IOP Publishing Inverse Problems
Inverse Problems **38** (2022) 125006 (40pp) https://doi.org/10.1088/1361-6420/ac99fa

Efficient derivative-free Bayesian inference for large-scale inverse problems

Daniel Zhengyu Huang[1,*], **Jiaoyang Huang**[2], **Sebastian Reich**[3] **and Andrew M Stuart**[1]

[1] California Institute of Technology, Pasadena, CA, United States of America
[2] New York University, New York, NY, United States of America
[3] Universität Potsdam, Potsdam, Germany

E-mail: dzhuang@caltech.edu, jh4427@nyu.edu, sebastian.reich@uni-potsdam.de and astuart@caltech.edu

Received 11 April 2022, revised 21 September 2022
Accepted for publication 13 October 2022
Published 31 October 2022

Abstract

We consider Bayesian inference for large-scale inverse problems, where computational challenges arise from the need for repeated evaluations of an expensive forward model. This renders most Markov chain Monte Carlo approaches infeasible, since they typically require $\mathcal{O}(10^4)$ model runs, or more. Moreover, the forward model is often given as a black box or is impractical to differentiate. Therefore derivative-free algorithms are highly desirable. We propose a framework, which is built on Kalman methodology, to efficiently perform Bayesian inference in such inverse problems. The basic method is based on an approximation of the filtering distribution of a novel mean-field dynamical system, into which the inverse problem is embedded as an observation operator. Theoretical properties are established for linear inverse problems, demonstrating that the desired Bayesian posterior is given by the steady state of the law of the filtering distribution of the mean-field dynamical system, and proving exponential convergence to it. This suggests that, for nonlinear problems which are close to Gaussian, sequentially computing this law provides the basis for efficient iterative methods to approximate the Bayesian posterior. Ensemble methods are applied to obtain interacting particle system approximations of the filtering distribution of the mean-field model; and practical strategies to further reduce the computational and memory cost of the methodology are presented, including low-rank approximation and a bi-fidelity approach. The effectiveness of the framework is demonstrated in several numerical experiments, including proof-of-concept linear/nonlinear examples and two large-scale applications: learning of permeability parameters in subsurface flow; and learning subgrid-scale parameters in a global climate model. Moreover, the stochastic

*Author to whom any correspondence should be addressed.

ensemble Kalman filter and various ensemble square-root Kalman filters are all employed and are compared numerically. The results demonstrate that the proposed method, based on exponential convergence to the filtering distribution of a mean-field dynamical system, is competitive with pre-existing Kalman-based methods for inverse problems.

Keywords: inverse problem, uncertainty quantification, Bayesian inference, derivative-free optimization, mean-field dynamical system, interacting particle system, ensemble Kalman filter

(Some figures may appear in colour only in the online journal)

1. Introduction

1.1. Orientation

The focus of this work is on efficient derivative-free Bayesian inference approaches for large scale inverse problems, in which the goal is to estimate probability densities for uncertain parameters, given noisy observations derived from the output of a model that depends on the parameters. Such approaches are highly desirable for numerous models arising in science and engineering applications, often defined through partial differential equations. These include, to name a few, global climate model calibration [1, 2], material constitutive relation calibration [3–5], seismic inversion in geophysics [6–10], and biomechanics inverse problems [11, 12]. Such problems may feature multiple scales, may include chaotic dynamics, or may involve turbulent phenomena; as a result the forward models are typically very expensive to evaluate. Moreover, the forward solvers are often given as a black box (e.g., off-the-shelf solvers [13] or multiphysics systems requiring coupling of different solvers [14, 15]), and may not be differentiable due to the numerical methods used (e.g., embedded boundary method [16, 17] and adaptive mesh refinement [18, 19]) or because of the inherently discontinuous physics (e.g. in fracture [20] or cloud modeling [21, 22]).

Traditional methods for derivative-free Bayesian inference to estimate the posterior distribution include specific instances of the Markov chain Monte Carlo methodology [23–26] (MCMC), such as random walk Metropolis or the preconditioned Crank–Nicolson (pCN) algorithm [26], and sequential Monte Carlo methods [27, 28] (SMC), which are in any case often interwoven with MCMC. These methods typically require $\mathcal{O}(10^4)$ iterations, or more, to reach statistical convergence for the complex forward models which motivate our work. Given that each forward run can be expensive, conducting $\mathcal{O}(10^4)$ runs is often computationally unfeasible. We present an approach based on the Kalman filter methodology, which aims to estimate the first two moments of the posterior distribution. We demonstrate that, in numerical tests across a range of examples, the proposed methodologies converge within $\mathcal{O}(10)$ iterations, using $\mathcal{O}(10)$ embarrassingly-parallel model evaluations per step, resulting in orders of magnitude reduction in cost over derivative-free MCMC and SMC methods. We also demonstrate favorable performance in comparison with existing Kalman-based Bayesian inversion techniques.

In subsection 1.2, we outline the Bayesian approach to inverse problems, describing various approaches to sampling, formulated as dynamical systems on probability measures, and introducing our novel mean field approach. In subsection 1.3, we discuss pathwise stochastic dynamical systems which realize such dynamics at the level of measures, and discuss filtering

algorithms which may be applied to them for the purposes of approximate inversion. Subsection 1.4 highlights the novel contributions in this paper, building on the context established in the two preceding subsections. Subsection 1.5 summarizes notational conventions that we adopt throughout.

1.2. Bayesian formulation of the inverse problem

Inverse problems can be formulated as recovering unknown parameters $\theta \in \mathbb{R}^{N_\theta}$ from noisy observation $y \in \mathbb{R}^{N_y}$ related through

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

Here $\mathcal{G}$ denotes a forward model mapping parameters to output observables, and η denotes observational noise; for simplicity we will assume known Gaussian statistics: $\eta \sim \mathcal{N}(0, \Sigma_\eta)$. In the Bayesian perspective, θ and y are treated as random variables. Given the prior $\rho_{\text{prior}}(\theta)$ on θ, the inverse problem can be formulated as finding the posterior $\rho_{\text{post}}(\theta)$ on θ given y [29–31]:

$$\rho_{\text{post}}(\theta) = \frac{1}{Z(y)} \mathrm{e}^{-\Phi(\theta,y)} \rho_{\text{prior}}(\theta), \quad \Phi(\theta, y) = \frac{1}{2} \|\Sigma_\eta^{-\frac{1}{2}}(y - \mathcal{G}(\theta))\|^2 \tag{2}$$

and $Z(y)$ is the normalization constant

$$Z(y) = \int \mathrm{e}^{-\Phi(\theta,y)} \rho_{\text{prior}}(\theta) \mathrm{d}\theta. \tag{3}$$

We focus on the case, where the prior ρ_{prior} is (or is approximated as) Gaussian with mean and covariance r_0 and Σ_0, respectively. Then the posterior $\rho_{\text{post}}(\theta)$ can be written as

$$\rho_{\text{post}}(\theta) = \frac{1}{Z(y)} \mathrm{e}^{-\Phi_R(\theta,y)}, \quad \Phi_R(\theta, y) = \Phi(\theta, y) + \frac{1}{2} \|\Sigma_0^{-\frac{1}{2}}(\theta - r_0)\|^2. \tag{4}$$

1.2.1. Computational approaches. Bayesian inference requires approximation of, or samples from, the posterior distribution given by equation (2). There are three major avenues to approximate the posterior distribution:

- Those based on variational inference [32, 33], where a parameterized approximate density is constructed and optimized to minimize the distance to the posterior density. They include Gaussian variational inference [34–36] and normalizing flows [37].
- Those based on sampling and more importantly the invariance of measures and ergodicity. They include MCMC [23, 24], Langevin dynamics [38, 39], and more recently interacting particle approaches [25, 40–42].

  At an abstract mathematical level, invariance and ergodicity-based approaches to sampling from the posterior ρ_{post} rely on the transition kernel $\psi_{\text{I}}(\theta', \theta)$ such that

  $$\rho_{\text{post}}(\theta) = \int \psi_{\text{I}}(\theta', \theta) \rho_{\text{post}}(\theta') \mathrm{d}\theta', \tag{5}$$

  that is, the posterior distribution $\rho_{\text{post}}(\theta)$ is invariant with respect to the transition kernel $\psi_{\text{I}}(\theta', \theta)$. Furthermore, starting from any initial distribution the associated Markov chain should approach the invariant measure $\rho_{\text{post}}(\theta)$ asymptotically.
- Those based on coupling ideas (mostly in the form of coupling the prior with the posterior). While several sequential data assimilation methods, such as importance sampling-resampling in SMC [43] and the ensemble Kalman filtering [44–46], can be viewed under

the coupling umbrella, the systematic exploitation/exposition of the coupling perspective in the context of Bayesian inference is more recent, including the ideas of transport maps [47–51].

At an abstract mathematical level, the coupling approach is based on a transition kernel $\psi_{\mathrm{C}}(\theta', \theta)$ such that

$$\rho_{\mathrm{post}}(\theta) = \int \psi_{\mathrm{C}}(\theta', \theta)\rho_{\mathrm{prior}}(\theta')\mathrm{d}\theta'. \tag{6}$$

The transition kernel forms a coupling between the prior and the posterior distribution and is applied only once. The induced transition from $\theta' \sim \rho_{\mathrm{prior}}$ to $\theta \sim \rho_{\mathrm{post}}$ is of the type of a McKean–Vlasov mean-field process and can be either deterministic or stochastic [45]. In practice the methodology is implemented via an approximate coupling, using linear transport maps:

$$\theta = A\theta' + b, \tag{7}$$

where the matrix A and the vector b depend on the prior distribution ρ_{prior}, the data likelihood Φ, and the data y, and are chosen such that the induced random variable θ approximately samples from the posterior distribution ρ_{post}. Many variants of the popular ensemble Kalman filter can be derived within this framework.

1.2.2. A novel algorithmic approach. The main contribution of this paper is to incorporate all three approaches from above by designing a particular (artificial) mean-field dynamical system and applying filtering methods, which employ a Gaussian ansatz, to approximate the filtering distribution resulting from partial observation of the system; the equilibrium of the filtering distribution is designed to be close to the desired posterior distribution. At an abstract level, we introduce a data-independent transition kernel, denoted by $\psi_{\mathrm{P}}(\theta'', \theta')$, and another data-dependent transition kernel, denoted by $\psi_{\mathrm{A}}(\theta'', \theta')$, such that the posterior distribution ρ_{post} remains invariant under the both transition kernels combined, that is,

$$\rho_{\mathrm{post}}(\theta) = \int \psi_{\mathrm{A}}(\theta', \theta)\left(\int \psi_{\mathrm{P}}(\theta'', \theta')\rho_{\mathrm{post}}(\theta'')\mathrm{d}\theta''\right)\mathrm{d}\theta'. \tag{8}$$

The first transition kernel, $\psi_{\mathrm{P}}(\theta'', \theta')$, corresponds to the prediction step in filtering methods and is chosen such that

$$\hat{\rho}_{n+1}(\theta) = \int \psi_{\mathrm{P}}(\theta', \theta)\rho_n(\theta')\mathrm{d}\theta' \quad \text{and} \quad \hat{\rho}_{n+1}(\theta) \propto \rho_n(\theta)^{1-\Delta\tau}, \tag{9}$$

where $0 < \Delta\tau < 1$ is the time-step size, a free parameter, and $\rho_n(\theta)$ denotes the current density. In other words, this transition kernel corresponds to a simple rescaling of a given density. The second transition kernel, $\psi_{\mathrm{A}}(\theta'', \theta')$, corresponds to the analysis step in filtering methods and has to satisfy

$$\rho_{n+1}(\theta) = \int \psi_{\mathrm{A}}(\theta', \theta)\hat{\rho}_{n+1}(\theta')\mathrm{d}\theta' \quad \text{and} \quad \rho_{n+1}(\theta) \propto \rho_{\mathrm{post}}(\theta)^{\Delta\tau}\hat{\rho}_{n+1}(\theta). \tag{10}$$

This transition kernel depends on the data and the posterior distribution and performs a suitably modified Bayesian inference step. Combining the two preceding displays yields

$$\rho_{n+1}(\theta) \propto \rho_{\mathrm{post}}(\theta)^{\Delta\tau}\rho_n(\theta)^{1-\Delta\tau}. \tag{11}$$

It is immediate that the overall transition $\rho_n \mapsto \rho_{n+1}$ is indeed invariant with respect to ρ_{post}; furthermore, by taking logarithms in the mapping from ρ_n to ρ_{n+1} it is possible to deduce exponential convergence to this steady state, for any $0 < \Delta\tau < 1$. In our concrete algorithm a mean field dynamical system is introduced for which equation (10) is satisfied exactly, while equation (9) is satisfied only in the linear, Gaussian setting; the resulting filtering distribution is approximated using Kalman methodology applied to filter the resulting partially observed mean-field dynamical system. We emphasize that the involved transition kernels are all of McKean–Vlasov type, that is, they depend on the distribution of the parameters θ.

There are several related approaches. We mention in this context in particular the recently proposed consensus-based methods. These sampling methods were analyzed in the context of optimization in [52]. Similar ideas were then developed for consensus based sampling (CBS) [53] based on the same principles employed here: to find a mean-field model which, in the linear Gaussian setting converges asymptotically to the posterior distribution, and then to develop implementable algorithms by employing finite particle approximations of the mean-field. Another related approach has been proposed in [54] where data assimilation algorithms are combined with stochastic dynamics in order to approximately sample from the posterior distribution ρ_{post}.

1.3. Filtering methods for inversion

Since filtering methods are at the heart of our proposed methodology, we provide here a brief summary of a few key concepts. Filtering methods may be deployed to approximate the posterior distribution given by equation (2). The inverse problem is first paired with a dynamical system for the parameter [55–58], leading to a hidden Markov model, to which filtering methods may be applied. In its most basic form, the hidden Markov model takes the form

$$\text{evolution}: \quad \theta_{n+1} = \theta_n, \tag{12a}$$

$$\text{observation}: \quad y_{n+1} = \mathcal{G}(\theta_{n+1}) + \eta_{n+1}; \tag{12b}$$

here θ_n is the unknown state vector, y_{n+1} is the output of the observation model, and $\eta_{n+1} \sim \mathcal{N}(0, \Sigma_\eta)$ is the observation error at the nth iteration. Any filtering method can be applied to estimate θ_n given observation data $\{y_\ell^\dagger\}_{\ell=1}$. The Kalman filter [59] can be applied to this setting provided the forward operator $\mathcal{G}$ is linear and the initial state θ_0 and the observation errors are Gaussian. The Kalman filter has been extended to nonlinear and non-Gaussian settings in manifold ways, including but not limited to, the extended Kalman filter (EKF, or sometimes ExKF) [60, 61], the ensemble Kalman filters (EnKF) [62–64], and the unscented Kalman (UKF) filter [58, 65]. We refer to the extended, ensemble and UKF filters as *approximate Kalman filters* to highlight the fact that, outside the linear setting where the Kalman filter [59] is exact, they are all uncontrolled approximations designed on the principle of matching first and second moments.

More precisely, the EnKF uses Monte Carlo sampling to estimate desired means and covariances empirically. Its update step is of the form (7) and can be either deterministic or stochastic. The ensemble adjustment/transform filters are particle approximations of square root filters, a deterministic approach to matching first and second moment information [66]. The UKF filter uses quadrature, and is also a deterministic method; it may also be viewed as approximating a square root filter. The stochastic EnKF on the other hand compares the data y to model generated data and its update step is intrinsically stochastic, that is, the vector b in (7) itself is random.

All of the filtering methods to estimate θ_n given $\{y_\ell^\dagger\}_{\ell=1}^n$ that we have described so far may be employed in the setting where $y_\ell^\dagger \equiv y$; repeated exposure of the parameter to the data helps the system to learn the parameter from the data. In order to maintain statistical consistency, an N-fold insertion of the same data y requires an appropriate modification of the data likelihood function and the resulting Bayesian inference step becomes

$$\rho_n(\theta) \to \rho_{n+1}(\theta) \propto \rho_n(\theta)\mathrm{e}^{-\frac{1}{N}\Phi(\theta,y)}. \tag{13}$$

Initializing with $\rho_0(\theta) = \rho_{\text{prior}}(\theta)$, after N iterations, $\rho_N(\theta)$ is equal to the posterior density. The filtering distribution for (12) recovers this exactly if $y_\ell^\dagger \equiv y$ and if the variance of η is rescaled by N; use of ensemble Kalman methods in this setting leads to approximate Bayesian inference, which is intuitively accurate when the posterior is close to Gaussian. We note that the resulting methodology can be viewed as a homotopy method, such as SMC [27] and transport variants [47], which seek to deform the prior into the posterior in one unit time with a finite number of inner steps N—foundational papers introducing ensemble Kalman methods in this context are [55, 56, 67]. Adaptive time-stepping strategies in this context are explored in [68–70]. Throughout this paper, we will denote the resulting methods as iterative extended Kalman filter, iterative ensemble Kalman filter (IEnKF), iterative unscented Kalman filter (IUKF), iterative ensemble adjustment Kalman filter (IEAKF) and iterative ensemble transport Kalman filter (IETKF).

We emphasize that multiple insertions of the same data y without the adjustment (13) of the data likelihood function, and/or over arbitrary numbers of steps, leads to the class of optimization-based Kalman inversion methods: EKI [57], Tikhonov-regularized EKI, termed TEKI [71] and unscented Kalman inversion, UKI [72]; see also [73] for recent adaptive methodologies which are variants on TEKI. These variants of the Kalman filter lead to efficient derivative-free optimization approaches to approximating the maximum likelihood estimator or maximum *a posteriori* estimator in the asymptotic limit as $n \to \infty$. The purpose of our paper is to develop similar ideas, based on iteration to infinity in n, but to tackle the problem of sampling from the posterior $\rho_{\text{post}}(\theta)$ rather than the optimization problem. To achieve these we introduce a novel mean-field stochastic dynamical system, generalizing (12) and apply ensemble Kalman methods to it. This leads to Bayesian analogues of EKI and the UKI. To avoid proliferation of nomenclature, we will also refer to these as EKI and UKI relying on the context to determine whether the optimization or Bayesian approach is being adopted; in this paper our focus is entirely on the Bayesian context. We will also use ensemble adjustment and transform filters, denoted as EAKF and ETKF, noting that these two may be applied in either the optimization (using (12)) or Bayesian (using the novel mean-field stochastic dynamical system introduced here) context, but that here we only study the Bayesian problem. The main conclusions of our work are two-fold, concerning the application of Kalman methods to solve the Bayesian inverse problem: that with carefully chosen underlying mean-field dynamical system, such that the prediction and analysis steps approximate equation (9) and replicate equation (10), iterating to infinity leads to more efficient and robust methods than the homotopy methods which transport prior to posterior in a finite number of steps; and that deterministic implementations of ensemble Kalman methods, and variants, are superior to stochastic methods.

