A Hybrid Gradient Method to Designing Bayesian Experiments for Implicit Models

Jiaxin Zhang[1*], Sirui Bi[2,3], Guannan Zhang[1]

[1]Computer Science and Mathematics Division, Oak Ridge National Laboratory

[2]Department of Civil and Systems Engineering, Johns Hopkins University, [3]Computational Sciences and Engineering Division, Oak Ridge National Laboratory,

*Email address: zhangj@ornl.gov

Introduction

Bayesian experimental design (BED) is to choose designs that maximize the information gathering. For implicit models, where the likelihood is intractable, but sampling is possible, conventional BED methods have difficulties in efficiently estimating the posterior distribution and maximizing the mutual information (MI) between data and parameters.

Recent work using gradient ascent to maximize a lower bound on MI was proposed to deal with the issues. However, the approach requires a sampling path to compute the pathwise gradient of the MI lower bound with respect to the design variables, and such a pathwise gradient is usually inaccessible for implicit models.

Novel contribution: **SAGABED** framework

- We propose a general unified framework that leverages **stochastic approximate gradient** without the requirement or assumption of pathwise gradients for implicit models.
- We introduce a **smoothed MI lower bound** to conduct robust MI estimation and optimization, which allows the **variance of the design and posterior distribution to be much smaller than existing approaches.**
- We show the superior performance of the approach through several experiments and demonstrate that the approach enables the optimization to be performed by stochastic gradient ascent algorithm and thus well scaled to **considerable high dimensional design problems.**

Bayesian experimental design (BED)

BED framework aims at choosing an experimental design ξ to maximize the information gained about some parameters of interest $\boldsymbol{\theta}$ from the outcome $\boldsymbol{y}$ of the experiment.

Expected information gain: $I(\xi) = \mathbb{E}_{p(y|\xi)}[\mathcal{Q}[p(\boldsymbol{\theta})] - \mathcal{Q}[p(\boldsymbol{\theta}|\boldsymbol{y},\xi)]]$

EIG can be interpreted as a mutual information between $\boldsymbol{\theta}$ and $\boldsymbol{y}$

$$I_{\mathrm{MI}}(\boldsymbol{\xi}) = \mathbb{E}_{p(\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\xi})}\left[\log \frac{p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\xi})}{p(\boldsymbol{y}|\boldsymbol{\xi})}\right]$$

MI estimate is the core

Bayesian optimal design is defined as $\boldsymbol{\xi}^* = \arg\max_{\boldsymbol{\xi}\in\Xi} I_{\mathrm{MI}}(\boldsymbol{\xi})$

Conventional approach: nested Monte Carlo (NMC) estimator,

$$I_{\mathrm{NMC}}(\boldsymbol{\xi}) = \frac{1}{N}\sum_{i=1}^{N} \log \frac{p(\boldsymbol{y}_i|\boldsymbol{\theta}_{i,0},\boldsymbol{\xi})}{\frac{1}{m}\sum_{j=1}^{M} p(\boldsymbol{y}_i|\boldsymbol{\theta}_{i,j},\boldsymbol{\xi})}$$

Computationally intensive!

Stochastic Approximate Gradient Ascent BED

Belghazi et al (2018) proposed to estimate MI by gradient ascent over neural networks and argued that the lower bound can be tightened by optimizing the neural network parameters, MINE

$$I_{\mathrm{MINE}}(\boldsymbol{\xi},\boldsymbol{\psi}) = \mathbb{E}_{p(\boldsymbol{\theta},\boldsymbol{y}|\boldsymbol{\xi})}[\mathcal{T}_{\psi}(\boldsymbol{\theta},\boldsymbol{y})] - \log \mathbb{E}_{p(\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\xi})}[e^{\mathcal{T}_{\psi}(\boldsymbol{\theta},\boldsymbol{y})}]$$

BED problem can be formulated by maximizing the overall objective

$$\boldsymbol{\xi}^* = \arg\max_{\boldsymbol{\xi}} \max_{\boldsymbol{\psi}} \{I_{\mathrm{MINE}}(\boldsymbol{\xi},\boldsymbol{\psi})\}.$$

