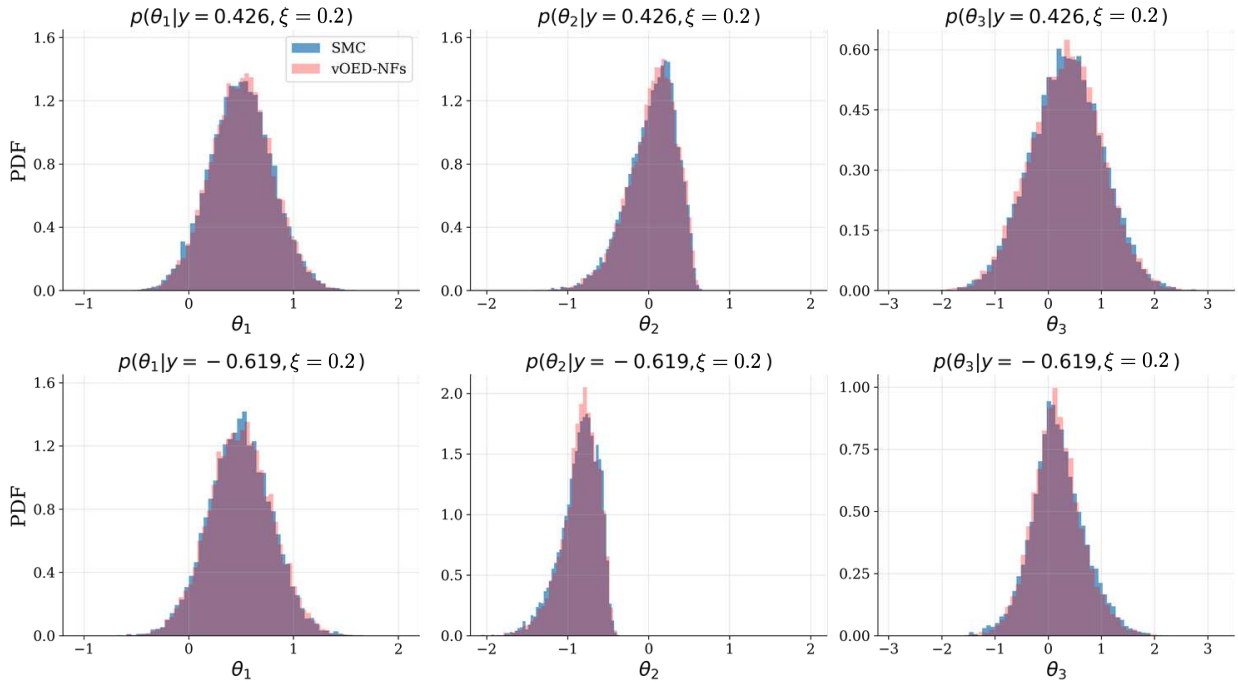


Figure 2.3: Case 1. Comparison of marginal posteriors obtained from SMC and vOED-NFs at $\xi = 0.2$.

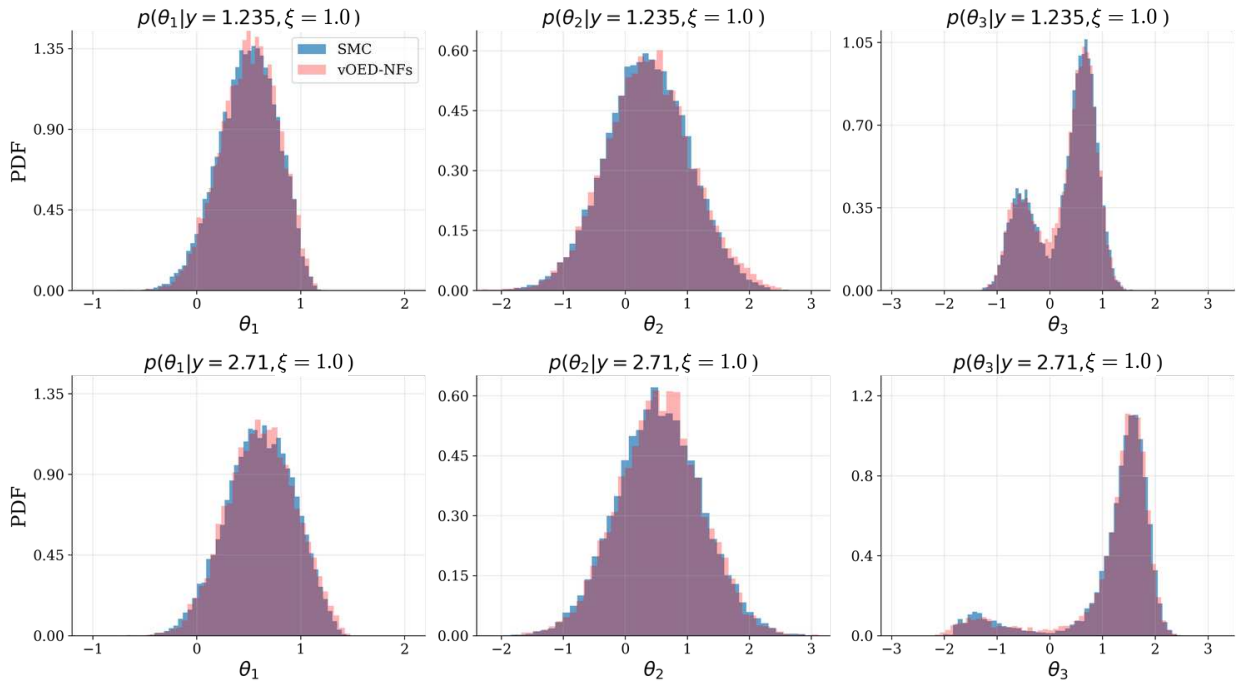


Figure 2.4: Case 1. Comparison of marginal posteriors obtained from SMC and vOED-NFs at $\xi = 1.0$.

Optimization results are presented in Table 2.1, with the middle column showing the mean ± one standard deviation of 10 lower bound estimates $\widehat{U}_L$ (each evaluated with $N = 10^4$ samples) at the optimal design found by vOED-GG and vOED-NFs, and the right column showing the mean ± one standard deviation of 10 high-quality NMC estimates (each using $N_{\text{out}} = 10^4$ and $N_{\text{in}} = 10^4$) at each method's optimal design to provide an accurate EIG estimate and using a common estimation method. We observe that vOED-NFs's optimal design achieves a tighter lower bound estimate and also a higher NMC estimate, indicating that a better design has been found than vOED-GG. However, the vOED-NFs lower bound estimate is not tight, and the remaining discrepancy between $\widehat{U}_L$ and $\widehat{U}_{\text{NMC}}$ may be contributed by a number of factors: (1) the functional form of cINN; (2) the number of cINN transformations; (3) the functional representation of s's and t's; (4) the possibility to converge to a local optimum during optimization; and (5) the finite sample sizes.

In order to compare the posterior approximation capabilities of the two approaches, we look at their approximate posteriors at a common design with each $\xi_j = \mathbf{e}_j$ set to the standard unit vector in the jth dimension, and at three y samples. Figures 2.6 and 2.7 present the marginal posteriors for $\{w_1, w_2, w_3, w_4, \sigma\}$ using No-U-Turn Hamiltonian Monte Carlo (NUTS-HMC) [73] as the reference posterior distribution, along with the approximations obtained from vOED-GG and vOED-NFs. From the figure, it is evident that the vOED-NFs posteriors can capture the non-Gaussian structures much better than vOED-GG's Gaussian/Gamma distributions, which is

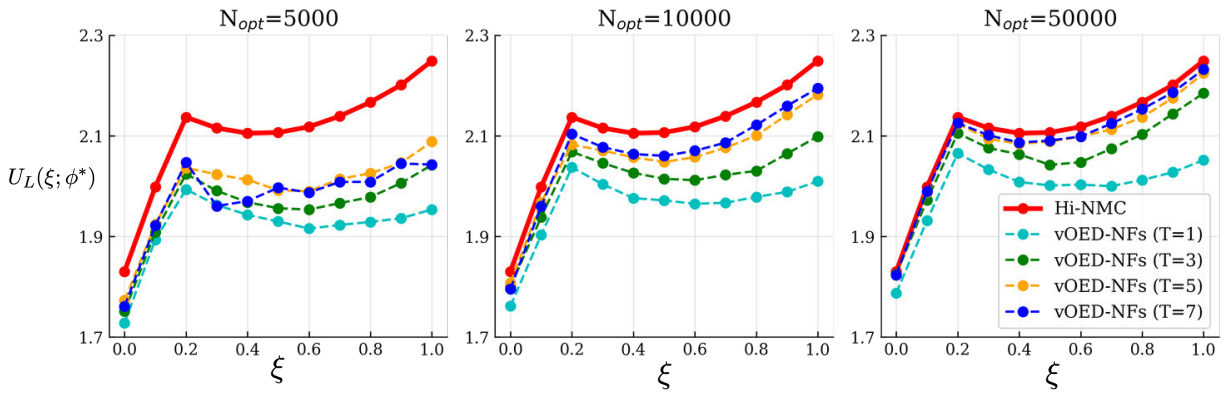


Figure 2.5: Case 1. EIG under different number of transformations T and optimization sample sizes N_{opt}.

Method	$\widehat{U}_L(\xi^*;\phi^*)$	$\widehat{U}_{\text{NMC}}(\xi^*)$
vOED-GG	12.07 (0.11)	24.28 (0.24)
vOED-NFs	20.94 (0.19)	24.91 (0.24)

Table 2.1: Case 2. Middle column shows the mean of 10 lower bound estimates $\widehat{U}_L$ (each is evaluated with $N = 10000$ samples) at the optimal design found by vOED-GG and vOED-NFs, and the right column shows the mean of 10 high-quality NMC estimates (each with $N_{\text{out}} = 10000$ and $N_{\text{in}} = 10000$) at the optimal designs. The $(\cdot)$ values are one standard deviation estimates for the estimators.

confirmed by the empirical KL divergence from the NFs posteriors to MCMC posteriors.

2.3.3 Case 3: Cathodic electrophoretic deposition (e-coating)

Cathodic electrophoretic deposition, commonly known as e-coating, is a technique for applying protective coatings to automobile surfaces. During the procedure, an anode and a cathode (the car body) are placed in a colloidal bath in which the electric current triggers an electrochemical reaction resulting in the deposition of material on the car body. A proper coating of the car body is achieved when a sufficient thickness of coating is reached. Here we apply vOED-NFs to design e-coating experiments, and employ both the lower and upper bound setups respectively in Eqn. (2.6) and (2.14).

The physical model adopted for the e-coating process, consisting of a PDE, is the baseline model from [81] and also summarized in A.5.1. The parameters of interest, $\theta = \{C_v, j_{\min}, Q_{\min}\}$, are the volumetric coulombic efficiency (m^3/C), minimum current (A/m^2), and minimum charge (C/m^2), respectively. Each parameter is endowed with a truncated normal prior, with detailed prior parameters provided in Appendix A.5.1. A measurement on the current (mA) can be obtained at

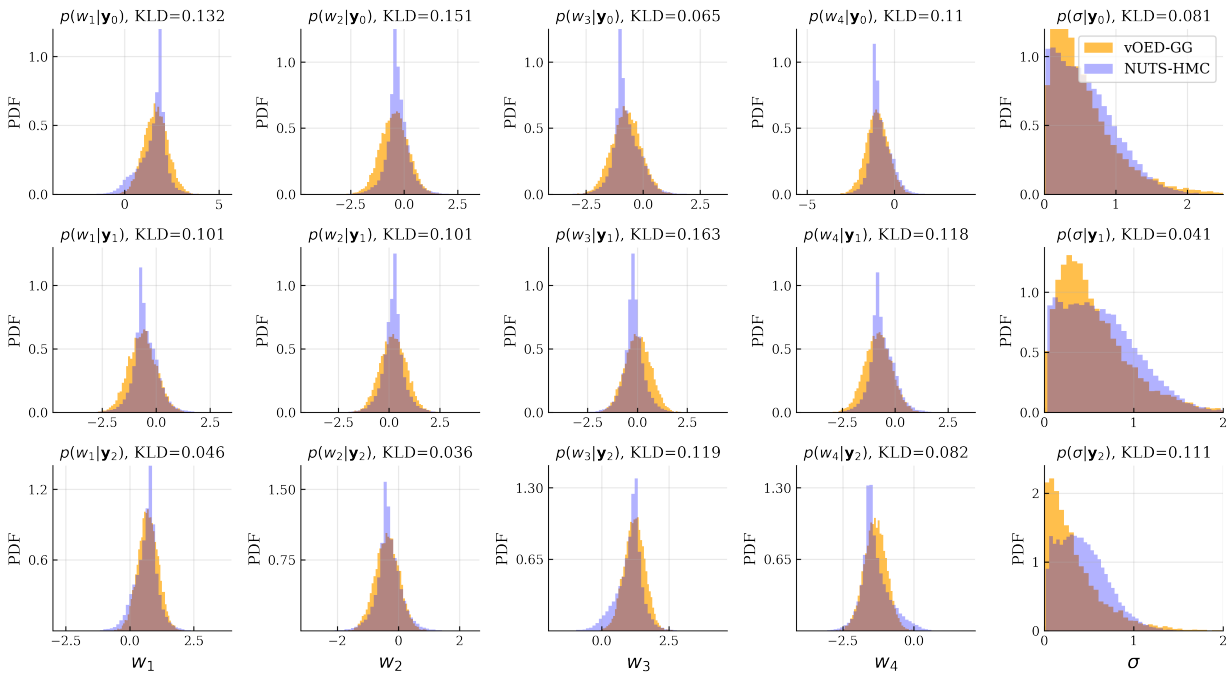


Figure 2.6: Case 2. Comparison of marginal posteriors obtained from NUTS-HMC and vOED-GG [50] at design where $\xi_j = \mathbf{e}_j$, with KLD denotes the empirical KL divergence from NFs posteriors to MCMC posteriors.

time t (s), modeled as

$$y_t = j(\theta, j_0, t)(1 + \epsilon), \tag{2.35}$$

where $j(\theta, j_0, t)$ is the forward model mapping from parameters to observables at time t and under constant current boundary condition j_0 (mA), and $\epsilon \sim \mathcal{N}(0, 0.1^2)$ represents a relative measurement noise. The design variable is j_0, and three candidate designs are considered: (1) $j_0 = 10$ mA with measurements made at $t = \{10, 20, \ldots, 100\}$ s; (2) $j_0 = 7.5$ mA with measurements made at $t = \{20, 40, \ldots, 200\}$ s; and (3) $j_0 = 5.0$ mA with measurements made at $t = \{30, 60, \ldots, 300\}$ s. Hence, all designs always involve $n_y = 10$ measurements at regular intervals; the lower j_0 designs have slower decays of the current (e.g., see Figure D.14 in [81]) and thus longer intervals are adopted.

Figure 2.8 presents the convergence of the upper and lower bound estimates from Eqn. (2.19) and (2.20), along with a reference EIG value, when optimizing over ϕ at design $j_0 = 5.0$ mA. The reference EIG is obtained from a high-quality NMC with the 'reuse' technique described in [77] in order to accommodate the relatively high computational cost for each forward model evaluation (i.e., a PDE solve). Thus, even with $N_{\text{out}} = N_{\text{in}} = 10^5$, the reuse technique only requires

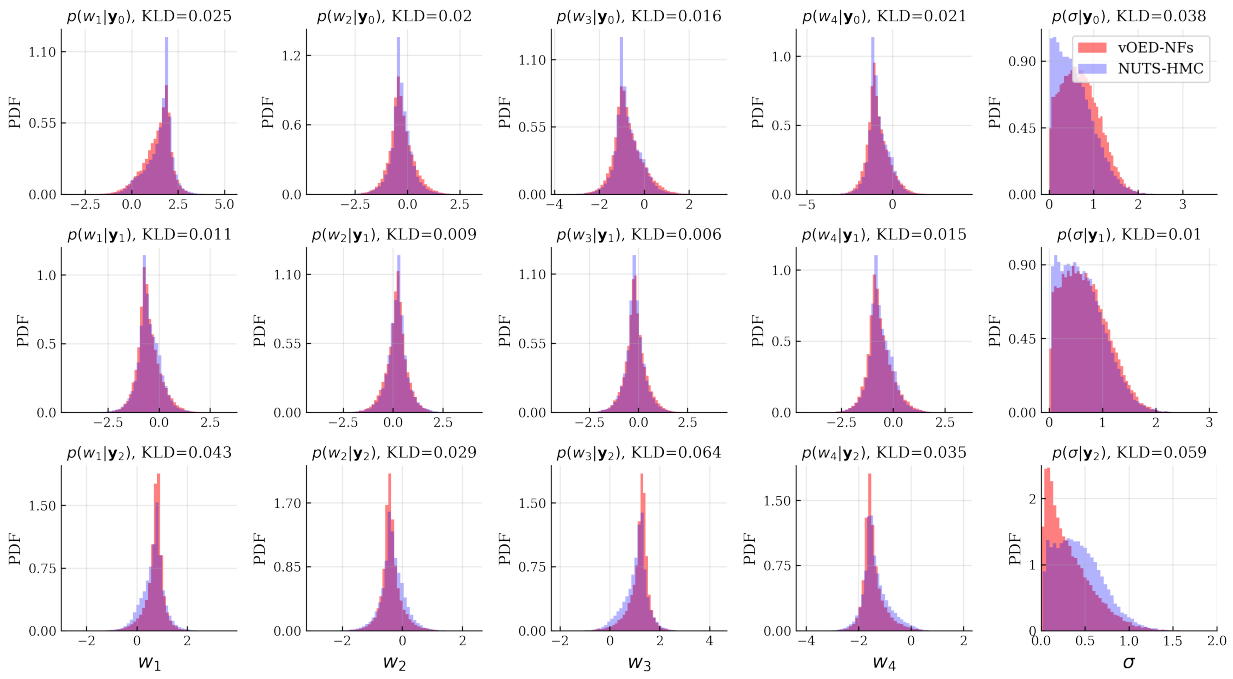


Figure 2.7: Case 2. Comparison of marginal posteriors obtained from NUTS-HMC and vOED-NFs [50] at design where $\xi_j = \mathbf{e}_j$, with KLD denotes the empirical KL divergence from NFs posteriors to MCMC posteriors.

10^5 total forward model evaluations (instead of 10^{10}, which would be impractical). We further provide comparison of the bounds between low-quality optimization that uses $N_{\text{opt}} = 10^4$ versus high-quality estimation that uses $N_{\text{opt}} = 8\times10^4$ (the evaluation of the bounds are always performed with $N = 10^4$). Results indicate that increasing N_{opt} tightens both bounds. Under the same N_{opt}, the upper bound appears to be tighter than the lower bound. To illustrate the ability of $q(y;\phi)$ in representing the marginal likelihood $p(y|\xi)$, their marginal distributions are shown in Fig. 2.9, where we see excellent agreement between the approximation and the true distributions obtained from direct Monte Carlo.

In scenarios where the dimension of $\boldsymbol{y}$ is high, the input size to the s and t networks of the cINN becomes large. To combat this high dimensionality, summary networks are employed to convert y to a lower-dimensional y', specifically through a long short-term memory (LSTM) network [58] as recommended in [124] to accommodate the sequential nature of y. Figure 2.10 presents the same convergence of bound estimates but now with $n_y = 50$ measurements taken in each experiment, i.e., $\boldsymbol{y} = \{y_{t_1}, \ldots, y_{t_{50}}\}$, and with a summary network that compresses it to a 10-dimensional y'. The figure indicates that the lower bound estimates are now tighter than their upper bound counterparts. Together with Fig. 2.8, these observations suggest that when the dimension of y significantly exceeds that of θ, EIG may be more accurately estimated though the lower bound estimate that approximates

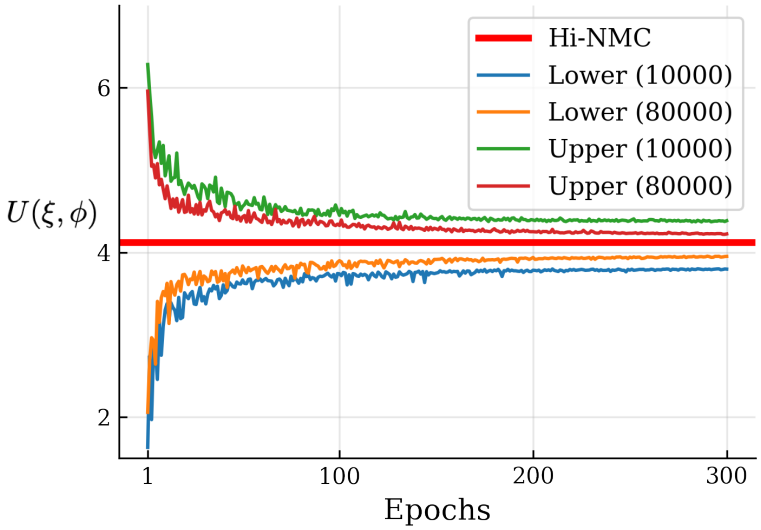


Figure 2.8: Case 3. Convergence history of the lower and upper bounds at $d = j_0 = 5$ mA and 10 observations when optimizing ϕ using different N_{opt}. Hi-NMC is the high-quality reference EIG estimate. Here the upper bound estimates appear to be tighter than the lower bound estimates.

the posterior, than the upper bound estimate that approximates the marginal likelihood.

Various lower bound estimates, using vOED-G, vOED-NFs, vOED-NFs-LSTM, and the high-quality NMC, are computed for all three candidate designs and shown in Fig. 2.11. All results indicate $j_0 = 0.75$ mA achieving the best design, although vOED-G incurs a notable error in estimating the EIG compared to the rest. The vOED-NFs lower bound estimate is further tightened when the LSTM summary network is used. The vOED-NFs-LSTM using different N_{opt} also further supports the benefit when optimizing using a larger sample size.

Figure 2.12 presents the marginal and pairwise posterior distributions obtained from using

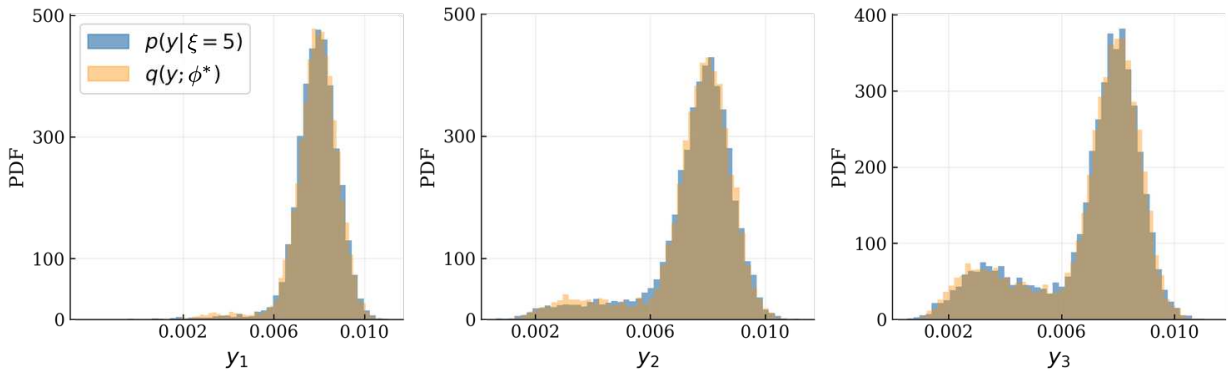


Figure 2.9: Case 3. Marginal distributions $p(y_1|\xi), p(y_2|\xi), p(y_3|\xi)$ at $\xi = j_0 = 5$ mA obtained using MC sampling and vOED-NFs with $N_{\text{opt}} = 10000$.

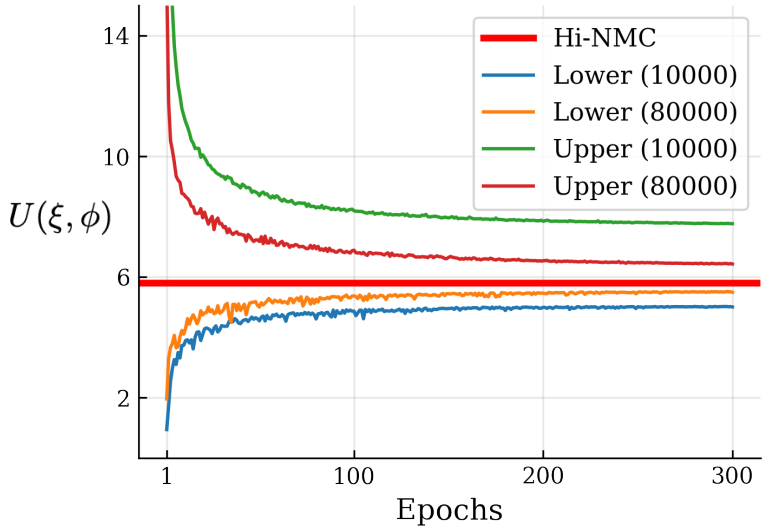


Figure 2.10: Case 3. Convergence history of the lower and upper bounds at $\xi = j_0 = 5$ mA and 50 observations when optimizing ϕ using different N_{opt}. Hi-NMC is the high-quality reference EIG estimate. Here the lower bound estimates appear to be tighter than the upper bound estimates.

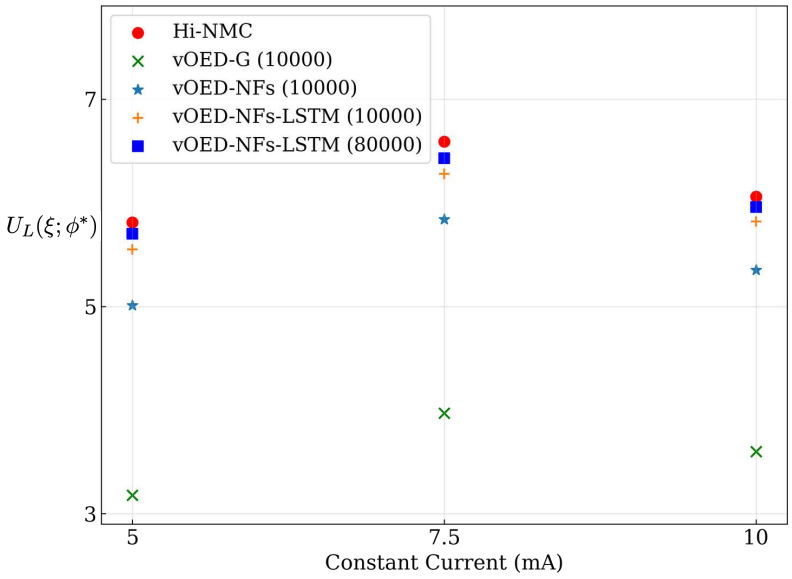


Figure 2.11: Case 3. Lower bound estimates for all designs using different methods. Hi-NMC is the high-quality reference EIG estimate.

vOED-NFs and vOED-NFs-LSTM lower bound estimates, along with the reference posteriors obtained from SMC. We observe that the posteriors obtained from vOED-NFs-LSTM in Fig. 2.12b offering improved approximation to the SMC reference than those from vOED-NFs in Fig. 2.12a. This improvement supports the effectiveness of employing the LSTM summary network for dimension reduction in this example.

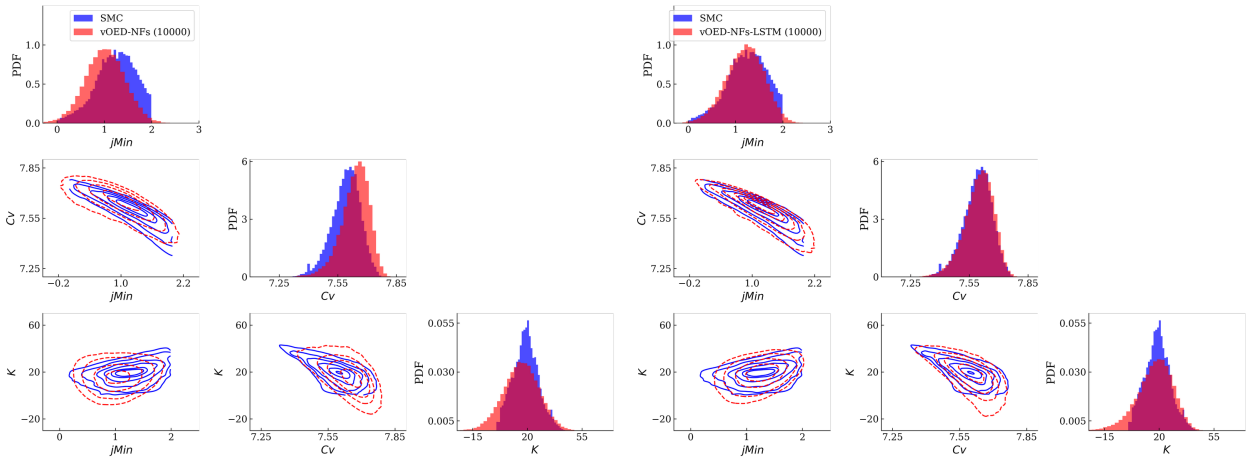
2.3.4 Case 4: aphid population

The last case involves a stochastic model depicting the growth of aphid population [102], which results in an implicit likelihood scenario where samples may be drawn from the likelihood but probability density cannot be evaluated. Denote $M(t)$ and $C(t)$ respectively the current and accumulative sizes of the aphid population, the probability that a birth or death occurs in a small time period δ_t is [60]:

$$\mathbb{P}\{M(t+\delta_t) = m+1, C(t+\delta_t) = c+1 \mid M(t) = m, C(t) = c\} = \alpha m \delta_t + o(\delta_t), \tag{2.36}$$

$$\mathbb{P}\{M(t+\delta_t) = m-1, C(t+\delta_t) = c \mid M(t) = m, C(t) = c\} = \beta m c \delta_t + o(\delta_t), \tag{2.37}$$

where $\theta = \{\alpha, \beta\}$ are the birth and death model parameters, respectively. In practice a non-zero δ_t must be used, leading to a discretized approximation to the continuum limit. We adopt a prior



(a) vOED-NFs (b) vOED-NFs-LSTM

Figure 2.12: Case 3. Marginal and pairwise joint posterior distributions obtained using SMC (high-quality reference) and vOED-NFs-LSTM.

distribution following [60]:

$$\begin{pmatrix}\alpha\\ \beta\end{pmatrix} \sim \mathcal{N}[\begin{pmatrix}0.246\\ 0.000136\end{pmatrix}, \begin{pmatrix}0.0079^2 & 5.8\times10^{-8}\\ 5.8\times10^{-8} & 0.00002^2\end{pmatrix}], \tag{2.38}$$

and an initial condition of $M(0) = C(0) = 28$. In each experiment, only $M(t)$ is observed at a particular observation time t and $C(t)$ is not observed. An exact likelihood evaluation of $\mathbb{P}(M(t) = m \mid \theta)$ thus requires the summation of probabilities along all paths that end at $M(t) = m$ and then marginalizing out $C(t)$; this is generally intractable to compute.

The OED problem is then to determine the optimal k measurement times, $\xi = \{t_1, t_2, ..., t_k\}$, $t_k \in [0, 50]$. Following [5], we solve the design optimization for $k = 1$ and 2 (i.e., designing for respectively 1 and 2 measurement times) through grid search, while for $k = 3$ and 4 we use the simultaneous perturbation stochastic approximation (SPSA) algorithm [138]. Table 2.2 presents the design results obtained using vOED-NFs and the LB-KLD method that was used in [5]. LB-KLD constructs another lower bound to EIG based on the entropy power inequality and entropy's concavity property, and adopts the nearest neighbor based entropy estimator within the lower bound estimator. Note that LB-KLD does not involve tightening the lower bound and therefore is not a variational lower bound, but instead directly estimates the lower bound, here using 3×10^4 samples. For vOED-NFs, we set $N_{\text{opt}} = 2\times10^4$ and $N = 10^4$ such that the total number of samples is the same. Implementation details can be found in Appendix A.6.1.

k	Method	ξ^*	$\widehat{U}_{\text{LB-KLD}}(\xi^*)$	$\widehat{U}_L(\xi^*;\phi^*)$
1	LB-KLD	(21)	1.18 (0.012)	1.22 (0.003)
	vOED-NFs	(21)	1.18 (0.012)	1.22 (0.003)
2	LB-KLD	(17, 28)	1.86 (0.021)	1.86 (0.018)
	vOED-NFs	(17, 27)	1.88 (0.008)	1.89 (0.004)
3	LB-KLD	(15.7, 22.7, 32.0)	2.00 (0.019)	2.08 (0.004)
	vOED-NFs	(14.6, 20.7, 28.6)	2.00 (0.018)	2.10 (0.005)
4	LB-KLD	(13.8, 19.1, 24.5, 30.6)	2.05 (0.017)	2.20 (0.006)
	vOED-NFs	(13.7, 19.2, 24.8, 32.7)	2.06 (0.014)	2.19 (0.006)

Table 2.2: Case 4. The third column shows the optimal design ξ^* found by LB-KLD and vOED-NFs. The fourth column shows the mean of 10 $\widehat{U}_{\text{LB-KLD}}$ estimates with different random seeds (each with 3×10^4 samples for Monte Carlo estimation) at the optimal designs, and $(\cdot)$ indicates the standard deviation originates from different random partitions of the same 30000 samples used for entropy estimation. The last column shows the mean of 10 $\widehat{U}_L$ estimates (each with $N_{\text{opt}} = 2\times10^4$ and $N = 10^4$) at the optimal designs but different ϕ^*. Here $(\cdot)$ indicates the standard deviation due to 10 random seeds for different initialization when optimizing ϕ^*, and evaluation is conducted on the same $N = 10^4$ samples.

Table 2.2 indicates that both methods generally propose comparable optimal designs for various k values and similar lower bound estimates. Notably at $k = 3$ and 4, vOED-NFs produce tighter bound estimators than LB-KLD. Samples from the approximate posterior in vOED-NFs for a simulated y at $k = 4$ are illustrated in Fig. 2.13, which agree well with reference posterior samples obtained from approximate Bayesian computation (ABC) [100]. Note that posterior information is only available owing to vOED-NFs's use of q to approximate the posterior density; in contrast, other lower bound approaches not based on posterior approximation, for example LB-KLD, do not offer any posterior information.

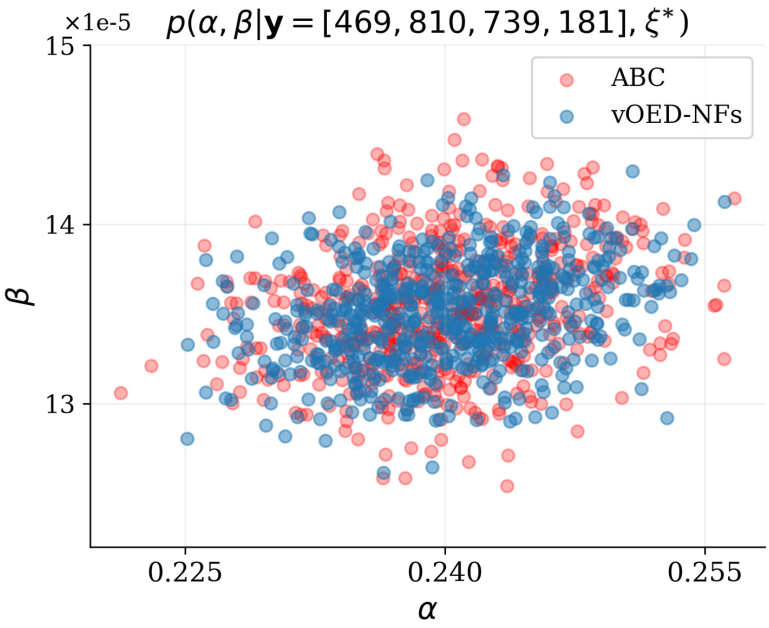


Figure 2.13: Case 4. Comparison of joint posteriors samples from ABC and vOED-NFs at the optimal design.

2.4 Conclusions

This work introduced vOED-NFs, a method to use normalizing flows to represent variational distributions in the context of Bayesian OED. When the expected utility of OED is chosen to be the EIG in the model parameters, the Barber–Agakov lower bound may be used to estimate the EIG and the bound tightened by optimizing its variational parameters. We presented Monte Carlo estimators to both lower and upper bound versions along with their gradient expressions. We then detailed the use of NFs, in particular with cINN architecture involving a composition of coupling layers and together with a summary network for dimension reduction, to approximate the posterior

or marginal likelihood distributions.

We validated vOED-NFs against established methods in two benchmark problems, and demonstrated vOED-NFs on a PDE-governed application of cathodic electrophoretic deposition and stochastic modeling of aphid population with an implicit likelihood. The findings suggested that 4–5 compositions of the coupling layers were adequate to achieve a lower bias compared to previous approaches. Furthermore, we illustrated that vOED-NFs produced approximate posteriors that matched very well with the true posteriors, able to capture non-Gaussian and multi-modal features effectively.

A limitation of vOED-NFs is the lack of analytical results connecting bound estimator quality to design optimization. For example, even if the lower bounds are not tight but the bound gap is consistent across $\mathcal{A}$, the design maximizer may still be similar to the true optimal design. Another limitation is that, to retain good accuracy under high-dimensional θ, complex NFs architectures with high-dimensional ϕ may be needed, making the optimization problem challenging. An associated potential difficulty arises from the general susceptibility of neural network training to local minima. Furthermore, although we have illustrated greater sampling efficiency of vOED-NFs compared to NMC, it still require a non-trivial number of forward model runs to tighten the lower bound and achieve high-quality posteriors; further improving sampling efficiency will be important. Future research exploring efficient and scalable transport map architectures in vOED, such as with rectification operator [10] and probability flow ODEs [137], will be useful. The adoption of vOED-NFs into other OED structures, such as goal-oriented OED [166] and sequential OED [134], will also be important.

More broadly, OED approaches that rely on EIG bounds can only be used when the expected utility is the EIG (possibly with other terms that do not depend on the posterior) and cannot be applied to other choices of design criteria. Moreover, the bounds (i.e., Eqn. (2.5), Eqn. (2.13)) cannot be computed in closed form, but only estimated through, for example, Monte Carlo sampling (i.e., Eqn. (2.19) and (2.20)). Consequently, the bound estimators can no longer provide inequality guarantees due to the sampling variance. Thus, as a candidate for future exploration, there is a great need for sampling efficiency for bound-based strategies, to keep the bound estimator variance lower than the bias (i.e., the bound gap).

CHAPTER 3

Variational Sequential Bayesian Optimal Experimental Design

In many scientific and engineering applications, experiments are conducted sequentially, with data collected at earlier stages informing the design of future experiments. Sequential experimental design provides a principled framework for exploiting this adaptivity, enabling more efficient learning under limited experimental budgets. By updating beliefs as observations are acquired, sequential strategies can substantially outperform one-shot designs, particularly in settings involving complex models, latent variables, or high-dimensional uncertainties.

Research gap and contributions. While prior work on sequential optimal experimental design (sOED) has established important foundations, existing approaches typically address only subsets of the challenges posed by realistic adaptive experimentation. In particular, there is a lack of a unified framework that simultaneously supports (i) goal-oriented objectives targeting user-specified quantities of interest rather than global model uncertainty, (ii) non-myopic sequential decision-making over entire experimental trajectories, and (iii) computational tractability for expensive or implicit likelihood models with flexible posterior approximations. The central research question addressed in this chapter is how to construct a scalable sequential experimental design framework that integrates these desiderata within a single principled formulation. To this end, this chapter introduces variational sequential optimal experimental design (vsOED), which formulates sOED as a finite-horizon Markov decision process and solves it using an actor–critic policy gradient approach combined with variational EIG bounds. The key contributions of this chapter include (i) a unified formulation of goal-oriented and sequential experimental design under a variational framework, (ii) numerical methods for learning non-myopic design policies using amortized inference and expressive posterior approximations, and (iii) empirical validation demonstrating improved performance under limited experimental budgets across a range of representative problem settings.

The contents of this chapter are published in [133]. The author gratefully acknowledges Wanggang Shen for his substantial contributions to the development of this work.

3.1 Problem formulation

3.1.1 Preliminaries

We focus on the design of a finite sequence of T experiments indexed by $t \in \{0, \dots, T-1\}$. We assume T is known and fixed. Each experiment is associated with a real-valued design vector $\xi_t \in \mathcal{A}_t \subseteq \mathbb{R}^{n_\xi}$ and observation vector $Y_t \in \mathcal{Y}_t \subseteq \mathbb{R}^{n_y}$. For simplicity, we assume n_ξ and n_y to be constant across t. At the tth experiment, the information sequence of all completed experiments' designs and observations is denoted by $I_t = \{\xi_0, Y_0, \dots, \xi_{t-1}, Y_{t-1}\}$, with $I_0 = \emptyset$.

Furthermore, let $\mathcal{M}_m$ be a countable set of models indexed by $m \in \{1, 2, \dots\}$ where each model has its own parameters of interest (PoIs) $\Theta_m \in \boldsymbol{\Theta}_m \subseteq \mathbb{R}^{n_{\Theta_m}}$, nuisance parameters $\Psi_m \in \boldsymbol{\Psi}_m \subseteq \mathbb{R}^{n_{\Psi_m}}$, and predictive quantities of interest (QoIs) $Z_m \in \mathcal{Z}_m \subseteq \mathbb{R}^{N_{Z_m}}$. The experiment observation is related to the parameters via an observation model, for example often in the form

$$Y_t = G_t(\Theta_m, \Psi_m, \xi_t; m, I_t) + \mathcal{E}_t, \tag{3.1}$$

where G_t is a nonlinear observation forward operator and ϵ_t is the observation noise. The QoIs are related to the parameters via a predictive model

$$Z_m = H(\Theta_m, \Psi_m, \mathcal{E}_Z; m), \tag{3.2}$$

where H is a nonlinear predictive stochastic forward operator with ϵ_Z representing the predictive stochastic source. If the predictive model is deterministic, we can write $Z_m = H(\Theta_m, \Psi_m; m)$. G_t and H often involve the most expensive computations (e.g., solving various differential equations), hence the number of these forward solves is often used as the unit for computational cost.

Under a Bayesian approach, after the tth experiment, the joint probability density on M, θ_m, and ψ_m updates following Bayes' rule:

$$p(m, \theta_m, \psi_m | I_{t+1}) = p(m, \theta_m, \psi_m | \xi_t, Y_t, I_t) = \frac{p(Y_t | m, \theta_m, \psi_m, \xi_t, I_t)\, p(m, \theta_m, \psi_m | I_t)}{p(Y_t | \xi_t, I_t)}, \tag{3.3}$$

where $p(m, \theta_m, \psi_m | \xi_t, Y_t, I_t)$ is the posterior which can also be written as $p(m, \theta_m, \psi_m | I_{t+1})$, $p(m, \theta_m, \psi_m | I_t)$ is the prior, $p(Y_t | m, \theta_m, \psi_m, \xi_t, I_t)$ is the likelihood, and $p(Y_t | \xi_t, I_t)$ is the marginal likelihood. In order to simplify notation in the remainder of this chapter, we adopt the convention where when m is not explicitly mentioned, conditioning on m is implied through other variables'

subscripts, e.g., $p(\theta_m, \psi_m|I_t) = p(\theta_m, \psi_m|m, I_t)$.

The posterior after the t-th experiment $p(m, \theta_m, \psi_m|I_{t+1})$ then serves as the prior for the $(t+1)$th experiment and is again applied to Eqn. (3.3). Bayes' rule can be recursively applied for sequential experiments.

Upon propagating the parameter distributions through H, the prior- and posterior-predictive densities for z_m can be respectively written as:

$$p(z_m) = \iint p(\theta_m, \psi_m)\, p(z_m|\theta_m, \psi_m)\, \mathrm{d}\theta_m\, \mathrm{d}\psi_m, \tag{3.4}$$

$$p(z_m|I_{t+1}) = \iint p(\theta_m, \psi_m|I_{t+1})\, p(z_m|\theta_m, \psi_m)\, \mathrm{d}\theta_m\, \mathrm{d}\psi_m. \tag{3.5}$$

If H is deterministic, $p(z_m|\theta_m, \psi_m)$ collapses to a Dirac delta centered at $z_m = H(\theta_m, \psi_m; m)$.

3.1.2 Sequential optimal experimental design (sOED)

We present sOED following [134] and earlier works [74, 78] that modeled it through a Markov decision process (MDP). The MDP is defined by a tuple $\left(\mathcal{S}, \{\mathcal{A}_t\}_t, s_0, \{r_t(\cdot)\}_t, \{K_t(\cdot)\}_t\right)$ consisting of a state space $\mathcal{S}$ for the state variable S_t that can take values $S_t \in \mathcal{S}$, action spaces $\mathcal{A}_t$ comprising possible actions (which here are designs) $\xi_t \in \mathcal{A}_t$, an initial state s_0, scalar-valued reward functions $r_t(S_t, \xi_t, Y_t)$ that evaluate the instantaneous reward when taking action ξ_t and observing Y_t at state S_t, and state transition kernels $K_t(\mathcal{S}_{t+1}|S_t, \xi_t)$ that evaluate the probability of transitioning to any set of states $\mathcal{S}_{t+1} \subseteq \mathcal{S}$ at stage $t+1$ having taken action ξ_t at state S_t. In the context of experimental design, the action being taken is the selection of a design; thus we use the terms 'action' and 'design' interchangeably.

State The state before the tth experiment is described by $S_t = \{S_t^b, S_t^p\}$, a quantity that summarizes all information deemed relevant to future design decisions. We split S_t into a 'belief state' S_t^b representing the state of knowledge/uncertainty in M, θ_m, ψ_m, and z_m, and a 'physical state' S_t^p comprising any other deterministic variables that may be relevant to the design process. A realization of the belief state, S_t^b, is simply the joint posterior distribution of M, θ_m, ψ_m, and z_m given all past experimental designs and realized observations, I_t. Numerically, it can be represented by, for example, a density or distribution function approximation, or an ensemble of particles, or by tracking I_t directly. Tracking I_t is easiest to implement since it does not require additional calculations translating $(\xi_i)_{i<t}$ and $(y_i)_{i<t}$ to another representation, but the dimension of I_t grows with t albiet bounded for finite T.

Maintaining only a belief state, in the form of the posterior, does not suffice to preserve the Markov property of the system if the likelihood depends on the history of past experiments I_t, as in

Eqn. (3.3). This can be fixed by introducing a physical state. With regard to the I_t-dependence in $p(Y_t|m, \theta, \psi_t, \xi_t, I_t)$, the physical state essentially extracts and tracks relevant features from I_t that allow the likelihood to be evaluated or the observations Y_t simulated. If I_t is adopted as the belief state, then information about the physical state is already contained in I_t, even if only implicitly. We adopt $S_t = I_t$ in this work and will use them interchangeably.

Action (design) and policy Sequential experimental design is adaptive in nature, and looks for a strategy, called a policy (or actor), describing how to choose the design depending on the current state. The policy is a collection of functions $\pi = \{\mu_t : \mathcal{S} \to \mathcal{A}_t, t = 0, \ldots, T-1\}$, where the policy function μ_t returns the design for the tth experiment given the current state, $\xi_t = \mu_t(S_t)$.

State transition dynamics When an experiment is performed, the state changes according to a transition kernel $T_t(\mathcal{S}_{t+1}|S_t, \xi_t)$ describing the probability of transitioning from the current state S_t, having chosen design ξ_t and observed the outcome of resulting experiment, to any set of states at stage $t+1$, $\mathcal{S}_{t+1} \subseteq \mathcal{S}$. This kernel is generally intractable to evaluate, but can be simulated by first sampling Y_t given the design ξ_t, and then applying Bayes' rule Eqn. (3.3). We denote the latter transition dynamics by $s_{t+1} = \mathcal{F}_t(S_t, \xi_t, Y_t)$; the function $\mathcal{F}_t$ encapsulates the transition from prior to posterior, given values of ξ_t and the realized data Y_t, following Eqn. (3.3). If adopting I_t as the state, then the transition is a simple concatenation $I_{t+1} = \{I_t, \xi_t, Y_t\}$.

Reward (utility) Here $r_t(S_t, \xi_t, Y_t) \in \mathbb{R}$ denotes the reward immediately obtained from the t-th experiment, and $r_T(s_T) \in \mathbb{R}$ is the terminal reward that can be computed only after all experiments are completed. Examples of information-theoretic rewards will be provided in Sec. 3.1.3 and 3.1.4.

sOED problem statement The sOED problem seeks a design policy that maximizes the expected utility $U(\pi)$:

$$\begin{aligned} \pi^* \in \underset{\pi=\{\mu_0,\ldots,\mu_{T-1}\}}{\operatorname{argmax}} \quad & \left\{ U(\pi) := \mathbb{E}_{y_{0:T-1}|\pi,s_0}\left[\sum_{t=0}^{T-1} r_t(S_t, \xi_t, Y_t) + r_T(S_T)\right]\right\} \\ \text{subject to} \quad & \xi_t = \mu_t(S_t) \in \mathcal{A}_t, \\ & S_{t+1} = \mathcal{F}_t(S_t, \xi_t, Y_t), \quad \text{for } t = 0, \ldots, T-1. \end{aligned} \tag{3.6}$$

Here $y_{0:T-1} = y_0, y_1, \ldots, y_{T-1}$. The initial state s_0 is assumed known; if it is not known, another expectation can be taken over S_0. While we strive to explicitly write out the conditioning on s_0, it should be interpreted that all terms in this chapter are implicitly conditioning on s_0 even if not written. When adopting $S_t = I_t$, the constraints can be written as $\xi_t = \mu_t(I_t) \in \mathcal{A}_t$ and

$I_{t+1} = \{I_t, \xi_t, Y_t\}$. This sOED formulation generalizes both the batch and greedy designs [134, Section 2.3].

3.1.3 Experimental design rewards

We propose two information gain (IG)-based [93, 61] reward formulations incorporating multiple design objectives. Non-IG-based reward terms (i.e., those that do not depend on the posteriors, such as a cost term of the experiment that depends only on ξ_t) may also be added without affecting the results below, but we omit them for simplicity of presentation. We will largely use I_t in place of S_t from hereon.

(1) Terminal-information-gain (TIG) TIG targets the overall IG (i.e., the Kullback–Leibler [KL] divergence from prior to posterior) from all T experiments via the terminal reward:

$$r_t(\mathrm{i}_t, \xi_t, y_t) = 0, \quad t = 0, \ldots, T-1, \tag{3.7}$$

$$\begin{aligned} r_N(\mathrm{i}_T) = {} & \alpha_M D_{\mathrm{KL}}\left(P_{M|\mathrm{i}_T} \,||\, P_M\right) \\ & + \mathbb{E}_{M|\mathrm{i}_T}\left[\alpha_\Theta D_{\mathrm{KL}}\left(p_{\Theta_m|\mathrm{i}_T} \,||\, p_{\Theta_m}\right) + \alpha_Z D_{\mathrm{KL}}\left(p_{Z_m|\mathrm{i}_T} \,||\, p_{Z_m}\right)\right], \end{aligned} \tag{3.8}$$

where in the KL divergence the upper case P denotes probability mass (for discrete random variables), subscript indicates the random variable that corresponds to the probability mass or density, and $\alpha_M, \alpha_\Theta, \alpha_Z \in [0, 1]$ are the weights for IG respectively coming from the model indicator, PoIs, and QoIs. For example, setting $\alpha_M = 1$ and $\alpha_\Theta = \alpha_Z = 0$ reduces to 'OED for model indicator' (i.e., design for model discrimination); $\alpha_\Theta = 1$ and $\alpha_M = \alpha_Z = 0$ reduces to 'OED for PoIs' (i.e., design for parameter inference); $\alpha_Z = 1$ and $\alpha_\Theta = \alpha_M = 0$ reduces to 'OED for QoIs' (i.e., design for goal-oriented prediction). In the special case when $\alpha_M = \alpha_\Theta = 1$ and $\alpha_Z = 0$, the terminal reward r_N becomes $D_{\mathrm{KL}}\left(p_{M,\Theta_m|\mathrm{i}_T} \,||\, p_{M,\Theta_m}\right)$ (Appendix B.1.1); when $\alpha_M = \alpha_Z = 1$ and $\alpha_\Theta = 0$, it similarly becomes $D_{\mathrm{KL}}\left(p_{M,Z_m|\mathrm{i}_T} \,||\, p_{M,Z_m}\right)$ (Appendix B.1.1). When nuisance parameters ψ_m are absent, one should not set both α_Θ and α_Z to 1 since the IG in z_m is already absorbed into the IG in Θ_m (Appendix B.1.2).

(2) Incremental-information-gain (IIG) IIG adopts incremental IG as the immediate rewards:

$$r_t(\mathsf{i}_t, \xi_t, y_t) = \alpha_M D_{\mathrm{KL}}\left(P_{M|\mathsf{i}_{t+1}} \,||\, P_{M|\mathsf{i}_t}\right) + \mathbb{E}_{M|I_{t+1}}\Big[\alpha_\Theta D_{\mathrm{KL}}\left(p_{\Theta_m|\mathsf{i}_{t+1}} \,||\, p_{\Theta_m|\mathsf{i}_t}\right) + \alpha_Z D_{\mathrm{KL}}\left(p_{Z_m|\mathsf{i}_{t+1}} \,||\, p_{Z_m|\mathsf{i}_t}\right)\Big], \quad t = 0, \ldots, T-1, \tag{3.9}$$

$$r_N(\mathsf{i}_T) = 0. \tag{3.10}$$

Note that i_{t+1} can be evaluated in Eqn. (3.9) since $\mathsf{i}_{t+1} = \{\mathsf{i}_t, \xi_t, y_t\}$.

Let $U_T(\pi)$ denote the resulting sOED expected utility in Eqn. (3.6) when adopting the TIG rewards in Eqn. (3.7) and (3.8), and $U_I(\pi)$ denote that when adopting the IIG rewards in Eqn. (3.9) and (3.10).

Theorem 3.1.1 (Terminal-incremental equivalence)**.** *$U_T(\pi) = U_I(\pi)$ for any policy π.*

A proof is provided in Appendix B.1.3. Hence, both formulations induce the same sOED problem.

3.1.4 One-point formulation of rewards

To compute the sOED objective in Eqn. (3.6) with the TIG or IIG rewards, an expectation needs to be taken over $y_{0:T-1}|\pi, s_0$. This requires sampling trajectories. For each trajectory, a model indicator and corresponding PoIs and nuisance parameters are drawn from the priors, $m_0^{(i)} \sim P_M, \theta_{m,0}^{(i)} \sim p\big(\theta_{m,0}|m_0^{(i)}\big), \psi_{m,0}^{(i)} \sim p\big(\psi_{m,0}|m_0^{(i)}\big)$, and a corresponding QoI sample is also drawn, $z_{m,0}^{(i)} \sim p\big(z_{m,0}|\theta_{m,0}^{(i)}, \psi_{m,0}^{(i)}, m_0^{(i)}\big)$ (if z_m has a deterministic H, then simply $z_{m,0}^{(i)} = H\big(\theta_{m,0}^{(i)}, \psi_{m,0}^{(i)}; m_0^{(i)}\big)$). Then, $m_0^{(i)}$, $\theta_{m,0}^{(i)}$, and $\psi_{m,0}^{(i)}$ generate a trajectory i_T. For any such trajectory, we substitute these 'oracle' values of the model indicator and model parameters that generated i_T into the integrands of the KL divergence terms to produce a 'one-point' approximation $\check{r}_t$ for each r_t.