The methods we propose are exact in the setting of linear $\mathcal{G}$ and Gaussian prior density ρ_{prior}; but, for nonlinear $\mathcal{G}$, the Kalman-based filters we employ generally do not converge to the exact posterior distribution, due to the Gaussian ansatz used when deriving the method; negative theoretical results and numerical evidence are reported in [74, 75]. Nonetheless, practical

experience demonstrates that the methodology can be effective for problems with distributions close to Gaussian, a situation which arises in many applications.

Finally, we note that we also include comparisons with the ensemble Kalman sampler [75–77], which we refer to as the EKS, an ensemble based Bayesian inversion method derived from discretizing a mean-field stochastic differential equation and which is also based on iteration to infinity, that is, on the invariance principle of the posterior distribution; and we include comparison with the CBS approach [53] mentioned above, another methodology which also iterates a mean-field dynamical system to infinity to approximate the posterior.

1.4. Our contributions

The key idea underlying this work is the development of an efficient derivative-free Bayesian inference approach based on applying Kalman-based filtering methods to a hidden Markov model arising from a novel mean-field dynamical system. Stemming from this, our main contributions are as follows[4]:

(a) In the setting of linear Gaussian inverse problems, we prove that the filtering distribution of the mean field model converges exponentially fast to the posterior distribution.

(b) We generalize the inversion methods EKI, UKI, EAKI and ETKI from the optimization to the Bayesian context by applying the relevant variants on Kalman methodologies to the novel mean-field dynamical system (Bayesian) rather than to (12) (optimization).

(c) We study and compare application of both deterministic and stochastic Kalman methods to the novel mean-field dynamical system, demonstrating that the deterministic methods (UKI, EAKI and ETKI) outperform the stochastic method (EKI); this may be attributed to smooth, noise-free approximations resulting from deterministic approaches.

(d) We demonstrate that the application of Kalman methods to the novel mean-field dynamical system outperforms the application of Kalman filters to transport/coupling models—the IEnKF, IUKF, IEAKF and IETKF approaches; this may be attributed to the exponential convergence underlying the filter for the novel mean-field dynamical system.

(e) We also demonstrate that the application of Kalman methods to the novel mean-field dynamical system outperforms the EKS, when Euler–Maruyama discretization is used, because the continuous-time formulation requires very small time-steps, and CBS which suffers from stochasticity, similarly to the EKI.

(f) We propose several strategies, including low-rank approximation and a bi-fidelity approach, to reduce the computational and memory cost.

(g) We demonstrate, on both linear and nonlinear model problems (including inference for subsurface geophysical properties in porous medium flow), that application of deterministic Kalman methods to approximate the filtering distribution of the novel mean-field dynamical system delivers mean and covariance which are close to the truth or to those obtained with the pCN MCMC method. The latter uses $\mathcal{O}(10^4)$ model evaluations or more whilst for our method only $\mathcal{O}(10)$ iterations are required with $\mathcal{O}(10)$ ensemble members, leading to only $\mathcal{O}(10^2)$ model evaluations, two orders of magnitude savings.

[4] In making these statements, we acknowledge that for linear Gaussian problems it is possible to solve the Bayesian inverse problem exactly in one step, or multiple steps, using the Kalman filter in transport/coupling mode, when initialized correctly and with a large enough ensemble. However, the transport/coupling methods are not robust to perturbations from initialization, non-Gaussianity and so forth, whereas the methods we introduce are. Our results substantiate this claim.

(h) The method is applied to perform Bayesian parameter inference of subgrid-scale parameters arising in an idealized global climate model, a problem currently far beyond the reach of state-of-the-art MCMC methods such as pCN and variants.

The remainder of the paper is organized as follows. In section 2, the mean field dynamical system, various algorithms which approximate its filtering distribution, and a complete analysis in the linear setting, are all presented. These correspond to our contributions (a) and (b). In section 3, strategies to speed up the algorithm and improve the robustness for real-world problems are presented. These correspond to our contribution (f). Numerical experiments are provided in section 4; these serve to empirically confirm the theory and demonstrate the effectiveness of the framework for Bayesian inference. These correspond to our contributions (c), (d), (e), (g) and (h). We make concluding remarks in section 5.

The code is accessible online:
https://github.com/Zhengyu-Huang/InverseProblems.jl.

1.5. Notational conventions

$A \succ B$ and $A \succeq B$ denote $A - B$ positive-definite or positive-semidefinite, for symmetric matrices A, B. $\|\cdot\|, \langle\cdot,\cdot\rangle$ denote Euclidean norm and inner-product. We use $\mathbb{Z}^+ = \{0, 1, 2, \ldots\}$ to denote the set of natural numbers; $\mathcal{N}(\cdot,\cdot)$ to denote Gaussian distributions; and $\varrho(\cdot)$ to denote the spectral radius. As encountered in subsection 1.2 we make use of the similar symbol ρ for densities; these should be easily distinguished from spectral radius by context and by a different font.

2. Novel algorithmic methodology

Our novel algorithmic methodology is introduced in this section. We first introduce the underlying mean-field dynamical system, which has prediction and analysis steps corresponding to the aforementioned transition kernels, in subsection 2.1. Then, in subsection 2.2, we introduce a class of conceptual Gaussian approximation algorithms found by applying Kalman methodology to the proposed mean-field dynamical system. Through linear analysis, we prove in subsection 2.3 that these algorithms converge exponentially to the posterior. For the nonlinear setting, a variety of nonlinear Kalman inversion methodologies are discussed in subsection 2.4.

2.1. Mean-field dynamical system

Following the discussion from section 1.2.2, we propose an implementation of (8) to solve inverse problems by pairing the parameter-to-data map with a dynamical system for the parameter, and then employ techniques from filtering to estimate the parameter given the data.

We introduce the prediction step

$$\theta_{n+1} = \theta_n + \omega_{n+1}. \tag{14}$$

Here θ_{n+1} is the unknown state vector and $\omega_{n+1} \sim \mathcal{N}(0, \Sigma_{\omega,n+1})$ is the independent, zero-mean Gaussian evolution error, which will be chosen such that (14) mimics (9) for Gaussian densities. The analysis step (10) follows exactly from the observation model (first introduced in [71])

$$x_{n+1} = \mathcal{F}(\theta_{n+1}) + \nu_{n+1}; \tag{15}$$

here we have defined the augmented forward map

$$\mathcal{F}(\theta) = \begin{bmatrix} \mathcal{G}(\theta) \\ \theta \end{bmatrix}, \tag{16}$$

with $\nu_{n+1} \sim \mathcal{N}(0, \Sigma_{\nu,n+1})$ the independent, zero-mean Gaussian observation error, and x_{n+1} the output of the observation model at time $n+1$. We define artificial observation $x^\dagger_{n+1}$ using the following particular instance of the data, constructed from the one observation y and the prior mean r_0 and assumed to hold for all $n \geqslant 1$:

$$x^\dagger_{n+1} = x := \begin{bmatrix} y \\ r_0 \end{bmatrix}. \tag{17}$$

We will apply filtering methods to condition θ_n on $Y_n := \{x^\dagger_1, x^\dagger_2, \ldots, x^\dagger_n\}$, the observation set at time n. As we will see later, the choice of $\{x^\dagger_\ell\}_{l=1}$ leads to the correct posterior.

Let C_n denote the covariance of the conditional random variable $\theta_n|Y_n$. Then the error covariance matrices $\{\Sigma_{\omega,n+1}\}$ and $\{\Sigma_{\nu,n+1}\}$ in the extended dynamical system (14) and (15) are chosen at the nth iteration, as follows:

$$\Sigma_{\nu,n+1} = \frac{1}{\Delta\tau}\begin{bmatrix} \Sigma_\eta & 0 \\ 0 & \Sigma_0 \end{bmatrix} \quad \text{and} \quad \Sigma_{\omega,n+1} = \frac{\Delta\tau}{1-\Delta\tau} C_n. \tag{18}$$

Here $0 < \Delta\tau < 1$, and in our numerical studies we choose $\Delta\tau = 1/2$, although other choices are possible. Since the artificial evolution error covariance $\Sigma_{\omega,n+1}$ in (14) is updated based on C_n, the conditional covariance of $\theta_n|Y_n$, it follows that (14) is a mean-field dynamical system: it depends on its own law, specifically on the law of $\theta_n|Y_n$. Details underpinning the choices of the error covariance matrices $\{\Sigma_{\omega,n}\}$ and $\{\Sigma_{\nu,n}\}$ are given in subsections 2.2 and 2.3: the matrices are chosen so that, for linear Gaussian problems, the prediction and analysis steps follow equations (9) and (10), and the converged mean and covariance of the resulting filtering distribution for $\theta_n|Y_n$ under the prediction step (14) and the observation model (15) match the posterior mean and covariance.

2.2. Gaussian approximation

Denote by ρ_n, the conditional density of $\theta_n|Y_n$. We first introduce a class of conceptual Kalman inversion algorithms which approximate ρ_n by considering only first and second order statistics (mean and covariance), and update ρ_n sequentially using the standard prediction and analysis steps [45, 46]: $\rho_n \mapsto \hat{\rho}_{n+1}$, and then $\hat{\rho}_{n+1} \mapsto \rho_{n+1}$, where $\hat{\rho}_{n+1}$ is the distribution of $\theta_{n+1}|Y_n$. The second analysis step is performed by invoking a Gaussian hypothesis. In subsequent subsections, we then apply different methods to approximate the resulting maps on measures, leading to unscented, stochastic ensemble Kalman and adjustment/transform square root Kalman filters.

In the prediction step, assume that $\rho_n \approx \mathcal{N}(m_n, C_n)$, then under equation (14), $\hat{\rho}_{n+1} = \mathcal{N}(\widehat{m}_{n+1}, \widehat{C}_{n+1})$ is also Gaussian and satisfies

$$\widehat{m}_{n+1} = \mathbb{E}[\theta_{n+1}|Y_n] = m_n \qquad \widehat{C}_{n+1} = \mathrm{Cov}[\theta_{n+1}|Y_n] = C_n + \Sigma_{\omega,n+1}. \tag{19}$$

In the analysis step, we assume that the joint distribution of $\{\theta_{n+1}, x_{n+1}\}|Y_n$ can be approximated by a Gaussian distribution

$$\mathcal{N}\left(\begin{bmatrix} \widehat{m}_{n+1} \\ \widehat{x}_{n+1} \end{bmatrix}, \begin{bmatrix} \widehat{C}_{n+1} & \widehat{C}^{\theta x}_{n+1} \\ \widehat{C}^{\theta x\ \mathrm{T}}_{n+1} & \widehat{C}^{xx}_{n+1} \end{bmatrix}\right), \tag{20}$$

where

$$\begin{aligned}\widehat{x}_{n+1} &= \mathbb{E}[x_{n+1}|Y_n] = \mathbb{E}[\mathcal{F}(\theta_{n+1})|Y_n],\\ \widehat{C}^{\theta x}_{n+1} &= \mathrm{Cov}[\theta_{n+1}, x_{n+1}|Y_n] = \mathrm{Cov}[\theta_{n+1}, \mathcal{F}(\theta_{n+1})|Y_n],\\ \widehat{C}^{xx}_{n+1} &= \mathrm{Cov}[x_{n+1}|Y_n] = \mathrm{Cov}[\mathcal{F}(\theta_{n+1})|Y_n] + \Sigma_{\nu,n+1}.\end{aligned} \tag{21}$$

These expectations are computed by assuming $\theta_{n+1}|Y_n \sim \hat{\rho}_{n+1}$ and noting that the distribution of (θ_{n+1}, x_{n+1}) is then defined by (14) and (15). This corresponds to projecting[5] the joint distribution onto the Gaussian which matches its mean and covariance. Conditioning the Gaussian in equation (20) to find $\theta_{n+1}|\{Y_n, x^\dagger_{n+1}\} = \theta_{n+1}|Y_{n+1}$, gives the following expressions for the mean m_{n+1} and covariance C_{n+1} of the approximation to ρ_{n+1}:

$$m_{n+1} = \widehat{m}_{n+1} + \widehat{C}^{\theta x}_{n+1}(\widehat{C}^{xx}_{n+1})^{-1}(x^\dagger_{n+1} - \widehat{x}_{n+1}), \tag{22a}$$

$$C_{n+1} = \widehat{C}_{n+1} - \widehat{C}^{\theta x}_{n+1}(\widehat{C}^{xx}_{n+1})^{-1}\widehat{C}^{\theta x\,\mathrm{T}}_{n+1}. \tag{22b}$$

Equations (19) to (21), (22a) and (22b) establish a class of conceptual algorithms for application of Gaussian approximation to solve the inverse problems. To make implementable algorithms a high level choice needs to be made: whether to work strictly within the class of Gaussians, that is to impose $\rho_n \equiv \mathcal{N}(m_n, C_n)$, or whether to allow non-Gaussian ρ_n but to insist that the second order statistics of the resulting measures agree with equations (19) to (21), (22a) and (22b). In what follows the UKI takes the first perspective; all other methods take the second perspective. For the UKI the method views equations (19) to (21), (22a) and (22b) as providing a nonlinear map $(m_n, C_n) \mapsto (m_{n+1}, C_{n+1})$; this map is then approximated using quadrature. For the remaining methods a mean-field dynamical system is used, which is non-Gaussian but matches the aforementioned Gaussian statistics; this mean-field model is then approximated by a finite particle system [79]. The dynamical system is of mean-field type because of the expectations required to calculate equations (20), (21) and (18). The continuous time limit of the evolution for the mean and covariance is presented in appendix A; this is obtained by letting $\Delta\tau \to 0$.

Remark 1. Consider the case, where $\rho_n = \mathcal{N}(m_n, C_n)$ is Gaussian. With the specific choice of $\{\Sigma_{\omega,n}\}$, we have $\hat{\rho}_{n+1} = \mathcal{N}(m_n, \frac{1}{1-\Delta\tau}C_n)$ from the prediction step equation (19), and hence the Gaussian density functions ρ_n and $\hat{\rho}_{n+1}$ fulfill equation (9). With the extended observation model (15) and the specific choice of $\{\Sigma_{\nu,n}\}$, the analysis step without Gaussian approximation can be written as

$$\begin{aligned}\rho(\theta_{n+1}|Y_{n+1}) &\propto \rho(\theta_{n+1}|Y_n)\rho(x^\dagger_{n+1}|\theta_{n+1}, Y_n)\\ &\propto \hat{\rho}_{n+1}(\theta_{n+1})\mathrm{e}^{-\Delta\tau\Phi_R(\theta_{n+1},y)}\\ &\propto \hat{\rho}_{n+1}(\theta_{n+1})\rho(\theta_{n+1})^{\Delta\tau}\end{aligned} \tag{23}$$

and hence the density functions $\hat{\rho}_{n+1}$ and ρ_{n+1} fulfill equation (10). Note, however, that ρ_{n+1} is not, in general, Gaussian, unless $\mathcal{G}$ is linear, the case studied in the next section. In the nonlinear case, we employ Kalman-based methodology which only employs first and second order statistics, and in effect projects ρ_{n+1} onto a Gaussian.

[5] We use the term 'projecting' as finding the Gaussian p which matches the first and second moments of a given measure π corresponds to finding the closest Gaussian p to π with respect to variation in the second argument of the (nonsymmetric) Kullback–Leibler divergence [78, theorem 4.5].

2.3. Linear analysis

In this subsection, we study the algorithm in the context of linear inverse problems, for which $\mathcal{G}(\theta) = G\theta$ for some matrix G. Furthermore we assume that ρ_{prior} is Gaussian $\mathcal{N}(r_0, \Sigma_0)$ and recall that the observational noise is $\mathcal{N}(0, \Sigma_\eta)$. Thanks to the linear Gaussian structure the posterior is also Gaussian with mean and precisions given by

$$\begin{aligned} m_{\text{post}} &= r_0 + \left(G^{\mathrm{T}}\Sigma_\eta^{-1}G + \Sigma_0^{-1}\right)^{-1} G^{\mathrm{T}}\Sigma_\eta^{-1}(y - Gr_0) \quad \text{and} \\ C_{\text{post}}^{-1} &= G^{\mathrm{T}}\Sigma_\eta^{-1}G + \Sigma_0^{-1}. \end{aligned} \tag{24}$$

Furthermore, the equation (21) reduce to

$$\begin{aligned} &\widehat{x}_{n+1} = Fm_n, \qquad \widehat{C}_{n+1}^{\theta x} = \widehat{C}_{n+1}F^{\mathrm{T}}, \quad \text{and} \\ &\widehat{C}_{n+1}^{xx} = F\widehat{C}_{n+1}F^{\mathrm{T}} + \Sigma_{\nu,n+1} \quad \text{where } F = \begin{bmatrix} G \\ I \end{bmatrix}. \end{aligned}$$

We note that

$$\langle Fv, Fv\rangle \geqslant \|v\|^2. \tag{25}$$

The update equations (22a) and (22b) become

$$m_{n+1} = m_n + \widehat{C}_{n+1}F^{\mathrm{T}}(F\widehat{C}_{n+1}F^{\mathrm{T}} + \Sigma_{\nu,n+1})^{-1}(x - Fm_n), \tag{26a}$$

$$C_{n+1} = \widehat{C}_{n+1} - \widehat{C}_{n+1}F^{\mathrm{T}}(F\widehat{C}_{n+1}F^{\mathrm{T}} + \Sigma_{\nu,n+1})^{-1}F\widehat{C}_{n+1}, \tag{26b}$$

with $\widehat{C}_{n+1} = C_n + \Sigma_{\omega,n+1}$. We have the following theorem about the convergence of the algorithm:

Theorem 1. *Assume that the error covariance matrices are as defined in equation* (18) *with* $0 < \Delta\tau < 1$ *and that the prior covariance matrix* $\Sigma_0 \succ 0$ *and initial covariance matrix* $C_0 \succ 0$. *The iteration for the conditional mean* m_n *and precision matrix* C_n^{-1} *characterizing the distribution of* $\theta_n|Y_n$ *converges exponentially fast to limit* m_∞, C_∞^{-1}. *Furthermore the limiting mean* m_∞ *and precision matrix* $C_\infty^{-1} = G^{\mathrm{T}}\Sigma_\eta^{-1}G + \Sigma_0^{-1}$ *are the posterior mean and precision matrix given by* (24).

