Unscented Kalman Inversion

Daniel Z. Huang[a], Tapio Schneider[a], Andrew M. Stuart[a]

[a] *California Institute of Technology, Pasadena, CA*

Abstract

A useful approach to solve inverse problems is to pair the parameter-to-data map with a stochastic dynamical system for the parameter, and then employ techniques from filtering to estimate the parameter given the data. Three classical approaches to filtering of nonlinear systems are the extended, ensemble and unscented Kalman filters. The extended Kalman filter (which we refer to as ExKI in the context of inverse problems) is impractical when the forward map is not readily differentiable and given as a black box, and also for high dimensional parameter spaces because of the need to propagate large covariance matrices. Ensemble Kalman inversion (EKI) has emerged as a useful tool which overcomes both of these issues: it is derivative free and works with a low-rank covariance approximation formed from the ensemble. In this paper, we demonstrate that unscented Kalman methods also provide an effective tool for derivative-free inversion in the setting of black-box forward models, introducing unscented Kalman inversion – UKI.

Theoretical analysis is provided for linear inverse problems, and a smoothing property of the data mis-fit, under the unscented transform used to define the UKI as a modification of the ExKI, is explained. We provide numerical experiments, including proof-of-concept linear examples and various applications: learning of permeability parameters in subsurface flow; learning the damage field from structure deformation; learning the Navier-Stokes initial condition from solution data at positive times; and learning subgrid-scale parameters in a general circulation model (GCM) from time-averaged statistics. The theory and experiments show that the UKI outperforms the EKI on parameter learning problems with moderate numbers of parameters and outperforms the ExKI on problems where the forward model is not readily differentiable, or where the derivative is very sensitive. In particular, UKI based methods are of particular value for parameter estimation problems in which the number of parameters is moderate but the forward model is expensive and provided as a black box which is impractical to differentiate.

Keywords: Inverse Problem, Optimization, Sampling, Filtering, Extended Kalman Methods, Ensemble Kalman Methods, Unscented Kalman Methods,

1. Introduction

1.1. Overview

This paper is concerned with inversion of the map $\mathcal{G} : \mathbb{R}^{N_\theta} \to \mathbb{R}^{N_y}$, in the presence of noise. We assume we are given $y \in \mathbb{R}^{N_y}$ and wish to recover $\theta \in \mathbb{R}^{N_\theta}$ from the relation

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

Email addresses: dzhuang@caltech.edu (Daniel Z. Huang), tapio@caltech.edu (Tapio Schneider), astuart@caltech.edu (Andrew M. Stuart)

The distribution of η, the observational noise, is assumed known, but not its value; to be concrete we will assume that it is drawn from a Gaussian with distribution $\mathcal{N}(0, \Sigma_\eta)$. Central to both the optimization and probabilistic approaches to inversion is the regularized objective function $\Phi_R(\theta)$ defined by

$$\Phi_R(\theta) := \Phi(\theta) + \frac{1}{2}\|\Sigma_0^{-\frac{1}{2}}(\theta - r_0)\|^2, \tag{2a}$$

$$\Phi(\theta) := \frac{1}{2}\|\Sigma_\eta^{-\frac{1}{2}}(y - \mathcal{G}(\theta))\|^2, \tag{2b}$$

where Σ_η normalizes the model-data misfit Φ by means of the known error statistics of the noise and r_0 and Σ_0 encode prior mean and covariance information about θ. We assume that Σ_η is strictly positive-definite, denoted $\Sigma_\eta \succ 0$, and similarly that $\Sigma_0 \succ 0$. [1]

The focus of this paper is mainly on derivative-free inversion by means of iterative techniques aimed at solving the optimization problem defined by minimization of Φ_R, or variants of this problem [1]. However, even in the optimization setting, the methods introduced in this paper are closely related to iterative methods applied in Bayesian (probabilistic) inversion which we first describe. In the Bayesian approach to the inverse problem [2, 3] the posterior distribution is given by

$$\mu(d\theta) = \frac{1}{Z}\exp\big(-\Phi(\theta)\big)\mu_0(d\theta), \tag{3}$$

where $\mu_0 = \mathcal{N}(r_0, \Sigma_0)$ is the prior and μ the posterior. A commonly adopted iterative approach to solve the problem of sampling from μ is sequential Monte Carlo (SMC) [4] in which the measures μ_n defined by

$$\mu_{n+1}(d\theta) = \frac{1}{Z_n}\exp\big(-h\Phi(\theta)\big)P_n\mu_n(d\theta) \tag{4}$$

are successively approximated by ensembles. In SMC P_n is chosen to be a μ_n-invariant Markov kernel so that $P_n\mu_n = \mu_n$. Note, then, that if $Nh = 1$ it follows that $\mu_N = \mu$. Thus the successive ensembles approximate μ_N after N steps. Application of this methodology to solve inverse problems may be found in [5]. On the other hand if $h = 1$ is fixed and the measures μ_n are studied in the limit of large n they will tend to concentrate on minimizers of Φ, restricted to the support of μ_0, providing an approach to solving an unregularized inverse problem; this follows from the identity

$$\mu_n(d\theta) = \frac{1}{\big(\Pi_{\ell=0}^{n-1} Z_\ell\big)}\exp\big(-n\Phi(\theta)\big)\mu_0(d\theta). \tag{5}$$

