Numerical Study of Gradient-Free Sampling Methods for Bayesian Inversion Problems

Yantong Xie*, Xuda Ye†

January 15, 2022

1 Introduction

Consider the inverse problem of finding the parameter $\theta \in \mathbb{R}^d$ from the observation $y \in \mathbb{R}^K$

$$y = G(\theta) + \eta, \tag{1}$$

where $G(\cdot) : \mathbb{R}^d \mapsto \mathbb{R}^K$ is the deterministic forward mapping and $\eta \sim \mathcal{N}(0, \Gamma)$ is the Gaussian noise. We adopt the Bayesian approach to the inversion problem [7] and assume that the prior distribution of the parameter θ is $\mathcal{N}(0, \Sigma)$ and independent of the noise η. By Bayes' Lemma, for given data $y \in \mathbb{R}^K$, the posterior distribution of θ is

$$\rho(\theta) \propto \exp(-f(\theta)), \tag{2}$$

where the potential function $f(\theta)$ is given by

$$f(\theta) = \frac{1}{2}|y - G(\theta)|_\Gamma^2 + \frac{1}{2}|\theta|_\Sigma^2. \tag{3}$$

For the sake of readers, we summarize the symbol in Table 1. Note that we use the following short hand notations: for a positive definite matrix A, we let

$$\langle x, y\rangle_A = x^{\mathrm{T}} A^{-1} y, \qquad |x|_A = \sqrt{x^{\mathrm{T}} A^{-1} x} \tag{4}$$

Our goal is to approximate the posterior distribution $\rho(\theta)$.

symbol	**dimension**	**meaning**
θ	d	parameter
η	K	noise of observation
y	K	observed data
Γ	$K \times K$	covariance of Gaussian noise
Σ	$d \times d$	covariance of Gaussian prior

Table 1: Symbol list of the inverse problem.

*Student ID: 1901110045. He writes the whole report except for Section 3.
†Student ID: 2001110036. He writes the codes and Section 3 of the report.

Many inverse problems arising in practical applications are involved with complex forward model G, some of which are only available as a black box. The derivatives of G are often unavailable or costly to obtain, hence effective sampling methods of the posterior distribution ρ without computing the gradient of G are highly in need. On the other hand, inspired by the Metropolis algorithm and various MCMC and SMC algorithms, a wide range of sampling methods that use a collective dynamics of particles have been brought into focus. There are two classes of methods emerging in this context: those arising form ensemble Kalman inversion (EKI) and those arising from consensus forming dynamics. Recently two novel gradient-free sampling algorithms using the two classes of dynamics of particles have been proposed respectively: the ensemble Kalman sampling (EKS) [6] based on the EKI and the consensus bases sampling (CBS) based on consensus forming dynamics [2]. Both algorithms show great potential in sampling the posterior, however, there is little literature on the choice of the ensemble size of the particles nor the comparison between the two algorithms. In this project, we aim to proceed numerical discussion on the algorithm parameters and compare the numerical results of the two sampling algorithms. With the numerical exploration on the EKS and CBS, we hope to enhance our understanding on the two algorithms treading different paths but leading to the same destination.

The rest of this report is outlined as follows: in Section 2, we introduced the EKS and the CBS and their associated dynamics of particle systems. In Section 3, we present and analyze the numerical results of the EKS and CBS through experiments.

2 Gradient-Free Sampling Methods

2.1 Ensemble Kalman sampler

The overdamped Langevin dynamics is a single particle dynamics in $\mathbb{R}^d$:

$$\dot{\theta} = -C\nabla f(\theta) + \sqrt{2C}\dot{W}, \tag{5}$$

where $W \in \mathbb{R}^d$ is a standard Wiener process. The symmetric matrix C is introduced as a preconditioning parameter to speed up the convergence. We remark that Equation (5) transforms an arbitrary distribution into the desired posterior distribution over an infinite time horizon.