(1) One-point-TIG One-point-TIG entails the following new reward terms:

$$\check{r}_t(\mathsf{i}_t, \xi_t, y_t) = 0, \quad t = 0, \ldots, T-1, \tag{3.11}$$

$$\check{r}_T(\mathsf{i}_T) = \alpha_M \log \frac{P(m_0|\mathsf{i}_T)}{P(m_0)} + \alpha_\Theta \log \frac{p(\theta_{m,0}|\mathsf{i}_T)}{p(\theta_{m,0})} + \alpha_Z \log \frac{p(z_{m,0}|\mathsf{i}_T)}{p(z_{m,0})}. \tag{3.12}$$

(2) One-point-IIG One-point-IIG entails the following new reward terms:

$$\check{r}_t(\mathrm{i}_t, \xi_t, y_t) = \alpha_M \log \frac{P(m_0|\mathrm{i}_{t+1})}{P(m_0|\mathrm{i}_t)} + \alpha_\Theta \log \frac{p(\theta_{m,0}|\mathrm{i}_{t+1})}{p(\theta_{m,0}|\mathrm{i}_t)} + \alpha_Z \log \frac{p(z_{m,0}|\mathrm{i}_{t+1})}{p(z_{m,0}|\mathrm{i}_t)}, \quad t = 0, \dots, T-1, \tag{3.13}$$

$$\check{r}_T(\mathrm{i}_T) = 0. \tag{3.14}$$

More precisely, the one-point reward terms are now functions of the trajectory-generating oracle model and parameter values, i.e., $\check{r}_t(\mathrm{i}_t, \xi_t, y_t, m_0, \theta_{m,0}, z_{m,0})$ and $\check{r}_T(\mathrm{i}_T, m_0, \theta_{m,0}, z_{m,0})$, but with the constraint that the $m_0, \theta_{m,0}, z_{m,0}$ arguments are those that generated the i_T argument; we omitted the oracle arguments in the $\check{r}_t$ expressions above for simplicity. Thus, we need to take expectation jointly over these inputs, to arrive at the one-point expected utility:

$$\check{U}(\pi) = \mathbb{E}_{M_0,\theta_{m,0},\psi_{m,0},z_{m,0}} \left[\mathbb{E}_{I_T|\pi,s_0,M_0,\theta_{m,0},\psi_{m,0}} \left[\sum_{t=0}^{T-1} \check{r}_t(I_t, \xi_t, Y_t) + \check{r}_T(I_T) \right] \right], \tag{3.15}$$

where we switch to the notation with $S_t = I_t$ to correspond to the I_t-based reward definitions, use the equivalence of $\mathbb{E}_{y_{0:T-1}|\pi,s_0,\dots}$ and $\mathbb{E}_{I_T|\pi,s_0,\dots}$, and follow the policy $\xi_t = \mu_t(I_t)$. Moreover, the inner expectation's conditioning on $z_{m,0}$ is dropped since the I_T sequence (and the Y_t's within I_T) does not depend on z_m when π is given. Furthermore, the expectation over $z_{m,0}$ can be ignored altogether for z_m's with deterministic H.

Let $\check{U}_T(\pi)$ denote the resulting sOED expected utility from Eqn. (3.15) when adopting the one-point-TIG rewards in Eqn. (3.11) and (3.12), and $\check{U}_I(\pi)$ denote that when adopting the one-point-IIG rewards in Eqn. (3.13) and (3.14).

Theorem 3.1.2 (One-point equivalence)**.** *$U_T(\pi) = \check{U}_T(\pi) = \check{U}_I(\pi) = U_I(\pi)$ for any policy π.*

A proof is provided in Appendix B.1.4. Hence, both the original sOED and one-point formulations, using either TIG or IIG, induce the same sOED problem.

Remark 3.1.3. For $\check{U}_I(\pi)$, under the summation, all intermediate posteriors cancel out and only the prior terms $p(\cdot)$ and the final posterior terms $p(\cdot|I_T)$ survive. However, working with intermediate posteriors in the incremental rewards can lead to denser rewards that improves numerical performance [18].

Remark 3.1.4. For any expected utility form, the prior terms $p(\cdot)$ may be omitted since they are independent of π and would only shift the expected utility without affecting the maximizer (Appendix B.1.5). Therefore, when the prior is difficult to compute (e.g., the prior-predictive $p(z_m)$ in Eqn. (3.4) that needs to marginalize out Θ_m and Ψ_m), we will drop that prior term when optimizing the policy.

3.1.5 Variational sequential optimal experimental design (vsOED)

Evaluating the probability terms of the one-point formulations in Sec. 3.1.4 remains highly challenging. For example, discretizing the Θ space and computing the density on a grid, as done in [134], can become computationally expensive even for moderately sized parameter spaces. To make the computation tractable, we replace all posterior terms $p(\cdot|\mathsf{i}_t)$ with variational posterior approximations $q(\cdot|\mathsf{i}_t;\phi_{(\cdot)})$ parameterized by $\phi_{(\cdot)}$. We assume the q's are always chosen such that the relevant KL divergence terms arising from the reward definitions below are well-defined.

(1) Variational-one-point-TIG Variational-one-point-TIG reward terms update to:

$$\tilde{r}_t(\mathsf{i}_t, \xi_t, y_t; \phi) = 0, \quad t = 0, \dots, T-1, \tag{3.16}$$

$$\tilde{r}_N(\mathsf{i}_T; \phi) = \alpha_M \log \frac{q(m_0|\mathsf{i}_T;\phi_M)}{P(m_0)} + \alpha_\Theta \log \frac{q(\theta_{m,0}|\mathsf{i}_T;\phi_{\Theta_m})}{p(\theta_{m,0})} + \alpha_Z \log \frac{q(z_{m,0}|\mathsf{i}_T;\phi_{z_m})}{p(z_{m,0})}, \tag{3.17}$$

where $\phi = \{\phi_M, \phi_{\Theta_m}, \phi_{z_m}\}$ is the full set of variational parameters.

(2) Variational-one-point-IIG Variational-one-point-IIG reward terms update to:

$$\begin{aligned}\tilde{r}_t(\mathsf{i}_t, \xi_t, y_t; \phi) = \alpha_M \log \frac{q(m_0|\mathsf{i}_{t+1};\phi_M)}{q(m_0|\mathsf{i}_t;\phi_M)} + \alpha_\Theta \log \frac{q(\Theta_{m,0}|\mathsf{i}_{t+1};\phi_{\Theta_m})}{q(\Theta_{m,0}|\mathsf{i}_t;\phi_{\Theta_m})} \\ + \alpha_Z \log \frac{q(z_{m,0}|\mathsf{i}_{t+1};\phi_{z_m})}{q(z_{m,0}|\mathsf{i}_t;\phi_{z_m})}, \quad t = 0, \dots, T-1,\end{aligned} \tag{3.18}$$

$$\tilde{r}_N(\mathsf{i}_T; \phi) = 0, \tag{3.19}$$

with the understanding that here $q(\cdot|\mathsf{i}_0;\phi_{(\cdot)})$ is simply the prior $p(\cdot)$ without any approximation. Similar to $\check{r}_t$, the new $\tilde{r}_t$ expressions are also functions of the oracle variables, but we omitted them in the expressions above for simplicity. Upon taking expectation over all random variables, the corresponding variational one-point expected utility becomes:

$$\tilde{U}(\pi, \phi) = \mathbb{E}_{M_0,\Theta_{m,0},\psi_{m,0},z_{m,0}}\left[\mathbb{E}_{I_T|\pi,s_0,M_0,\Theta_{m,0},\psi_{m,0}}\left[\sum_{t=0}^{T-1} \tilde{r}_t(I_t, \xi_t, Y_t; \phi) + \tilde{r}_N(I_T; \phi)\right]\right], \tag{3.20}$$

where $\xi_t = \mu_t(I_t)$, and again the outer expectation over $z_{m,0}$ can be ignored for z_m's with deterministic H.

Let $\tilde{U}_T(\pi)$ denote the resulting sOED expected utility from Eqn. (3.20) when adopting the variational-one-point-TIG rewards in Eqn. (3.16) and (3.17), and $\tilde{U}_I(\pi)$ denote that when adopting the variational-one-point-IIG rewards in Eqn. (3.18) and (3.19).

Theorem 3.1.5 (Variational lower bound)**.** $\tilde{U}_I(\pi;\phi) = \tilde{U}_T(\pi;\phi) \leq \check{U}_T(\pi) = \check{U}_I(\pi) = U_T(\pi) = U_I(\pi)$ *for any policy* π *and variational parameters* ϕ*. The bound is tight if and only if all final posteriors are perfectly approximated, i.e.,* $q(\cdot|\mathrm{i}_T;\phi_{(\cdot)}) = p(\cdot|\mathrm{i}_T)$ *(except the trivial case when* $\alpha_M = \alpha_\Theta = \alpha_Z = 0$*).*

A proof is provided in Appendix B.1.6.

Remark 3.1.6. In batch OED and when $\alpha_M = \alpha_Z = 0$ and $\alpha_\Theta = 1$, Eqn. (3.20) becomes the BA lower bound for mutual information between Θ and Y [12, 120, 50].

Remark 3.1.7. The results in all aforementioned theorems also hold when incorporating non-IG-based reward terms (i.e., those that do not depend on the posteriors, such as a cost term of the experiment that depends only on ξ_t).

Remark 3.1.8. The tightness of the bound does not depend on the quality of the intermediate variational posteriors (i.e., $q(\cdot|\mathrm{i}_t;\phi_{(\cdot)})$ for $t = 1,\ldots,T-1$) due to their cancellations, similar to Remark 3.1.3. Thus, low-quality intermediate posterior approximations may be used without affecting the theoretical value of the bound (see the first part of Appendix B.1.6). In the special case where all intermediate posteriors are approximated by the prior, variational-one-point-IIG collapses to variational-one-point-TIG. As we will show through numerical examples, however, good intermediate posterior approximations can lead to better computational performance.

Remark 3.1.9. Following [50, 41], it is also possible to construct a variational upper bound for $U_T(\pi)$ by first substituting $\frac{p(\cdot|I_N)}{p(\cdot)} = \frac{p(I_N|\cdot)}{p(I_N)}$ and then approximating $p(I_N)$ with $q(I_N)$. This upper bound is tight if and only if $p(I_N) = q(I_N)$. However, we choose to use the variational posterior (lower bound) approach for three main reasons. First, optimizing the policy using the tightest (i.e., lowest) upper bound would lead to a maximin saddle-point problem, which is more challenging to solve. Second, the dimension of I_N grows with T and typically lies in a high-dimensional space, making density estimation particularly difficult. Third, using variational posteriors can reduce the computational burden of subsequent posterior inference after the experiments are completed, since the "plug-in" variational posteriors are already available.

Variational sOED problem statement The **vsOED** problem seeks a design policy that maximizes the lower bound $\tilde{U}(\pi;\phi)$:

$$\begin{aligned} \{\pi^*,\phi^*\} \in \underset{\pi=\{\mu_0,\ldots,\mu_{T-1}\},\phi}{\operatorname{argmax}} \quad & \tilde{U}(\pi;\phi) \\ \text{subject to} \quad & \xi_t = \mu_t(I_t) \in \mathcal{A}_t, \\ & I_{t+1} = \{I_t,\xi_t,Y_t\}, \quad \text{for } t = 0,\ldots,T-1, \end{aligned} \tag{3.21}$$

where we used the notation adopting $S_t = I_t$.

3.2 Numerical methods for vsOED

Taking a similar approach as [134], we propose an actor-critic policy gradient method to solve the vsOED problem in Eqn. (3.21). The key to this method is to derive and compute gradients of the expected utility lower bound $\tilde{U}$ defined in Eqn. (3.20) with respect to the variational parameters and the policy, and use the gradients to improve the policy via, for example, gradient ascent.

To bring the policy (which entails functions) to a finite-dimensional space, we parameterize the policy π by $w \in \mathbb{R}^{N_w}$ and denote the parameterized policy as π_w. Forming the policy (i.e., the actor) explicitly in such a manner offers significantly faster online evaluation speeds [134, 49, 80, 18] compared to dynamic programming [78] and greedy design that require solving optimization problems on the fly. Upon replacing π with π_w, the vsOED problem from Eqn. (3.21) becomes:

$$\begin{aligned} \{w^*, \phi^*\} \in \quad & \underset{w,\phi}{\text{argmax}} \qquad \tilde{U}(w;\phi) \\ & \text{subject to} \qquad \xi_t = \mu_{t,w}(I_t) \in \mathcal{A}_t, \\ & \qquad\qquad\quad\; I_{t+1} = \{I_t, \xi_t, Y_t\}, \quad \text{for } t = 0, \ldots, T-1. \end{aligned} \tag{3.22}$$

The gradient of $\tilde{U}$ with respect to ϕ can be obtained trivially by applying the Leibniz rule:

$$\nabla_\phi \tilde{U}(w;\phi) = \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}}\left[\mathbb{E}_{I_T|w,s_0,M_0,\Theta_{m,0},\Psi_{m,0}}\left[\sum_{t=0}^{T-1} \nabla_\phi \tilde{r}_t(I_t,\xi_t,Y_t;\phi) + \nabla_\phi \tilde{r}_N(I_T;\phi)\right]\right], \tag{3.23}$$

where $\xi_t = \mu_{t,w}(I_t)$. The actor-critic policy gradient can be derived near-identically as the proof in Appendix B of [134] except that the expressions for vsOED now involve an additional outer expectation jointly over $M_0, \Theta_{m,0}, \Psi_{m,0}, Z_{m,0}$; therefore we do not repeat the derivation in this chapter. The final vsOED policy gradient expression is:

$$\nabla_w \tilde{U}(w;\phi) = \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}}\left[\sum_{t=0}^{T-1} \mathbb{E}_{I_t|w,s_0,M_0,\Theta_{m,0},\Psi_{m,0}}\left[\nabla_w \mu_{t,w}(I_t) \nabla_{\xi_t} \tilde{Q}_t^{\pi_w}(I_t,\xi_t;\phi)\right]\right], \tag{3.24}$$

where $\xi_t = \mu_{t,w}(I_t)$, and $\tilde{Q}_t^{\pi_w}$ is the action-value function (i.e., the critic) induced by the variational

one-point reward terms $\tilde{r}_t$ and $\tilde{r}_N$ and defined as:

$$\tilde{Q}_t^{\pi_w}(\mathsf{i}_t, \xi_t; \phi) = \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}|I_t}\Bigg[\mathbb{E}_{Y_{k:T-1}|w,s_0,\mathsf{i}_t,\xi_t,M_0,\Theta_{m,0},\Psi_{m,0}}\Bigg[\tilde{r}_t(\mathsf{i}_t, \xi_t, Y_t; \phi) + \sum_{t=t+1}^{T-1} \tilde{r}_t(I_t, \mu_{t,w}(I_t), Y_t; \phi) + \tilde{r}_N(I_T; \phi)\Bigg]\Bigg] \tag{3.25}$$

$$= \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}|\mathsf{i}_t}\Bigg[\mathbb{E}_{Y_k|w,s_0,\mathsf{i}_t,\xi_t,M_0,\Theta_{m,0},\Psi_{m,0}}\Bigg[\tilde{r}_t(\mathsf{i}_t, \xi_t, Y_t; \phi) + \tilde{Q}_{t+1}^{\pi_w}(I_{t+1}, \mu_{t+1,w}(I_{t+1}); \phi)\Bigg]\Bigg], \tag{3.26}$$

$$\tilde{Q}_N^{\pi_w}(\mathsf{i}_T, \cdot; \phi) = \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}|\mathsf{i}_T}\Big[\tilde{r}_N(\mathsf{i}_T; \phi)\Big], \tag{3.27}$$

for $t = 0, \ldots, T-1$ and subject to $I_{t+1} = \{I_t, \xi_t, Y_t\}$. The value of the critic $\tilde{Q}_t^{\pi_w}(\mathsf{i}_t, \xi_t; \phi)$ is the expected remaining cumulative reward (i.e., the expected sum of all remaining rewards) under variational parameters ϕ, for performing the tth experiment at design ξ_t from state i_t (i.e., from the latest posterior that is conditioned on I_t, as indicated by the outer conditional expectation) and thereafter following policy π_w. The critic can also be written in a recursive manner in Eqn. (3.26). To facilitate computing the $\nabla_{\xi_t}\tilde{Q}_t^{\pi_w}$ term in Eqn. (3.24), we also parameterize the critic functions by $\nu \in \mathbb{R}^{N_\nu}$ and denote the parameterized form as $\tilde{Q}_{t,\nu}^{\pi_w}$.

In summary, the two main steps involve: (a) forming variational posterior approximations q and estimating $\nabla_\phi \tilde{U}$ in Eqn. (3.23), and (b) forming approximate critic $\tilde{Q}_{t,\nu}^{\pi_w}$ and estimating $\nabla_w \tilde{U}$ in Eqn. (3.24). We detail these two steps below.

3.2.1 Variational gradient

Adopting the variational-one-point-TIG rewards into Eqn. (3.23), it is easy to verify that

$$\nabla_\phi \tilde{U}(w; \phi) = \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}}\Bigg[\mathbb{E}_{I_T|w,s_0,M_0,\Theta_{m,0},\Psi_{m,0}}\Big[\alpha_M \nabla_{\phi_M} \log q(M_0|I_T; \phi_M) + \alpha_\Theta \nabla_{\phi_{\Theta_m}} \log q(\Theta_{m,0}|I_T; \phi_{\Theta_m}) + \alpha_Z \nabla_{\phi_{Z_m}} \log q(Z_{m,0}|I_T; \phi_{Z_m})\Big]\Bigg], \tag{3.28}$$

where the log-prior terms disappear under the gradient operation since they do not depend on ϕ. The variational gradient for the variational-one-point IIG case is exactly the same, since the intermediate variational posteriors all cancel per Appendix B.1.6 and the gradient to those terms' ϕ's will always be zero. In order to obtain good intermediate posterior approximations in the variational-one-point-IIG case, we first note that the variational 'partial-length' expected utility up to stage t remains a lower bound to its non-variational counterpart for all t—that is, the result in Appendix B.1.6 remains true if we replace I_T with I_t, $\forall k$. Then, we can optimize the intermediate posteriors' ϕ's by successively maximizing these partial-length lower bounds.

We can then form a Monte Carlo (MC) estimator for the variational gradient in Eqn. (3.28):

$$\begin{aligned}\nabla_\phi \tilde{U}(w;\phi) \approx \frac{1}{n}\sum_{i=1}^{n}\Bigg[&\alpha_M \nabla_{\phi_M} \log q(m_0^{(i)}|\mathrm{i}_T^{(i)};\phi_M) \\ &+ \alpha_\Theta \nabla_{\phi_{\Theta_m}} \log q(\theta_{m,0}^{(i)}|\mathrm{i}_T^{(i)};\phi_{\Theta_m}) \\ &+ \alpha_Z \nabla_{\phi_{Z_m}} \log q(z_{m,0}^{(i)}|\mathrm{i}_T^{(i)};\phi_{Z_m})\Bigg],\end{aligned} \tag{3.29}$$

where the samples $m_0^{(i)}, \theta_{m,0}^{(i)}, \psi_{m,0}^{(i)}, z_{m,0}^{(i)}$ and corresponding trajectories $\mathrm{i}_T^{(i)}$ are generated following the same procedure described at the beginning of Sec. 3.1.4. In practice, multiple steps of gradient update can be applied to ϕ at a given w to more efficiently make use of the trajectory samples.

3.2.2 Policy gradient

In order to compute the policy gradient $\nabla_w \tilde{U}(w;\phi)$ in Eqn. (3.24), we need to first form the approximate critic $\tilde{Q}_{t,\nu}^{\pi_w}$.

This can be achieved by training $\tilde{Q}_{t,\nu}^{\pi_w}$, for any given ϕ, in a supervised learning manner to optimize ν towards the true $\tilde{Q}_k^{\pi_w}$ by minimizing a loss function based on the recursive definition in Eqn. (3.26). Since the true $\tilde{Q}_{t+1}^{\pi_w}$ values would not be available, they are replaced with the current approximations $\tilde{Q}_{t+1,\nu}^{\pi_w}$:

$$\begin{aligned}\nu^* \in \operatorname*{argmin}_\nu \Bigg\{\mathcal{L}_\phi(\nu) := \frac{1}{n}\sum_{i=1}^{n}\sum_{t=0}^{T-1}\Bigg[&\tilde{Q}_{k,\nu}^{\pi_w}(\mathrm{i}_t^{(i)}, \xi_t^{(i)};\phi) \\ &- \left(\tilde{r}_t(\mathrm{i}_t^{(i)}, \xi_t^{(i)}, y_t^{(i)};\phi) + \gamma \tilde{Q}_{t+1,\nu}^{\pi_w}(\mathrm{i}_{t+1}^{(i)}, \xi_{t+1}^{(i)};\phi)\right)\Bigg]^2\Bigg\},\end{aligned} \tag{3.30}$$

where $\gamma \in [0,1]$ is a weighing factor inserted for regularization. In practice, we first sample entire trajectories of $\mathrm{i}_T^{(i)}$ following the procedure described at the beginning of Sec. 3.1.4. Then, we extract

the partial sequences $\mathsf{i}_t^{(i)}$ and designs $\xi_t^{(i)}$ resulting from the same sample trajectory $\mathsf{i}_T^{(i)}$ to use for all t, instead of generating new $\mathsf{i}_{t+1}^{(i)}$ from their latest posteriors which would be much more expensive. The gradient $\nabla_\nu \mathcal{L}_\phi(\nu)$ can be obtained readily by differentiating $\pi_{wt,\nu}$ with respect to ν, while the contribution from $\tilde{Q}_{t+1,\nu}^{\pi_w}$ is typically omitted as it is a stand-in of the true $\tilde{Q}_{t+1}^{\pi_w}$ that does not depend on ν. The optimal value of ν^* needs not be precisely found, and often just a few steps of gradient update at this π_w can already lead to good performance, as verified in Table B.12.

The loss $\mathcal{L}_\phi$ from Eqn. (3.30) can be used for both variational-one-point-TIG and -IIG. However, the TIG case has $\tilde{r}_t = 0$ for $t = 0, \ldots, T-1$ (except for possible non-IG reward terms) and as a result it would take many ν updates from Eqn. (3.30)—especially when T is large—for the non-zero $\tilde{r}_T$ to propagate to the $\tilde{Q}_{t,\nu}^{\pi_w}$'s at early t's. This can lead to slow or even divergent policy gradient updates. To mitigate this effect, we follow the REINFORCE algorithm [160] and propose a modified loss as a more stable option for the TIG case:

$$\begin{aligned}\mathcal{L}_{\phi,T}(\nu) = \frac{1}{n}\sum_{i=1}^{n}\sum_{t=0}^{T-1}\Bigg[&\tilde{Q}_{k,\nu}^{\pi_w}(\mathsf{i}_t^{(i)}, \xi_t^{(i)}; \phi) \\ &- \eta\left(\tilde{r}_t(\mathsf{i}_t^{(i)}, \xi_t^{(i)}, y_t^{(i)}; \phi) + \gamma\tilde{Q}_{t+1,\nu}^{\pi_w}(\mathsf{i}_{t+1}^{(i)}, \xi_{t+1}^{(i)}; \phi)\right) \\ &- (1-\eta)\left(\sum_{t=k}^{T-1}\gamma^{t-k}\tilde{r}_t(\mathsf{i}_t^{(i)}, \xi_t^{(i)}, y_t^{(i)}; \phi) + \gamma^{N-k}\tilde{r}_N(\mathsf{i}_T^{(i)}; \phi)\right)\Bigg]^2, \end{aligned} \tag{3.31}$$

where η linearly increases from 0 to 1 during the training process.

Once $\tilde{Q}_{t,\nu}^{\pi_w}$ is obtained, we finally form the MC estimator for the policy gradient in Eqn. (3.24):

$$\nabla_w \tilde{U}(w; \phi) \approx \frac{1}{n}\sum_{i=1}^{n}\sum_{t=0}^{T-1}\nabla_w \mu_{t,w}(\mathsf{i}_t^{(i)})\nabla_{\xi_t}\tilde{Q}_{k,\nu}^{\pi_w}(\mathsf{i}_t^{(i)}, \xi_t^{(i)}; \phi), \tag{3.32}$$

where $\mathsf{i}_T^{(i)}$ is generated following the same procedure described at the beginning of Sec. 3.1.4 using policy π_w, and $I_t^{(i)}$ and $\xi_t^{(i)}$ are the partial sequences and designs, respectively, of the full sample sequence $\mathsf{i}_N^{(i)}$.

3.2.3 Implementation techniques and the overall algorithm

Variational posteriors We use a neural network (NN) to represent the approximate posterior of model indicator, $q(m_0|I_t; \phi_M)$; the NN takes I_t as input and uses a softmax output activation to output model probability. For the approximate posteriors of PoIs and QoIs, respectively $q(\theta_{m,0}|I_t; \phi_{\Theta_m})$ and $q(z_{m,0}|I_t; \phi_{Z_m})$, we adopt independent Gaussian mixture models (GMMs) with NNs predicting the GMM weights, means, and standard deviations. Truncated normal is used

for variables with compact support. We also explore the use of normalizing flows (NFs) [41] for posterior approximations. More details can be found in Appendix B.2.1 and B.2.2.

Policy and critic networks We employ NNs to represent both π_w and $\tilde{Q}_{t,\nu}^{\pi_w}$, therefore we also refer to these NNs as the policy network and critic network, respectively. The optimizations of the policy and critic networks are both carried out using Adam [83] with mini-batching. More details about these networks can be found in Appendix B.2.3 and B.2.4.

Target networks We adopt target networks for the actor and critic—i.e., secondary networks updated less frequently and with a damping factor—in order to promote stability in the training process by smoothing the learning target and mitigating drastic changes of Q-values and designs across training iterations. Target networks have been shown to greatly improve robustness and effectiveness of policy-gradient-based methods [92].

Replay buffer We use a replay buffer to store past trajectory samples from previously encountered policies, and resample these stored samples for use during the policy optimization process. Replay buffer permits off-policy learning, which entails optimizing the policy using data generated from policies that differ from the current one [159]. This can help stabilize the optimization and improve sample efficiency by allowing multiple uses of the simulation data, for example those generated from policies in previous iterations and from exploration policies [104].

Exploration policy A balance of exploration and exploitation is important for the numerical optimizer to identify a good policy. Insufficient exploration limits understanding of the objective function's global landscape, while too much exploration can delay convergence. To inject exploration, we make use of an exploration policy during the policy optimization (i.e., training) phase by adding perturbation to the deterministic base policy: $\xi_t = \mu_{t,w}(I_t) + \mathcal{E}_{k,\text{explore}}$, where $\mathcal{E}_{k,\text{explore}} \sim \mathcal{N}(0, \Sigma_{k,\text{explore}})$ and $\Sigma_{k,\text{explore}} = \mathbb{I}\sigma^2_{k,\text{explore}}$ is a diagonal covariance whose entries reflect the exploration length scale for each dimension of $\mathcal{A}_t$. A reasonable strategy is to set larger exploration in the early training iterations and reduce it gradually; details will be specified for each numerical case in Sec. 3.3. Once the final policy is obtained, its evaluation (i.e., testing phase) will use the deterministic policy only.

Overall algorithm Pseudocode for the overall vsOED algorithm is presented in Algorithm 1. While we write the simple gradient ascent update formulas in the pseudocode for illustration, it can be replaced by any other gradient-based updates.

Algorithm 1: The vsOED algorithm.

Initialize variational parameters ϕ, actor (policy) parameters w, critic (action-value function) parameters v;
for $l = 1, \ldots, n_{\text{updates}}$ **do**
  Simulate n_{traj} trajectories: for the ith trajectory, sample $m_0^{(i)} \sim P_M$, $\theta_{m,0}^{(i)} \sim p\big(\theta_{m,0}|m_0^{(i)}\big)$, $\psi_{m,0}^{(i)} \sim p\big(\psi_{m,0}|m_0^{(i)}\big)$, $z_{m,0}^{(i)} \sim p\big(z_{m,0}|\theta_{m,0}^{(i)}, \psi_{m,0}^{(i)}, m_0^{(i)}\big)$, and then for $t = 0, \ldots, T-1$ generate $\xi_t^{(i)} = \mu_{k,w_l}(I_t^{(i)}) + \epsilon_{k,\text{explore}}^{(i)}$ and $Y_t^{(i)} \sim p(Y_t|m_0^{(i)}, \theta_{m,0}^{(i)}, \psi_{m,0}^{(i)}, \xi_t^{(i)}, I_t^{(i)})$;
  Update newly generated information sequences $\left\{\mathrm{i}_T^{(i)}\right\}_{i=1}^{n_{\text{traj}}}$ into replay buffer;
  Sample n_{batch} trajectories from the replay buffer;
  Use batch trajectories to estimate $\nabla_\phi \tilde{U}$ following Eqn. (3.29), update $\phi_{l+1} = \phi_l + a_{\phi,l}\nabla_\phi \tilde{U}$ with learning rate $a_{\phi,l}$ (can be done multiple times per l-iteration), and calculate $\left\{\tilde{r}_t^{(i)}\right\}_{i=1}^{n_{\text{batch}}}$ with the new ϕ_{l+1};
  Update v towards the v^* in Eqn. (3.30) (or Eqn. (3.31)), e.g., through multiple steps of gradient ascent;
  Estimate $\nabla_w \tilde{U}$ following Eqn. (3.32), and then update $w_{l+1} = w_l + a_{w,l}\nabla_w \tilde{U}$ with learning rate $a_{w,l}$;
end for
Return final policy π_w;

3.3 Numerical experiments

To enhance clarity and organization, we have included a table of case-specific symbols and notation used in the numerical experiments in Appendix B.3.6.

We demonstrate vsOED and compare it against other state-of-the-art sequential experimental design methods across a number of problems with varying complexity and that illuminate different challenges. We first describe the demonstration setup in Sec. 3.3.1, and then present results for four cases: Case 1—source location finding in Sec. 3.3.2; Case 2—constant elasticity of substitution (CES) in Sec. 3.3.3; Case 3—SIR model for disease spread in Sec. 3.3.4; and Case 4—convection-diffusion-reaction in Sec. 3.3.5.

3.3.1 Demonstration setup

For **vsOED**, we employ GMMs and NFs for posterior approximation, and variational-one-point-TIG and -IIG for reward formulation: we adopt the naming convention where, for example, **vsOED-G-I** stands for **G**MM with **I**IG, and **vsOED-N-T** for **N**Fs with **T**IG. Other sequential experimental design algorithms being compared include **Random** design, **DAD** [49], **iDAD** [80], and a stochastic policy based **RL** method [18]. **RL** can also be combined with TIG and IIG, denoted by **RL-T** and

RL-I respectively. DAD and iDAD require the derivative of the forward model, vsOED and iDAD can accommodate implicit likelihood, and vsOED can handle multiple models through the model indicator. Table 3.1 summarizes which of these aspects have been considered or studied with these algorithms. For DAD, iDAD, and RL, we directly use code from their original publications' Github repositories.

Table 3.1: Properties of different sequential experimental design methods.

	Adaptive	Implicit likelihood	No model derivative	Multiple models
Random	✗	✓	✓	✓
DAD [49]	✓	✗	✗	✗
iDAD [80]	✓	✓	✗	✗
RL [18]	✓	✗	✓	✗
vsOED	✓	✓	✓	✓

In this work, we pay special attention to study algorithm efficiency in terms of trajectory samples since they entail the majority of forward model runs (i.e., evaluating G_t and H). We make comparisons under two computational settings: (i) 'fully trained' where the policies are trained using algorithms' default (usually a conservatively large) number of trajectory samples from their respective publications—8 trillion, 100 billion, and 200 million for RL, DAD, and iDAD, respectively—and we will use 10 million for vsOED; and (ii) 'limited budget' where we restrict to a significantly smaller total budget of 10 million training trajectories for all algorithms and cases (except for Case 3—SIR model where the budget is set to 1 million).

To ensure fair comparisons, it is important to evaluate the final policies resulting from different algorithms in a common manner. To this end, we adopt PCE [51, 49], specifically its sequential version, to evaluate the final expected utilities for policies resulting from 'OED for model indicator' and 'OED for PoIs' when likelihood is explicit and when there are no nuisance parameters (see Appendix B.3.1 for details on PCE, including a new variant we derive to handle 'OED for model indicator'). PCE itself is intractable to compute due to its outer expectation operator, and must be estimated numerically, for example by MC integration of the expectation. The final MC estimation of PCE results in a nested loop structure; in our numerical cases, we use 2000 trajectory samples for the outer loop and 10^6 samples for the inner loop to achieve a high-quality estimation; when multiple models are involved, we split the inner samples to $10^6/|\mathcal{M}|$ for each model so that the total number of inner loop samples remains the same overall. In the remainder of the chapter, we will refer to 'the MC estimate of PCE' simply as 'PCE' for brevity.

PCE cannot be used for other OED scenarios involving QoIs and nuisance parameters because the prior predictive densities and likelihood in those scenarios are intractable; similarly, PCE cannot be used for cases where likelihood is implicit. For those situations, we instead evaluate the policies

by calculating a high-quality MC estimate of $\tilde{U}$ using 10^6 trajectory samples (except for Case 3—SIR model where 3×10^5 samples are used). In the remainder of the chapter, we similarly refer to 'the MC estimate of $\tilde{U}$' simply as '$\tilde{U}$' for brevity. As we will demonstrate, in all cases where both $\tilde{U}$ and PCE are computed, the former consistently yields lower values than the latter. Since PCE is known to be a lower bound for the EIG, this observation serves as a numerical verification to Theorem 3.1.5.

All numerical experiments are conducted on the University of Michigan Great Lakes Slurm High Performance Computing Cluster nodes, where each node is equipped with a single Nvidia Tesla A40 or V100 GPU. All experiments are implemented in Python using PyTorch.

3.3.2 Case 1: source location finding

We adapt the source location finding problem from [49]. We enlist three candidate models with uniform model prior $P(m) = 1/3$ for $m = 1, 2, 3$. The mth model contains m sources randomly located in a two-dimensional (2D) domain, each emitting a signal that decays inversely with the square of the distance. The PoIs are the source locations $\Theta_m = \{\Theta_{m,i}\}_{i=1}^m$ where $\Theta_{m,i} = [\Theta_{m,i,x}, \Theta_{m,i,y}] \in \mathbb{R}^2$ denotes the location of the ith source and is endowed with independent priors $\Theta_{m,i,x}, \Theta_{m,i,y} \sim \mathcal{N}(0, 1^2)$. The total intensity at location $[x, y] \in \mathbb{R}^2$, aggregated from all sources, is

$$G(\Theta_m, x, y; m) = a_{\text{bg}} + \sum_{i=1}^{m} \frac{1}{a_{\max} + \|\Theta_{m,i} - [x, y]\|_2^2},$$

where $a_{\text{bg}} = 10^{-1}$ is the background signal and $a_{\max} = 10^{-4}$ is the maximum signal. The design variables $\xi_t = [\xi_{t,x}, \xi_{t,y}] \in [-4, 4]^2$ entail selecting the location of measurement within the allowable design space. The observation model is then $Y_t = G(\Theta_m, \xi_{t,x}, \xi_{t,y}; m) + \mathcal{E}_t$, where $\mathcal{E}_t \sim \mathcal{N}(0, \sigma_\epsilon^2)$ with $\sigma_\epsilon = 0.5$.

Additionally, we are also interested in a QoI that is the log of the total flux magnitude over an infinite vertical wall located at $x = 6$. The flux vector J at a spatial location $[x, y]$ is given by Fick's law:

$$J(\Theta_m, x, y; m) = -D\nabla_{(x,y)} G(\Theta_m, x, y; m),$$

where $D = 1$ is the scalar diffusivity. The total flux across the vertical wall at $x = 6$ is then

$$\varphi(\Theta_m; m) = \int_{-\infty}^{+\infty} J(\Theta_m, x = 6, y; m)\, \mathrm{d}y = \sum_{i=1}^{m} -\frac{\pi(\Theta_{m,i,x} - 6)}{\left(a_{\max} + (\Theta_{m,i,x} - 6)^2\right)^{\frac{3}{2}}}. \tag{3.33}$$

At last, the final QoI is $Z_m = \log |\varphi(\Theta_m; m)|$.

3.3.2.1 Case 1a: single-model

Let us first consider the single-model case where m is fixed at 2.

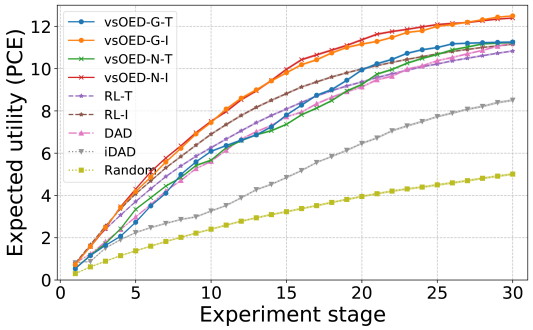
OED for PoIs (design for inference) The first example entails design for inferring the source location PoIs Θ, with $\alpha_\Theta = 1$ and $\alpha_Z = 0$; this setup is identical to those from previous literature [49, 80, 18]. Figure 3.1a presents the expected cumulative utility, evaluated using PCE, at different experiment stages t for policies optimized for a horizon of $T = 30$ experiments under the fully trained setting.

The plot suggests vsOED with IIG to achieve noticeably higher expected utilities throughout, while vsOED with TIG is lower in comparison but maintains a similar level as other algorithms at later stage numbers while lower in earlier stages. The intuition behind why IIG outperforms TIG, as mentioned just before Eq. (3.31), is that in TIG, the IG reward contribution only enters at the final stage, $k = N$. As a result, it would take many ν updates—especially when T is large—for the non-zero reward at $k = N$ to propagate to early t's. This can lead to slow policy gradient progress.

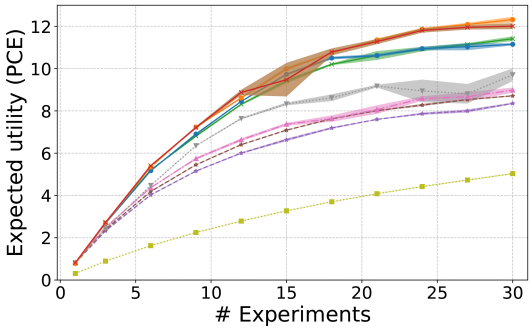
Figure 3.1b presents the average expected utility, averaged over four training replicates with different random seeds where each replicate is again evaluated using PCE, for policies trained under different design horizons T under the limited budget setting—that is, each point corresponds to policies optimized specifically for that horizon.

Here with limited budget, vsOED outperforms other algorithms throughout. IIG again performs better than TIG, especially for $T > 15$. The standard errors are plotted as shaded regions and illustrate general training robustness with vsOED-N-I exhibiting some instability at $T = 15$ and iDAD at $T = 24$ and higher. Lastly, Fig. 3.2 presents examples of approximate posteriors using GMMs and NFs, alongside the true posterior. The similarities between the approximations and the true posterior indicate the effectiveness of both GMMs and NFs in capturing complex, non-Gaussian posterior distributions.

OED for QoIs (design for prediction) The second example involves goal-oriented design for predicting the flux-based QoI Z introduced earlier, with $\alpha_\Theta = 0$ and $\alpha_Z = 1$. Only vsOED is used here since the other algorithms cannot accommodate 'OED for QoIs'; furthermore, only GMM version of vsOED is presented since the QoI Z is a scalar and unsuitable for the particular architecture of NFs we adopt that requires at least two dimensions for decomposition. Figure 3.3 plots the average $\tilde{U}$, averaged over four training replicates, versus T. The standard errors are plotted as shaded regions illustrating training robustness for vsOED. As observed in 'OED for PoIs', we again see higher expected utility estimates reached by IIG over TIG. Figure 3.4 presents

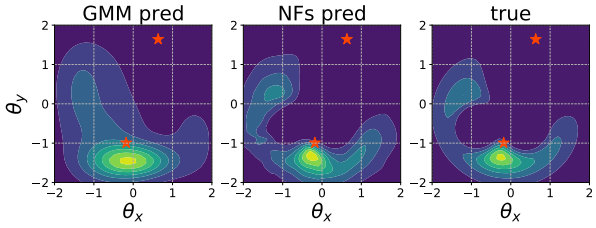


(a) Expected cumulative utility at different experiment stage t for policies optimized for $T = 30$

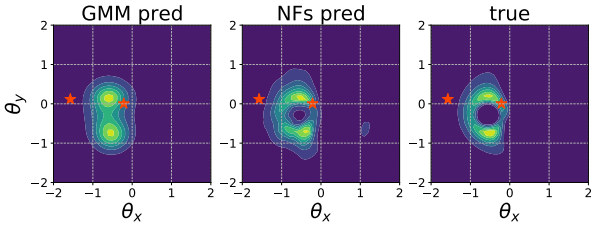


(b) Average expected utility over four training replicates versus design horizon T

Figure 3.1: Case 1a. Expected utility comparisons using policies resulting from different algorithms. The shaded regions represent the standard error.



(a) Example 1



(b) Example 2

Figure 3.2: Case 1a. Examples of GMM and NFs approximate posterior and true posterior for PoIs. The red stars indicate the true data-generating parameter values for these trajectory samples.

examples of trajectory using policies resulting from 'OED for PoIs' and 'OED for QoIs' for $T = 15$: the former tends to move toward the estimated source locations while the latter forms a roughly vertical design pattern. We can explain this behavior through physics. In Eqn. (3.33), since the integration is over the entire y, the flux is solely dependent on the x-position of the source (i.e., $\Theta_{m,i,x}$). Due to the isotropic nature of source emission, spreading measurements along a vertical line is thus more sensitive at detecting changes in θ_x. Figure 3.5a illustrates the QoI posteriors for both vertical and horizontal sensor designs. The results show a greater degree of QoI posterior shrinkage under the vertical sensor design compared to the horizontal design, which aligns with the previous interpretation. Figure 3.5b shows the approximate and true QoI posteriors. Their strong similarity supports that GMMs can also capture well the QoI posteriors.

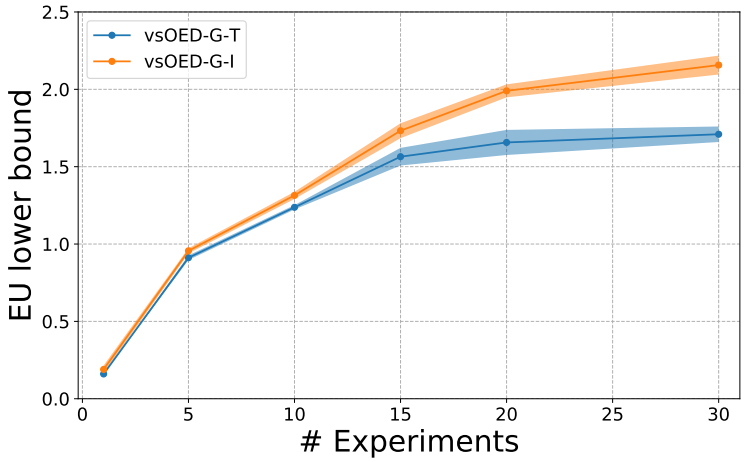
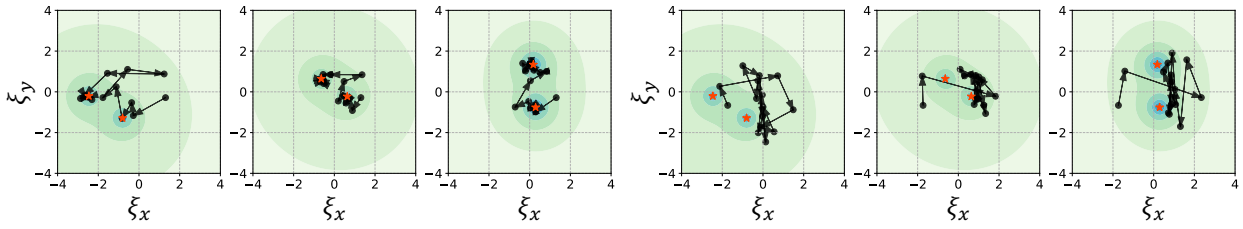


Figure 3.3: Case 1a. Average $\tilde{U}$ over four training replicates for 'OED for QoIs'. The shaded regions represent the standard error.

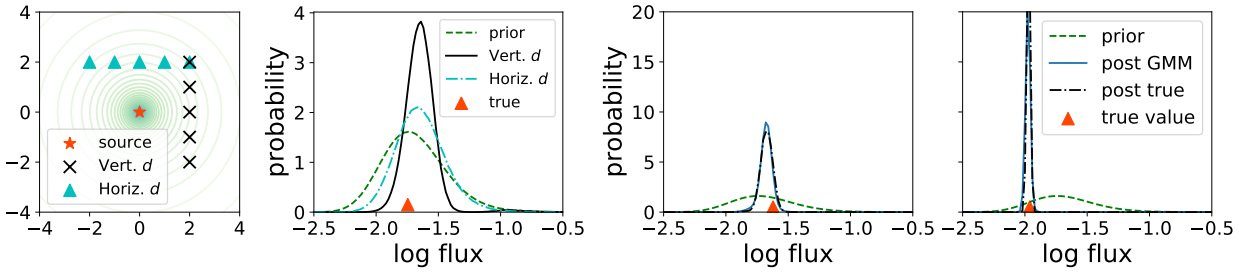


(a) OED for PoIs

(b) OED for QoIs

Figure 3.4: Case 1a. Examples of policy trajectory for $T = 15$. The contour background plots the true signal strength, and the red stars indicate the true source locations.

Hyperparameters and training stability Hyperparameter settings and training stability results can be found in Appendix B.3.2.



(a) Vertical and horizontal designs and posteriors

(b) Examples of QoI posterior at $T = 15$

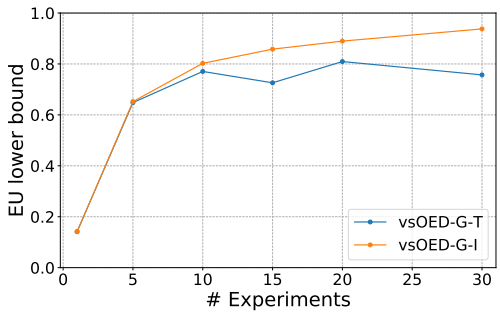
Figure 3.5: Case 1a. Design and posterior comparisons for 'OED for QoIs'.

3.3.2.2 Case 1b: multi-model

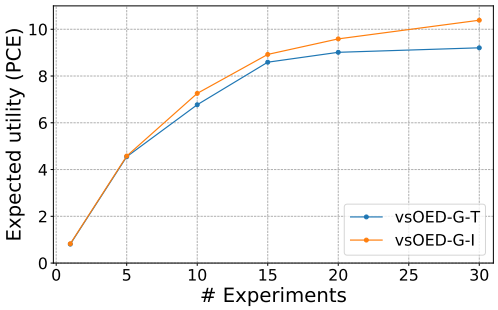
We now consider a multi-model case for the source location finding problem, with three candidate models. Five scenarios are considered: OED for model indicator (i.e., design for model discrimination, $\alpha_M = 1$, $\alpha_\Theta = \alpha_Z = 0$), OED for PoIs (i.e., design for parameter inference, $\alpha_\Theta = 1$, $\alpha_M = \alpha_Z = 0$), OED for QoIs (i.e., design for goal-oriented prediction, $\alpha_Z = 1$, $\alpha_M = \alpha_\Theta = 0$), OED for both model indicator and PoIs ('model-PoIs', $\alpha_M = \alpha_\Theta = 1$, $\alpha_Z = 0$), and OED for both model indicator and QoIs ('model-QoIs', $\alpha_M = \alpha_Z = 1$, $\alpha_\Theta = 0$).

Results Only vsOED is used in this case since the other algorithms cannot handle multiple models. Only the GMM version of vsOED is presented for brevity as this example focuses on vsOED capabilities across the five OED scenarios. Figure 3.6 presents the average expected utilities or $\tilde{U}$ over two training replicates for the five OED scenarios. IIG again demonstrates greater performance over TIG, especially for $T \geq 15$. Note that the loss of monotonicity for vsOED-G-T at larger values of T in some plots is attributable to numerical randomness; additional testing with more replicates suggests that the monotonic trend can be recovered. Figure 3.7 presents examples of trajectory using policies resulting from the five OED scenarios for $T = 30$. Generally, trajectories from 'OED for model indicator' tend to explore more and appear more diffuse, while those from 'OED for PoIs' tend to be more exploitative and remain closer to the estimated source location. Similar to the single-model case, 'OED for QoIs' promotes more vertically aligned design locations. The model-PoIs and model-QoIs scenarios appear more diffuse than 'OED for PoIs' and 'OED for QoIs', respectively, due to the addition of exploratory property from 'OED for model indicator'.

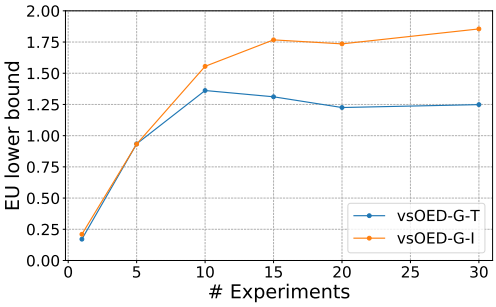
Figure 3.8 shows examples of approximate posterior and true posterior for model indicator using policy resulting from 'OED for model indicator' with vsOED-G-I for $T = 30$. The policy appears effective in discriminating the candidate models with highly concentrated posteriors, and



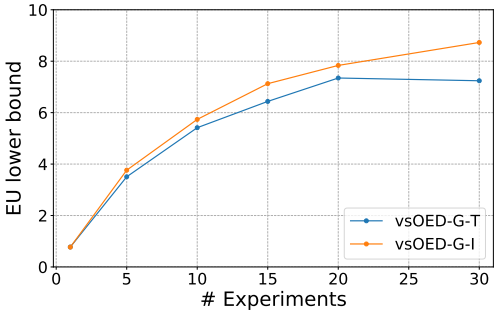
(a) OED for model indicator



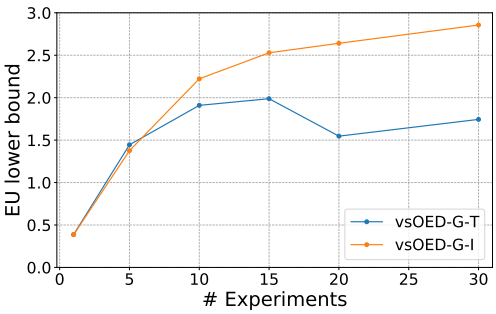
(b) OED for PoIs



(c) OED for QoIs

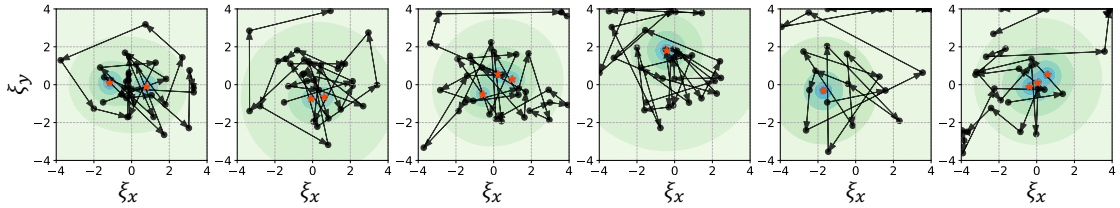


(d) Model-PoIs OED

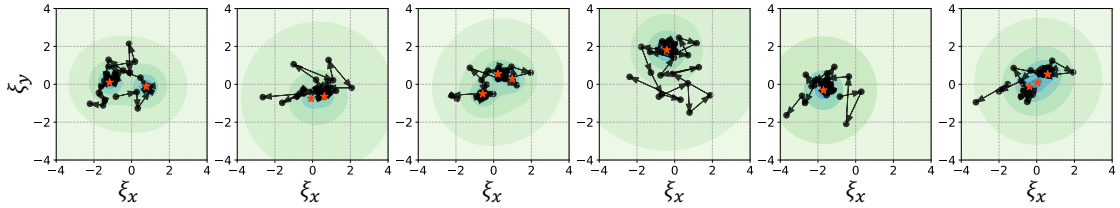


(e) Model-QoIs OED

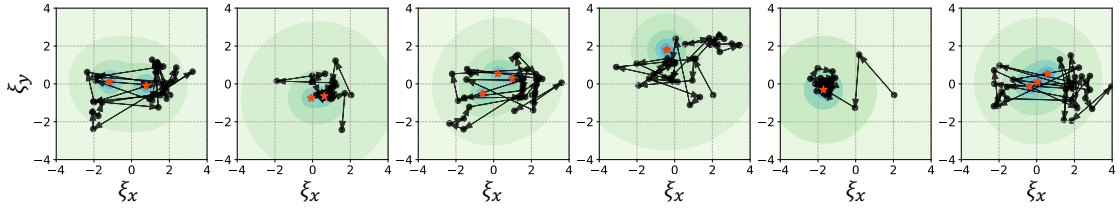
Figure 3.6: Case 1b. Average expected utility or $\tilde{U}$ over two training replicates versus design horizon T using policies resulting from the five OED scenarios.



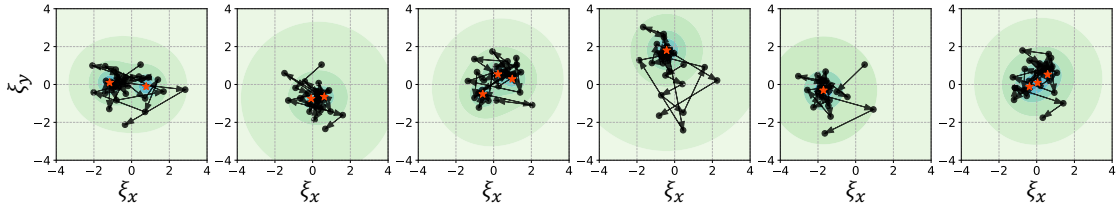
(a) OED for model indicator



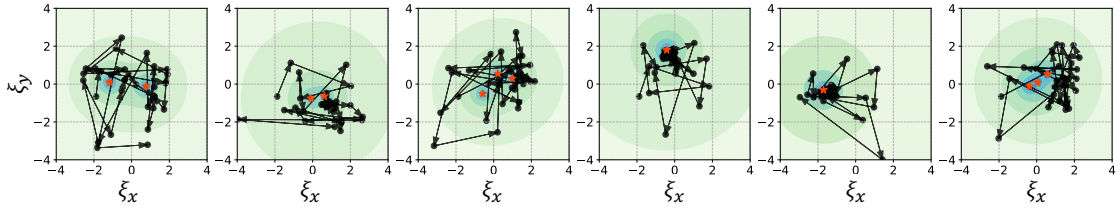
(b) OED for PoIs



(c) OED for QoIs



(d) Model-PoIs OED



(e) Model-QoIs OED

Figure 3.7: Case 1b. Examples of policy trajectory for $T = 30$. The contour background plots the true signal strength, and the red stars indicate the true source locations.

the approximate model posteriors match well with the true model posteriors. Table 3.2 presents the PCE estimates of the EIG for model indicator, PoIs, and QoIs, using policies resulting from the five OED scenarios, all using vsOED-G-I. The model discrimination OED finds the optimal policy in terms of maximizing the expected utility on model probability. Indeed, the 'OED for model indicator' policy achieves the highest EIG estimate for model indicator, while model-PoIs and model-QoIs achieve higher values compared to their counterparts of 'OED for PoIs' and 'OED for QoIs'. Similarly, the policy from 'OED for PoIs' achieves the highest EIG estimate on PoIs, with model-PoIs slightly lower; and the policy from 'OED for QoIs' achieves the highest EIG estimate on QoIs, with model-QoIs slightly lower.