We will build on the latter, optimization, approach to the problem. However we note that, other than restriction of μ_n to the support of μ_0, regularization is lost in this approach since it focuses on minimizing $\Phi(\cdot)$ and not Φ_R. To address this issue we will choose P_j to be the Markov kernel associated with a first-order autoregressive (AR1) process. The resulting dynamic on measures may be understood by considering the stochastic dynamical system

$$\text{evolution:} \qquad \theta_{n+1} = r + \alpha(\theta_n - r) + \omega_{n+1}, \qquad \omega_{n+1} \sim \mathcal{N}(0, \Sigma_\omega), \tag{6a}$$

$$\text{observation:} \qquad y_{n+1} = \mathcal{G}(\theta_{n+1}) + \nu_{n+1}, \qquad \nu_{n+1} \sim \mathcal{N}(0, \Sigma_\nu). \tag{6b}$$

We assume that the regularization parameter $\alpha \in (0, 1]$, the artificial evolution error covariance $\Sigma_\omega \succ 0$, and the artificial observation error covariance $\Sigma_\nu \succ 0$; r is an arbitrary vector. If $Y_n =$

[1] We will also write $A \preceq B$ when $B - A$ is positive semi-definite and $A \prec B$ when $B - A$ is positive definite.

$\{y_\ell\}_{\ell=1}^n$ then $\mu_n(d\theta)$, the conditional distribution of $\theta_n|Y_n$, is determined by the iteration (4) with $h = 1$ and P_n the Markov kernel associated with (6a), provided the choice $y_n \equiv y$ is made in the conditional measure. Note that μ_n is not invariant with respect to P_n with this AR1 choice; thus the method differs from SMC in this regard and the choice of P_n is made here in order to regularize the iterative optimization approach to inversion encapsulated in (5).

The method we introduce and study in this paper arises from application of ideas from Kalman filtering to the problem of approximating the distribution of $\theta_n|Y_n$. The Kalman filter itself applies to the case of linear $\mathcal{G}$ [6, 7]. When $\mathcal{G}$ is nonlinear the methods can be generalized by use of the extended Kalman filter (ExKF) [8] which is based on linearization and application of Kalman methodology. However this method suffers from two drawbacks which hamper its application in many large-scale applications: (a) it requires a derivative of the forward map $\mathcal{G}(\cdot)$; and (b) the approach scales poorly to high dimensional parameter spaces where $N_\theta \gg 1$, because of the need to sequentially update covariances in $\mathbb{R}^{N_\theta \times N_\theta}$. Thus, despite an early realization that Kalman-based methods could be useful for large-scale filtering problems arising in the geosciences [9], the methods did not become practical in this context until the work of Evensen [10]. This revolutionary paper introduced the ensemble Kalman filter (EnKF) the essence of which is to avoid the linearization of the dynamics and sequential updating of the covariance, and instead use a low rank approximation of the covariance found by maintaining an ensemble of estimates for $\theta_n|Y_n$ at every step n. These ensemble Kalman methods have been widely adopted in the geosciences, not only because they are effective for high dimensional parameter spaces, but also because they are derivative-free, requiring only $\mathcal{G}$ as a black box. Their use in the solution of inverse problems via iterative methods was pioneered in subsurface inversion [11, 12] where the perspective of fixing $h \ll 1$ and iterating until $n = N = 1/h$ was used, so that μ_N is viewed as an approximation of the posterior, provided μ_0 is chosen as the prior. These papers thus view the ensemble methodology as a way of sampling from the posterior and have elements in common with SMC; this idea is also implicit in the paper [13] which is focussed on data assimilation, and addresses solution of a Bayesian inverse problem each time new data is received.

In [14] the Kalman methodology for inversion was revisited from the optimization perspective, based on fixing $h = 1$ and iterating in n, leading to an algorithm we will refer to as ensemble Kalman inversion (EKI). The paper [15] introduced a novel approach to regularizing the iterative method, by drawing analogy with the Levenberg-Marquardt method [16]; see also [17]. Subsequent variants on the iterative optimization approach demonstrate how to introduce Tikhonov regularization into the EKI algorithm [18] and the paper [19] shows that adding noise to the iteration can lead to approximate Bayesian inversion, a method we will refer to as ensemble Kalman sampling (EKS) and which is further analyzed in [20, 21]. The EKS provides a different approach to the problem of Bayesian inversion from the ones pioneered in [11, 12] since it does not require starting with draws from the prior μ_0, but instead relies on ergodicity and iteration to large n; the methods in [11, 12] must be started with draws from the prior μ_0 and iterated for precisely $n = 1/h$ steps, and are hence more rigid in their requirements. Since the ensemble methods do not, in general, accurately approximate the true posterior distribution [22, 23] outside Gaussian scenarios, the derivative-free optimization perspective is arguably a more natural avenue within which to analyze ensemble inversion. However recent work demonstrates how a derivative-free multiscale stochastic sampling method can usefully take the output of EKS as a preconditioner for a method which provably approximates the true posterior distribution [24]; in that context the EKS is central to making the method efficient.

Within the control theory literature, and parallel to the development of the ensemble Kalman filter, the unscented Kalman filter (UKF) was introduced [25, 26]. Like the ensemble Kalman methods, this method also sidesteps the need to sequentially update the derivative of the forward

model as part of the covariance update; but, in the primary difference from ensemble Kalman methods, particles (sigma points) are chosen deterministically, and a quadrature rule is applied within a Gaussian approximation of the filter. The goal of this paper is to establish a framework for the development of unscented Kalman methods for inverse problems, based on (6): we formalize, and demonstrate the power of, unscented Kalman inversion (UKI) techniques. We also formalize extended Kalman inversion (ExKI) as a general purpose methodology for parameter learning and describe ExKI, UKI and UKI as different approximations of a general Gaussian methodology for the filtering problem defined by (6).