Given the particle system $\{\theta^{(j)}\}_{j=1}^J$, it is natural to evolve the system according to the following SDEs:

$$\dot{\theta}^{(j)} = -\mathcal{C}(\theta)\nabla f(\theta) + \sqrt{2\mathcal{C}(\theta)}\dot{W}^{(j)}, \tag{6}$$

where $\mathcal{C}(\theta)$ is chosen to be the empirical covariance of the particles

$$\mathcal{C}(\theta) = \frac{1}{J}\sum_{k=1}^{J}\left(\theta^{(k)} - \bar{\theta}\right) \otimes \left(\theta^{(k)} - \bar{\theta}\right) \in \mathbb{R}^{d\times d}, \quad \bar{\theta} = \frac{1}{J}\sum_{k=1}^{J}\theta^{(k)}. \tag{7}$$

Now we combine Equation (6) with the gradient of the potential function f, one arrives at

$$\dot{\theta}^{(j)} = -\frac{1}{J}\sum_{k=1}^{J}\langle \nabla G(\theta^{(j)})(\theta^{(j)} - \bar{\theta}), G(\theta^{(j)}) - y\rangle_\Gamma\, \theta^{(k)} - \mathcal{C}(\theta)\Sigma^{-1}\theta^{(j)} + \sqrt{2\mathcal{C}(\theta)}\dot{W}^{(j)} \tag{8}$$

where $\{W^{(j)}\}_{j=1}^J$ are independent Wiener processes in $\mathbb{R}^d$. In order to obtain a gradient-free algorithm, a common technique is to approximate the gradient by difference:

$$\nabla G(\theta^{(j)})(\theta^{(j)} - \bar{\theta}) \approx G(\theta^{(j)}) - \bar{G}, \quad \bar{G} = \frac{1}{J}\sum_{k=1}^{J} G(\theta^{(k)}). \tag{9}$$

Combine Equation (9) with Equation (8), we have

$$\dot{\theta}^{(j)} = -\frac{1}{J}\sum_{k=1}^{J}\langle G(\theta^{(k)}) - \bar{G}, G(\theta^{(j)}) - y\rangle_\Gamma \, \theta^{(k)} - \mathcal{C}(\theta)\Sigma^{-1}\theta^{(j)} + \sqrt{2\mathcal{C}(\theta)}\dot{W}^{(j)} \tag{10}$$

Equation (10) gives the **EKS dynamics**, which ideally transforms an arbitrary initial distribution of particle system into the desired posterior distribution over an infinite horizon. (10) can be directly integrated in the Euler-Maruyama scheme to produce the discrete-time dynamics.

Another prospective to understand the EKS dynamics is through the EKI [5], which is based on evolving a noisy set of interacting particles given by:

$$\dot{\theta}^{(j)} = -\frac{1}{J}\sum_{k=1}^{J}\langle G(\theta^{(k)}) - \bar{G}, G(\theta^{(j)}) - y\rangle_\Gamma \, \theta^{(k)} - \mathcal{C}^p(\theta)\Sigma^{-\frac{1}{2}}\dot{W}^{(j)}, \tag{11}$$

where $\mathcal{C}^p(\theta)$ denotes the empirical covariance matrix of the ensemble particles:

$$\mathcal{C}^p(\theta) = \frac{1}{J}\sum_{k=1}^{J}\left(G^{(k)} - \bar{G}\right) \otimes \left(\theta^{(k)} - \bar{\theta}\right) \in \mathbb{R}^{d\times d}. \tag{12}$$

Due to the drift term in Equation (11), the particles interact and acts to both drive towards consensus and fit the data y. By adding noise term in a manner of Equation (11), an astonishing feature of the EKI lies in that, in the linear setting, the dynamics (11) transforms the prior samples into the posterior samples in exactly one time unit [3]. Moreover, the resulting equations for the mean and covariance of Equation (11) can be viewed as Monte Carlo approximations of the Kalman filter [6]. In this context, the EKS dynamical system (10) can be viewed a modification of the standard EKI system (11) with a disparate noisy term, hence the EKS system possesses completely different properties.

2.2 Consensus-Based sampling

The consensus-based dynamics originates from the Cucker-Smale dynamics for opinion formation [4] as an optimization method. These dynamical particle systems describes the tendency of the particles to concentrate in certain variables modeling certain average properties, which can be view as consensus in opinion. To be more precise, the particles may exchange position information as they evolve and redirect towards a consensus position that is an average with respect to a Gibbs distribution [8]. Through carefully determining the noise term and the system parameters, the CBS is developed from the standard consensus-based optimization method in a manner similar with the development from the EKI to the EKS [2].

