ENSEMBLE KALMAN INVERSION FOR NONLINEAR PROBLEMS: WEIGHTS, CONSISTENCY, AND VARIANCE BOUNDS

ZHIYAN DING†∗
Department of Mathematics
University of Wisconsin-Madison
Madison, WI 53706 USA

QIN LI†
Department of Mathematics
University of Wisconsin-Madison
Madison, WI 53706 USA

JIANFENG LU
Department of Mathematics
Duke University
Durham, NC 27708 USA

Abstract. Ensemble Kalman Inversion (EnKI) [23] and Ensemble Square Root Filter (EnSRF) [36] are popular sampling methods for obtaining a target posterior distribution. They can be seen as one step (the analysis step) in the data assimilation method Ensemble Kalman Filter [17, 3]. Despite their popularity, they are, however, not unbiased when the forward map is nonlinear [12, 16, 25]. Important Sampling (IS), on the other hand, obtains the unbiased sampling at the expense of large variance of weights, leading to slow convergence of high moments.

We propose WEnKI and WEnSRF, the weighted versions of EnKI and EnSRF in this paper. It follows the same gradient flow as that of EnKI/EnSRF with weight corrections. Compared to the classical methods, the new methods are unbiased, and compared with IS, the method has bounded weight variance. Both properties will be proved rigorously in this paper. We further discuss the stability of the underlying Fokker-Planck equation. This partially explains why EnKI, despite being inconsistent, performs well occasionally in nonlinear settings. Numerical evidence will be demonstrated at the end.

2020 Mathematics Subject Classification. Primary: 62D05; Secondary: 82C31.
Key words and phrases. Ensemble Kalman inversion, Ensemble Square Root Filter, Importance sampling, Fokker-Planck equation, Bayesian inference.

† Zhiyan Ding and Qin Li are supported in part by NSF CAREER DMS-1750488, NSF TRIPODS 1740707 and Wisconsin Data Science Initiative. The work of Jianfeng Lu is supported in part by National Science Foundation via grants DMS-1454939 and DMS-2012286. All three authors thank the two anonymous referees for the very helpful suggestions.
∗ Corresponding author: Zhiyan Ding.
1. **Introduction.** How to sample from an intractable distribution is a classical challenge emerging from Bayesian statistics, machine learning [1], computational physics [28, 29], among many other areas. Denote \( G : X \to Y \) a forward map between separable Hilbert spaces \( X \) and \( Y \). While the forward problem amounts to finding \( G(u) \) for every \( u \in X \), the inverse problem amounts to reconstructing the unknown parameters \( u \) from the observation \( y \). Sampling provides a probability perspective for such reconstruction procedure. Throughout the paper we set \( X = \mathbb{R}^L \) and \( Y = \mathbb{R}^K \).

Let \( y \) be the collected data. It is generated from the forward map \( G \) acting on \( u \) with added Gaussian noise \( \eta \) that is assumed to be independent of \( u \):

\[
y = G(u) + \eta, \quad \text{with} \quad \eta \sim \mathcal{N}(0, \Gamma).
\]

Throughout, we assume \( G \) is sufficiently smooth and its gradient is denoted by

\[
[\nabla G(u)]_{i,j} = \partial_j G_i, \quad \forall 1 \leq i \leq K, 1 \leq j \leq L.
\]

To find \( u \) using \( y \), a typical approach is to perform minimization. We denote the least-squares functional \( \Phi(\cdot; y) : X \to \mathbb{R} \) by

\[
\Phi(u; y) = \frac{1}{2} |y - G(u)|_\Gamma^2 = \frac{1}{2} (y - G(u))^\top \Gamma^{-1} (y - G(u)),
\]

then the optimal solution \( u^* \) is simply the parameter that minimizes the mismatch:

\[
u^* = \arg\min_u \Phi(u; y).
\]

This approach however is unable to characterize the uncertainty of the estimation.

In the Bayesian formulation, one takes a probability point of view, and regards \( u \) as a random variable. The aim is to reconstruct the probability distribution of \( u \) that combines the prior knowledge and the information from the collected data \( y \). More explicitly, let \( \rho_{\text{prior}}(u) \) be the prior distribution, then the posterior distribution of \( u \), denoted by \( \rho_{\text{pos}} \), includes the prior distribution, modified by the likelihood function:

\[
\rho_{\text{pos}}(u) = \frac{1}{Z} \exp (-\Phi(u; y)) \rho_{\text{prior}}(u).
\]

The normalization constant \( Z \) is given by:

\[
Z := \int_X \exp (-\Phi(u; y)) \rho_{\text{prior}}(u) \, du,
\]

so that \( \int \rho_{\text{pos}}(u) \, du = 1 \).

This perspective provides the full landscape of \( u \). While it provides more information, the computational cost is certainly more demanding.

Sampling is one problem emerging under this framework: how to design a cheap numerical solver that generates (hopefully i.i.d.) samples from the target distribution (3)? In particular, suppose one can sample \( N \) particles in \( \{u^n\}_{n=1}^N \in X \), and each particle is associated with a weight \( w^n \), then how to design the values for \((u^n, w^n)\) so that, in some sense

\[
\sum_{n=1}^N w^n \delta_{u^n} \approx \rho_{\text{pos}}
\]

Many sampling algorithms have been proposed in literature, ranging from classical techniques such as Markov chain Monte Carlo to strategies based on interacting particles. Some set \( w^n = \frac{1}{N} \) for all \( n \), while others use \( u^n \)-dependent weights \( w^n \). We will explore the latter in this work.

There are two general guiding principles for designing sampling algorithms: consistency and small variance.
WEIGHTED ENSEMBLE KALMAN INVERSION

• Consistency means that the ensemble distribution is “equivalent” to the target posterior distribution, in the average sense: When tested on all smooth functions \( f \), we require
\[
E \left( \sum_{n=1}^{N} \omega^n f(u^n) \right) = E_{\rho_{\text{pos}}}(f) .
\] (5)

Here the \( E \) sign on the left hand side means taking expectation of all sampling configurations. Denote the ensemble distribution
\[
\mu = \sum_{n=1}^{N} \omega^n \delta_{u^n} ,
\] (6)
then we say \( \mu \) is consistent with \( \rho_{\text{pos}} \), or \( \mu \sim \rho_{\text{pos}} \), if (5) holds true. In some literature, this property is called unbiased sampling.

• Variance of the weights gives an indicator of the performance of the sampling algorithm, it measures how close each configuration of (6), from one run of the algorithm, is to the true, i.e. we would like an algorithm so that
\[
E \left( \sum_{n=1}^{N} \omega^n f(u^n) - E_{\rho_{\text{pos}}}(f) \right)^2 \text{ is small.} \quad (7)
\]
Once again the \( E \) sign takes expectation over all possible configurations from the sampling algorithm. For a bounded test function \( f \), if \( \{ (\omega^n, u^n) \}_{n=1}^{N} \) are i.i.d., then:
\[
E \left( \sum_{n=1}^{N} \omega^n f(u^n) - E_{\rho_{\text{pos}}}(f) \right)^2 = E \sum_{n=1}^{N} \left( \omega^n f(u^n) - \frac{1}{N} E_{\rho_{\text{pos}}}(f) \right)^2
\]
\[
= N E \left[ \left( \omega^1 - \frac{1}{N} \right) f(u^1) + \frac{1}{N} f(u^1) - \frac{1}{N} E_{\rho_{\text{pos}}}(f) \right]^2
\]
\[
\leq \frac{2}{N} \text{Var}(N \omega^1) \| f \|_{L^\infty}^2 + \frac{2}{N} E \| f(u^1) - E_{\rho_{\text{pos}}}(f) \|_{L^2}^2
\]
\[
\leq \frac{2}{N} \text{Var}(N \omega^1) \| f \|_{L^\infty}^2 + \frac{8}{N} \| f \|_{L^2}^2 ,
\] (8)
where we use i.i.d. in the second equality and \( E(N \omega^1) = 1 \) by consistency. This means the variance of the weight, \( \text{Var}(N \omega^1) \), serves as a measure of the performance. If (7) is small, the algorithm is regarded as a good one. We note that some sampling algorithms cannot provide i.i.d. \( \{ \omega^n, u^n \} \) pairs, making the inequality (8) not exactly true. Nevertheless the variance of the weights in some sense quantifies how well each run of the experiment approximates the target posterior distribution.

There have been many successful algorithms developed in literature that aim at achieving these two properties. Our algorithms are built upon ideas from some of these methods, including “Importance Sampling” (IS) and “Ensemble Kalman Inversion/Square Root Filter” (EnKI/EnSRF), all three of which will be briefly recalled below and reviewed in more details in Section 2.

Importance Sampling is a rather standard technique: it involves assigning weights to particles so that an easy-to-be-sampled distribution can be turned into the target distribution. The weight is simply the ratio of the two. Regarding the two guiding principles, IS always achieves consistency, but it may give rise to high variance, especially when the easy-to-be-sampled and the target distribution are very different.
Some approaches have been proposed to incorporate “re-sampling” to reduce the variance, such as the strategies used in [14, 15, 27]. We do not discuss the details.

EnKI and EnSRF are very different. These two algorithms, both trace the origins to the Kalman filter, require the motion of the particles. They can be seen as the “analysis step” of data assimilation. Roughly speaking, the samples are generated from an easy-to-be-sampled distribution, and some dynamics is injected to move the samples around so that after finite time (usually Time = 1) they look like i.i.d. samples from the posterior distribution. These two algorithms have completely the opposite properties, compared with IS. There are no weights involved at all, and each particles takes $w^n = \frac{1}{N}$, so the variance of weight is always 0. However, they are not consistent. This is a disadvantage inherited from Ensemble Kalman Filter: ensemble Kalman filter highly relies on the Gaussianity assumption, that furthermore requires linearity of the forward map – for nonlinear forward map, the sampling methods are not consistent, in the notion of (5).

Our goal in this work is to design algorithms by combining advantages of IS and EnKI/EnSRF. We rely on the introduction of the weights to achieve consistency, and the motion introduced in EnKI/EnSRF helps reducing the variance. In this way, we propose the Weighted-Ensemble-Kalman-Inversion (WEnKI) and Weighted-Ensemble-Square-Root-Filter (WEnSRF) as weighted versions of the EnKI and EnSRF. They achieve consistency for general nonlinear forward maps. We also establish theoretical bounds of the weight variance for the proposed methods. In some sense, this work can be viewed as a correction to EnKI/EnSRF to ensure consistency and an improvement over IS in terms of reducing the weight variance. A natural question then is: how much improvement do we get? As a comparison to EnKI/EnSRF, this amounts to analyzing the strength of the weight term. This is a side product of the paper: by quantifying the differences between EnKI and WEnKI by estimating the weight term, we give a control of the error for EnKI when the forward map is nonlinear.

We should emphasize that besides the two essential properties mentioned above that are theoretically important, there are a lot of practical concerns in implementing algorithms. For example, it would be ideal in real practical problems to design methods that are derivative free, and have low computational complexity. This partially explains the popularity of IS and EnKI. The algorithms are extremely simple, and no derivatives of $G$ are needed. The proposed new algorithms in this paper fail badly in this dimension: the newly introduced weight terms not only depend on derivatives, but also have very complicated formulation, as will be shown in Section 3. Although theoretically they are indeed consistent and achieve low variance, such high computational complexity will render them being of little practical use. How to build on top the results obtained in this paper for a practical useful sampling method that also enjoy good theoretical properties will be explored in the near future.

The rest of this paper is organized in the following: in Section 2, we give a brief review of the above mentioned three methods, Important Sampling, Ensemble Kalman Inversion, and Ensemble Square Root Filter. In Section 3 we propose our correction to EnKI and EnSRF with added weights. Proof of consistency and some discussion about the control of the variance of weights are presented in Section 4. In Section 5 we demonstrate numerical evidence. Some concluding remarks are presented at the end of the paper.
2. Importance sampling and ensemble Kalman filter. We review a few sampling strategies in this section. In particular, the Importance Sampling that involves adding weights to the particles to achieve consistency, Ensemble Kalman Inversion and Ensemble Square Root Filter that involve adding motions to the particles so that samples are moved to represent the support of the target.

2.1. Importance sampling. The first sampling method we will discuss is the Importance Sampling [21]. It is a fundamental step in Sequential Monte Carlo Methods [14, 15]. The idea is extremely simple: one samples a certain amount of particles from the prior distribution, and weight is then calculated based on the ratio of the posterior and the prior evaluation, so the samples with adjusted weights reflect the posterior distribution. The algorithm is summarized in Algorithm 1:

**Algorithm 1 Importance sampling**

**Preparation:**
1. Input: \( N \gg 1; \Gamma; \mathcal{G} \) (forward map) and \( y \) (data).
2. Initial: \( \{u^n\}^N_{n=1} \) i.i.d. sampled from the initial distribution \( \rho_{\text{prior}} \).

**Run:**
1. Calculate the weight, for all \( 1 \leq n \leq N \):
   \[
   \omega^{n,*} = \exp\{ -\Phi(u^n; y) \} = \exp\left( -\frac{1}{2} |y - \mathcal{G}(u^n)|^2_\Gamma \right);
   \]
2. Normalize weight:
   \[
   \omega^n = \frac{\omega^{n,*}}{\sum_{n=1}^N \omega^{n,*}}.
   \]

**Output:** \( \{\omega^n\}^N_{n=1}; \{u^n\}^N_{n=1} \).

It is expected that the newly updated distribution is consistent with the target distribution:
\[
\sum_{n=1}^N \omega^n \delta_{u^n} \sim \rho_{\text{pos}}
\]
in the sense that for any smooth test function \( f \):
\[
\mathbb{E}\left( \sum_{n=1}^N \omega^n f(u^n) \right) = \mathbb{E}_{\rho_{\text{pos}}}(f).
\]

However, the variance of the weights could be quite large, especially when \( \rho_{\text{pos}} \) and \( \rho_{\text{prior}} \) concentrate at different regions. According to the formulation of the method, this quantity can be explicitly computed:
\[
\mathbb{E}(\omega^n)^2 = \frac{1}{N^2} \int_{\mathbb{R}^L} \rho_{\text{pos}}^2(u) \rho_{\text{prior}}(u) \, du, \quad \text{and} \quad \text{Var}(N\omega) = \int_{\mathbb{R}^L} \rho_{\text{pos}}^2(u) \rho_{\text{prior}}(u) \, du - 1. \quad (9)
\]
Thus, if \( \rho_{\text{pos}} \) is non-trivial is the region where \( \rho_{\text{prior}} \) almost vanishes, the quantity can be extremely big, leading to poor performance of the algorithm. Various resampling strategies have been proposed [2] to reduce the high variance.

2.2. Ensemble Kalman filters. At the other end of the spectrum of sampling method is to not adjust weights at all. Every particle takes equal weight \( \frac{1}{N} \). Two typical examples are Ensemble Kalman Inversion and Ensemble Square Root Filter.

The link between sampling and the Kalman filter problem was drawn in an inspiring paper [32]. Kalman filter (or its more practical version: ensemble Kalman
filter [17]) is a class of data assimilation methods that combine data (usually collected at discrete time) with some underlying guessed system dynamics an estimation of parameters in dynamical systems. The dynamics is ran till discrete time when data is collected, and Bayes’ rule is applied to update the distribution of the unknown parameters. The paper views the application of the Bayes’ rule as an action at a delta function in time, and by inserting a mollifier, the updating process becomes continuous in time. There is a vast literation in Kalman filter and its variation [4, 3], some recent analysis can be found in [11, 22].

Such idea was elaborated and formulated into a minimization strategy in [23, 10]. In [35], the authors view the prior and the (modified) target distribution to be two functions on a function space, and designed a PDE that transforms one to another, either in finite time for the posterior distribution, or in the infinite time horizon for a delta function located at the minimizer. The sampling strategy is in some sense equivalent to the particle method for the PDE: the samples are drawn from the initial distribution, and follow the flow of the PDE by satisfying the associated coupled-ODE/SDE systems. The initial finite-time sampling method is termed “Ensemble Kalman Inversion (EnKI)” in [23], and some variations were developed that achieve the final distribution in infinite time, termed “Ensemble Kalman Sampling (EKS)” [13, 19, 20]. Since there are no adjustment of weights, the variance of weights keep being 0 throughout the dynamics. Indeed, upon the well-posedness results of the SDE obtained in [5, 35], in [12] the authors proved, using the mean-field argument [7, 30], that when the forward map $\mathcal{G}$ is linear, the method provides approximately i.i.d. samples for the posterior distribution (with $N^{-1/2}$ error in $L^2$-Wasserstein metric).

However, both the derivation of the PDE, and the mean-field limit argument, highly rely on Gaussianity. The forward map is required to be linear for the arguments to carry through. This is not a surprising property since the method was originally derived from Ensemble Kalman Filter and thus inherits its strong requirement: the “motion” of the particles only depend on the first two moments, and thus the method automatically fails when higher moments are necessary, as in the non-Gaussian case.

We describe both EnKI and EnSRF in details below.