Smoothed MI estimator: reduce the variance with clip operation

$$I_{\mathrm{SMILE}}(\boldsymbol{\xi},\boldsymbol{\psi}) = \mathbb{E}_{p(\boldsymbol{\theta},\boldsymbol{y}|\boldsymbol{\xi})}[\mathcal{T}_{\psi}(\boldsymbol{\theta},\boldsymbol{y})] - \log \mathbb{E}_{p(\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\xi})}[\mathrm{clip}(e^{\mathcal{T}_{\psi}(\boldsymbol{\theta},\boldsymbol{y})}, e^{-\tau}, e^{\tau})]$$

Stochastic gradient approximate: evolution strategies

$$f_{\sigma}(\boldsymbol{\xi}) = \mathbb{E}_{\epsilon\sim\mathcal{N}(0,\mathbf{I}_d)}[f(\boldsymbol{\xi}+\sigma\epsilon)]$$

$$\nabla f_{\sigma}(\boldsymbol{\xi}) = \frac{1}{\sigma}\mathbb{E}_{\epsilon\sim\mathcal{N}(0,\mathbf{I}_d)}[f(\boldsymbol{\xi}+\sigma\epsilon)\epsilon] \approx \frac{1}{M\sigma}\sum_{i=1}^{M} f(\boldsymbol{\xi}+\sigma\epsilon_i)\epsilon_i.$$

Guided ES algorithm

$$\Sigma = (\alpha/n)\cdot \mathbf{I}_n + (1-\alpha)/k \cdot UU^T$$

SAGABED for implicit models

- Unified framework vs. two-stage framework
- Scalability, portability, and parallelization
- Robust estimation with low variance