Given the parameter $\beta > 0$, consider the following McKean type SDE:

$$\dot{\theta} = -(\theta - \mathcal{M}_\beta(\theta)) + \sqrt{2(1+\beta)\mathcal{C}_\beta(\theta)}\dot{W}, \tag{13}$$

where the term $\mathcal{M}_\beta(\theta)$ and $\mathcal{C}_\beta(\theta)$ denote the mean and covariance of θ with respect to the Gibbs measure associated with the potential function f, i.e. $\frac{1}{Z}e^{-f(\theta)}$. Given the particle system $\{\theta^{(j)}\}_{j=1}^J$, the consensus-based particle system is given by

$$\dot{\theta}^{(j)} = -(\theta^{(j)} - \mathcal{M}_\beta(\theta)) + \sqrt{2(1+\beta)\mathcal{C}_\beta(\theta)}\dot{W}^{(j)} \tag{14}$$

where $\mathcal{M}_\beta(\theta)$ and $\mathcal{C}_\beta(\theta)$ are modified as the mean and covariance of the weighted empirical measure in $\mathbb{R}^d$, i.e.

$$\mathcal{M}_\beta(\theta) = \frac{\sum_{j=1}^J \omega^{(j)}\theta^{(j)}}{\sum_{j=1}^J \omega^{(j)}}, \quad \mathcal{C}_\beta(\theta) = \frac{\sum_{j=1}^J \omega^{(j)}\big(\theta^{(j)} - \mathcal{M}_\beta(\rho)\big)\big(\theta^{(j)} - \mathcal{M}_\beta(\rho)\big)^{\mathrm{T}}}{\sum_{j=1}^J \omega^{(j)}}. \tag{15}$$

The discrete version of CBS (14) is given by [2]

$$\theta_{n+1}^{(j)} = \mathcal{M}_\beta(\theta_n) + \alpha(\theta_n^{(j)} - \mathcal{M}_\beta(\theta_n)) + \sqrt{(1-\alpha^2)(1+\beta)\mathcal{C}_\beta(\theta_n)}\dot{W}^{(j)} \tag{16}$$

where $\alpha \in [0,1)$ is a constant. $\tau = 1-\alpha$ can be interpreted as the time step, and the discretized CBS (16) can be rewritten as

$$\theta_{n+1}^{(j)} = (1-\tau)\theta_n^{(j)} + \tau\mathcal{M}_\beta(\theta_n) + \sqrt{\tau(2-\tau)(1+\beta)\mathcal{C}_\beta(\theta_n)}\dot{W}^{(j)} \tag{17}$$

which is slightly different from the standard Euler-Maruyama scheme.

2.3 Complexity analysis

Both EKS and CBS relies on the evolution of the interacting particle system $\{\theta^{(j)}\}_{j=1}^J$. For these gradient-free sampling methods, the complexity in a single time step mainly comes from

1. Evaluation of the forward mappings $G(\theta^{(j)})$ for $j = 1, \cdots, J$;
2. Algebraic operations in computing the drift and diffusion terms.

In both EKS and CBS, the algebraic operations only involve matrix and vector multiplications, and the complexity is of order $\mathcal{O}(J)$. In some applications, the evaluation of $G(\theta)$ can be extremely costly, for example, $G(\theta)$ can be the numerical solution to a given PDE. In these cases, the calculation of $G(\theta^{(j)})$ dominates the complexity of the sampling methods, and we count the complexity by the number of times that $G(\theta)$ is evaluated.

3 Numerical experiments

3.1 Assessment of numerical efficiency

We test EKS and CBS on several typical inverse problems to compare their numerical performance. In order to assess the numerical efficiency of the sampling method, suppose $\{\theta_n\}_{n\geqslant 1}$ is the sequence of parameters produced by the sampling method, and we expect the empirical distribution

$$\hat{\rho}_N(\theta) := \frac{1}{N}\sum_{n=1}^N \delta(\theta - \theta_n) \tag{18}$$

to approximate the true posterior $\rho(\theta)$ as the number of samples $N \to \infty$. To measure the difference between $\hat{\rho}_N(\theta)$ and $\rho(\theta)$, introduce a series of test functions $\{A_s(\theta)\}_{s=1}^S$, and the error in the s-th test function $A_s(\theta)$ is defined as

$$e_s = \int_{\mathbb{R}^d} A_s(\theta)\hat{\rho}_N(\theta)\mathrm{d}\theta - \int_{\mathbb{R}^d} A_s(\theta)\rho(\theta)\mathrm{d}\theta, \quad s = 1, \cdots, S, \tag{19}$$

and the overall error of the empirical distribution $\hat{\rho}_N(\theta)$ is defined as

$$e[\hat{\rho}_N] := \left(\sum_{s=1}^S |e_s|^2\right)^{\frac{1}{2}}. \tag{20}$$

The asymptotic behavior of $e[\hat{\rho}_N]$ as $N \to \infty$ can then be used to assess the numerical efficiency of the sampling methods, from the following two aspects:

1. The limit of $e[\hat{\rho}_N]$ reveals the bias of the invariant distribution of the sampling method.
2. The decay rate of $e[\hat{\rho}_N]$ shows how fast the empirical distribution $\hat{\rho}_N$ converges.

That is, a good sampling scheme should have both small bias and fast convergence rate.

3.2 Low dimension: inverse problem for the elliptic equation

We adapt the example of elliptic equation from [2, 6]:

$$-\frac{\mathrm{d}}{\mathrm{d}x}\left(\exp(\theta_1)\frac{\mathrm{d}}{\mathrm{d}x}p(x)\right) = 1, \quad x \in [0,1] \tag{21}$$

with boundary values $p(0) = 0$ and $p(1) = \theta_2$, where $\theta = (\theta_1, \theta_2)$ is the parameter of the equation. The solution to the equation can be explicitly obtained as

$$p(x;\theta) = \theta_2 x + \exp(-\theta_1)\Big(-\frac{x^2}{2} + \frac{x}{2}\Big), \quad x \in [0,1], \tag{22}$$

and the forward mapping is defined by

$$G(\theta) = \begin{bmatrix} p(0.25;\theta) \\ p(0.75;\theta) \end{bmatrix}. \tag{23}$$

The Gaussian noise $\eta \sim \mathcal{N}(0,\Gamma)$ with $\Gamma = 0.1^2 I_2$ and the prior distribution is $\mathcal{N}(0,\Sigma)$ with $\Sigma = 10^2 I_2$. Our goal is to infer the parameter $\theta \in \mathbb{R}^2$ from the solution $p(x,\theta)$.

We employ EKS and CBS to generate the sequence $\{\theta_n\}_{n\geqslant 1}$ to approximate the posterior distribution. To measure the error of $\hat{\rho}_N$, introduce three test functions

$$\begin{aligned} A_1(\theta) &= \exp(-4(\theta_1 + 2.6)^2), \\ A_2(\theta) &= \exp(-2(\theta_2 - 104.5)^2), \\ A_3(\theta) &= \tanh(4(\theta_1 + 2.6)(\theta_2 - 104.5)), \end{aligned}$$

and define the error $e[\hat{\rho}_N]$ as in (20). In this low-dimensional example, the reference values of

$$\int_{\mathbb{R}^2} A_s(\theta)\rho(\theta)\mathrm{d}\theta, \quad s = 1,2,3 \tag{24}$$

are computed from numerical integration.

As an illustration, the particle systems $\{\theta^{(j)}\}_{j=1}^J$ generated by EKS and CBS are shown in Figure 1. Recall that we approximate the posterior $\rho(\theta)$ using the time average calculated from $\{\theta_n\}_{n\geqslant 1}$ rather than the particle system $\{\theta^{(j)}\}_{j=1}^J$ at some instant.

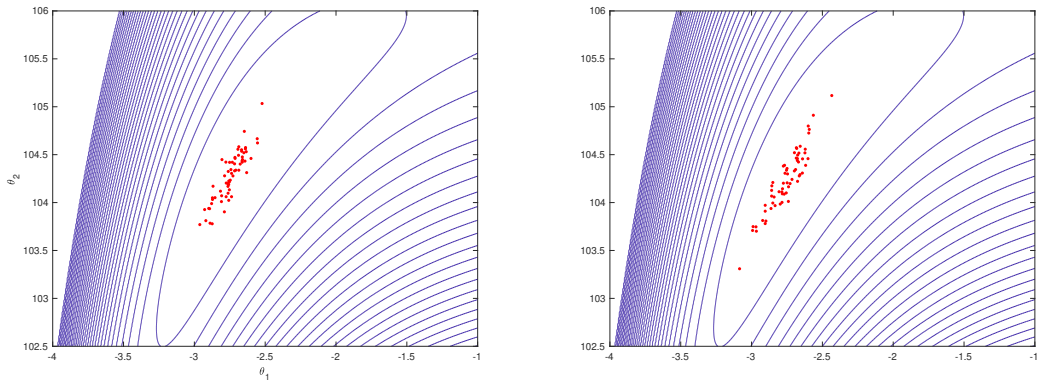

Figure 1: The particle system $\{\theta^{(j)}\}_{j=1}^J$ generated by EKS (left) and CBS (right) with the number of particles $J = 64$. The contour line of the potential function $f(\theta)$ is shown in the background.