2.2.1. Ensemble square root filter. The PDE for the ensemble square root filter (EnSRF) writes as the following:

\[
\begin{aligned}
\partial_t \varrho(u, t) - \frac{1}{2} \nabla \cdot \left( \text{Cov}_{u_p}^\varrho(t)^{-1} \left( \mathcal{G}(u) + \mathcal{G}^\varrho(t) - 2y \right) \varrho \right) &= 0, \\
\varrho(u, 0) &= \rho_{\text{prior}},
\end{aligned}
\]

(10)

where $\mathcal{G}^\varrho(t), \text{Cov}_{u_p}^\varrho(t)$ are expectation of $\mathcal{G}$ and the covariance of $(u, \mathcal{G}(u))$ in $\varrho(u, t)$:

$\mathcal{G}^\varrho(t) = \int \mathcal{G}(u) \varrho(t) \, du, \quad \text{Cov}_{u_p}^\varrho(t) = \int (u - \bar{u}) \otimes (\mathcal{G}(u) - \mathcal{G}) \varrho(t) \, du.$

For this particular PDE, one can show that if $\mathcal{G}$ is linear, namely:

$\mathcal{G}(u) = Au + b,$

(11)

the solution to the PDE (10) is the target posterior distribution at $t = 1$:

$\varrho(u, 1) = \rho_{\text{pos}}.$
Noting that the PDE (10) is essentially an advection-type PDE, it is easy to formulate the ODE system satisfied by the particles by simply following the trajectory:

\[
\frac{d}{dt} u^n_t = -\frac{1}{2} \text{Cov}_{u^n}(t)^{-1} \left( \mathcal{G}(u^n_t) + \mathcal{G}^N(t) - 2y \right),
\]

with \( \{u^n\} \), i.i.d. sampled from \( \rho_{\text{prior}} \) at \( t = 0 \). Since the particles \( \{u^n\} \) follow exactly the same flow as the PDE, it is straightforward to have, for \( \forall t \):

\[
\frac{1}{N} \sum_j \delta_{u^n(t)} \approx \rho(u,t).
\]

This approximation sign holds true in both weak sense, and in Wasserstein distance sense, for all \( t \leq 1 \):

- Weak convergence, for all \( f(u) \) bounded continuous:
  \[
  \mathbb{E} \left( \int \left( \frac{1}{N} \sum_{n=1}^{N} \delta_{u^n(t)} - \rho(u,t) \right) f(u) \, du \right) = 0,
  \]
  and
  \[
  \mathbb{E} \left( \int \left( \frac{1}{N} \sum_{n=1}^{N} \delta_{u^n(t)} - \rho(u,t) \right) f(u) \, du \right)^2 = \mathcal{O}(N^{-1});
  \]

- Convergence in \( L_2 \)-Wasserstein:
  \[
  \mathbb{E} \left( W_2 \left( \frac{1}{N} \sum_{n=1}^{N} \delta_{u^n(t)} ; \rho(u,t) \right) \right) \to 0.
\]

Note that the rate of convergence in \( L_2 \)-Wasserstein depends on the dimension. It is of \( \mathcal{O}(N^{-1/2}) \) if the dimension of \( u \) is smaller than 4. Details can be found in [18].

However, in the numerical experiment, since one does not have \( \rho(u,t), \mathcal{G} \) and \( \text{Cov}_{u^n}(t) \) are not available. In implementation these terms are replaced by the ensemble covariance and the ensemble mean:

\[
\mathcal{G}^N(t) = \frac{1}{N} \sum_{n=1}^{N} \mathcal{G}(u^n(t)), \quad \text{and} \quad \mathcal{G}(t) \to \mathcal{G}^N(t) = \frac{1}{N} \sum_{n=1}^{N} u^n(t),
\]

and

\[
\text{Cov}_{u^n}(t) \to \text{Cov}_{u^n}^N(t) = \frac{1}{N} \sum_{n=1}^{N} (u^n(t) - \mathcal{G}^N(t)) \otimes \left( \mathcal{G}(u^n(t)) - \mathcal{G}^N(t) \right).
\]

These replacements naturally bring error to realizations of (12). To prove such error is small, the classical mean-field argument is ran. The full recipe of the algorithm is summarized in Algorithm 2.

It is clear in the algorithm, (14) is simply the forward Euler solver applied on ODE (12) with time step being \( h = 1/M \), and the accuracy would be the standard \( \mathcal{O}(h) \). The method was proposed in papers [26, 32, 36] as a data assimilation method. The idea behind the scene is rather simple. Suppose a large number of particles are sampled from a normal distribution \( \mathcal{N}(\mu_1, \Sigma_1) \), and to form \( \mathcal{N}(\mu_2, \Sigma_2) \), one merely needs to adjust \( u^n \) to a new location:

\[
u^n \to \Sigma_2^{1/2} \Sigma_1^{-1/2} (u^n - \mu_1) + \mu_2.
\]

The newly formulated particles are then i.i.d. drawn from \( \mathcal{N}(\mu_2, \Sigma_2) \). The ODE (12) is the continuous in time version of this motion. It is immediate that since only the
Algorithm 2 Ensemble Square Root filter

**Preparation:**
1. Input: $N \gg 1; \Delta t \ll 1$ (time step); $M = 1/\Delta t$ (stopping index); $\mathcal{G}$ (forward map) and $y$ (data).
2. Initial: $\{u^n_0\}_{n=1}^N$ sampled from initial distribution $\rho_{\text{prior}}$.

**Run:** Set time step $m = 0$;

While $m < M$:
1. Define empirical means and covariance:
   
   $$
   \pi^n_m = \frac{1}{N} \sum_{n=1}^N u^n_m, \quad \mathcal{G}^N_m = \frac{1}{N} \sum_{n=1}^N \mathcal{G}(u^n_m)
   $$

   and
   
   $$
   \text{Cov}^{N, \text{up}}_{\pi} = \frac{1}{N} \sum_{n=1}^N (u^n_m - \pi^n_m) \otimes \left( \mathcal{G}(u^n_m) - \mathcal{G}^N_m \right).
   $$

2. Update (set $m \to m + 1$)

   $$
   u^n_{m+1} = u^n_m - \Delta t \frac{1}{2} \text{Cov}^{N, \text{up}}_{\pi} \Gamma^{-1} \left( \mathcal{G}(u^n_m) + \mathcal{G}^N_m - 2y \right), \quad \forall 1 \leq n \leq N.
   $$

**Output:** $\{u^n_M\}_{n=1}^N$.

Information of the first two moments is used, Gaussianity is crucial, meaning for consistency, the forward map $\mathcal{G}$ is necessary to be linear.

2.2.2. Ensemble Kalman inversion. A similar approach is used to derive another sampling method called Ensemble Kalman Inversion [17, 32]. The corresponding PDE is the following:

$$
\begin{cases}
\partial_t \varrho(u, t) + \nabla \cdot \left( (y - \mathcal{G}(u))^\top \Gamma^{-1} \text{Cov}^{\text{up}}_{\varrho} \varrho \right) = \frac{1}{2} \text{Tr} \left( \text{Cov}^{\text{up}}_{\varrho} \Gamma^{-1} \text{Cov}^{\text{up}}_{\varrho} \mathcal{H}_u \varrho \right), \\
\varrho(u, 0) = \rho_{\text{prior}}
\end{cases}
$$

where $\text{Cov}^{\text{up}}_{\varrho}$ and $\text{Cov}^{\text{up}}_{\varrho}$ are covariance of $(u, \mathcal{G})$ and $(\mathcal{G}, u)$ in $\varrho(u, t)$. $\mathcal{H}_u \varrho$ is the Hessian of $\varrho$. In [12] the authors showed that the solution to the PDE reconstructs the posterior distribution in finite time:

$$
\varrho(t = 1, u) = \rho_{\text{pos}}
$$

if the forward map $\mathcal{G}$ is linear (11). So the PDE provides a smooth path to transform the prior distribution to the target in the linear setting.

On the particle level, by following the trajectory of this PDE one has the following SDEs:

$$
\begin{align*}
    \text{d}u^n_t &= \text{Cov}^{\text{up}}_{\varrho}(t)^{-1} (y - \mathcal{G}(u^n_t)) \, \text{d}t + \text{Cov}^{\text{up}}_{\varrho}(t)^{-1/2} \, \text{d}W^n_t, \quad n = 1, \cdots N, \quad (18)
\end{align*}
$$

where $\text{d}W^n_t$ is the Brownian motion. In the implementation of this SDE, since $\varrho$ is not available, the covariance matrices need to be replaced by the ensemble versions, as is done in (13), meaning, in the real computation, we use the following coupled SDEs:

$$
\begin{align*}
    \text{d}u^n_t &= \text{Cov}^{N}_{\varrho}(t)^{-1} (y - \mathcal{G}(u^n_t)) \, \text{d}t + \text{Cov}^{N}_{\varrho}(t)^{-1/2} \, \text{d}W^n_t, \quad n = 1, \cdots N, \quad (19)
\end{align*}
$$
where \( \text{Cov}_{up}^N \) is the ensemble covariance matrix. Finally we use the ensemble distribution
\[
\frac{1}{N} \sum_{n=1}^{N} \delta_{u^n(t)} \approx \varrho(u,t),
\]
to approximate the PDE solution.

The discrete version of the coupled SDE (18) formulates Algorithm 3. It is apparent that (22) is simply the Euler-Maruyama method for (18), as rigorously justified in [6, 24].

**Algorithm 3 Ensemble Kalman Inversion**

**Preparation:**
1. Input: \( N \gg 1; \Delta t \ll 1 \) (time step); \( M = 1/\Delta t \) (stopping index); \( \Gamma; \mathcal{G} \) (forward map) and \( y \) (data).
2. Initial: \( \{u^n_0\}_{n=1}^{N} \) sampled from initial distribution \( \varrho_{\text{prior}} \).

**Run:** Set time step \( m = 0 \);

While \( m < M \):

1. Define empirical means and covariance:
   \[
   \overline{u}_m = \frac{1}{N} \sum_{n=1}^{N} u^n_m, \quad \text{and} \quad \text{Cov}_{up}^N = \frac{1}{N} \sum_{n=1}^{N} (u^n_m - \overline{u}_m) \otimes (\mathcal{G}(u^n_m) - \overline{\varrho}_m),
   \]
   \[
   \overline{\varrho}_m = \frac{1}{N} \sum_{n=1}^{N} \mathcal{G}(u^n_m), \quad \text{and} \quad \text{Cov}_{pp}^N = \frac{1}{N} \sum_{n=1}^{N} (\mathcal{G}(u^n_m) - \overline{\varrho}_m) \otimes (\mathcal{G}(u^n_m) - \overline{\varrho}_m).
   \]

2. Artificially perturb data (with \( \xi^n_{m+1} \) drawn i.i.d. from \( \mathcal{N}(0, (\Delta t)^{-1} \Gamma) \)):
   \[
   y^n_{m+1} = y + \xi^n_{m+1}, \quad n = 1, \ldots, N.
   \]

3. Update (set \( m \rightarrow m + 1 \))
   \[
   u^n_{m+1} = u^n_m + \text{Cov}_{up}^N \left( \text{Cov}_{pp}^N + (\Delta t)^{-1} \Gamma \right)^{-1} (y^n_{m+1} - \mathcal{G}(u^n_m)), \quad \forall 1 \leq n \leq N.
   \]

end

Output: \( \{u^n_M\}_{n=1}^{N} \).

The EKI method was initially proposed in [23], as a further development of [32], to find optimized parameter for inverse problem. Then continuous limit for the discretization in time was considered in [35] where the authors first wrote down and analyzed the SDE system (18). The well-posedness of this SDE system was shown [12]. The wellposedness of the new SDE system (19), which has the covariance replaced by its ensemble version, was proved in [5, 6]. In [12] the authors showed the mean-field limit of the new coupled SDE system (19) is the PDE (16) as \( N \rightarrow \infty \) (in \( L_2 \)-Wasserstein sense).

However, we would like to emphasize that in [12] it was shown the PDE provides the target distribution only in the linear setting while such convergence holds true for the relatively general weakly nonlinear case. Defending on the perspective, this is in fact a negative result for the nonlinear case: the target distribution is not the solution to the PDE, but the method nevertheless presents the flow to the PDE, so the method does not give a consistent sampling of the target distribution.
We also note that often in time, people view EnKI as an optimization algorithm instead of a sampling algorithm, and some relaxation terms have been added for convergence to the minimizer [8, 9].

2.3. Summary. It is rather clear that in IS, the particles are kept in the original location, and one merely adjusts the weights. This guarantees the consistency, namely, (5) always holds true for all bounded continuous functions. On the other hand, since the particles do not move, the weights could be largely suppressed or enlarged, leading to large variance of the weights even in the Gaussian case.

On the contrary, the later two algorithms, EnSRF and EnKI, move particles around to adjust the change of center and the variance. Since all particles are equally weighted, the variance of weight is kept at 0. However, the derivation of both methods assumes the Gaussianity, and thus the consistency fails for the nonlinear forward map.

3. Weighted ensemble Kalman inversion and square root filter. Our proposed algorithms combine the advantages of IS and EnKI/EnSRF, by including both weight and particle dynamics simultaneously, so that we guarantee the consistency at the expense of fairly small variance. The output of the algorithms would be an ensemble distribution having the format of

\[
M_{\text{en}} = \sum_{n=1}^{N} w_n \delta_{u_n},
\]

as an approximation to the target distribution \( \rho_{\text{pos}} \).

We call the proposed algorithms weighted-EnKI (WEnKI) and weighted-EnSRF (WEnSRF). As the names suggest, we largely keep the format of the flow (or the PDE) for EnKI and EnSRF, while we also add weights to achieve consistency. The underlying flow is designed so that the PDE solution provides a linear interpolation on the log-scale in a time parameter \( t \), from the prior to the posterior distributions [32, 35]:

\[
\rho(u, t) = \frac{1}{Z(t)} \exp\{-t\Phi(u; y)\} \rho_{\text{prior}}(u), \quad t \in [0, 1],
\]

where \( \Phi \) is the least-squares function defined in (1) and \( Z(t) \) is a function in time to normalize \( \rho(u, t) \) so that

\[
\int \rho(u, t) \, du = 1, \quad \forall t.
\]

It is clear that

\[
\rho(u, 0) = \rho_{\text{prior}}, \quad \rho(u, 1) = \rho_{\text{pos}},
\]

so the definition (24) provides a flow from the prior to the target posterior distribution. The prior distribution in our algorithm can be quite flexible, for example,

\[
\rho_{\text{prior}}(u) = \frac{1}{Z} \exp(-V(u))
\]

for a \( C^2 \) function \( V(u) \), with \( Z \) being the normalization factor. In practice, however, the prior distribution needs to be an distribution that is easy to sample, so for now we assume:

\[
\rho_{\text{prior}} = \mathcal{N}(u_0, \Gamma_0).
\]

The strategy we follow is divided into two steps:
Step 1: adjust the PDE (10) and (16) by adding weights so that (24) is a strong solution;

Step 2: design a corresponding particle system that carries out the flow of the PDE.

Before diving into details of the algorithms, we first introduce some notations. A straightforward but somewhat tedious calculation (see Appendix A) yields for ρ(u, t) defined in (24):

\[ \partial_t \rho(u, t) = \left[ -\frac{1}{2} |y - G(u)|^2_t + E_{\rho(t)} \left( \frac{1}{2} |y - G(u)|^2_t \right) \right] \rho(u, t), \tag{26} \]

\[ \nabla \rho(u, t) = \nabla u \rho(u, t), \tag{27} \]

\[ H_u \rho(u, t) = \left[ V(u, t)^T \Gamma^{-1} (y - G(u)) - \Gamma^{-1}_0 (u - u_0) \right], \tag{29} \]

where \( H_u \) denotes the Hessian with respect to \( u \), and \( V \in \mathbb{R}^{L \times 1} \), \( W \in \mathbb{R}^{L \times L} \) are defined as

\[ V(u, t) = t (\nabla G(u))^T \Gamma^{-1} (y - G(u)) - \Gamma^{-1}_0 (u - u_0), \]

\[ W(u) = \left[ (\partial_t \nabla G(u))^T \Gamma^{-1} (y - G(u)) \right] \cdot \left[ (\partial_t \nabla G(u))^T \Gamma^{-1} (y - G(u)) \right]. \tag{30} \]

3.1. Weighted ensemble square root filter (WEnSRF). Calculating the left hand side of (10) using the identities (26)-(28), we arrive at the PDE that ρ, defined in (24), satisfies:

\[ \partial_t \rho(u, t) - \frac{1}{2} \nabla \cdot \left( \text{Cov}_{u_p}^{\rho(t)} \Gamma^{-1} (G(u) + \tilde{G}^{\rho(t)} - 2y) \right) = [\mathcal{P}_1(u, t) + \mathcal{P}_2(u, t)] \rho, \tag{31} \]

where

\[ \mathcal{P}_1(u, t) = \frac{1}{2} \left\{ \left| y - \tilde{G}^{\rho(t)} \right|_{\Gamma} - |y - G(u)|_{\Gamma} \right\} + \frac{1}{2} \text{Tr} \left\{ \text{Cov}_{u_p}^{\rho(t)} \Gamma^{-1} \right\}, \]

\[ \mathcal{P}_2(u, t) = - \frac{1}{2} \text{Tr} \left\{ \text{Cov}_{u_p}^{\rho(t)} \Gamma^{-1} \nabla G(u) \right\} - \frac{1}{2} V^T(u, t) \text{Cov}_{u_p}^{\rho(t)} \Gamma^{-1} (G(u) + \tilde{G}^{\rho(t)} - 2y) \tag{32} \]

with shorthand notations

\[ \tilde{G}^{\rho(t)} = E_{\rho(t)}(u), \quad \tilde{G}^{\rho(t)} = E_{\rho(t)}(G), \]

\[ \text{Cov}_{u_u}^{\rho(t)} = E_{\rho(t)} \left( \left( u - \tilde{G}^{\rho(t)} \right) \otimes \left( u - \tilde{G}^{\rho(t)} \right) \right), \]

\[ \text{Cov}_{u_p}^{\rho(t)} = E_{\rho(t)} \left( \left( u - \tilde{G}^{\rho(t)} \right) \otimes \left( G - \tilde{G}^{\rho(t)} \right) \right), \]

\[ \text{Cov}_{u_p}^{\rho(t)} = E_{\rho(t)} \left( \left( G - \tilde{G}^{\rho(t)} \right) \otimes \left( G - \tilde{G}^{\rho(t)} \right) \right). \tag{33} \]

According to the derivation, it is a natural expectation that ρ is a strong solution. We will further show that the ensemble distribution of particles generated by the sampling method gives a weak solution to the PDE.