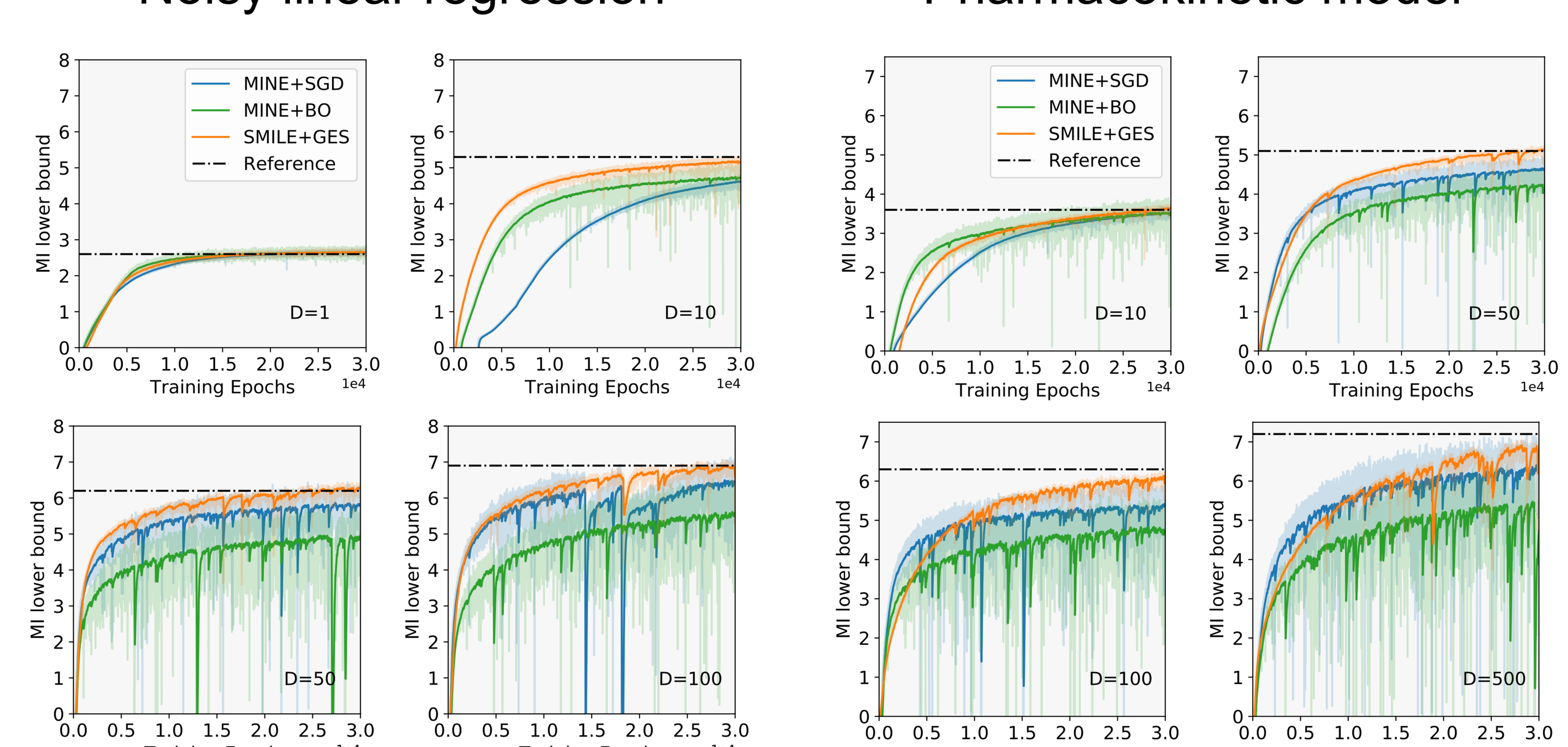
Algorithm 1: The `SAGABED` algorithm

1: **Require**: neural network architectures, learning rates ℓ_{ψ} and ℓ_{ξ}, τ in I_{SMILE}, total prior samples n, total iterations T, implicit model $\mathcal{M}$
2: **Process**:
3: Initialize a design $\boldsymbol{\xi}_0$ by random sampling
4: Initialize neural network parameter $\boldsymbol{\psi}_0$
5: **for** $t = 0 : T-1$ **do**
6: Draw n samples from the prior distribution of the model parameters $\boldsymbol{\theta}$: $\boldsymbol{\theta}^{(1)}, ..., \boldsymbol{\theta}^{(n)} \sim p(\boldsymbol{\theta})$
7: Compute the corresponding data samples $\boldsymbol{y}^{(i)}$, $i = 1, ..., n$ using the current design $\boldsymbol{\xi}_t$ and a implicit model $\mathcal{M}$
8: Evaluate the smoothed MI lower bound I_{SMILE} by Eq. (9) at the current design $\boldsymbol{\xi}_t$ and network parameters $\boldsymbol{\psi}_t$
9: Compute the approximate gradient estimator $\nabla_{\boldsymbol{\xi}} I_{\mathrm{SMILE}}(\boldsymbol{\xi}_t,\boldsymbol{\psi}_t)$ using the GES algorithm
10: Evaluate the gradient of the I_{SMILE} with respect to the network parameters $\nabla_{\boldsymbol{\psi}} I_{\mathrm{SMILE}}(\boldsymbol{\xi},\boldsymbol{\psi})$ using `PyTorch`
11: Update design $\boldsymbol{\xi}_t$ via gradient ascent: $\boldsymbol{\xi}_{t+1} = \boldsymbol{\xi}_t + \ell_{\xi}\nabla_{\boldsymbol{\xi}} I_{\mathrm{SMILE}}(\boldsymbol{\xi}_t,\boldsymbol{\psi}_t)$
12: Update neural network parameters $\boldsymbol{\psi}_t$ via gradient ascent: $\boldsymbol{\psi}_{t+1} = \boldsymbol{\psi}_t + \ell_{\psi}\nabla_{\boldsymbol{\psi}} I_{\mathrm{SMILE}}(\boldsymbol{\xi}_t,\boldsymbol{\psi}_t)$
13: **end for**