Proof. With the error covariance matrices defined in equation (18), the update equation for $\{C_n\}$ in equation (26b) can be rewritten as

$$\begin{aligned} C_{n+1}^{-1} &= F^{\mathrm{T}}\Sigma_{\nu,n+1}^{-1}F + (C_n + \Sigma_{\omega,n+1})^{-1} \\ &= \Delta\tau\left(G^{\mathrm{T}}\Sigma_\eta^{-1}G + \Sigma_0^{-1}\right) + (1 - \Delta\tau)C_n^{-1}. \end{aligned} \tag{27}$$

We thus have a closed formula for C_n^{-1}:

$$C_n^{-1} = \left[1 - (1 - \Delta\tau)^n\right]\left(G^{\mathrm{T}}\Sigma_\eta^{-1}G + \Sigma_0^{-1}\right) + (1 - \Delta\tau)^n C_0^{-1}. \tag{28}$$

Since $0 < \Delta\tau < 1$ this leads to the exponential convergence $\lim_{n\to\infty} C_n^{-1} = G^{\mathrm{T}}\Sigma_\eta^{-1}G + \Sigma_0^{-1} = C_{\text{post}}^{-1}$ given by (24).

Since we have made a choice independent of n we write $\Sigma_\nu := \Sigma_{\nu,n+1}$. Thus equations (27) and (28) lead to

$$F^{\mathrm{T}}\Sigma_\nu^{-1}F \preceq C_{n+1}^{-1} \preceq F^{\mathrm{T}}\Sigma_\nu^{-1}F + \Sigma_+ \quad \text{where } \Sigma_+ = \frac{1-\Delta\tau}{\Delta\tau}F^{\mathrm{T}}\Sigma_\nu^{-1}F + C_0^{-1}. \tag{29}$$

The update equation of m_n in equation (26a) can be rewritten as

$$m_{n+1} = m_n + C_{n+1}F^{\mathrm{T}}\Sigma_\nu^{-1}(x - Fm_n). \tag{30}$$

Note that $B := F^{\mathrm{T}}\Sigma_\nu^{-1}F$ is symmetric and that, as a consequence of (25) together with the fact that $\Sigma_\nu \succ 0$, it follows that $B \succ 0$; thus we have that $I - C_{n+1}B$ has the same spectrum as $I - B^{\frac{1}{2}}C_{n+1}B^{\frac{1}{2}}$. Using the upper bound on C_{n+1} appearing in equation (29), the spectral radius of the update matrix in equation (30) satisfies

$$\begin{aligned}
\varrho(\mathbb{I} - C_{n+1}F^{\mathrm{T}}\Sigma_\nu^{-1}F) &= \varrho(\mathbb{I} - C_{n+1}B) \\
&= \varrho\left(\mathbb{I} - B^{\frac{1}{2}}C_{n+1}B^{\frac{1}{2}}\right) \\
&\leqslant 1 - \varrho\left(B^{\frac{1}{2}}\left(B + \Sigma_+\right)^{-1}B^{\frac{1}{2}}\right) \\
&= 1 - \epsilon_0,
\end{aligned} \tag{31}$$

where $\epsilon_0 \in (0,1)$. Hence, we deduce that $\{m_n\}$ converges exponentially to the stationary point m_∞, which satisfies $F^{\mathrm{T}}\Sigma_\nu^{-1}(x - Fm_\infty) = 0$. Using the structure of F and Σ_ν the limiting mean can be written as the posterior mean given in (24):

$$m_\infty = r_0 + \left(G^{\mathrm{T}}\Sigma_\eta^{-1}G + \Sigma_0^{-1}\right)^{-1}G^{\mathrm{T}}\Sigma_\eta^{-1}(y - Gr_0) = m_{\mathrm{post}}. \tag{32}$$

□

Remark 2. Although this theorem applies only to the linear Gaussian setting we note that the premise of matching only first and second order moments is inherent to all Kalman methods. We demonstrate numerically in section 4 that application of the filtering methodology based on the proposed choices of covariances leads to approximated mean and covariances which are accurate for nonlinear inverse problems.

Remark 3. We note that the convergence of the means/covariances of the Kalman filter is a widely studied topic; and variants on some of our results can be obtained from the existing literature, for example, the use of contraction mapping arguments to study convergence of the Kalman filter is explored in [80, 81].

2.4. Nonlinear Kalman inversion methodologies

To make practical methods for solving nonlinear inverse problems (1) out of the foregoing, the expectations (integrals) appearing in the prediction step (19) as well as in the analysis step via equation (21) need to be approximated appropriately. While equation (19) can be implemented via a simple rescaling of the covariance matrix or ensemble, respectively, (we use both) the analysis step can be implemented using *any* nonlinear Kalman filter (we use a variety). In the present work, we focus on both the unscented and EnKF, which lead to the Bayesian

implementations of unscented Kalman inversion (UKI), stochastic ensemble Kalman inversion (EKI), ensemble adjustment Kalman inversion (EAKI), and ensemble transform Kalman inversion (ETKI). We now detail these methods[6].

2.4.1. Unscented Kalman inversion (UKI). UKI approximates the integrals in equation (21) by means of deterministic quadrature rules; this is the idea of the unscented transform [58, 65]. We now define this precisely in the versions used in this paper.

Definition 1 (Modified unscented transform [72]). Consider Gaussian random variable $\theta \sim \mathcal{N}(m, C) \in \mathbb{R}^{N_\theta}$. Define J sigma points $\{\theta^j\}_{j=0}^{J-1}$ according to the deterministic formulae

$$\theta^0 = m \qquad \theta^j = m + [\sqrt{C}]I_{N_\theta}[:, j] \quad (1 \leqslant j \leqslant J-1); \tag{33}$$

here $[\sqrt{C}]$ is the Cholesky factor of C and $I_{N_\theta}[:, j]$ is the jth column of the matrix $I_{N_\theta} \in \mathbb{R}^{N_\theta \times (J-1)}$. Consider any two real vector-valued functions $\mathcal{F}_1(\cdot)$ and $\mathcal{F}_2(\cdot)$ acting on $\mathbb{R}^{N_\theta}$. Using the sigma points we may define a quadrature rule approximating the mean and covariance of the random variables $\mathcal{F}_1(\theta)$ and $\mathcal{F}_2(\theta)$ as follows:

$$\begin{aligned} \mathbb{E}[\mathcal{F}_i(\theta)] &\approx \mathcal{F}_i(\theta^0) \\ \mathrm{Cov}[\mathcal{F}_1(\theta), \mathcal{F}_2(\theta] &\approx \sum_{j=1}^{J-1} a\big(\mathcal{F}_1(\theta^j) - \mathbb{E}\mathcal{F}_1(\theta)\big)\big(\mathcal{F}_2(\theta^j) - \mathbb{E}\mathcal{F}_2(\theta)\big)^{\mathrm{T}}. \end{aligned} \tag{34}$$

In the present work, we consider the following two variants,

- UKI-1 ($J = N_\theta + 2$) [12, 82]. I_{N_θ} is defined recursively as

$$I_1 = \begin{bmatrix} -\dfrac{1}{\sqrt{2a}} & \dfrac{1}{\sqrt{2a}} \end{bmatrix} \tag{35}$$

$$I_d = \begin{bmatrix} & & & 0 \\ & I_{d-1} & & \vdots \\ & & & 0 \\ \dfrac{1}{\sqrt{ad(d+1)}} & \cdots & \dfrac{1}{\sqrt{ad(d+1)}} & \dfrac{-d}{\sqrt{ad(d+1)}} \end{bmatrix}, \quad 2 \leqslant d \leqslant N_\theta \tag{36}$$

and the weight parameter is chosen as $a = \frac{N_\theta}{4(N_\theta+1)}$.

[6] Recall the discussion in subsection 1.3 about distinction between optimization and Bayesian implementations of all these methods.

- UKI-2 ($J = 2N_\theta + 1$) [72]. I_{N_θ} is defined as

$$I_{N_\theta} = \begin{bmatrix} \frac{1}{\sqrt{2a}} & & & & -\frac{1}{\sqrt{2a}} & & & \\ & \frac{1}{\sqrt{2a}} & & & & -\frac{1}{\sqrt{2a}} & & \\ & & \ddots & & & & \ddots & \\ & & & \frac{1}{\sqrt{2a}} & & & & -\frac{1}{\sqrt{2a}} \end{bmatrix} \tag{37}$$

  and the weight parameter is chosen as $a = \max\{\frac{1}{8}, \frac{1}{2N_\theta}\}$.

Consider the Gaussian approximation algorithm defined by equations (19) to (21), (22a) and (22b). By utilizing the aforementioned quadrature rule, the iteration procedure of the UKI becomes:

- Prediction step:

$$\widehat{m}_{n+1} = m_n \qquad \widehat{C}_{n+1} = \frac{1}{1-\Delta\tau} C_n. \tag{38}$$

- Generate sigma points:

$$\widehat{\theta}^0_{n+1} = \widehat{m}_{n+1}, \qquad \widehat{\theta}^j_{n+1} = \widehat{m}_{n+1} + [\sqrt{\widehat{C}_{n+1}}] I_{N_\theta}[:, j] \quad (1 \leqslant j \leqslant J-1). \tag{39}$$

- Analysis step:

$$\begin{aligned} \widehat{x}^j_{n+1} &= \mathcal{F}(\widehat{\theta}^j_{n+1}) \quad (0 \leqslant j \leqslant J-1), \\ \widehat{x}_{n+1} &= \widehat{x}^0_{n+1}, \\ \widehat{C}^{\theta x}_{n+1} &= \sum_{j=1}^{J-1} a(\widehat{\theta}^j_{n+1} - \widehat{m}_{n+1})(\widehat{x}^j_{n+1} - \widehat{x}_{n+1})^{\mathrm{T}}, \\ \widehat{C}^{xx}_{n+1} &= \sum_{j=1}^{J-1} a(\widehat{x}^j_{n+1} - \widehat{x}_{n+1})(\widehat{x}^j_{n+1} - \widehat{x}_{n+1})^{\mathrm{T}} + \Sigma_{\nu,n+1}, \\ m_{n+1} &= \widehat{m}_{n+1} + \widehat{C}^{\theta x}_{n+1}(\widehat{C}^{xx}_{n+1})^{-1}(x - \widehat{x}_{n+1}), \\ C_{n+1} &= \widehat{C}_{n+1} - \widehat{C}^{\theta x}_{n+1}(\widehat{C}^{xx}_{n+1})^{-1}\widehat{C}^{\theta x\ \mathrm{T}}_{n+1}. \end{aligned} \tag{40a}$$

2.4.2. Ensemble Kalman inversion. Ensemble Kalman inversion represents the distribution at each iteration by an ensemble of parameter estimates $\{\theta^j_n\}_{j=1}^J$ and approximates the integrals in equation (21) empirically. We describe three variants on this idea.

Stochastic ensemble Kalman inversion (EKI). The perturbed observations form of the ensemble Kalman filter [83] is applied to the extended mean-field dynamical system (14) and (15), which leads to the EKI:

- Prediction step:

$$\widehat{m}_{n+1} = m_n, \qquad \widehat{\theta}^j_{n+1} = \widehat{m}_{n+1} + \sqrt{\frac{1}{1-\Delta\tau}}(\theta^j_n - m_n). \tag{41}$$

- Analysis step:

$$\widehat{x}^j_{n+1} = \mathcal{F}(\widehat{\theta}^j_{n+1}) \qquad \widehat{x}_{n+1} = \frac{1}{J}\sum_{j=1}^{J} \widehat{x}^j_{n+1}, \tag{42a}$$

$$\widehat{C}^{\theta x}_{n+1} = \frac{1}{J-1}\sum_{j=1}^{J} (\widehat{\theta}^j_{n+1} - \widehat{m}_{n+1})(\widehat{x}^j_{n+1} - \widehat{x}_{n+1})^{\mathrm{T}}, \tag{42b}$$

$$\widehat{C}^{xx}_{n+1} = \frac{1}{J-1}\sum_{j=1}^{J} (\widehat{x}^j_{n+1} - \widehat{x}_{n+1})(\widehat{x}^j_{n+1} - \widehat{x}_{n+1})^{\mathrm{T}} + \Sigma_{\nu,n+1}, \tag{42c}$$

$$\theta^j_{n+1} = \widehat{\theta}^j_{n+1} + \widehat{C}^{\theta x}_{n+1}\left(\widehat{C}^{xx}_{n+1}\right)^{-1}(x - \widehat{x}^j_{n+1} - \nu^j_{n+1}), \tag{42d}$$

$$m_{n+1} = \frac{1}{J}\sum_{j=1}^{J} \theta^j_{n+1}. \tag{42e}$$

Here the superscript $j = 1, \ldots, J$ is the ensemble particle index, and $\nu^j_{n+1} \sim \mathcal{N}(0, \Sigma_{\nu,n+1})$ are independent and identically distributed random variables. The prediction step ensures the exactness of the predictive covariance equation (19).

Remark 4. The prediction step (41) is inspired by square root Kalman filters [63, 64, 66, 84] and covariance inflation [85]; these methods are designed to ensure that the mean and covariance of $\{\widehat{\theta}^j_{n+1}\}$ match $\widehat{m}_{n+1}$ and $\widehat{C}_{n+1}$ exactly. This is different from traditional stochastic ensemble Kalman inversion implementation, where i.i.d. Gaussian noises $\omega^j_{n+1} \sim \mathcal{N}(0, \Sigma_{\omega,n+1})$ are added. In the analysis step (42), we add noise in the $\{\theta^j_{n+1}\}$ update (42d) instead of the $\{\widehat{x}^j_{n+1}\}$ evaluation (42a); this ensures that $\widehat{C}^{xx}_{n+1}$ (42c) is symmetric positive definite.

Remark 5. As a precursor to understanding the adjustment and transform filters which follow this subsection, we show that the EKI does not exactly replicate the covariance update equation (22b). To this end, denote the matrix square roots $\widehat{Z}_{n+1}, Z_{n+1} \in \mathbb{R}^{N_\theta \times J}$ of $\widehat{C}_{n+1}, C_{n+1}$ and $\widehat{\mathcal{Y}}_{n+1}$ as follows:

$$\begin{aligned}
\widehat{Z}_{n+1} &= \frac{1}{\sqrt{J-1}}\left(\widehat{\theta}^1_{n+1} - \widehat{m}_{n+1} \quad \widehat{\theta}^2_{n+1} - \widehat{m}_{n+1} \quad \ldots \quad \widehat{\theta}^J_{n+1} - \widehat{m}_{n+1}\right), \\
Z_{n+1} &= \frac{1}{\sqrt{J-1}}\left(\theta^1_{n+1} - m_{n+1} \quad \theta^2_{n+1} - m_{n+1} \quad \ldots \quad \theta^J_{n+1} - m_{n+1}\right), \\
\widehat{\mathcal{Y}}_{n+1} &= \frac{1}{\sqrt{J-1}}\left(\widehat{x}^1_{n+1} - \widehat{x}_{n+1} \quad \widehat{x}^2_{n+1} - \widehat{x}_{n+1} \quad \ldots \quad \widehat{x}^J_{n+1} - \widehat{x}_{n+1}\right).
\end{aligned} \tag{43}$$

Then the covariance update equation (22b) does not hold exactly:

$$\begin{aligned}\widehat{C}_{n+1} - \widehat{C}^{\theta x}_{n+1}(\widehat{C}^{xx}_{n+1})^{-1}\widehat{C}^{\theta x\ \mathrm{T}}_{n+1} &= \widehat{Z}_{n+1}\widehat{Z}^{\mathrm{T}}_{n+1} - \widehat{Z}_{n+1}\widehat{\mathcal{Y}}^{\mathrm{T}}_{n+1} \\ &\quad \times (\widehat{\mathcal{Y}}_{n+1}\widehat{\mathcal{Y}}^{\mathrm{T}}_{n+1} + \Sigma_{\nu,n+1})^{-1}\widehat{\mathcal{Y}}_{n+1}\widehat{Z}^{\mathrm{T}}_{n+1} \\ &\neq Z_{n+1}Z^{\mathrm{T}}_{n+1} = C_{n+1}. \end{aligned} \tag{44}$$

Ensemble EAKI. Following the ensemble adjustment Kalman filter proposed in [63], the analysis step updates particles deterministically with a pre-multiplier A,

$$\theta^j_{n+1} - m_{n+1} = A(\widehat{\theta}^j_{n+1} - \widehat{m}_{n+1}). \tag{45}$$

Here $A = P\hat{D}^{\frac{1}{2}}UD^{\frac{1}{2}}\hat{D}^{-\frac{1}{2}}P^{\mathrm{T}}$ with

$$\begin{aligned} &\text{SVD}: \quad \widehat{Z}_{n+1} = P\hat{D}^{\frac{1}{2}}V^{\mathrm{T}}, \\ &\text{SVD}: \quad V^{\mathrm{T}}\left(\mathbb{I} + \widehat{\mathcal{Y}}^{\mathrm{T}}_{n+1}\Sigma^{-1}_{\nu,n+1}\widehat{\mathcal{Y}}_{n+1}\right)^{-1}V = UDU^{\mathrm{T}}, \end{aligned} \tag{46}$$

where both $\hat{D}$ and D are non-singular diagonal matrices, with dimensionality rank $(\widehat{Z}_{n+1})$, and $\widehat{Z}_{n+1}$ and $\widehat{\mathcal{Y}}_{n+1}$ are defined in equation (43). The analysis step becomes:

- Analysis step:

$$m_{n+1} = \widehat{m}_{n+1} + \widehat{C}^{\theta x}_{n+1}\left(\widehat{C}^{xx}_{n+1}\right)^{-1}(x - \widehat{x}_{n+1}), \tag{47a}$$

$$\theta^j_{n+1} = m_{n+1} + A(\widehat{\theta}^j_{n+1} - \widehat{m}_{n+1}). \tag{47b}$$

Remark 6. It can be verified that the covariance update equation (22b) holds:

$$\begin{aligned} C_{n+1} &= Z_{n+1}Z^{\mathrm{T}}_{n+1} \\ &= A\widehat{Z}_{n+1}\widehat{Z}^{\mathrm{T}}_{n+1}A^{\mathrm{T}} \\ &= P\hat{D}^{\frac{1}{2}}UDU^{\mathrm{T}}\hat{D}^{\frac{1}{2}}P \\ &= \widehat{Z}_{n+1}\left(\mathbb{I} + \widehat{\mathcal{Y}}^{\mathrm{T}}_{n+1}\Sigma^{-1}_{\nu,n+1}\widehat{\mathcal{Y}}_{n+1}\right)^{-1}\widehat{Z}^{\mathrm{T}}_{n+1} \\ &= \widehat{Z}_{n+1}\left(\mathbb{I} - \widehat{\mathcal{Y}}^{\mathrm{T}}_{n+1}(\widehat{\mathcal{Y}}_{n+1}\widehat{\mathcal{Y}}^{\mathrm{T}}_{n+1} + \Sigma_{\nu,n+1})^{-1}\widehat{\mathcal{Y}}_{n+1}\right)\widehat{Z}^{\mathrm{T}}_{n+1} \\ &= \widehat{C}_{n+1} - \widehat{C}^{\theta x}_{n+1}(\widehat{C}^{xx}_{n+1})^{-1}\widehat{C}^{\theta x\ \mathrm{T}}_{n+1}. \end{aligned} \tag{48}$$

ETKI. Following the ensemble transform Kalman filter proposed in [64, 66, 84], the analysis step updates particles deterministically with a post-multiplier T,

$$Z_{n+1} = \widehat{Z}_{n+1}T. \tag{49}$$

Here $T = P(\Gamma + I)^{-\frac{1}{2}}P^{\mathrm{T}}$, with

$$\text{SVD:} \quad \widehat{\mathcal{Y}}_{n+1}\Sigma^{-1}_{\nu,n+1}\widehat{\mathcal{Y}}_{n+1} = P\Gamma P^{\mathrm{T}}. \tag{50}$$

The analysis step becomes:

- Analysis step:

$$m_{n+1} = \widehat{m}_{n+1} + \widehat{C}^{\theta x}_{n+1}\left(\widehat{C}^{xx}_{n+1}\right)^{-1}(x - \widehat{x}_{n+1}), \tag{51a}$$

$$Z_{n+1} = \widehat{Z}_{n+1}T. \tag{51b}$$

Remark 7. It can be verified that the covariance update equation (22b) holds:

$$\begin{aligned} C_{n+1} &= Z_{n+1}Z_{n+1}^{\mathrm{T}} \\ &= \widehat{Z}_{n+1}TT^{\mathrm{T}}\widehat{Z}_{n+1}^{\mathrm{T}} \\ &= \widehat{Z}_{n+1}\left(\mathbb{I} + P\Gamma P^{\mathrm{T}}\right)^{-1}\widehat{Z}_{n+1}^{\mathrm{T}} \\ &= \widehat{Z}_{n+1}\left(\mathbb{I} + \widehat{\mathcal{Y}}_{n+1}^{\mathrm{T}}\Sigma_{\nu,n+1}^{-1}\widehat{\mathcal{Y}}_{n+1}\right)^{-1}\widehat{Z}_{n+1}^{\mathrm{T}} \\ &= \widehat{Z}_{n+1}\left(\mathbb{I} - \widehat{\mathcal{Y}}_{n+1}^{\mathrm{T}}(\widehat{\mathcal{Y}}_{n+1}\widehat{\mathcal{Y}}_{n+1}^{\mathrm{T}} + \Sigma_{\nu,n+1})^{-1}\widehat{\mathcal{Y}}_{n+1}\right)\widehat{Z}_{n+1}^{\mathrm{T}} \\ &= \widehat{C}_{n+1} - \widehat{C}^{\theta x}_{n+1}(\widehat{C}^{xx}_{n+1})^{-1}\widehat{C}^{\theta x\ \mathrm{T}}_{n+1}. \end{aligned} \tag{52}$$

Particles (ensemble members) updated by the basic form of the EKI algorithm through iterates are confined to the linear span of the initial ensemble [55, 57]. The same is true for both EAKI and ETKI:

Lemma 1. *For both EAKI and ETKI, all particles lie in the linear space $\mathcal{A}$ spanned by m_0 and the column vectors of Z_0.*

Proof. We will prove that m_n and column vectors of Z_n are in $\mathcal{A}$ by induction. We assume this holds for all $n \leqslant k$. Since $\widehat{m}_{k+1} = m_k$ and $\widehat{Z}_{k+1} = \sqrt{\frac{1}{1-\Delta\tau}}Z_k$ (see equation (41)), $\widehat{m}_{k+1}$ and column vectors of $\widehat{Z}_{k+1}$ are in $\mathcal{A}$. Combining the mean update equations (47a) and (51a) and the fact that $\widehat{C}^{\theta x}_{k+1} = \widehat{Z}_{k+1}\widehat{\mathcal{Y}}^{\mathrm{T}}_{k+1}$, we have m_{k+1} is in $\mathcal{A}$. For EAKI, since the pre-multiplier $A = P\hat{D}^{\frac{1}{2}}UD^{\frac{1}{2}}\hat{D}^{-\frac{1}{2}}P^{\mathrm{T}}$, and P is the left compact singular matrix of $\widehat{Z}_{k+1}$, it follows that the column vectors of A lie in $\mathcal{A}$; furthermore, the square root matrix update equation (47b), $Z_{k+1} = A\widehat{Z}_k$, has implication that the column vectors of Z_{k+1} lie in $\mathcal{A}$. For the ETKI, the square root matrix update equation (51b) implies that the column vectors of Z_{k+1} lie in $\mathcal{A}$. Since m_n and column vectors of Z_n are in $\mathcal{A}$, so are the particles $\{\theta_n^j\}_{j=1}^J$. □

3. Variants on the basic algorithm

In this section, we introduce three strategies to make the novel mean-field based methodology more efficient, robust and widely applicable in real large-scale problems. In subsection 3.1 we introduce low-rank approximation, in which the parameter space is restricted to a low-rank space induced by the prior; subsection 3.2 introduces a bi-fidelity approach in which multifidelity models are used for different ensemble members; and box constraints to enforce pointwise bounds on θ are introduced in subsection 3.3.

3.1. Low-rank approximation

When using ensemble methods for state estimation, the dimension of the ensemble space needed for successful state estimation may be much smaller than N_θ; a useful rule of thumb is that the ensemble space needs to be rich enough to learn about the unstable directions in the system. When using ensemble methods for inversion the situation is not so readily understood. The EKI algorithm presented here is limited to finding solutions in the linear span of the initial ensemble [55, 57] and we have highlighted a similar property for the EAKI and ETKI in lemma 1. While localization is often used to break this property [86] its use for this purpose is somewhat ad hoc. In this work we do not seek to break the subspace property. Indeed here we exploit low rank approximation within ensemble inversion techniques, a methodology which leads to solutions restricted to the linear span of a small number of dominant modes defined by the prior distribution.

Theorem 1 requires that the initial covariance matrix $C_0 \succ 0$ be strictly positive definite. To satisfy the assumption, the UKI requires $N_\theta + 2$ or $2N_\theta + 1$ forward problem evaluations and the storage of an $N_\theta \times N_\theta$ covariance matrix, and the EKI, EAKI and ETKI require $\mathcal{O}(N_\theta)$ forward problem evaluations and the storage of $\mathcal{O}(N_\theta)$ parameter estimates; some of the implications of these effects are numerically verified in section 4.4. Therefore, they are unaffordable for field inversion problems, where N_θ is large, typically from discretization of the $N_\theta = \infty$ limit. However, many physical phenomena or systems exhibit large-scale structure or finite-dimensional attractors, and in such situations the model error covariance matrices are generally low-rank; these low-rank spaces are spanned by, for example, the dominant Karhunen–Loève modes for random fields [87, 88] or the dominant spherical harmonics space on the sphere [63, 89]. We introduce a reparameterization strategy for this framework in order to leverage such low-rank structure when present, and thereby to reduce both computational and storage costs.

Given the prior distribution $\mathcal{N}(r_0, \Sigma_0)$, we assume Σ_0 is a low-rank matrix with the truncated singular value decomposition

$$\Sigma_0 \approx U D_0 U^{\mathrm{T}}.$$

Here $U = \{u_1, u_2, \ldots, u_{N_r}\}$ is the N_r-dominant singular vector matrix and D_0 is the singular value matrix. The discrepancy $\theta - r_0$ is assumed to be well-approximated in the linear space spanned by column vectors of U. Hence, the unknown parameters can be reparameterized as follows:

$$\theta = r_0 + \sum_{i=1}^{N_r} \tau_{(i)} u_i.$$

The aforementioned algorithm is then applied to solve for the vector $\tau = [\tau_{(1)}, \tau_{(2)}, \ldots, \tau_{(N_r)}]^{\mathrm{T}}$, which has prior mean 0 and prior covariance D_0. This reduces the computation and memory cost from $\mathcal{O}(N_\theta)$ and $\mathcal{O}(N_\theta^2)$ to $\mathcal{O}(N_r)$ and $\mathcal{O}(N_r N_\theta)$, where N_r is the rank of the covariance matrix.

More advanced approaches to extracting the low-rank space exist, including active subspace methods [90] and likelihood-informed subspace methods [91, 92]; however, they all require derivatives and so we do not pursue them here.

3.2. Bi-fidelity approach

For large-scale scientific or engineering problems, even with a small parameter number N_θ (or rank number N_r), the computational cost associated with these $\mathcal{O}(N_\theta)$ (or $\mathcal{O}(N_r)$) forward

model evaluations can be intractable; for example the number of parameters may be small, but the parameter-to-data map may require evaluation of a large, complex model. The bi-fidelity or multilevel strategy [93–96] is widely used to accelerate sampling-based methods; in particular it has been introduced in the context of ensemble methods in [97] and see [98] for a recent overview of developments in this direction.

We employ a particular bi-fidelity approach for the UKI algorithm. In this approach, low-fidelity models can be used to speed up forward model evaluations as follows. Consider equation (40a); evaluation of the mean $\mathcal{G}(\hat{\theta}^0_{n+1})$ can be performed using a high-fidelity model; meanwhile the other $J-1$ forward evaluations employed for covariance estimation, $\{\mathcal{G}(\hat{\theta}^j_{n+1})\}_{j=1}^{J-1}$, can use low-fidelity models.

3.3. Box constraints

Adding constraints to the parameters (for example, dissipation is non-negative) significantly improves the robustness of Kalman inversion [99–101]. In this paper, there are occasions where we impose element-wise box constraints of the form

$$0 \leqslant \theta \quad \text{or } \theta_{\min} \leqslant \theta \leqslant \theta_{\max}.$$

These are enforced by change of variables writing $\theta = \varphi(\tilde{\theta})$ where, for example, respectively,

$$\varphi(\tilde{\theta}) = \exp(\tilde{\theta}) \quad \text{or } \varphi(\tilde{\theta}) = \theta_{\min} + \frac{\theta_{\max} - \theta_{\min}}{1 + \exp(\tilde{\theta})}.$$

The inverse problem is then reformulated as

$$y = \mathcal{G}(\varphi(\tilde{\theta})) + \eta$$

and the proposed Kalman inversion methods are employed with $\mathcal{G} \mapsto \mathcal{G} \circ \varphi$.

4. Numerical experiments

In this section, we present numerical experiments demonstrating application of filtering methods to the novel mean-field dynamical system (equations (14), (15) and (18)) introduced in this paper[7]; the goal is to approximate the posterior distribution of unknown parameters or fields. The first subsection lists the five test problems, and the subsequent subsections consider them each in turn. In summary, our findings are as follows[8]:

- The proposed Kalman inversion methods based on (equations (14), (15) and (18)) are more efficient than transport/coupling methods based on (12) (i.e., iterative Kalman filter methods) on all the examples we consider. They remove the sensitivity to the initialization and, relatedly, they converge exponentially fast.
- The proposed Kalman inversion methods with deterministic treatment of stochastic terms, specifically UKI and EAKI, outperform other methods with stochastic treatments, such as EKI, EKS (with Euler–Maruyama discretization) and CBS. They do not suffer from the presence of noisy fluctuations and achieve convergence for both linear and nonlinear problems.

[7] We fix $\Delta\tau = 1/2$ based on the parameter study presented in appendix A and iterate $\mathcal{O}(10)$ iterations to demonstrate convergence. In practice, adaptive time stepping and increment-based stopping criteria can be applied.

[8] The footnote from subsection 1.4, appearing before the bulleted list of contributions, applies here too.

- The methodology is implementable for large-scale parameter identification problems, such as those arising in climate models.

4.1. Overview of test problems

The five test problems considered are:

(a) Linear-Gaussian two-parameter model problem: this problem serves as a proof-of-concept example, which demonstrates the convergence of the mean and the covariance as analyzed in subsection 2.3.

(b) Nonlinear two-parameter two-point boundary value problem: this example appears in [74] an important paper which demonstrates that the mean field limit of ensemble Kalman inversion methods may be far from the true posterior; it is also used as a test case in several other papers, such as [75, 102]. We show that by applying Kalman filtering techniques to the extended mean-field dynamical system (equations (14), (15) and (18)), we obtain methods which obtain accurate posterior approximation on this problem.

(c) Hilbert matrix problem: this high dimensional linear-Gaussian problem demonstrates the ability of the proposed Kalman inversion methodology to solve ill-conditioned inverse problems. In addition to testing the novel mean-field approach introduced in this paper, we also study the effect of the ensemble size on ensemble Kalman inversion, and in particular, the ensemble adjustment/transform Kalman inversions are examined in this context.

(d) Darcy flow inverse problem: this is an infinite dimensional field inversion problem (see [30] and the references therein); in addition to testing the novel mean-field approach introduced in this paper, we also demonstrate the low-rank approximation strategy in subsection 3.1.

(e) Idealized global climate model: this 3D Navier–Stokes problem, see [72] for background and references; in addition to testing the novel mean-field approach introduced in this paper, we also demonstrate the bi-fidelity approach introduced in subsection 3.2.

For the first and third problems, the Gaussian structure means that they are exactly solvable and this provides a benchmark against which we compare various methods. Markov chain Monte Carlo methods (MCMC), specifically the random walk Metropolis [24] and preconditioned Crank–Nicolson [26] methods, are used as the benchmark for the second and fourth problems respectively. Problem five is too large for the use of MCMC, and showcases the potential of the methodology studied here to solve problems otherwise beyond reach.

In the first two tests, we compare the proposed Kalman inversion methods (EKI, UKI, EAKI, ETKI applied to equations (14), (15) and (18) with other recently proposed Gaussian approximation algorithms, including the ensemble Kalman sampler (EKS) [75, 77],[9] and the consensus-based sampler (CBS) [52, 53, 103–105].[10] We also compare with variants of iterative Kalman filter methods, which seek to deform the prior into the posterior in one time unit (transport/coupling) using a finite number of intermediate steps (see appendix B) based on (12). They include iterative unscented Kalman filters (IUKF-1, IUKF-2), IEnKF, iterative ensemble adjustment Kalman filter (IEAKF), and iterative ensemble transform Kalman filter (IETKF) [55, 56, 106, 107] [108, algorithm 3]. Having shown the superiority of filtering based on our novel mean field dynamical system, we consider only this approach in the remaining examples. In the third test, we study the effect of the ensemble size on the proposed Kalman

[9] We follow the implementation in [75], which employs adaptive time-stepping.

[10] We follow the implementation in [53], setting $\alpha = 0.0$ and adaptively updating β with $\mu = 0.5$.

inversion methods, in particular for the EAKI/ETKI approaches to filtering, comparing with the UKI. In the fourth and fifth tests, we demonstrate the effectiveness of the proposed Kalman inversion methods for large-scale inverse problems and incorporate the low-rank and bi-fidelity approaches.

4.2. Linear two-parameter model problem

We consider the two-parameter linear inverse problem [72] of finding $\theta \in \mathbb{R}^2$ from y given by

$$y = G\theta + \eta. \tag{53}$$

Here the observation error noise is $\eta \sim \mathcal{N}(0, 0.1^2\mathbb{I})$. We explore the following two scenarios

- Over-determined system

$$y = \begin{bmatrix} 3 \\ 7 \\ 10 \end{bmatrix} \qquad G = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \rho_{\text{prior}} \sim \mathcal{N}(0, \mathbb{I}); \tag{54}$$

- Under-determined system

$$y = \begin{bmatrix} 3 \end{bmatrix} \qquad G = \begin{bmatrix} 1 & 2 \end{bmatrix} \quad \rho_{\text{prior}} \sim \mathcal{N}(0, \mathbb{I}). \tag{55}$$

We apply various Kalman inversions to our proposed novel mean-field dynamical system, including UKI-1 ($J = 4$), UKI-2 ($J = 5$), EKI, EAKI, and ETKI; we compare with pre-existing coupling/transport based iterative Kalman filters, including IUKF-1 ($J = 4$), IUKF-2 ($J = 5$), IEnKF, IEAKF and IETKF all with $J = 10$ ensemble members; and we compare with EKS, and CBS, again all with $J = 10$. All algorithms are initialized at the prior distribution; note however that the methods we introduce in this paper, and EKS and CBS, do not require this and indeed are robust to the use of different initializations. The iterative Kalman filters are discretized with $\Delta\tau = \frac{1}{30}$, and further correction (see appendix B.2) is applied on the initial ensemble members for the exactness of the initialization, except for the IEAKF*. Since the posterior distribution is Gaussian, we can compute the reference distribution analytically. The convergence of the posterior mean and posterior covariance are reported in figures 1 and 2. Because we use the same number of steps for all algorithms, and commensurate numbers of particles, the evaluation cost of all the methods studied are comparable; the size of the error discriminates between them.