The PDE (31) can be solved using standard method of characteristics, which gives rise to the following coupled ODE system for the particles, with \( u^n_t \) denoting the location of the \( n \)-th particle at time \( t \), and \( w^n_t \) the associated weight:

\[ \begin{aligned}
    \frac{du^n_t}{dt} &= -\frac{1}{2} \text{Cov}_{u_p}^{\rho(t)} \Gamma^{-1} \left( G(u^n_t) + \tilde{G}^{\rho(t)} - 2y \right) dt \\
    \frac{dw^n_t}{dt} &= \left( \mathcal{P}_1(u^n_t, t) + \mathcal{P}_2(u^n_t, t) \right) w^n_t dt
\end{aligned} \tag{34} \]

The initial condition is chosen so that \( \{u^n_0\}_{n=1}^{N} \) is i.i.d. sampled from \( \rho_{\text{prior}}(u) \) and \( w^n_0 = 1/N \), \( n = 1, \ldots, N \), to represent initial data \( \rho(u, 0) = \rho_{\text{prior}} \). The system
is decoupled, in the sense that when the underlying \( g(t) \) is given, each \( \{u^n, w^n\} \) pairs are independent from each other. The output of the algorithm is the empirical distribution:

\[
M_n(u) = \sum_{n=1}^{N} w_n \delta_{u^n}.
\] (35)

In practice, however, \( g(t) \) is not known and thus \( \text{Cov}^{\theta(t)} \) and \( \mathcal{G}^{\theta(t)} \) in (34) have to be replaced by the ensemble versions:

\[
\mathbf{\pi}^{\theta(t)} \to \mathbf{\pi}^N(t) = \sum_{n=1}^{N} w_n(t) u^n(t), \quad \mathbf{\bar{G}}^{\theta(t)} \to \mathbf{\bar{G}}^N(t) = \sum_{n=1}^{N} w_n(t) \mathcal{G}(u^n(t))
\] (36)

and

\[
\text{Cov}_{up}^{\theta(t)} \to \text{Cov}_{up}^N(t) = \sum_{n=1}^{N} w_n(t) (u^n(t) - \mathbf{\pi}^N(t)) \otimes \mathcal{G}(u^n(t) - \mathbf{\bar{G}}^N(t)).
\] (37)

This replacement makes the SDE system tangled up. We summarize the method in Algorithm 4. Note that due to numerical error, it is typically hard to keep the summation of the weight 1, and numerically one performs normalization at each time step. Some properties of the method such as the consistency and the boundedness of the variance will be shown in Section 4.

3.2. Weighted ensemble Kalman inversion (WEnKI). The same strategy can be applied to modify EnKI to deal with nonlinearity. Substituting (26)-(28) into (16), we have

\[
\partial_t \mathcal{G}(u, t) + \mathcal{L} [\mathcal{G}] = [\mathcal{R}_1(u, t) + \mathcal{R}_2(u, t) + \mathcal{R}_3(u, t)] \mathcal{G}(u, t),
\] (38)

where \( \mathcal{L} \) is a linear operator inherited from (16):

\[
\mathcal{L} [\mathcal{G}] = \nabla_u \cdot \left( (y - \mathcal{G}(u))^\top \Gamma^{-1} \text{Cov}_{pu}^{\theta(t)} \mathcal{G} \right) - \frac{1}{2} \text{Tr} \left( \text{Cov}_{up}^{\theta(t)} \Gamma^{-1} \text{Cov}_{pu}^{\theta(t)} \mathcal{H}_u(\mathcal{G}) \right),
\] (39)

and the remaining terms \( \mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3 \) are given by

\[
\mathcal{R}_1(u, t) = \frac{1}{2} \text{Tr} \left( \text{Cov}_{pp}^{\theta(t)} \Gamma^{-1} - 2 (\nabla \mathcal{G}(u) )^\top \Gamma^{-1} \text{Cov}_{pu}^{\theta(t)} \right)
\]

\[
+ \frac{1}{2} \text{Tr} \left( \text{Cov}_{up}^{\theta(t)} \Gamma^{-1} \text{Cov}_{pu}^{\theta(t)} \left[ t (\nabla \mathcal{G}(u) )^\top \Gamma^{-1} \nabla \mathcal{G}(u) + \Gamma_0^{-1} \right] \right),
\] (40)

\[
\mathcal{R}_2(u, t) = \frac{1}{2} \left[ (y - \mathcal{G}(u))^\top \Gamma - \frac{1}{2} \left| y - \mathcal{G}(u) - \text{Cov}_{pu}^{\theta(t)} \mathcal{V}(u, t) \right|_{\Gamma} \right],
\]

\[
\mathcal{R}_3(u, t) = - \frac{t}{2} \text{Tr} \left( \text{Cov}_{up}^{\theta(t)} \Gamma^{-1} \text{Cov}_{pu}^{\theta(t)} \mathcal{W}(u) \right),
\]

where \( \text{Cov}_{pp}^{\theta(t)}, \text{Cov}_{pu}^{\theta(t)}, \) and \( \text{Cov}_{pu}^{\theta(t)} \) are the corresponding covariance matrices, as defined in (33). Similar to WEnSRF, we arrive at the following decoupled SDE system:

\[
\begin{align*}
\frac{du^n_t}{dt} &= \text{Cov}_{up}^{\theta(t)} \Gamma^{-1} (y - \mathcal{G}(u^n_t)) \ dt + \text{Cov}_{up}^{\theta(t)} \Gamma^{-1/2} \ dW^n_t, \\
\frac{dw^n_t}{dt} &= (\mathcal{R}_1(u^n_t, t) + \mathcal{R}_2(u^n_t, t) + \mathcal{R}_3(u^n_t, t)) w^n_t \ dt, \\
\end{align*}
\] (41)

where the Brownian motion is introduced for the second order term in \( \mathcal{L} \). The initial condition is chosen so that \( \{u^n_0\}_{n=1}^{N} \) is i.i.d. sampled from \( \rho_{\text{prior}}(u) \) and \( w^n_0 = 1/N, \) \( n = 1, \ldots, N. \)
Algorithm 4 Weighted Ensemble Square Root Filter

Preparation:
1. Input: $N \gg 1; \Delta t \ll 1$ (time step); $M = 1/\Delta t$ (stopping index); $\Gamma; \mathcal{G}$ (forward map) and $y$ (data).
2. Initial: $\{u_0^n\}_{n=1}^N$ sampled from initial distribution $\rho_{\text{prior}}$. $\{w_0^n = \frac{1}{N}\}_{n=1}^N$ initial weight.

Run: Set time step $m = 0$;
While $m < M$:
1. Define empirical means and covariance:
   
   $$\bar{u}_m^N = \frac{1}{N} \sum_{n=1}^N u_m^n, \quad \bar{y}_m^N = \frac{1}{N} \sum_{n=1}^N \mathcal{G}(u_m^n),$$

   and
   
   $$\text{Cov}_{up}^N = \frac{1}{N} \sum_{n=1}^N (u_m^n - \bar{u}_m^N) \otimes \left( \mathcal{G}(u_m^n) - \bar{y}_m^N \right).$$

2. Update parameters:
   
   $$\mathcal{V}(u_m^n, t_m) = t_m \left( \nabla \mathcal{G}(u_m^n) \right)^\top \Gamma^{-1} \left( y - \mathcal{G}(u_m^n) \right) - \Gamma_0^{-1} (u_m^n - u_0),$$

   $$\mathcal{P}_{m,1}^n = \frac{1}{2} \left( |y - \bar{y}_m^N|_\Gamma - |y - \mathcal{G}(u_m^n)|_\Gamma + \text{Tr} \left\{ \text{Cov}_{pp}^N \Gamma^{-1} \right\} \right),$$

   $$\mathcal{P}_{m,2}^n = - \frac{1}{2} \text{Tr} \left\{ \text{Cov}_{up}^N \Gamma^{-1} \nabla \mathcal{G}(u_m^n) \right\}$$

   $$- \frac{1}{2} \mathcal{V}(u_m^n, t_m) \text{Cov}_{up}^N \Gamma^{-1} \left( \mathcal{G}(u_m^n) + \bar{y}_m^N - 2y \right).$$

3. Update (set $m \rightarrow m + 1$): for all $1 \leq n \leq N$:
   
   $$u_{m+1}^n = u_m^n - \Delta t \left( \frac{1}{2} \text{Cov}_{up}^N \Gamma^{-1} \left( \mathcal{G}(u_m^n) + \bar{y}_m^N - 2y \right) \right),$$

   $$w_{m+1}^{n,*} = w_m^n \exp \left( \Delta t \left( \mathcal{P}_{m,1}^n + \mathcal{P}_{m,2}^n \right) \right),$$

   $$w_{m+1}^n = \frac{w_{m+1}^{n,*}}{\sum_{n=1}^N w_{m+1}^{n,*}}.$$

end

Output: $\{w_m^n\}_{m=1}^N; \{u_m^n\}_{m=1}^N.$

Since $\sigma$ is unknown, as in (36)-(37), we once again replace the true covariance by the ensemble version, and define empirical distribution accordingly:

$$M_u(u) = \sum_{n=1}^N w_m^n \delta_{u_m^n}. \quad (42)$$

There are two sources of randomness involved in WEnKI: the initial sampling and the Brownian motion in (41). Let $\Omega$ be the sample space and $\mathcal{F}_0$ be the $\sigma$-algebra: $\sigma (u^n(t=0), 1 \leq n \leq N)$, then the filtration is introduced by the dynamics:

$$\mathcal{F}_t = \sigma (u^n(t=0), W_s^n, 1 \leq n \leq N, s \leq t).$$

It can be shown the SDE is well-posed in this $\sigma$-algebra [5, 12]. In next section, we will prove that the empirical distribution is consistent with $\rho(u, t)$ defined in (24) under the expectation sense in $\mathcal{F}_t$. We will also give control to the variance. The
method is summarized in Algorithm 5. As in the previous algorithm, the numerical error induces \( \sum w^n \neq 1 \), and an extra renormalization is conducted.

**Algorithm 5 Weighted Ensemble Kalman Inversion**

**Preparation:**
1. Input: \( N \gg 1; \Delta t \ll 1 \) (time step); \( M = 1/\Delta t \) (stopping index); \( \Gamma; \mathcal{G} \) (forward map) and \( y \) (data).
2. Initial: \( \{u_0^n\}_{n=1}^N \) sampled from initial distribution \( \mathcal{g}_{\text{prior}} \). \( \{w_0^n = \frac{1}{N}\}_{n=1}^N \) initial weight.

**Run:** Set time step \( m = 0 \):

**While** \( m < M \):
1. Define empirical means and covariance:
   \[
   \bar{\pi}_m^N = \frac{1}{N} \sum_{n=1}^N w_m^u u_m^n, \quad \text{and} \quad \bar{\mathcal{G}}_m^N = \frac{1}{N} \sum_{n=1}^N w_m^u \mathcal{G}(u_m^n),
   \]
   \[
   \text{Cov}_{pp}^N = \frac{1}{N} \sum_{n=1}^N w_m^u (\mathcal{G}(u_m^n) - \bar{\mathcal{G}}_m^N) \otimes (\mathcal{G}(u_m^n) - \bar{\mathcal{G}}_m^N),
   \]
   \[
   \text{Cov}_{up}^N = \frac{1}{N} \sum_{n=1}^N w_m^u (u_m^n - \bar{\pi}_m^N) \otimes (\mathcal{G}(u_m^n) - \bar{\mathcal{G}}_m^N).
   \]
2. Define first and second derivative:
   \[
   \mathcal{V}(u_m^n, t_m) = t_m (\nabla \mathcal{G}(u_m^n))^{\top} \Gamma^{-1} (y - \mathcal{G}(u_m^n)) - \Gamma_0^{-1} (u_m^n - u_0),
   \]
   \[
   \mathcal{W}(u_m^n) = [\partial_1 \nabla \mathcal{G}(u_m^n) \Gamma^{-1} (y - \mathcal{G}(u_m^n)) \cdots \partial_L \nabla \mathcal{G}(u_m^n) \Gamma^{-1} (y - \mathcal{G}(u_m^n))].
   \]
3. Define updated parameter:
   \[
   \mathcal{R}_{m,1}^n = \frac{1}{2} \text{Tr} \left\{ \text{Cov}_{pp}^N \Gamma^{-1} - 2 \text{Cov}_{up}^N \Gamma^{-1} \nabla \mathcal{G}(u_m^n) \right. \\
   \left. + \text{Cov}_{up}^N \Gamma^{-1} \text{Cov}_{pu}^N \left[ t_m (\nabla \mathcal{G}(u_m^n))^{\top} \Gamma^{-1} \nabla \mathcal{G}(u_m^n) + \Gamma_0^{-1} \right] \right\},
   \]
   \[
   \mathcal{R}_{m,2}^n = \frac{1}{2} \left| y - \mathcal{G}(u_m^n) \right|_{\Gamma} - \frac{1}{2} \left| y - \mathcal{G}(u_m^n) - \text{Cov}_{pu}^N \mathcal{W}(u_m^n, t_m) \right|_{\Gamma},
   \]
   \[
   \mathcal{R}_{m,3}^n = - \frac{t_m}{2} \text{Tr} \left\{ \text{Cov}_{up}^N \Gamma^{-1} \text{Cov}_{pu}^N \mathcal{W}(u_m^n) \right\}.
   \]
4. Artificially perturb data (with \( \xi_{m+1}^n \) drawn i.i.d. from \( \mathcal{N}(0, (\Delta t)^{-1} \Gamma) \)):
   \[
   y_{m+1}^n = y + \xi_{m+1}^n, \quad n = 1, \ldots, N.
   \]
5. Update (set \( m \rightarrow m + 1 \)): for all \( n \):
   \[
   u_{m+1}^n = u_m^n + \text{Cov}_{up}^N \left( \text{Cov}_{pp}^N + (\Delta t)^{-1} \Gamma \right)^{-1} (y_{m+1}^n - \mathcal{G}(u_m^n)),
   \]
   \[
   w_{m+1}^n = w_{m+1}^u \exp (\Delta t \left( \mathcal{R}_{m,1}^n + \mathcal{R}_{m,2}^n + \mathcal{R}_{m,3}^n \right)),
   \]
   \[
   w_{m+1} = \frac{w_{m+1}^{n,*}}{\sum_{n=1}^N w_{m+1}^n}. \quad \text{end}
   \]

**Output:** \( \{w_{m+1}^n\}_{n=1}^N, \{u_{m+1}^n\}_{n=1}^N \).

**Remark 1.** It is important to note that the method is different from running EnKI to time \( t = 1 \) and then apply Important Sampling. The latter was proposed
in [31] as a weighted version of Ensemble Kalman Filter [17], known as the Weighted Ensemble Kalman Filter (WEnKF).

Define the conditional mean
\[
E(u^n | u^n_0) = u^n_0 + \text{Cov}_{up}(u^n_0) (\text{Cov}_{pp}(u^n_0) + \Gamma)^{-1} (y - G(u^n_0)) ,
\]
and the conditional covariance:
\[
\text{Cov}(u^n | u^n_0) = \text{Cov}_{up}(u^n_0) (\text{Cov}_{pp}(u^n_0) + \Gamma)^{-1} \Gamma (\text{Cov}_{pp}(u^n_0) + \Gamma)^{-\top} \text{Cov}_{pu}(u^n_0) .
\]
In WEnKF, the particle weight is updated according to:
\[
\omega^n = \frac{1}{N} \times \frac{\rho_{\text{pos}}(u^n)}{\mathcal{N}(u^n; E(u^n | u^n_0), \text{Cov}(u^n | u^n_0))} ,
\]
where \(u^n_0\) are the initial samples according to the prior distribution, and \(\mathcal{N}\) is the density of Gaussian distribution centered at the conditional mean \(E(u^n | u^n_0)\). The major difference, compared with the one we propose in (41), is that the covariance used in (43) is calculated completely from the initial data. The updates along the evolution is entirely ignored. The updating formula in (41), however, involves the weights that evolve in time and is closer to the PDE solution (31).

Remark 2. We admit that the added weight terms are quite complicated, both in WEnKI and WEnSRF. In terms of the computational complexity, the new algorithms are far from being ideal. However, if we stick to the flow introduced by EnKI and EnSRF, it seems to difficult to avoid adding some cost to make the algorithm consistent. Another route to improve the computation is to modify the flow itself, see [34]. For example, we can involve derivatives in the flow by changing \(\text{Cov}_{up}\) to \(\text{Cov}_{uu}(\nabla G(u))^\top\). This potentially could lead to a better flow of the particles, and may potentially provide a less complicated weight term. We leave these to future works.

4. Properties of WEnKI and WEnSRF. We establish a few important properties of WEnKI and WEnSRF in this section. As argued in Section 2, the two guiding principles for the algorithm-design is consistency and small variance of the weights. These two properties are presented in §4.1 and §4.2 respectively. Furthermore, we study the difference between WEnKI and EnKI in §4.3, and provide some intuition for EnKI performing well sometimes, even when \(G\) is nonlinear.

We emphasize that in the proof below we use equation (34) and (41) where the covariance is provided by \(\rho(t)\). In numerics, these covariances matrices need to be replace by their ensemble versions, and another layer of error analysis needs to be added. This is beyond the scope of the current paper.

4.1. Consistency. The most important property is the consistency, namely, on average, the ensemble mean tested on any smooth function is the same as the real mean. Since the PDEs are obtained by forcing (24) to be the solution, the consistency is expected.

We first present the theorem for WEnSRF.