1.2. Our Contributions

We make the following contributions to the study of the UKI methodology:

- we establish the UKI algorithm as a general purpose derivative-free approach to solving the inverse problem (1), based on the novel stochastic dynamical system formulation (6);
- we derive the UKI methodology by introducing a useful conceptual algorithm, based on Gaussian approximation for the filtering distribution defined by (6) and, in the same framework, relate the UKI to the EKI and the ExKI;
- by studying linear problems we demonstrate that the methodology induces a form of Tikhonov regularization and we prove exponential convergence of the algorithm to the minimizer of the Tikhonov-regularized problem, in the linear case;
- we show that the method performs like a generalized Levenberg–Marquardt Algorithm for the ExKI and we demonstrate that the UKI may be viewed as an approximation of the generalized Levenberg–Marquardt Algorithm applied to a smoothed data-misfit;
- we show that UKI outperforms EKI for certain inverse problems with moderate-dimensional parameters, and for the UKI all tests converge within $\mathcal{O}(10)$ iterations with no empirical variance inflation or early stopping needed;
- the UKI is tested on a wide range of problems, including linear test problems, inversion for spatial fields in a variety of continuum mechanics applications and the learning of parameters in chaotic dynamical systems, using time-averaged data;
- we introduce the unscented Kalman sampler (UKS), an unscented Kalman method for generating approximate samples from the Bayesian posterior distribution (3).

Taken together, the theoretical framework we develop and the numerical results we present show that the UKI is a competitive methodology for solving inverse problems and parameter estimation problems defined by an expensive black-box forward model; indeed the UKI is shown to outperform the EKI in settings where the number of parameters N_θ is of moderate size and the black-box is not readily differentiable so that ExKI methods are not applicable.

We conclude this introductory section with a deeper literature review relating to the contributions we make in this paper, in Subsection 1.3. Then, in Section 2 we derive the algorithm from a Gaussian approximation of the filtering distribution associated with (6), and relate to the ExKI and EKI approaches to the problem. In Section 3 we study the methodology for linear problems, obtaining insight into the regularization conferred by (6a); study the relationship of the methodology to other gradient-based optimization techniques; and derive continuous time limits in the nonlinear setting. Section 4 describes variants on the basic UKI algorithm that we have found useful in practice, including the UKS, and in Section 5 we present numerical results demonstrating the performance of the UK approaches.

1.3. Literature Review

Inverse and parameter estimation problems are ubiquitous in engineering and scientific applications. Applications that motivate this work include global climate model calibration [27, 28, 29], material constitutive relation calibration [30, 31, 32], seismic inversion in geophysics [33, 34], and medical tomography [35, 36]. These problems are generally highly nonlinear, may feature multiple scales and may include chaotic and turbulent phenomena. Moreover, the observational data is often noisy and the inverse problem may be ill-posed. We note, also, that a number of inverse problems of interest may involve a moderate number of unknown parameters N_θ, yet may involve solution of a very expensive forward model $\mathcal{G}$ depending on those parameters; furthermore $\mathcal{G}$ may not be differentiable with respect to the parameters, or may be complex to differentiate as it is given as a black box.

In the nonlinear setting of state estimation, there are three primary Kalman filters [37, 38, 39]: the extended Kalman filter (ExKF), the unscented Kalman filter (UKF), and the ensemble Kalman filter (EnKF). The use of Kalman based methodology as a non-intrusive iterative method for parameter estimation originates in the papers [40, 41] which were based on the ExKF, hence requiring derivative $d\mathcal{G}$, and its adjoint, to propagate covariances; the use of derivative-free ensemble methods was then developed systematically in the papers [11, 12], leading to the EKI [14]. Derivative-free ensemble inversion and parameter estimation is particularly suitable for complex multiphysics problems requiring coupling of different solvers, such as fluid-structure interaction [42, 43, 44, 45] and general circulation models [46] and methods containing discontinuities such as the immersed/embedded boundary method [47, 48, 49, 50] and adaptive mesh refinement [51, 52]. Furthermore, derivative-free ensemble inversion and parameter estimation has been demonstrated to be effective in the context of forward models defined by chaotic dynamical systems [53] where adjoint-based methods fail to deliver meaningful sensitivities [54, 55]. These wide-ranging potential applications form motivation for developing other derivative-free Kalman based inversion and parameter estimation techniques, and in particular the unscented Kalman methods developed here.

There is already some work in which unscented Kalman methods are used for parameter inversion. Extended, ensemble and unscented Kalman inversions have been applied to train neural networks [40, 41, 26, 56] and EKI has been applied in the oil industry [57, 11, 12]. Dual and joint Kalman filters [58, 26] have been designed to simultaneously estimate the unknown states and the parameters [58, 59, 26, 60, 61] from noisy sequential observations. However, whilst the EKI has been systematically developed and analyzed as a general purpose methodology for the solution of inverse and parameter estimation problems, the same is not the case for UKI.