For both scenarios, UKI-1, UKI-2, EAKI, and ETKI converge exponentially fast. IUKF-1, IUKF-2, IEAKF, and IETKF reach exact posterior mean and covariance matrix at $T = 1$. However, IEAKF* does not converge due to the error introduced in the initialization. EKI and IEnKF do not converge, and suffer from the presence of random noise introduced in the analysis step. EKS and CBS do not converge, and suffer from the presence of random noise and the finite ensemble size.

4.3. Nonlinear two-parameter model problem

Consider the one-dimensional elliptic boundary-value problem

$$-\frac{\mathrm{d}}{\mathrm{d}x}\left(\exp(\theta_{(1)})\frac{\mathrm{d}}{\mathrm{d}x}p(x)\right) = 1, \quad x \in [0, 1] \tag{56}$$

with boundary conditions $p(0) = 0$ and $p(1) = \theta_{(2)}$. The solution for this problem is given by

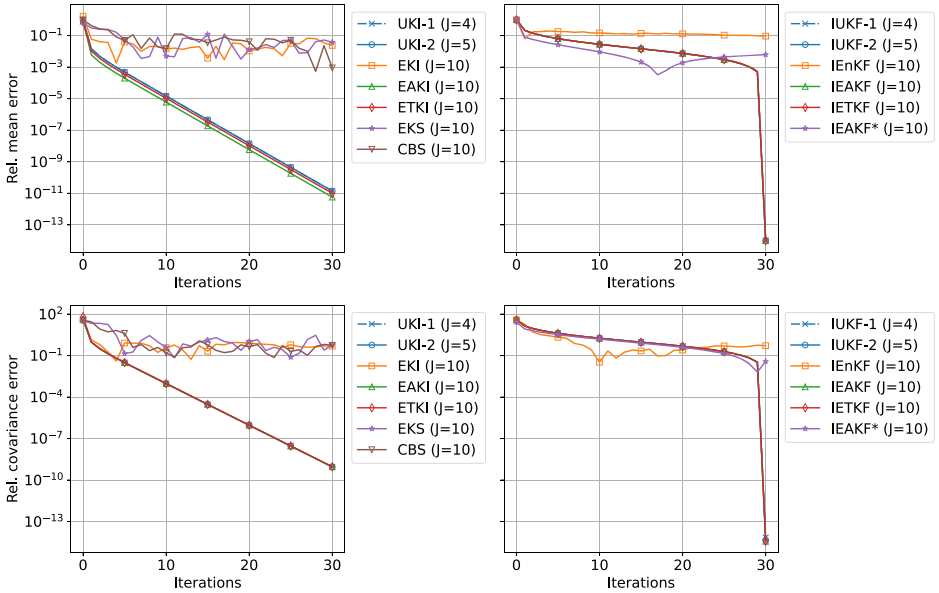

Figure 1. Linear two-parameter model problems: convergence of posterior mean (top) and posterior covariance (bottom) for the over-determined system.

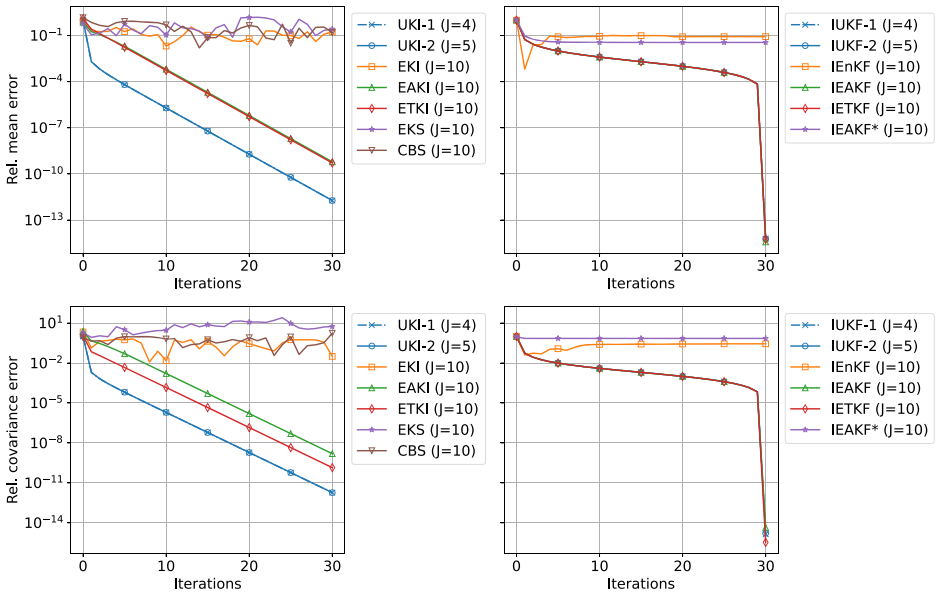

Figure 2. Linear two-parameter model problems: convergence of posterior mean (top) and posterior covariance (bottom) for the under-determined system (bottom).

$$p(x) = \theta_{(2)}x + \exp(-\theta_{(1)})\left(-\frac{x^2}{2} + \frac{x}{2}\right). \tag{57}$$

The Bayesian inverse problem is formulated as finding $\theta \in \mathbb{R}^2$ from y given by

$$y = \mathcal{G}(\theta) + \eta, \quad \theta = \begin{bmatrix} \theta_{(1)} \\ \theta_{(2)} \end{bmatrix} \quad \text{and} \quad \eta \sim \mathcal{N}(0, 0.1^2\mathbb{I}). \tag{58}$$

The observations comprise pointwise measurements of p and we consider well-determined and under-determined cases:

- Well-determined system

$$\mathcal{G}(\theta) = \begin{bmatrix} p(0.25, \theta) \\ p(0.75, \theta) \end{bmatrix} \qquad y = \begin{bmatrix} 27.5 \\ 79.7 \end{bmatrix} \quad \rho_{\text{prior}} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 100 \end{bmatrix}, \mathbb{I} \right). \tag{59}$$

- Under-determined system the observations

$$\mathcal{G}(\theta) = p(0.25, \theta) \qquad y = 27.5 \quad \rho_{\text{prior}} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 100 \end{bmatrix}, \mathbb{I} \right). \tag{60}$$

The reference posterior distribution is approximated by the random walk Metropolis algorithm with a step size 1.0 and 5×10^6 samples (with a 10^6 sample burn-in period). We compare the UKI-1 ($J = 4$), UKI-2 ($J = 5$), EKI, EAKI, and ETKI applied to (equations (14), (15) and (18)), iterative Kalman filters, including IUKF-1 ($J = 4$), IUKF-2 ($J = 5$), IEnKF, IEAKF, IETKF all with $J = 50$ ensemble members, applied to (12), and the EKS and CBS methods, also with $J = 50$. All algorithms are initialized at the prior distribution. The iterative Kalman filters are discretized with $\Delta\tau = \frac{1}{30}$, and further correction (see appendix B.2) is applied on the initial ensemble members for the exactness of the initialization.

Posterior distribution approximations obtained by different algorithms, all at the 30th iteration, are depicted in figures 3 and 4. Two common qualitative themes stand out from these figures: the iterative methods based on coupling/transport have difficulty covering the true posterior spread, especially in the under-determined case, when compared with the new methodologies based on our novel mean-field dynamical system; and application of ensemble transform methods in either coupling/transport or mean-field dynamical system suffers from a form of collapse. The first point may be seen quantitatively; the second does not show up so much quantitatively because collapse is in a direction in which there is less posterior spread. We now turn to quantitative comparisons. Again, because we use the same number of steps for all algorithms, and commensurate numbers of particles, the evaluation cost of all the methods studied are comparable; the size of the error discriminates between them.

The convergence of posterior mean and posterior covariance are reported in figures 5 and 6. For both scenarios, UKI-1, UKI-2, and EAKI converge exponentially fast at the beginning and then flatten out, since the posterior is not Gaussian [74]. The ETKI suffers from divergence for the under-determined scenario, and for this test, ETKI is less robust compared with UKI and EAKI. As in the linear two-parameter model problems (see subsection 4.2), EKI, EKS, CBS, and IEnKF suffer from random noise and finite ensemble sizes. Moreover, Kalman inversions, especially UKI and EAKI, outperform iterative Kalman filters, as measured by accuracy for commensurate cost, for these nonlinear tests.

4.4. Hilbert matrix problem

We define the Hilbert matrix $G \in \mathbb{R}^{N_\theta \times N_\theta}$ by its entries

$$G_{i,j} = \frac{1}{i + j - 1}, \tag{61}$$

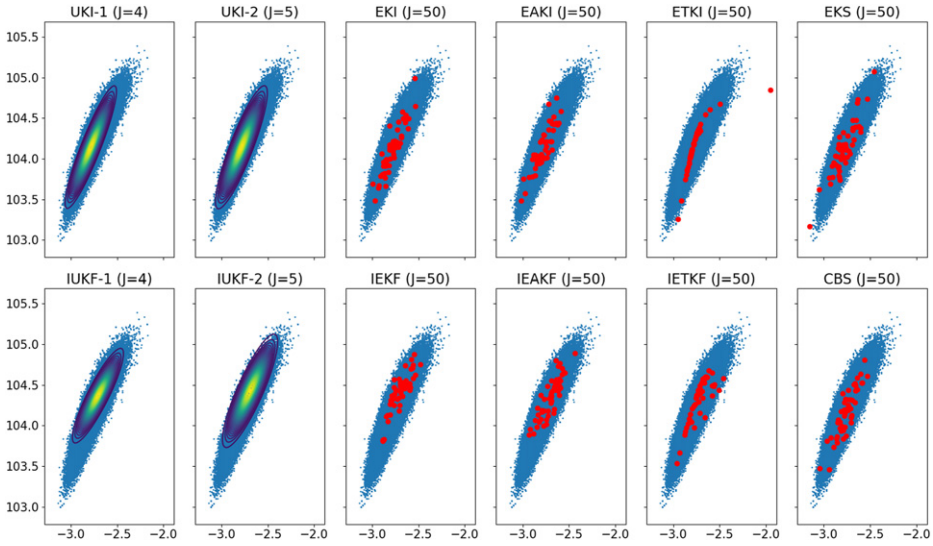

Figure 3. Nonlinear two-parameter model problem: posterior distribution approximated at the 30th iteration for the well-determined system. Blue dots represent the reference posterior distribution obtained by MCMC. x-axis is for $\theta_{(1)}$ and y-axis is for $\theta_{(2)}$.

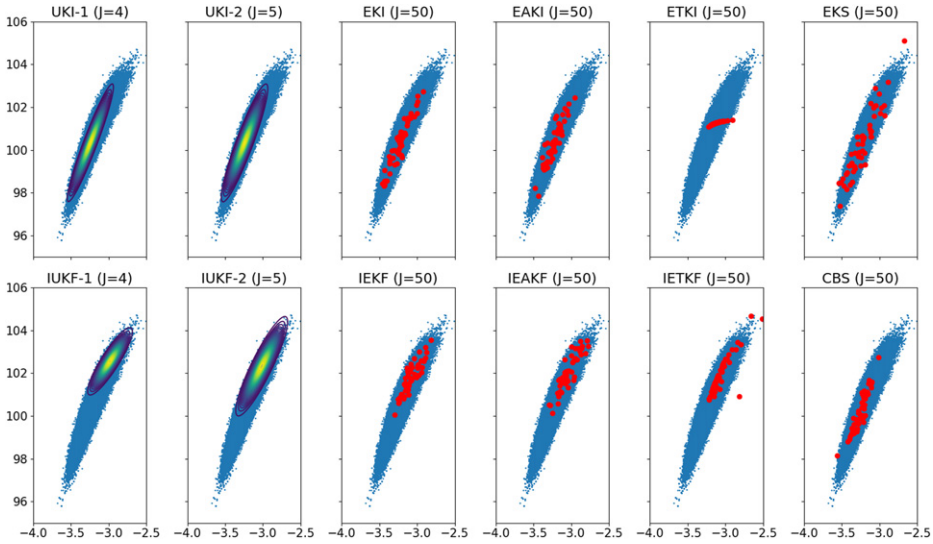

Figure 4. Nonlinear two-parameter model problem: posterior distribution approximated at the 30th iteration for the under-determined system. Blue dots represent the reference posterior distribution obtained by MCMC. x-axis is for $\theta_{(1)}$ and y-axis is for $\theta_{(2)}$.

with $N_\theta = 100$. We consider the inverse problem

$$y = G\mathbb{1} \qquad \eta \sim \mathcal{N}(0, 0.1^2\mathbb{I}) \quad \rho_{\text{prior}} \sim \mathcal{N}(0, \mathbb{I}). \tag{62}$$

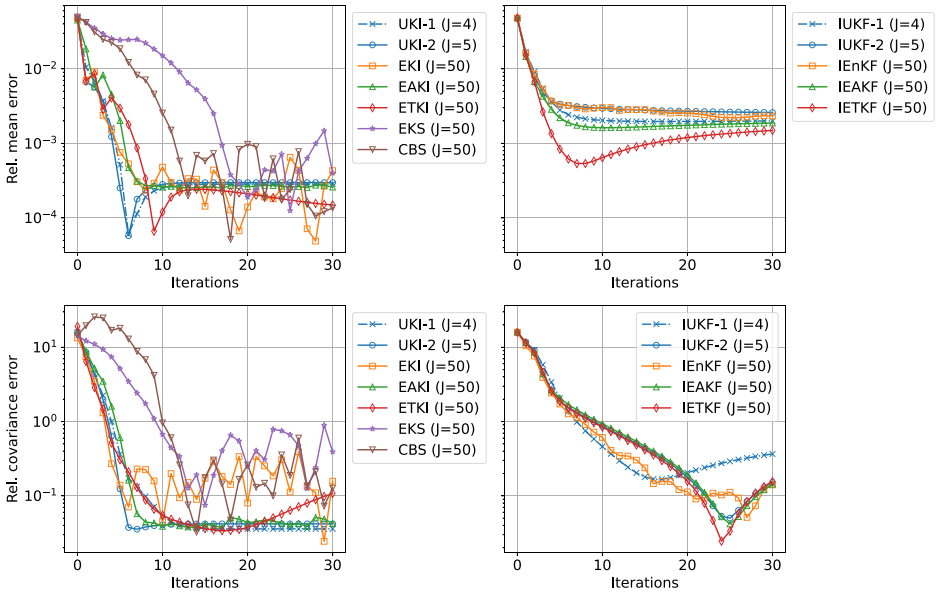

Figure 5. Nonlinear two-parameter model problem: convergence of posterior mean (top) and posterior covariance (bottom) for the well-determined system.

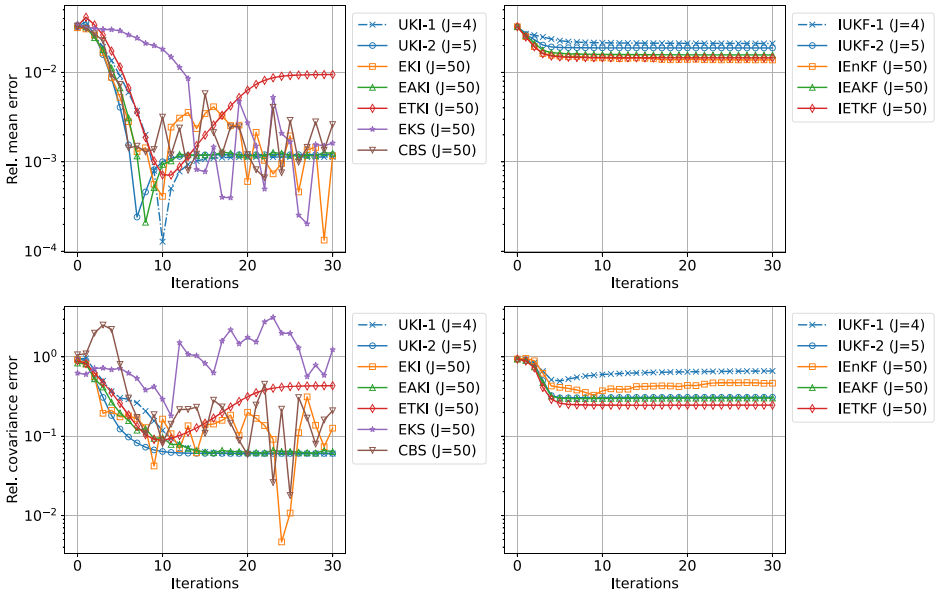

Figure 6. Nonlinear two-parameter model problem: convergence of posterior mean (top) and posterior covariance (bottom) for the under-determined system.

We no longer study iterated Kalman methods arising from coupling/transport as the preceding examples show that they are inefficient. Furthermore EKI, EKS, and CBS do not converge and suffer from random noise and/or finite ensemble sizes; these results are not shown. Instead,

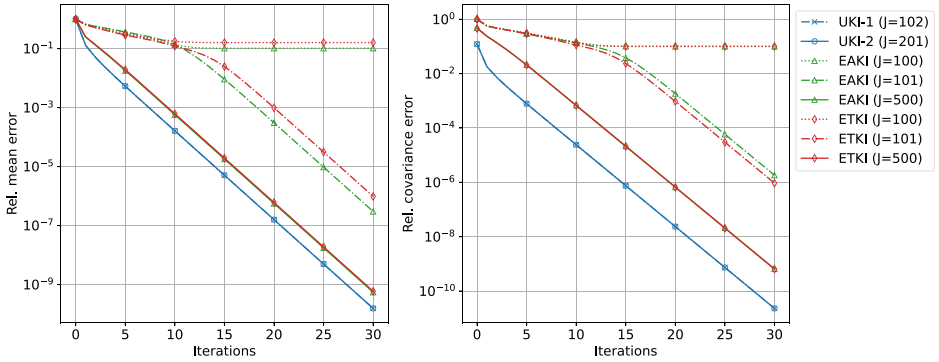

Figure 7. Hilbert matrix problem: convergence of posterior mean (left) and posterior covariance (right).

we focus on the effect of the ensemble size on EAKI and ETKI, comparing with UKI. To be concrete, we apply EAKI and ETKI with $J = N_\theta$, $N_\theta + 1$ and 500, and UKI-1 and UKI-2. Again, we initialize all algorithms at the prior distribution.