**Theorem 4.1.** Assume \(G : \mathbb{R}^L \rightarrow \mathbb{R}^K\) is a \(C^1\) function, then:

- the formula (24) is a strong solution to (31) with the initial condition \(\rho(u, 0) = \rho_{\text{prior}}\), namely, the PDE (31) smoothly connects the prior and the posterior distributions;
\begin{itemize}
\item the formula (34)-(35) is a weak solution to (31) with the initial condition \(\rho(u,0) = M_{u_0}(u)\), namely, the ODE system (34) follows the flow of the transition: for any smooth test function \(f : \mathbb{R}^L \to \mathbb{R}\) and \(0 \leq t \leq 1\), we have consistency:
\[
\mathbb{E}_{\rho(t)}(f) = \mathbb{E} \left( \sum_{n=1}^{N} w_n^n f(u_n^n) \right) = \mathbb{E} \left( \mathbb{E}_{M_{u_t}}(f) \right),
\]
where the \(\mathbb{E}\) on the outer layer of the right hand side comes from the random configuration of the initial condition for \(\{u_0^n\}\).
\end{itemize}

Proof. The first point is trivial: it amounts to substituting the solution (24) into the equation and balancing terms. To show that the empirical measure \(M_{u_t}\) is the weak solution to the PDE, we test it with a smooth function \(f(u)\). Note that
\[
\mathbb{E}_{M_{u_t}}(f) = \int \sum_{n=1}^{N} w_n^n \delta_{u_n^n} f(u) \, du = \sum_{n=1}^{N} w_n^n f(u_n^n),
\]
we have
\[
\frac{d}{dt} \mathbb{E}_{M_{u_t}}(f) = \frac{d}{dt} \sum_{n=1}^{N} w_n^n f(u_n^n) = \sum_{n=1}^{N} \frac{dw_n^n}{dt} f(u_n^n) + w_n^n \frac{df(w_n^n)}{dt}
\]
\[
= \sum_{n=1}^{N} \left[ \mathcal{P}_1(t, u_n^n) + \mathcal{P}_2(t, u_n^n) \right] w_n^n f(u_n^n)
\]
\[
- \frac{1}{2} w_n^n (\nabla f(u_n^n))^\top \text{Cov}_{u_n^n} \Gamma^{-1} (\mathcal{G}(u_n^n) + \mathcal{G} - 2y)
\]
\[
= \mathbb{E}_{M_{u_t}} \left( \left[ \mathcal{P}_1(u, t) + \mathcal{P}_2(u, t) \right] f(u) \right)
\]
\[
- \mathbb{E}_{M_{u_t}} \left( \frac{1}{2} (\nabla f(u))^\top \text{Cov}_{u_t} \Gamma^{-1} (\mathcal{G}(u) + \mathcal{G} - 2y) \right).
\]
This is exactly the weak formulation of (31) tested on \(f\) with the integration by parts applied on the advection term. To show (45), we simply note that both \(\rho\) and \(\varrho = M_{u_t}\) are weak solutions.

The same type of theorem holds true for WEnKI:

\begin{theorem}
If \(\mathcal{G} : \mathbb{R}^L \to \mathbb{R}^K\) is \(C^2\) and Lipschitz continuous,
\begin{itemize}
\item the formula (24) is a strong solution to (31) with the initial condition \(\rho(u,0) = \rho_{\text{prior}}\), namely, the PDE (38) characterizes the dynamics in (24) and connects the prior and the posterior distributions;
\item the formula (41)-(42), in expectation, is a weak solution to (31) with the initial condition \(\rho(u,0) = M_{u_0}(u)\), namely, the SDE system (41) follows the flow of the transition: for any smooth test function \(f : \mathbb{R}^L \to \mathbb{R}\) and \(0 \leq t \leq 1\),
\[
\mathbb{E}_{\rho(t)}(f) = \mathbb{E} \left( \sum_{n=1}^{N} w_n^n f(u_n^n) \right) = \mathbb{E} \left( \mathbb{E}_{M_{u_t}}(f) \right),
\]
where the outer-layer \(\mathbb{E}\) on the right hand side is taken in the probability space \((\Omega, \mathcal{F}_t, \mathbb{P})\).
\end{itemize}
\end{theorem}
Proof. The first part is again trivial. To show (46), we first realize that the equation holds trivially for \( t = 0 \) since \( \{w^n_n\}_{n=1}^{N} \) are i.i.d sampled from \( \rho_{\text{prior}}(u) \). For all \( t > 0 \), we plug in (41) and apply the Itô’s formula on
\[
d\sum_{n=1}^{N} w^n_t f(u^n_t) = \sum_{n=1}^{N} dw^n_t f(u^n_t) + w^n_t df(u^n_t)
\]
and use
\[
df(u^n_t) = (\nabla f(u^n_t))^\top \text{Cov}_{u^n_t}^{-1} (y - G(u^n_t)) \, dt + (\nabla f(u^n_t))^\top \text{Cov}_{u^n_t}^{-1/2} dW^n_t
\]
\[
+ \frac{1}{2} \text{Tr} \left\{ \mathcal{H}_v(f(u^n_t))\text{Cov}_{u^n_t}^{-1}\text{Cov}_v \right\} \, dt
\]
to get
\[
d\mathbb{E}(\mathbb{E}_{\text{M}_{n,t}}(f)) = d\mathbb{E} \left( \sum_{n=1}^{N} w^n_t f(u^n_t) \right)
\]
\[
= \mathbb{E} \sum_{n=1}^{N} \left[ (\mathcal{R}_1(t, u^n_t) + \mathcal{R}_2(t, u^n_t) + \mathcal{R}_3(t, u^n_t)) \right] w^n_t f(u^n_t) \, dt
\]
\[
+ w^n_t (\nabla f(u^n_t))^\top \text{Cov}_{u^n_t}^{-1} (y - G(u^n_t)) \, dt
\]
\[
+ \frac{1}{2} w^n_t \text{Tr} \left\{ \mathcal{H}_v(f(u^n_t))\text{Cov}_{u^n_t}^{-1}\text{Cov}_v \right\} \, dt
\]
\[
= \mathbb{E} \left( \mathbb{E}_{\text{M}_{n,t}} (\mathcal{R}_1 + \mathcal{R}_2 + \mathcal{R}_3) f(u) \right) \, dt
\]
\[
+ \mathbb{E} \left( \mathbb{E}_{\text{M}_{n,t}} \left( (\nabla f(u))^\top \text{Cov}_{u^n_t}^{-1} (y - G(u)) \right) \, dt
\]
\[
+ \frac{1}{2} \text{Tr} \left\{ \mathcal{H}_v(f(u))\text{Cov}_{u^n_t}^{-1}\text{Cov}_v \right\} \, dt \right) .
\]

In the second equation above, we used the fact that the SDE that \( u_t \) satisfies is wellposed due to the Lipschitz continuity of \( G \).

This formulation is exactly the weak formulation of (38) tested on \( f \) with integration by parts moving the \( \nabla \) and \( \mathcal{H} \) onto \( f \). The equality (46) follows as \( \rho(u, t) \) defined in (24) is also a weak solution. \( \square \)

4.2. Bounding the variance of weights. We investigate the behavior of the weights for a fairly large class of \( G \). Throughout this subsection, we will impose one of the two assumptions on the forward map \( \mathcal{G} \) below.

The first assumption is rather weak, and it only requires the boundedness of derivatives of \( \mathcal{G} \) up to second order.

**Assumption 4.1.** \( \mathcal{G} : \mathbb{R}^L \rightarrow \mathbb{R}^K \) is \( C^2 \) function and there exists \( \Lambda > 0 \) such that
\[
\| \nabla \mathcal{G} \|_2 \leq \Lambda, \quad \| \mathcal{H}(\mathcal{G}_i^2) \|_2 \leq \Lambda, \quad \| \partial_i \nabla \mathcal{G} \|_2 \leq \Lambda, \quad 1 \leq i \leq L . \quad (47)
\]

The second assumption is slightly stronger, and it asks for the structure of the range of the linear and nonlinear components of \( \mathcal{G} \).

**Assumption 4.2.** \( \mathcal{G} \) is weakly-nonlinear in the sense that there exists a matrix \( A \in \mathcal{L}(\mathbb{R}^L, \mathbb{R}^K) \) so that
\[
\mathcal{G}(u) = Au + m(u), \quad (48)
\]
where \( m(u) \) is a \( C^2 \) bounded Lipschitz function from \( \mathbb{R}^L \) to \( \mathbb{R}^K \) satisfying
\[
\Gamma^{-1/2} m(u) \perp \Gamma^{-1/2} Au, \quad \forall u \in \mathbb{R}^L ,
\]
and there exists constants $\Lambda, \Lambda_1$ and $M$ such that: for $1 \leq i \leq L$
\[ \|\nabla m\|_2 \leq \Lambda_1 \leq \Lambda, \quad |m| \leq M, \quad \|A\|_2 \leq \Lambda, \quad \|H(\|G\|_F^2)\|_2 \leq \Lambda, \quad \|\partial_i \nabla G\|_2 \leq \Lambda. \] (49)

If the second assumption holds true, we call the optimal solution for the linear part:
\[ u_A^* = \min_u \|y - Au\|_\Gamma, \] (50)
and the associated residue
\[ r = y - Au_A^*. \] (51)

It is then automatic that
\[ \Gamma^{-1/2}r \perp \Gamma^{-1/2}Au, \quad \forall u \in \mathbb{R}^L. \] (52)

We also define the Gaussian part of the distribution
\[ \rho_A(u, t) = \frac{1}{Z(t)} \exp \left( -\frac{1}{2} |Au_A^* - Au|_\Gamma^2 - \frac{1}{2} |u - u_0|_{\Gamma_0}^2 \right), \] (53)
so that we have
\[ \rho(u, t) \propto \rho_A(u, t) \exp \left( -\frac{1}{2} |r - m(u)|_{\Gamma_1}^2 \right). \]

This $\rho_A$ has expectation and the covariance matrix:
\[ u_A(t) = (tA^\top \Gamma^{-1}A + \Gamma_0^{-1})^{-1} (tA^\top \Gamma^{-1}Au_A^* + \Gamma_0^{-1}u_0), \] (54)
\[ \text{Cov}_A(t) = (tA^\top \Gamma^{-1}A + \Gamma_0^{-1})^{-1}. \]

It is immediate that the second assumption is stronger than the first one, and thus one would expect a tighter bound. Indeed, by comparing Theorem 4.5 and Theorem 4.6, we see that the variance of weights is bounded by a constant that exponentially grows with respect to $|y|$, the data, when only Assumption 4.1 holds true, but is bounded by a constant independent of $|y|$ when Assumption 4.2 also holds.

Since EnKI is a more popular method than EnSRF, the analysis is conducted on WEnKI mainly. Similar analysis could potentially be applied to deal with WEnSRF but could be more delicate. We do not pursue it in this paper.

We also note that the analysis is conducted on the dynamics (41) with coefficients calculated from the exact density, and thus each particle is evolved independently (this is in the spirit of the McKean-Vlasov dynamics or propagation of chaos, expected in the mean field limit). The analysis for the numerical version, with all the covariance matrices replaced by the ensemble ones as seen in (36)-(37) will be left for future works.

### 4.2.1. Bounded nonlinearity under assumption 4.1.

We first prove a lemma to bound covariance matrix of $\varrho(u, t)$.

**Lemma 4.3.** Under Assumption 4.1,
\[ \mathbb{E}^\varrho(t) (|y - G|_F^2) \]
declines in $t$, where $\varrho(u, t)$ is the solution to (38).
Proof. By Theorem 4.2, $\varrho = \rho(u, t)$ defined in (24) is a strong solution to the PDE. Taking partial derivative with respect to $t$ and rewriting (26), we get

$$
\partial_t \varrho = -\frac{1}{2} \left\{ |y - G|^2 - \mathbb{E}^{(t)} \left( |y - G|^2 \right) \right\} \varrho.
$$

(55)

Multiplying $|y - G|^2$ on both sides and taking integral yields

$$
\frac{d}{dt} \mathbb{E}^{(t)} |y - G|^2 = -\frac{1}{2} \left( \mathbb{E}^{(t)} \left( |y - G|^4 \right) - \left( \mathbb{E}^{(t)} \left( |y - G|^2 \right) \right)^2 \right) \leq 0,
$$

which concludes the lemma.

\[ \square \]

Lemma 4.4. Under Assumption 4.1, there exists a finite constant $C$ depending on $\Lambda$, $|u_0|$, $\|\Gamma_0^{-1}\|_2$, $\|\Gamma^{-1}\|_2$ and $|y|$ only such that for $0 \leq t \leq 1$:

$$
\mathbb{E}^{(t)} |u|^2 < C, \quad \mathbb{E}^{(t)} |G(u)|^2 < C.
$$

(56)

In the proof below, we use $C$ to denote a generic constant that changes from line to line, and we keep track of the constant’s dependence on different argument. However, we do not specify the form of the dependence.

Proof. Consider

$$
\varrho(u, 0) = \rho_{\text{prior}} = \exp \left( -|u - u_0|^2 \right),
$$

we expand $G$ around $\bar{0}$ and utilize the bound (47) for:

$$
\mathbb{E}^{(0)} |G(u)|^2 \leq \Lambda^2 \mathbb{E}^{(0)} |u|^2 + |G(\bar{0})|^2 \leq \Lambda^2 (|u_0|^2 + \text{Tr}(\Gamma_0)) + |G(\bar{0})|^2.
$$

Therefore,

$$
\begin{align*}
\mathbb{E}^{(0)} \left( |y - G|^2 \right) & \leq 2 \|\Gamma^{-1}\|_2 \left( |y|^2 + \mathbb{E}^{(0)} |G(u)|^2 \right) \\
& \leq 2 \|\Gamma^{-1}\|_2 \left( |y|^2 + \Lambda^2 (|u_0|^2 + \text{Tr}(\Gamma_0)) + |G(\bar{0})|^2 \right) \\
& =: C_1 |y|^2 + C_2,
\end{align*}
$$

(57)

where the last line defines constants $C_1$ and $C_2$, which only depend on $\Lambda$, $|u_0|$, $\|\Gamma_0^{-1}\|_2$, $\|\Gamma^{-1}\|_2$.

Multiplying $|G(u)|^2$ and $|u|^2$ on both sides of (55), we get

$$
\begin{align*}
\frac{d}{dt} \mathbb{E}^{(t)} |G(u)|^2 & = -\frac{1}{2} \int \left\{ |y - G|^2 - \mathbb{E}^{(t)} \left( |y - G|^2 \right) \right\} |G(u)|^2 \rho(t) \, du \\
& \leq -\frac{1}{2} \int \mathbb{E}^{(t)} \left( |y - G|^2 \right) |G(u)|^2 \rho(t) \, du \\
& = -\frac{1}{2} \mathbb{E}^{(t)} \left( |y - G|^2 \right) \mathbb{E}^{(t)} |G(u)|^2 \\
& \leq -\frac{1}{2} \mathbb{E}^{(0)} \left( |y - G|^2 \right) \mathbb{E}^{(t)} |G(u)|^2 \\
& \leq \frac{1}{2} \left( C_1 |y|^2 + C_2 \right) \mathbb{E}^{(t)} |G(u)|^2, \tag{57}
\end{align*}
$$

(57)
and
\[
\begin{align*}
\frac{d}{dt} \mathbb{E}^{(t)}|u|^2 &= -\frac{1}{2} \int \left\{ |y - G|^2 - \mathbb{E}^{(t)} \left( |y - G|^2 \right) \right\} |u|^2 \rho \, du \\
&\leq \frac{1}{2} \mathbb{E}^{(t)} \left( |y - G|^2 \right) \mathbb{E}^{(t)}|u|^2 \\
&\leq \frac{1}{2} \mathbb{E}^{(0)} \left( |y - G|^2 \right) \mathbb{E}^{(t)}|u|^2 \\
&\leq (C_1|y|^2 + C_2) \mathbb{E}^{(t)}|u|^2,
\end{align*}
\]
where we use Lemma 4.3 in the second inequalities. By Grönwall inequality, we have:
\[
\mathbb{E}^{(t)}|u|^2 \leq \mathbb{E}^{(0)}|u|^2 e^{(C_1|y|^2 + C_2)t} \leq (|u_0|^2 + \text{Tr}(\Gamma_0)) e^{(C_1|y|^2 + C_2)t},
\]
and
\[
\mathbb{E}^{(t)}|G(u)|^2 \leq \mathbb{E}^{(0)}|G(u)|^2 e^{(C_1|y|^2 + C_2)t} \leq \left( \Lambda^2 (|u_0|^2 + \text{Tr}(\Gamma_0)) + |G(0)|^2 \right) e^{(C_1|y|^2 + C_2)t}.
\]
(58)
Choose $C$ to be the bigger value of the two with $t = 1$, we conclude the lemma. □

The immediate consequence of Lemma 4.3 and Lemma 4.4 is the boundedness of the covariance matrices:

**Corollary 4.1.** Under Assumption 4.1, there exists a constant $C$ depending on $\Lambda$, $|u_0|$, $\|\Gamma_0^{-1}\|_2$, $\|\Gamma^{-1}\|_2$ and $|y|$ such that for $0 \leq t \leq 1$
\[
\|\text{Cov}^{(t)}_{uu}\|_2 \leq C, \quad \|\text{Cov}^{(t)}_{up}\|_2 \leq C, \quad \|\text{Cov}^{(t)}_{pp}\|_2 \leq C,
\]
(59)
where $\text{Cov}^{(t)}_{uu}$, $\text{Cov}^{(t)}_{up}$, $\text{Cov}^{(t)}_{pp}$ are the corresponding covariance matrices, as defined in (33).

These a priori estimates are now used to bound the variance of the weights.