Continuous time limits and gradient flow structure of the EKI have been introduced and studied in [13, 62, 63, 64, 65]. This work led to the development of variants on the EKI, such as the Tikhonov-EKI (TEKI) [18] and the EKS [19]. We will develop study of continuous time limits for the UKI, and variants including an unscented Kalman sampler (UKS), in this paper. There are interesting links to the Levenberg–Marquardt Algorithm [66, 16], as introduced in [15] and developed further in [67, 68, 17]. We will further refine the idea, which provides insights into understanding and improving UKI.

Finally we mention that there are other derivative-free optimization techniques which are based on interacting particle systems, but are not Kalman based. Rather these methods are based on consensus-forming mean-field models, and their particle approximations, leading to consensus-based optimization [69] and consensus-based sampling [70]. The paper [24] also provides an alternative derivative-free approach to optimization and sampling for inverse problems, using ideas from multiscale dynamical systems.

2. The Algorithm

Recall that the basic approach to inverse problems that we adopt in this paper is to pair the parameter-to-data relationship encoded in (1) with a stochastic dynamical system for the parameter, resulting in (6). We then employ techniques from filtering to approximate the distribution μ_n of $\theta_n|Y_n$. A useful way to think of updating μ_n is through the prediction and analysis steps [71, 72]: $\mu_n \mapsto \hat{\mu}_{n+1}$, and then $\hat{\mu}_{n+1} \mapsto \mu_{n+1}$, where $\hat{\mu}_{n+1}$ is the distribution of $\theta_{n+1}|Y_n$. In Subsection 2.1 we first introduce a Gaussian approximation of the analysis step, leading to an algorithm which maps the space of Gaussian measures into itself at each step of the iteration; it is not implementable in general, but it is a useful conceptual algorithm. Subsection 2.2 shows how this algorithm can be made practical, for low to moderate dimension N_θ and assuming that $d\mathcal{G}$ is available, by means of the ExKF, a form of linearization of the conceptual algorithm; we refer to this as ExKI. In Subsection 2.3 we show how the UKI algorithm, the main focus of this paper, may be derived by applying a quadrature rule to evaluate certain integrals appearing in the conceptual Gaussian approximation. Subsection 2.4 connects the conceptual algorithm with the EKI, an approach in which ensemble approximations of the integrals is used.

2.1. Gaussian Approximation

This conceptual algorithm maps Gaussians into Gaussians. Re refer to it henceforth as the Gaussian Approximation Algorithm (GAA). Assume that $\mu_n \approx \mathcal{N}(m_n, C_n)$. Note that, under (6a), $\hat{\mu}_{n+1}$ is also Gaussian. The algorithm proceeds by introducing the distribution of $\theta_{n+1}, y_{n+1}|Y_n$, projecting this onto a Gaussian by computing its mean and covariance, and then conditioning this Gaussian to obtain a Gaussian approximation $\mathcal{N}(m_{n+1}, C_{n+1})$ to μ_{n+1}.

In the analysis step, we assume that the joint distribution of $\{\theta_{n+1}, y_{n+1}\}|Y_n$ can be approximated by a Gaussian distribution [2]

$$\mathcal{N}\Big(\begin{bmatrix}\widehat{m}_{n+1}\\ \widehat{y}_{n+1}\end{bmatrix}, \begin{bmatrix}\widehat{C}_{n+1} & \widehat{C}^{\theta p}_{n+1}\\ \widehat{C}^{\theta p}_{n+1}{}^T & \widehat{C}^{pp}_{n+1}\end{bmatrix}\Big). \tag{7}$$

Use of (6a) shows that

$$\begin{aligned}\widehat{m}_{n+1} &= \mathbb{E}[\theta_{n+1}|Y_n] = r + \alpha(m_n - r),\\ \widehat{C}_{n+1} &= \mathrm{Cov}[\theta_{n+1}|Y_n] = \alpha^2 C_n + \Sigma_\omega.\end{aligned} \tag{8}$$

Then, with $\mathbb{E}$ denoting expectation with respect to $\theta_{n+1}|Y_n \sim \mathcal{N}(\widehat{m}_{n+1}, \widehat{C}_{n+1})$,

$$\begin{aligned}\widehat{y}_{n+1} =& \mathbb{E}[\mathcal{G}(\theta_{n+1})|Y_n],\\ \widehat{C}^{\theta p}_{n+1} =& \mathrm{Cov}[\theta_{n+1}, \mathcal{G}(\theta_{n+1})|Y_n],\\ \widehat{C}^{pp}_{n+1} =& \mathrm{Cov}[\mathcal{G}(\theta_{n+1})|Y_n] + \Sigma_\nu.\end{aligned} \tag{9}$$

Conditioning the Gaussian in (7) to find $\theta_{n+1}|\{Y_n, y_{n+1}\} = \theta_{n+1}|Y_{n+1}$ gives the following expressions for the mean m_{n+1} and covariance C_{n+1} of the approximation to μ_{n+1} :

$$\begin{aligned}m_{n+1} &= \widehat{m}_{n+1} + \widehat{C}^{\theta p}_{n+1}(\widehat{C}^{pp}_{n+1})^{-1}(y_{n+1} - \widehat{y}_{n+1}),\\ C_{n+1} &= \widehat{C}_{n+1} - \widehat{C}^{\theta p}_{n+1}(\widehat{C}^{pp}_{n+1})^{-1}\widehat{C}^{\theta p}_{n+1}{}^T.\end{aligned} \tag{10}$$

[2] The choice of indices θ and p in this, and what follows, is made to align with the notation in [14] where the data y comprised noisy observation of variable denoted by p.