Experiments: toy examples

Noisy linear regression

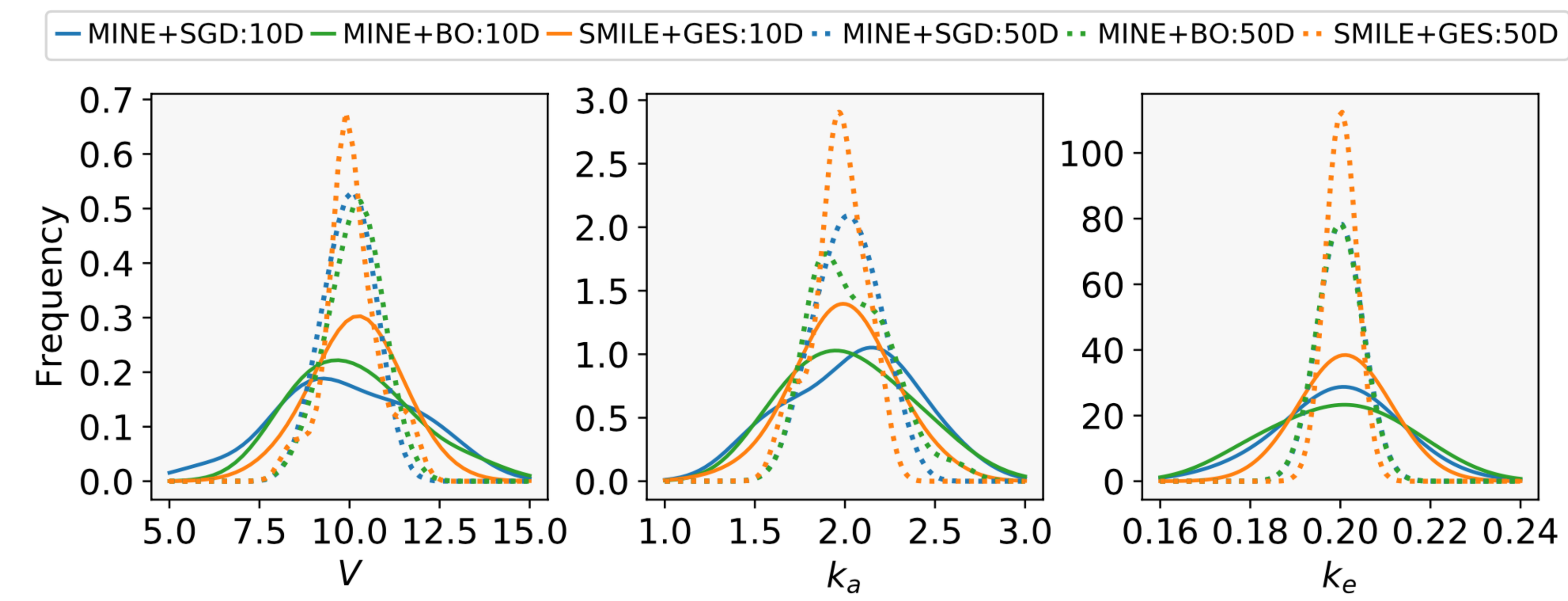
Pharmacokinetic model



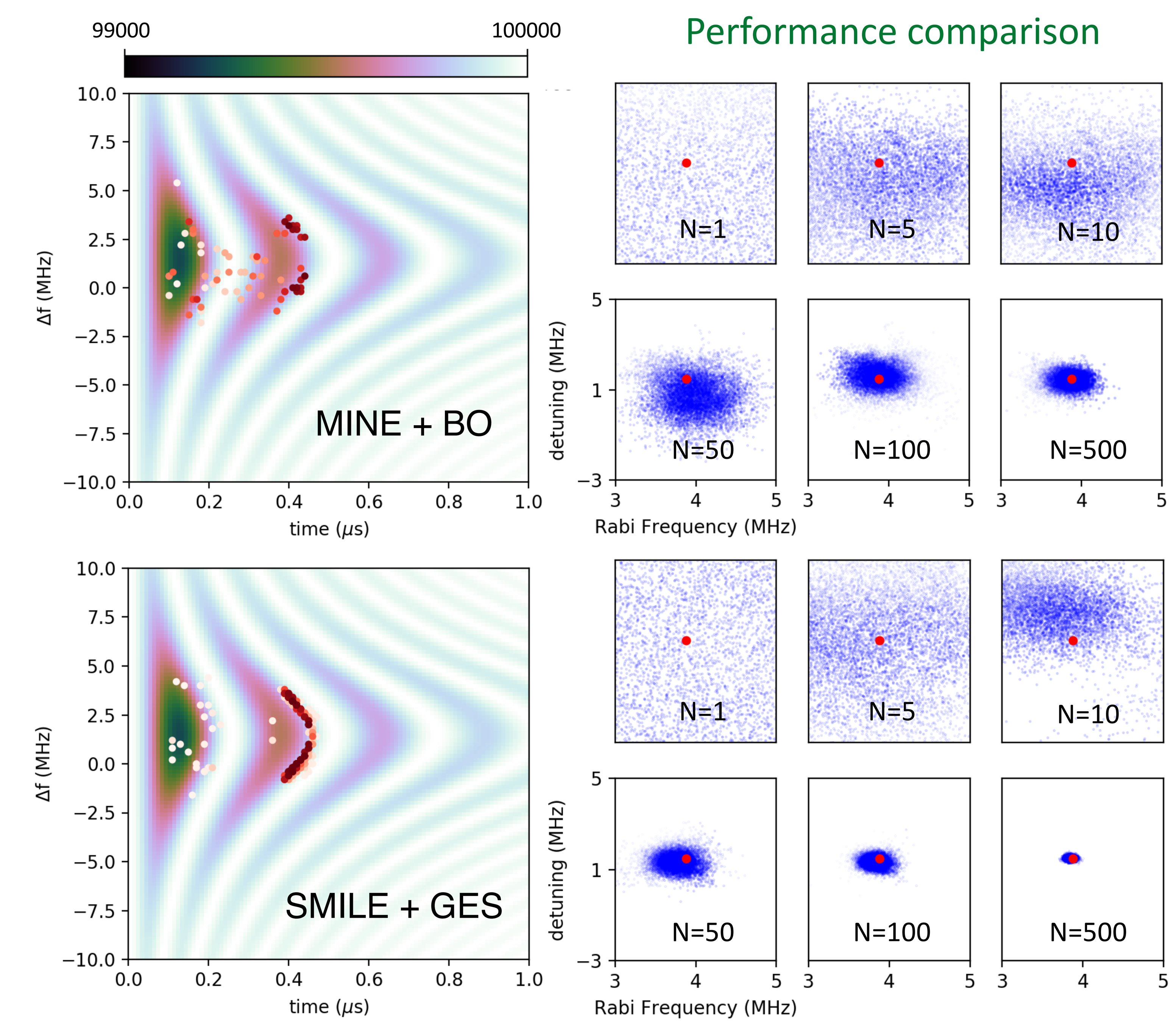
Statistical analysis of posterior samples

Table 1: Estimating mean and standard deviation of the posterior samples of the model parameters $\boldsymbol{\theta}$ using optimal designs $\boldsymbol{d}^*$ and real data observation $\boldsymbol{y}^*$ (use $\boldsymbol{\theta}_{\mathrm{true}} = [1,4]$ to generate $\boldsymbol{y}^*$)

Method	D=1		D=10		D=50		D=100	
	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$
MINE + SGD	-1.39±2.54	6.03±**0.93**	0.51±**0.44**	2.99±0.67	1.20±0.18	3.79±0.23	0.97±0.05	4.04±0.04
MINE + BO	-1.42±**0.81**	2.98±1.19	1.22±0.58	4.93±0.91	0.71±0.25	3.66±0.40	1.35±0.21	4.79±0.26
SMILE + GES	2.76±1.36	5.74±3.08	0.83±0.56	4.69±**0.58**	1.11±**0.13**	4.25±**0.19**	1.02±**0.04**	3.98±**0.03**



Real-world application: tuning for quantum control



Conclusion and future work

Develop a general unified framework that utilizes the stochastic approximate gradient for BED with implicit models. Our approach allows to scaled to substantial high dimensional design problems. **The future work** will focus on the extension of our proposed framework to sequential Bayesian optimization design (SBED).