The difference in trajectories across five OED scenarios, as shown in Fig. 3.7, along with the superior EIG performance of each single-goal OED scenario for its respective target, as illustrated in Table 3.2, both highlight the importance of selecting the correct design criterion as intended for the experiments being designed.

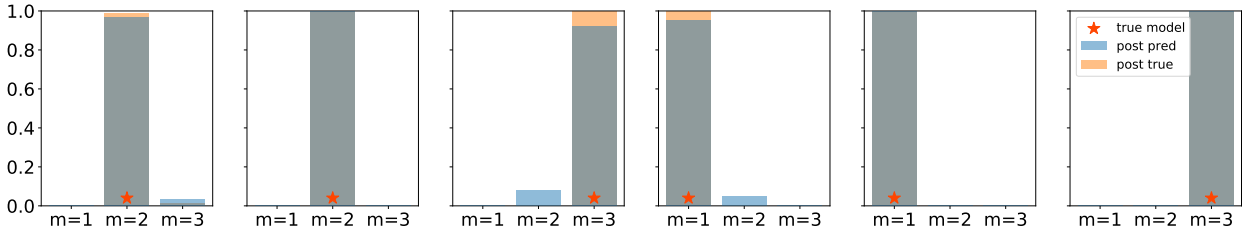


Figure 3.8: Case 1b. Examples of approximate posterior and true posterior for model indicator using policy resulting from 'OED for model indicator', for $T = 30$.

Table 3.2: Case 1b. PCE estimate of the EIG on model indicator, PoIs, and QoIs, using policies resulting from the five OED scenarios. ± represents standard error.

	EIG on model indicator	EIG on PoIs	EIG on QoIs
OED for model indicator	**1.020** ± 0.003	5.956 ± 0.065	1.539 ± 0.027
OED for PoIs	0.896 ± 0.005	**10.567** ± 0.048	1.587 ± 0.081
OED for QoIs	0.815 ± 0.005	6.999 ± 0.053	**2.388** ± 0.037
Model-PoIs OED	0.950 ± 0.005	10.330 ± 0.049	2.026 ± 0.065
Model-QoIs OED	0.967 ± 0.004	7.830 ± 0.052	2.298 ± 0.032

Hyperparameters and training stability Hyperparameter settings and training stability results can be found in Appendix B.3.2.

3.3.3 Case 2: constant elasticity of substitution

The CES case involves a single model and 'OED for PoIs' only, and has been previously studied in [50, 51, 18]. CES originates from behavioral economics, where participants are presented with two baskets of goods and asked to assess the subjective difference in utility between the two baskets. The CES model [7] represents the underlying utility function with latent PoIs $\Theta = \{\rho, \beta, \log u\}$, where $\beta \in \mathbb{R}^3$ while ρ and u are scalars. We adopt independent priors $\rho \sim \text{Beta}(1, 1), \beta \sim \text{Dirichlet}([1, 1, 1]), \log u \sim \mathcal{N}(1, 3^2)$. Note that the degree of freedom for β is 2 since the sum of its components is constraint to $\beta_1 + \beta_2 + \beta_3 = 1$; hence, only β_1 and β_2 are included in Θ. The design variables are $\xi_t = [\xi_{t,x}, \xi_{t,x'}]$ and constrained to the design space $\xi_{t,x} \in [0, 100]^3$ and $\xi_{t,x'} \in [0, 100]^3$. The observation model is

$$Y_t = \text{clip}\Big(\text{sigmoid}(\varphi_t), \epsilon, 1 - \epsilon\Big), \tag{3.34}$$

where the clipping threshold parameter is set to $\epsilon = 2^{-22}$, and $\varphi_t \sim \mathcal{N}\left(\mu_\varphi, \sigma_\varphi^2\right)$ with

$$\mu_\varphi = u\left[\left(\sum_{i=1}^{3}(\xi_{t,x,i})^\rho \beta_i\right)^{\frac{1}{\rho}} - \left(\sum_{j=1}^{3}(\xi_{t,x',j})^\rho \beta_j\right)^{\frac{1}{\rho}}\right], \qquad \sigma_\varphi = \tau u\Big(1 + \|\xi_{t,x} - \xi_{t,x'}\|_2\Big),$$

and $\tau = 0.005$. While the observation model in Eqn. (3.34) differs from the form in Eqn. (3.1), the likelihood remains explicit and can be evaluated by applying a change of variable to the sigmoid function and integrating the tail probabilities beyond the clipped thresholds; we do not write the likelihood formula here for brevity, and refer interested readers to the associated vsOED code implementation on Github.

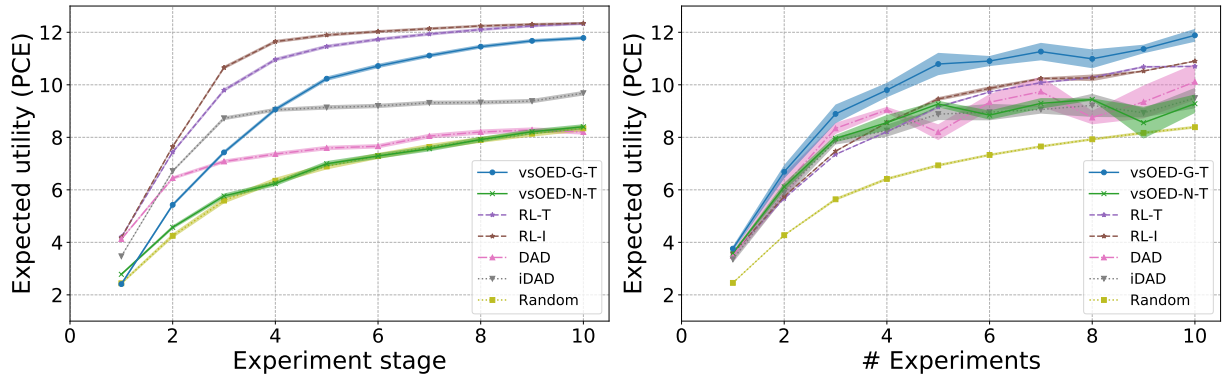
Results We use vsOED only with TIG since the horizons of this problem are relatively short, at most $T = 10$, and Fig. 3.1b indicates that TIG and IIG perform similarly under those horizons. Figure 3.9a presents the expected cumulative utility, evaluated using PCE, at different experiment stages t for policies optimized for a horizon of $T = 10$ experiments under the fully trained setting. The plot suggests that vsOED with GMM and TIG is slightly inferior to fully-trained RL but significantly better than fully-trained DAD and iDAD. The lower values for vsOED with GMM and TIG in the earlier stages suggest that the policy prioritizes long-term expected utility accumulation over short-term rewards.

Figure 3.9b presents the average expected utility, averaged over four training replicates where each replicate is again evaluated using PCE, for policies trained under different design horizons T under the limited budget setting. Here vsOED with GMM and TIG outperforms other approaches

for all values of T. The standard errors are plotted as shaded regions and illustrate a greater robustness of vsOED-G-T and RL over vsOED-N-T and DAD.

For this case, vsOED with NFs appears significantly worse compared to vsOED with GMM. This is like due to NFs having more difficulty in representing distributions of random variables with compact support, ρ and β in this case.

Figure 3.10 provides examples comparing the GMM approximate posterior and the true posterior. GMM performs well for $\log u$, but tends to produce wider posteriors for ρ and β. This is likely due to that many observations are clipped at the two ends per Eqn. (3.34) which leads to numerous observations with identical values, and the non-uniqueness makes it difficult for GMM to learn the mapping from designs and observations to the posterior. Despite this challenge, vsOED with GMM is still able to find good policies.



(a) Expected cumulative utility at different experiment stage t for policies optimized for $N = 10$

(b) Average expected utility over four training replicates versus design horizon T

Figure 3.9: Case 2. Expected utility comparisons using policies resulting from different algorithms. The shaded regions represent the standard error.

Hyperparameters and training stability Hyperparameter settings and training stability results can be found in Appendix B.3.3.

3.3.4 Case 3: SIR model for disease spread

The SIR case involves a single model and 'OED for PoIs' only, and has been previously studied in [80]. As the model involves solving a stochastic differential equation (SDE), the case tests the ability of vsOED to handle an expensive forward model and implicit likelihood. SIR is a stochastic model [87, 34] describing the spread of infectious diseases in a population. Individuals in the population are divided into three categories: susceptible, infected, and recovered. The number

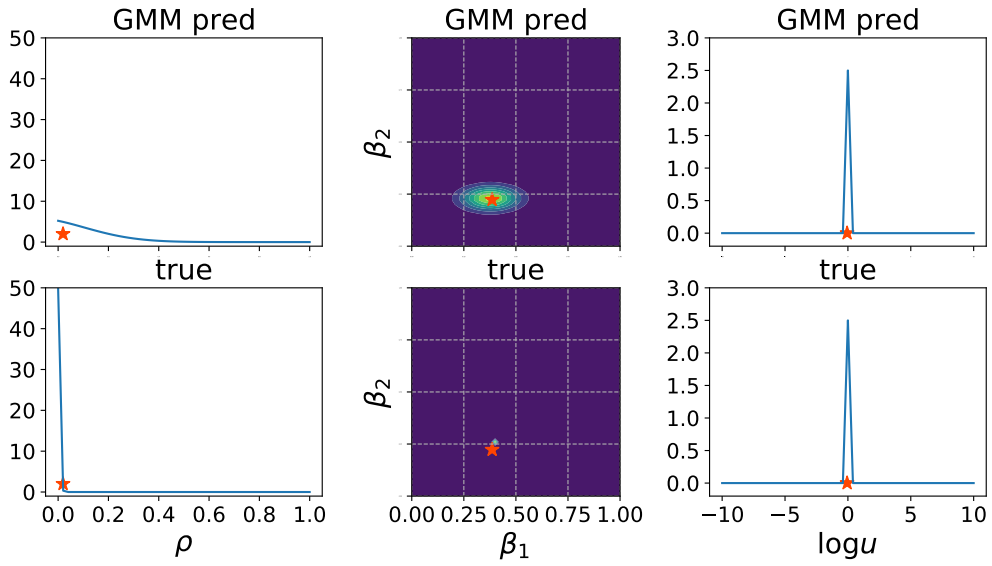


Figure 3.10: Case 2. Examples of GMM approximate posterior and true posterior at $T = 10$. The red stars indicate the true data-generating parameter values for these trajectory samples.

of individuals in these categories at any given time k are respectively represented by $S(k)$, $I(k)$, and $R(k)$, which always sum to a fixed population size $S(k) + I(k) + R(k) = N_p, \forall t$. Thus, the population state vector is defined to be $X(k) = [S(k), I(k)]$, where $R(k)$ is implied by the population constraint and can be ignored.

An individual at time k in the susceptible category has a probability to become infected controlled by rate parameter β, an individual who is infected has a probability to recover controlled by rate parameter ρ, and an individual who is already recovered remains recovered. The PoIs $\Theta = \{\log \beta, \log \rho\}$ entail the logarithm of the rates to ensure their non-negativity. We adopt independent priors $\log \beta \sim \mathcal{N}\left(\log 0.5, 0.5^2\right), \log \rho \sim \mathcal{N}\left(\log 0.1, 0.5^2\right)$. The design variable for the tth experiment is $\xi_t \in [0, 100]$, which is the time chosen for observing the number of infected $I(k = \xi_t)$. The designs are also constrained by $\xi_t \leq \xi_{t+1}$. The overall SIR model is often defined by a continuous-time Markov chain and can be sampled via the Gillespie algorithm [2], but it generally yields discrete population states that have undefined gradients. To circumvent these limitations, we follow [80] and adopt a continuous-state alternative (i.e., where $X(t)$ can be real-valued) form based on an SDE formulation.

In this SIR model formulation, the state dynamics follows the Itô SDE:

$$\mathrm{d}X(k) = f(X(k))\,\mathrm{d}k + D(X(k))\,\mathrm{d}W(k), \tag{3.35}$$

where $W(k)$ is a vector of independent Wiener processes (i.e., Brownian motion), and f and D are

the state-dependent drift vector and diffusion matrix, respectively, defined as [87]

$$f(X(k)) = \begin{bmatrix} -\beta \frac{S(k)I(k)}{N_p} \\ \beta \frac{S(k)I(k)}{N_p} - \rho I(k) \end{bmatrix}, \qquad D(X(k)) = \begin{bmatrix} -\sqrt{\beta \frac{S(k)I(k)}{N_p}} & 0 \\ \sqrt{\beta \frac{S(k)I(k)}{N_p}} & -\sqrt{\rho I(k)} \end{bmatrix}. \tag{3.36}$$

Given Eqn. (3.35) and (3.36), we can simulate state $X(k)$ by solving the SDE via finite-differencing, such as the Euler–Maruyama method; here we directly adopt the solver from [80]. For a fair comparison, we follow [80] and use the solutions of Eqn. (3.35) as data and do not consider an additional Poisson observational model that increases the noise in simulated data as suggested in [87]—that is, $Y_k = I(\xi_k)$. The likelihood is implicit for this case, since computing the probability of all stochastic transitions leading to the observed value would be intractable. Hence, we can sample from the observation model, but not evaluate its probability density.

Solving the SDE is quite computationally expensive, we thus limit the computational budget to 1 million trajectory samples for both vsOED and iDAD. To accelerate the training process, we pre-generate and store 1 million simulations and the access the stored simulations during the training. A new set of 3×10^5 simulations are used as evaluation data.

Results Only vsOED and iDAD are used in this case since the other algorithms cannot handle implicit likelihood. We use vsOED only with TIG since the horizons of this problem are relatively short, at most $T = 10$, and Fig. 3.1b indicates that TIG and IIG perform similarly under those horizons. Figure 3.11 presents the average $\tilde{U}$, averaged over four training replicates, for policies trained under different design horizons T. vsOED and iDAD appear to perform similarly, with vsOED slightly better at some spots. However, we note that the comparison is not entirely commensurable since iDAD is trained based on a different EIG lower bound estimator than $\tilde{U}$ (which vsOED uses and is also used here for evaluation), and that iDAD additionally uses the forward model derivatives. The absence of requiring model derivatives in vsOED can be potentially valuable in situations where model derivatives are inaccessible.

Figure 3.12 presents examples of infected state trajectory $I(k)$ and corresponding designs (observation times) ξ_t for $T = 10$ with three realizations of (β, ρ) and different ratios $r = \beta/\rho$. We observe that smaller r corresponds to a more spread out design of observation times, which aligns with the results in [80].

Hyperparameters and training stability Hyperparameter settings and training stability results can be found in Appendix B.3.4.

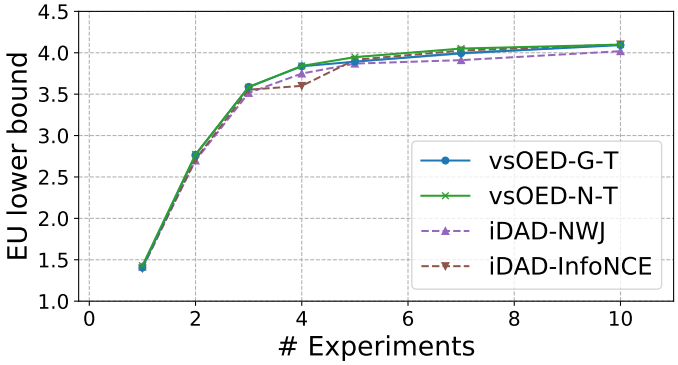
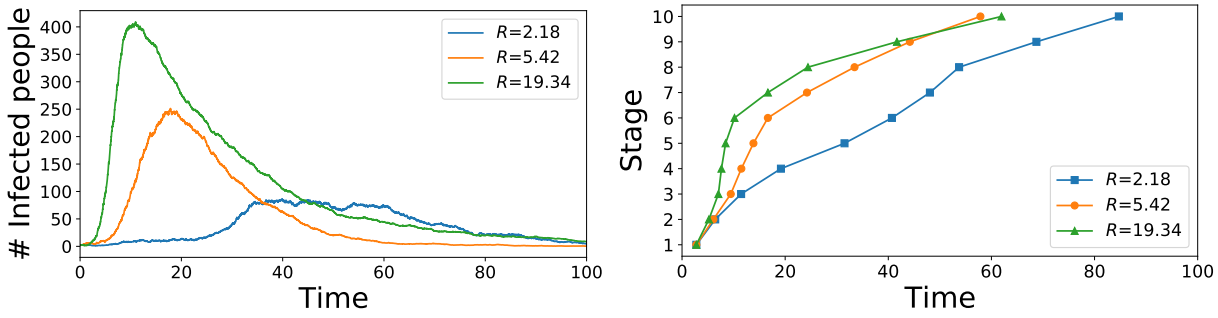


Figure 3.11: Case 3. Average $\tilde{U}$ over four training replicates versus design horizon T. The shaded regions represent the standard error.



(a) $I(t)$ (b) ξ_t

Figure 3.12: Case 3. Examples of infected state trajectory $I(k)$ and corresponding designs (observation times) ξ_t for $T = 10$ with three realizations of (β, ρ) and different ratios $r = \beta/\rho$.

3.3.5 Case 4: convection-diffusion-reaction

The convection-diffusion-reaction case involves nuisance parameters and physical state, and has been previously studied in [134] under the single-model setting. The case entails designing mobile sensor movements within a contaminant plume whose dynamics are governed by a 2D convection-diffusion-reaction partial differential equation (PDE). In this work, we consider three candidate models with uniform model prior $P(m) = 1/3$ for $m = 1, 2, 3$. The mth model contains m Gaussian-profiled contaminant sources randomly located in a 2D domain. The PoIs are the source locations $\Theta_m = \{\Theta_{m,i}\}_{i=1}^m$ where $\Theta_{m,i} = [\Theta_{m,i,x}, \Theta_{m,i,y}] \in \mathbb{R}^2$ denotes the location of the ith source and is endowed with independent priors $\Theta_{m,i,x}, \Theta_{m,i,y} \sim \mathcal{U}(0, 1^2)$. For the mth model, the contaminant concentration G at location $[x, y]$ and time k is governed by

$$\frac{\partial G(\Theta_m, \Psi_m, x, y, k; m)}{\partial k} = \nabla^2_{(x,y)} G - u(\Psi_m) \cdot \nabla_{(x,y)} G + S(\Theta_m, x, y, k; m),$$

for $[x, y] \in [-1, 2]^2$ and $0 < k \leq 0.2$. Here $u = [v \cos\beta, v \sin\beta]^\top \in \mathbb{R}^2$ is a time-invariant convection velocity that is described by nuisance parameters $\Psi_m = \{v, \beta\}$ with independent priors on the convection speed (magnitude) $v \sim \mathcal{U}(0, 20)$ and convection angle $\beta \sim \mathcal{U}(0, 2\pi)$. The source function is

$$S(\Theta_m, x, y, k; m) = \sum_{i=1}^{m} \frac{s}{2\pi h^2} \exp\left(-\frac{\|\Theta_{m,i} - [x, y]\|_2^2}{2h^2}\right),$$

where $s = 2$ is the known source strength and $h = 0.05$ is the known source width. The initial condition is $G(\Theta_m, \Psi_m, x, y, 0; m) = 0$ and homogeneous Neumann boundary conditions are applied to all sides of the computational domain. The design variables $\xi_t = [\xi_{t,x}, \xi_{t,y}] \in [0, 1]^2$ entail selecting the location of concentration measurement within the allowable design space at prescribed time intervals $k_t = 0.01(t + 1)$. The observation model is then $Y_t = G(\Theta_m, \Psi_m, \xi_{t,x}, \xi_{t,y}, k_t; m) + \mathcal{E}_t$ where $\mathcal{E}_t \sim \mathcal{N}(0, \sigma_\epsilon^2)$ with $\sigma_\epsilon = 0.05$. Moreover, the measurement sensor is initially located at $[x_0, y_0] = [0.5, 0.5]$, and a sensor movement penalty of $-0.1\|\xi_t - \xi_{t-1}\|_2$ for $t = 1, \ldots, T - 1$ (and $-0.1\|\xi_t - [x_0, y_0]\|_2$ for $t = 0$) is incurred in the immediate rewards to reflect the cost of moving the sensor.

Similar to Sec. 3.3.2, we are interested in a QoI that is the log of the integrated flux magnitude crossing the right boundary $x = 1$ of the design space (i.e., spanning from $y = 0$ to $y = 1$) at a future time $k = 0.2$. The QoI is $Z_m = \log(\|\varphi(\Theta_m, \Psi_m; m)\|)$, where

$$\varphi(\Theta_m, \Psi_m; m) = \int_0^1 -\frac{\partial G(\Theta_m, \Psi_m, x = 1, y, k = 0.2; m)}{\partial x} \, \mathrm{d}y.$$

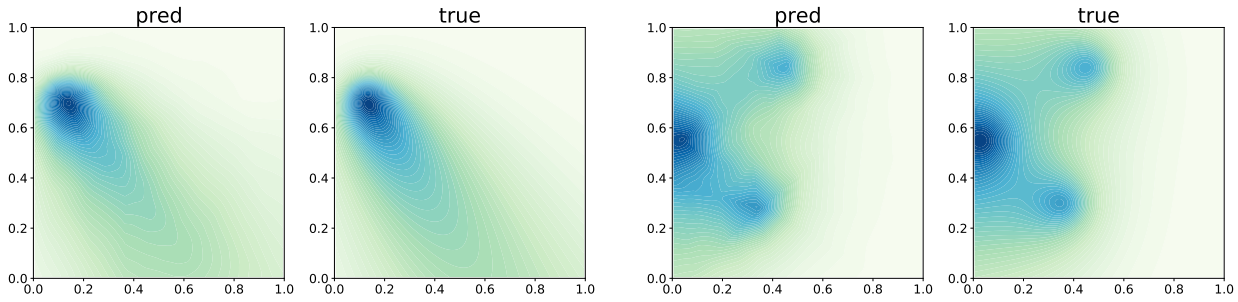
We apply a second-order finite volume method to solve the PDE numerically. The computational domain $[-1, 2]^2$ is discretized into a uniform grid with cell size $\Delta x = \Delta y = 0.01$. Second-order fractional step method is used for time integration, with a step size $\Delta k = 5.0 \times 10^{-4}$. The integrated flux φ is computed by first estimating the derivative term using finite difference on the grid values at the target boundary, and then numerically integrated using midpoint rule.

While we can directly use the finite volume solver as the forward model, it can be expensive and inefficient since each solve calculates the concentrations for all $[x, y]$ and all k, while in the OED problem only a small subset of these values are needed for each forward model evaluation. Therefore, to accelerate computations, we pre-build NN surrogate models to replace $G(\Theta_m, \Psi_m, \xi, k_t; m)$ and $\varphi(\Theta_m, \Psi_m; m)$, one for each m and k_t. The architectures of these NN surrogates can be found in Appendix B.3.5.

A separate G surrogate is built for each m and k_t, and a separate φ surrogate is built for each m. For each m, 20,000 simulations are generated with random samples of Θ_m and Ψ_m, with 18,000 used for training and 2,000 for testing. The testing mean squared errors (MSE) are shown in Table 3.3 for G surrogates at the end time $k = 0.2$ and the three φ surrogates, illustrating excellent accuracy. Figure 3.13 presents examples of the true and surrogate concentration fields G at the end time $k = 0.2$; they demonstrate excellent agreement.

Table 3.3: Case 4. Testing MSE of the surrogate models.

Model	Surrogate G at $k = 0.2$	Surrogate φ
$m = 1$	3.094×10^{-5}	4.141×10^{-5}
$m = 2$	3.284×10^{-4}	4.986×10^{-4}
$m = 3$	1.650×10^{-3}	2.080×10^{-3}



(a) $m = 1$ (b) $m = 3$

Figure 3.13: Case 4. Examples of the true and surrogate concentration fields G at $k = 0.2$.

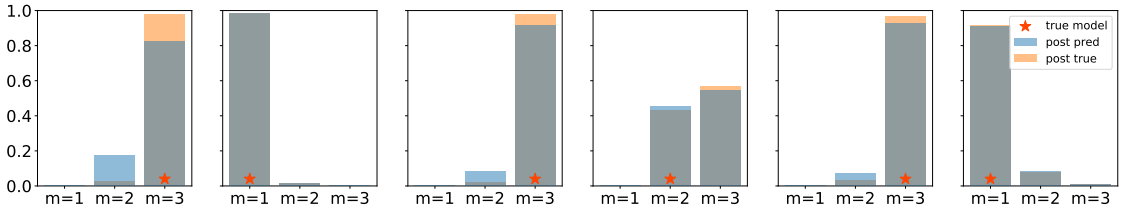
Results Only vsOED is used in this case since the other algorithms cannot handle nuisance parameters and multiple models. We use vsOED only with TIG since the horizons of this problem are relatively short, at most $T = 15$, and Fig. 3.1b indicates that TIG and IIG perform similarly under those horizons. Only the GMM version of vsOED is presented for brevity. Similar to Case 1b, we consider five OED scenarios: OED for model indicator (i.e., design for model discrimination, $\alpha_M = 1, \alpha_\Theta = \alpha_Z = 0$), OED for PoIs (i.e., design for parameter inference, $\alpha_\Theta = 1, \alpha_M = \alpha_Z = 0$), OED for QoIs (i.e., design for goal-oriented prediction, $\alpha_Z = 1, \alpha_M = \alpha_\Theta = 0$), OED for both model indicator and PoIs ('model-PoIs', $\alpha_M = \alpha_\Theta = 1, \alpha_Z = 0$), and OED for both model indicator and QoIs ('model-QoIs', $\alpha_M = \alpha_Z = 1, \alpha_\Theta = 0$).

Figure 3.14 presents examples comparing the GMM approximate posterior and true posterior for model indicator and PoIs (from $m = 1$) using policies respectively resulting from 'OED for model indicator' and 'OED for PoIs' for $T = 10$; the GMMs again approximate the posteriors well for this case. Notably, an example PoI posterior and the corresponding nuisance parameter posterior are shown in Fig. 3.14b, both plotted over the full support of their respective uniform priors. We see that the PoI posterior has shrunk more substantially than that for the nuisance parameter, which is consistent with the expected behavior of 'OED for PoIs' that targets to reduce the PoI uncertainty. Figure 3.15 presents examples of trajectory using policies from the five OED scenarios for $T = 10$. The overall policy behavior is similar to the multi-model source location finding problem of Case 1b. Trajectories from 'OED for model indicator' tend to explore more and appear more diffuse compared to the other scenarios, often extending to the boundaries of the design space and leading to narrow posteriors of model indicator as shown in Fig. 3.14a. Trajectories from 'OED for PoIs' tend to be more exploitative, remaining closer to the estimated sources while leveraging the background convection. Trajectories from 'OED for QoIs' exhibit a vertical design tendency, due to the same effects as explained in Case 1b.

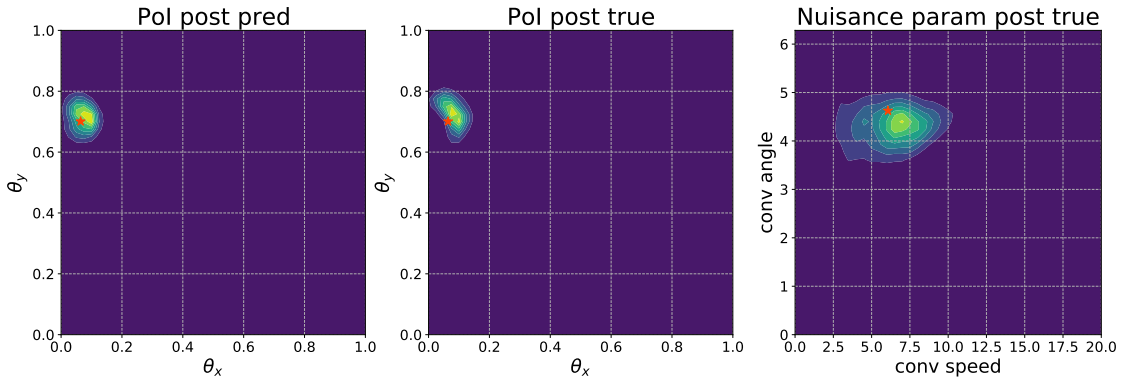
Hyperparameters, training stability, and surrogate architectures Hyperparameter settings, training stability, and surrogate architectures can be found in Appendix B.3.5.

3.4 Conclusions

We introduced vsOED, a novel method for solving Bayesian sequential OED problem using an actor-critic reinforcement learning framework powered by policy gradient techniques and the variational Barber–Agakov bound to the EIG. vsOED is capable of accommodating nuisance parameters, implicit likelihoods, and multiple candidate models, while supporting a flexible design criterion that can target designs for model discrimination, parameter inference, goal-oriented prediction, or their weighted combinations.

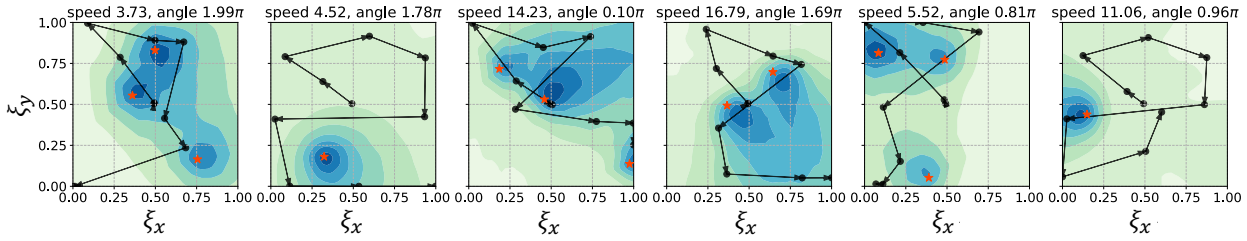


(a) Posteriors for model indicator from 'OED for model indicator'.

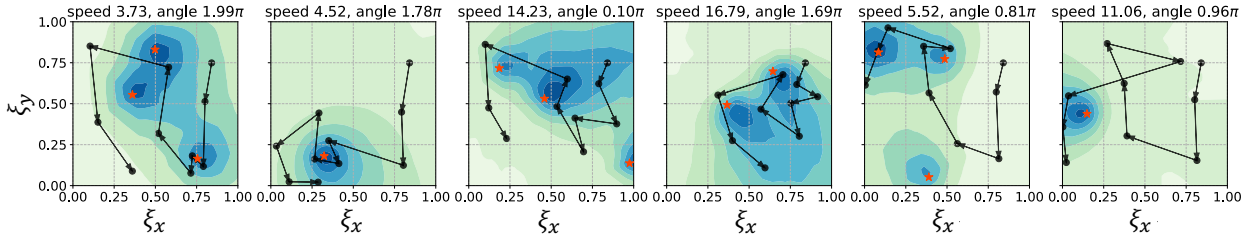


(b) Posteriors for PoIs from 'OED for PoIs' for $m = 1$, along with the true posterior for the nuisance parameters.

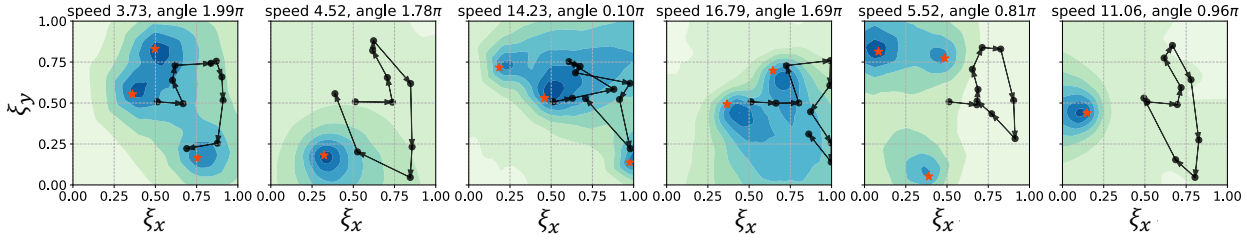
Figure 3.14: Case 4. Examples of approximate posterior and true posterior for model indicator and PoIs using policy resulting from 'OED for model indicator' and 'OED for PoIs', respectively, for $T = 10$.



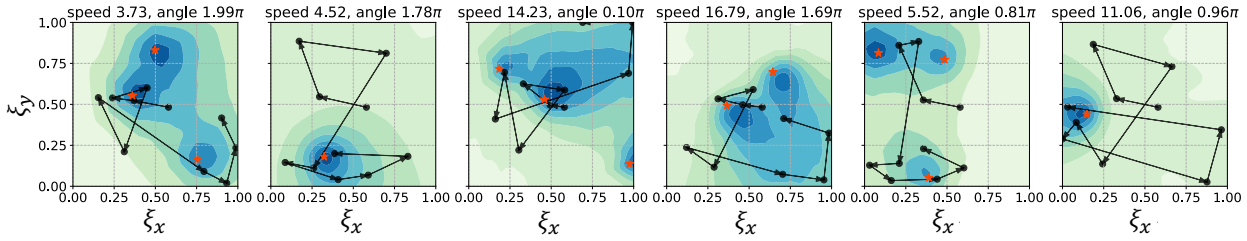
(a) OED for model indicator



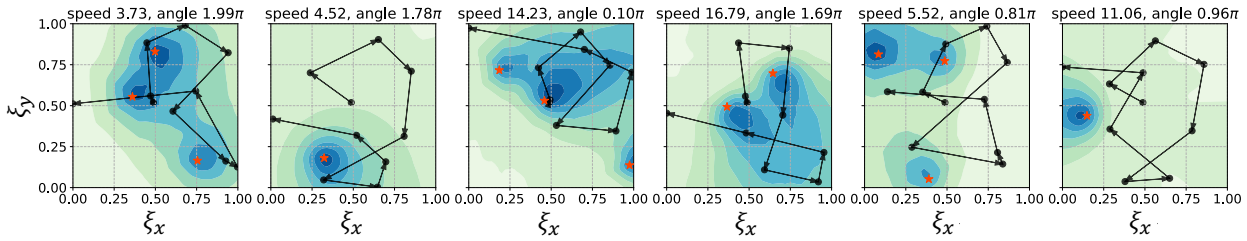
(b) OED for PoIs



(c) OED for QoIs



(d) Model-PoIs OED



(e) Model-QoIs OED

Figure 3.15: Case 4. Examples of policy trajectory for $T = 10$. The background contour plots the true contaminant concentration at the end time $t = 0.2$, and the red stars indicate the true source locations.

We provided key theoretical results including theorems proving the equivalence between incremental and terminal information gain rewards, equivalence between EIG and its one-point reward formulations, and the lower bound property of vsOED when posteriors are approximated. We then developed numerical methods under an actor-critic framework, deriving and estimating the policy gradient and utilizing Gaussian mixture models and normalizing flows to approximate the posteriors.

Finally, we demonstrated vsOED across four numerical cases—source location finding, constant elasticity of substitution, SIR model for disease spread, and convection-diffusion-reaction. These scenarios involved challenges of multiple models, implicit likelihoods, and nuisance parameters. Comparisons with existing algorithms (DAD, iDAD, and RL) showed that vsOED achieves superior sample efficiency under limited budgets and avoids reliance on model derivatives. The comparison of trajectories and EIGs across various design scenarios further underscored the importance of adopting a generalized OED framework, as focusing on the correct design criterion can lead to vastly different policy outcomes.

Despite its strengths, vsOED has limitations. Its performance is sensitivity to inaccurate posterior approximations, especially for posteriors with compact support, multiple modes, and highly non-Gaussian features. Another limitation is that the current examples involve parameters in relatively low-dimensional spaces, motivating the need to explore and strengthen vsOED to high-dimensional settings. For instance, [4] applied the variational posterior from [98] to sequential causal discovery, where dimensionality scales quadratically with the number of nodes in a causal graph due to the binary adjacency matrix representation. Elsewhere, [113] demonstrated non-sequential EIG estimation for high-dimensional parameters (e.g., 320×320 images) using conditional NFs. Together, these examples highlight a key direction for future research: the development of more accurate and scalable posterior representations to enable sequential OED in high-dimensional regimes. At the same time, as with all EIG-bound approaches, a tighter bound does not automatically translate to a "more optimal" policy. Further investigation into the relationship between the quality of a bound and the optimality of the resulting policy would be highly valuable. Additionally, vsOED currently does not handle discrete designs or stochastic policies, which could expand its applicability, as demonstrated by [18]. Further enhancements could also stem from advanced reinforcement learning techniques, such as proximal policy optimization, trust region policy optimization, and soft actor-critic [131, 130, 65, 54].

CHAPTER 4

Goal-Oriented Sequential Bayesian Experimental Design for Causal Learning

The vsOED framework developed in the previous chapter provides a general methodology for learning adaptive experimental policies under uncertainty. While this framework is applicable to a broad class of inference problems, its flexibility becomes particularly valuable in causal learning settings, where experiments correspond to interventions and the space of scientific objectives is often task-dependent.

In causal applications, experimental design is rarely driven by the goal of fully recovering the underlying causal model. Instead, practitioners are frequently interested in answering specific causal questions, such as estimating the effect of a particular intervention, identifying a subset of relevant causal mechanisms, or resolving ambiguity among competing causal hypotheses. These objectives naturally motivate a goal-oriented formulation of experimental design, in which experiments are selected to maximally reduce uncertainty with respect to a targeted causal quantity rather than the entire model.

Research gap and contributions. Existing Bayesian optimal experimental design (BOED) approaches for causal learning predominantly focus on maximizing information about the full causal model, even when the scientific objective concerns a specific causal query such as an intervention effect or counterfactual outcome. Moreover, many causal design strategies rely on greedy or one-step lookahead policies, which fail to exploit long-horizon complementarities across interventions and are poorly suited to mixed discrete–continuous design spaces and high-dimensional relational observations. The central research question addressed in this chapter is how to design a scalable, non-myopic experimental design framework that directly targets user-defined causal quantities of interest while remaining computationally viable for complex causal models. To address this challenge, this chapter introduces goal-oriented causal Bayesian experimental design (GO-CBED),

which integrates causal semantics, variational EIG objectives, and sequential decision-making within a unified framework. The key contributions include (i) a principled formulation of goal-oriented EIG for causal queries under interventions, (ii) a sequential, non-myopic policy-learning approach that plans over entire intervention trajectories, and (iii) an amortized, transformer-based policy architecture that enables real-time intervention selection in high-dimensional causal learning settings.

The contents of this chapter can be found in the manuscript [165]. The author gratefully acknowledges Zheyu Zhang for his substantial contributions to the development of this work.

4.1 Introduction

A structural causal model (SCM) provides a mathematical framework for representing causal relationships via a directed acyclic graph (DAG). SCMs are foundational across domains such as genomics and precision medicine [144], economics [151], and the social sciences [135, 79], where understanding the cause-effect relationships is central to scientific inquiry. Key tasks in causal modeling include: *causal discovery*, which learns the DAG structure; *causal mechanism identification*, which estimates functional dependencies; and *causal reasoning*, which answers interventional and counterfactual queries. All such tasks depend on data. While (passive) observational data can reveal correlational structures, they often fail to identify the true causal model [152]. In contrast, (active) *interventional* data are essential for uncovering causal relationships and estimating causal effects—but such experiments are inherently expensive and limited, making careful experimental design essential.

BOED addresses this challenge by selecting interventions that maximize the EIG and provides a principled framework for handling uncertainty in both causal structure and mechanisms. However, most existing causal BOED methods focus on *learning the full model*—for causal discovery or mechanism identification—regardless of the scientific goal.

In many real-world applications, the objective is more focused on *causal reasoning*: researchers aim to estimate the effect of a specific intervention, rather than recover the entire causal system. For instance, in drug discovery, it is often more important to understand how particular molecular targets influence disease pathways than to map the full biological network. Shown in Figure 4.1, optimizing for full model parameters (middle) leads to experiments that are misaligned with such targeted goals (left), resulting in inefficient use of resources (compared to right). This motivates a **goal-oriented** approach to BOED—one that tailors interventions to the specific causal queries that matter the most.

Recent work by [149] begins to address goal-oriented causal design, but adopts a myopic strategy—selecting only the next experiment without planning ahead. More broadly, most causal

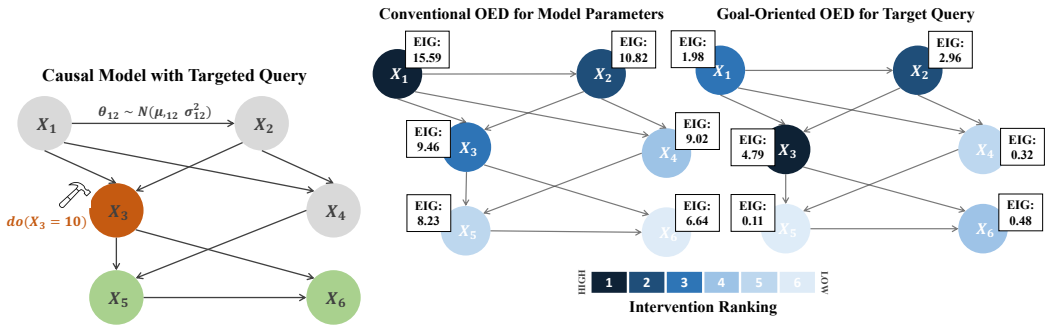


Figure 4.1: Illustration of goal-oriented versus conventional BOED for causal learning. *Left*: A linear Gaussian SCM with six nodes; the experimental goal is to estimate the causal effect of the intervention $\text{do}(Y_3 = 10)$ on node Y_5 and Y_6. *Middle*: Conventional BOED selects interventions that maximize EIG over all model parameters, resulting in Y_1 and Y_2 being selected as the best. *Right*: Our GO-CBED approach selects interventions by directly maximizing EIG for the specific causal query, leading to a different intervention that prioritizes nodes most relevant to the query, i.e., Y_3 and Y_2.

BOED approaches are greedy, optimizing interventions one step at a time without accounting for how early decisions influence future learning.

Overcoming this limitation requires a *non-myopic* framework, which we formulate as a Markov decision process and solve using tools from reinforcement learning (RL) [126, 75].

To address these challenges, we propose **Goal-Oriented Causal Bayesian Experimental Design (GO-CBED)**, a novel framework for sequential, non-myopic causal experimental design. Our key contributions include: a **goal-oriented framework** that substantially improves experimental efficiency for specific causal queries; a **sequential, non-myopic strategy** that captures synergies between interventions; and empirical results showing that **GO-CBED outperforms existing methods**.

4.2 Preliminaries

Structural Causal Models SCMs [116] provide a rigorous mathematical framework for representing and reasoning about cause-effect relationships. An SCM defines a collection of random variables $Y = \{Y_1, \ldots, Y_d\}$, structured by a DAG $G := \{V, E\}$. The SCM is denoted as $\mathcal{M} := \{G, \Theta\}$, where G encodes the causal structure and Θ parameterizes the causal mechanisms. Each variable Y_i is determined by its parents in the graph and an exogenous noise term via a structural equation:

$$Y_i = f_i(Y_{\text{pa}(i)}, \Theta_i; \mathcal{E}_i), \quad \forall i \in V. \tag{4.1}$$

Here, $Y_{\mathrm{pa}(i)}$ denotes the parent variables of Y_i in G, f_i is a causal mechanism parameterized by Θ_i, and $\mathcal{E}_i \sim P_{\mathcal{E}_i}$ is an independent noise variable. The SCM thus defines a joint distribution over Y, enabling causal reasoning and interventional analysis.

Interventions (Experimental Designs) SCMs support formal reasoning about interventions—i.e., external manipulations to the system. A *perfect* (or *hard*) intervention on a subset of variables Y_I, denoted by $\mathrm{do}(Y_I = s_I)$ [116], replaces the corresponding structural equations with fixed values s_I, modifying the data-generation process. This introduces an *interventional SCM*, which leads to a new distribution over the variables. Assuming causal sufficiency and independent noise [139], the interventional distribution follows the Markov factorization:

$$p(Y|\mathcal{M}, \xi) = \prod_{j \in V \setminus I} p(Y_j | Y_{\mathrm{pa}(j)}, \Theta_j, \mathrm{do}(Y_I = s_I)), \tag{4.2}$$

where the design variable $\xi := \{I, s_I\}$ encodes both the intervention target I and the intervention value s_I. Interventions form the foundation for both causal discovery (i.e., identifying G) and causal reasoning (i.e., estimating interventional effects).

Goal-Oriented Sequential Bayesian Framework Conventional causal BOED methods typically follow a two-step procedure: first, learn the full model, and then use it to answer causal queries. Such an approach can be inefficient when only a small subset of causal QoIs matter, as it may spend significant resources learning aspects of the model irrelevant to the target queries.

To address this inefficiency, we adopt a goal-oriented perspective: rather than learning the entire model, we design experiments to directly improve our ability to answer specific causal queries. We formalize this using a query function H that maps the causal model $\mathcal{M}$ to the desired quantity $Z = H(\mathcal{M}; \epsilon_Z)$, where ϵ_Z captures any inherent stochasticity in the query. For example, setting $z = G$ corresponds to causal discovery, while $Z = Y_i^{\mathrm{do}(Y_j = \psi_j)}$ corresponds to estimating the causal effect of setting $Y_j = \psi_j$ on Y_i from a distribution of possible intervention values $\psi_j \sim p(\psi_j)$.

At experiment stage t of a sequence of T experiments, let the history be $I_t := \{\xi_{0:t-1}, Y_{0:t-1}\}$, where ξ_τ and Y_τ denote the design and outcome of the τ-th interventional experiment. The belief over the causal model $\mathcal{M}$ is updated via Bayes' rule:[1]

$$p(G|I_{t+1}) = \frac{p(G|I_t)\, p(Y_t|G, I_t, \xi_t)}{p(Y_t|I_t, \xi_t)}, \qquad p(\theta|G, I_{t+1}) = \frac{p(\theta|G, I_t)\, p(Y_t|G, \theta, I_t, \xi_t)}{p(Y_t|G, I_t, \xi_t)},$$

where the marginal likelihood $p(Y_t|G, I_t, \xi_t)$ is computed by integrating over Θ.

[1] When observational data $\mathcal{D}$ is available prior to designing interventions, all distributions are implicitly conditioned on $\mathcal{D}$. See Appendix C.3.4 for further details.

However, our primary interest lies not in inferring the full model $\mathcal{M}$, but in updating beliefs about the target query Z. This is captured by the posterior-predictive distribution:

$$p(z|I_t) = \sum_G \int p(z|G, \theta, I_t)\, p(G|I_t)\, p(\theta|G, I_t)\, \mathrm{d}\theta. \tag{4.3}$$

4.3 Goal-Oriented sequential causal Bayesian experimental design

Problem Statement GO-CBED seeks an optimal policy $\pi : I_t \to \xi_t$ that maximizes the EIG on the target causal QoI Z over a sequence of T experiments, for simplicity, just consider the terminal reward:

$$\pi^* \in \underset{\pi}{\operatorname{argmax}} \left\{ \mathcal{I}_T(\pi) := \mathbb{E}_{p(\mathcal{M})p(I_T|\mathcal{M},\pi)p(Z|\mathcal{M})} \left[\log \frac{p(Z|I_T)}{p(Z)} \right] \right\}, \tag{4.4}$$

subject to the constraint that designs following the policy: $\xi_t = \pi(I_{t-1})$ for all t.

This formulation is *goal-oriented*, as it directly targets EIG on specific causal QoIs rather than the full model, and *non-myopic*, as it optimizes the entire sequence of interventions rather than selecting each greedily. An equivalent formulation based on incremental (stage-wise) EIG after each experiment is also possible; see Sec. C.1.1 for details.

The EIG on Z defined in (4.4), $\mathcal{I}_T$, is also the mutual information between Z and I_T. When Z is a bijective function of $\mathcal{M}$, maximizing EIG on Z is equivalent to maximizing it on $\mathcal{M}$ [15]. However, when Z is not invertible with respect to $\mathcal{M}$, directly maximizing EIG on Z becomes more efficient. It avoids expending effort on irrelevant parts of $\mathcal{M}$, reducing both computational and experimental costs—especially beneficial when dealing with large causal graphs or tight intervention budgets.

4.3.1 Variational lower bound

Evaluating and optimizing the EIG in (4.4) requires estimating the posterior density $p(Z|I_T)$, which is generally intractable for complex causal models. To address this, we adopt a *variational approach* that approximates the posterior using $q(Z|\pi, I_T; \phi)$, where ϕ is the variational posterior and embedding network parameter for historical interventional data:

$$\mathcal{I}_{T;\,L}(\pi; \phi) := \mathbb{E}_{p(\mathcal{M})p(I_T|\mathcal{M},\pi)p(Z|\mathcal{M})} \left[\log \frac{q(Z|I_T; \phi)}{p(Z)} \right], \tag{4.5}$$

subject to $\xi_t = \pi(I_t)$ for all t.

Theorem 4.3.1 (Variational Lower Bound)**.** *For any policy π, variational and embedding parameter ϕ, the EIG satisfies $\mathcal{I}_T(\pi) \geq \mathcal{I}_{T;L}(\pi;\phi)$. The bound is tight if and only if $p(z|\mathfrak{i}_T) = q(z|\mathfrak{i}_T;\phi)$ for all z and $\mathfrak{i}_T$.*

A proof is provided in Appendix C.1.2.

Since $p(z)$ is independent of π and ϕ, it can be omitted from the optimization statement without affecting the argmax. Thus, maximizing the EIG lower bound reduces to maximizing the *prior-omitted* EIG bound:

$$\pi^*, \phi^* \in \underset{\pi,\phi}{\operatorname{argmax}} \left\{ \mathcal{R}_{T;L}(\pi;\phi) := \mathbb{E}_{p(\mathcal{M})p(I_T|\mathcal{M},\pi)p(Z|\mathcal{M})}\big[\log q(Z|I_T;\phi)\big] \right\}, \tag{4.6}$$

where $\mathcal{R}_{T;L}(\pi;\phi) \leq \mathcal{R}_T(\pi) := \mathbb{E}_{p(\mathcal{M})p(I_T|\mathcal{M},\pi)p(Z|\mathcal{M})}[\log p(Z|I_T)]$ is a lower bound to the prior-omitted EIG $\mathcal{R}_T(\pi)$. See Appendix C.1.3 for additional information on $\mathcal{R}_{T;L}(\pi;\phi)$.

4.3.2 Variational posteriors and policy network

Having established the theoretical foundation of GO-CBED, we now describe its implementation. Our approach comprises two key components: (1) variational posteriors for establishing the EIG lower bound, and (2) a policy network that guides the intervention selection process.

Variational Posteriors While GO-CBED supports arbitrary causal queries, we focus on two fundamental tasks that form the basis of our experimental evaluation: causal reasoning (i.e., estimating interventional effects) and causal discovery (i.e., learning graph structure).

For causal reasoning tasks, where the query takes the form $Z = Y_i^{\mathrm{do}(Y_j=\psi_j)}$, the posterior distribution $p(Z|I_T)$ is often complex and multimodal due to the structural uncertainty—different causal graphs can imply different causal effects for the same intervention. To capture this complexity, we parameterize the variational posterior $q(z|I_T;\phi)$ using normalizing flows (NFs).

For causal discovery tasks, where the query is the graph itself, $z = G$, we model the posterior over graph structures using an independent Bernoulli distribution for each potential edge [98]:

$$q(G|I_T;\phi) = \prod_{i,j} q(G_{i,j}|I_T;\phi), \tag{4.7}$$

where each $q(G_{i,j}|\cdot)$ is approximated using Bernoulli distribution. This parameterization allows efficient modeling of the posterior over DAG structures, while maintaining scalability and differentiability.

Algorithm 2: The GO-CBED algorithm.

Input: H, $p(\Psi)$; prior $p(G)$, $p(\Theta|G)$; likelihood $p(Y_t|G, \Theta, \xi_t)$; number of experiments T;
Initialize policy network parameters γ, variational and embedding parameters ϕ;
for $l = 1, \ldots, n_{\text{step}}$ **do**
    Simulate n_{env} samples of G, θ, ψ and z;
    for $t = 0, \ldots, T$, **do**
        Compute $\xi_t = \pi(I_{t-1})$, then sample $Y_t \sim p(Y_t|G, \theta, \xi_t)$;
    end for
    Update γ, and ϕ following gradient ascent, where gradient obtained from auto-grad on $\mathcal{R}_{T;L}$;
end for
Output: Optimal π^* parameterized by γ^*, and ϕ^*;

Policy Network We represent the policy π using a neural network with parameters γ. The policy network selects the next intervention by mapping the history I_{t-1} to a design ξ_t at stage t. The architecture is designed to satisfy two symmetry properties that have been shown to improve performance [4]: permutation invariance across history samples and permutation equivariance across variables.

The network is composed of L transformer layers that alternate between attention over variable and observation dimensions. This alternating structure enables rich and efficient information flow across the entire history of interventions and outcomes. The final embedding is passed through two output heads: one that produces the intervention targets I_t, using the Gumbel-softmax trick to enable differentiability for discrete variables, and the other predicts the corresponding intervention values s_{I_t}. The architecture is illustrated in Appendix C.3.1.

Training Procedure Algorithm 2 outlines our training procedure, which jointly optimizes the policy parameters and variational parameters by maximizing the variational lower bound. In each training iteration, we sample causal models $\mathcal{M} = (G, \Theta)$ from the prior, derive the target QoIs Z, simulate intervention trajectories, and update all network parameters via gradient ascent. At deployment time, only forward passes through the policy network are required, eliminating the need for online Bayesian inference. This enables real-time decision-making with constant computational complexity, independent of experiment sequence length.

4.4 Numerical experiments

Our numerical experiments demonstrate how GO-CBED advances causal experimental design through goal-oriented optimization. We begin in Section 4.4.1 to recap the motivating example from Figure 4.1 that illustrates the fundamental advantage of targeting specific causal queries over full model learning. We then focus on two key causal tasks: Section 4.4.2 examines causal reasoning, where we evaluate performance in estimating targeted causal effects across both synthetic causal models and semi-synthetic gene regulatory networks derived from the Dialogue for Reverse Engineering Assessments and Methods (DREAM) benchmarks [64]; Section 4.4.3 then turns to causal discovery, comparing GO-CBED against existing causal BOED baselines on similar synthetic and semi-synthetic settings.