Equations (8), (9) and (10) define the GAA. As a method for solving the inverse problem (1), the GAA is implemented by assuming all observations $\{y_n\}$ are identical to y and iterating in n., With this assumption, we may write the algorithm as[3]

$$(m_{n+1}, C_{n+1}) = F(m_n, C_n; \mathcal{G}, r, \Sigma_\omega), \tag{11}$$

noting that the mapping is dependent on $\mathcal{G}$ and on the mean and covariance of the assumed auto-regressive dynamics for $\{\theta_n\}$.

In the setting where $\mathcal{G}$ is linear, the Gaussian ansatz used in the derivation of the conceptual algorithm is exact, the integrals appearing in (9) have closed form, and the algorithm reduces to the Kalman filter applied to (6), with a particular assumption on the data stream $\{y_n\}$. In Subsection 3.1 we will show that the mean of this iteration converges to a minimizer of Φ_R given by (2), in which the prior mean of the regularization is $r_0 = r$ and the prior covariance of the regularization Σ_0 is defined by solution of a linear equation depending on the choices of α, Σ_ω, and Σ_ν.

In the nonlinear setting, to make an implementable algorithm from the GAA encapsulated in equations (8) to (10), it is necessary to approximate the integrals appearing in (9). When extended, unscented and ensemble Kalman filters are applied, respectively, to make such approximation, we obtain the ExKI, UKI and EKI algorithms. The extended, unscented and ensemble approaches to this are detailed in the following three subsections. Underlying all of them is the following property of the GAA encapsulated in Proposition 1.

We recall the idea of affine invariance, introduced for MCMC methods in [73], motivated by the attribution of the empirical success of the Nelder-Mead algorithm [74] for optimization to a similar property; further development of the method in the context of sampling algorithms may be found in [75, 20]. In words an iteration is affine invariant if an invertible linear transformation of the variable being iterated makes no difference to the algorithm and hence to the convergence properties of the algorithm; this has the desirable consequence that coordinates in which the problem is well-conditioned may be used to understand convergence of algorithms for ill-conditioned problems.

Consider the invertible mapping from $x \in \mathbb{R}^{N_\theta}$ to ${}^*x \in \mathbb{R}^{N_\theta}$ defined by ${}^*x = Ax + b$. Then define ${}^*\mathcal{G}(\theta) = \mathcal{G}\big(A^{-1}(\theta - b)\big)$, ${}^*r = Ar + b$ and ${}^*\Sigma_\omega = A\Sigma_\omega A^T$.

Proposition 1. *Define, for all $n \in \mathbb{Z}^{0+}$,*

$$ {}^*m_n = Am_n + b \qquad {}^*C_n = AC_nA^T. $$

Then

$$({}^*m_{n+1}, {}^*C_{n+1}) = F({}^*m_n, {}^*C_n; {}^*\mathcal{G}, {}^*r, {}^*\Sigma_\omega). \tag{12}$$

Proof. The proof is in Appendix A. □

This establishes the property of affine invariance, noting that only $\mathcal{G}, r, \Sigma_\omega$ need to be transformed as the affine map applies only on the signal space for $\{\theta_n\}$ and not the observation space for $\{y_n\}$.

2.2. Extended Kalman Inversion

Consider the algorithm defined by equations (8) to (10). The ExKI algorithm follows from invoking the approximations

$$\mathcal{G}(\theta_{n+1}) \approx \mathcal{G}(\widehat{m}_{n+1}) + d\mathcal{G}(\widehat{m}_{n+1})(\theta_{n+1} - \widehat{m}_{n+1}) \tag{13}$$

[3]Dependencies on other matrices entering the specification of (6) are not included in this notation as they remain unchanged in the discussion of affine invariance which follows below.

in the analysis updates for the mean and covariance respectively. In particular both the mean and the covariances in (9) can be evaluated in closed form with the approximation (13). The approximations are valid if the fluctuations around the mean state are small, say of $\mathcal{O}(\epsilon) \ll 1$, and all the covariances are $\mathcal{O}(\epsilon^2)$. This results in the following algorithm:

- Prediction step :

$$\begin{aligned} \widehat{m}_{n+1} =& r + \alpha(m_n - r), \\ \widehat{C}_{n+1} =& \alpha^2 C_n + \Sigma_\omega. \end{aligned} \tag{14}$$

- Analysis step :

$$\begin{aligned} &\widehat{y}_{n+1} = \mathcal{G}(\widehat{m}_{n+1}), \\ &\widehat{C}^{\theta p}_{n+1} = \widehat{C}_{n+1} d\mathcal{G}(\widehat{m}_{n+1})^T, \\ &\widehat{C}^{pp}_{n+1} = d\mathcal{G}(\widehat{m}_{n+1}) \widehat{C}_{n+1} d\mathcal{G}(\widehat{m}_{n+1})^T + \Sigma_\nu, \\ &m_{n+1} = \widehat{m}_{n+1} + \widehat{C}^{\theta p}_{n+1} (\widehat{C}^{pp}_{n+1})^{-1} \big(y - \widehat{y}_{n+1}\big), \\ &C_{n+1} = \widehat{C}_{n+1} - \widehat{C}^{\theta p}_{n+1} (\widehat{C}^{pp}_{n+1})^{-1} \widehat{C}^{\theta p}_{n+1}{}^T. \end{aligned} \tag{15}$$

2.3. Unscented Kalman Inversion

UKI approximates the conceptual Gaussian algorithm as does ExKI, but it approximates the integrals appearing in Equations (9) by means of deterministic quadrature rules which are exact when evaluating means and covariances of variables defined as linear transformations of the random variable in question. This is the idea of the unscented transform [25, 26] which we now define.