We compute the reference distribution analytically. The convergence of posterior mean and posterior covariance are reported in figure 7. UKI-1, UKI-2, and EAKI and ETKI with more than N_θ ensemble particles, converge exponentially fast. The relatively poor performance of EAKI and ETKI with a smaller number of ensemble particles is related to theorem 1, since EAKI and ETKI require at least $N_\theta + 1$ particles to ensure that the initial covariance matrix C_0 is strictly positive definite.

4.5. Darcy flow problem

The two-dimensional Darcy flow equation describes the pressure field $p(x)$ in a porous medium defined by a parameterized, positive permeability field $a(x, \theta)$:

$$\begin{aligned} -\nabla \cdot (a(x,\theta)\nabla p(x)) &= f(x), \quad x \in D, \\ p(x) &= 0, \qquad x \in \partial D. \end{aligned} \tag{63}$$

Here the computational domain is $D = [0, 1]^2$, Dirichlet boundary conditions are applied on ∂D, and f defines the source of the fluid:

$$f(x_1, x_2) = \begin{cases} 1000 & 0 \leqslant x_2 \leqslant \dfrac{4}{6} \\ 2000 & \dfrac{4}{6} < x_2 \leqslant \dfrac{5}{6} \\ 3000 & \dfrac{5}{6} < x_2 \leqslant 1 \end{cases}. \tag{64}$$

The inverse problem of interest is to determine parameter θ of the field $a(\cdot;\ \theta)$ from observation y_{ref}, which consists of pointwise measurements of the pressure value $p(\cdot)$ at 49 equidistant points in the domain (see figure 8), corrupted with observation error $\eta \sim \mathcal{N}(0, \mathbb{I})$. We now describe how a depends on θ, and specify a standard Gaussian prior on θ.

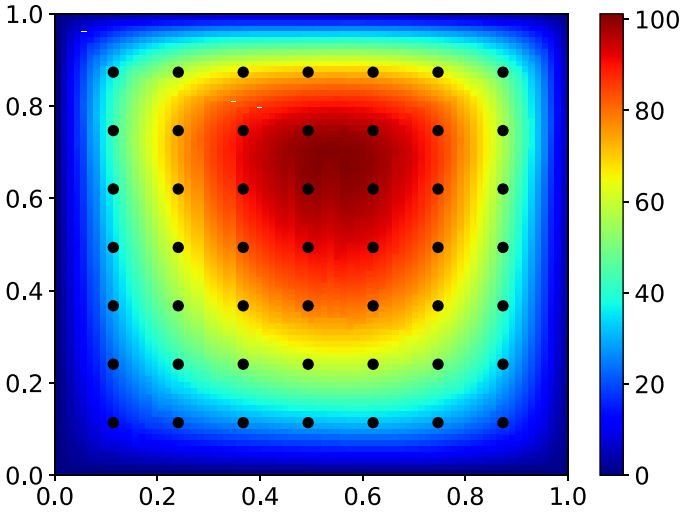

Figure 8. The reference pressure field with 49 equidistant pointwise measurements of the Darcy flow problem.

We write

$$\log\, a(x,\theta) = \sum_{l\in K} \theta_{(l)} \sqrt{\lambda_l}\psi_l(x), \tag{65}$$

where $K = \mathbb{Z}^+ \times \mathbb{Z}^+ \backslash \{0,0\}$, and

$$\psi_l(x) = \begin{cases} \sqrt{2}\,\cos(\pi l_1 x_1) & l_2 = 0 \\ \sqrt{2}\,\cos(\pi l_2 x_2) & l_1 = 0 \\ 2\,\cos(\pi l_1 x_1)\cos(\pi l_2 x_2) & \text{otherwise} \end{cases}, \quad \lambda_l = (\pi^2|l|^2 + \tau^2)^{-d} \tag{66}$$

and $\theta_{(l)} \sim \mathcal{N}(0,1)$ i.i.d. The expansion equation (65) can be rewritten as a sum over $\mathbb{Z}^+$ rather than a lattice:

$$\log\, a(x,\theta) = \sum_{k\in\mathbb{Z}^+} \theta_{(k)} \sqrt{\lambda_k}\psi_k(x), \tag{67}$$

where the eigenvalues λ_k are in descending order. We note that these considerations amount to assuming that $\log a(x,\theta)$ is a mean zero Gaussian random field with covariance

$$\mathsf{C} = (-\Delta + \tau^2)^{-d}, \tag{68}$$

with $-\Delta$ the Laplacian on D subject to homogeneous Neumann boundary conditions on the space of spatial-mean zero functions; hyperparameter τ denotes the inverse length scale of the random field and hyperparameter d determines its Sobolev and Hölder regularity, which is $d-1$ in our two dimensional setting [30].

In this work, we take $\tau = 3$ and $d = 2$. In practice, we truncate the sum (67) to N_θ terms, based on the largest N_θ eigenvalues, and hence $\theta \in \mathbb{R}^{N_\theta}$. The forward problem is solved by a finite difference method on a 80×80 grid. To create the data y_{ref} referred to above, we generate a truth random field $\log a(x,\theta_{\text{ref}})$ with $N_\theta = 128$ and $\theta_{\text{ref}} \sim \mathcal{N}(0, \mathbb{I}^{128})$.

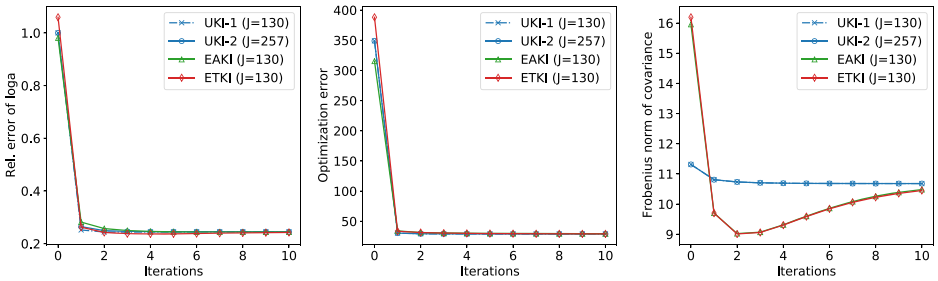

Figure 9. Darcy flow problem: the relative error of the permeability field, the optimization error $\frac{1}{2}\|\Sigma_\eta^{-\frac{1}{2}}(y_{\text{obs}} - \widehat{y}_n)\|^2 + \frac{1}{2}\|\Sigma_0^{-\frac{1}{2}}(m_n - r_0)\|^2$ and the Frobenius norm $\|C_n\|$ (from left to right).

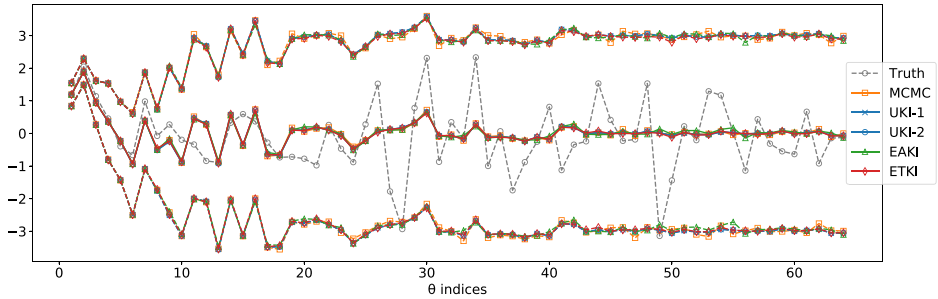

Figure 10. The estimated KL expansion parameters $\theta_{(i)}$ and the associated $3-\sigma$ confidence intervals obtained by UKI-1 ($J = 130$), UKI-2 ($J = 257$), EAKI ($J = 130$), ETKI ($J = 130$) and MCMC for the Darcy flow problem.

The benchmark posterior distribution is approximated by the preconditioned Crank–Nicolson algorithm with 2×10^6 samples (with a 5×10^5 sample burn-in period) with the step size 0.04. Since the preceding examples have shown the benefits of using UKI, EAKI and ETKI over all other methods considered, we compare only these approaches with the benchmark. Specifically, we apply UKI-1, UKI-2, and EAKI and ETKI with $J = N_\theta + 2$, again initialized at the prior distribution.

The convergence of the relative L_2 error of the mean of the $\log a$ field, the optimization errors, and the Frobenius norm of the estimated posterior covariance, as the iteration progresses, are depicted in figure 9. This clearly shows that all four Kalman inversion techniques converge within 10 iterations.

Figure 10 shows the properties of the converged posterior distribution, after the 10th iteration, comparing them with MCMC and with the truth (referred to as 'Truth'). The information is broken down according to recovery of the $\{\theta_{(i)}\}$, visualizing only the first 64 modes, since the statistical estimates of other modes obtained by MCMC and by our Kalman inversion methodologies are close to the prior $\mathcal{N}(0, 1)$—the data does not inform them. We first note that the truth values lie in the confidence intervals determined by MCMC, with high probabilities. Secondly, we note that all four Kalman methods reproduce the posterior mean and confidence intervals computed by MCMC accurately. The estimated log-permeability fields $\log a$ and the truth are depicted in figure 11. The mean estimations obtained by the MCMC and these Kalman inversions match well, and they both capture the main feature of the truth log-permeability field.

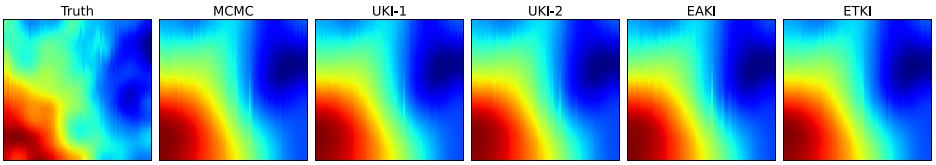

Figure 11. The truth log permeability field $\log a$, and log permeability fields obtained by MCMC, UKI-1 ($J = 130$), UKI-2 ($J = 257$), EAKI ($J = 130$), ETKI ($J = 130$) (left to right).

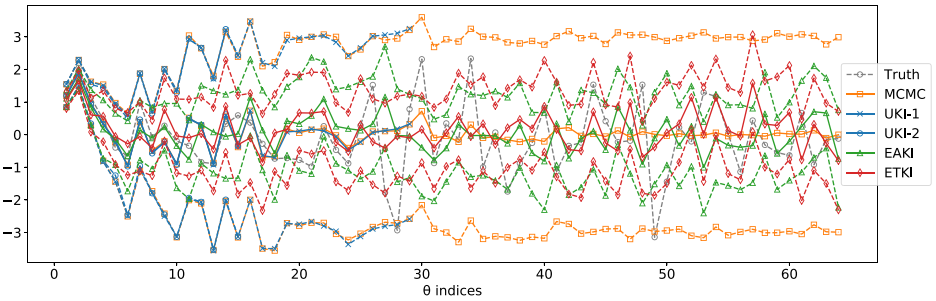

Figure 12. The estimated KL expansion parameters $\theta_{(i)}$ and the associated $3-\sigma$ confidence intervals obtained by UKI-1, UKI-2, EAKI, ETKI with $J = 31$ ensemble members and MCMC for the Darcy flow problem.

Remark 8. In practice, for many inversion problems for fields, the realistic number of ensemble members is much smaller than the dimension of the state space. To probe this setting, we repeat the test by using UKI-1, UKI-2, ETKI, and EAKI with $J = 31$ ensemble members; for UKI-1 and UKI-2, we invert for the first 29 and 15 coefficients of $\{\theta_{(k)}\}$, respectively, and for ETKI and EAKI we invert for all 128 coefficients. The estimated KL expansion parameters $\{\theta_{(i)}\}$ for the log-permeability field and the associated $3-\sigma$ confidence intervals obtained by MCMC, and different Kalman inversions at the 10th iteration, are depicted in figure 12. The mean and standard deviation of the coefficients associated with these dominant modes obtained by both UKIs match well with those obtained by MCMC. The results indicate that the 'truncate then invert' strategy used by UKIs outperforms the 'direct inversion' strategy, used here by EAKI and ETKI, when only small ensemble numbers are feasible.

4.6. Idealized global climate model

Finally, we consider using low-fidelity model techniques to speed up an idealized global climate model inverse problem. The model is based on the 3D Navier–Stokes equations, making the hydrostatic and shallow-atmosphere approximations common in atmospheric modeling. Specifically, we test on the notable Held–Suarez test case [109], in which a detailed radiative transfer model is replaced by Newtonian relaxation of temperatures toward a prescribed 'radiative equilibrium' $T_{\text{eq}}(\phi, p)$ that varies with latitude ϕ and pressure p. Specifically, the

thermodynamic equation for temperature T

$$\frac{DT}{\partial t} - \frac{RT\omega}{C_p p} = Q \tag{69}$$

(including advective and pressure work terms) contains a diabatic heat source

$$Q = -k_T(\phi, p, p_s)\big(T - T_{\text{eq}}(\phi, p)\big), \tag{70}$$

with relaxation coefficient (inverse relaxation time)

$$k_T = k_a + (k_s - k_a)\max\left(0, \frac{\sigma - \sigma_b}{1 - \sigma_b}\right)\cos^4\phi. \tag{71}$$

Here, $\sigma = p/p_s$, pressure p normalized by surface pressure p_s, is the vertical coordinate of the model, and

$$T_{\text{eq}} = \max\left\{200\ \text{K}, \left[315\ \text{K} - \Delta T_y \sin^2\phi - \Delta\theta_z \log\left(\frac{p}{p_0}\right)\cos^2\phi\right]\left(\frac{p}{p_0}\right)^{\kappa}\right\} \tag{72}$$

is the equilibrium temperature profile ($p_0 = 10^5$ Pa is a reference surface pressure and $\kappa = 2/7$ is the adiabatic exponent).

The inverse problem of interest here is to determine the parameters $(k_a, k_s, \Delta T_y, \Delta\theta_z)$ from statistical averages of the temperature field T. We impose the following constraints:

$$0\ \text{day}^{-1} < k_a < 1\ \text{day}^{-1}, \quad 0\ \text{day}^{-1} < k_s < 1\ \text{day}^{-1},$$
$$0\ \text{K} < \Delta T_y < 100\ \text{K}, \quad 0\ \text{K} < \Delta\theta_z < 50\ \text{K}.$$

The inverse problem is formed as follows [72],

$$y = \mathcal{G}(\theta) + \eta \quad \text{with } \mathcal{G}(\theta) = \overline{T}(\phi, \sigma) \tag{73}$$

with the parameter transformation

$$\theta : (k_a, k_s, \Delta T_y, \Delta\theta_z) = \left(\frac{1}{1 + \exp(\theta^{(1)})}, \frac{1}{1 + \exp(\theta^{(2)})}, \times \frac{100}{1 + \exp(\theta^{(3)})}, \frac{50}{1 + \exp(\theta^{(4)})}\right) \tag{74}$$

enforcing the constraints. The observation mapping $\mathcal{G}$ is defined by mapping from the unknown θ to the 200 days zonal mean of the temperature ($\overline{T}$) as a function of latitude (ϕ) and height (σ), after an initial spin-up of 200 days.

Default parameters used to generate the data in our simulation study are

$$k_a = (40\ \text{day})^{-1}, \qquad k_s = (4\ \text{day})^{-1}, \qquad \Delta T_y = 60\ \text{K}, \qquad \Delta\theta_z = 10\ \text{K}.$$

For the numerical simulations, we use the spectral transform method in the horizontal, with T42 spectral resolution (triangular truncation at wavenumber 42, with 64×128 points on the latitude-longitude transform grid); we use 20 vertical levels equally spaced in σ. With the default parameters, the model produces an Earth-like zonal-mean circulation, albeit without moisture or precipitation. The truth observation is the 1000 days zonal mean of the temperature

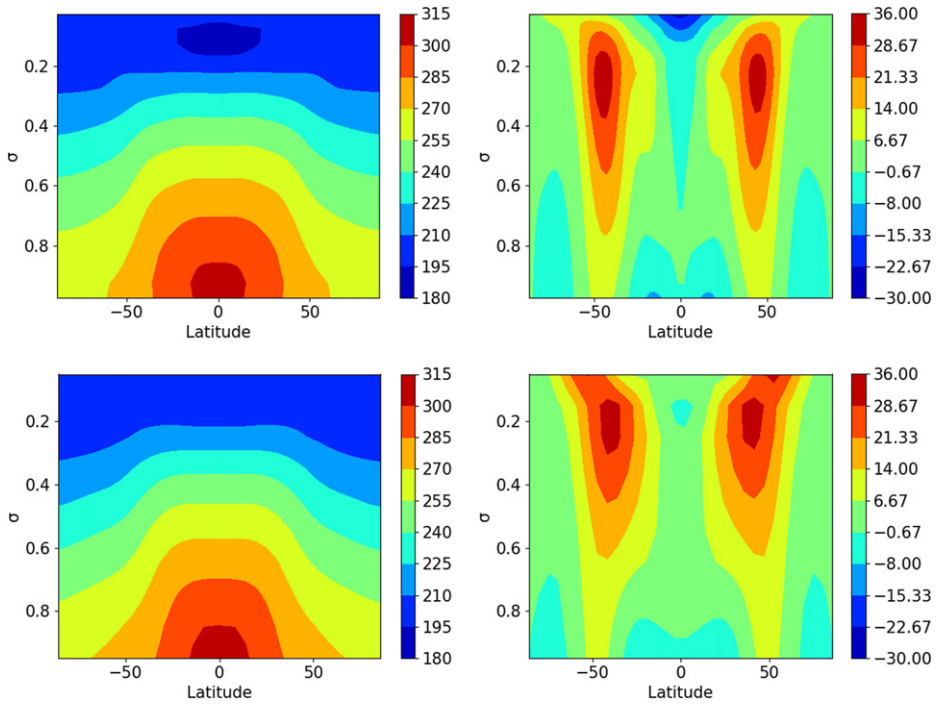

Figure 13. Zonal mean temperature (left) and zonal wind velocity (right) obtained with the T42 grid (top) and the T21 grid (bottom).

(see figure 13-top-left), after an initial spin-up, also of 200 days, to eliminate the influence of the initial condition. Because the truth observations come from an average 5 times as long as the observation window used for parameter learning, the chaotic internal variability of the model introduces noise in the observations.