4.4.1 Motivating example with fixed graph structure

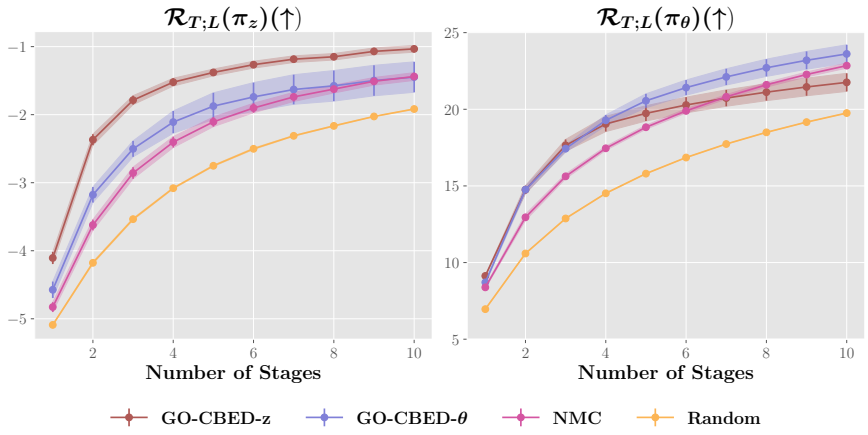


Figure 4.2: Performance comparison of policies trained for $T = 10$, evaluated across different stage lengths. *Left*: Performance on causal query $z = \{Y_5, Y_6 \mid \mathrm{do}(Y_3 = 10)\}$. *Right*: Performance on model parameters $z = \{\theta \setminus \theta_{\mathrm{pa}(3)}\}$. While GO-CBED-$\theta$ achieves higher EIG on the task centering parameters (right), it performs significantly worse than GO-CBED-z on the causal reasoning task (left). Shaded regions represent ± 1 standard error across 4 random seeds.

We first evaluate the benefits of goal-oriented policies on the motivating example with a fixed graph structure in Figure 4.1. This setup assumes a linear-Gaussian relationship between variables, allowing for analytical posterior computation and accurate EIG estimation. We compare four policies: **GO-CBED**-z, which is optimized for the specific causal query; **GO-CBED**-θ, which targets model parameters; **NMC**, the NMC estimator for the prior-omitted EIG on QoIs [149]; and **Random**, which selects both intervention targets and values uniformly at random. We evaluate their performance on the prior-omitted EIG (or lower bound) for Z. Full experiment details can be

found in Appendix C.3.2.

Figure 4.2 reveals a key insight: although GO-CBED-θ achieves higher EIG on model parameters, its performance on the actual causal query is substantially worse than that of GO-CBED-z. This supports our central argument—when the goal is to answer specific causal queries, policies that directly target those queries are significantly more efficient than those optimized for general model learning. Moreover, GO-CBED's variational formulation consistently outperforms the sampling-based NMC. This advantage is especially pronounced when the inner-loop sample size in NMC is small, where the estimator suffers from high bias (see Appendix C.4.1).

4.4.2 Causal reasoning tasks

We evaluate GO-CBED's ability to design interventions that maximize EIG with respect to specific causal queries, now no longer fixing the graph structure. We compare three policies: **GO-CBED**-z optimized directly for causal queries, **GO-CBED**-G trained for causal discovery, and **Random** selection. Additional experiment details are provided in Appendix C.3.3.

Synthetic SCMs Figure 4.3 compares performance using Erdös–Rényi (ER) and Scale-Free (SF) graph priors with linear and nonlinear mechanisms, detailed in Appendix C.3.3. In linear settings, GO-CBED-z and GO-CBED-G perform similarly, indicating that causal structure alone is often sufficient to accurate estimate simple causal effects. However, in nonlinear settings, GO-CBED-z significantly outperforms GO-CBED-G, despite the latter's strong performance on causal discovery tasks (see Appendix C.3.3). These results indicate a key insight: as the complexity of causal mechanisms increases, directly targeting causal queries becomes increasingly advantageous. Moreover, in nonlinear mechanisms parameterized by neural networks, the dimensionality of the weights θ is large and graph-dependent, making it challenging to generate sufficient samples to effectively tighten the EIG lower bound for full graphs and parameters. These are precisely the advantages that goal-oriented design aims to provide.

Semi-Synthetic Gene Regulatory Networks To assess real-world applicability, we evaluate GO-CBED on semi-synthetic gene regulatory networks derived from the DREAM [64] benchmarks, detailed in Appendix C.3.3. Figure 4.4 presents results on *E. coli* networks with nonlinear causal mechanisms ($d = 10$, $T = 10$). GO-CBED-z, which directly targets causal query, significantly outperforms both GO-CBED-G and Random baselines, especially after the initial stages of intervention. In addition, GO-CBED-G performs worse than Random despite achieving better causal discovery performance (see Appendix C.3.3), indicating that recovering the causal graph does not necessarily provide good uncertainty reduction for specific causal queries. This performance gap on biologically-inspired networks has important practical implications. In real biological research,

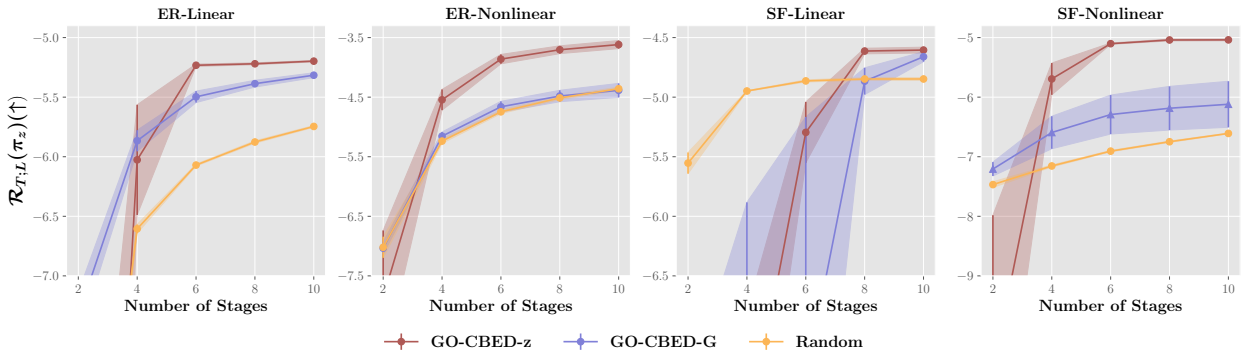


Figure 4.3: Performance comparison of policies trained for $T = 10$ on causal queries, using ER and SF graph priors with linear and nonlinear mechanisms (3 interventions per stage, $d = 10$). While GO-CBED-G—which targets structure learning—is a natural baseline, it underperforms on causal queries, particularly in nonlinear settings. In contrast, GO-CBED-z, which directly targets causal QoI, consistently achieves higher EIG, especially under nonlinear mechanisms. Shaded regions represent ±1 standard error over 4 random seeds.

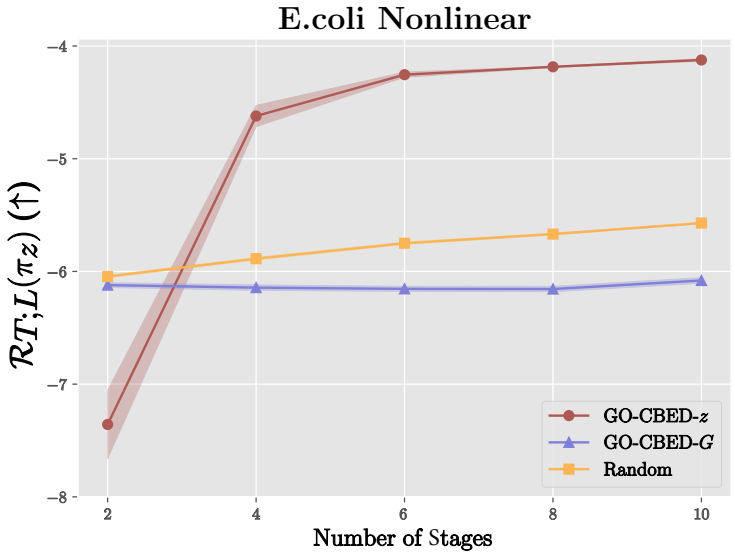


Figure 4.4: Performance comparison of policies trained for $T = 10$ on *E. coli* gene regulatory networks with nonlinear causal mechanisms ($d = 10$). GO-CBED-z performs comparably to baselines in early stages but exhibits rapid improvement after stage 3, ultimately achieving substantially higher EIG than both GO-CBED-G and Random. These results highlight the value of goal-oriented experimental design in realistic biological settings with complex nonlinear causal mechanisms. Shaded regions represent ±1 standard error across 4 random seeds.

experimental resources are often limited, and researchers typically seek to answer specific causal questions rather than infer the entire network structure. GO-CBED's ability to efficiently target such queries highlights its promise for applications such as gene regulatory network analysis and drug target identification, where maximizing information about specific causal effects is critical. We further validate GO-CBED's effectiveness through additional experiments on Yeast networks and with diverse goal specifications in Appendix C.4.2.

Additional evaluations of GO-CBED's robustness to distributional shifts in observation noise are presented in Appendices C.4.2 and C.4.5 for the linear and nonlinear cases, respectively.

4.4.3 Causal discovery tasks

While GO-CBED is primarily designed for general goal-oriented experimental design, we also apply it on specific causal discovery tasks, where the target QoI is the causal graph $z = G$. This enables comparison with existing causal BOED methods specifically designed for structure learning. We consider synthetic settings using ER and SF graph priors, and semi-synthetic settings based on real gene regulatory networks from DREAM [64]. In all cases, we simulate linear and nonlinear causal mechanisms with additive noise. See Appendix C.3.3 for more details.

Baselines We benchmark GO-CBED against four methods: **CAASL** [4], which uses offline RL method with a fixed pre-trained posterior network; **Random**, which uniformly selects interventions at random; **Soft-CBED** [147], which employs Bayesian optimization for single-step EIG; and **DiffCBED** [146], which learns a non-adaptive policy through gradient-based optimization.

Metrics We evaluate performance using three metrics: prior-omitted EIG lower bound $\mathcal{R}_{T;L}$; expected structural Hamming distance $\mathbb{E}$-SHD [37] between posterior graph samples and the ground truth; and F_1-score for edge prediction. To ensure a fair comparison across policies, we train a dedicated posterior network for each policy. This isolates the contribution of the policy itself and avoids confounding effects from differing posterior approximation methods. For example, while CAASL relies on a fixed pre-trained posterior network from [98], other baselines use DAG-bootstrap [52, 69] for linear SCMs and DiBS [97] for nonlinear SCMs. In our evaluation, we adopt posterior networks for inference across all baselines, as they produce higher-quality posteriors than those used in the original works. For completeness, results using each method's original inference setup are included in Appendix C.4.3. All results are averaged over 10 random seeds.

Synthetic SCMs Figure 4.5 presents GO-CBED's performance using ER and SF graph priors with linear ($d = 30$, $T = 15$) and nonlinear ($d = 20, T = 10$) mechanisms. Across all metrics and graph types, GO-CBED consistently outperforms baseline methods, showing especially large

gains in F_1-scores (reaching up to 0.75 on SF graphs). Although GO-CBED initially performs comparably to some baselines, it steadily surpasses them as more interventions are collected. This highlights its strength in optimizing long-term information gain rather than short-term or greedy improvements. While the advantage is most prominent in linear settings, GO-CBED still achieves strong performance in the more challenging nonlinear cases.

Semi-Synthetic Gene Regulatory Networks Figure 4.6 evaluates GO-CBED on 20-node networks derived from the DREAM *E. coli* gene regulatory benchmark with linear mechanisms. Our method consistently outperforms all baselines across all evaluation metrics. By the final intervention stage, GO-CBED achieves high $\mathcal{R}_{T;L}$ values, low $\mathbb{E}$-SHD scores, and high F_1-scores, indicating accurate recovery of the true causal structure with minimal posterior uncertainty. This strong performance on biologically-inspired networks demonstrates GO-CBED's ability to handle the complex dependencies and noise typically encountered in gene regulatory systems. Additional experiments on Yeast networks and with nonlinear mechanisms presented in Appendix C.4.4 further support GO-CBED's effectiveness in biologically relevant settings.

4.5 Conclusions

We presented GO-CBED, a goal-oriented Bayesian framework for sequential causal experimental design. Unlike conventional approaches that aim to learn the full causal model, GO-CBED directly maximizes the EIG on specific causal QoIs, enabling more targeted and efficient experimentation. The framework is both non-myopic, optimizing over entire sequences of interventions, and goal-oriented, focusing on model aspects relevant to the causal query. To overcome the intractability of exact EIG computation, we introduced a variational lower bound, optimized jointly over policy and variational parameters. Our implementation leveraged NFs for flexible posterior approximations and a transformer-based policy network that captures symmetry and structure in the intervention history. Numerical experiments demonstrated that GO-CBED outperforms baseline methods in multiple causal tasks, with gains increasing as causal mechanisms become more complex. Crucially, the joint training of intervention policies and variational posteriors enabled adaptive, goal-oriented exploration of the causal model.

While GO-CBED demonstrates strong empirical performance, several limitations remain. Its scalability is constrained by the complexity of both the underlying causal models and the neural networks architectures. Additionally, its effectiveness depends on the availability of prior knowledge over causal structures and mechanisms. Future work includes incorporating foundation models as high-fidelity world simulators for offline policy training. Recent advances in biological foundation models [145, 35] offer a promising avenue simulating complex, realistic causal mechanisms, which

could significantly enhance policy learning without relying on costly real-world experimentation. Other valuable extensions include generalizing GO-CBED to support multi-target intervention settings and non-differentiable likelihoods, as well as improving policy robustness to changing experimental horizons and dynamic model updates during experimentation.

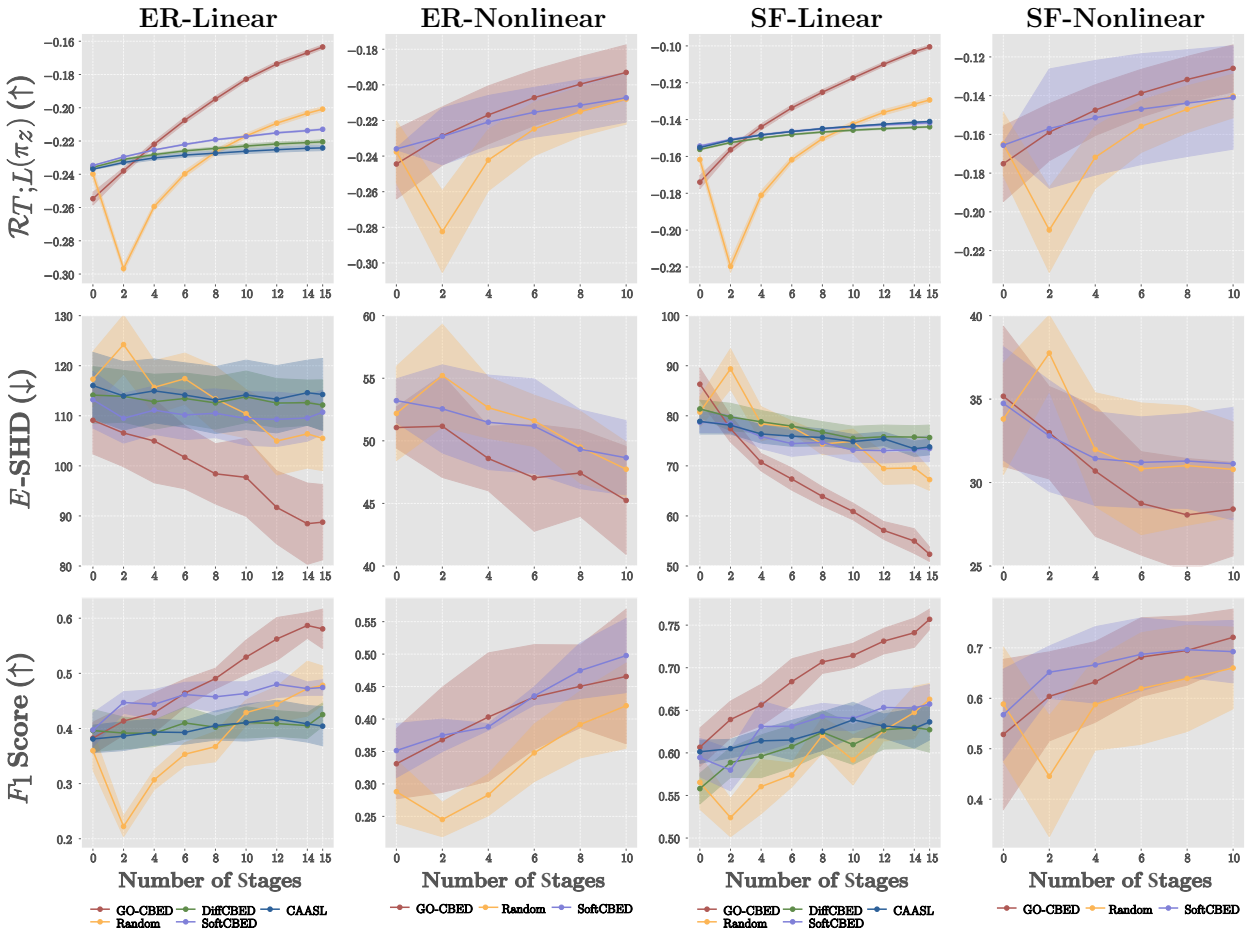


Figure 4.5: Performance comparison on synthetic SCMs, using ER and SF graph priors with linear and nonlinear mechanisms. Metrics include prior-omitted EIG lower bound $\mathcal{R}_{T;L}$, expected structural Hamming distance $\mathbb{E}$-SHD, and F_1-score. GO-CBED performs better or comparatively in terms of uncertainty reduction ($\mathcal{R}_{T;L}$), structural recovery ($\mathbb{E}$-SHD), and structural accuracy (F_1-score) compared to all baselines. Shaded regions indicate ±1 standard error across 10 random seeds.

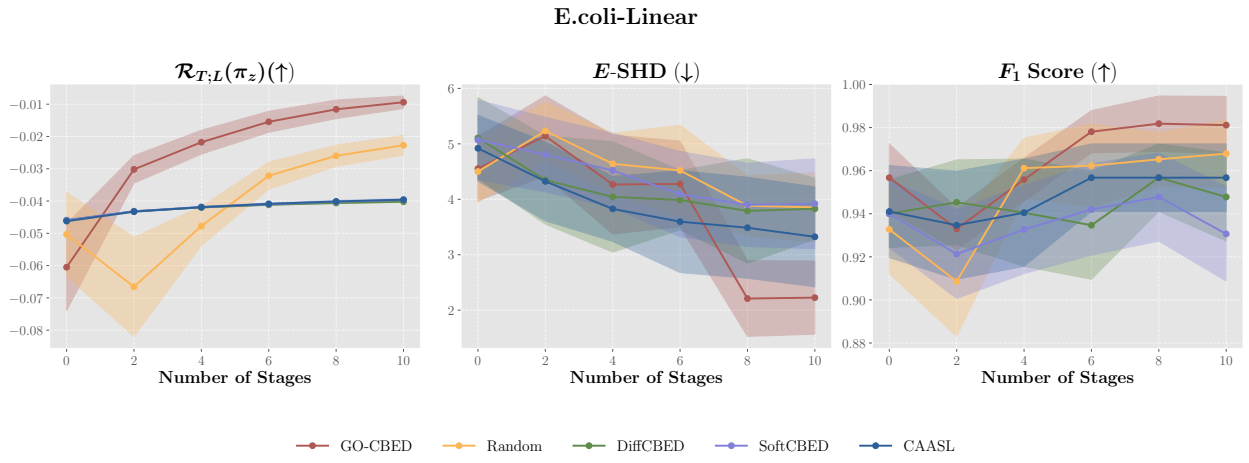


Figure 4.6: GO-CBED outperforms all baselines on semi-synthetic *E. coli* gene regulatory networks ($d = 20$, $T = 10$) with linear mechanisms. Our method achieves: (*Left*) near-zero $\mathcal{R}_{T;L}(\pi_z; \phi)$, (*Middle*) significant lower $\mathbb{E}$-SHD, and (*Right*) superior F_1 score for edge prediction. This demonstrates GO-CBED's ability to efficiently identify the true causal structure in biologically-inspired networks, with the variational posterior tightly concentrated around the ground truth after 10 stages.

CHAPTER 5

Symmetrized Expected Information Gain for Bayesian Optimal Experimental Design

While the expected information gain (EIG) is a natural and widely used utility for Bayesian optimal experimental design (BOED), its practical implementation is often limited by the computational burden associated with nested expectation estimators. This limitation motivates the exploration of alternative information-based utilities that retain the essential properties required for principled experimental design while offering improved computational efficiency.

More generally, the choice of a utility function in BOED is not unique. As established by [61], a valid measure of information for an experiment needs to satisfy a fundamental set of criteria. Within this admissible class, different choices may lead to different computational tradeoffs and interpretations.

Research gap and contributions. In this chapter, we introduce symmetrized expected information gain (sEIG) as an alternative expected utility for BOED. Prior to this work, sEIG has not been systematically studied in the BOED literature, and its theoretical validity and computational advantages relative to EIG were not established. In particular, it was unclear whether sEIG satisfies the axiomatic requirements of an information-based design utility, or whether it offers meaningful computational benefits in practical design optimization.

The central research question addressed in this chapter is whether sEIG can serve as a principled and computationally efficient alternative to EIG for BOED. To answer this question, this chapter (i) establishes sEIG as a valid information-based utility by verifying its compliance with the axioms required for measure of information, (ii) develops efficient single-loop Monte Carlo estimators for sEIG, and (iii) empirically compares sEIG with EIG under different application cases.

5.1 Problem formulation

5.1.1 Symmetric Expected Information Gain

Following Eqn. (2.3), given an observation y, the utility $u(y|\xi)$ is commonly chosen to be the Kullback–Leibler (KL) divergence from the prior to the posterior $u(y|\xi) = D_{\mathrm{KL}}(\, p_{\Theta|Y,\xi} \,||\, p_\Theta \,)$. The EIG at a design ξ then emerges from taking the expectation of this utility over all possible outcomes with respect to the likelihood distribution. From Eqn. (2.18), the nested Monte Carlo (NMC) estimator entails evaluating the likelihood—each entails running a forward model—$N_{\mathrm{out}} \times N_{\mathrm{in}}$ times, which is expensive when the forward solver is already expensive (i.e., from solving a system of differential equations).

Instead, we investigate the use of symmetrized EIG (sEIG) as the expected utility for BOED. Here, the utility is chosen to be the simple average of the forward and reverse prior-to-posterior KL divergence terms:

$$\begin{aligned}
U_S(\xi) &= \mathbb{E}_{Y|\xi}\left[\frac{1}{2}D_{\mathrm{KL}}(p_{\Theta|Y,\xi} \,||\, p_\Theta) + \frac{1}{2}D_{\mathrm{KL}}(p_\Theta \,||\, p_{\Theta|Y,\xi})\right] \\
&= \mathbb{E}_{Y|\xi}\left[\frac{1}{2}\mathbb{E}_{\Theta|Y,\xi}\left[\log\frac{p(\Theta|Y,\xi)}{p(\Theta)}\right] + \frac{1}{2}\mathbb{E}_\Theta\left[\log\frac{p(\Theta)}{p(\Theta|Y,\xi)}\right]\right] \\
&= \mathbb{E}_{Y|\xi}\left[\frac{1}{2}\mathbb{E}_{\Theta|Y,\xi}\left[\log\frac{p(Y|\Theta,\xi)}{p(Y|\xi)}\right] + \frac{1}{2}\mathbb{E}_\Theta\left[\log\frac{p(Y|\xi)}{p(Y|\Theta,\xi)}\right]\right] \\
&= \frac{1}{2}\Big\{\mathbb{E}_{Y,\Theta|\xi}\left[\log p(Y|\Theta,\xi)\right] - \mathbb{E}_{Y,\Theta|\xi}\left[\log p(Y|\xi)\right] \\
&\qquad\qquad + \mathbb{E}_{Y|\xi}\left[\mathbb{E}_\Theta\left[\log p(Y|\xi)\right]\right] - \mathbb{E}_{Y|\xi}\left[\mathbb{E}_\Theta\left[\log p(Y|\Theta,\xi)\right]\right]\Big\} \\
&= \frac{1}{2}\Big\{\mathbb{E}_{Y,\Theta|\xi}\left[\log p(Y|\Theta,\xi)\right] - \mathbb{E}_{Y|\xi}\left[\mathbb{E}_\Theta\left[\log p(Y|\Theta,\xi)\right]\right]\Big\},
\end{aligned} \tag{5.1}$$

where in the fourth equality, the second and third terms cancel since

$$-\mathbb{E}_{Y,\Theta|\xi}\left[\log p(Y|\xi)\right] + \mathbb{E}_{Y|\xi}\left[\mathbb{E}_\Theta\left[\log p(Y|\xi)\right]\right] = -\mathbb{E}_{Y|\xi}\left[\log p(Y|\xi)\right] + \mathbb{E}_{Y|\xi}\left[\log p(Y|\xi)\right] = 0. \tag{5.2}$$

A non-nested (i.e., single-loop) Monte Carlo estimator can then be formed:

$$\widehat{U}_S^{N,M}(\xi) = \frac{1}{2N}\sum_{i=1}^{N}\log p(y^{(i)}|\theta^{(i)},\xi) - \frac{1}{2M}\sum_{j=1}^{M}\log p(y^{(j)}|\theta'^{(j)},\xi), \tag{5.3}$$

where in the first term $\theta^{(i)} \sim p(\theta)$ and $y^{(i)} \sim p(y^{(i)}|\theta^{(i)},\xi)$ are sampled jointly, and in the second term $\theta^{(j)} \sim p(\theta)$ and $y^{(j)} \sim p(y^{(j)}|\theta^{(j)},\xi)$ are sampled jointly while only $y^{(j)}$ is used in Eqn. (5.3),

and $\theta'^{(j)} \sim p(\theta)$ is generated independently. The total number of unique likelihood evaluations (and hence forward solves) is therefore $N + 2M$.

Estimator mean. Compared to the traditional NMC estimator that is biased for EIG [125], it is trivial to show that Eqn. (5.3) is an unbiased estimator for sEIG.

$$\begin{aligned}\mathbb{E}\big[\hat{U}_S^{N,M}(\xi)\big] &= \frac{1}{2}\int p(\theta)\left(\int p(y|\theta,\xi)\log p(y|\theta,\xi)\,dy\right)\mathrm{d}\theta \\ &\quad - \frac{1}{2}\int p(y|\xi)\left(\int p(\theta)\log p(y|\theta,\xi)\,d\theta\right)\mathrm{d}y. \\ &= \frac{1}{2}\,\mathbb{E}_\Theta\mathbb{E}_{Y|\Theta,\xi}\big[\log p(Y|\Theta,\xi)\big] - \frac{1}{2}\,\mathbb{E}_{Y|\xi}\mathbb{E}_{\Theta'}\big[\log p(Y|\Theta',\xi)\big] = U_S(\xi).\end{aligned}$$

Estimator variance. Define random variables $A := \log p(Y|\Theta,\xi)$ and $B := \log p(Y|\Theta',\xi)$. Since $\hat{U}_S^{N,M}(\xi) = \frac{1}{2N}\sum_{i=1}^N A_i - \frac{1}{2M}\sum_{j=1}^M B_j$ where $\{A_i\}_{i=1}^N$ are i.i.d. copies of A and $\{B_j\}_{j=1}^M$ are i.i.d. copies of B, the two collections are independent. Then,

$$\mathrm{Var}\Big(\hat{U}_S^{N,M}(\xi)\Big) = \frac{\mathrm{Var}(A)}{4N} + \frac{\mathrm{Var}(B)}{4M},$$

implying an asymptotic rate $\mathrm{Var}\Big(\hat{U}_S^{N,M}(\xi)\Big) = O\Big(\frac{1}{N}\Big) + O\Big(\frac{1}{M}\Big)$. This linear variance decay matches standard Monte Carlo rates and allows independent control of accuracy through N and M.

5.1.2 Valid measure of information

We provide a justification below showing that $U_S(\xi)$ is a valid measure of information. As stated in [61], results established for finite Θ spaces remain valid in countable and continuous settings. Thus, we focus on the finite case here for simplicity, without loss of generality.

Intuitively, reasonable measure of information should assign zero value to an experiment that conveys no information about the unknown parameter. To make this precise, we introduce the notion of a **totally non-informative experiment**: ξ_{tni}: the experiment whose distribution of y (i.e., likelihood) is independent of $\theta \in \Theta$, hence observing y cannot provide information about θ.

Definition: Valid Measure of Information [61] A measure of the information about Θ in an experiment ξ assigns a value $U(\xi)$ such that:

1. $U(\xi)$ is a real number, (5.4)
2. $U(\xi_{\text{tni}}) = 0$, (5.5)
3. Whenever ξ_1 is sufficient for ξ_2 then $U(\xi_1) \geq U(\xi_2)$. (5.6)

Conditions (1) and (2) ensure normalization and the "non-informative" intuition, while condition (3) enforces monotonicity with respect to experiment informativeness. A formal definition of sufficiency is given in Appendix D. Intuitively, sufficiency means that any information about Θ available from ξ_2 can be recovered from ξ_1 via a parameter-independent stochastic transformation; consequently, ξ_1 cannot be less informative than ξ_2.

A specific example satisfying the above conditions is the ϕ-divergence measure [61]. The generalized ϕ-divergence measure of the information about Θ in an experiment ξ is:

$$U(\xi) = \int p(y)\phi\left(\frac{p(y|\theta_1)}{p(y)}, \ldots, \frac{p(y|\theta_k)}{p(y)}\right) \mathrm{d}y, \tag{5.7}$$

where $\phi(u)$ is a real valued convex function on $K_\Theta = \{u = (u_1, \ldots, u_k) \in R^k : u_i \geq 0, \sum_{i=1}^k p(\theta_i)u_i = 1\}$ with $\phi(1, \ldots, 1) = 0$. The ϕ-divergence measure can be equivalently expressed as:

$$U(\xi) = E_Y[\varphi(p(\theta|y))] = \int p(y)\varphi(p(\theta_1|y), \ldots, p(\theta_k|y))\,\mathrm{d}y \tag{5.8}$$

with $\varphi(u)$ a real valued convex function on K_Θ and $\varphi(p(\theta_1), \ldots, p(\theta_k)) = 0$. Any information measure that can be expressed in the form of ϕ-divergence measure is a valid measure of information [61]: conditions (1) and (2) satisfy trivially, and the proof that the ϕ-divergence measure satisfying condition (3) is given in Appendix D.

It is easy to verify that the traditional EIG is a ϕ-divergence measure with $\phi^F(u) = \sum_{i=1}^k p(\theta_i)u_i \log u_i$:

$$\begin{aligned} U_F(\xi) &= \int p(y)\phi\left(\frac{p(y|\theta_1)}{p(y)}, \ldots, \frac{p(y|\theta_k)}{p(y)}\right) \mathrm{d}y \\ &= \int p(y) \sum_{i=1}^k p(\theta_i)\frac{p(y|\theta_i)}{p(y)} \log\left(\frac{p(y|\theta_i)}{p(y)}\right) \mathrm{d}y \\ &= \int p(y) \sum_{i=1}^k p(\theta_i|y) \log\left(\frac{p(\theta_i|y)}{p(\theta_i)}\right) \mathrm{d}y. \end{aligned} \tag{5.9}$$

Specifically, under the assumption that the posterior is absolutely continuous with respect to the prior so that all logarithmic ratios are well defined, the expected KL divergence from prior to

posterior (reverse) is also a valid information measure, with $\phi^R(u) = \sum_{i=1}^k p(\theta_i)\log(1/u_i)$:

$$U_R(\xi) = \int p(y)\phi\left(\frac{p(y|\theta_1)}{p(y)}, \ldots, \frac{p(x|\theta_k)}{p(y)}\right) \mathrm{d}y = \int p(y) \sum_{i=1}^k p(\theta_i)\log\left(\frac{p(y)}{p(y|\theta_i)}\right) \mathrm{d}y$$
$$= \int p(y) \sum_{i=1}^k p(\theta_i)\log\left(\frac{p(\theta_i)}{p(\theta_i|y)}\right) \mathrm{d}y. \tag{5.10}$$

Therefore, under the same absolute continuity assumption, as sEIG is the average of Eqn. (5.9) and Eqn. (5.10), then with $\phi^S(u) = \sum_{i=1}^k p(\theta_i)[u_i\log(u_i) + \log(1/u_i)]$:

$$\begin{aligned} U_S(\xi) &= \int p(y)\phi\left(\frac{p(y|\theta_1)}{p(y)}, \ldots, \frac{p(y|\theta_k)}{p(y)}\right) \mathrm{d}y \\ &= \int p(y) \sum_{i=1}^k \left[p(\theta_i)\frac{p(y|\theta_i)}{p(y)}\log(\frac{p(y|\theta_i)}{p(y)}) + p(\theta_i)\log\left(\frac{p(y)}{p(y|\theta_i)}\right)\right] \mathrm{d}y \\ &= \int p(y) \sum_{i=1}^k \left[p(\theta_i|y)\log(\frac{p(\theta_i|y)}{p(\theta_i)}) + p(\theta_i)\log\left(\frac{p(\theta_i)}{p(\theta_i|y)}\right)\right] \mathrm{d}y. \end{aligned} \tag{5.11}$$

It is trivial to verify that ϕ^F, ϕ^R, ϕ^S are all convex functions on K_θ with $\phi^F(1, \ldots, 1) = 0$, $\phi^R(1, \ldots, 1) = 0$, and $\phi^S(1, \ldots, 1)$. Hence, they satisfy all conditions.

5.2 Numerical experiments

Throughout this section, although we compare the values and optimal designs obtained from EIG and sEIG, it is important to keep in mind that both are equally valid information measures for OED, as established in the previous section. In particular, EIG should not be regarded as a "ground truth" with sEIG viewed as an approximation. Rather, our goal is to contrast their outcomes and to offer insight into their respective properties and interpretations.

5.2.1 Case 1: Linear-Gaussian examples

We first consider the classical linear-Gaussian model for illustrating sampling efficiency of the sEIG estimator. Here the unknown parameter $\Theta \in \mathbb{R}^d$ takes a standard normal prior $\mathcal{N}(0, I_d)$, and the likelihood takes the form:

$$y|\theta \sim \mathcal{N}(\xi\theta, \sigma^2 I_m), \tag{5.12}$$

with design matrix $\xi \in \mathbb{R}^{m \times d}$. For a given design ξ, the EIG admits the analytic form:

$$\mathrm{EIG}(\xi) \;=\; \tfrac{1}{2} \log \det\!\left(I_\xi + \tfrac{1}{\sigma^2} \xi^\top \xi \right). \tag{5.13}$$

Similarly, the sEIG admits an analytic form:

$$\mathrm{sEIG}(\xi) \;=\; \frac{1}{2\sigma^2} \|\xi\|_F^2 \;=\; \frac{1}{2\sigma^2} \operatorname{tr}\!\left(\xi^\top \xi\right) \;=\; \frac{1}{2\sigma^2} \sum_{i=1}^{\min(m,d)} s_i^2, \tag{5.14}$$

where $\|\xi\|_F$ denotes the Frobenius norm of ξ, and $\{s_i\}$ are the singular values of ξ.

Figure 5.1 compares the analytic values of EIG and sEIG (solid lines) against their Monte Carlo estimators. The standard NMC estimator is employed for EIG, which involves an inner loop of size M approximation to the marginal log-likelihood $\log p(y)$.

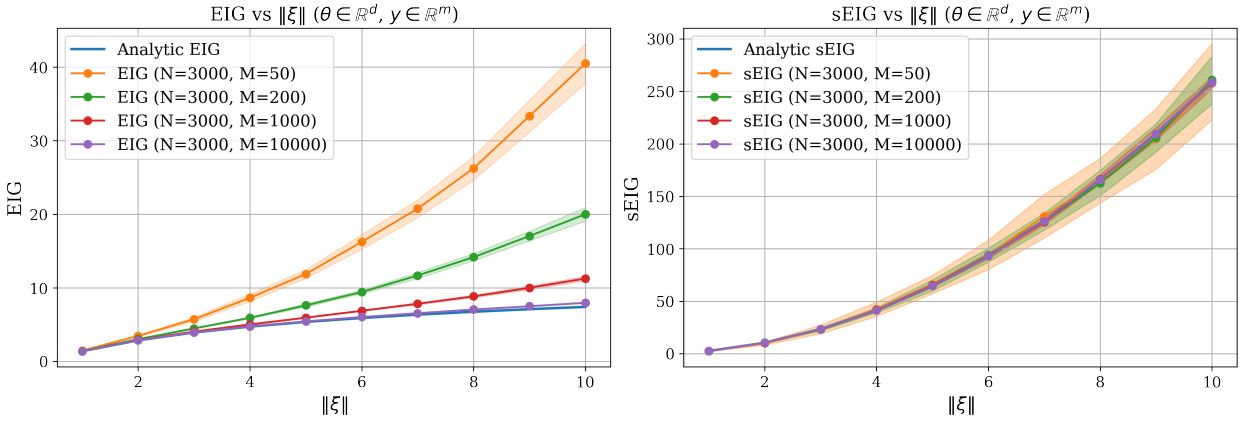


Figure 5.1: Comparison of analytic and Monte Carlo estimators for EIG (left) and sEIG (right) in the linear-Gaussian model with $d = 10$ and $m = 3$. Shaded regions show ± 1 standard deviation over 30 independent runs. NMC estimates of EIG suffer from high bias for small inner sample sizes M, whereas the single-loop Monte Carlo estimator of sEIG is unbiased.

The NMC estimator exhibits significant bias when the inner sample size M is small[1], especially for larger design strengths $\|\xi\|$. This bias diminishes as M increases, with the estimates approaching the analytic curve. In contrast, the estimator for sEIG is unbiased across all values of M, with variability decreasing as M grows. Specifically, Figure 5.2 plots the relative difference, defined as the ratio of the standard deviation to the mean across 30 runs, for $N = 3000$, $M = 200$ in the case of EIG and $N = 3000$, $M = 10{,}000$ for sEIG. Notably, the sEIG estimator remains markedly more stable even with a total forward model budget of 13,000, which is substantially smaller than the comparison NMC case with a much larger budget of $3000 \times 200 = 600,000$.

[1]Here we use the same notation of N and M in place of n_{out} and n_{in} for NMC, in order to draw contrast with the sample sizes for sEIG. However, the respective sample sizes of the two methods do not have inherent ties.

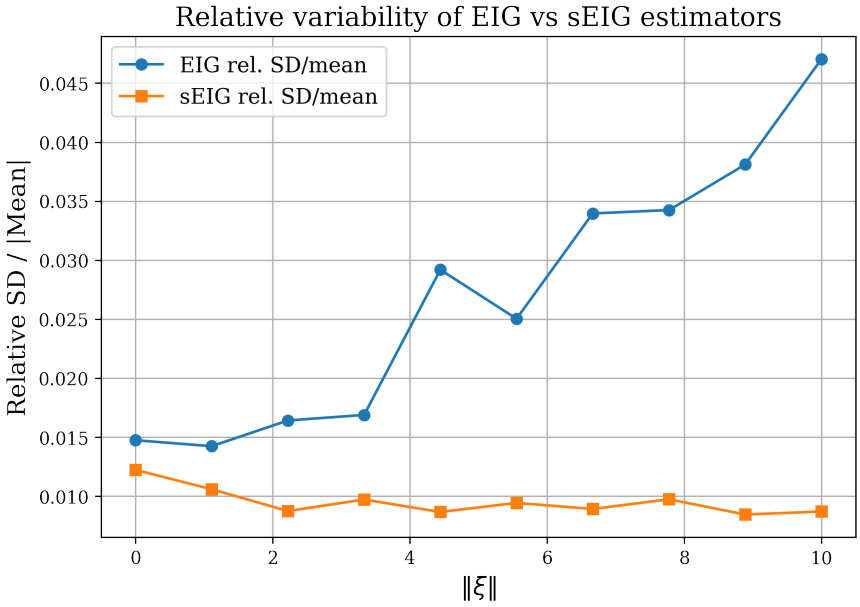


Figure 5.2: Relative variability (SD/mean) of EIG versus sEIG estimators as a function of design strength $\|\xi\|$, calculated over 30 runs. The NMC estimator for EIG ($N = 3000$, $M = 200$ for a total of 600,000 forward runs) exhibits higher relative variability despite using a larger budget, which increases with $\|\xi\|$, while the sEIG estimator ($N = 3000$, $M = 200$ for a total of 13,000 forward runs) remains substantially more stable despite using a much smaller computational budget.

5.2.2 Case 2: Convection-diffusion with building obstacles

The progression of a contaminant's concentration in a 2D square domain $[0, 1]^2$ may be described by the scalar diffusion partial differential equation (PDE):

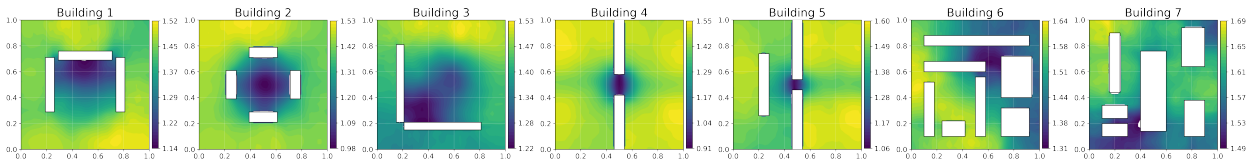
$$\frac{\partial G(z, t; \theta)}{\partial t} = \nabla^2 G + S(z, t; \theta), \qquad z \in [0, 1]^2, \quad t > 0, \tag{5.15}$$

where $\theta = [\theta_{Y_2}, \theta_y] \in \mathbb{R}^2$ represents the source location, which is also the unknown parameter to be inferred and endowed with a uniform prior $\mathcal{U}([0, 1]^2)$. The source term has the form

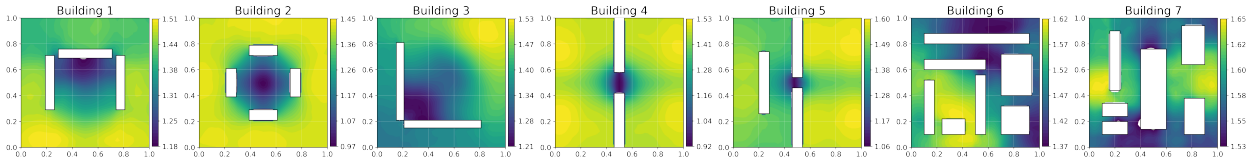
$$S(z, t; \theta) = \frac{s}{2\pi h^2} \exp\left(-\frac{\|\theta - z\|^2}{2h^2}\right), \tag{5.16}$$

where $s = 2$ and $h = 0.05$ indicate the source strength and source width, respectively. The initial condition is $G(z, 0; \theta) = 0$ and we impose Neumann boundary conditions on all sides of the square domain. The PDE is solved using a second-order finite volume method on a uniform grid with $\Delta z_{Y_2} = \Delta z_y = 0.01$, and the time marching uses a second-order fractional step method with $\Delta t = 5 \times 10^{-4}$.

The design variable is the sensor location used to measure the contaminant concentration. We



(a) Estimated EIG using NMC with $10^3 \times 10^3$ samples.



(b) Estimated EIG using NMC with $3 \times 10^4 \times 3 \times 10^4$ samples.

Figure 5.3: Comparison of EIG landscapes under identical building configurations obtained with different nested Monte Carlo budgets. The top row uses a smaller Monte Carlo budget and exhibits noticeable stochastic variability, while the bottom row uses a substantially larger budget, resulting in smoother utility surfaces and reduced estimator variance.

conduct a single experiment at $t = 0.16$, so the dimension of the design variable depends on the number of sensors placed in the domain. If m sensors are used, then the dimension of design variables is $N_d = 2m$, and the dimension of observations is $N_y = m$, while the dimension of parameters is always $N_\theta = 2$. We assume additive Gaussian noise on the measurements:

$$y(z) = G(z, t = 0.16; \theta) + \epsilon, \tag{5.17}$$

where $\epsilon \sim \mathcal{N}(0, 0.05^2)$. The design variable d is thus a set of sensor locations $\xi = [z^{(1)}, \ldots, z^{(m)}]$, and the observation vector is $y = [y(z^{(1)}), \ldots, y(z^{(m)})]$. For multiple sensors, we also assume that the measurement noise is independent across sensors.

Extension with building obstacles. To make the source inversion domain more realistic, we further incorporate additional building obstacles into the 2D domain. The prior over the source location remains a uniform distribution on $[0, 1]^2$, except that the prior density is set to zero inside the building regions. This imposes hard constraints on feasible source locations while leaving the PDE model unchanged.

Results Figure 5.3a and Figure 5.3b illustrate the EIG under several building configurations, estimated using NMC approximations with different sample sizes. In Figure 5.3a, the EIG is estimated using $10^3 \times 10^3$ samples in NMC, while Figure 5.3b employs a substantially larger budget of $3 \times 10^4 \times 3 \times 10^4$ samples.

When the sample sizes are relatively small, the estimated EIG surfaces exhibit noticeable

stochastic variability, leading to optimal sensor locations that can differ appreciably from those obtained with larger sample sizes. As the number of samples increases, the EIG estimates become smoother and more consistent, and the locations of the optimal designs converge toward those identified by the higher-fidelity NMC approximation. These results highlight the trade-off between computational cost and robustness of the resulting experimental designs.

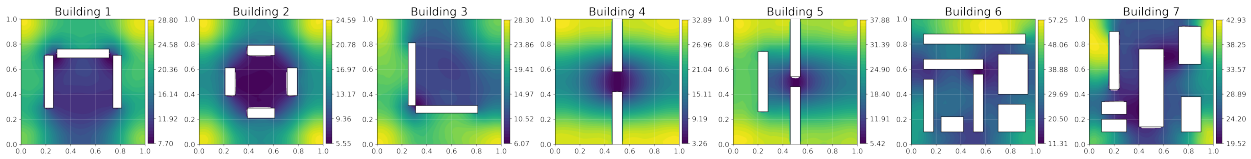


Figure 5.4: Estimated sEIG with $N = M = 10^4$.

Comparison with sEIG. We further evaluate the sEIG estimator under the same building configurations, using $N = M = 10^4$ samples. Figure 5.4 illustrates the resulting sEIG landscapes across the same building geometries. Compared to the standard EIG estimator, sEIG exhibits substantially smoother utility landscapes despite using a much smaller computational budget. This highlights the reduced estimator variance afforded by the symmetrized formulation.

Beyond variance reduction, the qualitative structure of the optimal designs also differs between EIG and sEIG. Standard EIG tends to favor sensor locations where measurements are expected to be large under most source locations, effectively selecting regions where the sensor is likely to receive strong signals over a large portion of the prior support. In contrast, sEIG emphasizes designs that are informative in a contrastive sense: it prefers sensor locations where an observation can strongly rule out specific regions of the prior. As a result, sEIG tends to select sensor locations in regions where the contaminant signal is unlikely to reach based on the prior, leading to dominant contributions from likelihood–prior contrast in the symmetrized KL divergence.

5.3 Conclusions

In this chapter, we investigated the sEIG as an expected utility for BOED. We established that sEIG is a valid measure of information, being a member under the broader family of ϕ-divergences, and that it admits a single loop Monte Carlo estimator requiring only $N + 2M$ likelihood (forward model) evaluations. This stands in contrast to the NMC estimator for the standard EIG, which requires $N \times M$ evaluations and suffers from high bias and variance at small inner sample sizes.

Preliminary experiments in the linear-Gaussian model confirm these advantages: the sEIG estimator is unbiased, exhibits lower relative variability, reveals its optimal designs and achieves stable performance with substantially smaller computational budgets. These findings provide strong motivation for adopting sEIG in settings where forward model evaluations are costly.

Further work entails expanding the capabilities of the methodology to higher dimensions, sequential experiments, and demonstrating under more complex application problems.

CHAPTER 6

Conclusions and Future Work

6.1 Conclusions

This thesis develops principled and computationally efficient methods for Bayesian optimal experimental design (BOED) in both goal-oriented and sequential settings, with applications to causal and physical systems. Across the thesis, we address a recurring bottleneck in BOED: while expected information gain (EIG) provides a rigorous utility for experimental design, its evaluation and optimization are often prohibitively expensive for complex models, long horizons, and structured scientific goals. To overcome these challenges, we propose new formulations, estimators, and amortized algorithms that enable efficient design under uncertainty, and we validate them on both synthetic and realistic scientific benchmarks.

We begin by formulating variational optimal experimental design (vOED) for static experimental design. In vOED, the EIG objective is replaced by a provable variational lower bound, yielding a tractable utility that can be optimized efficiently. We show that this lower bound admits a closed-form objective via the change-of-variables formula, enabling an estimator that requires only $O(N)$ forward model evaluations. We then introduce coupling-layer–based normalizing flows as an expressive variational family. Their triangular structure enables efficient density evaluation and exact inverse sampling, making them particularly well suited for posterior approximation within BOED. Together, these components support implicit likelihood models, scalable optimization, and efficient posterior sampling within a unified variational framework. We demonstrate the effectiveness of vOED through multiple examples. First, we consider nonlinear benchmark problems to evaluate the quality of the variational lower bound. We show that normalizing flow–based variational families achieve substantially higher sampling efficiency than nested Monte Carlo estimators and markedly lower bias than simpler Gaussian variational approximations. In addition, qualitative comparisons of the amortized posteriors show that the variational posteriors closely match the

reference posteriors obtained via sequential Monte Carlo (SMC), confirming the expressiveness of the normalizing flows. Second, we demonstrate that, beyond improved estimator efficiency, the variational formulation enables direct gradient-based optimization of the design variables. Because the lower bound itself is differentiable with respect to the design, optimal experimental configurations can be identified efficiently via gradient descent. Third, we deploy vOED in a realistic electro-coating (e-coating) application, where observations are high-dimensional and sequential. We show that adopting efficient data embeddings, such as LSTM-based encoders, significantly tightens the variational bound by retaining task-relevant information while reducing inference complexity. Finally, we extend vOED to implicit likelihood models. Across these examples, vOED consistently outperforms existing benchmarks in both computational efficiency and design quality, demonstrating its flexibility and practical utility for goal-oriented experimental design in complex physical systems.

We next extend vOED to the sequential setting, resulting in variational sequential OED (vsOED). We formulate the design of a finite horizon T as an MDP in which the state summarizes the experiment history I_t, the action is the next design ξ_t, and the dynamics are induced by Bayes' rule through the data-generating forward operator. Within this framework, we derive a policy-level variational lower bound on the EIG accrued over an entire experiment sequence, preserving the information-theoretic foundation while enabling tractable optimization of adaptive policies. A key theoretical result is the equivalence between terminal-information-gain (TIG) and incremental-information-gain (IIG) formulations, and we further show that both admit a one-point reward representation that eliminates the need for explicit KL integration over posteriors. Importantly, the vsOED objective is flexible and unifies multiple experimental goals—parameter inference, goal-oriented prediction via QoIs, model discrimination, and settings with nuisance parameters. We demonstrate vsOED across various benchmarks. For the source location and CES problems which are well established benchmarks in the sequential BOED literature, vsOED achieves substantially higher sampling efficiency than existing approaches while maintaining accurate posterior approximation. Although TIG and IIG are theoretically equivalent in terms of the resulting sequential information objective, IIG consistently yields stronger empirical performance, particularly for longer horizons. We also apply vsOED in the SIR epidemic model with implicit likelihoods, vsOED remains competitive with existing methods while requiring only forward simulations and no model derivatives. Finally, we apply vsOED to a realistic convection–diffusion problem involving nuisance parameters, physical state dynamics, multiple competing models, and surrogate-based forward solvers. This example highlights the practical applicability of the framework to real-world scientific and engineering problems, where computational cost, model uncertainty, and heterogeneous experimental objectives must be handled simultaneously. Taken together, these results demonstrate that the vsOED framework provides a unified, scalable, and flexible approach to sequential experimental design,

supporting parameter inference, prediction, model discrimination, and real-world physics-based applications within a single principled methodology.

Building on these developments, we further apply the proposed variational BOED framework to causal experimental design, where focusing on task-relevant causal quantities of interest can be substantially more efficient than full parameter or structure inference. We develop a goal-oriented causal Bayesian experimental design (GO-CBED) formulation that explicitly targets causal QoIs, encompassing both causal discovery and causal reasoning tasks. For causal discovery, we consider recovery of causal structures using a Bernoulli variational family over graph edges, while for causal reasoning we still employ normalizing flow–based variational posteriors to infer interventional effects. To support scalable and expressive sequential decision making in causal settings, we adopt transformer-based architectures for both the policy network and the data embedding module.

In a linear–Gaussian setting, where exact posteriors are tractable, we motivate the proposed GO-CBED framework through analytical solutions that validate the correctness and tightness of the variational objectives. This example demonstrates both that directly optimizing experimental designs for causal quantities of interest is substantially more sample-efficient than full parameter inference, and that nested Monte Carlo estimators for QoI-based utilities can be highly biased and less efficienct under budget limit, whereas the proposed GO-CBED variational objective remains significantly more stable. Across a range of graph priors and parent–child generative mechanisms, GO-CBED consistently outperforms existing benchmarks. These results highlight the advantages of goal-oriented causal design and reinforce the broader thesis theme that information-theoretic variational formulations, combined with expressive amortized models, enable scalable and principled experimental design for complex causal and physical systems.

Finally, we explore the symmetrized expected information gain (sEIG) as a complementary information-theoretic objective for experimental design. We show that sEIG admits an unbiased Monte Carlo estimator with computational complexity of order $O(M+N)$, in contrast to the $O(MN)$ nested structure required by standard EIG estimators. We further establish that sEIG satisfies the conditions of a valid measure of information, providing a principled theoretical foundation for its use in experimental design. Through numerical studies, we demonstrate that the sEIG estimator exhibits substantially improved sampling efficiency and reliability compared to nested Monte Carlo estimators for EIG, as quantified by markedly lower relative variability measured by the ratio of standard deviation to mean. Finally, we observe that, due to their differing formulations, sEIG and EIG can favor distinct optimal experimental designs despite both being valid information measures, highlighting an important and previously underexplored trade-off between information criteria and experimental objectives in practical design problems.

6.2 Limitations and future work

Despite the significant gains in efficiency and flexibility achieved by the proposed variational BOED frameworks, several limitations and open challenges remain. A primary consideration is scalability in high-dimensional and structurally complex settings [63]. While amortization, expressive variational families, and transformer-based policies substantially reduce the computational burden relative to nested Monte Carlo approaches, scaling to very large systems remains challenging. This is particularly pronounced in causal discovery problems [67], where the number of latent structural parameters grows quadratically with the number of nodes. Further progress in this direction will likely require additional structural constraints, sparsity-inducing priors, or hierarchical representations that can better exploit problem-specific structure.

More broadly, all information-theoretic design objectives considered in this thesis are inherently Bayesian and therefore depend on the assumed prior and model class. When the prior is misspecified or the generative model does not adequately capture the true data-generating process, the resulting designs may be suboptimal or overly confident. While variational formulations reduce computational bias and improve tractability, they do not eliminate epistemic uncertainty arising from incorrect modeling assumptions. An important direction for future work is the development of design criteria and learning-based policies that are robust to prior misspecification and model mismatch, potentially through adaptive model correction mechanisms that update or refine the model class as new data are acquired [48, 142].

Another fundamental assumption underlying the proposed methods is the availability of a forward model or simulator that faithfully represents the underlying physical or causal system [99, 153, 157]. As large foundation models increasingly serve as approximate world simulators, an open challenge is how to integrate such models into principled Bayesian experimental design pipelines. In this setting, issues of calibration, uncertainty quantification, distributional shift, and adaptive model correction become especially critical. Understanding how variational BOED objectives behave when the simulator itself is learned, imperfect, or evolving over time—and how experimental design can actively drive model refinement—remains an important direction for future research.