Definition 1 (Modified Unscented Transform)**.** *Let denote Gaussian random variable $\theta \sim \mathcal{N}(m, C) \in \mathbb{R}^{N_\theta}$, $2N_\theta + 1$ symmetric sigma points are chosen deterministically:*

$$\begin{aligned} \theta^0 &= m, \\ \theta^j &= m + c_j [\sqrt{C}]_j \quad (1 \le j \le N_\theta), \\ \theta^{j+N_\theta} &= m - c_j [\sqrt{C}]_j \quad (1 \le j \le N_\theta), \end{aligned} \tag{16}$$

where $[\sqrt{C}]_j$ is the jth column of the Cholesky factor of C. The quadrature rule approximates the mean and covariance of the transformed variable $\mathcal{G}_i(\theta)$ as follows,

$$\mathbb{E}[\mathcal{G}_i(\theta)] \approx \mathcal{G}_i(\theta^0) \qquad \mathrm{Cov}[\mathcal{G}_1(\theta), \mathcal{G}_2(\theta)] \approx \sum_{j=0}^{2N_\theta} W_j^c (\mathcal{G}_1(\theta^j) - \mathbb{E}\mathcal{G}_1(\theta))(\mathcal{G}_2(\theta^j) - \mathbb{E}\mathcal{G}_2(\theta))^T. \tag{17}$$

Here these constant weights are

$$\begin{aligned} &c_1 = c_2 \cdots = c_{N_\theta} = \sqrt{N_\theta + \lambda} \quad \lambda = a^2(N_\theta + \kappa) - N_\theta, \\ &W_0^c = \frac{\lambda}{N_\theta + \lambda} + (1 - a^2 + b) \quad W_j^c = \frac{1}{2(N_\theta + \lambda)} \ (j = 1, \cdots, 2N_\theta). \end{aligned}$$

The original unscented transform leads to 2nd-order accuracy [76] with respect to small covariance C: the approximation introduces errors in estimating the mean and covariance at the forth and

higher orders in the Taylor series. The modification we employ here replaces the original 2nd-order approximation of the $\mathbb{E}[\mathcal{G}_i(\theta)]$ with its 1st-order counterpart. We do this to avoid negative weights; it also has ramifications for the optimization process which we discuss in Remark 8. Finally we mention that our modified unscented transform retains the properties of exactness for mean and covariance under arbitrary linear transformations $\mathcal{G}_1$ and $\mathcal{G}_2$.

In this paper, the hyper-parameters are chosen to be [4]

$$\kappa = 0 \quad a = \min\{\sqrt{\frac{4}{N_\theta + \kappa}}, 1\} \quad b = 2.$$

Consider the algorithm defined by equations (8) to (10). By utilizing the aforementioned quadrature rule, we obtain the following UKI algorithm:

- Prediction step :

$$\begin{aligned} \widehat{m}_{n+1} =& r + \alpha(m_n - r), \\ \widehat{C}_{n+1} =& \alpha^2 C_n + \Sigma_\omega. \end{aligned} \tag{18}$$

- Generate sigma points :

$$\begin{aligned} &\widehat{\theta}^0_{n+1} = \widehat{m}_{n+1}, \\ &\widehat{\theta}^j_{n+1} = \widehat{m}_{n+1} + c_j[\sqrt{\widehat{C}_{n+1}}]_j \quad (1 \le j \le N_\theta), \\ &\widehat{\theta}^{j+N_\theta}_{n+1} = \widehat{m}_{n+1} - c_j[\sqrt{\widehat{C}_{n+1}}]_j \quad (1 \le j \le N_\theta). \end{aligned} \tag{19}$$

- Analysis step :

$$\begin{aligned} &\widehat{y}^j_{n+1} = \mathcal{G}(\widehat{\theta}^j_{n+1}) \qquad \widehat{y}_{n+1} = \widehat{y}^0_{n+1}, \\ &\widehat{C}^{\theta p}_{n+1} = \sum_{j=0}^{2N_\theta} W^c_j (\widehat{\theta}^j_{n+1} - \widehat{m}_{n+1})(\widehat{y}^j_{n+1} - \widehat{y}_{n+1})^T, \\ &\widehat{C}^{pp}_{n+1} = \sum_{j=0}^{2N_\theta} W^c_j (\widehat{y}^j_{n+1} - \widehat{y}_{n+1})(\widehat{y}^j_{n+1} - \widehat{y}_{n+1})^T + \Sigma_\nu, \\ &m_{n+1} = \widehat{m}_{n+1} + \widehat{C}^{\theta p}_{n+1}(\widehat{C}^{pp}_{n+1})^{-1}(y - \widehat{y}_{n+1}), \\ &C_{n+1} = \widehat{C}_{n+1} - \widehat{C}^{\theta p}_{n+1}(\widehat{C}^{pp}_{n+1})^{-1}\widehat{C}^{\theta p}_{n+1}{}^T. \end{aligned} \tag{20}$$

2.4. Ensemble Kalman Inversion

Consider the conceptual Gaussian approximation algorithm defined by equations (8) to (10). The EKI approach to making this implementable is to work with an ensemble of parameter estimates and approximate the covariances $\widehat{C}^{\theta p}_{n+1}$ and $\widehat{C}^{pp}_{n+1}$ empirically:

[4] We note that the papers [76, 26, 39], suggest using a small positive value of a. We find in the numerical examples considered in this paper that our proposed choice of a outperforms the choice $a = \min\{\sqrt{\frac{4}{N_\theta+\kappa}}, 0.01\}$), which builds in the idea of using a small positive value of a.