To perform the inversion, we set the prior $\rho_{\text{prior}} \sim \mathcal{N}(0, 10^2\mathbb{I})$. Within the algorithm, we assume that the observation error satisfies $\eta \sim \mathcal{N}(0\ \text{K}, 3^2\mathbb{I}\ \text{K}^2)$. All these Kalman inversions are initialized with $\theta_0 \sim \mathcal{N}(0, 0.1^2\mathbb{I})$, since initializing at the prior leads to unstable simulations at the first iteration. The bi-fidelity approach discussed in subsection 3.2 is applied to speed up both UKI-1 and UKI-2. These $J-1$ forward model evaluations are computed on a T21 grid (triangular truncation at wavenumber 21, with 32×64 points on the latitude-longitude transform grid) with 10 vertical levels equally spaced in σ (twice coarser in all three directions). They are abbreviated as UKI-1-BF and UKI-2-BF. The computational cost of the high-fidelity (T42) and low-fidelity (T21) models are about four-CPU hour and 0.5-CPU hour, and therefore the bi-fidelity approach effectively reduces CPU costs. The 1000 days zonal mean of the temperature and velocity predicted by the low-resolution model with the truth parameters are shown in figure 13-bottom. It is worth mentioning there are significant discrepancies comparing with results computed on the T42 grid (figure 13-top). Whether these would be tolerable will depend on the use to which the posterior inference is put.

The estimated parameters and associated $3-\sigma$ confidence intervals for each component at each iteration are depicted in figure 14. Since the prior covariance is large and the problem is over-determined it is natural to expect that the posterior mean should be close to the true

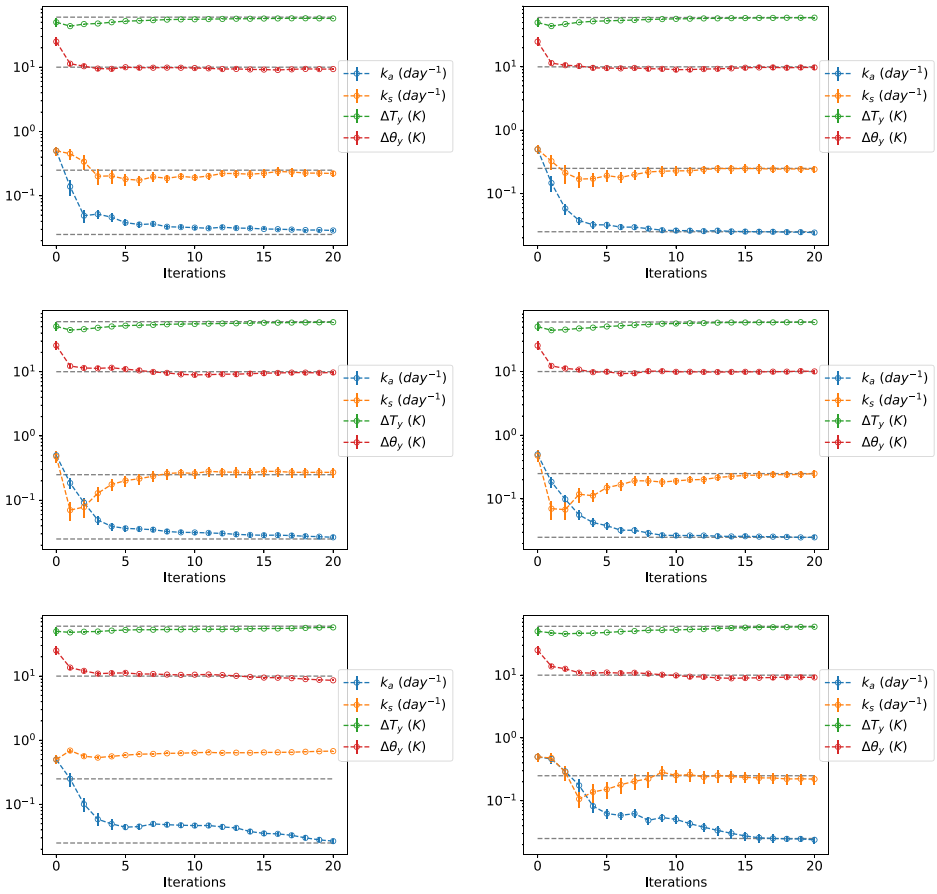

Figure 14. Convergence of the idealized global climate model inverse problem. True parameter values are represented by dashed grey lines. Top: UKI-1 ($J = 6$) and UKI-2 ($J = 9$); middle: EAKI ($J = 9$) and ETKI ($J = 9$); bottom: UKI-1-BF ($J = 6$) and UKI-2-BF ($J = 9$).

parameters. Indeed all the different Kalman inversions, except UKI-1-BF ($J = 6$), do indeed converge to the true parameters.

5. Conclusion

Kalman-based inversion has been widely used to construct derivative-free optimization and sampling methods for nonlinear inverse problems. In this paper, we developed new Kalman-based inversion methods, for Bayesian inference and uncertainty quantification, which build on the work in both optimization and sampling. We propose a new method for Bayesian inference based on filtering a novel mean-field dynamical system subject to partial noisy observations, and which depends on the law of its own filtering distribution, together with application of the Kalman methodology. Theoretical guarantees are presented: for linear inverse problems, the mean and covariance obtained by the method converge exponentially fast to the posterior mean and covariance. For nonlinear inverse problems, numerical studies indicate the method delivers an excellent approximation of the posterior distribution for problems which are not too

far from Gaussian. The methods are shown to be superior to existing coupling/transport methods, under the umbrella of iterative Kalman methods; and deterministic rather than stochastic implementations of Kalman methodology are found to be favorable. We further propose several simple strategies, including low-rank approximation and a bi-fidelity approach, to reduce the computational and memory cost of the proposed methodology.

There are numerous directions for future study in this area. On the theoretical side, it would be of value to obtain a theoretical understanding of the fixed point iteration (11) underpinning the present method, and Gaussian and/or particle approximations of it; and it is important to obtain theoretical guarantees concerning the accuracy of the present method when applied to nonlinear inverse problems, which are close to Gaussian. On the methodological side, the presented method is based on a Gaussian approximation of the fixed point iteration (11). However, many posterior distributions of interest are far from Gaussian; this can happen, for example, when the inverse problem (1) has multiple solutions and the prior is uninformative; or when the inverse problem has a unique solution, but the prior is uninformative and the noise error is large. To be concrete, consider the nonlinear two-parameter model problem discussed in section 4.3:

- At the same noise level, when the prior becomes uninformative $\rho_{\text{prior}} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 100 \end{bmatrix}, 10^2\mathbb{I}\right)$, the posterior is far from Gaussian in the under-determined case;
- For the same prior, when the noise level becomes moderately[11] large, $\Sigma_\eta = \mathbb{I}$, the posterior is again far from Gaussian for both well-determined and under-determined systems.

The present method then still converges, but finds an approximately Gaussian distribution which is far from the true posterior. To overcome this it will be of interest to replace the Gaussian ansatz with other distributions, such as Gaussian mixtures, to create efficient and derivative-free sampling techniques for non-Gaussian posterior distributions.

Acknowledgments

AMS and DZH are supported by the generosity of Eric and Wendy Schmidt by recommendation of the Schmidt Futures program; AMS is also supported by the Office of Naval Research (ONR) through Grant N00014-17-1-2079. JH is supported by the Simons Foundation as a Junior Fellow at New York University. SR is supported by Deutsche Forschungsgemeinschaft (DFG)—Project-ID 318763901—SFB1294.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://github.com/Zhengyu-Huang/InverseProblems.jl.

Appendix A. Continuous time limit

To derive a continuous-time limit of the novel mean-field dynamical system (equations (14), (15) and (18)) we define $\tau_n = n\Delta\tau$ and define $\{z_n\}$ by $x_n = \Delta\tau^{-1}(z_{n+1} - z_n)$. Then $\theta_n \approx \theta(\tau_n)$ and $z_n \approx z(\tau_n)$. Let $W \in \mathbb{R}^{N_\theta}$ and $B \in \mathbb{R}^{N_y+N_\theta}$ be standard unit Brownian motions.

[11] When the noise level is very large, the posterior becomes the Gaussian prior.

Letting $\Delta\tau \to 0$ in (equations (14), (15) and (18)), we then obtain

$$\begin{aligned} \dot{\theta} &= C^{\frac{1}{2}}\dot{W}, \\ \dot{z} &= \mathcal{F}(\theta) + \begin{bmatrix} \Sigma_\eta & 0 \\ 0 & \Sigma_0 \end{bmatrix}^{\frac{1}{2}} \dot{B}. \end{aligned} \tag{A.1}$$

We are interested in the filtering problem of finding the distribution of $\theta(\tau)$ given $\{z(s)\}_{s=0}^{\tau}$ and then evaluating this distribution in the setting where $\dot{z}(s) \equiv x$ defined in (17).

Under a similar limiting process, the Gaussian approximation algorithm defined by (22) becomes

$$\begin{aligned} \dot{m} &= \widehat{C}^{\theta x} \begin{bmatrix} \Sigma_\eta & 0 \\ 0 & \Sigma_0 \end{bmatrix}^{-1} (x - \widehat{x}), \\ \dot{C} &= C - \widehat{C}^{\theta x} \begin{bmatrix} \Sigma_\eta & 0 \\ 0 & \Sigma_0 \end{bmatrix}^{-1} \widehat{C}^{\theta x \mathrm{T}}, \end{aligned} \tag{A.2}$$

where $\widehat{x} = \mathbb{E}(\mathcal{F}(\theta))$, $\widehat{C}^{\theta x} = \mathbb{E}((\theta - m) \otimes (\mathcal{F}(\theta) - \mathbb{E}\mathcal{F}(\theta)))$, and expectation $\mathbb{E}$ is with respect to the distribution $\theta \sim \mathcal{N}(m, C)$. Define $\hat{\Phi}_R(m, C) = (x - \widehat{x})^{\mathrm{T}} \begin{bmatrix} \Sigma_\eta & 0 \\ 0 & \Sigma_0 \end{bmatrix}^{-1} (x - \widehat{x})$, and note that $\frac{\partial \widehat{x}}{\partial m} = \widehat{C}^{\theta x \mathrm{T}} C^{-1}$ [72, lemma 2]. It follows that equation (A.2) can be rewritten as

$$\begin{aligned} \dot{m} &= -\frac{1}{2} C \frac{\partial \hat{\Phi}_R}{\partial m}, \\ \dot{C}^{-1} &= -C^{-1} + \left(\frac{\partial \widehat{x}}{\partial m}\right)^{\mathrm{T}} \begin{bmatrix} \Sigma_\eta & 0 \\ 0 & \Sigma_0 \end{bmatrix}^{-1} \left(\frac{\partial \widehat{x}}{\partial m}\right). \end{aligned} \tag{A.3}$$

The stationary points of equation (A.3) satisfy

$$\frac{\partial \hat{\Phi}_R}{\partial m} = 0 \quad \text{and} \quad C^{-1} = \left(\frac{\partial \widehat{x}}{\partial m}\right)^{\mathrm{T}} \begin{bmatrix} \Sigma_\eta & 0 \\ 0 & \Sigma_0 \end{bmatrix}^{-1} \left(\frac{\partial \widehat{x}}{\partial m}\right) \succeq \Sigma_0^{-1}. \tag{A.4}$$

Although the $\Delta\tau$ plays no role in the continuous time limit, the original $\Delta\tau$ parameter does have a marked effect on discrete-time algorithms used in practice. To highlight this, we study the empirical effect of $\Delta\tau$, in the context of the nonlinear two-parameter model problem discussed in section 4.3. The UKI-2 is applied with $\Delta\tau = 1/5$, $\Delta\tau = 1/3$, $\Delta\tau = 1/2$, $\Delta\tau = 2/3$, and $\Delta\tau = 3/4$. Empirically we observe that all solutions considered converge to approximately the same equilibrium point; but the convergence properties depend on $\Delta\tau$. The convergence of the posterior mean and covariance are reported in figure A1. The best convergence rate is achieved with $\Delta\tau$ around $1/2$ for this test. We note that the case $\Delta\tau = 1/2$ also appears in [54] under the notion of ensemble transform Langevin dynamics.

Appendix B. Iterative Kalman filter

B.1. Gaussian approximation

Iterative Kalman filters for inverse problems are obtained by applying filtering methods, over $N = \Delta\tau^{-1}$ steps, to the dynamical system equation (12) with scaled observation error $\eta_{n+1} \sim$

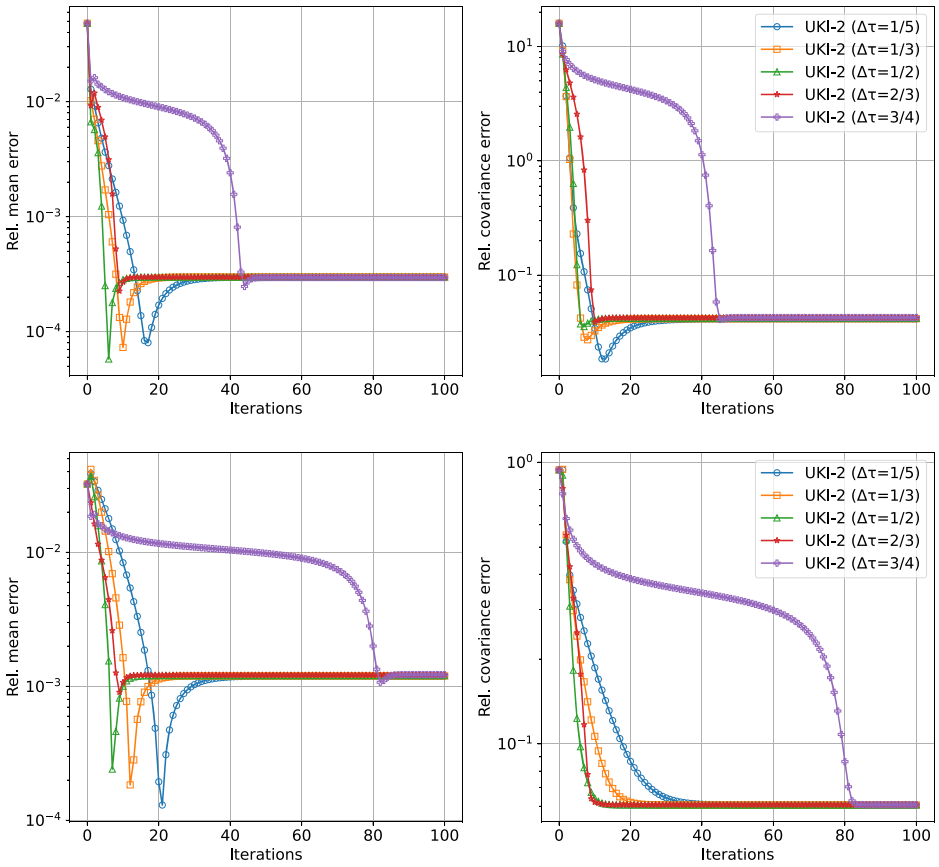

Figure A1. Nonlinear two-parameter model problem: convergence of posterior mean (left) and posterior covariance (right) for the well-determined system (top) and the under-determined system (bottom) with different $\Delta\tau$.

$\mathcal{N}(0, \Delta\tau^{-1}\Sigma_\eta)$. Following the discussion in subsection 2.2, the conceptual algorithm can be written as

$$m_{n+1} = m_n + \widehat{C}^{\theta y}_{n+1}(\widehat{C}^{yy}_{n+1})^{-1}(y - \widehat{y}_{n+1}), \tag{B.1a}$$

$$C_{n+1} = C_n - \widehat{C}^{\theta y}_{n+1}(\widehat{C}^{yy}_{n+1})^{-1}\widehat{C}^{\theta y\,\mathrm{T}}_{n+1}, \tag{B.1b}$$

where

$$\begin{aligned}
\widehat{y}_{n+1} &= \mathbb{E}[y_{n+1}|Y_n] = \mathbb{E}[\mathcal{G}(\theta_{n+1})|Y_n], \\
\widehat{C}^{\theta y}_{n+1} &= C[\theta_{n+1}, y_{n+1}|Y_n] = C[\theta_{n+1}, \mathcal{G}(\theta_{n+1})|Y_n], \\
\widehat{C}^{yy}_{n+1} &= \mathrm{Cov}[y_{n+1}|Y_n] = \mathrm{Cov}[\mathcal{G}(\theta_{n+1})|Y_n] + \Delta\tau^{-1}\Sigma_\eta,
\end{aligned} \tag{B.2}$$

and $Y_n := \{y_1^\dagger, y_2^\dagger, \ldots, y_n^\dagger\}$, the observation set at time n. Different Kalman filters (see subsection 2.4) can be applied, which lead to the IUKF, IEnKF, IEAKF and IETKF, all considered here. We have the following theorem about the convergence of the conceptual algorithm:

Theorem 2. *Consider the linear case $\mathcal{G}(\theta) = G\theta$. Assume that the initial mean and covariance for* (B.1) *are the prior mean and covariance. Then the output of the iteration* (B.1) *at $N = \Delta\tau^{-1}$ equals the posterior mean and covariance.*

Related calculations in continuous time may be found in section 2.2 of [75].

B.2. Gaussian initialization

Theorem 2 can be extended to apply to the ensemble approximations of the mean field dynamics (B.1), provided that the initial ensemble represents the prior mean and covariance exactly. For iterative ensemble filters, we first sample $\{\theta_j\}_{j=1}^J \sim \mathcal{N}(m, C)$ $(J \geqslant N_\theta + 1)$, and then correct them as follows.