Finally, while the sEIG offers an unbiased and computationally efficient alternative to standard EIG estimation, its symmetrized formulation can favor experimental designs that differ from those selected by EIG. Although both measures satisfy principled information-theoretic criteria, they need not align with the same scientific objectives in practice. A deeper understanding of when sEIG best reflects the goals of a given experiment, how to choose between EIG and sEIG, and how these utilities compare with other valid information measures [72] in terms of theoretical properties and empirical performance remains an open question and a promising avenue for future work.

Despite these limitations, the methods developed in this thesis provide a unified view of information-theoretic design that is both computationally tractable and aligned with targeted scientific questions. We expect that goal-oriented, non-myopic sequential causal BOED, together with sample-efficient information utilities such as sEIG, will become increasingly important as experimentation moves toward complex, high-dimensional systems where resources are limited and scientific objectives are inherently query-driven.

CHAPTER A

Appendix for Variational BOED with Normalizing Flows

A.1 Proof for Proposition 1

Proof. Starting from Eqn. (2.8), we have

$$
\begin{aligned}
& U_L(\xi;\phi_1) \le U_L(\xi;\phi_2) \\
\iff & \mathbb{E}_{\Theta,Y|\xi}\left[\log \frac{q(\Theta|Y;\phi_1)}{p(\Theta)}\right] \le \mathbb{E}_{\Theta,Y|\xi}\left[\log \frac{q(\Theta|Y;\phi_2)}{p(\Theta)}\right] \\
\iff & \mathbb{E}_{\Theta,Y|\xi}\left[\log q(\Theta|Y;\phi_1) - \log p(\Theta)\right] \le \mathbb{E}_{\Theta,Y|\xi}\left[\log q(\Theta|Y;\phi_2) - \log p(\Theta)\right] \\
\iff & \mathbb{E}_{\Theta,Y|\xi}\left[\log q(\Theta|Y;\phi_1) - \log p(\Theta|Y,\xi)\right] \le \mathbb{E}_{\Theta,Y|\xi}\left[\log q(\Theta|Y;\phi_2) - \log p(\Theta|Y,\xi)\right] \\
\iff & \mathbb{E}_{\Theta,Y|\xi}\left[\log \frac{q(\Theta|Y;\phi_1)}{p(\Theta|Y,\xi)}\right] \le \mathbb{E}_{\Theta,Y|\xi}\left[\log \frac{q(\Theta|Y;\phi_2)}{p(\Theta|Y,\xi)}\right] \\
\iff & -\mathbb{E}_{Y|\xi}\mathbb{E}_{\Theta|Y,\xi}\left[\log \frac{p(\Theta|Y,\xi)}{q(\Theta|Y;\phi_1)}\right] \le -\mathbb{E}_{Y|\xi}\mathbb{E}_{\Theta|\xi}\left[\log \frac{p(\Theta|Y,\xi)}{q(\Theta|Y;\phi_2)}\right] \\
\iff & \mathbb{E}_{Y|\xi}\left[D_{\mathrm{KL}}(p_{\Theta|Y,\xi} \,||\, q_{\Theta|Y;\phi_1})\right] \ge \mathbb{E}_{Y|\xi}\left[D_{\mathrm{KL}}(p_{\Theta|Y,\xi} \,||\, q_{\Theta|Y;\phi_2})\right] \qquad \text{(A.1)}
\end{aligned}
$$

ending at Eqn. (2.9). The fourth line is from adding $\mathbb{E}_{\Theta,Y|\xi}\left[\log p(\Theta) - \log p(\Theta|Y,\xi)\right]$, which is a constant, on both sides to the preceeding line. □

A.2 Additional Discussions for Invertible Neural Networks

By incorporating the data y into the coupling layer, the vOED-NFs can then produce a posterior distribution conditioned on that data. Correspondingly, Eqn. (2.27) and (2.28) become

$$f_1(\widetilde{\Theta}|y) = \begin{bmatrix} \widetilde{\Theta}_1 \\ Z_2 = \widetilde{\Theta}_2 \odot \exp(s_1(\widetilde{\Theta}_1, y)) + t_1(\widetilde{\Theta}_1, y) \end{bmatrix}, \tag{A.2}$$

$$f_2(f_1(\widetilde{\Theta}|y)) = \begin{bmatrix} Z_1 = \widetilde{\Theta}_1 \odot \exp(s_2(Z_2), y) + t_2(Z_2, y) \\ Z_2 \end{bmatrix}. \tag{A.3}$$

The invertibility is ensured as follows. Once a sample for $\widetilde{\Theta}$ is obtained, the outputs from $s_1(\cdot), s_2(\cdot)$ and $t_1(\cdot), t_2(\cdot)$ are then fixed. The resulting transformation for $Z_2 = \widetilde{\Theta}_2 \odot \exp(s_1(\widetilde{\Theta}_1, y)) + t_1(\widetilde{\Theta}_1, y)$ is then a linear function with respect to $\widetilde{\Theta}_2$ and therefore invertible.

A.3 Case 1

A.3.1 Hyperparameters

The hyperparameters for Sec. 5.2.1 are given in Tables A.1 and A.2. The number of vOED-NFs transformation is also compared across $T = \{1, 3, 5, 7\}$.

Hyperparameter	vOED-G
N_{opt}	2×10^4
N	10^4
N_{batch}	1000
Initial learning rate	10^{-2}
Learning rate decay	0.99
Network structure	{32, 32}
Training epochs	301
Activation	ReLU

Table A.1: Case 1. Hyperparameters for vOED-G.

Hyperparameter	vOED-NFs
N_{opt}	5000 / 10^4 / 2×10^4 / 5×10^4
N	10^4
N_{batch}	250 / 500 / 1000 / 2500
Initial learning rate	5×10^{-3} / 5×10^{-3} / 10^{-2} / 10^{-2}
Learning rate decay	0.99
Network structure (s & t)	$\{32, 32\}$
Training epochs	301
Activation	ELU

Table A.2: Case 1. Hyperparameters for vOED-NFs.

A.3.2 EIG estimated on training sample

Figure A.1 shows the EIG evaluated using the same samples it used for optimization (i.e., it is 'testing' on the same samples that it used for 'training'). We see that in this case, smaller N_{opt} can actually yield higher EIG estimates than Hi-NMC that uses $N_{\text{out}} = N_{\text{in}} = 2 \times 10^4$ samples, indicating an overfitting phenomenon where $\widehat{U}_L(\xi; \phi^*)$ could overestimate the EIG. Though in this case it does alter the location of the optimal design, the observation suggests the importance of evaluating EIG on a separate sample set than those used for optimization in order to reduce the bias in estimating the lower bound, especially when N_{opt} is small.

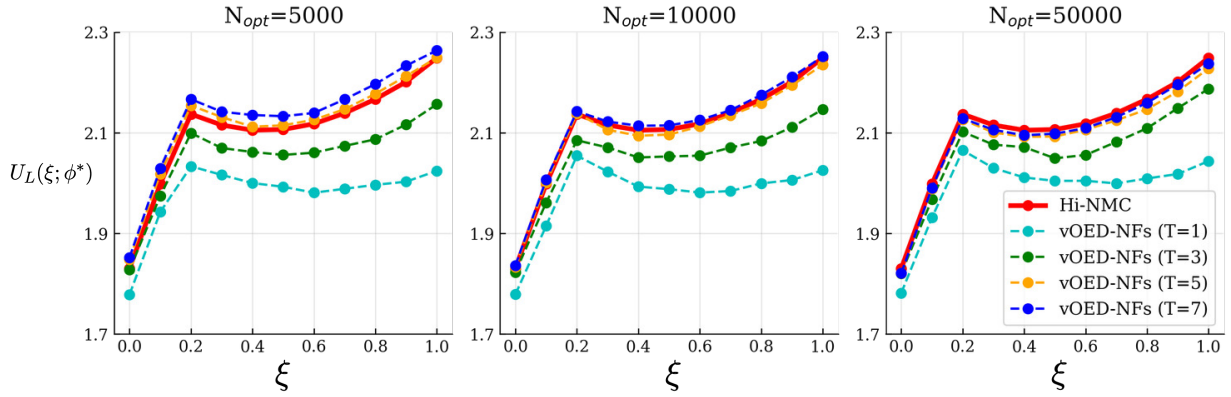


Figure A.1: Case 1. EIG under different number of transformations T and optimization sample sizes N_{opt}, evaluated using the same samples that it used for performing optimization.

A.4 Case 2

A.4.1 Hyperparameters

The hyperparameters for Sec. 2.3.2 are given in Tables A.3 and A.4.

Hyperparameter	vOED-GG
N_{opt}	5×10^4
N	10^4
N_{batch}	2048
Initial learning rate	5×10^{-3}
Learning rate decay	0.99
Network structure	$\{64, 64\}$
Training epochs	201 (Table 2.1) / 301 (Fig. 2.6)
Activation for vOED-GG	ReLU

Table A.3: Case 2. Hyperparameters for vOED-GG.

Hyperparam	vOED-NFs
N_{opt}	5×10^4
N	10^4
N_{batch}	2048 (Table 2.1) / 1024 (Fig. 2.7)
Initial learning rate	5×10^{-3}
Learning rate decay	0.99
Network structure vOED-NFs (s & t)	$\{32, 32\}$
T (numebr of transformations)	4
Training epochs	201 (Table 2.1) / 301 (Fig. 2.7)
Activation for vOED-NFs	ELU

Table A.4: Case 2. Hyperparameters for vOED-NFs.

A.5 Case 3

A.5.1 Forward model

This section introduces the forward model used in Sec. 2.3.3. The model solve for e-coating can be described by three steps. First, the electric field within the bath is computed using the conservation of current density. For a constant current e-coating, a Poisson PDE with Robin boundary condition at the interface bath/film and Neumann condition at the anode is solved and described by the following equations:

$$\nabla \cdot \mathbf{j} = 0 \tag{A.4}$$

$$\mathbf{j} = \sigma_{\text{bath}} \nabla \phi \tag{A.5}$$

$$\phi|_{\Gamma} = R_{\text{film}}\, j_n \qquad \text{at the interface film-bath} \tag{A.6}$$

$$j_n = j_0 \qquad \text{at the anode} \tag{A.7}$$

where $\mathbf{j}$ is the current density, $j_n = \mathbf{j} \cdot \mathbf{n}$ is the normal component of the current density, j_0 is the prescribed density current at the anode, ϕ is the electrical potential, σ_{bath} is the bath conductivity, R_{film} is the coating film resistance, and Γ represents the interface between the coating film and the bath. Second, the film deposition rate is computed using

$$\frac{dh}{dt} = C_v j_n, \tag{A.8}$$

where h is the film thickness and C_v is the Coulombic efficiency. Third, the film resistance is found by solving the following equation:

$$\frac{dR_{\text{film}}}{dt} = \rho(\mathbf{j}) \frac{dh}{dt}, \tag{A.9}$$

where $\rho(\mathbf{j})$ is the film resistivity. The coulombic efficiency C_v is assumed to be constant and the film resistivity $\rho(\mathbf{j})$ is a decreasing function of the current density.

The onset condition is critical for the prediction of the film deposition. Two criteria are used to evaluate the deposition onset. The first one considers a minimal value of the current density j_{min} that will trigger the deposition:

$$\frac{dh}{dt} = C_v\, j_n \text{ for } j > j_{\text{min}}. \tag{A.10}$$

The second condition is a minimum charge condition Q_{min} and assumes that the deposition starts

when the accumulative charge on the cathode reaches a minimum value as follows:

$$\frac{dh}{dt} = C_v\, j_n \text{ for } Q > Q_{\min}, \tag{A.11}$$

where the electric charge Q is defined as $Q(t) = \int_t j_n dt$. The minimum charge can be expressed as a function of the constant current j_0, with K a constant for a given bath and materials:

$$Q_{\min} = \frac{K^2}{j_0}. \tag{A.12}$$

The truncated normal priors for $\{C_v, j_{\min}, Q_{\min}\}$ are:

$$p(C_v) \sim p_{TN}(\mu = 7,\ \sigma = 0.5,\ l = 6,\ u = 8), \tag{A.13}$$

$$p(j_{\min}) \sim p_{TN}(\mu = 1,\ \sigma = 0.5,\ l = 0,\ u = 2), \tag{A.14}$$

$$p(Q_{\min}) \sim p_{TN}(\mu = 50, \sigma = 25,\ l = 0,\ u = 100), \tag{A.15}$$

where

$$p_{TN}(x;, \mu, \sigma, l, u) = \frac{1}{\sigma} \frac{\phi(\frac{x-\mu}{\sigma})}{\Phi(\frac{u-x}{\sigma}) - \Phi(\frac{l-x}{\sigma})}$$

for $x \in [l, u]$, and equals 0 otherwise. Here $\phi(\cdot)$ denotes the PDF for a standard normal distribution and $\Phi(\cdot)$ is its cumulative distribution function.

A.5.2 Hyperparameters for the 10-dimensional case

For y with dimension 10, Table A.5 lists the hyperparameters used for the lower and upper bound estimators in vOED-NFs.

Hyperparam	vOED-NFs
N_{opt}	10^4 / 8×10^4
N	10^4
N_{batch}	512 / 2048
Initial learning rate	5×10^{-3}
Learning rate decay	0.99
Network structure (s & t)	$\{16, 16\}$
T	5
Training epochs	301
Activation for vOED-NFs	ELU

Table A.5: Case 3. Hyperparameters for the 10-dimensional case, both lower and upper bound estimators.

A.5.3 Hyperparameters for the 50-dimensional case

For y with dimension 50, Table A.6 lists the hyperparameters used for the lower bound estimators in vOED-G, and Table A.7 lists the hyperparameters used for the lower and upper bound estimators in vOED-NFs, without the LSTM summary network.

Hyperparameter	vOED-G
N_{opt}	10^4
N	10^4
N_{batch}	256
Initial learning rate	1×10^{-3}
Learning rate decay	0.99
Network structure	$\{32, 32\}$
Training epochs	301
Activation for vOED-G	ReLU

Table A.6: Case 3. Hyperparameters for vOED-G for the 50-dimensional case, for the lower bound estimator.

Hyperparameter	vOED-NFs
N_{opt}	10^4 / 8×10^4
N	10^4
N_{batch}	512 / 2048
Initial learning rate	$1 \times 10^{-3}/5 \times 10^{-3}$
Learning rate decay	0.99
Network structure (s & t)	$\{16, 16\}$
T	5
Training epochs	301
Activation for vOED-NFs	ELU

Table A.7: Case 3. Hyperparameters for vOED-NFs for the 50-dimensional case, for both lower and upper bound estimators without LSTM summary network.

Table A.8 lists the hyperparameters for vOED-NFs with the LSTM summary network. The initial learning rates reported in both tables with or without LSTM empirically yield the highest lower bound estimates among the learning rate values $\{5\times10^{-4}, 1\times10^{-3}, 5\times10^{-3}, 1\times10^{-2}\}$. To get a dimension reduced 10-dimensional y', we first feed y as input to a bidirectional LSTM network, and then concatenate the resulting context vectors, which is further fed into the final embedding network emanating y'.

A.5.4 Posteriors from vOED-NFs with LSTM summary network and $N_{\text{opt}} = 8 \times 10^4$

Figure A.2 plots vOED-NFs posteriors when using LSTM summary network and $N_{\text{opt}} = 8 \times 10^4$. The posteriors match the SMC posteriors (reference) better compared to those trained with smaller N_{opt} and when y is higher dimensional (i.e., from Fig. 2.12).

Hyperparameters	vOED-NFs-LSTM
N_{opt}	10^4 / 8×10^4
N	10^4
N_{batch}	512 / 2048
Initial learning rate	5×10^{-3}/1×10^{-2}
Learning rate decay	0.99
Network structure (s & t)	{16, 16}
T	5
Training epochs	301
Activation for vOED-NFs-LSTM	ELU
LSTM-n_feature	1
LSTM-hidden_dim	20
LSTM-num_layers	1
Embedding network structure	40, 20
Activation for embedding network	ELU

Table A.8: Case 3. Hyperparameters for vOED-NFs for the 50-dimensional case, for lower bound estimator with LSTM summary network.

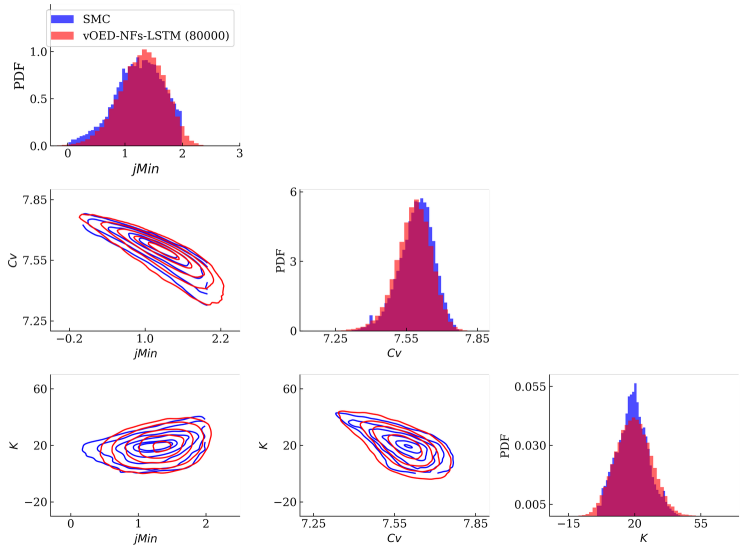


Figure A.2: Case 3. Marginal and pairwise joint posterior distributions obtained using SMC (high-quality reference) and vOED-NFs-LSTM with $N_{\text{opt}} = 80000$.

A.6 Case 4

A.6.1 Hyperparameters

The hyperparameters for Sec. 2.3.4 are given in Table A.9. Note LB-KLD [5] is computed using 10^4 prior samples, each associated with 3 data samples such that the overall number of forward model runs is 3×10^4. For vOED-NFs, we keep the total number of forward model runs at 3×10^4 ($N_{\text{opt}} = 2 \times 10^4$ and $N = 10^4$).

Hyperparameters	vOED-NFs
N_{opt}	2×10^4
N	10^4
N_{batch}	2048
Initial learning rate	1×10^{-2}
Learning rate decay	0.99
Network structure (s & t)	$\{16, 16\}$
T	4
Training epochs	51
Activation for vOED-NFs	ELU

Table A.9: Case 4. Hyperparameters for vOED-NFs.

A.7 Training Time for EIG Estimation

This section provides a sense of the computational cost in terms of "running time" for EIG evaluation at a specific design. We note that these figures are dependent on the implementation, such as which packages are being used, which training hyperparameters are adopted, whether code is optimized or parallelized, etc. Therefore, we suggest it to be used for awareness only, rather than direct comparisons. All numerical comparisons are run on the University of Michigan Great Lakes Slurm HPC Cluster nodes, with each node equipped with a single Nvidia Tesla A40 or V100 GPU.

For case 1 in Sec. 5.2.1, Fig. 2.2a, the EIG evaluation time at a specific design for approaches (a)–(e) are approximately: (a) (Hi-NMC) 57 seconds, (b) (Lo-NMC) less than 0.1 seconds, (c) (Re-NMC) 16.5 seconds, (d) (vOED-G) 43 seconds, and (e) 1 minute 38 seconds. For this case, the forward model run is extremely fast and thus the computational cost from vOED is mainly contributed by the optimization process.

For case 2 in Sec. 2.3.2, Table 2.1, the improvement in EIG and posterior approximation of vOED-NFs over vOED-GG comes at the expense of increased training time, with vOED-GG requiring approximately 1 minute and 39 seconds per EIG evaluation, while vOED-NFs takes around 5 minutes and 56 seconds under the selected hyperparameters.

For case 3 in Sec. 2.3.3, Fig. 2.8, the training times for (a) vsOED-NFs (upper), $N_{\text{opt}} = 10^4$, (b) vsOED-NFs (upper), $N_{\text{opt}} = 8 \times 10^4$, (c) vsOED-NFs (lower), $N_{\text{opt}} = 10^4$, (d) vsOED-NFs (lower), $N_{\text{opt}} = 8 \times 10^4$ are respectively 2 minutes 32 seconds, 5 minutes 59 seconds, 3 minutes 50 seconds, and 6 minutes 48 seconds under the selected hyperparameters. Whereas for the Hi-NMC estimator which requires 10^5 forward model runs in total, each run involves solving a PDE and takes approximately 1.5 seconds. Therefore, the additional time induced by training NFs is trivial compared to the overall computational cost of running the forward models for Hi-NMC.

For case 3, Fig. 2.12, with LSTM incorporated via summary network for evaluating the EIG lower bounds, the training times for (a) vOED-NFs, $N_{\text{opt}} = 10^4$, (b) vOED-NFs, $N_{\text{opt}} = 8 \times 10^4$, (c) vOED-NFs-LSTM, $N_{\text{opt}} = 10^4$, (d) vOED-NFs-LSTM, $N_{\text{opt}} = 8 \times 10^4$ are respectively 4 minutes 33 seconds, 6 minutes 19 seconds, 5 minutes 13 seconds, and 7 minutes 38 seconds under the selected hyperparameters.

Lastly, for case 4 in Sec. 2.3.4, Fig. 2.13, the training time for vOED-NFs is approximately 15 seconds.

CHAPTER B

Appendix for vsOED

B.1 Proofs

B.1.1 Information gain jointly with model indicator

Akin to the total entropy in [19], the IG jointly on the model indicator and PoIs is:

$$
\begin{aligned}
&D_{\mathrm{KL}}\left(P_{M,\Theta_m|\mathrm{i}_{t_2}} \,||\, P_{M,\Theta_m|\mathrm{i}_{t_1}}\right) \\
&= \sum_{m=1}^{|\mathcal{M}_m|} \int p(m,\theta_m|\mathrm{i}_{t_2}) \log \frac{p(m,\theta_m|\mathrm{i}_{t_2})}{p(m,\theta_m|\mathrm{i}_{t_1})} \,\mathrm{d}\theta_m \\
&= \sum_{m=1}^{|\mathcal{M}_m|} P(m|\mathrm{i}_{t_2}) \int p(\theta_m|m,\mathrm{i}_{t_2}) \log \frac{P(m|\mathrm{i}_{t_2})\, p(\theta_m|m,\mathrm{i}_{t_2})}{P(m|\mathrm{i}_{t_1})\, p(\theta_m|m,\mathrm{i}_{t_1})} \,\mathrm{d}\theta_m \\
&= \sum_{m=1}^{|\mathcal{M}_m|} P(m|\mathrm{i}_{t_2}) \log \frac{P(m|\mathrm{i}_{t_2})}{P(m|\mathrm{i}_{t_1})} + \sum_{m=1}^{|\mathcal{M}_m|} P(m|\mathrm{i}_{t_2}) \int p(\theta_m|m,\mathrm{i}_{t_2}) \log \frac{p(\theta_m|m,\mathrm{i}_{t_2})}{p(\theta_m|m,\mathrm{i}_{t_1})} \,\mathrm{d}\theta_m \\
&= D_{\mathrm{KL}}\left(P_{M|\mathrm{i}_{t_2}} \,||\, P_{M|\mathrm{i}_{t_1}}\right) + \mathbb{E}_{M|\mathrm{i}_{t_2}}\left[D_{\mathrm{KL}}\left(p_{\Theta_m|\mathrm{i}_{t_2}} \,||\, p_{\Theta_m|\mathrm{i}_{t_1}}\right)\right],
\end{aligned}
$$

where $0 \le t_1 \le t_2 \le T$, and $p(\theta_m|m,\mathrm{i}_t) = p(\theta_m|\mathrm{i}_t)$ per our notation convention. When setting $t_1 = 0$ and $t_2 = T$, we recover the terminal reward in Eqn. (3.8) under the special case of $\alpha_M = \alpha_\Theta = 1$

and $\alpha_Z = 0$. Similarly, the IG jointly on the model indicator and QoIs is:

$$
\begin{aligned}
&D_{\mathrm{KL}}\left(p_{M,Z_m|\mathrm{i}_{t_2}} \,||\, p_{M,Z_m|\mathrm{i}_{t_1}}\right) \\
&= \sum_{m=1}^{|\mathcal{M}_m|} \int p(m, z_m|\mathrm{i}_{t_2}) \log \frac{p(m, z_m|\mathrm{i}_{t_2})}{p(m, z_m|\mathrm{i}_{t_1})} \,\mathrm{d}z_m \\
&= \sum_{m=1}^{|\mathcal{M}_m|} P(m|\mathrm{i}_{t_2}) \int p(z_m|m, \mathrm{i}_{t_2}) \log \frac{P(m|\mathrm{i}_{t_2})\, p(z_m|m, \mathrm{i}_{t_2})}{P(m|\mathrm{i}_{t_1})\, p(z_m|m, \mathrm{i}_{t_1})} \,\mathrm{d}z_m \\
&= \sum_{m=1}^{|\mathcal{M}_m|} P(m|\mathrm{i}_{t_2}) \frac{P(m|\mathrm{i}_{t_2})}{P(m|\mathrm{i}_{t_1})} + \sum_{m=1}^{|\mathcal{M}_m|} P(m|\mathrm{i}_{t_2}) \int p(z_m|m, \mathrm{i}_{t_2}) \log \frac{p(z_m|m, \mathrm{i}_{t_2})}{p(z_m|m, \mathrm{i}_{t_1})} \,\mathrm{d}z_m \\
&= D_{\mathrm{KL}}\left(P_{M|\mathrm{i}_{t_2}} \,||\, P_{M|\mathrm{i}_{t_1}}\right) + \mathbb{E}_{M|\mathrm{i}_{t_2}}\left[D_{\mathrm{KL}}\left(p_{Z_m|\mathrm{i}_{t_2}} \,||\, p_{Z_m|\mathrm{i}_{t_1}}\right)\right],
\end{aligned}
$$

where $0 \le t_1 \le t_2 \le T$, and $p(z_m|m, \mathrm{i}_t) = p(z_m|\mathrm{i}_t)$ per our notation convention. When setting $t_1 = 0$ and $t_2 = T$, we recover the terminal reward in Eqn. (3.8) under the special case of $\alpha_M = \alpha_Z = 1$ and $\alpha_\Theta = 0$.

B.1.2 Information gain jointly on PoIs and QoIs

When nuisance parameters Ψ_m are absent, the IG jointly on the PoIs and QoIs given model m is:

$$
\begin{aligned}
&D_{\mathrm{KL}}\left(p_{\Theta_m,Z_m|\mathrm{i}_{t_2}} \,||\, p_{\Theta_m,Z_m|\mathrm{i}_{t_1}}\right) \\
&= \iint p(\theta_m, z_m|\mathrm{i}_{t_2}) \log \frac{p(\theta_m, z_m|\mathrm{i}_{t_2})}{p(\theta_m, z_m|\mathrm{i}_{t_1})} \,\mathrm{d}z_m \,\mathrm{d}\theta_m \\
&= \iint p(\theta_m, z_m|\mathrm{i}_{t_2}) \log \frac{p(\theta_m|\mathrm{i}_{t_2})\, p(z_m|\theta_m, \mathrm{i}_{t_2})}{p(\theta_m|\mathrm{i}_{t_1})\, p(z_m|\theta_m, \mathrm{i}_{t_1})} \,\mathrm{d}z_m \,\mathrm{d}\theta_m \\
&= \iint p(\theta_m, z_m|\mathrm{i}_{t_2}) \log \frac{p(\theta_m|\mathrm{i}_{t_2})\, p(z_m|\theta_m)}{p(\theta_m|\mathrm{i}_{t_1})\, p(z_m|\theta_m)} \,\mathrm{d}z_m \,\mathrm{d}\theta_m \\
&= \iint p(\theta_m, z_m|\mathrm{i}_{t_2}) \log \frac{p(\theta_m|\mathrm{i}_{t_2})}{p(\theta_m|\mathrm{i}_{t_1})} \,\mathrm{d}z_m \,\mathrm{d}\theta_m \\
&= \int p(\theta_m|\mathrm{i}_{t_2}) \log \frac{p(\theta_m|\mathrm{i}_{t_2})}{p(\theta_m|\mathrm{i}_{t_1})} \,\mathrm{d}\theta_m \\
&= D_{\mathrm{KL}}\left(p_{\Theta_m|\mathrm{i}_{t_2}} \,||\, p_{\Theta_m|\mathrm{i}_{t_1}}\right),
\end{aligned}
$$

where $0 \le t_1 \le t_2 \le T$, and the third equality is due to Z_m only depending on Θ_m when Ψ_m is absent (see Eqn. (3.2)). Hence, the IG on the QoIs is fully absorbed into the IG on the PoIs when nuisance parameters are absent.

B.1.3 Proof of Theorem 1 (Terminal-incremental equivalence)

Proof. We first decompose $U_T(\pi)$ into four summing parts:

$$U_T(\pi) = U_{T,\mathrm{NIG}}(\pi) + U_{T,M}(\pi) + U_{T,\Theta}(\pi) + U_{T,Z}(\pi),$$

where $U_{T,\mathrm{NIG}}(\pi)$ captures any non-IG reward contributions, and the other three parts are (with the conditioning on I_0 written out explicitly for the prior terms)

$$\begin{aligned}
U_{T,M}(\pi) &= \alpha_M \mathbb{E}_{Y_{0:T-1}|\pi,s_0}\left[D_{\mathrm{KL}}\left(P_{M|I_T} \,||\, P_{M|I_0}\right)\right], \\
U_{T,\Theta}(\pi) &= \alpha_\Theta \mathbb{E}_{Y_{0:T-1}|\pi,s_0}\mathbb{E}_{M|I_T}\left[D_{\mathrm{KL}}\left(p_{\Theta_m|I_T} \,||\, p_{\Theta_m|I_0}\right)\right], \\
U_{T,Z}(\pi) &= \alpha_Z \mathbb{E}_{Y_{0:T-1}|\pi,s_0}\mathbb{E}_{M|I_T}\left[D_{\mathrm{KL}}\left(p_{Z_m|I_T} \,||\, p_{Z_m|I_0}\right)\right].
\end{aligned}$$

Similarly, $U_I(\pi)$ is also decomposed into four summing parts:

$$U_I(\pi) = U_{I,\mathrm{NIG}}(\pi) + U_{I,M}(\pi) + U_{I,\Theta}(\pi) + U_{I,Z}(\pi),$$

where $U_{I,\mathrm{NIG}}(\pi)$ captures any non-IG reward contributions, and the other three parts are

$$\begin{aligned}
U_{I,M}(\pi) &= \alpha_M \mathbb{E}_{Y_{0:T-1}|\pi,s_0}\left[\sum_{t=0}^{T-1} D_{\mathrm{KL}}\left(P_{M|I_{t+1}} \,||\, P_{M|I_t}\right)\right] \\
U_{I,\Theta}(\pi) &= \alpha_\Theta \mathbb{E}_{Y_{0:T-1}|\pi,s_0}\left[\sum_{t=0}^{T-1} \mathbb{E}_{M|I_{t+1}}\left[D_{\mathrm{KL}}\left(p_{\Theta_m|I_{t+1}} \,||\, p_{\Theta_m|I_t}\right)\right]\right] \\
U_{I,Z}(\pi) &= \alpha_Z \mathbb{E}_{Y_{0:T-1}|\pi,s_0}\left[\sum_{t=0}^{T-1} \mathbb{E}_{M|I_{t+1}}\left[D_{\mathrm{KL}}\left(p_{Z_m|I_{t+1}} \,||\, p_{Z_m|I_t}\right)\right]\right].
\end{aligned}$$

Since TIG and IIG formulations only entail the IG contributions, the non-IG reward contributions are not affected by this choice, and hence

$$U_{T,\mathrm{NIG}}(\pi) = U_{I,\mathrm{NIG}}(\pi).$$

For the part corresponding to the IG contribution from model indicator:

$$
\begin{aligned}
&U_{I,M}(\pi) - U_{T,M}(\pi) \\
&= \alpha_M \mathbb{E}_{Y_{0:T-1}|\pi,s_0}\left[\sum_{t=0}^{T-1} D_{\mathrm{KL}}\left(P_{M|I_{t+1}} \,||\, P_{M|I_t}\right) - D_{\mathrm{KL}}\left(P_{M|I_T} \,||\, P_{M|I_0}\right)\right] \\
&= \alpha_M \mathbb{E}_{Y_{0:T-1}|\pi,s_0}\left[\sum_{t=0}^{T-1}\sum_{m=1}^{|\mathcal{M}_m|} P(m|I_{t+1}) \log\frac{P(m|I_{t+1})}{P(m|I_t)} - \sum_{m=1}^{|\mathcal{M}_m|} P(m|I_T)\log\frac{P(m|I_T)}{P(m|I_0)}\right] \\
&= \alpha_M \sum_{m=1}^{|\mathcal{M}_m|} \mathbb{E}_{Y_{0:T-1}|\pi,s_0}\left[\sum_{t=0}^{T-1} P(m|I_{t+1}) \log\frac{P(m|I_{t+1})}{P(m|I_t)} - P(m|I_T)\log\frac{P(m|I_T)}{P(m|I_0)}\right] \\
&= \alpha_M \sum_{m=1}^{|\mathcal{M}_m|} \mathbb{E}_{Y_{0:T-1}|\pi,s_0}\left[\sum_{t=0}^{T-2} P(m|I_{t+1}) \log\frac{P(m|I_{t+1})}{P(m|I_t)} + P(m|I_T)\log\frac{P(m|I_T)}{P(m|I_{T-1})}\right. \\
&\qquad\qquad \left. - P(m|I_T)\log\frac{P(m|I_T)}{P(m|I_0)}\right] \\
&= \alpha_M \sum_{m=1}^{|\mathcal{M}_m|} \mathbb{E}_{Y_{0:T-1}|\pi,s_0}\left[\sum_{t=0}^{T-2} P(m|I_{t+1}) \log\frac{P(m|I_{t+1})}{P(m|I_t)} - P(m|I_T)\log\frac{P(m|I_{T-1})}{P(m|I_0)}\right] \\
&= \alpha_M \sum_{m=1}^{|\mathcal{M}_m|} \mathbb{E}_{Y_{0:T-2}|\pi,s_0}\left[\sum_{t=0}^{T-2} P(m|I_{t+1}) \log\frac{P(m|I_{t+1})}{P(m|I_t)}\right. \\
&\qquad\qquad \left. - \mathbb{E}_{Y_{T-1}|Y_{0:T-2},\pi,s_0}\left[P(m|I_T)\log\frac{P(m|I_{T-1})}{P(m|I_0)}\right]\right] \\
&= \alpha_M \sum_{m=1}^{|\mathcal{M}_m|} \mathbb{E}_{Y_{0:T-2}|\pi,s_0}\left[\sum_{t=0}^{T-2} P(m|I_{t+1}) \log\frac{P(m|I_{t+1})}{P(m|I_t)} - P(m|I_{T-1})\log\frac{P(m|I_{T-1})}{P(m|I_0)}\right] \\
&\;\;\vdots \\
&= \alpha_M \sum_{m=1}^{|\mathcal{M}_m|} \mathbb{E}_{Y_0|\pi,s_0}\left[\sum_{t=0}^{0} P(m|I_{t+1}) \log\frac{P(m|I_{t+1})}{P(m|I_t)} - P(m|I_1)\log\frac{P(m|I_1)}{P(m|I_0)}\right] \\
&= 0
\end{aligned}
$$

where the seventh equality is due to

$$
\begin{aligned}
&\mathbb{E}_{Y_{T-1}|Y_{0:T-2},\pi,s_0}\left[P(m|I_T)\log\frac{P(m|I_{T-1})}{P(m|I_0)}\right] \\
&= \int p(y_{T-1}|Y_{0:T-2},\pi,s_0)P(m|I_T)\log\frac{P(m|I_{T-1})}{P(m|I_0)}\,\mathrm{d}y_{T-1} \\
&= \int p(y_{T-1}|\pi,I_{T-1})P(m|y_{T-1},\pi,I_{T-1})\log\frac{P(m|I_{T-1})}{P(m|I_0)}\,\mathrm{d}y_{T-1} \\
&= \int p(y_{T-1},m|\pi,I_{T-1})\log\frac{P(m|I_{T-1})}{P(m|I_0)}\,\mathrm{d}y_{T-1} \\
&= P(m|I_{T-1})\log\frac{P(m|I_{T-1})}{P(m|I_0)},
\end{aligned}
$$

and the eighth equality results from repeatedly applying the steps between the third and seventh equalities $T-1$ times.

For the part corresponding to the IG contribution from PoIs:

$$
\begin{aligned}
&U_{I,\Theta}(\pi) - U_{T,\Theta}(\pi) \\
&= \alpha_\Theta \mathbb{E}_{Y_{0:T-1}|\pi,s_0}\left[\sum_{t=0}^{T-1} \mathbb{E}_{M|I_{t+1}} D_{\mathrm{KL}}\left(p_{\Theta_m|I_{t+1}} \,||\, p_{\Theta_m|I_t}\right) - \mathbb{E}_{M|I_T} D_{\mathrm{KL}}\left(p_{\Theta_m|I_T} \,||\, p_{\Theta_m|I_0}\right)\right] \\
&= \alpha_\Theta \sum_{m=1}^{|\mathcal{M}_m|} \mathbb{E}_{Y_{0:T-1}|\pi,s_0}\Bigg[\sum_{t=0}^{T-1} P(m|I_{t+1}) D_{\mathrm{KL}}\left(p_{\Theta_m|I_{t+1}} \,||\, p_{\Theta_m|I_t}\right) \\
&\qquad\qquad - P(m|I_T) D_{\mathrm{KL}}\left(p_{\Theta_m|I_T} \,||\, p_{\Theta_m|I_0}\right)\Bigg] \\
&= \alpha_\Theta \sum_{m=1}^{|\mathcal{M}_m|} \mathbb{E}_{Y_{0:T-1}|\pi,s_0}\Bigg[\sum_{t=0}^{T-2} P(m|I_{t+1}) D_{\mathrm{KL}}\left(p_{\Theta_m|I_{t+1}} \,||\, p_{\Theta_m|I_t}\right) \\
&\qquad\qquad + P(m|I_T) D_{\mathrm{KL}}\left(p_{\Theta_m|I_T} \,||\, p_{\Theta_m|I_{T-1}}\right) \\
&\qquad\qquad - P(m|I_T) D_{\mathrm{KL}}\left(p_{\Theta_m|I_T} \,||\, p_{\Theta_m|I_0}\right)\Bigg] \\
&= \alpha_\Theta \sum_{m=1}^{|\mathcal{M}_m|} \mathbb{E}_{Y_{0:T-1}|\pi,s_0}\Bigg[\sum_{t=0}^{T-2} P(m|I_{t+1}) D_{\mathrm{KL}}\left(p_{\Theta_m|I_{t+1}} \,||\, p_{\Theta_m|I_t}\right) \\
&\qquad\qquad + P(m|I_T) \int p(\theta_m|I_T) \log \frac{p(\theta_m|I_0)}{p(\theta_m|I_{T-1})} \,\mathrm{d}\theta_m\Bigg] \\
&= \alpha_\Theta \sum_{m=1}^{|\mathcal{M}_m|} \mathbb{E}_{Y_{0:T-2}|\pi,s_0}\Bigg[\sum_{t=0}^{T-2} P(m|I_{t+1}) D_{\mathrm{KL}}\left(p_{\Theta_m|I_{t+1}} \,||\, p_{\Theta_m|I_t}\right) \\
&\qquad\qquad + \mathbb{E}_{Y_{T-1}|Y_{0:T-2},\pi,s_0}\left[P(m|I_T) \int p(\theta_m|I_T) \log \frac{p(\theta_m|I_0)}{p(\theta_m|I_{T-1})} \,\mathrm{d}\theta_m\right]\Bigg] \\
&= \alpha_\Theta \sum_{m=1}^{|\mathcal{M}_m|} \mathbb{E}_{Y_{0:T-2}|\pi,s_0}\Bigg[\sum_{t=0}^{T-2} P(m|I_{t+1}) D_{\mathrm{KL}}\left(p_{\Theta_m|I_{t+1}} \,||\, p_{\Theta_m|I_t}\right) \\
&\qquad\qquad - P(m|I_{T-1}) D_{\mathrm{KL}}\left(p_{\Theta_m|I_{T-1}} \,||\, p_{\Theta_m|I_0}\right)\Bigg] \\
&\quad\vdots \\
&= \alpha_\Theta \sum_{m=1}^{|\mathcal{M}_m|} \mathbb{E}_{Y_0|\pi,s_0}\left[\sum_{t=0}^{0} P(m|I_{t+1}) D_{\mathrm{KL}}\left(p_{\Theta_m|I_{t+1}} \,||\, p_{\Theta_m|I_t}\right) - P(m|I_1) D_{\mathrm{KL}}\left(p_{\Theta_m|I_1} \,||\, p_{\Theta_m|I_0}\right)\right] \\
&= 0,
\end{aligned}
$$

where the sixth equality is due to

$$
\begin{aligned}
&\mathbb{E}_{Y_{T-1}|Y_{0:T-2},\pi,s_0}\left[P(m|I_T)\int p(\theta_m|I_T)\log\frac{p(\theta_m|I_0)}{p(\theta_m|I_{T-1})}\,\mathrm{d}\theta_m\right] \\
&= \int P(y_{T-1}|Y_{0:T-2},\pi,s_0)P(m|I_T)\int p(\theta_m|I_T)\log\frac{p(\theta_m|I_0)}{p(\theta_m|I_{T-1})}\,\mathrm{d}\theta_m\,\mathrm{d}y_{T-1} \\
&= \int P(y_{T-1}|\pi,I_{T-1})P(m|y_{T-1},\pi,I_{T-1})\int p(\theta_m|m,y_{T-1},\pi,I_{T-1})\log\frac{p(\theta_m|I_0)}{p(\theta_m|I_{T-1})}\,\mathrm{d}\theta_m\,\mathrm{d}y_{T-1} \\
&= \iint P(y_{T-1},m,\theta_m|\pi,I_{T-1})\log\frac{p(\theta_m|I_0)}{p(\theta_m|I_{T-1})}\,\mathrm{d}\theta_m\,\mathrm{d}y_{T-1} \\
&= P(m|I_{T-1})\int P(\theta_m|I_{T-1})\log\frac{p(\theta_m|I_0)}{p(\theta_m|I_{T-1})}\,\mathrm{d}\theta_m \\
&= -P(m|I_{T-1})D_{\mathrm{KL}}\left(p_{\Theta_m|I_{T-1}}\,||\,p_{\Theta_m|I_0}\right),
\end{aligned}
$$

and the seventh equality results from repeatedly applying the steps between the second and sixth equalities $T-1$ times.

For the part corresponding to the IG contribution from QoIs, the derivation is identical as above for the PoIs except Θ is replaced with Z, to arrive at

$$U_{I,Z}(\pi) - U_{T,Z}(\pi) = 0.$$

Combining the equivalence results from all four parts, we obtain

$$U_I(\pi) = U_T(\pi)$$

for any policy π. □

B.1.4 Proof of Theorem 2 (One-point equivalence)

We begin by proving the equivalence of expected utility under TIG and one-point-TIG:

Proof.

$$
\begin{aligned}
U_T(\pi) &= \mathbb{E}_{Y_{0:T-1}|\pi,s_0}\left[\sum_{t=0}^{T-1} r_t(S_t,\xi_t,y_t) + r_T(S_T)\right] \\
&= \mathbb{E}_{Y_{0:T-1}|\pi,s_0}\Bigg[\alpha_M D_{\mathrm{KL}}(\, P_{M|I_T} \,||\, P_M \,) \\
&\qquad\qquad + \mathbb{E}_{M|I_T}\left[\alpha_\Theta D_{\mathrm{KL}}(\, p_{\Theta_m|I_T} \,||\, p_{\Theta_m} \,) + \alpha_Z D_{\mathrm{KL}}(\, p_{Z_m|I_T} \,||\, p_{Z_m} \,)\right] \\
&= \mathbb{E}_{Y_{0:T-1}|\pi,s_0}\Bigg[\alpha_M \mathbb{E}_{M|I_T} \log \frac{P(M|I_T)}{P(M)} \\
&\qquad\qquad + \mathbb{E}_{M|I_T}\left[\alpha_\Theta \mathbb{E}_{\Theta_m|I_T} \log \frac{p(\Theta_m|I_T)}{p(\Theta_m)} + \alpha_Z \mathbb{E}_{Z_m|I_T} \log \frac{p(Z_m|I_T)}{p(Z_m)}\right]\Bigg] \\
&= \mathbb{E}_{M,Y_{0:T-1}|\pi,s_0}\Bigg[\alpha_M \log \frac{P(M|I_T)}{P(M)} \\
&\qquad\qquad + \alpha_\Theta \mathbb{E}_{\Theta_m|I_T} \log \frac{p(\Theta_m|I_T)}{p(\Theta_m)} + \alpha_Z \mathbb{E}_{Z_m|I_T} \log \frac{p(Z_m|I_T)}{p(Z_m)}\Bigg] \\
&= \mathbb{E}_{M,Y_{0:T-1}|\pi,s_0}\Bigg[\alpha_M \mathbb{E}_{\Theta_m,Z_m|I_T} \log \frac{P(M|I_T)}{P(M)} \\
&\qquad\qquad + \alpha_\Theta \mathbb{E}_{\Theta_m,Z_m|I_T} \log \frac{p(\Theta_m|I_T)}{p(\Theta_m)} + \alpha_Z \mathbb{E}_{\Theta_m,Z_m|I_T} \log \frac{p(Z_m|I_T)}{p(Z_m)}\Bigg] \\
&= \mathbb{E}_{M,\Theta_m,Z_m,Y_{0:T-1}|\pi,s_0}\left[\alpha_M \log \frac{P(M|I_T)}{P(M)} + \alpha_\Theta \log \frac{p(\Theta_m|I_T)}{p(\Theta_m)} + \alpha_Z \log \frac{p(Z_m|I_T)}{p(Z_m)}\right] \\
&= \mathbb{E}_{M,\Theta_m,\Psi_m,Z_m,Y_{0:T-1}|\pi,s_0}\left[\alpha_M \log \frac{P(M|I_T)}{P(M)} + \alpha_\Theta \log \frac{p(\Theta_m|I_T)}{p(\Theta_m)} + \alpha_Z \log \frac{p(Z_m|I_T)}{p(Z_m)}\right] \\
&= \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}}\mathbb{E}_{Y_{0:T-1}|\pi,s_0,M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}}\Bigg[\alpha_M \log \frac{P(M_0|I_T)}{P(M_0)} \\
&\qquad\qquad + \alpha_\Theta \log \frac{p(\Theta_{m,0}|I_T)}{p(\Theta_{m,0})} + \alpha_Z \log \frac{p(Z_{m,0}|I_T)}{p(Z_{m,0})}\Bigg] \\
&= \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}}\mathbb{E}_{I_T|\pi,s_0,M_0,\Theta_{m,0},\Psi_{m,0}}\Bigg[\alpha_M \log \frac{P(M_0|I_T)}{P(M_0)} \\
&\qquad\qquad + \alpha_\Theta \log \frac{p(\Theta_{m,0}|I_T)}{p(\Theta_{m,0})} + \alpha_Z \log \frac{p(Z_{m,0}|I_T)}{p(Z_{m,0})}\Bigg] \\
&= \check{U}_T(\pi),
\end{aligned}
$$

where at several occasions we use the equivalence between $\mathbb{E}_{Y_{0:t-1}|\pi,s_0,\dots}$ and $\mathbb{E}_{I_t|\pi,s_0,\dots}$, and between $\mathbb{E}_{\dots|I_t}$ and $\mathbb{E}_{\dots|Y_{0:t-1},\pi,s_0}$, in the eighth equality the random variables M, Θ_m, Ψ_m, Z_m are replaced with their oracle versions M_0, $\Theta_{m,0}$, $\Psi_{m,0}$, $Z_{m,0}$ which have identical distributions, and the ninth equality omits the conditioning on $Z_{m,0}$ in the inner expectation since I_T does not depend on $Z_{m,0}$.

Next, we have $U_T(\pi) = U_I(\pi)$ through Theorem 3.1.1 and its proof in Sec. B.1.3. Finally, we show the equivalence between $\check{U}_I(\pi)$ and $\check{U}_T(\pi)$ by cancelling out all intermediate posteriors:

$$
\begin{aligned}
&\check{U}_I(\pi) \\
&= \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}} \mathbb{E}_{I_T|\pi,s_0,M_0,\Theta_{m,0},\Psi_{m,0}} \sum_{t=0}^{T-1} \left[\alpha_M \log \frac{P(M_0|I_{t+1})}{P(M_0|I_t)} + \alpha_\Theta \log \frac{p(\Theta_{m,0}|I_{t+1})}{p(\Theta_{m,0}|I_t)} \right. \\
&\qquad\qquad \left. + \alpha_Z \log \frac{p(Z_{m,0}|I_{t+1})}{p(Z_{m,0}|I_t)} \right] \\
&= \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}} \mathbb{E}_{I_T|\pi,s_0,M_0,\Theta_{m,0},\Psi_{m,0}} \left[\alpha_M \log \frac{P(M_0|I_T)}{P(M_0)} + \alpha_\Theta \log \frac{p(\Theta_{m,0}|I_T)}{p(\Theta_{m,0})} \right. \\
&\qquad\qquad \left. + \alpha_Z \log \frac{p(Z_{m,0}|I_T)}{p(Z_{m,0})} \right] \\
&= \check{U}_T(\pi).
\end{aligned}
$$

Combining the above equivalence results together, we arrive at

$$
U_T(\pi) = \check{U}_T(\pi) = \check{U}_I(\pi) = U_I(\pi)
$$

for any policy π. □

B.1.5 Omitting prior terms in the expected utility

When the prior terms $p(\cdot)$ are omitted, the expected utility is shifted by the amount:

$$
\begin{aligned}
&\mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}} \mathbb{E}_{I_T|\pi,s_0,M_0,\Theta_{m,0},\Psi_{m,0}} \left[\alpha_M \log P(M_0) + \alpha_\Theta \log p(\Theta_{m,0}) + \alpha_Z \log p(Z_{m,0}) \right] \\
&= \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}} \left[\alpha_M \log P(M_0) + \alpha_\Theta \log p(\Theta_{m,0}) + \alpha_Z \log p(Z_{m,0}) \right],
\end{aligned}
$$

which does not depend on the policy π. In other words, the expected utility is shifted by a constant. Hence, whether including or omitting the prior terms will not affect the optimized policy. The same result holds for both TIG and IIG rewards, and their one-point versions and variational one-point versions.

B.1.6 Proof of Theorem 3 (Variational lower bound)

Proof. First, we prove the equivalence between the expected utility under variational-one-point-TIG and variational-one-point-IIG:

$$
\begin{aligned}
&\tilde{U}_I(\pi;\phi) \\
&= \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}}\mathbb{E}_{I_T|\pi,s_0,M_0,\Theta_{m,0},\Psi_{m,0}} \sum_{t=0}^{T-1} \Bigg[\alpha_M \log \frac{q(M_0|I_{t+1};\phi_M)}{q(M_0|I_t;\phi_M)} \\
&\qquad\qquad + \alpha_\Theta \log \frac{q(\Theta_{m,0}|I_{t+1};\phi_{\Theta_m})}{q(\Theta_{m,0}|I_t;\phi_{\Theta_m})} + \alpha_Z \log \frac{q(Z_{m,0}|I_{t+1};\phi_{Z_m})}{q(Z_{m,0}|I_t;\phi_{Z_m})} \Bigg] \\
&= \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}}\mathbb{E}_{I_T|\pi,s_0,M_0,\Theta_{m,0},\Psi_{m,0}} \Bigg[\\
&\qquad \alpha_M \left(\log \frac{q(M_0|I_1;\phi_M)}{P(M_0)} + \log \frac{q(M_0|I_2;\phi_M)}{q(M_0|I_1;\phi_M)} + \cdots + \log \frac{q(M_0|I_T;\phi_M)}{q(M_0|I_{T-1};\phi_M)} \right) \\
&\qquad + \alpha_\Theta \left(\log \frac{q(\Theta_{m,0}|I_1;\phi_{\Theta_m})}{p(\Theta_{m,0})} + \log \frac{q(\Theta_{m,0}|I_2;\phi_{\Theta_m})}{q(\Theta_{m,0}|I_1;\phi_{\Theta_m})} + \cdots + \log \frac{q(\Theta_{m,0}|I_T;\phi_{\Theta_m})}{q(\Theta_{m,0}|I_{T-1};\phi_{\Theta_m})} \right) \\
&\qquad + \alpha_Z \left(\log \frac{q(Z_{m,0}|I_1;\phi_{Z_m})}{p(Z_{m,0})} + \log \frac{q(Z_{m,0}|I_2;\phi_{Z_m})}{q(Z_{m,0}|I_1;\phi_{Z_m})} + \cdots + \log \frac{q(Z_{m,0}|I_T;\phi_{Z_m})}{q(Z_{m,0}|I_{T-1};\phi_{Z_m})} \right) \Bigg] \\
&= \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}}\mathbb{E}_{I_T|\pi,s_0,M_0,\Theta_{m,0},\Psi_{m,0}} \Bigg[\alpha_M \log \frac{q(M_0|I_T;\phi_M)}{P(M_0)} \\
&\qquad\qquad + \alpha_\Theta \log \frac{q(\Theta_{m,0}|I_T;\phi_{\Theta_m})}{p(\Theta_{m,0})} + \alpha_Z \log \frac{q(Z_{m,0}|I_T;\phi_{Z_m})}{p(Z_{m,0})} \Bigg] \\
&= \tilde{U}_T(\pi;\phi).
\end{aligned}
$$

Due to the cancellation of all intermediate variational posteriors, only the prior and final variational posterior terms survive. Since the prior density is usually accessible analytically or omitted per Sec. B.1.5, the accuracy of the variational expected utility only depends on the quality of the final variational posterior approximations, $q(\cdot|I_T;\phi_{(\cdot)})$.