In the following experiments, the parameter β is fixed at 2 in CBS, and the number of particles J varies in $8, 16, 32, 64$. The time step τ is chosen as $2^{-11}, 2^{-12}$ in EKS and $2^{-8}, 2^{-9}$ in CBS. Choosing smaller time steps for numerical integration results in instability. We plot the overall error $e[\hat{\rho}_N]$ as the function of complexity with different choices of parameters, as shown in Figure 2 and 3:

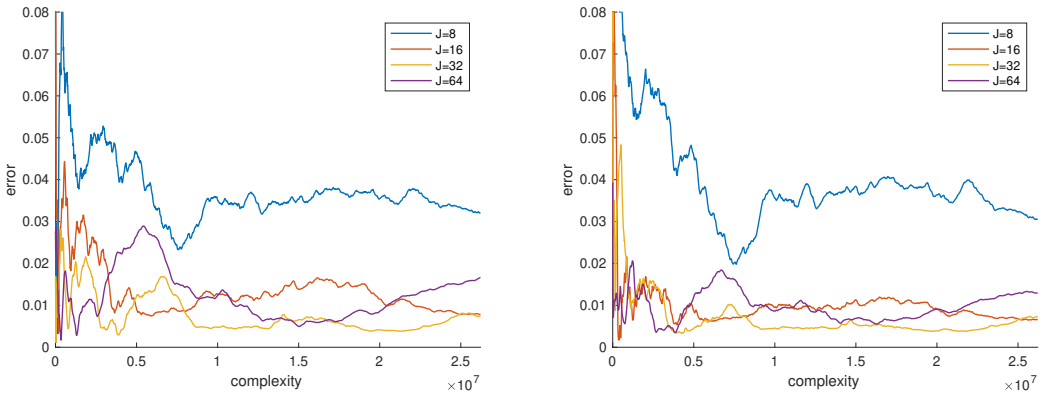

Figure 2: Overall error $e[\hat{\rho}_N]$ of EKS as the function of complexity with different numbers of particles $J = 8, 16, 32, 64$. The time step $\tau = 2^{-11}$ and 2^{-12} in the left and right figures.

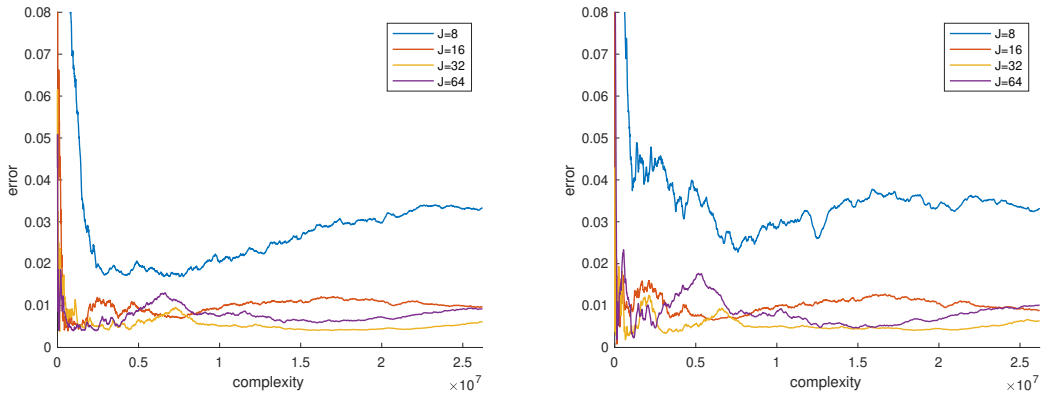

Figure 3: Overall error $e[\hat{\rho}_N]$ of CBS as the function of complexity with different numbers of particles $J = 8, 16, 32, 64$. The time step $\tau = 2^{-8}$ and 2^{-9} in the left and right figures.