**Theorem 4.5.** Under Assumptions 4.1, let $\{u^n_i, \omega^n_i\}_{i=1}^N$ solve (41). Then there exists a constant $C$ only depending on $\Lambda$, $|u_0|$, $\|\Gamma_0^{-1}\|_2$, $\|\Gamma^{-1}\|_2$ and $|y|$ such that for any $0 \leq t \leq 1$.
\[
\text{Var}(N\omega^n_i) \leq C.
\]

**Remark 3.** We note that the result in Theorem 4.5 is not optimal. The constant, if traced carefully, blows up as $|y| \to \infty$ with a rate of at least $e^{|y|^2}$, as suggested in (58). Essentially this result does not demonstrate WEnKI superior than the classical IS. However, as will be shown in Theorem 4.6, under a stronger assumption (Assumption 4.2), the dependence on $y$ could be removed.

**Proof.** Note that
\[
\text{Var}(N\omega^n_i) = \mathbb{E}(N\omega^n_i - 1)^2 = N^2 \left( \mathbb{E}|\omega^n_i|^2 - \frac{1}{N^2} \right),
\]
thus to prove the theorem, it suffices to show that
\[
N^2\mathbb{E}|\omega^n_i|^2 \leq C
\]
(60)
with $C$ depending on $\Lambda$, $|u_0|$, $\|\Gamma_0^{-1}\|_2$, $\|\Gamma^{-1}\|_2$ and $|y|$.
Multiplying $\omega^n$ on both sides of the second equation of (41) and taking expectation, we have
\[
\frac{d}{dt} \mathbb{E}[|\omega^n|^2] \leq 2 \mathbb{E} \left\{ |R_1(u^n, t) + R_2(u^n, t) + R_3(u^n, t)| |w^n|^2 \right\}
\leq 2 \left( \| R_1 \|_\infty + \frac{1}{2} (y - \overline{\mathcal{G}}^t)^\top \mathcal{G}^{-1} (y - \overline{\mathcal{G}}^t) + \| R_3 \|_\infty \right) \mathbb{E}[|\omega^n|^2],
\] (61)
where we have omitted the last three terms in $R_2$ because the sum of them is negative. We then bound the three terms in bracket separately. As a preparation, we note that
\[
\text{Tr} \left\{ \text{Cov}^{(t)}_{pp} \mathcal{G}^{-1} \right\} = \int (\mathcal{G}(u) - \overline{\mathcal{G}})^\top \mathcal{G}^{-1} (\mathcal{G}(u) - \overline{\mathcal{G}}) \varphi(u, t) \, du \leq \mathbb{E}^{(t)} |\mathcal{G}(u) - \overline{\mathcal{G}}|^2 \|\mathcal{G}^{-1}\|_2,
\]
and
\[
\text{Tr} \left\{ \text{Cov}^{(t)}_{up} \mathcal{G}^{-1} \right\} = \int (u - \overline{u})^\top \mathcal{G}^{-1} (\mathcal{G}(u) - \overline{\mathcal{G}}) \varphi(u, t) \, du \leq \left( \mathbb{E}^{(t)} |u - \overline{u}| |\mathcal{G}(u) - \overline{\mathcal{G}}| \right) \|\mathcal{G}^{-1}\|_2.
\]
Apply these inequalities to estimate $R_k$ defined in (40), we arrive at the following bounds.
\[
\| R_1(u, t) \|_2 \leq \frac{\| \mathcal{G}^{-1} \|_2}{2} \left\{ \mathbb{E}^{(t)} |\mathcal{G}(u) - \overline{\mathcal{G}}^{(t)}|^2 + \left[ 2 \mathcal{A} + \| \text{Cov}^{(t)}_{up} \|_2 (t \| \mathcal{G}^{-1} \|_2 \mathcal{A}^2 + \| \mathcal{G}^{-1}_0 \|_2) \right] \times \left( \mathbb{E}^{(t)} |u - \overline{u}^{(t)}| |\mathcal{G}(u) - \overline{\mathcal{G}}^{(t)}| \right) \right\}
\]
(62)
\[
\leq \frac{\| \mathcal{G}^{-1} \|_2}{2} \left\{ \text{Tr}(\text{Cov}^{(t)}_{pp}) + \left[ 2 \mathcal{A} + \| \text{Cov}^{(t)}_{up} \|_2 (t \| \mathcal{G}^{-1} \|_2 \mathcal{A}^2 + \| \mathcal{G}^{-1}_0 \|_2) \right] \times \text{Tr}(\text{Cov}^{(t)}_{pp})^{1/2} \text{Tr}(\text{Cov}^{(t)}_{up})^{1/2} \right\}
\]
\[
\leq C,
\]
where the last inequality comes from Corollary 4.1.

For the non-negative contribution from $R_2$, we have
\[
(y - \overline{\mathcal{G}}^{(t)})^\top \mathcal{G}^{-1} (y - \overline{\mathcal{G}}^{(t)}) \leq \| \mathcal{G}^{-1} \|_2 \| y - \overline{\mathcal{G}}^{(t)} \|^2 \leq 2 \| \mathcal{G}^{-1} \|_2 (|y|^2 + \mathbb{E}^{(t)} |\mathcal{G}(u)|^2) \leq C,
\] (63)
where the last inequality comes from Lemma 4.4.

Finally, for $R_3$, we have
\[
\| R_3(u, t) \|_2 \leq \frac{1}{2} t \| \mathcal{G}^{-1} \|_2 \| \text{Cov}^{(t)}_{up} \|_2 \| \mathcal{W}(u) \|_2 (\mathbb{E}^{(t)} |u - \overline{u}^{(t)}| |\mathcal{G}(u) - \overline{\mathcal{G}}^{(t)}|)
\]
\[
\leq \frac{1}{2} t \| \mathcal{G}^{-1} \|_2 \| \text{Cov}^{(t)}_{up} \|_2 \| \mathcal{W}(u) \|_2 \text{Tr}(\text{Cov}^{(t)}_{pp})^{1/2} \text{Tr}(\text{Cov}^{(t)}_{up})^{1/2}
\]
\[
\leq C,
\] (64)
where we have used Corollary 4.1, and that, by definition of \( W \),
\[
\|W(u)\|_2 \leq \|W(u)\|_F \\
\leq \frac{L}{2} (\|\mathcal{H}(\|\mathcal{G}\|_F^2)\|_2 + \|\Gamma^{-1}\|_2 \|\nabla\mathcal{G}\|_2^2) + \|y\| \|\Gamma^{-1}\|_2 \max_{1 \leq t \leq L} \{\|\partial_t \nabla \mathcal{G}\|_2\} \leq C. 
\] (65)

All the constants above depend on \( \Lambda, |u_0|, \|\Gamma^{-1}\|_2, \|\Gamma^{-1}\|_2 \) and \( |y| \). Substitute these into (61), we have
\[
dE|\omega_t^n|^2 \leq C E|\omega_t^n|^2.
\]
Realizing that \( \omega_t^n = \frac{1}{N} \) so that \( E|\omega_t^n|^2 = \frac{1}{N^2} \), we obtain
\[
E|\omega_t^n|^2 \leq \frac{e^{Ct}}{N^2}.
\]
This concludes (60) and this theorem. \( \Box \)

4.2.2. Weak nonlinearity under assumption 4.2. The variance bound can be improved when we assume further structure of the nonlinearity, namely, when the nonlinear component \( m(u) \) is perpendicular to the range of the linear component \( A \), weighted by \( \Gamma^{-1/2} \). In particular, the bound becomes independent of \( y \), as shown in the following theorem.

**Theorem 4.6.** Under Assumption 4.2, there exists a finite constant \( C \) depending on \( \Lambda_1, \Lambda, \|\Gamma^{-1}\|_2, \|\Gamma^{-1}\|_2, |r|, M, \) and \( |u_A^*| \), such that
\[
\|\text{Var}(N\omega_t^n)\|_{L^\infty[0,1]} \leq C.
\] (66)
Furthermore,
\[
\lim_{\Lambda_1 \to 0} C \leq C_1, \tag{67}
\]
where \( C_1 \) only depends on \( \Lambda, \|\Gamma^{-1}\|_2, \|\Gamma^{-1}\|_2 \).

**Remark 4.** This theorem is a counterpart of Theorem 4.5, but stronger assumption on the nonlinearity is added. As a result, the variance of weight is bounded, independent of \( y \). In the most extreme case, suppose \( \mathcal{G} \) is entirely linear, \( \Lambda_1 = 0 \), then according to the theorem, the variance is bounded by a fixed constant. As a comparison, if one applies Important Sampling directly, for large \( y \) and thus large \( u_* \), the variance blows up at the order of \( \mathcal{O}(e^{\|u^*\|^2}) \), equivalently to \( \mathcal{O}(e^{\|w^*\|^2}) \) for reasonably conditioned \( A \). This means that under mild conditions (Assumption 4.2), the newly proposed WEnKI method significantly reduces the weight variance from the classical method IS.

The proof of the theorem is largely based on the following calculation.

**Proposition 1.** Under Assumption 4.2, let \( \{u_n^*, \omega_t^n\}_{n=1}^N \) solve (41), we have
\[
\frac{d}{dt} \left( \frac{E|\omega_t^n|^2(N\omega_t^n)^2}{E|\omega_t^n|^2(N\omega_t^n)^2} \right) \leq CW(t) \left( \frac{E|\omega_t^n|^2(N\omega_t^n)^2}{E|\omega_t^n|^2(N\omega_t^n)^2} \right), \quad \forall 0 \leq t \leq 1. \tag{68}
\]
where \( W(t) \) is a \( 2 \times 2 \) matrix defined by
\[
W_{1,1}(t) = C \left( \text{Var}^\theta(t)(u)^2 + \text{Var}^\theta(t)(u) + |u_A^*| \text{Var}^\theta(t)(u) + |u_A^* - \bar{w}^\theta(t)|^2 + 1 \right),
\]
\[
W_{1,2}(t) = C |\text{Var}^\theta(t)(u)| \left[ \text{Var}^\theta(t)(u) + |u_A^*| + 1 \right],
\]
\[
W_{2,1}(t) = C \left( \|\bar{w}^\theta(t) - u_A^*\|_2 \left| I - (\text{Cov}_A)^{-1}\text{Cov}_{u_A^*}^\theta(t) \right|_2 + \|\text{Cov}_{u_A^*}^\theta(t)\|_2 \right),
\]
\[
W_{2,2}(t) = C \left( \|\bar{w}^\theta(t) - u_A^*\|_2 \left| I - (\text{Cov}_A)^{-1}\text{Cov}_{u_A^*}^\theta(t) \right|_2 + \|\text{Cov}_{u_A^*}^\theta(t)\|_2 \right).
\]
and

\[
W_{2,2}(t) = C \left[ \left\| \overline{\rho}(t) - u_A \right\| \left( \left\| \overline{\rho}(t) - \text{Cov}_{u,u}(\text{Cov}_A(t))^{-1} u_A \right\| \right. \\
+ \left\| I - \text{Cov}_A^{-1} \text{Cov}_{u,u} \right\|_2 + \left\| \text{Cov}_{u,u} \right\|_2 \Lambda_1 \right) \\
+ (\text{Var}_t^u(u))^2 + \text{Var}_t^u(u) + \left\| \text{Cov}_{m,u} \left( |u_A| + \Lambda_1 + 1 \right) \right\|, \\
\right]
\]

where \( \text{Var}_t^u(u) = \text{Tr} \left( \text{Cov}_t^u \right) \) and \( C \) is a constant depending on \( \Lambda, \left\| \Gamma^{-1} \right\|_2, \left\| \Gamma^{-1} \right\|_2, |r|, \) and \( M. \)

The proof for the proposition is deferred to Appendix B. We now give the proof for Theorem 4.6 based on the above proposition.

Proof of Theorem 4.6. For fixed \( 1 \leq n \leq N \), let

\[
p(t) = \mathbb{E} \left| u^n_t \right|^2 (N \omega^n_t)^2, \quad q(t) = \text{Var}(N \omega^n_t) + 1.
\]

Since \( \omega^n_0 = \frac{1}{N} \),

\[
p(0) = \mathbb{E}_{\rho_{\text{prior}}} \left| u \right|^2, \quad q(0) = 1.
\]

According to Proposition 1,

\[
\frac{d}{dt} \left( \begin{array}{c} p(t) \\ q(t) \end{array} \right) \leq W(t) \left( \begin{array}{c} p(t) \\ q(t) \end{array} \right),
\]

which implies (66). If \( \Lambda_1 \to 0 \), nonlinear function \( m(u) \) is almost a constant. Therefore, we also have \( \overline{\rho}(t) \to u_A(t) \), \( \text{Cov}_{u,u}(t) \to \text{Cov}_A(t) \) and \( \left\| \text{Cov}_{m,u} \right\|_2 \to 0 \), then the coefficients for \( q \) satisfy:

\[
\lim_{\Lambda_1 \to 0} W_{2,1}(t) = 0, \quad \lim_{\Lambda_1 \to 0} W_{2,2}(t) = (\text{Tr}(\text{Cov}_A))^2 + \text{Tr}(\text{Cov}_A).
\]

Then (67) is a direct consequence, concluding the theorem.

4.3. EnKI with nonlinear forward map. In this section, we study a slightly different topic: how different are WEnKI and EnKI? In fact, it was proved in [12] that EnKI is not a consistent sampling method when the forward map is nonlinear. The algorithm, without the weight, can be regarded as the discrete version of PDE (16), but the target distribution \( \rho(u,t) \) is not the solution to the PDE, and hence EnKI is inconsistent.

It is numerically observed, however, that despite being inconsistent, EnKI mysteriously performs rather well [32], especially when the target distribution is almost Gaussian-like, no matter how nonlinear \( G \) is, also see the book [33] for more examples. To the best of our knowledge, such discrepancy in terms of theoretical and practical performance, has not been addressed in literature. In this subsection, as a first attempt to explain it, we provide one criterion, under which, EnKI performs similarly well as WEnKI.

The argument in the end comes down to comparing the continuous version of WEnKI and EnKI, two Fokker-Planck equations, with the former one having a weight term while the latter not.

Once again we denote \( \rho \) the target distribution, defined in (24) and proved to be the solution to equation (38) in Theorem 4.2, and let \( \varrho \) the solution to the
Proposition 2. Under Assumption 4.1, there exists a constant $\Lambda$, such that

$$d = \rho$$

where $\text{Cov}_{\rho}^{(t)}$ and $\text{Cov}_{\rho}^{(t)}$ are covariance of $(u, \mathcal{G})$ and $(\mathcal{G}, u)$ in $\rho(u, t)$. It was proved in [12] that (69) is the mean-field limit of EnKI.

We will now show that $\rho$ and $\varrho$ are close when the weight term (defined in (40))

$$\mathcal{W}(u, t) = \mathcal{R}_1(u, t) + \mathcal{R}_2(u, t) + \mathcal{R}_3(u, t)$$

is small. This means that WEnKI and EnKI give more or less the same results when the weight term is small. We recall the bounded Lipschitz metric ($d_{BL}$) between probability measures:

$$d_{BL}(\mu, \nu) = \sup_{f \in \text{Lip}(\mathbb{R}^L)} \left| \int_{\mathbb{R}^L} f d\mu - \int_{\mathbb{R}^L} f d\nu \right|,$$

where

$$\text{Lip}(\mathbb{R}^L) = \left\{ f \in C_b : \sup_x |f(x)| \leq 1, \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|} \leq 1 \right\}.$$  

Since the admissible set in the supremum is smaller than the class of Lipschitz-1 function and 1-bounded function, this metric can be bounded by $L^2$-Wasserstein distance $W_2(\mu, \nu)$ and total variation $TV(\mu, \nu)$

$$d_{BL}(\mu, \nu) \leq W_2(\mu, \nu), \quad d_{BL}(\mu, \nu) \leq TV(\mu, \nu). \quad (70)$$

We have the following theorem characterizing the difference between $\varrho$ and $\rho$, i.e., EnKI and WEnKI (that is consistent to the target distribution).

**Theorem 4.7.** Under Assumption 4.1, there exists a constant $C$ depending on $\Lambda$, $|u_0|$, $\|\Gamma_0^{-1}\|_2$, $\|\Gamma_1^{-1}\|_2$, $|y|$, such that

$$d_{BL}(\rho(u, t) du, \rho(u, t) du) \leq C \int_0^1 \int (1 + |u|^2)|\mathcal{W}|\rho du ds \quad (71)$$

for all $0 \leq t \leq 1$.

This theorem states that the size of the weight gives control over the distance between $\rho$ and $\varrho$. To compare them, we introduce an intermediate surrogate $\tilde{\rho}$, given by

$$\begin{cases}
\partial_t \tilde{\rho}(u, t) + \nabla_u \cdot \left( (y - \mathcal{G}(u))^\top \Gamma^{-1} \text{Cov}_{\rho}^{(t)} \tilde{\rho} \right) = \frac{1}{2} \text{Tr} \left( \text{Cov}_{\rho}^{(t)} \Gamma^{-1} \text{Cov}_{\rho}^{(t)} H_u \tilde{\rho} \right), \\
\tilde{\rho}(u, 0) = \rho_{\text{prior}}
\end{cases} \quad (72)$$

where $\text{Cov}_{\rho}^{(t)}$ and $\text{Cov}_{\rho}^{(t)}$ are given by $\rho(u, t)$. We will bound $d_{BL}(\rho, \tilde{\rho})$ and $d_{BL}(\tilde{\rho}, \varrho)$ in the following two propositions. The theorem is a direct consequence of the two.

**Proposition 2.** Under Assumption 4.1, there exists a constant $C$ depending on $\Lambda$, $|u_0|$, $\|\Gamma_0^{-1}\|_2$, $\|\Gamma_1^{-1}\|_2$ and $|y|$, such that

$$d_{BL}(\tilde{\rho}(u, t) du, \rho(u, t) du) \leq TV(\tilde{\rho}(u, t) du, \rho(u, t) du) \leq C \int_0^1 \int |\mathcal{W}|\rho du ds \quad (73)$$

for all $0 \leq t \leq 1$.  