Next, we prove the lower bound for variational-one-point-TIG:

$$
\begin{aligned}
&U_T(\pi) - \tilde{U}_T(\pi;\phi) \\
&= \check{U}_T(\pi) - \tilde{U}_T(\pi;\phi) \\
&= \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}} \mathbb{E}_{I_T|\pi,s_0,M_0,\Theta_{m,0},\Psi_{m,0}} \Bigg[\alpha_M \log \frac{P(M_0|I_T)}{q(M_0|I_T;\phi_M)} + \alpha_\Theta \log \frac{p(\Theta_{m,0}|I_T)}{q(\Theta_{m,0}|I_T;\phi_{\Theta_m})} \\
&\qquad\qquad + \alpha_Z \log \frac{p(Z_{m,0}|I_T)}{q(Z_{m,0}|I_T;\phi_{Z_m})} \Bigg] \\
&= \alpha_M \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}} \mathbb{E}_{I_T|\pi,s_0,M_0,\Theta_{m,0},\Psi_{m,0}} \left[\log \frac{P(M_0|I_T)}{q(M_0|I_T;\phi_M)} \right] \\
&\quad + \alpha_\Theta \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}} \mathbb{E}_{I_T|\pi,s_0,M_0,\Theta_{m,0},\Psi_{m,0}} \left[\log \frac{p(\Theta_{m,0}|I_T)}{q(\Theta_{m,0}|I_T;\phi_{\Theta_m})} \right] \\
&\quad + \alpha_Z \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}} \mathbb{E}_{I_T|\pi,s_0,M_0,\Theta_{m,0},\Psi_{m,0}} \left[\log \frac{p(Z_{m,0}|I_T)}{q(Z_{m,0}|I_T;\phi_{Z_m})} \right] \\
&= \alpha_M \mathbb{E}_{M_0,I_T|\pi,s_0} \left[\log \frac{P(M_0|I_T)}{q(M_0|I_T;\phi_M)} \right] \\
&\quad + \alpha_\Theta \mathbb{E}_{M_0,\Theta_{m,0},I_T|\pi,s_0} \left[\log \frac{p(\Theta_{m,0}|I_T)}{q(\Theta_{m,0}|I_T;\phi_{\Theta_m})} \right] \\
&\quad + \alpha_Z \mathbb{E}_{M_0,Z_{m,0},I_T|\pi,s_0} \left[\log \frac{p(Z_{m,0}|I_T)}{q(Z_{m,0}|I_T;\phi_{Z_m})} \right] \\
&= \alpha_M \mathbb{E}_{I_T|\pi,s_0} \mathbb{E}_{M_0|I_T} \left[\log \frac{P(M_0|I_T)}{q(M_0|I_T;\phi_M)} \right] \\
&\quad + \alpha_\Theta \mathbb{E}_{M_0,I_T|\pi,s_0} \mathbb{E}_{\Theta_{m,0}|M_0,I_T} \left[\log \frac{p(\Theta_{m,0}|I_T)}{q(\Theta_{m,0}|I_T;\phi_{\Theta_m})} \right] \\
&\quad + \alpha_Z \mathbb{E}_{M_0,I_T|\pi,s_0} \mathbb{E}_{Z_{m,0}|M_0,I_T} \left[\log \frac{p(Z_{m,0}|I_T)}{q(Z_{m,0}|I_T;\phi_{Z_m})} \right] \\
&= \alpha_M \mathbb{E}_{I_T|\pi,s_0} \left[D_{\mathrm{KL}} \left(P_{M_0|I_T} \,||\, q_{M_0|I_T;\phi_M} \right) \right] \\
&\quad + \alpha_\Theta \mathbb{E}_{M_0,I_T|\pi,s_0} \left[D_{\mathrm{KL}} \left(p_{\Theta_{m,0}|I_T} \,||\, q_{\Theta_{m,0}|I_T;\phi_{\Theta_m}} \right) \right] \\
&\quad + \alpha_Z \mathbb{E}_{M_0,I_T|\pi,s_0} \left[D_{\mathrm{KL}} \left(p_{Z_{m,0}|I_T} \,||\, q_{Z_{m,0}|I_T;\phi_{Z_m}} \right) \right] \\
&\geq 0,
\end{aligned}
$$

where the first equality invokes Theorem 3.1.2, the sixth equality is due to $p(\Theta_{m,0}|I_T)$ being equivalent to $p(\Theta_{m,0}|M_0, I_T)$ and $p(Z_{m,0}|I_T)$ being equivalent to $p(Z_{m,0}|M_0, I_T)$, and the final inequality is due to the non-negativity of KL divergence terms and that we require $\alpha_M, \alpha_\Theta, \alpha_Z \in [0, 1]$. The bound is tight if and only if $q(\cdot|I_T; \phi_{(\cdot)}) = p(\cdot|I_T)$ and so all the KL divergence terms become zero; (except the trivial case when $\alpha_M = \alpha_\Theta = \alpha_Z = 0$, under which U, $\check{U}$, and $\tilde{U}$ will always be identically zero).

Putting everything together, we arrive at

$$\tilde{U}_I(\pi; \phi) = \tilde{U}_T(\pi; \phi) \leq \check{U}_T(\pi) = \check{U}_I(\pi) = U_T(\pi) = U_I(\pi),$$

for any π and ϕ, where the last three equalities result from Theorem 3.1.2. □

B.1.7 Summary of general symbols and notation

Please see Table B.1 for the general symbols and notation used in this paper.

B.2 Algorithm details

B.2.1 Neural network architecture for approximate posteriors of model indicator

The overall architecture for a NN-based approximate posterior of model indicator, $q(m|\mathsf{i}_t; \phi_M)$, is shown in Table B.2; the same architecture is used for all numerical cases in this paper. The NN takes i_t as input, and outputs the log-probabilities of each candidate model, $\log q(m|\mathsf{i}_t; \phi_M)$. Separate NNs are trained for each stage t when the IIG formulation is used. As shown in the first part of Sec. B.1.6, the quality of the intermediate approximate posteriors does not directly contribute to the accuracy of the overall variational expected utility, and thus one may train these intermediate approximate posteriors more 'roughly', for example, by using simpler NN architectures and with shared weights among the NNs.

B.2.2 Neural network architectures for approximate posteriors of PoIs and QoIs

We use GMMs and NFs as approximate posteriors for PoIs, $q(\theta_m|\mathsf{i}_t; \phi_{\Theta_m})$, and for QoIs, $q(z_m|\mathsf{i}_t; \phi_{Z_m})$[1]. The same architecture is used for both PoI and QoI cases, and we introduce

[1] In this section, we have dropped the oracle subscript 0 (from $\theta_{m,0}$ and $z_{m,0}$) to simplify notation.

below only in the context of PoIs for simplicity. Separate GMMs/NFs are trained for each stage t when the IIG formulation is used.

B.2.2.1 Independent Gaussian mixture models

An independent GMM approximates a complex distribution through a weighted sum of multiple independent Gaussians:

$$q(\theta_m|\mathsf{i}_t;\phi_{\Theta_m}) = \sum_{i=1}^{n_{\text{mix}}} w_i(\mathsf{i}_t;\phi_{\Theta_m})\,\mathcal{N}(\theta_m;\mu_i(\mathsf{i}_t;\phi_{\Theta_m}),\Sigma_i(\mathsf{i}_t;\phi_{\Theta_m})), \tag{B.1}$$

where for the ith Gaussian, $w_i(\mathsf{i}_t;\phi_{\Theta_m})$ is its mixture weight, $\mu_i(\mathsf{i}_t;\phi_{\Theta_m}) \in \mathbb{R}^{N_{\theta_m}}$ is its mean, and $\Sigma_i(\mathsf{i}_t;\phi_{\Theta_m}) \in \mathbb{R}^{N_{\theta_m}\times N_{\theta_m}}$ is its diagonal covariance matrix with square root of the diagonal terms being the standard deviations. The weights, means, and standard deviations of the GMM are predicted using NNs, together referred to as the GMM net. These NNs share a common backend network that learns shared features. The architectures for the feature net, weight net, mean net and standard deviation net are provided in Tables B.3 to B.5. The 'Linear mapping' in Table B.5 refers to the process of mapping the output to a specific range that is problem dependent. This mapping ensures that the predicted means and standard deviations fall within the desired range. Additionally, a 'nugget' of 10^{-27} is added to Eqn. (B.1) to prevent numerical underflow. When certain PoIs have compact support, independent truncated normal distributions [23] are used for the dimensions corresponding to those PoIs. The specific ranges of the linear mapping and the usage of truncated normal will be stated in each numerical case. The same GMM net architecture is used across all numerical cases.

B.2.2.2 Normalizing flows

To incorporate the i_t-dependence into the approximate posteriors $q(\theta_m|\mathsf{i}_t;\phi_{\Theta_m})$, the s and t functions are designed to additionally take i_t as input, leading to a form of conditional INNs (cINNs) [114]. Similar to the GMM setup, i_t is first fed into a feature network whose output has the same dimension as i_t. The architectures of the feature network, and the s_1, t_1, s_2, t_2 networks in NFs are provided in Tables B.6 to B.8.

B.2.3 Neural network architectures for actor and critic

The architectures for the actor (policy) and critic (action-value function) networks from [134] are adopted in this work. The actor $\mu_{t,w}$ is a mapping from i_t to design ξ_t. Instead of learning separate actor networks for each stage t, we combine them into a single actor. The overall input to the actor

network takes the form

$$\mathsf{i}_t^{\text{actor}} = [e_t, \tilde{\mathsf{i}}_t],$$

where e_t is an 0-indexed one-hot encoding vector of size T that represents the current experiment stage:

$$e_t = [0, \ldots, 0, \underbrace{1}_{t\text{th}}, 0, \ldots, 0],$$

and $\tilde{\mathsf{i}}_t$ is a vector of fixed size $(T-1)(n_\xi + n_y)$ obtained by extending i_t with zero-padding:

$$\tilde{\mathsf{i}}_t = [\overbrace{\xi_0}^{n_\xi}, \ldots, \xi_{t-1}, \underbrace{0, \ldots, 0}_{n_\xi(T-1-t)}, \overbrace{y_0}^{n_y}, \ldots, y_{t-1}, \underbrace{0, \ldots, 0}_{n_y(T-1-t)}].$$

The total dimension of $\mathsf{i}_t^{\text{actor}}$ is $T + (T-1)(n_\xi + n_y)$. Similarly, the overall input to the critic network is

$$\mathsf{i}_t^{\text{critic}} = [\mathsf{i}_t^{\text{actor}}, \xi_t],$$

with total dimension $T + (T-1)(n_\xi + n_y) + n_\xi$. The output of the critic network is a scalar. We note that without the presence of e_t, it would not be possible for the actor or the critic to distinguish between whether the state is at stage t, or at a later stage but with ξ_t and y_t actually being zero (i.e., whether zero values are padding or actual results). The architectures for the actor and critic networks are presented in Tables B.9 and B.10, respectively, with 'Linear mapping' in Table B.9 indicating the mapping of the output value to be within the design bounds. The same actor and critic architectures are used across all numerical cases.

B.2.4 More about the critic

While we introduced the critic $\tilde{Q}_t^{\pi_w}(\mathsf{i}_t, \xi_t; \phi)$ in Eqn. (3.25) to (3.27) for the variational one-point reward terms $\tilde{r}_t$ and $\tilde{r}_T$ from Sec. 3.1.5, a corresponding critic can be formed for the one-point

reward terms $\check{r}_t$ and $\check{r}_T$ from Sec. 3.1.4:

$$\begin{aligned}
\check{Q}_t^{\pi_w}(\mathsf{i}_t, \xi_t) &= \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}|\mathsf{i}_t}\Bigg[\mathbb{E}_{Y_{t:T-1}|w,s_0,\mathsf{i}_t,\xi_t,M_0,\Theta_{m,0},\Psi_{m,0}}\Bigg[\check{r}_t(\mathsf{i}_t, \xi_t, Y_t) \\
&\qquad + \sum_{k=t+1}^{T-1} \check{r}_t(I_t, \mu_{t,w}(I_t), Y_t) + \check{r}_T(I_T)\Bigg]\Bigg] \\
&= \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}|\mathsf{i}_t}\Bigg[\mathbb{E}_{Y_k|w,s_0,\mathsf{i}_t,\xi_t,M_0,\Theta_{m,0},\Psi_{m,0}}\Bigg[\check{r}_t(\mathsf{i}_t, \xi_t, Y_t) \\
&\qquad + \check{Q}_{t+1}^{\pi_w}(I_{t+1}, \mu_{k+1,w}(I_{t+1}))\Bigg]\Bigg], \\
\check{Q}_T^{\pi_w}(\mathsf{i}_T, \cdot) &= \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}|\mathsf{i}_T}\Big[\check{r}_T(\mathsf{i}_T)\Big],
\end{aligned}$$

for $t = 0, \ldots, T-1$ and subject to $I_{t+1} = \{I_t, \xi_t, Y_t\}$.

When using the TIG formulation, these critics become

$$\begin{aligned}
\check{Q}_{T,t}^{\pi_w}(\mathsf{i}_t, \xi_t) &= \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}|\mathsf{i}_t}\Bigg[\mathbb{E}_{Y_{t:T-1}|w,s_0,\mathsf{i}_t,\xi_t,M_0,\Theta_{m,0},\Psi_{m,0}}\Bigg[\\
&\qquad \alpha_M \log \frac{P(M_0|I_T)}{P(M_0)} + \alpha_\Theta \log \frac{p(\Theta_{m,0}|I_T)}{p(\Theta_{m,0})} + \alpha_Z \log \frac{p(Z_{m,0}|I_T)}{p(Z_{m,0})}\Bigg]\Bigg] \\
\check{Q}_{T,N}^{\pi_w}(\mathsf{i}_T, \cdot) &= \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}|\mathsf{i}_T}\Bigg[\alpha_M \log \frac{P(M_0|\mathsf{i}_T)}{P(M_0)} \\
&\qquad + \alpha_\Theta \log \frac{p(\Theta_{m,0}|\mathsf{i}_T)}{p(\Theta_{m,0})} + \alpha_Z \log \frac{p(Z_{m,0}|\mathsf{i}_T)}{p(Z_{m,0})}\Bigg],
\end{aligned}$$

for $t = 0, \ldots, T-1$ and subect to $I_{t+1} = \{I_t, \xi_t, Y_t\}$, and

$$\begin{aligned}
\tilde{Q}_{T,t}^{\pi_w}(\mathsf{i}_t, \xi_t; \phi) &= \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}|\mathsf{i}_t}\Bigg[\mathbb{E}_{Y_{t:T-1}|w,s_0,\mathsf{i}_t,\xi_t,M_0,\Theta_{m,0},\Psi_{m,0}}\Bigg[\\
&\qquad \alpha_M \log \frac{q(M_0|I_T;\phi_M)}{P(M_0)} + \alpha_\Theta \log \frac{q(\Theta_{m,0}|I_T;\phi_{\Theta_m})}{p(\Theta_{m,0})} + \alpha_Z \log \frac{q(Z_{m,0}|I_T;\phi_{Z_m})}{p(Z_{m,0})}\Bigg]\Bigg], \\
\tilde{Q}_{T,T}^{\pi_w}(\mathsf{i}_T, \cdot; \phi) &= \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}|\mathsf{i}_T}\Bigg[\alpha_M \log \frac{q(M_0|\mathsf{i}_T;\phi_M)}{P(M_0)} \\
&\qquad + \alpha_\Theta \log \frac{q(\Theta_{m,0}|\mathsf{i}_T;\phi_{\Theta_m})}{p(\Theta_{m,0})} + \alpha_Z \log \frac{q(Z_{m,0}|\mathsf{i}_T;\phi_{Z_m})}{p(Z_{m,0})}\Bigg],
\end{aligned}$$

for $t = 0, \ldots, T-1$ and subject to $I_{t+1} = \{I_t, \xi_t, Y_t\}$.

Similarly, when using the IIG formulation, these critics become

$$\begin{aligned}
\check{Q}^{\pi_w}_{I,t}(\mathsf{i}_t, \xi_t) &= \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}|\mathsf{i}_t}\Bigg[\mathbb{E}_{Y_{t:T-1}|w,s_0,\mathsf{i}_t,\xi_t,M_0,\Theta_{m,0},\Psi_{m,0}}\Bigg[\\
&\qquad \alpha_M \log \frac{P(M_0|I_{t+1})}{P(M_0|\mathsf{i}_t)} + \alpha_\Theta \log \frac{p(\Theta_{m,0}|I_{t+1})}{p(\Theta_{m,0}|\mathsf{i}_t)} + \alpha_Z \log \frac{p(Z_{m,0}|I_{t+1})}{p(Z_{m,0}|\mathsf{i}_t)} \\
&\qquad + \sum_{k=t+1}^{T-1} \alpha_M \log \frac{P(M_0|I_{t+1})}{P(M_0|I_t)} + \alpha_\Theta \log \frac{p(\Theta_{m,0}|I_{t+1})}{p(\Theta_{m,0}|I_t)} + \alpha_Z \log \frac{p(Z_{m,0}|I_{t+1})}{p(Z_{m,0}|I_t)}\Bigg]\Bigg], \\
&= \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}|\mathsf{i}_t}\Bigg[\mathbb{E}_{Y_{t:T-1}|w,s_0,\mathsf{i}_t,\xi_t,M_0,\Theta_{m,0},\Psi_{m,0}}\Bigg[\\
&\qquad \alpha_M \log \frac{P(M_0|I_T)}{P(M_0|\mathsf{i}_t)} + \alpha_\Theta \log \frac{p(\Theta_{m,0}|I_T)}{p(\Theta_{m,0}|\mathsf{i}_t)} + \alpha_Z \log \frac{p(Z_{m,0}|I_T)}{p(Z_{m,0}|\mathsf{i}_t)}\Bigg]\Bigg], \\
\check{Q}^{\pi_w}_{I,N}(\mathsf{i}_t, \cdot) &= 0,
\end{aligned}$$

for $t = 0, \ldots, T-1$ and subject to $I_{t+1} = \{I_t, \xi_t, Y_t\}$, and

$$\begin{aligned}
\tilde{Q}^{\pi_w}_{I,t}(\mathsf{i}_t, \xi_t; \phi) &= \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}|\mathsf{i}_t}\Bigg[\mathbb{E}_{Y_{t:T-1}|w,s_0,\mathsf{i}_t,\xi_t,M_0,\Theta_{m,0},\Psi_{m,0}}\Bigg[\\
&\quad \alpha_M \log \frac{q(M_0|I_{t+1};\phi_M)}{q(M_0|\mathsf{i}_t;\phi_M)} + \alpha_\Theta \log \frac{q(\Theta_{m,0}|I_{t+1};\phi_{\Theta_m})}{q(\Theta_{m,0}|\mathsf{i}_t;\phi_{\Theta_{m,0}})} + \alpha_Z \log \frac{q(Z_{m,0}|I_{t+1};\phi_{Z_m})}{q(Z_{m,0}|\mathsf{i}_t;\phi_{Z_m})} \\
&\quad + \sum_{k=t+1}^{T-1} \alpha_M \log \frac{q(M_0|I_{t+1};\phi_M)}{q(M_0|I_t;\phi_M)} + \alpha_\Theta \log \frac{q(\Theta_{m,0}|I_{t+1};\phi_{\Theta_m})}{q(\Theta_{m,0}|I_t;\phi_{\Theta_m})} + \alpha_Z \log \frac{q(Z_{m,0}|I_{t+1};\phi_{Z_m})}{q(Z_{m,0}|I_t;\phi_{Z_m})}\Bigg]\Bigg], \\
&= \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}|\mathsf{i}_t}\Bigg[\mathbb{E}_{Y_{t:T-1}|w,s_0,\mathsf{i}_t,\xi_t,M_0,\Theta_{m,0},\Psi_{m,0}}\Bigg[\\
&\quad \alpha_M \log \frac{q(M_0|I_T;\phi_M)}{q(M_0|\mathsf{i}_t;\phi_M)} + \alpha_\Theta \log \frac{q(\Theta_{m,0}|I_T;\phi_{\Theta_m})}{q(\Theta_{m,0}|\mathsf{i}_t;\phi_{\Theta_{m,0}})} + \alpha_Z \log \frac{q(Z_{m,0}|I_T;\phi_{Z_m})}{q(Z_{m,0}|\mathsf{i}_t;\phi_{Z_m})}\Bigg]\Bigg], \\
\tilde{Q}^{\pi_w}_{I,N}(\mathsf{i}_t, \cdot; \phi) &= 0,
\end{aligned}$$

for $t = 0, \ldots, T-1$ and subject to $I_{t+1} = \{I_t, \xi_t, Y_t\}$.

Remark B.2.1. The differences between the one-point-TIG and one-point-IIG critics are

$$\begin{aligned}
\check{Q}^{\pi_w}_{T,t}(\mathsf{i}_t, \xi_t) - \check{Q}^{\pi_w}_{I,t}(\mathsf{i}_t, \xi_t) &= \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}|\mathsf{i}_t}\Bigg[\mathbb{E}_{Y_{t:T-1}|w,s_0,\mathsf{i}_t,\xi_t,M_0,\Theta_{m,0},\Psi_{m,0}}\Bigg[\\
&\qquad \alpha_M \log\frac{P(M_0|\mathsf{i}_t)}{P(M_0)} + \alpha_\Theta \log\frac{p(\Theta_{m,0}|\mathsf{i}_t)}{p(\Theta_{m,0})} + \alpha_Z \log\frac{p(Z_{m,0}|\mathsf{i}_t)}{p(Z_{m,0})}\Bigg]\Bigg] \\
&= \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}|\mathsf{i}_t}\Bigg[\\
&\qquad \alpha_M \log\frac{P(M_0|\mathsf{i}_t)}{P(M_0)} + \alpha_\Theta \log\frac{p(\Theta_{m,0}|\mathsf{i}_t)}{p(\Theta_{m,0})} + \alpha_Z \log\frac{p(Z_{m,0}|\mathsf{i}_t)}{p(Z_{m,0})}\Bigg], \\
\check{Q}^{\pi_w}_{T,N}(\mathsf{i}_T, \cdot) - \check{Q}^{\pi_w}_{I,N}(\mathsf{i}_T, \cdot) &= \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}|\mathsf{i}_T}\Bigg[\\
&\qquad \alpha_M \log\frac{P(M_0|\mathsf{i}_T)}{P(M_0)} + \alpha_\Theta \log\frac{p(\Theta_{m,0}|\mathsf{i}_T)}{p(\Theta_{m,0})} + \alpha_Z \log\frac{p(Z_{m,0}|\mathsf{i}_T)}{p(Z_{m,0})}\Bigg],
\end{aligned}$$

for $t = 0, \ldots, T-1$ and subject to $I_{t+1} = \{I_t, \xi_t, Y_t\}$. Notably, all difference expressions are independent of ξ_t. When using the one-point reward formulations, the same policy gradient expression as Eqn. (3.24) will emerge but with $\nabla_{\xi_t}\check{Q}^{\pi_w}_t(\mathsf{i}_t, \xi_t)$. Since the difference of the above critics do not depend on ξ_t, $\nabla_{\xi_t}\check{Q}^{\pi_w}_t(\mathsf{i}_t, \xi_t)$ will be identical for one-point-TIG and one-point-IIG, and hence their policy gradients will also be identical.

Remark B.2.2. The differences between the one-point-TIG and variational one-point-TIG critics

are

$$
\begin{aligned}
&\check{Q}^{\pi_w}_{T,t}(\mathsf{i}_t, \xi_t) - \tilde{Q}^{\pi_w}_{T,t}(\mathsf{i}_t, \xi_t; \phi) \\
&= \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}|\mathsf{i}_t}\Bigg[\mathbb{E}_{Y_{t:T-1}|w,s_0,\mathsf{i}_t,\xi_t,M_0,\Theta_{m,0},\Psi_{m,0}}\Bigg[\\
&\qquad \alpha_M \log \frac{P(M_0|I_T)}{q(M_0|I_T;\phi_M)} + \alpha_\Theta \log \frac{p(\Theta_{m,0}|I_T)}{q(\Theta_{m,0}|I_T;\phi_{\Theta_m})} + \alpha_Z \log \frac{p(Z_{m,0}|I_T)}{q(Z_{m,0}|I_T;\phi_{Z_m})}\Bigg]\Bigg] \\
&= \mathbb{E}_{I_T|w,s_0,\mathsf{i}_t,\xi_t}\Bigg[\alpha_M D_{\mathrm{KL}}\left(P_{M_0|I_T} \,||\, q_{M_0|I_T;\phi_M}\right) \\
&\qquad + \mathbb{E}_{M_0|I_T}\Bigg[\alpha_\Theta D_{\mathrm{KL}}\left(p_{\Theta_{m,0}|I_T} \,||\, q_{\Theta_{m,0}|I_T;\phi_{\Theta_m}}\right) + \alpha_Z D_{\mathrm{KL}}\left(p_{Z_{m,0}|I_T} \,||\, q_{Z_{m,0}|I_T;\phi_{Z_m}}\right)\Bigg]\Bigg], \\
&\check{Q}^{\pi_w}_{T,N}(\mathsf{i}_T, \cdot) - \tilde{Q}^{\pi_w}_{T,N}(\mathsf{i}_T, \cdot; \phi) \\
&= \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}|\mathsf{i}_T}\Bigg[\\
&\qquad \alpha_M \log \frac{P(M_0|\mathsf{i}_T)}{q(M_0|\mathsf{i}_T;\phi_M)} + \alpha_\Theta \log \frac{p(\Theta_{m,0}|\mathsf{i}_T)}{q(\Theta_{m,0}|\mathsf{i}_T;\phi_{\Theta_m})} + \alpha_Z \log \frac{p(Z_{m,0}|\mathsf{i}_T)}{q(Z_{m,0}|\mathsf{i}_T;\phi_{Z_m})}\Bigg] \\
&= \alpha_M D_{\mathrm{KL}}\left(P_{M_0|\mathsf{i}_T} \,||\, q_{M_0|\mathsf{i}_T;\phi_M}\right) \\
&\qquad + \mathbb{E}_{M_0|\mathsf{i}_T}\Bigg[\alpha_\Theta D_{\mathrm{KL}}\left(p_{\Theta_{m,0}|\mathsf{i}_T} \,||\, q_{\Theta_{m,0}|\mathsf{i}_T;\phi_{\Theta_m}}\right) + \alpha_Z D_{\mathrm{KL}}\left(p_{Z_{m,0}|\mathsf{i}_T} \,||\, q_{Z_{m,0}|\mathsf{i}_T;\phi_{Z_m}}\right)\Bigg],
\end{aligned}
$$

for $t = 0, \ldots, T-1$ and subject to $I_{t+1} = \{I_t, \xi_t, Y_t\}$. Notably, all difference expressions are expectations of weighted sum of KL divergence terms with non-negative weights. The difference expressions are zero if and only if the variational posterior approximations $q(\cdot|I_T;\phi_{(\cdot)})$ are equal to the true posteriors $p(\cdot|I_T)$ (except the trivial case when $\alpha_M = \alpha_\Theta = \alpha_Z = 0$). Hence, $\tilde{Q}^{\pi_w}_{T,t}(\mathsf{i}_t, \xi_t; \phi)$ forms a lower bound to $\check{Q}^{\pi_w}_{T,t}(\mathsf{i}_t, \xi_t)$ for all t, and learning an accurate variational posterior approximation would help reduce the error in the critic.

Remark B.2.3. The differences between the one-point-IIG and variational one-point-IIG critics are

$$
\begin{aligned}
&\check{Q}^{\pi_w}_{I,t}(\mathsf{i}_t, \xi_t) - \tilde{Q}^{\pi_w}_{I,t}(\mathsf{i}_t, \xi_t; \phi) \\
&= \mathbb{E}_{M_0,\Theta_{m,0},\Psi_{m,0},Z_{m,0}|\mathsf{i}_t}\Bigg[\mathbb{E}_{Y_{t:T-1}|w,s_0,\mathsf{i}_t,\xi_t,M_0,\Theta_{m,0},\Psi_{m,0}}\Bigg[\\
&\qquad \alpha_M \log\frac{P(M_0|I_T)}{q(M_0|I_T;\phi_M)} + \alpha_\Theta \log\frac{p(\Theta_{m,0}|I_T)}{q(\Theta_{m,0}|I_T;\phi_{\Theta_m})} + \alpha_Z \log\frac{p(Z_{m,0}|I_T)}{q(Z_{m,0}|I_T;\phi_{Z_m})} \\
&\qquad - \alpha_M \log\frac{P(M_0|\mathsf{i}_t)}{q(M_0|\mathsf{i}_t;\phi_M)} - \alpha_\Theta \log\frac{p(\Theta_{m,0}|\mathsf{i}_t)}{q(\Theta_{m,0}|\mathsf{i}_t;\phi_{\Theta_m})} - \alpha_Z \log\frac{p(Z_{m,0}|\mathsf{i}_t)}{q(Z_{m,0}|\mathsf{i}_t;\phi_{Z_m})}\Bigg]\Bigg] \\
&= \mathbb{E}_{I_T|w,s_0,\mathsf{i}_t,\xi_t}\Bigg[\alpha_M D_{\mathrm{KL}}\left(P_{M_0|I_T} \,||\, q_{M_0|I_T;\phi_M}\right) \\
&\qquad + \mathbb{E}_{M_0|I_T}\Bigg[\alpha_\Theta D_{\mathrm{KL}}\left(p_{\Theta_{m,0}|I_T} \,||\, q_{\Theta_{m,0}|I_T;\phi_{\Theta_m}}\right) + \alpha_Z D_{\mathrm{KL}}\left(p_{Z_{m,0}|I_T} \,||\, q_{Z_{m,0}|I_T;\phi_{Z_m}}\right)\Bigg] \\
&\qquad\quad - \alpha_M D_{\mathrm{KL}}\left(P_{M_0|\mathsf{i}_t} \,||\, q_{M_0|\mathsf{i}_t;\phi_M}\right) \\
&\qquad - \mathbb{E}_{M_0|\mathsf{i}_t}\Bigg[\alpha_\Theta D_{\mathrm{KL}}\left(p_{\Theta_{m,0}|\mathsf{i}_t} \,||\, q_{\Theta_{m,0}|\mathsf{i}_t;\phi_{\Theta_m}}\right) + \alpha_Z D_{\mathrm{KL}}\left(p_{Z_{m,0}|\mathsf{i}_t} \,||\, q_{Z_{m,0}|\mathsf{i}_t;\phi_{Z_m}}\right)\Bigg], \\
&\check{Q}^{\pi_w}_{I,N}(\mathsf{i}_T, \cdot) - \tilde{Q}^{\pi_w}_{I,N}(\mathsf{i}_T, \cdot; \phi) = 0,
\end{aligned}
$$

for $t = 0, \ldots, T-1$ and subject to $I_{t+1} = \{I_t, \xi_t, Y_t\}$. Applying triangle inequality, these difference expressions become bounded by

$$
\begin{aligned}
&|\check{Q}^{\pi_w}_{I,t}(\mathsf{i}_t, \xi_t) - \tilde{Q}^{\pi_w}_{I,t}(\mathsf{i}_t, \xi_t; \phi)| \\
&\leq \mathbb{E}_{I_T|w,s_0,\mathsf{i}_t,\xi_t}\Bigg[\alpha_M D_{\mathrm{KL}}\left(P_{M_0|I_T} \,||\, q_{M_0|I_T;\phi_M}\right) \\
&\qquad + \mathbb{E}_{M_0|I_T}\Bigg[\alpha_\Theta D_{\mathrm{KL}}\left(p_{\Theta_{m,0}|I_T} \,||\, q_{\Theta_{m,0}|I_T;\phi_{\Theta_m}}\right) + \alpha_Z D_{\mathrm{KL}}\left(p_{Z_{m,0}|I_T} \,||\, q_{Z_{m,0}|I_T;\phi_{Z_m}}\right)\Bigg]\Bigg] \\
&\qquad\quad + \alpha_M D_{\mathrm{KL}}\left(P_{M_0|\mathsf{i}_t} \,||\, q_{M_0|\mathsf{i}_t;\phi_M}\right) \\
&\qquad + \mathbb{E}_{M_0|\mathsf{i}_t}\Bigg[\alpha_\Theta D_{\mathrm{KL}}\left(p_{\Theta_{m,0}|\mathsf{i}_t} \,||\, q_{\Theta_{m,0}|\mathsf{i}_t;\phi_{\Theta_m}}\right) + \alpha_Z D_{\mathrm{KL}}\left(p_{Z_{m,0}|\mathsf{i}_t} \,||\, q_{Z_{m,0}|\mathsf{i}_t;\phi_{Z_m}}\right)\Bigg], \\
&|\check{Q}^{\pi_w}_{I,N}(\mathsf{i}_T, \cdot) - \tilde{Q}^{\pi_w}_{I,N}(\mathsf{i}_T, \cdot; \phi)| = 0,
\end{aligned}
$$

for $t = 0, \ldots, T-1$ and subject to $I_{t+1} = \{I_t, \xi_t, Y_t\}$. Therefore, the error of the critic is contributed from both the error in the final variational posterior $q(\cdot|I_T; \phi_{(\cdot)})$ and the errors in the intermediate variational posteriors $q(\cdot|I_t; \phi_{(\cdot)})$. The critic errors become zero if all the variational posteriors

equal their corresponding true posteriors (except the trivial case when $\alpha_M = \alpha_\Theta = \alpha_Z = 0$).

B.2.4.1 Hyperparameter tuning

Our main strategy for hyperparameter tuning is to start with a relatively large hyperparameter value and gradually decrease it.

For optimizing the GMM and NFs approximate posteriors for model indicator and PoIs, we start with the initial learning rate of 10^{-3} and an exponential learning rate decay rate 0.9999. For optimizing the critic network, we use an initial learning rate of 10^{-3} and a learning rate decay rate of 0.9999 across all numerical cases. Both posterior approximation and critic network optimizations are updated 5 steps (i.e., applying gradient ascent 5 times) within each outer iteration. Making too many update steps within each outer iteration may result in overestimation of the value function and adversely affect the policy search [68].

For optimizing the actor network, we use a learning rate decay rate of 0.9999. However, the choice of the initial learning rate is more problem-dependent. Typically, we start with an initial learning rate of 10^{-3} and gradually decrease it to 5×10^{-4} or 2×10^{-4} if divergence occurs. For the IIG formulation, an initial learning rate of 10^{-3} works well. For TIG, a smaller learning rate is generally required. This is likely due to the slower propagation of critic values from T to the earlier t stages under TIG, and a large learning rate may induce divergence in the early iterations of training.

For other hyperparameters including the number of updates n_{update}, number of sample trajectories n_{traj}, batch size n_{batch}, and replay buffer size n_{buffer}, a number of combinations are tested to identify the optimal setting. Their values are specified for each numerical case in Sec. B.3.

B.3 Numerical experiment details

B.3.1 Prior contrastive estimator

The PCE for batch (non-sequential) OED is introduced in [51]:

$$U^{\text{PCE}}(\xi) = \mathbb{E}_{Y_0|\Theta_0,\xi}\mathbb{E}_{\Theta_0}\mathbb{E}_{\Theta_{1:L}}\left[\log \frac{p(Y_0|\Theta_0,\xi)}{\frac{1}{L+1}\sum_{j=1}^{L+1} p(Y_0|\Theta_j,\xi)}\right],$$

where the expectation is over $\Theta_0, Y_0|\xi \sim p(\Theta, Y|\xi)$ and $\Theta_{1:L} \overset{\text{iid}}{\sim} p(\Theta)$, and the subscripts here in the non-sequential setting represent multi-sample indexing. One can show that $U^{\text{PCE}}(\xi)$ is a lower bound to the mutual information between Y and Θ, $\mathcal{I}(Y;\Theta|\xi)$, (i.e., the EIG of Θ)—that is, $U^{\text{PCE}}(\xi) \leq \mathcal{I}(Y;\Theta|\xi)$—for any $L > 0$, and the bound becomes tight as $L \to \infty$ [51, Theorem 1].

However, the expectation in U^{PCE} is generally intractable to evaluate, and U^{PCE} needs to be estimated numerically, for example through MC:

$$U^{\text{PCE}}(\xi) \approx \frac{1}{n_{\text{out}}} \sum_{i=1}^{n_{\text{out}}} \log \frac{p(y_0^{(i)}|\theta_0^{(i)}, \xi)}{\frac{1}{L+1}\sum_{j=1}^{L+1} p(y_0^{(i)}|\theta_j^{(i)}, \xi)},$$

where $(y_0^{(i)}, \theta_0^{(i)}, \theta_{1:L}^{(i)})$ are n_{out} independent and identically distributed (i.i.d.) realizations of $(Y_0, \Theta_0, \Theta_{1:L})$.

For sequential OED, the PCE becomes [49]:

$$U^{\text{sPCE}}(\pi) = \mathbb{E}_{Y_{0:T-1}|\pi,\Theta_0}\mathbb{E}_{\Theta_0}\mathbb{E}_{\Theta_{1:L}}\left[\log \frac{p(Y_{0:T-1}|\Theta_0, \xi_{0:T-1})}{\frac{1}{L+1}\sum_{j=1}^{L+1} p(Y_{0:T-1}|\Theta_j, \xi_{0:T-1})}\right]$$

where Θ_0 is the data-generating parameter for $\xi_{0:T-1}, Y_{0:T-1}$ with $\xi_k = \mu_t(I_t)$ following the given policy π; the subscripts for Y and ξ refer to the experiment (stage) index, while those for Θ refer to the multi-sample index. One can show that $U^{\text{sPCE}}(\pi) \leq \mathcal{I}(Y_{0:T-1}; \Theta|\pi)$ for any π and $L > 0$, and the bound becomes tight as $L \to \infty$ [49, Theorem 1]. Similarly, U^{sPCE} can be estimated using MC:

$$U^{\text{sPCE}}(\pi) \approx \frac{1}{n_{\text{out}}} \sum_{i=1}^{n_{\text{out}}} \log \frac{p(y_{0:T-1}^{(i)}|\theta_0^{(i)}, \xi_{0:T-1})}{\frac{1}{L+1}\sum_{j=1}^{L+1} p(y_{0:T-1}^{(i)}|\theta_j^{(i)}, \xi_{0:T-1})},$$

where $(y_{0:T-1}^{(i)}, \theta_0^{(i)}, \theta_{1:L}^{(i)})$ are n_{out} i.i.d. realizations of $(Y_{0:T-1}, \Theta_0, \Theta_{1:L})$. We use this estimator for evaluating policies from 'OED for PoIs' in this paper when applicable.

From the expressions above, we can see that PCE cannot be used when nuisance parameters Ψ are present since the term $p(y_{0:T-1}|\theta_0, \xi_{0:T-1})$ would be intractable as it needs to marginalize out Ψ. Similarly, PCE cannot be used for 'OED for QoIs' since in that scenario, the role of Θ is replaced by Z and $p(y_{0:T-1}|\theta_0, \xi_{0:T-1})$ becomes $p(y_{0:T-1}|z_0, \xi_{0:T-1})$, which is intractable to evaluate. Lastly, PCE cannot be used for implicit likelihood cases since $p(y_{0:T-1}|\theta_0, \xi_{0:T-1})$ would not be accessible.

We can write an analogous PCE expression for model indicator by replacing Θ with M and use the exact expression for the 'marginal-likelihood' term in the denominator:

$$\begin{aligned} U^{\text{sPCE},M}(\pi) &= \mathbb{E}_{Y_{0:T-1}|\pi,M_0}\mathbb{E}_{M_0}\left[\log \frac{p(Y_{0:T-1}|M_0, \xi_{0:T-1})}{\frac{1}{|\mathcal{M}_m|}\sum_{j=1}^{|\mathcal{M}_m|} P(m_j)p(Y_{0:T-1}|m_j, \xi_{0:T-1})}\right] \\ &\approx \frac{1}{n_{\text{out}}} \sum_{i=1}^{n_{\text{out}}} \log \frac{p(y_{0:T-1}^{(i)}|m_0^{(i)}, \xi_{0:T-1})}{\frac{1}{|\mathcal{M}_m|}\sum_{j=1}^{|\mathcal{M}_m|} P(m_j)p(y_{0:T-1}^{(i)}|m_j, \xi_{0:T-1})}, \end{aligned}$$

where M_0 is the data-generating parameter for $\xi_{0:T-1}, Y_{0:T-1}$ with $\xi_k = \mu_t(I_t)$ following the given

policy π, and $(y_{0:T-1}^{(i)}, m_0^{(i)})$ are n_{out} i.i.d. realizations of $(Y_{0:T-1}, M_0)$. The 'likelihood' term can be estimated, for example, via

$$\begin{aligned} p(y_{0:T-1}|m, \xi_{0:T-1}) &= \iint p(\theta_m, \psi_m|m)\, p(y_{0:T-1}|m, \theta_m, \psi_m, \xi_{0:T-1})\, \mathrm{d}\theta_m\, \mathrm{d}\psi_m \\ &\approx \frac{1}{L} \sum_{l=1}^{L} p(y_{0:T-1}|m, \theta_m^{(l)}, \psi_m^{(l)}, \xi_{0:T-1}), \end{aligned}$$

where $\theta_m^{(l)}, \psi_m^{(l)} \sim p(\theta_m, \psi_m|m)$. We use this estimator for evaluating policies from 'OED for model indicator' in this paper.

B.3.2 Case 1: source location finding

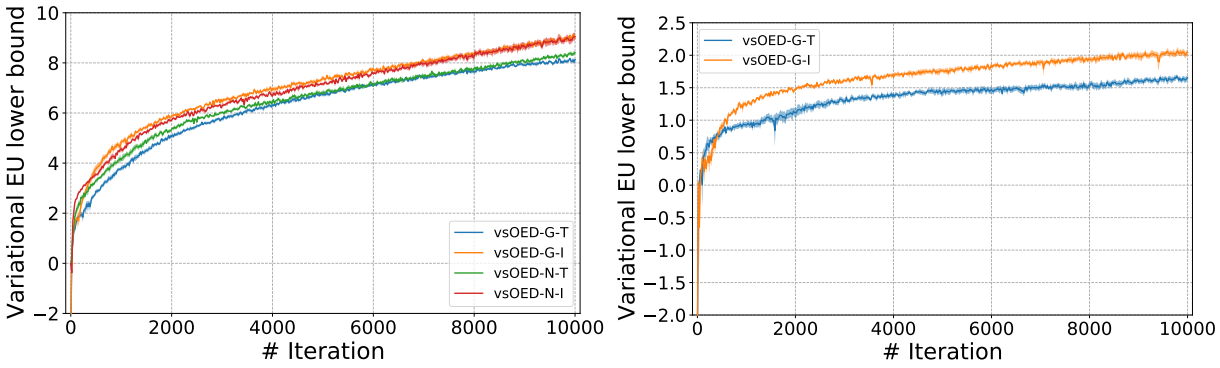
Hyperparameters Tables B.11 and B.13 present the hyperparameter settings for Case 1a and 1b, respectively. In both cases, for the 'Linear mapping' in the GMM net, we map the PoI GMM mean to $[-6, 6]$, PoI GMM standard deviation to $[10^{-5}, 1]$, QoI GMM mean to $[-6, 6]$, and QoI GMM standard deviation to $[10^{-5}, 2]$. The truncated normal distribution is not used. Table B.12 presents PCE estimates of the EIG using policies for $T = 30$ and vsOED-G-T with varying number of critic update. The results indicate that both too few or too many critic updates can lead to deteriorated performance.

Training stability For Case 1b, we only illustrate training stability results for 'OED for PoIs' for brevity. Figures B.1 and B.2 present the training history of average $\tilde{U}$ over four training replicates for Case 1a and 1b, respectively, for $T = 30$. The shaded regions represent the standard error. Tables B.14 and B.15 present the expected utility and standard error ($\pm$) for four training replicates for 'OED for PoIs', evaluated using PCE, for $T = 30$. The last columns display the average of the four replicates along with its standard error. The results indicate a good level of training robustness.

B.3.3 Case 2: constant elasticity of substitution

Hyperparameters Table B.16 presents the hyperparameter settings for this case. For the 'Linear mapping' in the GMM net, we map the PoI GMM mean to $[-1, 2]$ for ρ and β and to $[-17, 19]$ for $\log u$, and PoI GMM standard deviation to $[10^{-5}, 3]$ for all variables. The truncated normal distribution is used on ρ and β with support $[0, 1]$.

Training stability Figure B.3 presents the training history of average $\tilde{U}$ over four training replicates for $T = 10$, with the shaded regions representing the standard error. Table B.17 presents the expected utility and standard error ($\pm$) for four training replicates, evaluated using PCE, for $T = 10$.



(a) OED for PoIs

(b) OED for QoIs

Figure B.1: Case 1a. Training history of average $\tilde{U}$ over four training replicates for $T = 30$. The shaded regions represent standard error.

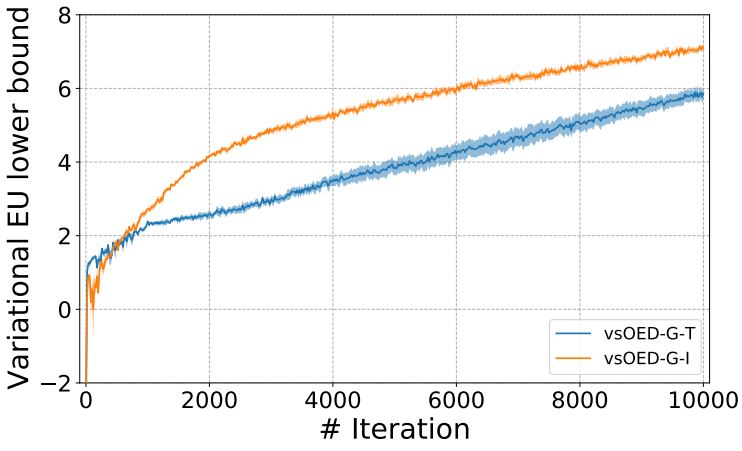


Figure B.2: Case 1b. Training history of average $\tilde{U}$ over four training replicates for 'OED for PoIs' and $T = 30$. The shaded regions represent standard error.

The last column displays the average of the four replicates along with its standard error. The results exhibits slightly more variation compared to Case 1, but overall remains robust.

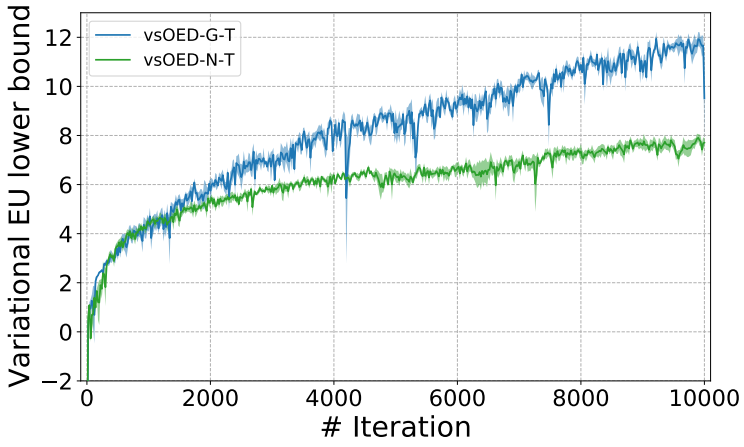


Figure B.3: Case 2. Training history of average $\tilde{U}$ over four training replicates for $T = 10$. The shaded regions represent the standard error.

B.3.4 Case 3: SIR model for disease spread

Hyperparameters Table B.18 presents the hyperparameter settings for this case. For the 'Linear mapping' in the GMM net, we map the PoI GMM mean to $[-6, 4]$, and PoI GMM standard deviation to $[10^{-5}, 0.5]$. The truncated normal distribution is not used.

Training stability Figure B.4 presents the training history of average $\tilde{U}$ over four training replicates for $T = 10$, with the shaded regions representing the standard error. Table B.19 presents the expected utility and standard error ($\pm$) for four training replicates, evaluated using PCE, for $T = 10$. The last column displays the average of the four replicates along with its standard error.

The training appears highly stable in this case with consistent performance across different random seeds, except for a dip in the training history for vsOED-G-T.

B.3.5 Case 4: convection-diffusion-reaction

Hyperparameters Table B.20 presents the hyperparameter settings for this case. For the 'Linear mapping' in the GMM net, we map the PoI GMM mean to $[-1, 2]$, PoI GMM standard deviation to $[10^{-5}, 1]$, QoI GMM mean to $[-15, 3]$, and QoI standard deviation net to $[10^{-5}, 4]$. The truncated normal distribution is used on all PoIs with support $[0, 1]$.

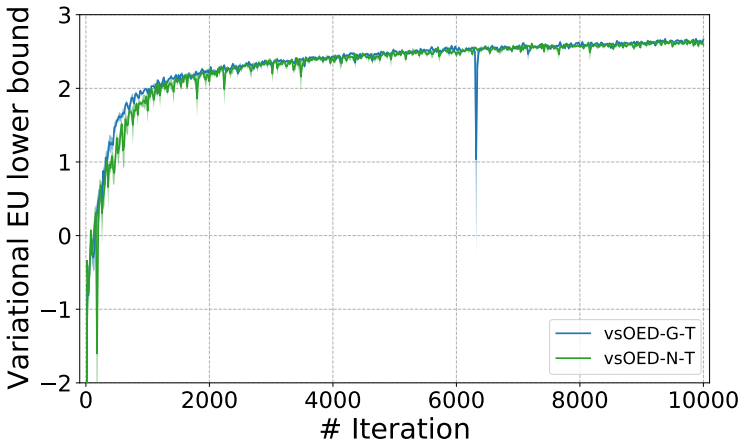


Figure B.4: Case 3. Training history of average $\tilde{U}$ over four training replicates for $T = 10$. The shaded regions represent the standard error.

Training stability We only illustrate training stability results for 'OED for PoIs' for brevity. Figure B.5 presents the training history of average $\tilde{U}$ (but with the prior term omitted due to the presence of nuisance parameters, per Sec. B.1.5) over four training replicates, with the shaded regions representing standard error. Table B.21 presents the prior-omitted $\tilde{U}$ and standard error ($\pm$) for four training replicates for $T = 10$. The last column displays the average of the four replicates along with its standard error. These results indicate vsOED to have excellent training robustness.

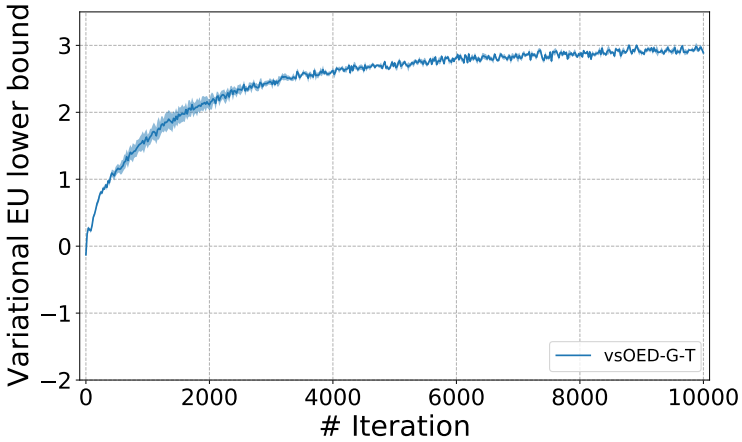


Figure B.5: Case 4. Training history of average $\tilde{U}$ over four training replicates for $T = 10$. The shaded regions represent the standard error.

Neural network architectures for surrogate models The architecture for the NN-based surrogate models of G and φ are provided in Tables B.22 and B.23.

B.3.6 Summary of case-specific symbols and notations

Please see Table B.24 for the case-specific symbols and notation used in the numerical experiments.

	Description
$\xi_t \in \mathcal{A}_t \subseteq \mathbb{R}^{n_\xi}$	Design vector at step t
$Y_t(y_t) \in \mathcal{Y}_t \subseteq \mathbb{R}^{n_y}$	Observation vector at step t
$I_t(\mathsf{i}_t)$	Information set up to step t: $\{\xi_0, Y_0, \ldots, \xi_{t-1}, Y_{t-1}\}$
$m \in \mathcal{M}$	Model index
$\Theta_m(\theta_m) \in \mathbf{\Theta}_m \subseteq \mathbb{R}^{N_{\theta_m}}$	Parameters of interest (PoIs)
$\Psi_m(\psi_m) \in \mathbf{\Psi}_m \subseteq \mathbb{R}^{N_{\psi_m}}$	Nuisance parameters
$Z_m(z_m) \in \mathcal{Z}_m \subseteq \mathbb{R}^{N_{z_m}}$	Predictive quantities (QoIs)
G_t	Forward operator for observation
H	Forward operator for prediction
$\mathcal{E}_t$	Observation noise
$\mathcal{E}_Z$	Predictive stochastic source
$\alpha_M, \alpha_\theta, \alpha_Z \in [0, 1]$	Weights of IG associated with model indicator, PoIs and QoIs
$S_t(s_t) = \{S_k^b(s_t^b), S_k^p(s_t^p)\}$	State at step t (belief state, physical state)
μ_t	Policy function s.t. $\xi_t = \mu_t(s_t)$; $\mu_t : \mathcal{S}_t \to \mathcal{A}_t$
π	Policy: $\{\mu_t\}_{t=0}^{T-1}$
π_w	Policy parametrized by w
$\mathcal{F}_t$	A transition function that maps the current state s_t, design ξ_t, and observed data y_t to the next state s_{t+1}, encapsulating the update of belief
$r_t(s_t, \xi_t, y_t) \in \mathbb{R}$	Reward at $k-th$ experiment, $t = 0, \ldots, T-1$
$r_T(s_T) \in \mathbb{R}$	Terminal reward
$U(\pi)$	Expected utility for policy π
$U_T(\pi)$	Expected utility for policy π with TIG reward
$U_I(\pi)$	Expected utility for policy π with IIG reward
$\check{r}_t(\mathsf{i}_t, \xi_t, y_t), \check{r}_T(s_T) \in \mathbb{R}$	One-point approximation of the reward
$\check{U}(\pi)$	One-point expected utility for policy π
$\check{U}_T(\pi)$	One-point expected utility for policy π with one TIG reward
$\check{U}_I(\pi)$	One-point expected utility for policy π with one IIG reward
$q(\cdot\|\cdot; \phi)$	Variational posterior parameterized by ϕ
$\tilde{r}_t(\mathsf{i}_t, \xi_t, y_t), \tilde{r}_T(s_T) \in \mathbb{R}$	Variational one-point reward
$\tilde{U}(\pi)$	Expected utility for policy π with variational posterior approximation
$\tilde{U}_T(\pi)$	Expected utility for policy π with variational one-point TIG
$\tilde{U}_I(\pi)$	Expected utility for policy π with variational one-point IIG
$\widetilde{Q}_t^{\pi_w}$	Critic (action-value) function: expected remaining cumulative reward under variational parameters ϕ, after taking design ξ_t from state i_t, and subsequently following policy π_w, the expectation is taken over the posterior predictive distribution conditioned on i_t
$\tilde{Q}_{k,\nu}^{\pi_w}$	Critic network parameterized by ν
$\mathcal{L}_\phi(\nu)$	Critic training loss
$\gamma \in [0, 1]$	Discount factor (for regularization)
$\eta \in [0, 1]$	Interpolation weight in critic loss for TIG case

Table B.1: Summary of general symbols and notation used in the paper.

Table B.2: Architecture for the NN-based approximate posteriors of model indicator.

Layer	Description	Dimension	Activation
Input	i_t	$t(n_\xi + n_y)$	-
Hidden 1	Dense	256	ReLU
Hidden 2	Dense	256	ReLU
Hidden 3	Dense	256	ReLU
Output	Dense	$\lvert\mathcal{M}_m\rvert$	LogSoftmax

Table B.3: Architecture for the feature net within the GMM net.

Layer	Description	Dimension	Activation
Input	i_t	$t(n_\xi + n_y)$	-
Hidden 1	Dense	256	ReLU
Output	Dense	256	ReLU

Table B.4: Architecture for the weight net within the GMM net.

Layer	Description	Dimension	Activation
Input	Feature(i_t)	256	-
Hidden 1	Dense	256	ReLU
Hidden 2	Dense	256	ReLU
Output	Dense	n_{mix}	Softmax

Table B.5: Architecture for the mean net and standard deviation net within the GMM net.

Layer	Description	Dimension	Activation
Input	Feature(i_t)	256	-
Hidden 1	Dense	256	ReLU
Hidden 2	Dense	256	ReLU
Hidden 3	Dense	$n_{\text{mix}}N_{\theta_m}$	Sigmoid
Output	Identity	$n_{\text{mix}}N_{\theta_m}$	Linear mapping

Table B.6: Architecture for the feature net within the NFs. The first value under the 'Dimension' column is used for the source location problem and the CES problem; the second value is used for the SIR problem.