From the evolution of the error $e[\hat{\rho}_N]$ in EKS and CBS, we conclude the following results:

1. EKS and CBS have similar bias in their numerical invariant distributions.

   For any choice of J, the error $e[\hat{\rho}_N]$ tends to be a stable value as $N \to \infty$, and this stable value reveals how the numerical invariant distributions of EKS and CBS deviate from the true posterior. In this example, choosing the number of particles $J = 16$ is sufficient keep the bias smaller than 0.02. Also, larger J does not enhance the accuracy of EKS or CBS evidently, which is not surprising since neither EKS or CBS can fully capture the non-Gaussian posterior.

2. The stability of EKS and CBS dynamics largely relies on the time step τ. If EKS or CBS has been stable, reducing the time step τ does not improve the accuracy evidently.

   This result indicates that the error due to numerical discretization does not dominate in the overall error $e[\hat{\rho}_N]$. Therefore, the choice of the time step τ only needs to ensure the stability of the dynamics.

3. In this example, CBS converges faster than EKS.

   From the initial stage of the evolution of $e[\hat{\rho}_N]$, it can be observed that the fluctuation of $e[\hat{\rho}_N]$ in CBS is much smaller than that of EKS. As indicated in [2], in the sense of mean-field limit, the convergence rate of CBS to its numerical invariant distribution can be explicitly obtained, while such result is not verified for EKS yet.

3.3 High dimension: inverse problem for the heat equation

Consider the heat equation with Neumann boundary conditions:

$$\begin{cases} u_t = u_{xx} + f(u), & x \in (0,\pi), \quad t \geqslant 0 \\ u_x(0,t) = u_x(\pi,t) = 0, & t \geqslant 0 \\ u(x,0) = u_0(x), & x \in (0,\pi) \end{cases} \tag{25}$$

where $f(u) = u\exp(-u^2)$ is a nonlinear diffusion term. Decompose $u_0(x)$ as Fourier cosine waves:

$$u_0(x) = \sum_{l=1}^{d} \theta_l \cos(l-1)x, \qquad x \in (0, \pi) \tag{26}$$

then $u_0(x)$ is parameterized by $\theta = \{\theta_k\}_{k=1}^d$. Given $T > 0$, define the forward mapping

$$G(\theta) = \begin{bmatrix} u(x_1, T) \\ \vdots \\ u(x_K, T) \end{bmatrix} \in \mathbb{R}^K \tag{27}$$

where $u(x,t)$ is the solution to the heat equation and $x_1, \cdots, x_K$ are grid points in $[-\pi, \pi]$.

Our goal is to infer the initial value $u_0(x)$ (parameter θ) from the solution $u(x,T)$. The parameters in the inverse problem for the heat equation is chosen as

- $d = 8$, $K = 20$;
- prior: $\theta_l \sim \mathcal{N}(0, 4/(l+1)^2)$, $l = 1, \cdots, d$;
- grid points: $x_k = (k-1)\pi/(K-1)$, $k = 1, \cdots, K$;
- observation noise: $\eta_k \sim \mathcal{N}(0, 0.1^2)$, $k = 1, \cdots, K$;
- evolution time: $T = 0.2$;
- exact solution: $\theta^* = (0.1, -0.3, 0.2, 0.1, -0.15, 0.02, 0, 0.01)$;
- observed data: $y = G(\theta^*)$.

In the paramater setting above, the initial value $u_0(x)$ corresponding to the exact solution θ^* and the solution $u(x,T)$ are shown in Figure 4:

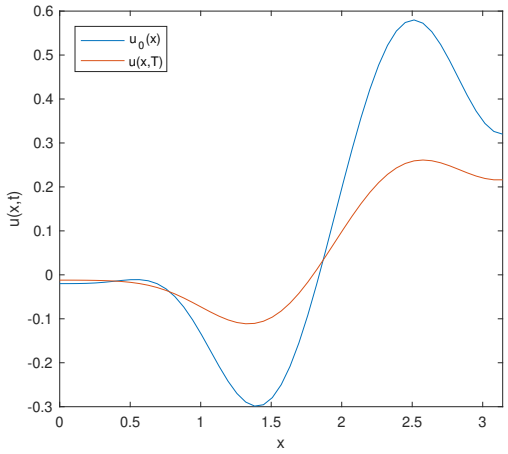

Figure 4: The initial value $u_0(x)$ and the solution $u(x,T)$ to the heat equation.