Proposition 3. Under Assumption 4.1, there exists a constant \( C \) depending on \( \Lambda \), \( |u_0| \), \( \|\Gamma^{-1}\|_2 \), \( \|\Gamma^{-1}\|_2 \) and \( |y| \), such that

\[
d_{\text{BL}}(\varrho(u,t) \, du, \tilde{\rho}(u,t) \, du) \leq W_2(\varrho(u,t) \, du, \tilde{\rho}(u,t) \, du) \leq C \int_0^1 \int (1+|u|^2)|\mathcal{W}||\rho| \, du \, ds
\]

(74)

for all \( 0 \leq t \leq 1 \).

Proof of Proposition 2. The proof is based on the following construction of particle system. Let

\[
\begin{aligned}
\begin{cases}
  du_t &= \text{Cov} \rho_{\varrho(t)} \Gamma^{-1} (y - \mathcal{G}(u_t)) \, dt + \text{Cov} \rho_{\varrho(t)}^{\frac{1}{2}} \, dW_t, \\ 
  dw_t &= \mathcal{W}(u,t) w_t \, dt,
\end{cases}
\end{aligned}
\]

(75)

with initial data \( u_0 \) sampled from \( \mu_{\text{prior}} \) and \( w_0 = 1 \). This is a Langevin dynamics, so that for any test function \( f \):

\[
E(f(u_t)) = E_{\tilde{\rho}(t)} f, \quad E(w_t f(u_t)) = E_{\rho(t)} f.
\]

It is clear from second equality in (75):

\[
w_t > 0
\]

for all \( t \), and that

\[
d|w_t - 1| \leq |dw_t - 1| \leq |\mathcal{W}(u_t,t)| w_t \, dt.
\]

This means

\[
\frac{d}{dt} E(|w_t - 1|) \leq E(|\mathcal{W}(u_t,t)| w_t) = \int |\mathcal{W}| \rho \, du
\]

and

\[
E(|w_t - 1|) \leq \int_0^1 \int |\mathcal{W}| \rho \, du \, dt, \quad \forall 0 \leq t \leq 1.
\]

The boundedness (73) is a direct result by using \( L^\infty \) test function to bound total variation:

\[
\text{TV}(\tilde{\rho} \, du, \rho \, du) \leq \sup_{\|f\|_\infty = 1} \int f(\tilde{\rho} - \rho) \, du \leq \sup_{\|f\|_\infty = 1} |E(w_t - 1) f(u_t)|
\]

\[
\leq \sup_{\|f\|_\infty = 1} \|f\|_\infty E|w_t - 1|
\]

\[
\leq \int_0^1 \int |\mathcal{W}| \rho \, du \, ds.
\]

\[
\Box
\]

Proof of Proposition 3. We first state and prove an estimate of the difference of covariance

\[
\|\text{Cov}_{\varrho(t)} - \text{Cov}_{\tilde{\varrho}(t)}\|_2 \leq C \int_0^1 \int (1+|u|^2)|\mathcal{W}||\rho| \, du \, ds
\]

(76)

for all \( 0 \leq t \leq 1 \). For this, we first bound \( E|u_t|^2|w_t - 1| \) using Itô’s formula and (75):

\[
d|u_t|^2|w_t - 1| = 2 \langle du_t, u_t \rangle |w_t - 1| + \langle du_t, du_t \rangle |w_t - 1| + |u_t|^2 d|w_t - 1|.
\]
Taking expectation on both sides, we have
\[
\begin{align*}
d\mathbb{E}[u_t^2] & \leq \mathbb{E} \left[ \text{Cov}_{up}^{\rho(t)} \Gamma^{-1} (y - \mathcal{G}(u_t)) , u_t \right] |w_t - 1| dt + C \mathbb{E}[w_t - 1] dt \\
& \quad + \mathbb{E} [u_t^2] |\mathcal{W}(u_t, t)| w_t dt \\
& \leq C \mathbb{E} \left[ (|u_t|^2 + |u_t|)|w_t - 1| \right] dt + C \int_0^1 \int (1 + |u|^2) |\mathcal{W}| \rho \, du \, ds \\
& \leq C \mathbb{E} \left[ |u_t|^2 |w_t - 1| \right] + C (\mathbb{E}[w_t - 1])^{1/2} (\mathbb{E}[|u_t|^2 |w_t - 1|])^{1/2} \\
& \quad + C \int_0^1 \int (1 + |u|^2) |\mathcal{W}| \rho \, du \, ds \\
& \leq C \mathbb{E}|u_t|^2 |w_t - 1| \\
& \quad + C \left( \int_0^1 \int (1 + |u|^2) |\mathcal{W}| \rho \, du \, dt \right)^{1/2} (\mathbb{E}|u_t|^2 |w_t - 1|)^{1/2} \\
& \quad + C \int_0^1 \int (1 + |u|^2) |\mathcal{W}| \rho \, du \, ds,
\end{align*}
\]
where we use Corollary 4.1 and equation (59).

By Grönwall’s inequality and \( w_0 = 1 \), we get
\[
\left\| \int u \otimes u(\tilde{\rho} - \rho) \, du \right\|_2 \leq \mathbb{E} \left[ |u_t|^2 |w_t - 1| \right] \leq C \int_0^1 \int (1 + |u|^2) |\mathcal{W}| \rho \, du \, ds \tag{77}
\]
and
\[
\left\| \int u(\tilde{\rho} - \rho) \, du \right\|_2 \leq \mathbb{E} |u_t| |w_t - 1| \leq (\mathbb{E}[w_t - 1])^{1/2} (\mathbb{E}|u_t|^2 |w_t - 1|)^{1/2} \tag{78}
\]
for any \( t \leq 1 \). Combining (77) and (78), we have
\[
\|\text{Cov}_{up}^{\rho(t)} - \text{Cov}_{up}^{\tilde{\rho}(t)}\|_2 \leq \left\| \int u \otimes u(\tilde{\rho} - \rho) \, du \right\|_2 + \left\| \int u(\tilde{\rho} - \rho) \, du \right\|_2 \left\| \int u(\tilde{\rho} + \rho) \, du \right\|_2 \leq C \int_0^1 \int (1 + |u|^2) |\mathcal{W}| \rho \, du \, ds,
\]
which proves (76).

We now come back to the Proposition to prove (74). We use two particle systems to represent (72) and (69). Let
\[
\begin{align*}
dv_t & = \text{Cov}_{up}^{\rho(t)} \Gamma^{-1} (y - \mathcal{G}(u_t)) \, dt + \text{Cov}_{up}^{\rho(t)} \Gamma^{-1/2} \, dW_t, \\
dv_t & = \text{Cov}_{up}^{\rho(t)} \Gamma^{-1} (y - \mathcal{G}(u_t)) \, dt + \text{Cov}_{up}^{\rho(t)} \Gamma^{-1/2} \, dW_t
\end{align*}
\]
where the initial data \( u_0 \) is sampled from \( \rho_{\text{prior}}(u) \), and let
\[
\begin{align*}
dv_t & = \text{Cov}_{up}^{\rho(t)} \Gamma^{-1} (y - \mathcal{G}(v_t)) \, dt + \text{Cov}_{up}^{\rho(t)} \Gamma^{-1/2} \, dW_t \\
W_2(\tilde{\rho}, \rho) & \leq (\mathbb{E}|u_t - v_t|^2)^{1/2}.
\end{align*}
\]
To show the theorem, it suffices to prove
\[
\mathbb{E}|u_t - v_t|^2 \leq C \sup_{t \in [0,1]} \|\text{Cov}_{up}^{\rho(t)} - \text{Cov}_{up}^{\tilde{\rho}(t)}\|_2 \tag{81}
\]
for all \( t \) and then utilize (76).
Let $\gamma_t = u_t - v_t$, one subtracts (80) from (79) and uses Itô’s formula to obtain
\[
\frac{d}{dt} \mathbb{E}|\gamma_t|^2 \leq \mathbb{E} \left\langle \text{Cov}_{up}^t \Gamma^{-1} (y - \mathcal{G}(u_t)) - \text{Cov}_{up}^t \Gamma^{-1} (y - \mathcal{G}(v_t)) , \gamma_t \right\rangle dt \\
+ \frac{1}{2} \text{Tr} \left\{ \left( \text{Cov}_{up}^t - \text{Cov}_{up}^t \right) \Gamma^{-1} \left( \text{Cov}_{up}^t - \text{Cov}_{up}^t \right) \right\} dt
\]
\[
= \mathbb{E} \left\langle \left( \text{Cov}_{up}^t - \text{Cov}_{up}^t \right) \Gamma^{-1} (y - \mathcal{G}(u_t)) , \gamma_t \right\rangle dt \\
- \mathbb{E} \left\langle \text{Cov}_{up}^t \Gamma^{-1} (\mathcal{G}(u_t) - \mathcal{G}(v_t)) , \gamma_t \right\rangle dt \\
+ \frac{1}{2} \text{Tr} \left\{ \left( \text{Cov}_{up}^t - \text{Cov}_{up}^t \right) \Gamma^{-1} \left( \text{Cov}_{up}^t - \text{Cov}_{up}^t \right) \right\} dt
\]
\[
\leq C||\text{Cov}_{up}^t - \text{Cov}_{up}^t||_2 (|y|^2 + \mathbb{E}|u_t|^2)^{1/2} (\mathbb{E}|\gamma_t|^2)^{1/2} + C||\text{Cov}_{up}^t||_2 \mathbb{E}|\gamma_t|^2 \\
+ C||\text{Cov}_{up}^t - \text{Cov}_{up}^t||_2^2.
\]
Since
\[
||\text{Cov}_{up}^t - \text{Cov}_{up}^t||_2 \leq (\mathbb{E}|\gamma_t|^2)^{1/2},
\]
we have
\[
||\text{Cov}_{up}^t - \text{Cov}_{up}^t||_2 \leq ||\text{Cov}_{up}^t - \text{Cov}_{up}^t||_2 + ||\text{Cov}_{up}^t - \text{Cov}_{up}^t||_2 \\
\leq ||\text{Cov}_{up}^t - \text{Cov}_{up}^t||_2 + C(\mathbb{E}|\gamma_t|^2)^{1/2}.
\]
Therefore
\[
\frac{d}{dt} \mathbb{E}|\gamma_t|^2 \leq C\mathbb{E}|\gamma_t|^2 + C||\text{Cov}_{up}^t - \text{Cov}_{up}^t||_2 (\mathbb{E}|\gamma_t|^2)^{1/2} + ||\text{Cov}_{up}^t - \text{Cov}_{up}^t||_2.
\]
Since $\gamma_0 = 0$, by Grönwall’s inequality, we finally arrive at
\[
\mathbb{E}|\gamma_t|^2 \leq C||\text{Cov}_{up}^t - \text{Cov}_{up}^t||_2 \|_{L^\infty[0,1]}
\]
for all $t < 1$, which proves (81), concluding the proposition.

5. Numerical results. In this section, we show some numerical evidence to demonstrate the superiority of the proposed method. All numerical examples are highly nonlinear, so we are away from the known “safe zone” where EnKI and EnSRF work both perfect. We remark that we conduct numerical experiments only in low dimensional setting to have a clear illustration of the behavior of the algorithm.

5.1. One dimension example. As a start, we first test out the 1D case. We set the normal distribution $\mathcal{N}(0,1)$ as the prior distribution.

- Example 1: In this example we set $\mathcal{G}(u) = 4 \cos(2(u - 3)) + \sin(u - 3)$ and the data (with only one observation) is given at $y = 0$. The posterior distribution is a multimodal distributions, as shown in Figure 1. The number of samples is set to be $N = 1000$, and in WEnKI and WEnSRF, we choose the time step $\Delta t = 10^{-3}$. As a comparison, we plot the result using WEnKF (Remark 1) and Important Sampling, EnKI and EnSRF. In this example, the prior and the posterior distributions share supports, so IS and WEnKF, the two methods that achieve consistency, behave relatively well. But due to nonlinearity and non-Gaussianity, EnKI and EnSRF, the two methods that tend to give one-mode Gaussian-like profile, fail.

- Example 2: This is a highly nonlinear example with 4-th power in $\mathcal{G}$: $\mathcal{G}(u) = (u - 3)^4 - 1$, and data is still set to be $y = 0$. $N = 2000$ and $\Delta t = 10^{-5}$. WEnKI and WEnSRF clearly outperform the others, see Figure 2. Note that in the experiment
we find that for stability of the Euler solver (34), and (41), the time step is chosen to be rather small.

- Example 3: In this example we set \( G(u) = (u - 5)^2 \) and data \( y = 0 \). \( N = 2000 \) and \( \Delta t = 10^{-3} \). The posterior distribution has one peak, but is non-Gaussian. While the center of the prior is at 0, the center of the likelihood function is at \( u = 5 \): so there is a big shift of support from the prior to the posterior distribution. As seen in Figure 3, both WEnKI and WEnSRF still capture the posterior distribution rather well. EnKI and EnSRF cannot capture the entire profile, but at least can move to fit relatively accurate support. WEnKF and IS completely fail.

To quantitatively study the behavior, we numerically compute the weights variance

\[
\text{Var}(Nw(t)) \approx \frac{1}{N} \sum_{n=1}^{N} |Nw^n(t)|^2 - 1
\]

(82)

of all three methods (WEnKI, WEnSRF, and IS). (For IS, the weight at \( t \) is calculated using \( \rho(u,t) \) (defined in (24)).) In Figure 4, we plot evolution of the weight variance with respect to \( t \) in log scale (shifted by 1 for positivity). It shows the variance of weights in IS quickly blows up in time, while the quantity for the other two keep reasonably bounded.

As a demonstration of consistency, we compare moments computed using the four methods, and the reference solution, as shown in Table 1. It is clear that despite EnKI and EnSRF are visually close to the groundtruth solution, the errors in the moments are still rather large. This is expected: the groundtruth solution is
not a Gaussian distribution, but the underlying assumption for EnKI and EnSRF to be valid is the Gaussianity.

As noted before, the weight terms for the two methods are very complicated. The updating formula also requires the computation of the derivatives of $\mathcal{G}$. This will introduce a high cost in practice. We document the cost in Table 2. The weighted version of the algorithms almost double the cost. This is understandable. The number of ODEs that we need to compute is doubled: instead of computing $u_t$ only, we compute both $u_t$ and $w_t$.

5.2. Two dimension example. We also present some 2-D examples. Normal distribution $\mathcal{N}(0, I_2)$ is chosen as the prior distribution.

- Example 4: We consider likelihood function

$$\exp(-\Phi(u; y)) = \frac{1}{4} \sum_{i,j=0}^1 \exp \left( -\frac{(u_1 - a_{i,j})^2 + (u_2 - b_{i,j})^2}{0.2} \right),$$

with

$$a = \begin{bmatrix} 6 & 3 \\ 3 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 3 & 6 \\ 0 & 3 \end{bmatrix}.$$

This design of likelihood function induces two separate centers, as shown in Figure 5. Here, we use $N = 2000$ and choose $\Delta t = 10^{-4}$ for WEnKI and WEnSRF. They capture the motion of the particles accurately. In comparison, IS loses a lot of particles. In this multimodal example, EnKI and EnSRF fail as expected due to the Gaussian assumption. We should emphasize that WEnSRF and WEnKI are not perfect in practice. Indeed, they are nevertheless the corrected version of EnKI.
and EnSRF, the two methods that drove most of the particles to one (the red) block. The weighted correction makes sampling theoretically unbiased, but in the end more particles are left in this particular block.
Table 1. Error of moments estimation in Example 3

|     | WEnKI     | WEnSRF    |
|-----|-----------|-----------|
|     | Moments   | Est. | Re. Error | Est. | Re. Error |
| $E|u^1|$ | 3.84      | 3.82   | 0.0056   | 3.88 | 0.0098    |
| $E|u^2|$ | 14.90     | 14.73  | 0.0114   | 15.19| 0.0192    |
| $E|u^3|$ | 58.22     | 57.19  | 0.0177   | 59.86| 0.0281    |
| $E|u^4|$ | 229.36    | 223.79 | 0.0243   | 237.75| 0.0366   |
| $E|u^5|$ | 911.22    | 882.83 | 0.0312   | 951.95| 0.0447   |

|     | EnKI     | EnSRF    |
|-----|----------|----------|
|     | Moments   | Est. | Re. Error | Est. | Re. Error |
| $E|u^1|$ | 3.84      | 3.69   | 0.0413   | 3.70 | 0.0391    |
| $E|u^2|$ | 14.90     | 13.66  | 0.0833   | 13.73| 0.0785    |
| $E|u^3|$ | 58.22     | 50.90  | 0.1258   | 51.35| 0.1181    |
| $E|u^4|$ | 229.36    | 190.68 | 0.1687   | 193.24| 0.1575   |
| $E|u^5|$ | 911.22    | 718.31 | 0.2117   | 732.17| 0.1965   |

|     | WEnKF | IS    |
|-----|-------|-------|
|     | Moments   | Est. | Re. Error | Est. | Re. Error |
| $E|u^1|$ | 3.84      | 3.40  | 0.1156   | 3.52 | 0.0858    |
| $E|u^2|$ | 14.90     | 11.65 | 0.2181   | 12.37| 0.1699    |
| $E|u^3|$ | 58.22     | 40.22 | 0.3093   | 43.57| 0.2517    |
| $E|u^4|$ | 229.36    | 139.72| 0.3908   | 153.56| 0.3305   |
| $E|u^5|$ | 911.22    | 488.51| 0.4639   | 541.71| 0.4055   |

Table 2. Simulation time in Example 1-3

| Case      | WEnKI | WEnSRF | EnKI | EnSRF |
|-----------|-------|--------|------|-------|
| Example 1 | 0.362s| 0.197s | 0.138s| 0.178s |
| Example 2 | 50.041s| 41.739s| 26.564s| 18.518s |
| Example 3 | 0.198s| 0.115s | 0.120s| 0.072s |

• Example 5: In this case we consider $G(u_1, u_2) = (g_1(u_1, u_2), g_2(u_1, u_2))$ and $y = (0, 0)$, where

$$g_1(u_1, u_2) = (u_1 - 3)^2 + \frac{(u_2 - 3)^2}{2}, \quad g_2(u_1, u_2) = \frac{(u_1 - 3)^2}{2} + (u_2 - 3)^2.$$ 

$N = 1000$ and $\Delta t = 10^{-3}$ for WEnKI and WEnSRF. Results are presented in Figure 6. Due to the form of $G$, the center of the likelihood function is $(2, 2)$ instead of $(0, 0)$ for the prior distribution. Such transition of support is hard for IS to capture. After resampling, only a few samples survive. Visually EnKI and EnSRF still give satisfying results.