Layer	Description	Dimension	Activation
Input	i_t	$t(n_\xi + n_y)$	-
Hidden 1	Dense	256 / 128	ReLU
Hidden 2	Dense	256 / 128	ReLU
Hidden 3	Dense	256 / None	ReLU
Output	Dense	$t(n_\xi + n_y)$	-

Table B.7: Architecture for the s_1 and t_1 nets within the NFs. The first value under the 'Dimension' column is used for the source location problem and the CES problem; the second value is used for the SIR problem.

Layer	Description	Dimension	Activation
Input	Feature(i_t) + θ_1	$t(n_\xi + n_y) + n_{\Theta_{m,1}}$	-
Hidden 1	Dense	256 / 128	ReLU
Hidden 2	Dense	256 / 128	ReLU
Hidden 3	Dense	256 / 128	ReLU
Output	Dense	$n_{\Theta_{m,2}}$	-

Table B.8: Architecture for the s_2 and t_2 nets within the NFs. The first value under the 'Dimension' column is used for the source location problem and the CES problem; the second value is used for the SIR problem.

Layer	Description	Dimension	Activation
Input	Feature(i_t) + $\tilde{\theta}_2$	$t(n_\xi + n_y) + n_{\Theta_{m,2}}$	-
Hidden 1	Dense	256 / 128	ReLU
Hidden 2	Dense	256 / 128	ReLU
Hidden 3	Dense	256 / 128	ReLU
Output	Dense	$n_{\Theta_{m,1}}$	-

Table B.9: Architecture for the actor network.

Layer	Description	Dimension	Activation
Input	$\mathfrak{i}_t^{\text{actor}}$	$T + (T-1)(n_\xi + n_y)$	-
Hidden 1	Dense	256	ReLU
Hidden 2	Dense	256	ReLU
Hidden 3	Dense	256	ReLU
Hidden 4	Dense	n_ξ	Sigmoid
Output	Identity	n_ξ	Linear mapping

Table B.10: Architecture for the critic network.

Layer	Description	Dimension	Activation
Input	$\mathfrak{i}_t^{\text{critic}}$	$T + (T-1)(n_\xi + n_y) + n_\xi$	-
Hidden 1	Dense	256	ReLU
Hidden 2	Dense	256	ReLU
Hidden 3	Dense	256	ReLU
Output	Dense	1	-

Table B.11: Case 1a. Hyperparameter settings. In the table, 'lr' stands for 'learning rate'.

	vsOED-G-T	vsOED-G-I	vsOED-N-T	vsOED-N-I
n_{updates}	10001	10001	10001	10001
n_{traj}	1000	1000	1000	1000
n_{batch}	10000	10000	10000	10000
$a_{\phi,0}$	10^{-3}	10^{-3}	10^{-3}	10^{-3}
a_ϕ decay	0.9999	0.9999	0.9999	0.9999
# of ϕ updates per l-iteration	5	5	5	5
n_{mixture}	8	8	N/A	N/A
n_{trans}	N/A	N/A	4	4
$a_{w,0}$	5×10^{-4}	10^{-3}	10^{-3}	10^{-3}
a_w decay	0.9999	0.9999	0.9999	0.9999
Initial critic lr	10^{-3}	10^{-3}	10^{-3}	10^{-3}
Critic lr decay	0.9999	0.9999	0.9999	0.9999
# of ν updates per l-iteration	5	5	5	5
Max buffer size	10^6	10^6	10^6	10^6
γ	1	0.9	1	0.9
Initial $\sigma_{t,\text{explore}}$	0.5	0.5	0.5	0.5
$\sigma_{t,\text{explore}}$ decay	0.9999	0.9999	0.9999	0.9999
Target network lr	0.1	0.1	0.1	0.1

Table B.12: Case 1a. Case 1a. PCE estimate of the EIG for $T = 30$ using vsOED-G-T with varying number of critic update per iteration. ± represents standard error.

# of ν updates per l-iteration	2	5	10	20
PCE-evaluated EIG on PoIs	10.909 ± 0.101	11.176 ± 0.077	11.122 ± 0.107	10.224 ± 0.455

Table B.13: Case 1b. Hyperparameter settings. In the table, 'lr' stands for 'learning rate'.

	vsOED-G-T	vsOED-G-I
n_{update}	10001	10001
n_{traj}	1000	1000
n_{batch}	10000	10000
$a_{\phi,0}$	10^{-3}	10^{-3}
a_ϕ decay	0.9999	0.9999
# of ϕ updates per l-iteration	5	5
n_{mixture}	8	8
$a_{w,0}$	2×10^{-4}	10^{-3}
a_w decay	0.9999	0.9999
Initial critic lr	10^{-3}	10^{-3}
Critic lr decay	0.9999	0.9999
# of ν updates per l-iteration	5	5
Max buffer size	10^6	10^6
γ	1	0.9
Initial $\sigma_{t,\text{explore}}$	0.5	0.5
$\sigma_{t,\text{explore}}$ decay	0.9999	0.9999
Target network lr	0.1	0.1

Table B.14: Case 1a. Expected utility and standard error (±) for four training replicates for 'OED for PoIs', evaluated using PCE, for $T = 30$.

	Replicate 1	Replicate 2	Replicate 3	Replicate 4	Average
vsOED-G-T	11.26 ± 0.05	11.09 ± 0.05	11.24 ± 0.05	11.02 ± 0.04	11.15 ± 0.05
vsOED-G-I	12.50 ± 0.04	12.31 ± 0.04	12.58 ± 0.04	11.90 ± 0.04	12.32 ± 0.13
vsOED-N-T	11.24 ± 0.05	11.67 ± 0.05	11.15 ± 0.05	11.59 ± 0.05	11.42 ± 0.11
vsOED-N-I	12.39 ± 0.04	12.34 ± 0.04	12.16 ± 0.04	12.54 ± 0.04	12.36 ± 0.07

Table B.15: Case 1b. Expected utility and standard error (±) for four training replicates for 'OED for PoIs', evaluated using PCE, for $T = 30$.

	Replicate 1	Replicate 2	Replicate 3	Replicate 4	Average
vsOED-G-T	9.67 ± 0.06	9.26 ± 0.06	9.06 ± 0.06	8.83 ± 0.06	9.21 ± 0.15
vsOED-G-I	10.57 ± 0.05	10.09 ± 0.05	10.46 ± 0.05	10.43 ± 0.05	10.39 ± 0.09

Table B.16: Case 2. Hyperparameter settings. In the table, 'lr' stands for 'learning rate'.

	vsOED-G-T	vsOED-N-T
n_{update}	10001	10001
n_{traj}	1000	1000
n_{batch}	10000	10000
$a_{\phi,0}$	10^{-3}	10^{-3}
a_ϕ decay	0.9999	0.9999
# of ϕ updates per l-iteration	5	5
n_{mixture}	8	N/A
n_{trans}	N/A	4
$a_{w,0}$	10^{-3}	10^{-3}
a_w decay	0.9999	0.9999
Initial critic lr	10^{-3}	10^{-3}
Critic lr decay	0.9999	0.9999
# of ν updates per l-iteration	5	5
Max buffer size	10^6	10^6
γ	1	1
Initial $\sigma_{t,\text{explore}}$	5	5
$\sigma_{t,\text{explore}}$ decay	0.9998	0.9998
Target network lr	0.1	0.1

Table B.17: Case 2. Expected utility and standard error (±) for four training replicates, evaluated using PCE, for $T = 10$.

	Replicate 1	Replicate 2	Replicate 3	Replicate 4	Average
vsOED-G-T	11.79 ± 0.07	12.34 ± 0.06	12.29 ± 0.06	11.13 ± 0.08	11.89 ± 0.25
vsOED-N-T	8.40 ± 0.10	9.51 ± 0.09	8.91 ± 0.09	10.30 ± 0.08	9.28 ± 0.35

Table B.18: Case 3. Hyperparameter settings. In the table, 'lr' stands for 'learning rate'.

	vsOED-G-T	vsOED-N-T
n_{update}	10001	10001
n_{traj}	1000	1000
n_{batch}	10000	10000
$a_{\phi,0}$	5×10^{-4}	10^{-3}
a_ϕ decay	0.9999	0.9999
# of ϕ updates per l-iteration	5	5
n_{mixture}	8	N/A
n_{trans}	N/A	4
$a_{w,0}$	5×10^{-4}	5×10^{-4}
a_w decay	0.9999	0.9999
Initial critic lr	10^{-3}	10^{-3}
Critic lr decay	0.9999	0.9999
# of ν updates per l-iteration	5	5
Max buffer size	10^6	10^6
γ	1	1
Initial $\sigma_{t,\text{explore}}$	5	5
$\sigma_{t,\text{explore}}$ decay	0.9999	0.9999
Target network lr	0.1	0.1

Table B.19: Case 3. Expected utility and standard error (±) for four training replicates, evaluated using PCE, for $T = 10$.

	Replicate 1	Replicate 2	Replicate 3	Replicate 4	Average
vsOED-G-T	4.091 ± 0.002	4.093 ± 0.002	4.090 ± 0.001	4.092 ± 0.001	4.092 ± 0.001
vsOED-N-T	4.097 ± 0.002	4.100 ± 0.002	4.091 ± 0.002	4.106 ± 0.002	4.099 ± 0.003

Table B.20: Case 4. Hyperparameter settings. In the table, 'lr' stands for 'learning rate'.

	vsOED-G-T
n_{update}	10001
n_{traj}	1000
n_{batch}	10000
$a_{\phi,0}$	10^{-3}
a_ϕ decay	0.9999
# of ϕ updates per l-iteration	5
n_{mixture}	8
$a_{w,0}$	5×10^{-4}
a_w decay	0.9999
Initial critic lr	10^{-3}
Critic lr decay	0.9999
# of ν updates per l-iteration	5
Max buffer size	10^6
γ	1
Initial $\sigma_{t,\text{explore}}$	0.05
$\sigma_{t,\text{explore}}$ decay	0.9999
Target network lr	0.1

Table B.21: Case 4. Prior-omitted $\tilde{U}$ and standard error (±) for four training replicates and $T = 10$.

	Replicate 1	Replicate 2	Replicate 3	Replicate 4	Average
vsOED-G-T	2.998 ± 0.002	3.057 ± 0.002	2.857 ± 0.002	3.039 ± 0.002	2.99 ± 0.04

Table B.22: Case 4. Architecture for the surrogate model of G.

Layer	Description	Dimension	Activation
Input	$[\theta_m, \psi_m, x, y]$	$2m + 4$	-
Hidden 1	Dense	256	ReLU
Hidden 2	Dense	256	ReLU
Hidden 3	Dense	256	ReLU
Hidden 4	Dense	256	ReLU
Output	Dense	1	-

Table B.23: Case 4. Architecture for the surrogate model of φ.

Layer	Description	Dimension	Activation
Input	$[\theta_m, \psi_m]$	$2m + 2$	-
Hidden 1	Dense	256	ReLU
Hidden 2	Dense	256	ReLU
Hidden 3	Dense	256	ReLU
Output	Dense	1	-

	Description
Case 1:	
a_{bg}	Background signal
$a_{\max}$	Maximum signal amplitude
$\varphi(\Theta_m; m)$	Total flux across the vertical wall at $x = 6$
Case 2:	
$\Theta = [\alpha, \beta, \log u]$	Latent PoIs
$Y_t = \text{clip}(\text{sigmoid}(\varphi_t), \epsilon, 1 - \epsilon)$	Observation model, $\varphi_t \sim \mathcal{N}\left(\mu_\varphi, \sigma_\varphi^2\right)$
ϵ	Clipping threshold
$\mu_\varphi, \sigma_\varphi$	Mean and std of the score φ
τ	Controls clipping slope
Case 3:	
$X(k) = [S(k), I(k)]$	State vector: susceptible and infected individuals at k
$R(k)$	Recovered individuals at k
N_p	Fixed population size = $S(k) + I(k) + R(k)$
$\Theta = [\log \beta, \log \rho]$	Log of infection and recovery rates
$Y_t = I(k_t)$	Observed infected count
$dX(k)$	Follows Itô SDE for disease dynamics
$W(k)$	Independent Wiener processes
$f(X(k))$, $D(X(k))$	Drift and diffusion terms
Case 4:	
$\Psi_m = [v, \beta]$	Nuisance: convection speed and direction
$G(\Theta_m, \Psi_m, x, y, k; m)$	Contaminant concentration at location $[x, y]$ and time k
$S(\Theta_m, x, y, k; m)$	Source function
$u(\Psi_m)$	Convection velocity vector: $[v \cos \beta, v \sin \beta]$
$\varphi(\Theta_m, \Psi_m)$	Integrated flux at $x = 1$ over $y \in [0, 1]$

Table B.24: Summary of case-specific symbols and notation used in the numerical experiments.

CHAPTER C

Appendix for GO-CBED

C.1 Theoretical and Numerical Formulations

C.1.1 Incremental EIG Formulation

The incremental EIG on the target query z resulting from an experiment at stage t with design ξ_t, given the intervention history I_{t-1}, is defined as:

$$\mathcal{I}_t(\xi_t, I_{t-1}) := \mathbb{E}_{p(\mathcal{M}|I_{t-1})p(Y_t|\mathcal{M},\xi_t)p(Z|\mathcal{M})}\left[\log \frac{p(Z|I_t)}{p(Z|I_{t-1})}\right]. \tag{C.1}$$

Proposition C.1.1. *The total EIG of a policy π on the target query z over a sequence of T experiments can be written as:*

$$\mathcal{I}_T(\pi) = \mathbb{E}_{p(I_T|\pi)}\left[\sum_{t=1}^{T} I_t(\xi_t, I_{t-1})\right]. \tag{C.2}$$

Proof. Beginning from the right-hand side, we have:

$$
\begin{aligned}
&\mathbb{E}_{p(I_T|\pi)}\left[\sum_{t=1}^{T} \mathcal{I}_t(\xi_t, I_{t-1})\right] \\
&= \mathbb{E}_{p(I_T|\pi)}\left[\sum_{t=1}^{T} \mathbb{E}_{p(\mathcal{M}|I_{t-1})p(Y_t|\mathcal{M},\xi_t)p(Z|\mathcal{M})}\left[\log \frac{p(Z|I_t)}{p(Z|I_{t-1})}\right]\right] \\
&= \sum_{t=1}^{T}\left[\mathbb{E}_{p(I_{t-1}|\pi)p(\mathcal{M}|I_{t-1})p(Y_t|\mathcal{M},\xi_t)p(Z|\mathcal{M})}\left[\log \frac{p(Z|I_t)}{p(Z|I_{t-1})}\right]\right] \\
&= \sum_{t=1}^{T}\left[\mathbb{E}_{p(\mathcal{M},Y_t,I_{t-1},Z|\pi)}\left[\log \frac{p(Z|I_t)}{p(Z|I_{t-1})}\right]\right] \\
&= \sum_{t=1}^{T}\left[\mathbb{E}_{p(\mathcal{M},I_t|\pi)p(Z|\mathcal{M})} \log p(Z|I_t) - \mathbb{E}_{p(\mathcal{M},I_{t-1}|\pi)p(Z|\mathcal{M})} \log p(Z|I_{t-1})\right] \\
&= \mathbb{E}_{p(\mathcal{M},I_T|\pi)p(Z|\mathcal{M})}\left[\log \frac{p(Z|I_T)}{p(Z)}\right] \\
&= \mathcal{I}_T(\pi),
\end{aligned} \tag{C.3}
$$

where in the third equality, the joint expectation is formed using $p(z|\mathcal{M}) = p(z|\mathcal{M}, I_{t-1})$, and the fifth equality follows from the cancellation of all terms in the summation except the first and last.

□

C.1.2 Proof for Theorem 4.3.1

Proof for Theorem 4.3.1. Following (4.4) and (4.5), the difference

$$
\begin{aligned}
&\mathcal{I}_T(\pi) - \mathcal{I}_{T;\,L}(\pi;\phi) \\
&= \mathbb{E}_{p(\mathcal{M})p(I_T|\mathcal{M},\pi)p(z|\mathcal{M})}\left[\log \frac{p(z|I_T)}{p(z)}\right] - \mathbb{E}_{p(\mathcal{M})p(I_T|\mathcal{M},\pi)p(z|\mathcal{M})}\left[\log \frac{q(z|I_T;\phi)}{p(z)}\right] \\
&= \mathbb{E}_{p(\mathcal{M})p(I_T|\mathcal{M},\pi)p(z|\mathcal{M})}\left[\log \frac{p(z|I_T)}{q(z|I_T;\phi)}\right] \\
&= \mathbb{E}_{p(\mathcal{M},I_T,z|\pi)}\left[\log \frac{p(z|I_T)}{q(z|I_T;\phi)}\right] \\
&= \mathbb{E}_{p(I_T,z|\pi)}\left[\log \frac{p(z|I_T)}{q(z|I_T;\phi)}\right] \\
&= \mathbb{E}_{p(I_T|\pi)p(z|I_T)}\left[\log \frac{p(z|I_T)}{q(z|I_T;\phi)}\right] \\
&= \mathbb{E}_{p(I_T|\pi)}\left[D_{\mathrm{KL}}\Big(p(z|I_T) \;||\; q(z|I_T;\phi)\Big)\right]
\end{aligned} \tag{C.4}
$$

is an expectation of a KL divergence, which is always non-negative. Hence, $\mathcal{I}_T(\pi) \geq \mathcal{I}_{T;\,L}(\pi;\phi)$ for any π and ϕ. The bound is tight if and only if the KL divergence equals zero, which occurs when $q(z|I_T);\phi) = p(z|I_T)$ for all z and I_T. □

C.1.3 Prior-omitted EIG

We note that

$$
\begin{aligned}
\mathcal{I}_T(\pi) &= \mathbb{E}_{p(\mathcal{M})p(I_T|\mathcal{M},\pi)p(z|\mathcal{M})}\left[\log p(z|I_T)\right] - c \\
&= \mathcal{R}_T(\pi) - c,
\end{aligned} \tag{C.5}
$$

where $c := \mathbb{E}_{p(\mathcal{M})p(z|\mathcal{M})}[\log p(z)]$ is independent of π. Similarly,

$$
\begin{aligned}
\mathcal{I}_{T;\,L}(\pi;\phi) &= \mathbb{E}_{p(\mathcal{M})p(I_T|\mathcal{M},\pi)p(z|\mathcal{M})}\left[\log q(z|I_T;\phi)\right] - c \\
&= \mathcal{R}_{T;L}(\pi;\phi) - c.
\end{aligned} \tag{C.6}
$$

Proposition C.1.2. *For any policy π, variational parameter ϕ, the prior-omitted EIG satisfies $\mathcal{R}_T(\pi) \geq \mathcal{R}_{T;\,L}(\pi;\phi)$. The bound is tight if and only if $p(z|I_T) = q(z|I_T;\phi)$ for all z and I_T.*

Proof. $\mathcal{R}_T(\pi) = \mathcal{I}_T(\pi) + c \geq \mathcal{I}_{T;L}(\pi;\phi) + c = \mathcal{R}_{T;L}(\pi;\phi)$, making use of $\mathcal{I}_T(\pi) \geq \mathcal{I}_{T;L}(\pi;\phi)$ from Theorem 4.3.1. □

We adopt standard Monte Carlo to estimate $\mathcal{R}_{T;L}(\pi;\phi)$:

$$\begin{aligned}\mathcal{R}_{T;L}(\pi;\phi) &= \mathbb{E}_{p(\mathcal{M})p(I_T|\mathcal{M},\pi)p(z|\mathcal{M})}\left[\log q(z|I_T;\phi)\right] \\ &\approx \frac{1}{N}\sum_{i=1}^{N} \log q(z^i|I_T^i;\phi),\end{aligned} \tag{C.7}$$

where $\mathcal{M}^i \sim p(\mathcal{M})$, $I_T^i \sim p(I_T|\mathcal{M}^i,\pi)$, and $z^i \sim p(z|\mathcal{M}^i)$.

We further propose a NMC estimator to estimate $\mathcal{R}_T(\pi)$:

$$\begin{aligned}\mathcal{R}_T(\pi) &= \mathbb{E}_{p(\mathcal{M})p(I_T|\mathcal{M},\pi)p(z|\mathcal{M})}\left[\log p(z|I_T)\right] \\ &= \mathbb{E}_{p(\mathcal{M})p(I_T|\mathcal{M},\pi)p(z|\mathcal{M})}\left[\log p(z,I_T|\pi) - \log p(I_T|\pi)\right] \\ &= \mathbb{E}_{p(\mathcal{M})p(I_T|\mathcal{M},\pi)p(z|\mathcal{M})}\left[\log \mathbb{E}_{p(\boldsymbol{\mathcal{M}}')}[p(z,I_T|\boldsymbol{\mathcal{M}}',\pi)] - \log \mathbb{E}_{p(\boldsymbol{\mathcal{M}}'')}[p(I_T|\boldsymbol{\mathcal{M}}'',\pi)]\right] \\ &\approx \frac{1}{N}\sum_{i=1}^{N}\left[\log \frac{1}{M_1}\sum_{j_1=1}^{M_1} p(z^i,I_T^i|\mathcal{M}^{j_1},\pi) - \log \frac{1}{M_2}\sum_{j_2=1}^{M_2} p(I_T^i|\mathcal{M}^{j_2},\pi)\right],\end{aligned} \tag{C.8}$$

where $\mathcal{M}^i \sim p(\mathcal{M})$, $I_T^i \sim p(I_T|\mathcal{M}^i,\pi)$, $z^i \sim p(z|\mathcal{M}^i)$, and $\mathcal{M}^{j_1} \sim p(\boldsymbol{\mathcal{M}}')$ and $\mathcal{M}^{j_2} \sim p(\boldsymbol{\mathcal{M}}'')$. This NMC estimator is only used in Section 4.4.1 as a baseline comparison.

C.2 Detailed Related Work

Our work on GO-CBED builds upon several related lines of research.

Causal Bayesian Experimental Design Experimental design for causal discovery within a BOED framework was initially explored by [108] and [148] for discrete variables with single-target acquisition. Subsequent research extended this approach to continuous variables within BOED [1, 154, 149, 32] and alternative frameworks [89, 55, 59, 112]. Notable non-BOED methods include strategies for cyclic structures [105] and latent variables [90]. Within BOED, [147] proposed selecting single target-state pairs via stochastic batch acquisition, later extending this to gradient-based optimization to multiple target-state pairs [146]. [140] introduced a greedy method for selecting multi-target experiments without specifying intervention states. More recently, [4] proposed adaptive sequential experimental design for causal structure learning, although their objective—minimizing graph prediction error—is different from traditional BOED. [56] developed a reinforcement learning method for sequential experimental design using Prior Contrastive Estimation [49] as a reward function; however, their approach relies on initial observational data

and is computationally intensive. In contrast, our method uses direct policy optimization with differentiable rewards, enabling more efficient training without needing initial observational data.

Bayesian Causal Discovery Causal discovery has been extensively studied in machine learning and statistics [62, 71, 119, 156]. Traditional causal discovery methods typically infer a single causal graph from observational data [22, 69, 94, 117, 118, 71]. In contrast, Bayesian causal discovery [53, 70, 148] seeks to infer a posterior distribution over SCMs. Recent work [36, 97, 3] has introduced variational approximations of the DAG posterior, enabling representation of uncertainty by a full distribution rather than a point estimate. Addressing the discrete nature of DAGs—which prevents straightforward gradient-based optimization—[97] used Stein variational gradient descent (SVGD) [95] in a continuous latent embedding space, enabling efficient Bayesian inference over DAG structures.

Goal-Oriented BOED Goal-oriented BOED extends classical optimal design principles—such as L-, D_A-, I-, V-, and G-optimality [8]—by shifting the objective from general parameter estimation to directly maximizing utility for specific, downstream QoIs. The concept of tailoring experiments for particular goals, especially in nonlinear settings, was formulated by [15]. Recent work has primarily focused on addressing computational challenges associated with optimizing the EIG for QoIs [9, 161, 16, 166, 29]. For example, [9] introduced gradient-based optimization approaches for D_A- and L-optimal designs in linear models, while [161] developed scalable low-rank approximations for high-dimensional QoIs. In nonlinear scenarios, methods leveraging Markov chain Monte Carlo combined with kernel density estimation [166] and likelihood-free density-ratio estimation approaches [29] have been proposed. Although powerful for general predictive tasks, these methods typically do not accommodate the unique structure of causal queries derived from SCMs or the interventional nature inherent to causal experimentation.

Non-Myopic Sequential BOED Non-myopic sequential BOED addresses the limtiations of greedy, single-step experimental strategies by planning optimal sequences of interventions. Such methods have been broadly explored in various general settings [49, 80, 18, 134], including goal-oriented extensions such as vsOED [132]. Within causal BOED specifically, non-myopic approaches have predominantly focused on causal discovery tasks aimed at learning the graph structures [4, 56]. Although active learning methods targeting specific causal reasoning queries have also been proposed [149], these typically employ myopic (single-step) intervention designs, thus limiting their ability to strategically plan for long-term gains.

C.3 Experiment Details

C.3.1 Hyperparameter Settings For Policy and Posterior Networks

The input to the policy network has shape ($n = n_{\text{int}} \times T$, d, 2), where the last dimension encodes the intervention data and binary intervention masks. The policy network architecture (see Figure C.1) proceeds as follows:

1. The input is passed through a fully connected layer, transforming it to shape ($n_{\text{int}} \times T$, d, $n_{\text{embedding}}$).

2. The embedded representation is processed through L stacked Transformer layers. Each layer includes:

   - Two multi-head self-attention sublayers, each preceded by layer normalization and followed by dropout.
   - A feedforward fully-connected (FFN) sublayer, also preceded by layer normalization and followed by dropout.

   Residual connections are applied after each sublayer. This output retains the shape ($n_{\text{int}} \times T, d, n_{\text{embedding}}$).

3. A max-pooling operation is applied across the $n_{\text{int}} \times T$ dimension, yielding a compressed representation of shape ($d, n_{\text{embedding}}$).

4. The pooled representation is passed through:

   - A target prediction layer, followed by a Gumbel-softmax transformation with temperature τ, producing a discrete intervention target vector.
   - A separate value layer, with final outputs scaled to fall within a specific range $\min_{val}$ and $\max_{val}$.

The detailed implementation setup is provided in Table C.1. The "step" associated with τ refers to the current training step, and the values of T, n_{step}, and n_{envs} per training step are kept to be the same as those used for training the posterior networks (see below).

For the posterior networks, the initial input has shape (n_{envs}, $n_{\text{int}} \times T$, d, 2), representing full trajectories. The processing steps follows the same as those of the policy network up to step 3, resulting in a max-pooled output of shape (n_{envs}, d, $n_{\text{embedding}}$). Specific to the causal discovery case, starting from step 4:

4. The pooled representation is processed as follows:

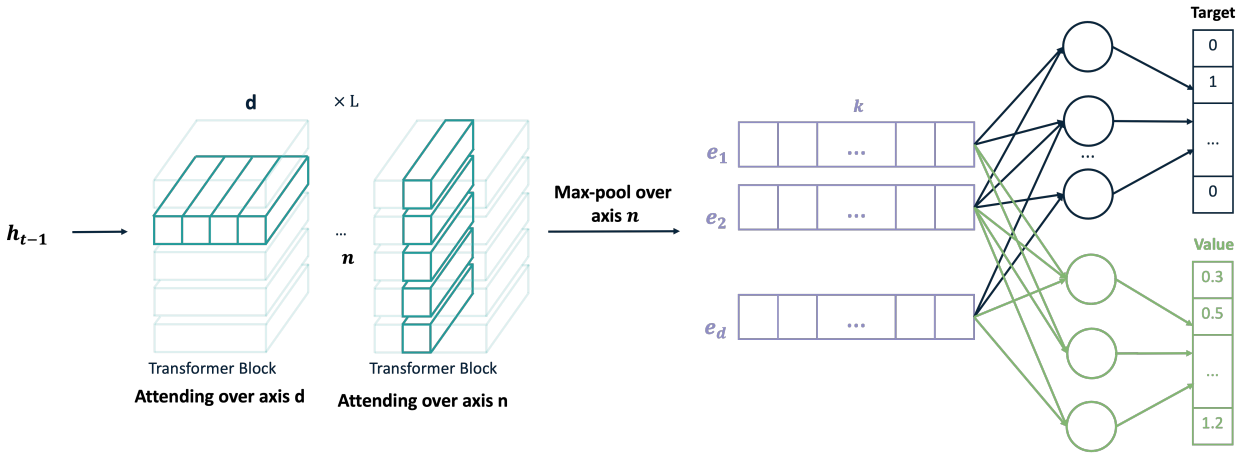


Figure C.1: Policy network architecture. The model takes as input a three-dimensional tensor of shape $n \times d \times 2$, where $n = n_{\text{int}} \times T$. It is permutation-invariant along the n-axis and permutation-equivariant along the d-axis. Each of the L layers first applies self-attention across the d-axis, followed by attention across the n-axis, with shared parameters across the non-attended axis.

Table C.1: Hyperparameter settings for the policy network.

Hyperparameter	**Value**
Embedding dimension $n_{\text{embedding}}$	32
Number of transformer layers (L)	4
Key size in self-attention	16
Number of attention heads	8
FFN dimensions	$(n_{\text{embedding}}, 4 \times n_{\text{embedding}}, n_{\text{embedding}})$
Activation	ReLU
Dropout rate	0.05
$\max_{val}$	10
$\min_{val}$	-10
τ	$\min(5 \times 0.9995^{\text{step}}, 0.1)$
Initial learning rate	5×10^{-4} or 10^{-4}
Scheduler	ExponentialLR with $\gamma = 0.8$, step every 1000 training steps
T	10 when $d = 10, 20$ 15000 when $d = 30$
n_{step}	10000 when $d = 10$ 15000 when $d = 20, 30$
n_{env} per training step	10

- Two independent linear transformations are applied to produce vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, each of shape $(n_{\text{envs}}, d, n_{\text{out}})$.
- Both $\boldsymbol{u}$ and $\boldsymbol{v}$ are normalized using their ℓ_2-norm along the last dimension.

5. Pairwise edge logits are computed:
   - A dot product between every pair of variables $\boldsymbol{u}_i$ and $\boldsymbol{v}_j$, resulting in a tensor of shape (n_{envs}, d, d).
   - The logits are scaled by a learnable temperature parameter "temp" via the operation $\text{logit}_{ij} \times \exp(\text{temp})$, which is then added element-wise with a learnable term, "bias".

The detailed implementation setup is provided in Table C.2.

Table C.2: Hyperparameter settings for the posterior network in the causal discovery case.

Hyperparameter	**Value**
Embedding dimension $n_{\text{embedding}}$	128
Number of transformer layers (L)	8
Key size in self-attention	64
Number of attention heads	8
FFN dimensions	$(n_{\text{embedding}}, 4 \times n_{\text{embedding}}, n_{\text{embedding}})$
Activation	ReLU
Dropout rate	0.05
Bias	-3
Temp	2
Initial learning rate	10^{-4}
Scheduler	ExponentialLR with $\gamma = 0.8$, step every 1000 training steps

Specific to the causal reasoning case, starting from step 4:

4. The pooled representation is flattened to shape $(n_{\text{envs}}, d \times n_{\text{embedding}})$ and passed into the $s(\cdot)$ and $t(\cdot)$ networks, with n_{trans} transformations in total. The final output has shape (n_{envs}, n_z).

The detailed implementation setup is provided in Table C.3.

C.3.2 Example in Section 4.4.1

In this example, we consider a fixed causal graph structure with Gaussian priors on parameters:

$$\theta_{12} \sim \mathcal{N}(0.1, 1), \quad \theta_{23} \sim \mathcal{N}(1, 0.2^2), \quad \theta_{35} \sim \mathcal{N}(0.2, 0.5^2), \quad \theta_{45} \sim \mathcal{N}(-0.5, 0.5^2),$$
$$\theta_{13} \sim \mathcal{N}(-0.2, 0.5^2), \quad \theta_{24} \sim \mathcal{N}(0.3, 0.3^2), \quad \theta_{36} \sim \mathcal{N}(0, 0.5^2), \quad \theta_{56} \sim \mathcal{N}(0, 0.5^2),$$
$$\theta_{14} \sim \mathcal{N}(-0.5, 0.3^2),$$

Table C.3: Hyperparameter settings for the posterior network in the causal reasoning case.

Hyperparameter	**Value**
Embedding dimension $n_{\text{embedding}}$	16
Number of transformer layers (L)	8
Key size in self-attention	16
Number of attention heads	8
FFN dimensions	$(n_{\text{embedding}}, 4 \times n_{\text{embedding}}, n_{\text{embedding}})$
Activation	ReLU
Dropout rate	0.05
n_{trans}	4
$s(\cdot)$ and $t(\cdot)$ dimensions	(256, 256, 256)
Initial learning rate	5×10^{-4} or 10^{-3}
Scheduler	ExponentialLR with $\gamma = 0.8$, step every 1000 training steps

with observation model $Y_i = \theta_i^\top Y_{\text{pa}(i)} + \epsilon_i$ where $\theta_i = [\theta_{ij}]^\top$, $j \in \text{pa}(i)$, and additive Gaussian noise $\epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$ with standard deviations $\sigma = \{0.2, 0.2, 0.2, 0.2, 0.3, 0.3\}$. This linear-Gaussian setup enables analytical posterior computations and efficient estimation of the variational lower bound $\mathcal{R}_{T;L}(\pi; \phi)$.

Figure C.2 provides a qualitative assessment of the posterior approximation $q(z|I_T; \phi)$ achieved by NFs. The NF-based approximations closely align with the true posterior predictive distributions $p(z|h_T)$ across these two examples, demonstrating a high-quality posterior approximation.

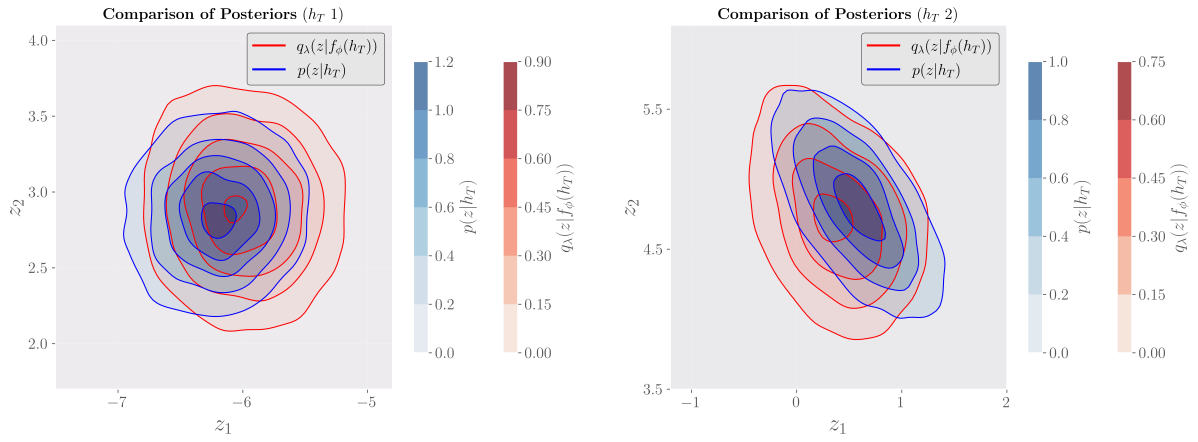


Figure C.2: Comparison between the true posterior predictive distribution $p(z|I_T)$ and the variational approximation $q(z|I_T; \phi)$) for two simulated trajectories. The approximate posterior closely aligns with the true posterior.

C.3.3 Examples in Sections 4.4.2 and 4.4.3

In the synthetic experiments, we consider Erdös–Rényi (ER) and Scale-Free (SF) random graphs as priors over graph structures. For semi-synthetic experiments, we utilize gene regulatory networks derived from the DREAM benchmarks [64], which reflect realistic biological scenarios.

C.3.3.1 Priors over Graph Structures

Erdös–Rényi In the ER model, each potential edge between node pairs is included independently with a fixed probability p. Given n nodes, the resulting random undirected graph has a number of edges that follows a binomial distribution, with the expected number of edges equal to $p \times \binom{n}{2}$. Following [98], we scale p such that the expected number of edges is $O(d)$, where d is the desired average degree. To obtain a DAG, we first retain only the lower triangular portion of the adjacency matrix and then apply a random permutation to the node indices to break symmetry.

Scale-Free SF graphs exhibit a power-law degree distribution, where the probability that a node has degree k is proportional to $k^{-\gamma}$, with an exponent $\gamma > 1$ [11]. Consequently, a small subset of nodes ("hubs") has a very high number of connections, while most nodes have relatively few. Such structures are commonly observed in biological and social networks. We generate SF graphs using the Barabási–Albert preferential attachment model implemented in NetworkX, which iteratively adds nodes by connecting them preferentially to existing high-degree nodes.

Realistic Gene Regulatory Networks For semi-synthetic scenarios, we employ networks from the DREAM benchmarks [64], widely used for evaluating computational approaches to reverse-engineering biological systems. DREAM datasets provide realistic simulations of gene regulatory and protein signaling networks generated by GeneNetWeaver v3.12. Specifically, our experiments focus on two DREAM subnetworks—the *E. coli* and Yeast networks—following the setup described in [146].

C.3.3.2 Mechanisms

Linear Model In the linear setting, each variable Y_i is modeled as a linear function of its parent variables $Y_{\mathrm{pa}(i)}$ according to

$$Y_i = \theta_i^\top Y_{\mathrm{pa}(i)} + b_i + \epsilon_i, \tag{C.9}$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ with fixed variance $\sigma^2 = 0.1$. The parameters have priors $\theta_i \sim \mathcal{N}(0, 2)$ and $b_i \sim \mathcal{U}(-1, 1)$.

Nonlinear Model For the nonlinear setting, the functional relationship between each child and its parent is modeled using a feedforward neural network with two hidden layers, each containing 8 ReLU-activated neurons. All weights and biases have standard normal priors.

Queries for Causal Reasoning For the causal reasoning experiments shown in Figure 4.3, the query QoIs for the four panels are: $z = \{Y_6, Y_8 \,|\, \mathrm{do}(Y_2 \sim \mathcal{N}(5, 2^2))\}$; $\{X_0, Y_5 \,|\, \mathrm{do}(Y_6 \sim \mathcal{N}(3, 1))\}$; $\{Y_3, Y_5 \,|\, \mathrm{do}(Y_5 \sim \mathcal{N}(6, 0.5^2))\}$; and $\{Y_3, Y_4 \,|\, \mathrm{do}(Y_9 \sim \mathcal{N}(4, 1))\}$. For the *E. coli* case in Figure 4.4, the QoI is $z = \{Y_6, Y_8 \,|\, \mathrm{do}(Y_7 \sim \mathcal{N}(4, 2^2))\}$.

Comparison to Discovery-Oriented Policy To contextualize the performance of GO-CBED-z, we also include the performance of the structure-learning-oriented policy $\pi^*_{\boldsymbol{G}}$, evaluated on $\mathcal{R}_{T;L}(\pi^*_{\boldsymbol{G}})$, as shown in in Figures C.3 and C.4. While GO-CBED-G is effective for causal discovery, it is consistently outperformed by GO-CBED-z when the objective is to estimate specific causal inquiries, as seen by comparing to Figures 4.3 and 4.4.

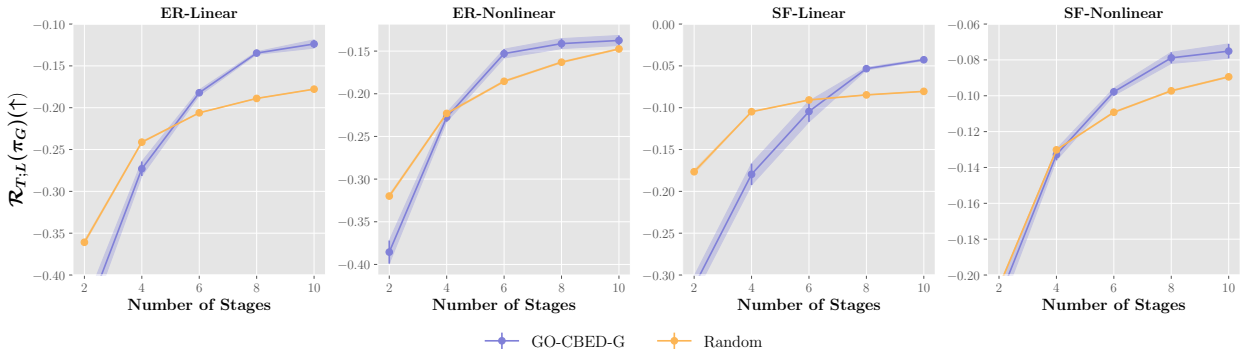


Figure C.3: Evaluation of policies on ER and SF graphs with both linear and nonlinear causal mechanisms. The $\pi^*_{\boldsymbol{G}}$ demonstrates strong performance in accurately identifying the underlying causal graph across all settings.

C.3.4 Incorporating Existing Observational Data into the Prior

In practical settings, it is common to have access to a set of observational data $\mathcal{D}$ prior to designing interventions. This data can be used to update the prior into a posterior, which then serves as an informative prior for the subsequent experimental design.

We infer both the posterior over the graph structure $p(G|\mathcal{D})$ and the parameters $p(\theta|\mathcal{D}, G)$ in two stages. First, since the realized data may not be available during posterior construction, we treat $\mathcal{D}$ as a random variable. We infer the graph structure using the approach of [98], and train an

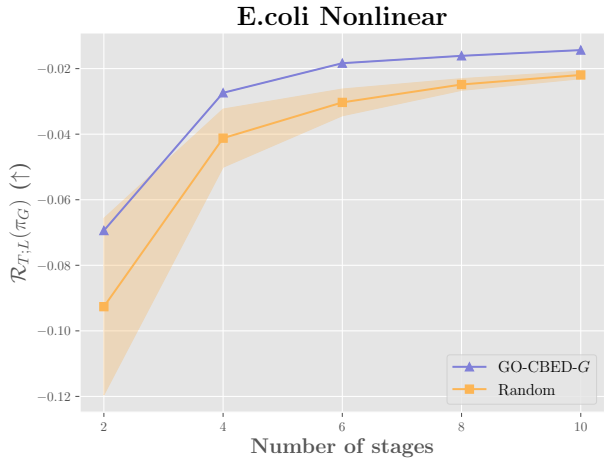


Figure C.4: Evaluation of policies on *E. coli* graphs with nonlinear causal mechanisms. The $\pi^*_{\boldsymbol{G}}$ demonstrates strong performance in accurately identifying the underlying causal graph.

amortized variational posterior that generalizes across potential realizations of $\mathcal{D}$, by minimizing

$$\mathbb{E}_{p(\mathcal{D})}\left[D_{\mathrm{KL}}\left(p(G|\mathcal{D})\,||\,q(\mathcal{D};\phi)\right]\right. \tag{C.10}$$

with respect to ϕ. The approximate posterior $q(\cdot)$ is modeled as a product of independent Bernoulli distributions over potential edges. Once a specific realization $\mathcal{D}^*$ becomes available, the posterior is instantiated via substitution as $q(\mathcal{D}^*;\phi)$. Samples from this distribution are drawn from the Bernoulli marginals and retaining only acyclic graphs to ensure valid DAGs.

Second, to perform inference over the parameters θ, we exploit the conditional independence structure of the posterior:

$$p(\theta|\mathcal{D},G) = \prod_j p(\theta_j|\mathcal{D}_{\mathrm{pa}(j)},\mathcal{D}_j,G). \tag{C.11}$$

This factorization enables efficient sampling of θ by decomposing the joint posterior into node-wise conditionals.

For linear models, we sample directly from the posterior $p(\theta_j|G,\mathcal{D})$ using Markov chain Monte Carlo. For nonlinear models, we apply Pyro's Stochastic Variational Inference (SVI) [17] to learn a mean-field Gaussian approximation to the posterior.

C.4 Additional Experiments

C.4.1 Higher Bias in the NMC Estimator

We provide a qualitative comparison between the NMC estimator and the GO-CBED approach. The experiment follows the setup in Figure C.5, which assumes a fixed graph with $T = 1$ and additive Gaussian noise $\epsilon_i \sim \mathcal{N}(0, 0.3^2)$ for all observations. Interventions are uniformly selected in integers from -5 to 5, and the $\mathcal{R}_T$ or $\mathcal{R}_{T;L}$ is evaluated using the NMC and GO-CBED estimators and presented in Figure C.6. Since there is no policy optimization in this setting, GO-CBED reduces to training a variational posterior network using NFs.

The NMC estimator uses an outer loop size of $5,000$ samples, and the inner loop sample size is indicated in the parenthesis in the legend of Figure C.6. This sample size is also used as the training sample size for GO-CBED. Despite using significantly fewer samples, GO-CBED consistently identifies the optimal EIG near the boundary of the design space. This finding reinforces our observation from Section 4.4.1: variational approximations via GO-CBED can offer more efficient and reliable EIG estimation compared to NMC, especially in causal inference tasks involving large graphs or high-dimensional parameter spaces, where traditional sampling becomes computationally and memory intensive.

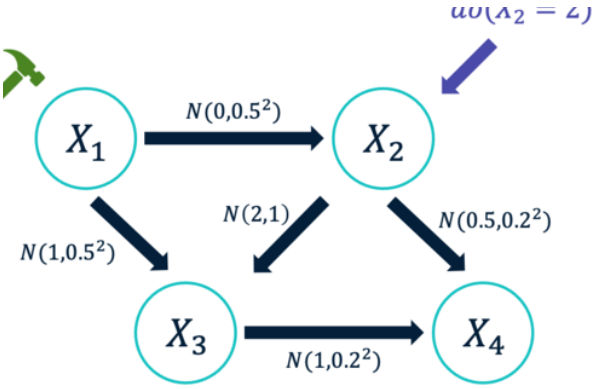


Figure C.5: Evaluation of interventions on node 1 using integers from -5 to 5, with the causal query defined as $z = \{Y_3, Y_4 \mid \mathrm{do}(Y_2 = 2)\}$.

C.4.2 Causal Reasoning on Diverse Realistic Graph Structures

In Section 4.4.2, we presented causal reasoning results using the *E. coli* gene regulatory network. Here, we extend the analysis to include additional tasks based on both *E. coli* and Yeast gene regulatory networks, each incorporating nonlinear mechanisms. These supplementary experiments further demonstrate the robustness and generality of GO-CBED across a range of causal graph

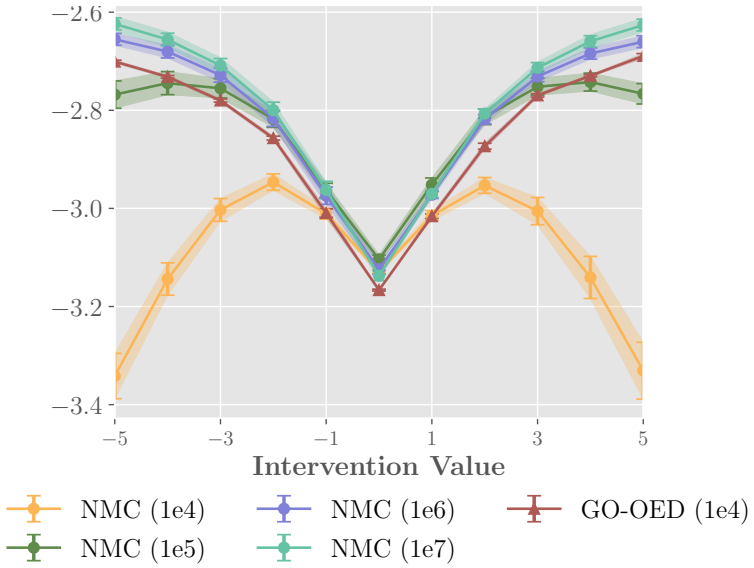


Figure C.6: Prior-omitted EIG lower bound estimates, with parenthesis values denoting the inner loop sample size for NMC and training sample size for GO-CBED. Shaded regions represent ±1 standard error across 4 random seeds.

structures and varying complexities of intervention-target relationships (see Figure C.7).

Specifically, Figure C.7(a) corresponds to the main result in the paper. Figures C.7)(b) and (c) illustrate additional scenarios, highlighting GO-CBED's consistent advantages across diverse topologies. In Figure C.7)(b), GO-CBED achieves approximately twice the EIG compared to baselines. This improvement likely stems from the complex paths linking intervention nodes to targets, where goal-oriented strategies more effectively exploit structural dependencies.

In contrast, Figure C.7)(c) shows a reduced performance gap. This is likely due to node Y_5 being highly informative for both causal discovery and targeted queries, aligning the objective of structure learning and query-specific inference. Notably, the random policy also performs competitively in this setting, likely benefiting from the high-quality variational posterior achievable even under random interventions. We leave a deeper investigation of this phenomenon as an interesting direction for future work.

In the causal discovery setting, we additionally assess robustness under distributional shift in observation noise at deployment. Posterior networks are pre-trained under random intervention with shifted noise, using linear models with ER and SF priors (10 nodes). As shown in Figure C.8, GO-CBED continues to outperform the random baseline, demonstrating its reliability across multiple causal tasks and noise conditions.

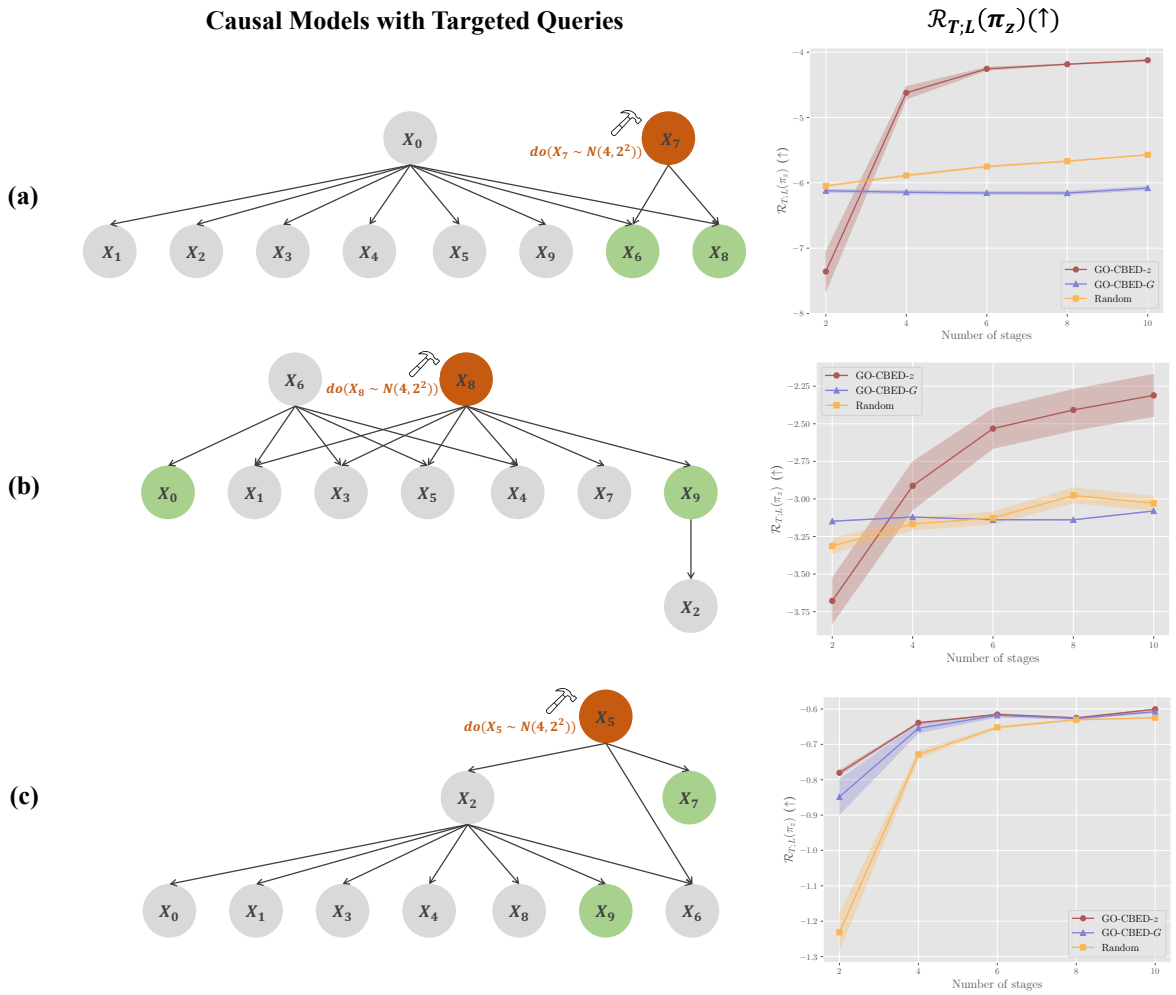


Figure C.7: Additional causal reasoning experiments on nonlinear gene regulatory networks. **(a)** *E. coli* network in Section 4.4.2: intervention on node 7 targeting nodes 6 and 8. **(b)** Yeast network: intervention on node 8 targeting nodes 0 and 9. **(c)** *E. coli* network: intervention on node 5 targeting nodes 7 and 9. GO-CBED consistently outperforms baseline methods, with performance gains varying based on structural complexity of the intervention-target relationships. Shaded regions represent ±1 standard error over 4 random seeds.

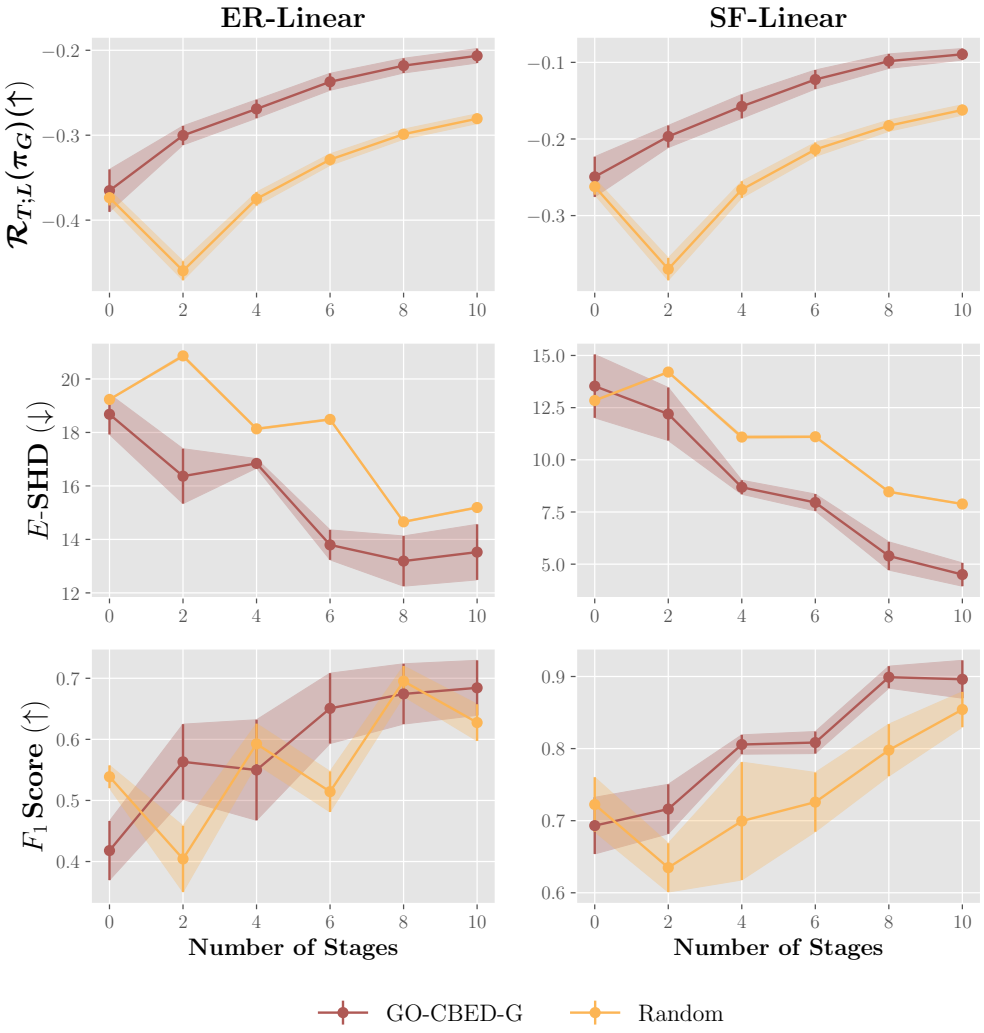


Figure C.8: Evaluation of GO-CBED under a distributional shift in observation noise at deployment for linear causal reasoning tasks. GO-CBED consistently outperforms the random baseline that is using posterior networks pre-trained under the shifted noise, demonstrating the robustness of its jointly optimized policy and posterior networks.