Define

$$\Theta' = [\theta_1 - m'; \theta_2 - m'; \dots \theta_J - m']^{\mathrm{T}} \quad m' = \frac{1}{J}\sum_{j=1}^{J} \theta_j, \tag{B.3}$$

and correct Θ' with a $N_\theta \times N_\theta$ matrix X to keep zero mean and match the covariance

$$\Theta := \Theta' X + 1 \otimes m^{\mathrm{T}}, \tag{B.4}$$

where

$$\begin{aligned} \mathrm{SVD}: \Theta' &= U_1 S_1 V_1^{\mathrm{T}}, \\ \mathrm{SVD}: (J-1)C &= U_2 S_2 U_2^{\mathrm{T}}, \\ X &= V_1 S_1^{-1} S_2^{\frac{1}{2}} U_2. \end{aligned} \tag{B.5}$$

ORCID iDs

Daniel Zhengyu Huang https://orcid.org/0000-0001-6072-9352

References

[1] Sen M K and Stoffa P L 2013 *Global Optimization Methods in Geophysical Inversion* (Cambridge: Cambridge University Press)
[2] Schneider T, Lan S, Stuart A and Teixeira J 2017 Earth system modeling 2.0: a blueprint for models that learn from observations and targeted high-resolution simulations *Geophys. Res. Lett.* **44** 12–396
[3] Huang D Z, Xu K, Farhat C and Darve E 2020 Learning constitutive relations from indirect observations using deep neural networks *J. Comput. Phys.* **416** 109491
[4] Xu K, Huang D Z and Darve E 2021 Learning constitutive relations using symmetric positive definite neural networks *J. Comput. Phys.* **428** 110072
[5] Avery P, Huang D Z, He W, Ehlers J, Derkevorkian A and Farhat C 2021 A computationally tractable framework for nonlinear dynamic multiscale modeling of membrane woven fabrics *Int. J. Numer. Methods Eng.* **122** 2598–625
[6] Russell B H 1988 *Introduction to Seismic Inversion Methods* (Tulsa, OK: Society of Exploration Geophysicists)

[7] Bunks C, Saleck F M, Zaleski S and Chavent G 1995 Multiscale seismic waveform inversion *Geophysics* **60** 1457–73

[8] Bui-Thanh T, Ghattas O, Martin J and Stadler G 2013 A computational framework for infinite-dimensional Bayesian inverse problems: I. The linearized case, with application to global seismic inversion *SIAM J. Sci. Comput.* **35** A2494–523

[9] Martin J, Wilcox L C, Burstedde C and Ghattas O 2012 A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion *SIAM J. Sci. Comput.* **34** A1460–87

[10] Wang L 2022 Joint inversion of receiver function and surface wave dispersion based on innocent unscented Kalman methodology (arXiv:2202.09544)

[11] Bertoglio C, Moireau P and Gerbeau J-F 2012 Sequential parameter estimation for fluid-structure problems: application to hemodynamics *Int. J. Numer. Methods Biomed. Eng.* **28** 434–55

[12] Moireau P and Chapelle D 2011 Reduced-order unscented Kalman filtering with application to parameter identification in large-dimensional systems *ESAIM Control, Optim. Calc. Var.* **17** 380–405

[13] Jasak H *et al* 2007 OpenFOAM: a C++ library for complex physics simulations *Int. Workshop Coupled Methods in Numerical Dynamics* vol 1000 (IUC Dubrovnik Croatia) pp 1–20

[14] Huang D Z, Avery P, Farhat C, Rabinovitch J, Derkevorkian A and Peterson L D 2020 Modeling, simulation and validation of supersonic parachute inflation dynamics during Mars landing *AIAA Scitech 2020 Forum* p 0313

[15] Cao S and Huang D Z 2022 Bayesian calibration for large-scale fluid structure interaction problems under embedded/immersed boundary framework *Int. J. Numer. Methods Eng.* **123** 1791–812

[16] Peskin C S 1977 Numerical analysis of blood flow in the heart *J. Comput. Phys.* **25** 220–52

[17] Huang D Z, De Santis D and Farhat C 2018 A family of position- and orientation-independent embedded boundary methods for viscous flow and fluid-structure interaction problems *J. Comput. Phys.* **365** 74–104

[18] Berger M J *et al* 1989 Local adaptive mesh refinement for shock hydrodynamics *J. Comput. Phys.* **82** 64–84

[19] Borker R, Huang D, Grimberg S, Farhat C, Avery P and Rabinovitch J 2019 Mesh adaptation framework for embedded boundary methods for computational fluid dynamics and fluid-structure interaction *Int. J. Numer. Methods Fluids* **90** 389–424

[20] Moës N, Dolbow J and Belytschko T 1999 A finite element method for crack growth without remeshing *Int. J. Numer. Methods Eng.* **46** 131–50

[21] Tan Z, Kaul C M, Pressel K G, Cohen Y, Schneider T and Teixeira J 2018 An extended Eddy-diffusivity mass-flux scheme for unified representation of subgrid-scale turbulence and convection *J. Adv. Model. Earth Syst.* **10** 770–800

[22] Lopez-Gomez I, Christopoulos C D, Ervik H L L, Dunbar O, Cohen Y and Schneider T 2022 Training physics-based machine-learning parameterizations with gradient-free ensemble Kalman methods *J. Adv. Model. Earth Syst.* **14** e2022MS003105

[23] Geyer C J 1992 Practical Markov Chain Monte Carlo *Stat. Sci.* **7** 473–83

[24] Gelman A, Gilks W R and Roberts G O 1997 Weak convergence and optimal scaling of random walk Metropolis algorithms *Ann. Appl. Probab.* **7** 110–20

[25] Goodman J and Weare J 2010 Ensemble samplers with affine invariance *Commun. Appl. Math. Comput. Sci.* **5** 65–80

[26] Cotter S L, Roberts G O, Stuart A M and White D 2013 MCMC methods for functions: modifying old algorithms to make them faster *Stat. Sci.* **28** 424–46

[27] Del Moral P, Doucet A and Jasra A 2006 Sequential Monte Carlo samplers *J. R. Stat. Soc.* B **68** 411–36

[28] Beskos A, Jasra A, Muzaffer E A and Stuart A M 2015 Sequential Monte Carlo methods for Bayesian elliptic inverse problems *Stat. Comput.* **25** 727–37

[29] Kaipio J and Somersalo E 2006 *Statistical and Computational Inverse Problems* vol 160 (Berlin: Springer)

[30] Dashti M and Stuart A M 2013 The Bayesian approach to inverse problems (arXiv:1302.6989)

[31] Dashti M, Law K J H, Stuart A M and Voss J 2013 MAP estimators and their consistency in Bayesian nonparametric inverse problems *Inverse Problems* **29** 095017

[32] Anderson J R and Peterson C 1987 A mean field theory learning algorithm for neural networks *Complex Syst.* **1** 995–1019

[33] Parisi G and Shankar R 1988 Statistical field theory *Phys. Today* **41** 110

[34] Opper M and Archambeau C 2009 The variational Gaussian approximation revisited *Neural Comput.* **21** 786–92
[35] Quiroz M, Nott D J and Kohn R 2018 Gaussian variational approximation for high-dimensional state space models (arXiv:1801.07873)
[36] Galy-Fajou T, Perrone V and Opper M 2021 Flexible and efficient inference with particles for the variational Gaussian approximation *Entropy* **23** 990
[37] Rezende D and Mohamed S 2015 Variational inference with normalizing flows *Int. Conf. Machine Learning (PMLR)* pp 1530–8
[38] Rossky P J, Doll J D and Friedman H L 1978 Brownian dynamics as smart Monte Carlo simulation *J. Chem. Phys.* **69** 4628–33
[39] Roberts G O and Tweedie R L 1996 Exponential convergence of Langevin distributions and their discrete approximations *Bernoulli* **2** 341–63
[40] Vrugt J A, ter Braak C J F, Diks C G H, Robinson B A, Hyman J M and Higdon D 2009 Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling *Int. J. Nonlinear Sci. Numer. Simul.* **10** 273–90
[41] Foreman-Mackey D, Hogg D W, Lang D and Goodman J 2013 EMCEE: the MCMC hammer *Publ. Astron. Soc. Pac.* **125** 306
[42] Leimkuhler B, Matthews C and Weare J 2018 Ensemble preconditioning for Markov chain Monte Carlo simulation *Stat. Comput.* **28** 277–90
[43] Smith A 2013 *Sequential Monte Carlo Methods in Practice* (Berlin: Springer)
[44] Evensen G 2009 The ensemble Kalman filter for combined state and parameter estimation *IEEE Control Syst.* **29** 83–104
[45] Reich S and Cotter C 2015 *Probabilistic Forecasting and Bayesian Data Assimilation* (Cambridge: Cambridge University Press)
[46] Law K, Stuart A and Zygalakis K 2015 *Data Assimilation* (Berlin: Springer)
[47] Reich S 2011 A dynamical systems framework for intermittent data assimilation *BIT Numer. Math.* **51** 235–49
[48] Moselhy T A E and Marzouk Y M 2012 Bayesian inference with optimal maps *J. Comput. Phys.* **231** 7815–50
[49] Reich S 2013 A nonparametric ensemble transform method for Bayesian inference *SIAM J. Sci. Comput.* **35** A2013–24
[50] Marzouk Y, Moselhy T, Parno M and Spantini A 2016 Sampling via measure transport: an introduction *Handbook of Uncertainty Quantification* (New York: Springer) pp 1–41
[51] Ruchi S, Dubinkina S and Iglesias M A 2019 Transform-based particle filtering for elliptic Bayesian inverse problems *Inverse Problems* **35** 115005
[52] Carrillo J A, Choi Y-P, Totzeck C and Tse O 2018 An analytical framework for consensus-based global optimization method *Math. Models Methods Appl. Sci.* **28** 1037–66
[53] Carrillo J A, Hoffmann F, Stuart A M and Vaes U 2021 Consensus-based sampling *Stud. Appl. Math.* **148** 1069–140
[54] Pidstrigach J and Reich S 2021 Affine-invariant ensemble transform methods for logistic regression (arXiv:2104.08061)
[55] Emerick A A and Reynolds A C 2013 Investigation of the sampling performance of ensemble-based methods with a simple reservoir model *Comput. Geosci.* **17** 325–50
[56] Chen Y and Oliver D S 2012 Ensemble randomized maximum likelihood method as an iterative ensemble smoother *Math. Geosci.* **44** 1–26
[57] Iglesias M A, Law K J H and Stuart A M 2013 Ensemble Kalman methods for inverse problems *Inverse Problems* **29** 045001
[58] Wan E A and Van Der Merwe R 2000 The unscented Kalman filter for nonlinear estimation *Proc. IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symp. (Cat. No. 00EX373)* (Piscataway, NJ: IEEE) pp 153–8
[59] Kalman R E 1960 A new approach to linear filtering and prediction problems *J. Basic Eng. Mar* **82** 35–45
[60] Sorenson H W 1985 *Kalman Filtering: Theory and Application* (Piscataway, NJ: IEEE)
[61] Jazwinski A H 2007 *Stochastic Processes and Filtering Theory* (New York: Dover)
[62] Evensen G 1994 Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics *J. Geophys. Res.* **99** 10143–62
[63] Anderson J L 2001 An ensemble adjustment Kalman filter for data assimilation *Mon. Weather Rev.* **129** 2884–903

[64] Bishop C H, Etherton B J and Majumdar S J 2001 Adaptive sampling with the ensemble transform Kalman filter: I. Theoretical aspects *Mon. Weather Rev.* **129** 420–36
[65] Julier S J, Uhlmann J K and Durrant-Whyte H F 1995 A new approach for filtering nonlinear systems *Proc. 1995 American Control Conf. ACC'95* vol 3 (Piscataway, NJ: IEEE) pp 1628–32
[66] Tippett M K, Anderson J L, Bishop C H, Hamill T M and Whitaker J S 2003 Ensemble square root filters *Mon. Weather Rev.* **131** 1485–90
[67] Bocquet M and Sakov P 2013 Joint state and parameter estimation with an iterative ensemble Kalman smoother *Nonlinear Process. Geophys.* **20** 803–18
[68] Iglesias M and Yang Y 2021 Adaptive regularisation for ensemble Kalman inversion *Inverse Problems* **37** 025008
[69] Matveev M Y, Endruweit A, Long A C, Iglesias M A and Tretyakov M V 2021 Bayesian inversion algorithm for estimating local variations in permeability and porosity of reinforcements using experimental data *Composites* A **143** 106323
[70] Tso C-H M, Iglesias M, Wilkinson P, Kuras O, Chambers J and Binley A 2021 Efficient multiscale imaging of subsurface resistivity with uncertainty quantification using ensemble Kalman inversion *Geophys. J. Int.* **225** 887–905
[71] Chada N K, Stuart A M and Tong X T 2020 Tikhonov regularization within ensemble Kalman inversion *SIAM J. Numer. Anal.* **58** 1263–94
[72] Huang D Z, Schneider T and Stuart A M 2022 Iterated Kalman methodology for inverse problems *J. Comput. Phys.* **463** 111262
[73] Weissmann S, Chada N K, Schillings C and Tong X T 2022 Adaptive Tikhonov strategies for stochastic ensemble Kalman inversion *Inverse Problems* **38** 045009
[74] Ernst O G, Sprungk B and Starkloff H-J 2015 Analysis of the ensemble and polynomial chaos Kalman filters in Bayesian inverse problems *SIAM/ASA J. Uncertain. Quantification* **3** 823–51
[75] Garbuno-Inigo A, Hoffmann F, Li W and Stuart A M 2020 Interacting Langevin diffusions: gradient structure and ensemble Kalman sampler *SIAM J. Appl. Dyn. Syst.* **19** 412–41
[76] Garbuno-Inigo A, Nüsken N and Reich S 2020 Affine invariant interacting Langevin dynamics for Bayesian inference *SIAM J. Appl. Dyn. Syst.* **19** 1633–58
[77] Nüsken N and Reich S 2019 Note on interacting Langevin diffusions: gradient structure and ensemble Kalman sampler by Garbuno-Inigo, Hoffmann, Li and Stuart (arXiv:1908.10890)
[78] Sanz-Alonso D, Stuart A M and Taeb A 2018 Inverse problems and data assimilation (arXiv:1810.06191)
[79] Calvello E, Reich S and Stuart A M 2022 Ensemble Kalman methods: a mean field perspective (arXiv:2209.11371)
[80] Lancaster P and Rodman L 1995 *Algebraic Riccati Equations* (Oxford: Clarendon)
[81] Bougerol P 1993 Kalman filtering with random coefficients and contractions *SIAM J. Control Optim.* **31** 942–59
[82] Julier S J and Uhlmann J K 2002 Reduced sigma point filters for the propagation of means and covariances through nonlinear transformations *Proc. 2002 American Control Conf. (IEEE Cat. No. CH37301)* vol 2 (Piscataway, NJ: IEEE) pp 887–92
[83] Evensen G 2009 *Data Assimilation: The Ensemble Kalman Filter* (Berlin: Springer)
[84] Wang X and Bishop C H 2003 A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes *J. Atmos. Sci.* **60** 1140–58
[85] Anderson J L 2007 An adaptive covariance inflation error correction algorithm for ensemble filters *Tellus* A **59** 210–24
[86] Chen Y and Oliver D S 2017 Localization and regularization for iterative ensemble smoothers *Comput. Geosci.* **21** 13–30
[87] Ghanem R G and Spanos P D 2003 *Stochastic Finite Elements: A Spectral Approach* (New York: Dover)
[88] Stuart A M 2010 Inverse problems: a Bayesian perspective *Acta Numer.* **19** 451–559
[89] Ehrendorfer M 2011 *Spectral Numerical Weather Prediction Models* (Philadelphia, PA: SIAM)
[90] Constantine P G, Dow E and Wang Q 2014 Active subspace methods in theory and practice: applications to Kriging surfaces *SIAM J. Sci. Comput.* **36** A1500–24
[91] Cui T, Martin J, Marzouk Y M, Solonen A and Spantini A 2014 Likelihood-informed dimension reduction for nonlinear inverse problems *Inverse Problems* **30** 114015
[92] Cui T and Tong X T 2021 A unified performance analysis of likelihood-informed subspace methods (arXiv:2101.02417)

[93] Heinrich S 2001 Multilevel Monte Carlo methods *Int. Conf. Large-Scale Scientific Computing* (Springer) pp 58–67
[94] Giles M B 2015 Multilevel Monte Carlo methods *Acta Numer.* **24** 259–328
[95] Gao H and Wang J-X 2021 A bi-fidelity ensemble Kalman method for PDE-constrained inverse problems in computational mechanics *Comput. Mech.* **67** 1115–31
[96] Fairbanks H R, Jofre L, Geraci G, Iaccarino G and Doostan A 2020 Bi-fidelity approximation for uncertainty quantification and sensitivity analysis of irradiated particle-laden turbulence *J. Comput. Phys.* **402** 108996
[97] Hoel H, Law K J H and Tempone R 2016 Multilevel ensemble Kalman filtering *SIAM J. Numer. Anal.* **54** 1813–39
[98] Chada N K, Jasra A and Yu F 2020 Multilevel ensemble Kalman–Bucy filters *SIAM/ASA J. Uncertain. Quantification* **10** 584–618
[99] Albers D J, Blancquart P-A, Levine M E, Seylabi E E and Stuart A 2019 Ensemble Kalman methods with constraints *Inverse Problems* **35** 095007
[100] Chada N K, Schillings C and Weissmann S 2019 On the incorporation of box-constraints for ensemble Kalman inversion (arXiv:1908.00696)
[101] Schneider T, Stuart A M and Wu J-L 2020 Ensemble Kalman inversion for sparse learning of dynamical systems from time-averaged data (arXiv:2007.06175)
[102] Herty M and Visconti G 2018 Kinetic methods for inverse problems (arXiv:1811.09387)
[103] Pinnau R, Totzeck C, Tse O and Martin S 2017 A consensus-based model for global optimization and its mean-field limit *Math. Models Methods Appl. Sci.* **27** 183–204
[104] Carrillo J A, Totzeck C and Vaes U 2021 Consensus-based optimization and ensemble Kalman inversion for global optimization problems with constraints (arXiv:2111.02970)
[105] Carrillo J A, Jin S, Li L and Zhu Y 2021 A consensus-based global optimization method for high dimensional machine learning problems *ESAIM Control, Optim. Calc. Var.* **27** S5
[106] Asch M, Bocquet M and Nodet M 2016 *Data Assimilation: Methods, Algorithms, and Applications* (Philadelphia, PA: SIAM)
[107] Oliver D S, Reynolds A C and Liu N 2008 *Inverse Theory for Petroleum Reservoir Characterization and History Matching* (Cambridge: Cambridge University Press)
[108] Ding Z, Qin L and Lu J 2020 Ensemble Kalman inversion for nonlinear problems: weights, consistency, and variance bounds (arXiv:2003.02316)
[109] Held I M and Suarez M J 1994 A proposal for the intercomparison of the dynamical cores of atmospheric general circulation models *Bull. Am. Meteorol. Soc.* **75** 1825–30