To quantitatively understand the performance of the algorithms, we compute the variance of weight (82) and accuracy of moments estimation in this example. The variance of the weight, as a function of time, is plotted in Figure 7 in log scale (shifted by 1 for positivity). As can be seen clearly, the weight of IS blows up quickly while the two newly proposed methods stay reasonable. In Table 3, we tabulate the error of higher moments. Even though EnKI and EnSRF are visually
good methods, in comparison, they do not capture the moments as well as their weighted versions.

6. Conclusion. We conclude the paper with a few remarks of the proposed algorithms.

Since EnKI was proposed in [23], the mystery of if and how it works for the nonlinear case has attracted a lot of attention. The surrounding work, such as the wellposedness of the coupled SDE [6], the wellposedness of the PDE [5], the mean-field limit of SDE to the PDE [12, 13], and the convergence rate [12, 18], and convergence as an optimization method [8, 9], have all been studied in depth, and the use of similar idea leads to development of new algorithms [27, 13]. The investigation into the core sampling problem with nonlinear forward map, however, is thin.

In this paper, by adding the weights to the particles, we are able to correct EnKI (and similarly EnSRF) to ensure the consistency of the algorithm for nonlinear $G$. The derivation, though tedious, is mathematically straightforward. The resulting “weight” factor (40), however, is mathematically messy and physically not intuitive at all. We would like to emphasize that this nonphysical weight term is uniquely determined once the flow is set, namely, if one follows the flow of EnKI,

$$du^n = \text{Cov}_{\text{up}} \Gamma^{-1} \left( y - G(u^n) \right) dt + \text{Cov}_{\text{up}} \Gamma^{-1/2} dW_t,$$

so that in time, the flow provides a linear interpolation between the origin and the target on the logarithmic scale:

$$\rho(u, t) \sim \mu_{\text{prior}} \exp \left\{ -t |y - G(u)|^2 / 2 \right\},$$
then there will be no other ways to define the weight term, and it has to be as tedious and nonphysical as was derived in this paper. This naturally leads to the question if some modifications to the flow can result a physically more meaningful weight function. This, however, is beyond the scope of the current paper.
Table 3. Error of moments estimation in Example 5

|Moments| WEnKI| WEnSRF|
|---|---|---|
|\(\mathbb{E}|u|^{1} = 3.32\)| 3.30| 3.32|
|\(\mathbb{E}|u|^{2} = 11.16\)| 10.99| 11.19|
|\(\mathbb{E}|u|^{3} = 38.05\)| 36.99| 38.12|
|\(\mathbb{E}|u|^{4} = 131.45\)| 125.53| 131.47|
|\(\mathbb{E}|u|^{5} = 460.56\)| 429.99| 459.16|

|Moments| EnKI| EnSRF|
|---|---|---|
|\(\mathbb{E}|u|^{1} = 3.32\)| 2.96| 3.28|
|\(\mathbb{E}|u|^{2} = 11.16\)| 9.07| 11.04|
|\(\mathbb{E}|u|^{3} = 38.05\)| 29.17| 38.25|
|\(\mathbb{E}|u|^{4} = 131.45\)| 100.32| 137.43|
|\(\mathbb{E}|u|^{5} = 460.56\)| 379.73| 516.22|

|Moments| WEnKF| IS|
|---|---|---|
|\(\mathbb{E}|u|^{1} = 3.32\)| 3.40| 3.24|
|\(\mathbb{E}|u|^{2} = 11.16\)| 7.72| 10.50|
|\(\mathbb{E}|u|^{3} = 38.05\)| 21.74| 34.10|
|\(\mathbb{E}|u|^{4} = 131.45\)| 61.62| 110.81|
|\(\mathbb{E}|u|^{5} = 460.56\)| 175.99| 360.27|

Appendix A. Derivatives of \(\rho\). Recall the definition of \(\rho\), we compute its time derivative to have:

\[
\partial_t \rho = -Z'(t) \exp(-t\Phi(u; y))\rho_{\text{prior}}(u) - \Phi(u; y) \frac{Z(t)}{Z(t)} \rho - \Phi(u; y) \rho.
\]

Since \(\Phi(u; y) = \frac{1}{2}|y - G(u)|^2\),

\[
Z'(t) = -\int -\Phi(u; y) \rho(u, t) du = \mathbb{E}_{\rho(t)} \left( \frac{1}{2} |y - G(u)|^2 \right),
\]

we obtain (26).

Compute the \(u\) derivative, we have:

\[
\nabla \rho = \frac{1}{Z(t)} \left( \nabla \exp(-t\Phi(u; y)) \right) \rho_{\text{prior}}(u) + \frac{1}{Z(t)} \nabla \exp(-t\Phi(u; y)) \nabla \rho_{\text{prior}}(u).
\] (83)

Noticing

\[
\nabla \exp(-t\Phi(u; y)) = -t \nabla \Phi(u; y) \exp(-t\Phi(u; y))
\]

and

\[
\nabla \rho_{\text{prior}}(u) = -\Gamma_0^{-1}(u - u_0) \rho_{\text{prior}}(u),
\]

we plug them in (83) to obtain (27) for \(\nabla \rho = \mathcal{V} \rho\).

To compute the hessian, we repeat the process above for

\[
\mathcal{H}_u \rho = \nabla \mathcal{V} u + \mathcal{V} (\nabla \rho)^\top.
\]
While $\nabla (\nabla \rho)^\top$ contributes $\nabla \nabla^\top \rho$, the derivative of $\nabla$ become $-t (\nabla G)^\top \Gamma^{-1} \nabla G + t \nabla W$, leading to (28) in the end.

Appendix B. Proof of proposition 1. In this appendix, we derive the explicit bound for $R_t$ in the following lemma, and show Proof of Proposition 1. It plays the crucial role in Theorem 4.6.

Lemma B.1. Under Assumption 4.2, for all $0 < t < 1$, $R_1$, $R_2$ and $R_3$ defined in (40) satisfy:

- For $R_1$
  \[ |R_1(u, t)| \leq C_2(\text{Var}^{\rho(t)}(u))^2 + C_1 \text{Var}^{\rho(t)}(u). \]  
  (84)

- For $R_2$
  \[ 2|R_2(u^n_t, t)| \leq \| \Gamma^{-1} \|_2 (A^2 |u_A^* - \overline{\rho}(t)|^2 + |r|^2 + M^2). \]  
  (85)

- For $R_3$
  \[ |R_3(u, t)| \leq C_3(\text{Var}^{\rho(t)}(u))^2. \]  
  (86)

- More carefully:
  \[ |R_2(u^n_t, t)| \leq C_4 \left| \overline{\rho}(t) - u_A^* \right| \left[ \left| \overline{\rho}(t) - \text{Cov}^{\rho(t)}_{u,u}(\text{Cov}_A(t))^{-1} u_A \right| + \| \text{Cov}^{\rho(t)}_{u,u} \|_2 A_1 \right] \]
  \[ + C_4 \| \text{Cov}^{\rho(t)}_{m,u} \|_2 (|u_A| + A_1) \]
  \[ + C_4 \left| \overline{\rho}(t) - u_A^* \right| \left| \left| I - (\text{Cov}_A)^{-1} \text{Cov}^{\rho(t)}_{u,u} \right|_2 + \| \text{Cov}^{\rho(t)}_{m,u} \|_2 \right| u^n_t |. \]  
  (87)

In the equation $\text{Var}^{\rho(t)}(u) = \text{Tr} \left( \text{Cov}^{\rho(t)}_{u,u} \right)$ and all constants $C_1, \ldots, C_4$ are constants depending on $\Lambda$, $\| \Gamma^{-1} \|_2$, $\| \Gamma_0^{-1} \|_2$, $|r|$, and $M$.

Proof. This comes from direct calculation. Firstly to show (84), we plug (48) into (40):\[
R_1(u, t) = \frac{1}{2} \text{Tr} \left\{ \text{Cov}^{\rho(t)}_{u,u} \Gamma^{-1} A - 2 A^\top \Gamma^{-1} A \text{Cov}^{\rho(t)}_{u,u} \right\} \]
\[ + \frac{1}{2} \text{Tr} \left\{ \text{Cov}^{\rho(t)}_{u,u} A^\top \Gamma^{-1} A \text{Cov}^{\rho(t)}_{u,u} \left( t (\nabla G(u))^\top \Gamma^{-1} \nabla G(u) + \Gamma_0^{-1} \right) \right\} \]
\[ + \frac{1}{2} \text{Tr} \left\{ \text{Cov}^{\rho(t)}_{u,m} \Gamma^{-1} - 2 (\nabla m)^\top \Gamma^{-1} \text{Cov}^{\rho(t)}_{m,u} \right\} \]
\[ + \frac{1}{2} \text{Tr} \left\{ \text{Cov}^{\rho(t)}_{m,u} \Gamma^{-1} \text{Cov}^{\rho(t)}_{m,u} \left( t (\nabla G(u))^\top \Gamma^{-1} \nabla G(u) + \Gamma_0^{-1} \right) \right\}, \]
where we use (49) and the first term is less than 0. Notice
\[
\text{Var}^{\rho(t)}(m(u)) = \text{Tr} \left( \text{Cov}^{\rho(t)}_{m,m} \right) = E^{\rho(t)} |m(u) - m|^2 \]
\[ \leq E^{\rho(t)} |m(u) - m|^2 \leq \Lambda^2 E^{\rho(t)} |u - \overline{\mu}|^2 = \Lambda^2 \text{Var}^{\rho(t)}(u), \]
we have the following five inequalities:

\[
\begin{align*}
\text{Tr} \left\{ \text{Cov}^{\rho(t)}_{uu} A^\top \Gamma^{-1} A \right\} & \leq \Lambda^2 \|\Gamma^{-1}\|_2 \text{Var}^{\rho(t)}(u), \\
\text{Tr} \left\{ \text{Cov}^{\rho(t)}_{mm} \Gamma^{-1} \right\} & \leq \|\Gamma^{-1}\|_2 \text{Var}^{\rho(t)}(m(u)) \leq \Lambda^2 \|\Gamma^{-1}\|_2 \text{Var}^{\rho(t)}(u), \\
\left| \text{Tr} \left\{ \Gamma^{-1} \text{Cov}^{\rho(t)}_{mu} \right\} \right| & = \mathbb{E} (m(u) - \bar{m})^\top \Gamma^{-1} (u - \bar{u}) \leq \|\Gamma^{-1}\|_2 \mathbb{E} \|m(u) - \bar{m}\|_2 \|u - \bar{u}\|_2 \\
& \leq \|\Gamma^{-1}\|_2 (\text{Var}^{\rho(t)}(m(u)))^{1/2} (\text{Var}^{\rho(t)}(u))^{1/2} \\
& \leq \Lambda \|\Gamma^{-1}\|_2 \text{Var}^{\rho(t)}(u),
\end{align*}
\]

and furthermore

\[
\begin{align*}
\text{Tr} \left\{ \text{Cov}^{\rho(t)}_{uu} A^\top \Gamma^{-1} A \text{Cov}^{\rho(t)}_{uu} \right\} & \leq \text{Tr} \left\{ \text{Cov}^{\rho(t)}_{uu} A^\top \Gamma^{-1} A \right\} \|\text{Cov}^{\rho(t)}_{uu}\|_2 \\
& \leq \Lambda^2 \|\Gamma^{-1}\|_2 (\text{Var}^{\rho(t)}(u))^2,
\end{align*}
\]

and

\[
\begin{align*}
\text{Tr} \left\{ \text{Cov}^{\rho(t)}_{um} \Gamma^{-1} \text{Cov}^{\rho(t)}_{mu} \right\} & \leq \|\Gamma^{-1}\|_2 \|\text{Cov}^{\rho(t)}_{um}\|_2^2 \\
& \leq \|\Gamma^{-1}\|_2 (\text{Var}^{\rho(t)}(m(u))) (\text{Var}^{\rho(t)}(u)) \\
& \leq \Lambda^2 \|\Gamma^{-1}\|_2 (\text{Var}^{\rho(t)}(u))^2,
\end{align*}
\]

we have

\[
|\mathcal{R}_1(u, t)| \leq \frac{3}{2} \Lambda^2 \|\Gamma^{-1}\|_2 \text{Var}^{\rho(t)}(u) + [t \Lambda^2 \|\Gamma^{-1}\|_2 + \|\Gamma_0^{-1}\|_2] \Lambda^2 \|\Gamma^{-1}\|_2 (\text{Var}^{\rho(t)}(u))^2.
\]

Let 

\[
C_1 = \frac{3}{2} \Lambda^2 \|\Gamma^{-1}\|_2, \quad C_2 = [t \Lambda^2 \|\Gamma^{-1}\|_2 + \|\Gamma_0^{-1}\|_2] \Lambda^2 \|\Gamma^{-1}\|_2,
\]

we obtain (84). To bound $\mathcal{R}_3$, we first notice

\[
\mathcal{W}(u) = \left[ (\partial_1 \nabla m(u))^\top \Gamma^{-1} (r - \bar{m}(u)), \ldots, (\partial_L \nabla m(u))^\top \Gamma^{-1} (r - \bar{m}(u)) \right].
\]

by plugging in (48),(51),(52). Then we have

\[
|\mathcal{R}_3(u, t)| \leq \frac{t \text{Tr} \left\{ \text{Cov}^{\rho(t)}_{uu} A^\top \Gamma^{-1} A \text{Cov}^{\rho(t)}_{uu} + \text{Cov}^{\rho(t)}_{um} \Gamma^{-1} \text{Cov}^{\rho(t)}_{mu} \right\} \|\mathcal{W}(u)\|_2}{2} \\
\leq C'(\Lambda, \|\Gamma^{-1}\|_2)(1 + |r|) \Lambda^2 \|\Gamma^{-1}\|_2 (\text{Var}^{\rho(t)}(u))^2,
\]

where we use (89),(90) and

\[
\|\mathcal{W}(u)\|_2 \leq \|\mathcal{W}(u)\|_F \leq \frac{L}{2} \left( \|\mathcal{H}([m_i^2])\|_2 + \|\Gamma^{-1}\|_2 \|\nabla m\|_2^2 \right) \\
+ |r| \max_{1 \leq i \leq L} \{ \|\partial_i \nabla m\|_2 \} \\
\leq C'(\Lambda, \|\Gamma^{-1}\|_2)(1 + |r|).
\]

We obtain (86) by defining $C_3 = C'(\Lambda, \|\Gamma^{-1}\|_2)(1 + |r|) \Lambda^2 \|\Gamma^{-1}\|_2$.