C.4.3 Performance Comparison with Original Posterior Inference Methods

In the main paper, we establish a fair comparison among policies by evaluating them using their respectively pre-trained variational posteriors, thereby isolating the impact of policy optimization. However, a key advantage of GO-CBED is its joint training of both the policy and posterior networks. Specifically, the posterior networks trained in GO-CBED can themselves serve as a highly efficient inference tool, even independent of the policy.

Here, we provide additional results in which each baseline method is evaluated using its original posterior inference procedure, as proposed in their respective papers. Specifically, CAASL [4] uses a fixed pre-trained AVICI posterior [97]; DiffCBED [146] and SoftCBED [147] rely on DAG-Bootstrap [52, 69] for linear SCMs and DiBS [97] for nonlinear SCMs.

Figures C.9 and C.10 present results on synthetic and semi-synthetic SCMs with both linear and nonlinear mechanisms. GO-CBED consistently outperforms baselines across both the $\mathbb{E}$-SHD and F_1 score metrics. Interestingly, even the random policy—when paired with pre-trained variational posteriors—achieves competitive performance, highlighting the advantage variational posteriors bring to causal learning.

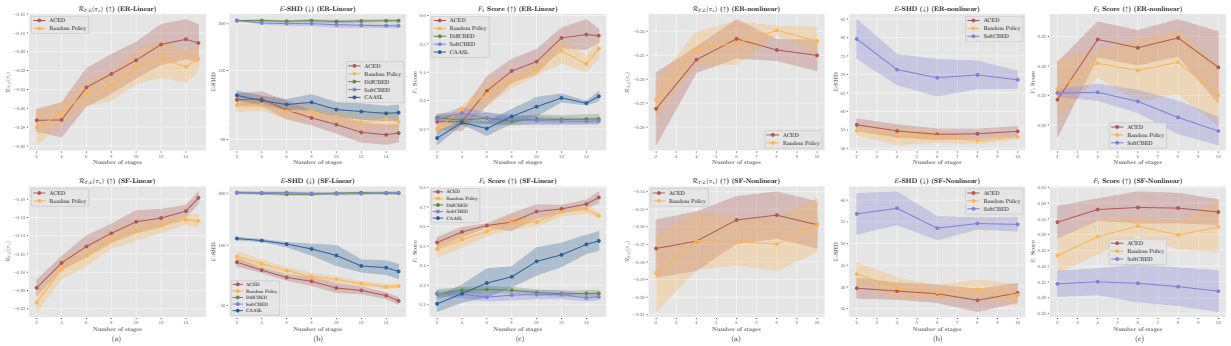


Figure C.9: Performance comparison on synthetic SCMs, with each method using its originally proposed posterior inference approach. GO-CBED consistently outperforms all baselines across both linear and nonlinear settings, demonstrating the advantage of jointly optimizing policy and posterior networks. Shaded regions represent ±1 standard error across 10 random seeds.

C.4.4 Extended Evaluation on Causal Discovery Tasks

We extend our analysis with additional causal discovery experiments on more complex settings, with nonlinear mechanisms in *E. coli* networks (Figure C.11), and both linear (Figure C.12) and nonlinear (Figure C.13) mechanisms in Yeast gene regulatory networks. While GO-CBED continues to outperforms baseline methods, the performance gap narrows in the nonlinear setting, reflecting the increased difficulty of recovering the full causal structure under complex settings.

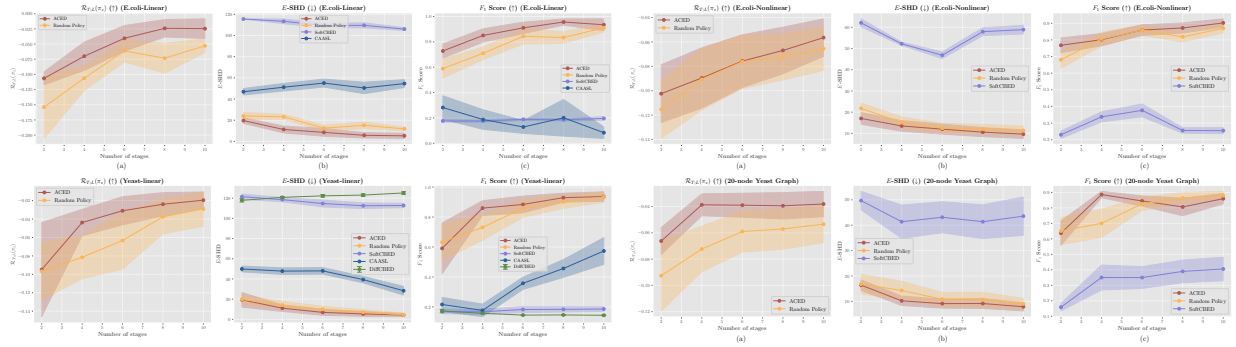


Figure C.10: Performance comparison on semi-synthetic gene regulatory (*E. coli* and Yeast) networks, with each method using its originally proposed posterior inference approach. GO-CBED demonstrates strong performance on causal tasks in biologically inspired settings. Shaded regions represent ±1 standard error across 10 random seeds.

These results align with our motivation for goal-oriented BOED: when the objective is to answer specific causal queries rather than reconstruct the entire model, targeted policies like GO-CBED-z offer greater efficiency and practical relevance.

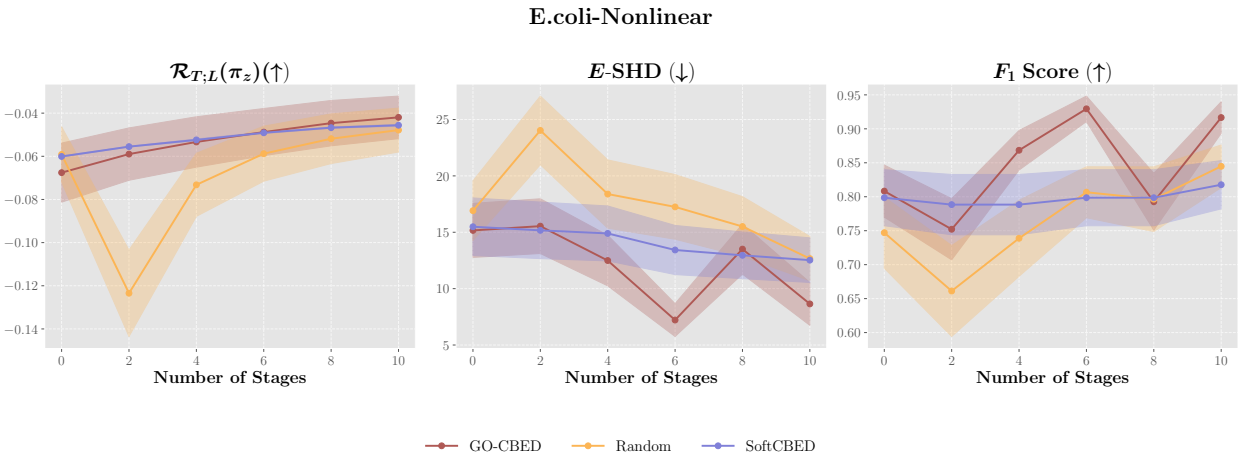


Figure C.11: Causal discovery performance on nonlinear *E. coli* gene regulatory networks. GO-CBED performs better or comparatively in terms of uncertainty reduction ($\mathcal{R}_{T;L}$), structural recovery ($\mathbb{E}$-SHD), and structural accuracy (F_1-score) compared to all baselines. Shaded regions indicate ±1 standard error across 10 random seeds.

C.4.5 Distributional Shift in Observation Noise

We evaluate the robustness of GO-CBED's policy and posterior networks under distributional shift in observation noise. At deployment, the noise variance σ_i^2 is sampled from an inverse Gamma distribution, $\sigma_i^2 \sim \text{InverseGamma}(10, 1)$, in contrast to the fixed variance ($\sigma_i^2 = 0.1$) assumed

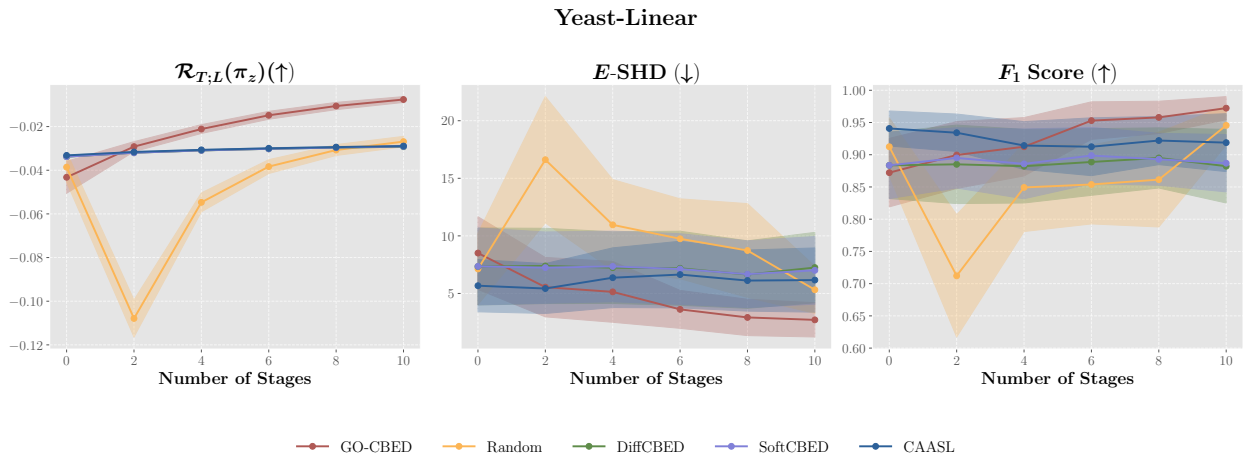


Figure C.12: Causal discovery performance on linear Yeast gene regulatory networks. GO-CBED performs better or comparatively in terms of uncertainty reduction ($\mathcal{R}_{T;L}$), structural recovery ($\mathbb{E}$-SHD), and structural accuracy (F_1-score) compared to all baselines. Shaded regions represent ± 1 standard error across 10 random seeds.

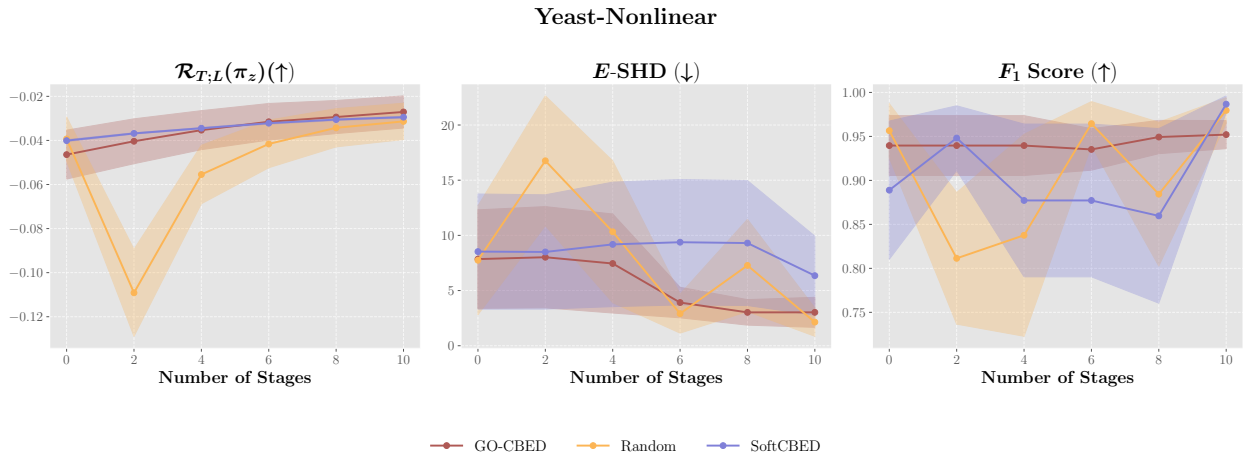


Figure C.13: Causal discovery performance on nonlinear Yeast gene regulatory networks. GO-CBED performs better or comparatively in terms of uncertainty reduction ($\mathcal{R}_{T;L}$), structural recovery ($\mathbb{E}$-SHD), and structural accuracy (F_1-score) compared to all baselines. Shaded regions represent ± 1 standard error across 10 random seeds.

during training. For comparison, we include a random intervention policy baseline, paired with a posterior network trained specifically on data with the shifted noise distribution.

The results for linear causal reasoning tasks were previously presented in Figure C.8. We now focus on nonlinear causal reasoning tasks, with ER and SF graph priors over 10-node networks. As shown in Figure C.14, GO-CBED maintains strong performance, demonstrating the robustness of both its policy and posterior networks in the presence of heteroskedastic noise.

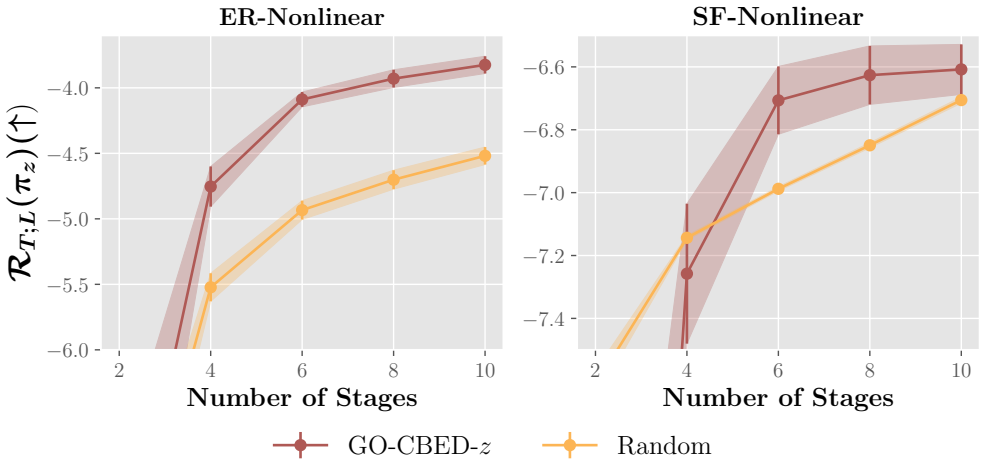


Figure C.14: Evaluation of GO-CBED under a distributional shift in observation noise at deployment for nonlinear causal reasoning tasks. GO-CBED consistently outperforms the random baseline that is using posterior networks pre-trained under the shifted noise, demonstrating the robustness of its jointly optimized policy and posterior networks.

CHAPTER D

Appendix for Symmetrized EIG

D.1 Sufficient Experiment

Def: Sufficiency of Experiment An experiment/design ξ_1 observing Y_1 is sufficient for ξ_2 observing Y_2, if there exist a non-negative function h satisfying the following conditions:

$$p_{Y_2}(y_2|\theta) = \int_{Y_1} h(y_2, y_1) p_{Y_1}(y_1|\theta) dy_1 \quad \text{for } \theta \in \Theta,\ y_2 \in \mathcal{Y}_2 \tag{D.1}$$

$$\int_{Y_2} h(y_2, y_1) dy_2 = 1 \qquad \text{for } y_1 \in \mathcal{Y}_1 \tag{D.2}$$

$$0 < \int_{Y_1} h(y_2, y_1) dy_1 < \infty \qquad \text{for } y_2 \in \mathcal{Y}_2 \tag{D.3}$$

then h is called a stochastic transformation from y_1 to y_2 [38].

We first introduce the Likelihood to Evidence Ratio: **Likelihood to Evidence Ratio:** $T(y) = \{\frac{p(y|\theta_1)}{p(y)}, ..., \frac{p(y|\theta_k)}{p(y)}\}$ is the likelihood to evidence ratio with $E_Y(T(y)) = \{1, ...1\}$. According to Bayes Rule, $T(y) = \{\frac{p(\theta_1|y)}{p(\theta_1)}, ..., \frac{p(\theta_k|y)}{p(\theta_k)}\}$ is also the **Posterior to Prior Ratio**. The range of values taken by $T(y)$ is a subset of the set: $K_\Theta = \{u = (u_1, ..., u_k) \in R^k : u_i \geq 0, \sum_{i=1}^k p(\theta_i) u_i = 1\}$.

Proposition D.1.1. *Experiment ξ_1 observing Y_1 is sufficient for experiment observing Y_2 if and only if*

$$E_{Y_1}[\phi(T(y_1))] \geq E_{Y_2}[\phi(T(y_2))] \quad \textit{for every convex function } \phi \textit{ on } K_\Theta \tag{D.4}$$

or

$$E_{Y_1}[\varphi(p(\theta|y_1))] \geq E_{Y_2}[\varphi(p(\theta|y_2))] \quad \textit{for every convex function } \varphi \textit{ on the simplex of } \mathbb{R}^k \tag{D.5}$$

Proof. (for sufficiency)

For vector $a, b \in \mathbb{R}^k$. , let $a \otimes b = \frac{1}{a \cdot b}(a_1 b_1, ..., a_k b_k)$. Let $\gamma \in K_\Theta$ be any fixed vector, and the function $v(\cdot)$ defined on $\mathbb{R}^k$ by the following relation:

$$v(a) = (\gamma \cdot a)\phi(\gamma \otimes a) \tag{D.6}$$

where ϕ a convex function on K_Θ.

First to show that $v(\cdot)$ is a convex function, which requires showing for $\alpha + \beta = 1, v(\alpha a + \beta b) \leq \alpha v(a) + \beta v(b)$. If either $\gamma \cdot a > 0$ or $\gamma \cdot b > 0$, then

$$v(\alpha a + \beta b) = [\alpha(\gamma \cdot a) + \beta(\gamma \cdot b)]\phi(\alpha^*(\gamma \otimes a) + \beta^*(\gamma \cdot b)) \tag{D.7}$$

where $\alpha^* = \frac{\alpha(\gamma \cdot a)}{\alpha(\gamma \cdot a) + \beta(\gamma \cdot b)}, \beta^* = \frac{\beta(\gamma \cdot b)}{\alpha(\gamma \cdot a) + \beta(\gamma \cdot b)}$, as ϕ is convex, then:

$$\begin{aligned} v(\alpha a + \beta b) &\leq [\alpha(\gamma \cdot a) + \beta(\gamma \cdot b)](\alpha^* \phi(\gamma \otimes a) + \beta^* \phi(\gamma \cdot b)) \\ &= \alpha(\gamma \cdot a)\phi(\gamma \otimes a) + \beta(\gamma \cdot b)\phi(\gamma \cdot b)) \end{aligned} \tag{D.8}$$

Since equality is achieved when $\gamma \cdot a = \gamma \cdot b = 0$, the convexity is proven. Then define $\psi(\cdot|y)$:

$$\psi(\theta_i|y_2) = \frac{\int_{Y_1} h(y_2, y_1) p_{Y_1}(y_1|\theta_i) dy_1}{\int_{Y_1} h(y_2, y_1) dy_1}, \quad i = 1, ..., k \tag{D.9}$$

since

$$\begin{aligned} E_{Y_2}[\varphi(p(\theta|y_2))] &= \int_{Y_2} p(y_2)\phi[p(\theta_1|y_2), ...p(\theta_k|y_2)]dy_2 \\ &= \int_{Y_2} (\sum_{i=1}^{k} p(\theta_i)p(y_2|\theta_i))\phi[\frac{p(\theta_1)p(y_2|\theta_1)}{p(y_2)}, ...\frac{p(\theta_k)p(y_2|\theta_k)}{p(y_2)}]dy_2 \end{aligned} \tag{D.10}$$

Thus $E_{Y_2}[\varphi(p(\theta|Y_2))]$ is in the form of $\int_{Y_2} v(p_{Y_2}(y_2|\cdot))dy_2 = \int_{Y_2} v(\psi(\cdot|y_2))[\int_{Y_1} h(y_2, y_1)dy_1]dy_2$ (by D.3). By defining

$$p^*_{Y_1}(y_1|y_2) = \frac{h(y_2, y_1)}{\int_{Y_1} h(y_2, y_1)dy_1}, \quad y_1 \in \mathcal{Y}_1 \tag{D.11}$$

Then $\psi(\theta_i|y_2) = E[p(y_1|\theta_i)]$ over $p^*_{Y_1}(y_1|y_2)$, and by Jensen's Inequality:

$$v(\psi(\theta_i|y_2)) \leq \int_{Y_1} v(p(y_1|\theta_i))p^*_{Y_1}(y_1|y_2)dy_1, \tag{D.12}$$

which leads to

$$
\begin{aligned}
E_{Y_2}[\varphi(p(\theta|y_2))] &\leq \int_{Y_2}\int_{Y_1} v(p_{Y_1}(y_1|\cdot))p^*_{Y_1}(y_1|y_2)dy_1[\int_{Y_1} h(y_2,y_1)dy_1]dy_2 \\
&= \int_{Y_2}\int_{Y_1} v(p_{Y_1}(y_1|\cdot))h(y_2,y_1)dy_1dy_2 \\
&= \int_{Y_1} v(p_{Y_1}(y_1|\cdot))[\int_{Y_2} h(y_2,y_1)dy_2]dy_1 \\
&= E_{Y_1}[\varphi(p(\theta|y_1))] \qquad \text{(D.13)}
\end{aligned}
$$

Thus proven. □

BIBLIOGRAPHY

[1] Raj Agrawal, Chandler Squires, Karren Yang, Karthikeyan Shanmugam, and Caroline Uhler. Abcd-strategy: Budgeted experimental design for targeted causal structure discovery. In The 22nd International Conference on Artificial Intelligence and Statistics, pages 3400–3409. PMLR, 2019.

[2] Linda JS Allen. A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. Infectious Disease Modelling, 2(2):128–142, 2017.

[3] Yashas Annadani, Jonas Rothfuss, Alexandre Lacoste, Nino Scherrer, Anirudh Goyal, Yoshua Bengio, and Stefan Bauer. Variational causal networks: Approximate bayesian inference over causal structures. arXiv preprint arXiv:2106.07635, 2021.

[4] Yashas Annadani, Panagiotis Tigas, Stefan Bauer, and Adam Foster. Amortized active causal induction with deep reinforcement learning. Advances in Neural Information Processing Systems, 37:44216–44239, 2024.

[5] Ziqiao Ao and Jinglai Li. An approximate KLD based experimental design for models with intractable likelihoods. In Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics, volume 108 of Proceedings of Machine Learning Research, pages 3241–3251. PMLR, 26–28 Aug 2020.

[6] Lynton Ardizzone, Jakob Kruse, Carsten Rother, and Ullrich Köthe. Analyzing inverse problems with invertible neural networks. In International Conference on Learning Representations, 2019.

[7] Kenneth J Arrow, Hollis B Chenery, Bagicha S Minhas, and Robert M Solow. Capital-labor substitution and economic efficiency. The Review of Economics and Statistics, 43(3):225–250, 1961.

[8] Anthony Atkinson, Alexander Donev, and Randall Tobias. Optimum Experimental Designs, with SAS. Oxford University Press, 05 2007.

[9] Ahmed Attia, Alen Alexanderian, and Arvind Saibaba. Goal-oriented optimal design of experiments for large-scale bayesian linear inverse problems. Inverse Problems, 34, 02 2018.

[10] Ricardo Baptista, Youssef Marzouk, and Olivier Zahm. On the representation and learning of monotone triangular transport maps. Foundations of Computational Mathematics, pages 1–46, 2023.

[11] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. Science, 286(5439):509–512, 1999.

[12] David Barber and Felix Agakov. The IM algorithm: A variational approach to information maximization. In Advances in Neural Information Processing Systems 16, pages 201–208. MIT Press, 2003.

[13] Joakim Beck, Ben Mansour Dia, Luis FR Espath, Quan Long, and Raul Tempone. Fast Bayesian experimental design: Laplace-based importance sampling for the expected information gain. Computer Methods in Applied Mechanics and Engineering, 334:523–553, 2018.

[14] Mohamed Ishmael Belghazi, Aristide Baratin, Sai Rajeswar, Sherjil Ozair, Yoshua Bengio, Aaron Courville, and R Devon Hjelm. Mine: mutual information neural estimation. arXiv preprint arXiv:1801.04062, 2018.

[15] Jose M. Bernardo. Expected information as expected utility. The Annals of Statistics, 7(3):686–690, 1979.

[16] Freddie Bickford Smith, Andreas Kirsch, Sebastian Farquhar, Yarin Gal, Adam Foster, and Tom Rainforth. Prediction-oriented bayesian active learning. In Francisco Ruiz, Jennifer Dy, and Jan-Willem van de Meent, editors, Proceedings of the 26th International Conference on Artificial Intelligence and Statistics, volume 206 of Proceedings of Machine Learning Research, pages 7331–7348. PMLR, 25–27 Apr 2023.

[17] Eli Bingham, Jonathan P Chen, Martin Jankowiak, Fritz Obermeyer, Neeraj Pradhan, Theofanis Karaletsos, Rohit Singh, Paul Szerlip, Paul Horsfall, and Noah D Goodman. Pyro: Deep universal probabilistic programming. Journal of Machine Learning Research, 20(28):1–6, 2019.

[18] Tom Blau, Edwin V. Bonilla, Iadine Chades, and Amir Dezfouli. Optimizing sequential experimental design with deep reinforcement learning. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato, editors, Proceedings of the 39th International Conference on Machine Learning (ICML 2022), volume 162 of Proceedings of Machine Learning Research, pages 2107–2128. PMLR, 2022.

[19] David M Borth. A total entropy criterion for the dual problem of model discrimination and parameter estimation. Journal of the Royal Statistical Society: Series B (Methodological), 37(1):77–87, 1975.

[20] George Edward Pelham Box. Sequential experimentation and sequential assembly of designs. Quality Engineering, 5(2):321–330, 1992.

[21] Anthony E. Brockwell and Joseph B. Kadane. A gridding method for Bayesian sequential decision problems. Journal of Computational and Graphical Statistics, 12(3):566–584, 2003.

[22] Philippe Brouillard, Sébastien Lachapelle, Alexandre Lacoste, Simon Lacoste-Julien, and Alexandre Drouin. Differentiable causal discovery from interventional data. Advances in Neural Information Processing Systems, 33:21865–21877, 2020.

[23] John Burkardt. The truncated normal distribution. Technical report, Florida State University, 2023. available at https://people.sc.fsu.edu/~jburkardt/presentations/truncated_normal.pdf.

[24] T. Butler, J. Jakeman, and T. Wildey. Combining push-forward measures and Bayes' rule to construct consistent solutions to stochastic inverse problems. SIAM Journal on Scientific Computing, 40(2):A984–A1011, 2018.

[25] T. Butler, J. D. Jakeman, and T. Wildey. Optimal experimental design for prediction based on push-forward probability measures. Journal of Computational Physics, 416:109518, 2020.

[26] Troy Butler, J. Jakeman, and Timothy Wildey. Convergence of probability densities using approximate models for forward and inverse problems in uncertainty quantification. SIAM Journal on Scientific Computing, 40(5):A3523–A3548, 2018.

[27] Bradley P. Carlin, Joseph B. Kadane, and Alan E. Gelfand. Approaches for optimal sequential decision analysis in clinical trials. Biometrics, 54(3):964–975, 1998.

[28] Daniel R. Cavagnaro, Jay I. Myung, Mark A. Pitt, and Janne V. Kujala. Adaptive design optimization: A mutual information-based approach to model discrimination in cognitive science. Neural Computation, 22(4):887–905, 2010.

[29] Atlanta Chakraborty, Xun Huan, and Tommie Catanach. A likelihood-free approach to goal-oriented Bayesian optimal experimental design. arXiv preprint arXiv:2408.09582, 2024.

[30] Kathryn Chaloner and Isabella Verdinelli. Bayesian Experimental Design: A Review. Statistical Science, 10(3):273–304, 1995.

[31] Xinyue Chen, Che Wang, Zijian Zhou, and Keith Ross. Randomized ensembled double Q-learning: Learning fast without a model. In 9th International Conference on Learning Representations (ICLR 2021), 2021. available at https://openreview.net/forum?id=AY8zfZm0tDd.

[32] Hyunghoon Cho, Bonnie Berger, and Jian Peng. Reconstructing causal biological networks through active learning. PloS One, 11(3):e0150611, 2016.

[33] J. Andrés Christen and Miguel Nakamura. Sequential stopping rules for species accumulation. Journal of Agricultural, Biological & Environmental Statistics, 8(2):184–195, 2003.

[34] Alex R Cook, Gavin J Gibson, and Christopher A Gilligan. Optimal observation times in experimental epidemic processes. Biometrics, 64(3):860–868, 2008.

[35] Haotian Cui, Chloe Wang, Hassaan Maan, Kuan Pang, Fengning Luo, Nan Duan, and Bo Wang. scGPT: toward building a foundation model for single-cell multi-omics using generative AI. Nature Methods, pages 1–11, 2024.

[36] Chris Cundy, Aditya Grover, and Stefano Ermon. Bcd nets: Scalable variational approaches for bayesian causal discovery. Advances in Neural Information Processing Systems, 34:7095–7110, 2021.

[37] Martijn de Jongh and Marek J Druzdzel. A comparison of structural distance measures for causal Bayesian network models. Recent Advances in Intelligent Information Systems, Challenging Problems of Science, Computer Science Series, pages 443–456, 2009.

[38] Morris H DeGroot. Optimal statistical decisions. John Wiley & Sons, 2005.

[39] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using Real NVP. arXiv preprint arXiv:1605.08803, 2016.

[40] Jiayuan Dong, Christian Jacobsen, Mehdi Khalloufi, Maryam Akram, Wanjiao Liu, Karthik Duraisamy, and Xun Huan. Variational bayesian optimal experimental design with normalizing flows. Computer Methods in Applied Mechanics and Engineering, 433:117457, 2025.

[41] Jiayuan Dong, Christian Jacobsen, Mehdi Khalloufi, Maryam Akram, Wanjiao Liu, Karthik Duraisamy, and Xun Huan. Variational Bayesian optimal experimental design with normalizing flows. Computer Methods in Applied Mechanics and Engineering, 433:117457, 2025.

[42] Arnaud Doucet, Nando Freitas, Kevin Murphy, and Stuart Russell. Sequential Monte Carlo Methods in Practice. Springer New York, 2013.

[43] Felix Draxler, Stefan Wahl, Christoph Schnörr, and Ullrich Köthe. On the universality of coupling-based normalizing flows. arXiv preprint arXiv:2402.06578, 2024.

[44] Hovav A. Dror and David M. Steinberg. Sequential experimental designs for generalized linear models. Journal of the American Statistical Association, 103(481):288–298, 2008.

[45] Christopher C. Drovandi, James M. McGree, and Anthony N. Pettitt. Sequential Monte Carlo for Bayesian sequentially designed experiments for discrete data. Computational Statistics & Data Analysis, 57(1):320–335, 2013.

[46] Christopher C. Drovandi, James M. McGree, and Anthony N. Pettitt. A sequential Monte Carlo algorithm to incorporate model uncertainty in Bayesian sequential design. Journal of Computational and Graphical Statistics, 23(1):3–24, 2014.

[47] Chi Feng and Youssef M. Marzouk. A layered multiple importance sampling scheme for focused optimal Bayesian experimental design. arXiv preprint arXiv:1903.11187, 2019.

[48] Alex Forster, Desi R. Ivanova, and Tom Rainforth. Improving robustness to model misspecification in bayesian experimental design. In 7th Symposium on Advances in Approximate Bayesian Inference – Workshop Track, 2025.

[49] Adam Foster, Desi R. Ivanova, Ilyas Malik, and Tom Rainforth. Deep adaptive design: Amortizing sequential Bayesian experimental design. In M. Meila and T. Zhang, editors, Proceedings of the 38th International Conference on Machine Learning (ICML 2021), volume 139 of Proceedings of Machine Learning Research, pages 3384–3395. PMLR, 2021.

[50] Adam Foster, Martin Jankowiak, Elias Bingham, Paul Horsfall, Yee Whye Teh, Thomas Rainforth, and Noah Goodman. Variational Bayesian optimal experimental design. In Advances in Neural Information Processing Systems 32, page 14036–14047, 2019.

[51] Adam Foster, Martin Jankowiak, Matthew O'Meara, Yee Whye Teh, and Tom Rainforth. A unified stochastic gradient approach to designing Bayesian-optimal experiments. In Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics, volume 108 of Proceedings of Machine Learning Research, pages 2959–2969. PMLR, 2020.

[52] Nir Friedman, Moises Goldszmidt, and Abraham Wyner. Data analysis with Bayesian networks: A bootstrap approach. arXiv preprint arXiv:1301.6695, 2013.

[53] Nir Friedman and Daphne Koller. Being Bayesian about network structure: A Bayesian approach to structure discovery in Bayesian networks. Machine Learning, 50:95–125, 2003.

[54] Scott Fujimoto, Herke van Hoof, and David Meger. Addressing function approximation error in actor-critic methods. 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 1587–1596. PMLR, 10–15 Jul 2018.

[55] Juan L Gamella and Christina Heinze-Deml. Active invariant causal prediction: Experiment selection through stability. Advances in Neural Information Processing Systems, 33:15464–15475, 2020.

[56] Heyang Gao, Zexu Sun, Hao Yang, and Xu Chen. Policy-based bayesian active causal discovery with deep reinforcement learning. In Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pages 839–850, 2024.

[57] Roland Gautier and Luc Pronzato. Adaptive control for sequential design. Discussiones Mathematicae Probability and Statistics, 20(1):97–113, 2000.

[58] Felix Gers, Jürgen Schmidhuber, and Fred Cummins. Learning to forget: continual prediction with LSTM. Neural Computation, 12:2451–2471, 10 2000.

[59] AmirEmad Ghassami, Saber Salehkaleybar, Negar Kiyavash, and Elias Bareinboim. Budgeted experiment design for causal structure learning. In International Conference on Machine Learning, pages 1724–1733. PMLR, 2018.

[60] Colin S Gillespie and Andrew Golightly. Bayesian inference for generalized stochastic population growth models with application to aphids. Journal of the Royal Statistical Society Series C: Applied Statistics, 59(2):341–357, 2010.

[61] Josep Ginebra. On the measure of the information in a statistical experiment. Bayesian Analysis, 2(1):167–212, 2007.

[62] Clark Glymour, Kun Zhang, and Peter Spirtes. Review of causal discovery methods based on graphical models. Frontiers in Genetics, 10:524, 2019.

[63] Jinwoo Go and Peng Chen. Sequential infinite-dimensional bayesian optimal experimental design with derivative-informed latent attention neural operator. Journal of Computational Physics, 532:113976, 2025.

[64] Alex Greenfield, Aviv Madar, Harry Ostrer, and Richard Bonneau. DREAM4: Combining genetic and dynamic information to identify biological networks and dynamical models. PloS one, 5(10):e13397, 2010.

[65] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. 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 1861–1870. PMLR, 10–15 Jul 2018.

[66] Markus Hainy, Christopher C Drovandi, and James M McGree. Likelihood-free extensions for Bayesian sequentially designed experiments. In J. Kunert, C. Müller, and A. Atkinson, editors, mODa 11: Advances in Model-Oriented Design and Analysis, Contributions to Statistics, pages 153–161. Springer, 2016.

[67] Seong Woo Han, Daniel Duy Vo, and Brielin C. Brown. Large-scale bayesian causal discovery with interventional data, 2025.

[68] Hado Hasselt. Double Q-learning. In J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, editors, Advances in Neural Information Processing Systems, volume 23. Curran Associates, Inc., 2010.

[69] Alain Hauser and Peter Bühlmann. Characterization and greedy learning of interventional Markov equivalence classes of directed acyclic graphs. The Journal of Machine Learning Research, 13(1):2409–2464, 2012.

[70] David Heckerman, Christopher Meek, and Gregory Cooper. A Bayesian approach to causal discovery. Innovations in Machine Learning: Theory and Applications, pages 1–28, 2006.

[71] Christina Heinze-Deml, Marloes H Maathuis, and Nicolai Meinshausen. Causal structure learning. Annual Review of Statistics and Its Application, 5:371–391, 2018.

[72] Tapio Helin, Youssef Marzouk, and Jose Rodrigo Rojo-Garcia. Bayesian optimal experimental design with wasserstein information criteria, 2025.

[73] Matthew D Hoffman, Andrew Gelman, et al. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research, 15(1):1593–1623, 2014.

[74] Xun Huan. Numerical approaches for sequential Bayesian optimal experimental design. PhD thesis, Massachusetts Institute of Technology, 2015.

[75] Xun Huan, Jayanth Jagalur, and Youssef Marzouk. Optimal experimental design: Formulations and computations. Acta Numerica, 33:715–840, 2024.

[76] Xun Huan and Youssef Marzouk. Gradient-based stochastic optimization methods in Bayesian experimental design. International Journal for Uncertainty Quantification, 4(6), 2014.

[77] Xun Huan and Youssef M Marzouk. Simulation-based optimal bayesian experimental design for nonlinear systems. Journal of Computational Physics, 232(1):288–317, 2013.

[78] Xun Huan and Youssef M. Marzouk. Sequential Bayesian optimal experimental design via approximate dynamic programming. 2016.

[79] Guido W Imbens. Causal inference in the social sciences. Annual Review of Statistics and Its Application, 11, 2024.

[80] Desi R Ivanova, Adam Foster, Steven Kleinegesse, Michael U Gutmann, and Thomas Rainforth. Implicit deep adaptive design: Policy-based experimental design without likelihoods. Advances in Neural Information Processing Systems 34, pages 25785–25798, 2021.

[81] Christian Jacobsen, Jiayuan Dong, Mehdi Khalloufi, Xun Huan, Karthik Duraisamy, Maryam Akram, and Wanjiao Liu. Enhancing dynamical system modeling through interpretable machine learning augmentations: a case study in cathodic electrophoretic deposition. arXiv preprint arXiv:2401.08414, 2024.

[82] Woojae Kim, Mark A. Pitt, Zhong-Lin Lu, Mark Steyvers, and Jay I. Myung. A hierarchical adaptive approach to optimal experimental design. Neural Computation, 26:2565–2492, 2014.

[83] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. 2014.

[84] Durk P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. In Advances in Neural Information Processing Systems 31, 2018.

[85] Steven Kleinegesse, Christopher Drovandi, and Michael U. Gutmann. Sequential Bayesian experimental design for implicit models via mutual information. Bayesian Analysis, 16(3):773–802, sep 2021.

[86] Steven Kleinegesse and Michael U Gutmann. Bayesian experimental design for implicit models by mutual information neural estimation. In Proceedings of the 37th International Conference on Machine Learning, volume 119 of Proceedings of Machine Learning Research, pages 5316–5326. PMLR, 2020.

[87] Steven Kleinegesse and Michael U Gutmann. Gradient-based Bayesian experimental design for implicit models using mutual information lower bounds. arXiv preprint arXiv:2105.04379, 2021.

[88] Ivan Kobyzev, Simon JD Prince, and Marcus A Brubaker. Normalizing flows: An introduction and review of current methods. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(11):3964–3979, 2020.

[89] Murat Kocaoglu, Alex Dimakis, and Sriram Vishwanath. Cost-optimal learning of causal graphs. In International Conference on Machine Learning, pages 1875–1884. PMLR, 2017.

[90] Murat Kocaoglu, Karthikeyan Shanmugam, and Elias Bareinboim. Experimental design for learning causal graphs with latent variables. Advances in Neural Information Processing Systems, 30, 2017.

[91] Jakob Kruse, Gianluca Detommaso, Ullrich Köthe, and Robert Scheichl. HINT: Hierarchical invertible neural transport for density estimation and Bayesian inference. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 35, pages 8191–8199, 2021.

[92] Timothy P Lillicrap, Jonathan J Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning. 2015.

[93] D. V. Lindley. On a measure of the information provided by an experiment. The Annals of Mathematical Statistics, 27(4):986–1005, 1956.

[94] Phillip Lippe, Taco Cohen, and Efstratios Gavves. Efficient neural causal discovery without acyclicity constraints. arXiv preprint arXiv:2107.10483, 2021.

[95] Qiang Liu and Dilin Wang. Stein variational gradient descent: A general purpose Bayesian inference algorithm. Advances in Neural Information Processing Systems, 29, 2016.

[96] Quan Long, Marco Scavino, Raúl Tempone, and Suojin Wang. Fast estimation of expected information gains for Bayesian experimental designs based on Laplace approximations. Computer Methods in Applied Mechanics and Engineering, 259:24–39, 2013.

[97] Lars Lorch, Jonas Rothfuss, Bernhard Schölkopf, and Andreas Krause. Dibs: Differentiable bayesian structure learning. Advances in Neural Information Processing Systems, 34:24111–24123, 2021.

[98] Lars Lorch, Scott Sussex, Jonas Rothfuss, Andreas Krause, and Bernhard Schölkopf. Amortized inference for causal structure learning. Advances in Neural Information Processing Systems, 35:13104–13118, 2022.

[99] Yuchen Ma, Dennis Frauen, Emil Javurek, and Stefan Feuerriegel. Foundation models for causal inference via prior-data fitted networks, 2025.

[100] Jean-Michel Marin, Pierre Pudlo, Christian P. Robert, and Robin J. Ryder. Approximate Bayesian computational methods. Statistics and Computing, 22:1167–1180, 2011.

[101] Youssef Marzouk, Tarek Moselhy, Matthew Parno, and Alessio Spantini. Sampling via measure transport: An introduction. In Handbook of Uncertainty Quantification, pages 1–41. Springer International Publishing, Cham, 2016.

[102] James Matis, Thomas Kiffe, Timothy Matis, and Douglass Stevenson. Nonlinear stochastic modeling of aphid population growth. Mathematical Biosciences, 198(2):148–168, 01 2006.

[103] Thomas P. Minka. Expectation Propagation for approximate Bayesian inference. In Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence (UAI 2001), pages 362–369, 2001.

[104] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. Nature, 518:529–533, 2015.

[105] Ehsan Mokhtarian, Saber Salehkaleybar, AmirEmad Ghassami, and Negar Kiyavash. A unified experiment design approach for cyclic and acyclic causal models. arXiv preprint arXiv:2205.10083, 2022.

[106] Peter Müller, Donald A. Berry, Andy P. Grieve, Michael Smith, and Michael Krams. Simulation-based sequential Bayesian design. Journal of Statistical Planning and Inference, 137(10):3140–3150, 2007.

[107] Peter Müller, Yunshan Duan, and Mauricio Garcia Tec. Simulation-based sequential design. Pharmaceutical Statistics, 21(4):729–739, jul 2022.

[108] Kevin P Murphy. Active learning of causal bayes net structure. Technical report, technical report, UC Berkeley, 2001.

[109] S. A. Murphy. Optimal dynamic treatment regimes. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 65(2):331–355, 2003.

[110] Robert O Ness, Karen Sachs, Parag Mallick, and Olga Vitek. A Bayesian active learning experimental design for inferring signaling networks. Journal of Computational Biology, 25(7):709–725, 2018.

[111] XuanLong Nguyen, Martin J. Wainwright, and Michael I. Jordan. Estimating divergence functionals and the likelihood ratio by convex risk minimization. IEEE Transactions on Information Theory, 56(11):5847–5861, 2010.

[112] Mateusz Olko, Micha l Zajac, Aleksandra Nowak, Nino Scherrer, Yashas Annadani, Stefan Bauer, Lukasz Kuciński, and Piotr Mi loś. Trust your \∇: Gradient-based intervention targeting for causal discovery. Advances in Neural Information Processing Systems, 36, 2024.

[113] Rafael Orozco, Felix J Herrmann, and Peng Chen. Probabilistic Bayesian optimal experimental design using conditional normalizing flows. arXiv preprint arXiv:2402.18337, 2024.

[114] Govinda Anantha Padmanabha and Nicholas Zabaras. Solving inverse problems using conditional invertible neural networks. Journal of Computational Physics, 433:110194, 2021.

[115] George Papamakarios, Eric T Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. Journal of Machine Learnine Research, 22(57):1–64, 2021.

[116] Judea Pearl. Causality. Cambridge University Press, 2009.

[117] Ronan Perry, Julius Von Kügelgen, and Bernhard Schölkopf. Causal discovery in heterogeneous environments under the sparse mechanism shift hypothesis. Advances in Neural Information Processing Systems, 35:10904–10917, 2022.

[118] Jonas Peters, Peter Bühlmann, and Nicolai Meinshausen. Causal inference by using invariant prediction: identification and confidence intervals. Journal of the Royal Statistical Society Series B: Statistical Methodology, 78(5):947–1012, 2016.

[119] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. Elements of causal inference: foundations and learning algorithms. The MIT Press, 2017.

[120] Ben Poole, Sherjil Ozair, Aaron Van Den Oord, Alex Alemi, and George Tucker. On variational bounds of mutual information. In Proceedings of the 36th International Conference on Machine Learning, volume 97 of Proceedings of Machine Learning Research, pages 5171–5180. PMLR, 09–15 Jun 2019.

[121] Ryan Prenger, Rafael Valle, and Bryan Catanzaro. Waveglow: a flow-based generative network for speech synthesis. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 3617–3621. IEEE, 2019.

[122] Luc Pronzato and Éric Thierry. Sequential experimental design and response optimisation. Statistical Methods and Applications, 11(3):277–292, 2002.

[123] Friedrich Pukelsheim. Optimal Design of Experiments. Society for Industrial and Applied Mathematics, 2006.

[124] Stefan T Radev, Ulf K Mertens, Andreas Voss, Lynton Ardizzone, and Ullrich Köthe. BayesFlow: learning complex stochastic models with invertible neural networks. IEEE Transactions on Neural Networks and Learning Systems, 33(4):1452–1466, 2020.

[125] Tom Rainforth, Rob Cornish, Hongseok Yang, Andrew Warrington, and Frank Wood. On nesting monte carlo estimators. In Proceedings of the 35th International Conference on Machine Learning, volume 80 of Proceedings of Machine Learning Research, pages 4267–4276. PMLR, 2018.

[126] Tom Rainforth, Adam Foster, Desi R Ivanova, and Freddie Bickford Smith. Modern Bayesian experimental design. Statistical Science, 39(1):100–114, 2024.

[127] Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In Proceedings of the 32nd International Conference on Machine Learning, volume 37 of Proceedings of Machine Learning Research, pages 1530–1538. PMLR, 07–09 Jul 2015.

[128] Elizabeth G. Ryan, Christopher C. Drovandi, James M. McGree, and Anthony N. Pettitt. A review of modern computational algorithms for Bayesian optimal design. International Statistical Review, 84(1):128–154, 2016.

[129] Kenneth J. Ryan. Estimating expected information gains for experimental designs with application to the random fatigue-limit model. Journal of Computational and Graphical Statistics, 12(3):585–603, 2003.

[130] John Schulman, Sergey Levine, Pieter Abbeel, Michael Jordan, and Philipp Moritz. Trust region policy optimization. In Francis Bach and David Blei, editors, Proceedings of the 32nd International Conference on Machine Learning, volume 37 of Proceedings of Machine Learning Research, pages 1889–1897. PMLR, 07–09 Jul 2015.

[131] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. 2017.

[132] Wanggang Shen, Jiayuan Dong, and Xun Huan. Variational sequential optimal experimental design using reinforcement learning. Computer Methods in Applied Mechanics and Engineering, (444):118068, 2025.

[133] Wanggang Shen, Jiayuan Dong, and Xun Huan. Variational sequential optimal experimental design using reinforcement learning. Computer Methods in Applied Mechanics and Engineering, 444:118068, 2025.

[134] Wanggang Shen and Xun Huan. Bayesian sequential optimal experimental design for nonlinear models using policy gradient reinforcement learning. Computer Methods in Applied Mechanics and Engineering, 416:116304, 2023.

[135] Michael E Sobel. Causal inference in the social sciences. Journal of the American Statistical Association, 95(450):647–651, 2000.

[136] Antti Solonen, Heikki Haario, and Marko Laine. Simulation-based optimal design using a response variance criterion. Journal of Computational and Graphical Statistics, 21(1):234–252, 2012.

[137] Yang Song, Conor Durkan, Iain Murray, and Stefano Ermon. Maximum likelihood training of score-based diffusion models. In Advances in Neural Information Processing Systems 34, pages 1415–1428, 2021.

[138] James C. Spall. Implementation of the simultaneous perturbation algorithm for stochastic optimization. IEEE Transactions on Aerospace and Electronic Systems, 34(3):817–823, 1998.

[139] Peter Spirtes, Clark N Glymour, and Richard Scheines. Causation, Prediction, and Search. The MIT press, 2000.

[140] Scott Sussex, Caroline Uhler, and Andreas Krause. Near-optimal multi-perturbation experimental design for causal structure learning. Advances in Neural Information Processing Systems, 34:777–788, 2021.

[141] Esteban G Tabak and Eric Vanden-Eijnden. Density estimation by dual ascent of the log-likelihood. Communications in Mathematical Sciences, 8(1):217–233, 2010.

[142] Roubing Tang, Sabina Sloman, and Samuel Kaski. Generalization analysis for bayesian optimal experiment design under model misspecification, 06 2025.

[143] Mauricio Tec, Yunshan Duan, and Peter Müller. A comparative tutorial of Bayesian sequential design and reinforcement learning. The American Statistician, 77(2):223–233, apr 2023.

[144] Alejandro Tejada-Lapuerta, Paul Bertin, Stefan Bauer, Hananeh Aliee, Yoshua Bengio, and Fabian J Theis. Causal machine learning for single-cell genomics. arXiv preprint arXiv:2310.14935, 2023.

[145] Christina V Theodoris, Ling Xiao, Anant Chopra, Mark D Chaffin, Zeina R Al Sayed, Matthew C Hill, Helene Mantineo, Elizabeth M Brydon, Zexian Zeng, X Shirley Liu, et al. Transfer learning enables predictions in network biology. Nature, 618(7965):616–624, 2023.

[146] Panagiotis Tigas, Yashas Annadani, Desi R Ivanova, Andrew Jesson, Yarin Gal, Adam Foster, and Stefan Bauer. Differentiable multi-target causal Bayesian experimental design. In International Conference on Machine Learning, pages 34263–34279. PMLR, 2023.

[147] Panagiotis Tigas, Yashas Annadani, Andrew Jesson, Bernhard Schölkopf, Yarin Gal, and Stefan Bauer. Interventions, where and how? Experimental design for causal models at scale. Advances in Neural Information Processing Systems, 35:24130–24143, 2022.

[148] Simon Tong and Daphne Koller. Active learning for structure in Bayesian networks. In International Joint Conference on Artificial Intelligence, volume 17, pages 863–869, 2001.

[149] Christian Toth, Lars Lorch, Christian Knoll, Andreas Krause, Franz Pernkopf, Robert Peharz, and Julius Von Kügelgen. Active Bayesian causal inference. Advances in Neural Information Processing Systems, 35:16261–16275, 2022.

[150] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748, 2018.

[151] Hal R Varian. Causal inference in economics and marketing. Proceedings of the National Academy of Sciences, 113(27):7310–7315, 2016.

[152] Thomas S Verma and Judea Pearl. Equivalence and synthesis of causal models. In Probabilistic and Causal Inference: The works of Judea Pearl, pages 221–236. 2022.

[153] Julius Vetter, Manuel Gloeckler, Daniel Gedon, and Jakob H. Macke. Effortless, simulation-efficient bayesian inference using tabular foundation models, 2025.

[154] Julius von Kügelgen, Paul K Rubenstein, Bernhard Schölkopf, and Adrian Weller. Optimal experimental design via Bayesian optimization: active causal structure learning for gaussian process networks. arXiv preprint arXiv:1910.03962, 2019.

[155] Udo Von Toussaint. Bayesian inference in physics. Reviews of Modern Physics, 83:943–999, 2011.

[156] Matthew J Vowels, Necati Cihan Camgoz, and Richard Bowden. D'ya like dags? a survey on structure learning and causal discovery. ACM Computing Surveys, 55(4):1–36, 2022.

[157] Xiangyu Wang, Taiyu Ban, Lyuzhou Chen, Derui Lyu, Qinrui Zhu, and Huanhuan Chen. Large-scale hierarchical causal discovery via weak prior knowledge. IEEE Transactions on Knowledge and Data Engineering, 37(5):2695–2711, 2025.

[158] J. Kyle Wathen and J. Andrés Christen. Implementation of backward induction for sequentially adaptive clinical trials. Journal of Computational and Graphical Statistics, 15(2):398–413, 2006.

[159] Christopher JCH Watkins and Peter Dayan. Q-learning. Machine learning, 8(3-4):279–292, 1992.

[160] Ronald J Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. Machine learning, 8(3):229–256, 1992.

[161] Keyi Wu, Peng Chen, and Omar Ghattas. An efficient method for goal-oriented linear Bayesian optimal experimental design: Application to optimal sensor placement. arXiv preprint arXiv:2102.06627, 2021.

[162] Keyi Wu, Peng Chen, and Omar Ghattas. A fast and scalable computational framework for large-scale high-dimensional Bayesian optimal experimental design. SIAM/ASA Journal on Uncertainty Quantification, 11(1):235–261, 2023.

[163] Keyi Wu, Peng Chen, and Omar Ghattas. An offline-online decomposition method for efficient linear Bayesian goal-oriented optimal experimental design: Application to optimal sensor placement. SIAM Journal on Scientific Computing, 45(1):B57–B77, 2023.

[164] Jing Yu, Victor M. Zavala, and Mihai Anitescu. A scalable design of experiments framework for optimal sensor placement. Journal of Process Control, 67:44–55, 2018.

[165] Zheyu Zhang, Jiayuan Dong, Jie Liu, and Xun Huan. Goal-oriented sequential bayesian experimental design for causal learning, 2025.

[166] Shijie Zhong, Wanggang Shen, Tommie Catanach, and Xun Huan. Goal-oriented Bayesian optimal experimental design for nonlinear models using Markov Chain Monte Carlo. arXiv preprint arXiv:2403.18072, 2024.