Given the values $u(x,T)$ at the grid points $\{x_k\}_{k=1}^K$, our goal is to recover the parameter θ in the initial value $u_0(x)$. Here are some other comments on this inverse problem:

1. Although it is well-known that the inverse problem for the heat equation is ill-posed (see [1] for example), we may impose a prior knowledge on the initial value $u_0(x)$ to force the inverse problem to be well-posed.

2. The larger the evolution time T is, the more difficult the inverse problem will be. This is due to the diffusive behavior of the solution.

3. If the source term $f(u) = 0$, then $G(\theta)$ will be a simple linear mapping. By choosing a nonlinear $f(u)$, $G(\theta)$ is a small perturbation of the linear mapping, thus we may expect the posterior is approximately Gaussian.

To approximate the posterior distribution, we choose the parameters in sampling methods as follows:

- EKS: number of particles $J = 32$, time step $\tau = 2^{-9}$, simulation time $T = 1000$;
- CBS: number of particles $J = 32$, time step $\tau = 2^{-5}$, simulation time $T = 16000$, $\beta = 2$.

The time steps τ are chosen to ensure the stability of the dynamics. The mean and standard deviation (std) of the posterior distribution obtained by EKS and CBS are shown in Table 2:

parameter	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8
exact	0.1	-0.3	0.2	0.1	-0.15	0.02	0	0.01
mean (EKS)	0.0893	-0.2986	0.1952	0.0945	-0.1377	0.0092	0.0344	0.0284
std (EKS)	0.0419	0.0585	0.0744	0.0951	0.1229	0.1415	0.1232	0.1895
mean (CBS)	0.0918	-0.3003	0.1983	0.0746	-0.1519	0.0847	-0.0215	0.0074
std (CBS)	0.0330	0.0286	0.0475	0.0884	0.1008	0.1115	0.1427	0.1121

Table 2: Mean and standard deviation (std) obtained by EKS and CBS

Table 2 shows that numerical performances of EKS and CBS are both satisfactory, while CBS is slightly better than EKS in the first three components of θ. For both methods, θ_l with larger index l has larger variance, because it is in general difficult to extract high dimensional information from the solution to the heat equation.

4 Conclusion

In this project we compare the numerical performance of two gradient-free sampling methods: EKS and CBS. Both methods rely on the evolution of an interacting particle system, where each particle is a duplicate of the original model. In the numerical experiments, we test EKS and CBS with different choices of the number of particles J and time steps τ. Both methods require J to be large enough ($J = 16$ in our numerical experiments) to have an accurate estimation of the posterior distribution. Error due to time discretization does not dominant in the overall error, which implies the time step τ only needs to ensure the stability of the dynamics. In general, EKS requires smaller time steps than CBS, and CBS has better accuracy in our numerical experiments. It is still an open question how to characterize the error of these gradient-free sampling methods for nonlinear models.

References

[1] Beck, J. V., Blackwell, B., & Clair Jr, C. R. S. (1985). Inverse heat conduction: Ill-posed problems. James Beck.

[2] Carrillo, J. A., Hoffmann, F., Stuart, A. M., Vaes, U. (2021). Consensus Based Sampling. arXiv preprint arXiv:2106.02519.

[3] Chen, Y., Oliver, D. S. (2012). Ensemble randomized maximum likelihood method as an iterative ensemble smoother. Mathematical Geosciences, 44(1), 1-26.

[4] Cucker, F., Smale, S. (2007). On the mathematics of emergence. Japanese Journal of Mathematics, 2(1), 197-227.

[5] Emerick, A. A., Reynolds, A. C. (2013). Investigation of the sampling performance of ensemble-based methods with a simple reservoir model. Computational Geosciences, 17(2), 325-350.

[6] Garbuno-Inigo, A., Hoffmann, F., Li, W., Stuart, A. M. (2020). Interacting Langevin diffusions: Gradient structure and ensemble Kalman sampler. SIAM Journal on Applied Dynamical Systems, 19(1), 412-441.

[7] Kaipio, J., Somersalo, E. (2006). Statistical and computational inverse problems (Vol. 160). Springer Science & Business Media.

[8] Pinnau, R., Totzeck, C., Tse, O., Martin, S. (2017). A consensus-based model for global optimization and its mean-field limit. Mathematical Models and Methods in Applied Sciences, 27(01), 183-204.