To show (85), we simply plug in the weak nonlinearity assumption for:

\[
2 |\mathcal{R}_2(u^*, t)| \leq \left( y - \overline{y}^{\rho(t)} \right)^\top \Gamma^{-1} \left( y - \overline{y}^{\rho(t)} \right) = \left| A \left( u^*_A - \bar{y}^{\rho(t)} \right) \right|_\Gamma^2 + |r - \bar{y}^{\rho(t)}|_\Gamma^2 \\
\leq \|\Gamma^{-1}\|_2 (\Lambda^2 |u_A^* - \bar{y}^{\rho(t)}|^2 + |r|^2 + M^2).
\]
Finally,
\[
R_2(u^n_t, t) = \frac{1}{2} \left| A(\pi^{\rho(t)} - u^*_A) \right|_\Gamma^2 - \frac{1}{2} \left| A(u^*_A - u^n_t) - ACov^{\rho(t)}_{u, u}\mathcal{V}(u^n_t, t) \right|_\Gamma^2
\]
\[
+ \frac{1}{2} \left| r - m(u^n_t) \right|_\Gamma^2 - \frac{1}{2} \left| r - m(u^n_t) - Cov^{\rho(t)}_{m, u}\mathcal{V}(u^n_t, t) \right|_\Gamma^2.
\]  
(91)

According to the definition of \( \mathcal{V} \),
\[
\mathcal{V}(u^n_t, t) = tA^\top \Gamma^{-1}(u^*_A - u^n_t) - \Gamma_0^{-1}(u^n_t - u_0) + t(\nabla m)^\top \Gamma^{-1}(r - m(u^n_t))
\]
\[
= (\text{Cov}_A(t))^{-1}(u^*_A - u^n_t) + t(\nabla m)^\top \Gamma^{-1}(r - m(u^n_t)),
\]  
(92)

and thus
\[
\frac{1}{2} \left| A(u^*_A - u^n_t) - ACov^{\rho(t)}_{u, u}\mathcal{V}(u^n_t, t) \right|_\Gamma^2
\]
\[
= \frac{1}{2} \left| A \left[ (I - Cov^{\rho(t)}_{u, u}(\text{Cov}_A(t))^{-1})u^n_t + Cov^{\rho(t)}_{u, u}(\text{Cov}_A(t))^{-1}u_A(t) - u^*_A + \mathcal{M}(u) \right] \right|_\Gamma^2,
\]
where
\[
\mathcal{M}(u) = tCov^{\rho(t)}_{u, u}(\nabla m)^\top \Gamma^{-1}(r - m(u)).
\]

This means the first two terms in (91) are controlled by:
\[
\frac{1}{2} \left| A(\pi^{\rho(t)} - u^*_A) \right|_\Gamma^2 - \frac{1}{2} \left| A(u^*_A - u^n_t) - ACov^{\rho(t)}_{u, u}\mathcal{V}(u^n_t, t) \right|_\Gamma^2
\]
\[
\leq \frac{1}{2} \left| A(\pi^{\rho(t)} - u^*_A) \right|_\Gamma^2 - \frac{1}{2} \left| A \left[ (I - Cov^{\rho(t)}_{u, u}(\text{Cov}_A(t))^{-1})u^n_t + Cov^{\rho(t)}_{u, u}(\text{Cov}_A(t))^{-1}u_A(t) - u^*_A + \mathcal{M}(u) \right] \right|_\Gamma^2
\]
\[
\leq C' \left| A(\pi^{\rho(t)} - u^*_A) \right|_\Gamma 
\cdot \left| A \left[ (I - Cov^{\rho(t)}_{u, u}(\text{Cov}_A(t))^{-1})u^n_t + Cov^{\rho(t)}_{u, u}(\text{Cov}_A(t))^{-1}u_A(t) - u^*_A + \mathcal{M}(u) \right] \right|_\Gamma
\]
\[
\leq C' \left| \pi^{\rho(t)} - u^*_A \right|_\Gamma \left[ \left| \pi^{\rho(t)} - Cov^{\rho(t)}_{u, u}(\text{Cov}_A(t))^{-1}u_A(t) \right| + \left| Cov^{\rho(t)}_{u, u}(\text{Cov}_A(t))^{-1}u_A(t) \right|_2 \right]
\]
\[
+ C' \left| \pi^{\rho(t)} - u^*_A \right|_\Gamma \left| I - (\text{Cov}_A(t))^{-1}Cov^{\rho(t)}_{u, u}(\text{Cov}_A(t))^{-1}u_A(t) \right|_2 \left| u^n_t \right|_2,
\]  
(93)

where \( C' \) is a constant depending on \( \Lambda, \left\| \Gamma^{-1} \right\|_2, \left\| \Gamma_0^{-1} \right\|_2, \left\| r \right\| \) and \( M \). Furthermore, the latter two terms in (91) are bounded by:
\[
\frac{1}{2} \left| r - m(u^n_t) \right|_\Gamma^2 - \frac{1}{2} \left| r - m(u^n_t) - Cov^{\rho(t)}_{m, u}\mathcal{V}(u^n_t, t) \right|_\Gamma^2
\]
\[
\leq \left( \Gamma^{-1}(r - m(u^n_t)), Cov^{\rho(t)}_{m, u}(\text{Cov}_A(t))^{-1}(u_A - u^n_t) \right)
\]
\[
+ \left( \Gamma^{-1}(r - m(u^n_t)), tCov^{\rho(t)}_{m, u}(\nabla m)^\top \Gamma^{-1}(r - m(u^n_t)) \right)
\]
\[
\leq C'' \left| Cov^{\rho(t)}_{m, u}(\text{Cov}_A(t))^{-1}(u_A - u^n_t) \right|_2
\]
\[
+ C'' \left| Cov^{\rho(t)}_{m, u}(\text{Cov}_A(t))^{-1}(u_A - u^n_t) + A_1 \right| + C'' \left| Cov^{\rho(t)}_{m, u}(\text{Cov}_A(t))^{-1}(u_A - u^n_t) \right|_2
\]
\[
\leq C'' \left| Cov^{\rho(t)}_{m, u}(\text{Cov}_A(t))^{-1}(u_A - u^n_t) \right|_2,
\]  
(94)

where \( C'' \) is a constant depending on \( \Lambda, \left\| \Gamma^{-1} \right\|_2, \left\| \Gamma_0^{-1} \right\|_2, \left\| r \right\| \) and \( M \).

These together give the upper bound for \( R_2 \). Call the constant \( C_4 \), we finish the proof.

Now we are ready to prove Proposition 1.
We first estimate \( \frac{d}{dt} \mathbb{E}[u_t^n](\omega_t^n)^2 \). Using (41) and Itô’s formula, we obtain
\[
\frac{d}{dt} \mathbb{E}[u_t^n](\omega_t^n)^2 = 2 \mathbb{E}[\omega_t^n]^2 \langle du_t^n, u_t^n \rangle + \mathbb{E}[u_t^n]^2 \omega_t^n \, du_t^n,
\]
and thus
\[
\frac{d}{dt} \mathbb{E}[u_t^n](\omega_t^n)^2 = 2 \mathbb{E} \left( \omega_t^n \right)^2 \left( \text{Cov}_{u_t}^{(t)} \Gamma^{-1}(y - G(u_t^n)) , u_t^n \right) + 2 \mathbb{E} \left( \omega_t^n \right)^2 \text{Tr} \left( \text{Cov}_{u_t}^{(t)} \Gamma^{-1} \text{Cov}_{u_t}^{(t)} \right) + 2 \mathbb{E} \left( u_t^n \right)^2 \omega_t^n \mathbb{E} \left( \omega_t^n \right)^2.
\]

We now bound the two terms respectively. For the first:

\[
\mathbb{E} \left( \omega_t^n \right)^2 \left( \left\langle \text{Cov}_{u_t}^{(t)} \Gamma^{-1}(y - G(u_t^n)) , u_t^n \right\rangle + \frac{1}{2} \text{Tr} \left( \text{Cov}_{u_t}^{(t)} \Gamma^{-1} \text{Cov}_{u_t}^{(t)} \right) \right) \leq \Lambda^2 \|\Gamma^{-1}\|_2 \text{Var}^{(t)}(u) \mathbb{E} \left( [u_t^n]^2 + |r| |u_t^n| + M |u_t^n|^2 \right) \omega_t^n \mathbb{E} \left( \omega_t^n \right)^2 \]

where we use (48) in the first equality and that
\[
\left\langle \text{Cov}_{u_t}^{(t)} A^\Gamma^{-1} A (u_t^n - u_t^n), u_t^n \right\rangle \leq \|\text{Cov}_{u_t}^{(t)} A^\Gamma^{-1} A\|_2 \mathbb{E} \left( [u_t^n]^2 + |u_t^n|^2 \right) \leq \Lambda^2 \|\Gamma^{-1}\|_2 \text{Var}^{(t)}(u) \mathbb{E} \left( [u_t^n]^2 + |u_t^n|^2 \right)
\]

and
\[
\langle \text{Cov}_{u_t}^{(t)}(r - m(u)), u_t^n \rangle \leq \mathbb{E} \left| m(u) - m \right| \|\Gamma^{-1}\|_2 |u_t^n| \mathbb{E} \left( |r| |u_t^n| + M |u_t^n|^2 \right) \leq \Lambda^2 \|\Gamma^{-1}\|_2 \text{Var}^{(t)}(u) \mathbb{E} \left( |r| |u_t^n| + M |u_t^n|^2 \right),
\]

where (89) and (90) are applied.

For second term in (95), we simply apply the inequalities (84)-(85). These together provide the estimate of (95) as

\[
\frac{d}{dt} \mathbb{E}[u_t^n](\omega_t^n)^2 \leq \mathcal{C} \left[ \text{Var}^{(t)}(u) \mathbb{E} \left( [u_t^n]^2 \omega_t^n \right)^2 \right] + \mathcal{C} \left[ u_t^n - \bar{m}^{(t)} \right]^2 + 1 \mathbb{E} \left( [u_t^n]^2 \omega_t^n \right)^2 \right]
\]

where \( \mathcal{C} \) is a constant depends on \( \Lambda, \|\Gamma^{-1}\|_2, \|\Gamma_0^{-1}\|_2, |r| \) and \( M \).

Next we estimate \( \mathbb{E}[(u_t^n)^2] \). Note that, according to (41), one has

\[
\frac{1}{2} \frac{d}{dt} \mathbb{E}[u_t^n]^2 \leq (\|\mathcal{R}_1\|_\infty + |\mathcal{R}_2(u_t^n, t)| + \|\mathcal{R}_3\|_\infty) \mathbb{E}[u_t^n]^2.
\]
To control these terms we apply (84), (86) and (87), which leads to:

\[
\frac{d}{dt} \mathbb{E}(\omega_t^n)^2 \\
\leq C \left[\Var^{(t)}(u) + \Var^{(t)}(u)\right] \mathbb{E}(\omega_t^n)^2 \\
+ C \left[|\pi_t^{(t)} - u_A| \left[|\pi_t^{(t)} - \text{Cov}_{u,u}^{(t)}(\text{Cov}_{A}(t))^{-1}u_A| + \|\text{Cov}_{u,u}^{(t)}\|_2 A_1\right] \mathbb{E}(\omega_t^n)^2 \\
+ C \|\text{Cov}_{A}^{(t)}\|_2 (|u_A| + A_1) \mathbb{E}(\omega_t^n)^2 \\
+ C \left[|\pi_t^{(t)} - u_A| \left[|I - (\text{Cov}_{A})^{-1}\text{Cov}_{u,u}^{(t)}|\right]_2 + \|\text{Cov}_{u,u}^{(t)}\|_2 \mathbb{E} \left[|u_t^n| (\omega_t^n)^2\right] \right] \\
\leq C \left[\Var^{(t)}(u) + \Var^{(t)}(u)\right] \mathbb{E}(\omega_t^n)^2 \\
+ C \left[|\pi_t^{(t)} - u_A| \left[|\pi_t^{(t)} - \text{Cov}_{u,u}^{(t)}(\text{Cov}_{A}(t))^{-1}u_A| + \|\text{Cov}_{u,u}^{(t)}\|_2 A_1\right] \mathbb{E}(\omega_t^n)^2 \\
+ C \|\text{Cov}_{A}^{(t)}\|_2 (|u_A| + A_1) \mathbb{E}(\omega_t^n)^2 \\
+ C \left[|\pi_t^{(t)} - u_A| \left[|I - (\text{Cov}_{A})^{-1}\text{Cov}_{u,u}^{(t)}|\right]_2 + \|\text{Cov}_{u,u}^{(t)}\|_2 \mathbb{E} \left[|u_t^n| (\omega_t^n)^2\right] \right] / 2, \\
\right.
\]

(99)

where \( C \) is a constant depends on \( \Lambda, \|\Gamma^{-1}\|_2, \|\Gamma_0^{-1}\|_2, |r| \) and \( M \).

Combine (97) and (99), we have

\[
\frac{d}{dt} \left( \frac{\mathbb{E}|u_t^n|^2(\omega_t^n)^2}{\mathbb{E}(\omega_t^n)^2} \right) \leq W(t) \frac{\mathbb{E}|u_t^n|^2(\omega_t^n)^2}{\mathbb{E}(\omega_t^n)^2},
\]

Multiply \( N^2 \) on both sides of this inequality and notice \( \mathbb{E}(N\omega_t^n) = 1 \), we have, for \( 0 \leq t \leq 1 \)

\[
\frac{d}{dt} \left( \frac{\mathbb{E}|u_t^n|^2(N\omega_t^n)^2}{\mathbb{E}(N\omega_t^n)^2 - (\mathbb{E}N\omega_t^n)^2 + 1} \right) \leq W(t) \frac{\mathbb{E}|u_t^n|^2(N\omega_t^n)^2}{\mathbb{E}(N\omega_t^n)^2 - (\mathbb{E}N\omega_t^n)^2 + 1},
\]

which concludes (68), hence the proposition.

\[\square\]

REFERENCES

[1] C. Andrieu, N. Freitas, A. Doucet and M. Jordan, An introduction to MCMC for machine learning, Machine Learning, 50 (2003), 5–43.
[2] A. Bain and D. Crisan, Fundamentals of Stochastic Filtering, Stochastic Modelling and Applied Probability, Springer New York, 2008.
[3] K. Bergemann and S. Reich, A localization technique for ensemble Kalman filters, Quarterly Journal of the Royal Meteorological Society, 136 (2010), 701–707.
[4] K. Bergemann and S. Reich, A mollified ensemble Kalman filter, Quarterly Journal of the Royal Meteorological Society, 136 (2010), 1636–1643.
[5] D. Blömker, C. Schillings, P. Wacker and S. Weissmann, Well posedness and convergence analysis of the ensemble Kalman inversion, Inverse Problems, 35 (2019), 085007.
[6] D. Blömker, C. Schillings and P. Wacker, A strongly convergent numerical scheme from ensemble Kalman inversion, SIAM Journal on Numerical Analysis, 56 (2018), 2537–2562.
[7] J. A. Cañizo, J. A. Carrillo and J. Rosado, A well-posedness theory in measures for some kinetic models of collective motion, Mathematical Models and Methods in Applied Sciences, 21 (2011), 515–539.
[8] N. Chada, A. Stuart and X. Tong, Tikhonov regularization within ensemble Kalman inversion, SIAM Journal on Numerical Analysis, 58 (2019), 1263–1294.
[9] N. Chada and X. T. Tong, Convergence acceleration of ensemble Kalman inversion in nonlinear settings, preprint, arXiv:1911.02424.
[10] N. K. Chada, M. A. Iglesias, L. Roininen and A. M. Stuart, Parameterizations for ensemble Kalman inversion, *Inverse Problems*, 34 (2018), 055009.

[11] J. de Wiljes, S. Reich and W. Stannat, Long-time stability and accuracy of the ensemble Kalman–Bucy filter for fully observed processes and small measurement noise, *SIAM Journal on Applied Dynamical Systems*, 17 (2018), 1152–1181.

[12] Z. Ding and Q. Li, Ensemble Kalman inversion: Mean-field limit and convergence analysis, arXiv:1908.05575.

[13] Z. Ding and Q. Li, Ensemble Kalman sampling: Mean-field limit and convergence analysis, preprint, arXiv:1910.12923.

[14] A. Doucet, N. De Freitas and N. Gordon, *An Introduction to Sequential Monte Carlo Methods*, Springer New York, New York, NY, 2001.

[15] A. Doucet, N. De Freitas and N. Gordon, *Sequential Monte Carlo Methods in Practice*, Springer New York, London, 2001.

[16] O. G. Ernst, B. Sprungk and H.-J. Starkloff, Analysis of the ensemble and polynomial chaos Kalman filters in bayesian inverse problems, *SIAM/ASA Journal on Uncertainty Quantification*, 3 (2015), 823–851.

[17] G. Evensen, *Data Assimilation: The Ensemble Kalman Filter*, Springer-Verlag, Berlin, Heidelberg, 2006.

[18] N. Fournier and A. Guillin, On the rate of convergence in Wasserstein distance of the empirical measure, *Probability Theory and Related Fields*, 162 (2015), 707–738.

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

[20] A. Garbuno-Inigo, N. Nüsken and S. Reich, Affine invariant interacting Langevin dynamics for Bayesian inference, CoRR, abs/1912.02859.

[21] J. Geweke, Bayesian inference in econometric models using Monte Carlo integration, *Econometrica*, 57 (1989), 1317–1339.

[22] M. Herty and G. Visconti, Continuous limits for constrained ensemble Kalman filter, *Inverse Problems*, 36 (2020), 075006.

[23] M. A. Iglesias, K. J. H. Law and A. M. Stuart, Ensemble Kalman methods for inverse problems, *Inverse Problems*, 29 (2013), 045001.

[24] T. Lange and W. Stannat, On the continuous time limit of the ensemble Kalman filter, preprint, arXiv:1901.05204.

[25] K. J. H. Law, H. Tembine and R. Tempone, Deterministic mean-field ensemble Kalman filtering, *SIAM Journal on Scientific Computing*, 38 (2016), A1251–A1279.

[26] D. M. Livings, S. L. Dance and N. K. Nichols, Unbiased ensemble square root filters, *Physica D: Nonlinear Phenomena*, 237 (2008), 1021–1028.

[27] Y. Lu, J. Lu and J. Nolen, Accelerating Langevin sampling with birth-death, preprint, arXiv:1905.09863.

[28] J. Martin, L. Wilcox, C. Burstedde and O. Ghattas, A stochastic newton MCMC method for large-scale statistical inverse problems with application to seismic inversion, *SIAM Journal on Scientific Computing*, 34 (2012), A1460–A1487.

[29] Y. M. Marzouk, H. N. Najm and L. A. Rahn, Stochastic spectral methods for efficient Bayesian solution of inverse problems, *Journal of Computational Physics*, 224 (2007), 560–586.

[30] A. Muntean, J. Rademacher and A. Zagaris, *Macroscopic and Large Scale Phenomena: Coarse Graining, Mean Field Limits and Ergodicity*, LAMM, 3, Springer, Cham, 2016.

[31] N. Papadakis, E. Mém-in, A. Cuzol and N. Gengembre, Data assimilation with the weighted ensemble Kalman filter, *Tellus A*, 62 (2010), 673–697.

[32] S. Reich, A dynamical systems framework for intermittent data assimilation, *BIT Numerical Mathematics*, 51 (2011), 235–249.

[33] S. Reich and C. Cotter, *Probabilistic Forecasting and Bayesian Data Assimilation*, Cambridge University Press, 2015.

[34] S. Reich and S. Weissmann, Fokker-planck particle systems for bayesian inference: Computational approaches, preprint, arXiv:1911.10832.

[35] C. Schillings and A. M. Stuart, Analysis of the ensemble Kalman filter for inverse problems, *SIAM Journal on Numerical Analysis*, 55 (2017), 1264–1290.

[36] M. Tippett, J. Anderson, C. Bishop, T. Hamill and J. Whitaker, Ensemble square root filters, *Monthly Weather Review*, 131 (2003), 1485–1490.
Received May 2020; revised July 2020.

E-mail address: zding49@math.wisc.edu
E-mail address: qinli@math.wisc.edu
E-mail address: jianfeng@math.duke.edu