- Prediction step :

$$\begin{aligned} \widehat{\theta}_{n+1}^j &= r + \alpha(\theta_n^j - r) + \omega_{n+1}^j, \\ \widehat{m}_{n+1} &= \frac{1}{J}\sum_{j=1}^J \widehat{\theta}_{n+1}^j. \end{aligned} \tag{21}$$

- Analysis step :

$$\begin{aligned} &\widehat{y}_{n+1}^j = \mathcal{G}(\widehat{\theta}_{n+1}^j) \qquad \widehat{y}_{n+1} = \frac{1}{J}\sum_{j=1}^J \widehat{y}_{n+1}^j, \\ &\widehat{C}_{n+1}^{\theta p} = \frac{1}{J-1}\sum_{j=1}^J (\widehat{\theta}_{n+1}^j - \widehat{m}_{n+1})(\widehat{y}_{n+1}^j - \widehat{y}_{n+1})^T, \\ &\widehat{C}_{n+1}^{pp} = \frac{1}{J-1}\sum_{j=1}^J (\widehat{y}_{n+1}^j - \widehat{y}_{n+1})(\widehat{y}_{n+1}^j - \widehat{y}_{n+1})^T + \Sigma_\nu, \\ &\theta_{n+1}^j = \widehat{\theta}_{n+1}^j + \widehat{C}_{n+1}^{\theta p}\left(\widehat{C}_{n+1}^{pp}\right)^{-1}(y - \widehat{y}_{n+1}^j - \nu_{n+1}^j), \\ &m_{n+1} = \frac{1}{J}\sum_{j=1}^J \theta_{n+1}^j. \end{aligned} \tag{22}$$

Here the superscript $j = 1, \cdots, J$ is the ensemble particle index, $\omega_{n+1}^j \sim \mathcal{N}(0, \Sigma_\omega)$ and $\nu_{n+1}^j \sim \mathcal{N}(0, \Sigma_\nu)$ are independent and identically distributed random variables.

3. Theoretical Insights

Recall that we view the GAA as an underlying conceptual algorithm which gives insight into the ExKI, UKI and EKI algorithms. The ExKI is itself an approximation of the GAA, found by linearizing $\mathcal{G}$ around the predictive mean and the UKI and EKI algorithms are approximations of the resulting ExKI. Thus study of the GAA and ExKI give insights into the UKI and EKI algorithms. This section is devoted to such studies. In Subsection 3.1 we consider behaviour of the GAA in the linear setting, where it is identical to the ExKI. In Subsection 3.2 we show that the ExKI may be viewed as a generalization of the Levenberg-Marquardt method for optimization. Subsection 3.3 exhibits an averaging property induced by the unscented approximation, indicating how this may help in solving problems with rough energy landscapes. And in Subsection 3.4 we study a continuous time limit of the GAA, which may itself be approximated to obtain continuous time limits of the ExKI, UKI and EKI algorithms; this provides insight into the discrete algorithms as implemented in practice.

3.1. The Linear Setting

In the linear setting the stochastic dynamical system for state $\{\theta_n\}$ and observations $\{y_n\}$ is given by

$$\text{evolution:} \qquad \theta_{n+1} = r + \alpha(\theta_n - r) + \omega_{n+1}, \qquad \omega_{n+1} \sim \mathcal{N}(0, \Sigma_\omega), \tag{23a}$$

$$\text{observation:} \qquad y_{n+1} = G\theta_{n+1} + \nu_{n+1}, \qquad \nu_{n+1} \sim \mathcal{N}(0, \Sigma_\nu). \tag{23b}$$

Thanks to the linearity, equations (9) reduce to

$$\widehat{y}_{n+1} = Gm_n, \quad \widehat{C}^{\theta p}_{n+1} = \widehat{C}_{n+1}G^T, \quad \text{and} \quad \widehat{C}^{pp}_{n+1} = G\widehat{C}_{n+1}G^T + \Sigma_\nu.$$

The update equations (10) become

$$\begin{aligned} \widehat{m}_{n+1} &= r + \alpha(m_n - r), \\ \widehat{C}_{n+1} &= \alpha^2 C_n + \Sigma_\omega, \end{aligned} \tag{24}$$

and

$$m_{n+1} = \widehat{m}_{n+1} + \widehat{C}_{n+1}G^T(G\widehat{C}_{n+1}G^T + \Sigma_\nu)^{-1}\Big(y - G\widehat{m}_{n+1}\Big), \tag{25a}$$

$$C_{n+1} = \widehat{C}_{n+1} - \widehat{C}_{n+1}G^T(G\widehat{C}_{n+1}G^T + \Sigma_\nu)^{-1}G\widehat{C}_{n+1}. \tag{25b}$$

We have the following theorem about the convergence of the UKI in the setting of the linear forward model:

Theorem 1. *Assume that $\Sigma_\omega \succ 0$. Consider the iteration (24), (25) mapping (m_n, C_n) into (m_{n+1}, C_{n+1}). Assume further $\alpha = 1$ and Range$(G^T) = \mathbb{R}^{N_\theta}$ or that $\alpha \in (0,1)$. Then the steady state equation of equation (25b)*

$$C_\infty^{-1} = G^T\Sigma_\nu^{-1}G + (\alpha^2 C_\infty + \Sigma_\omega)^{-1} \tag{26}$$

has a unique solution $C_\infty \succ 0$. The pair (m_n, C_n) converges exponentially fast to limit (m_∞, C_∞). Furthermore the limiting mean m_∞ is the minimizer of the Tikhonov regularized least squares functional Φ_R given by

$$\Phi_R(\theta) := \frac{1}{2}\|\Sigma_\nu^{-\frac{1}{2}}(y - G\theta)\|^2 + \frac{1-\alpha}{2}\|\widehat{C}_\infty^{-\frac{1}{2}}(\theta - r)\|^2, \tag{27}$$

where

$$\widehat{C}_\infty = \alpha^2 C_\infty + \Sigma_\omega. \tag{28}$$

Proof. The proof is in Appendix A. □

Remark 1. *The theorem suggests the importance of choosing $\alpha \in (0,1)$ unless the forward operator has empty null-space. In the case $\alpha \in (0,1)$ the preceding theorem demonstrates the regularization which underlies the proposed iterative method.*

Remark 2. *The free parameters $r, \Sigma_\nu, \Sigma_\omega$ define the prior mean and covariance r_0 and Σ_0 appearing in (2). However it is important to realize that, despite the clear parallels with Tikhonov regularization [1], there is an important difference: the matrix $\widehat{C}_\infty$ defining the* implied prior covariance *in the regularization term depends on the forward model. This may be seen by noting that it is defined by (28) in terms of the steady state covariance C_∞ satisfying (26). To get some insight into the implications of this, we consider the over-determined linear system in which $G^T\Sigma_\eta^{-1}G$ is invertible and we may define*

$$C_* = (G^T\Sigma_\eta^{-1}G)^{-1}. \tag{29}$$

If we choose $r = r_0$, the desired prior mean, and the artificial evolution and observation error covariances

$$\Sigma_\nu = 2\Sigma_\eta, \tag{30a}$$

$$\Sigma_\omega = \big(2 - \alpha^2\big)C_*, \tag{30b}$$

then straightforward calculation with (26), (28) *shows that*

$$C_\infty = C_*, \quad \widehat{C}_\infty = 2C_*.$$

From (27) *it follows that*

$$\Phi_R(\theta) = \frac{1}{4}\left\|\Sigma_\eta^{-\frac{1}{2}}(y - G\theta)\right\|^2 + \frac{(1-\alpha)}{4}\left\|\Sigma_\eta^{-\frac{1}{2}}G(\theta - r_0)\right\|^2. \tag{31}$$

This calculation clearly demonstrates the dependence of the second (regularization) term on the forward model and that choosing $\alpha \in (0, 1]$ *allows different weights on the regularization term. Indeed by allowing a multiplicative factor different from* 2 *in (30a) the regularization term can be given arbitrarily large weight relative to the data misfit.*

Remark 3. *The prior regularization is defined through the steady state of the iterative process, in contrast to SMC where the prior is specified as the initial condition.*

Remark 4. *Let* $\alpha = 1$ *and consider the setting where the forward operator* G *has null space, so that the problem is under-determined. Then the steady state precision matrix* C_∞^{-1} *of equation* (26) *is singular. Moreover,* $\{C_n\}$ *diverges to* $+\infty$ *with the following linear bound*

$$C_n \preceq C_0 + n\Sigma_\omega,$$

but $\{m_n\}$ *still converges to a stationary point of* $\Phi(\theta)$*, which for* $\alpha = 1$ *has no regularizing term; this phenomenon is illustrated in the numerical experiments presented in Subsection 5.3.*

Remark 5. *When* $\alpha \in (0, 1)$*, the exponential convergence rates of the mean and covariance are independent of the condition number of* $G^T\Sigma_\nu^{-1}G$*.*

Remark 6. *When* $\alpha \in (0, 1)$*,* $\widehat{C}_\infty$ *is bounded,* $\Sigma_\omega \preceq \widehat{C}_\infty \preceq \dfrac{\Sigma_\omega}{1-\alpha^2}$*, since* $0 \preceq C_\infty \preceq \alpha^2 C_\infty + \Sigma_\omega$*.*

Remark 7. *When* $\Sigma_\omega = 0$*,* $\lim_{n\to\infty} C_n = 0$*.*

3.2. ExKI: Levenberg–Marquardt Connection

In the nonlinear setting, our numerical results will demonstrate the implicit regularization and linear (sometimes superlinear) convergence of ExKI and UKI. This desirable features can be understood by the analogy with the Levenberg–Marquardt Algorithm (LMA). We focus this discussion on the particular case $\alpha = 1$ as we find that, for over-determined problems, this choice often produces the best results.

Consider the non-regularized nonlinear least-squares objective function Φ, defined in (2b). The key step in the Levenberg–Marquardt Algorithm (LMA) is to solve the minimization problem for(2b) by a preconditioned gradient descent procedure which maps θ_n to $\theta_n + \delta\theta_n$ and where $\delta\theta_n$ solves

$$(d\mathcal{G}(\theta_n)^T\Sigma_\nu^{-1}d\mathcal{G}(\theta_n) + \lambda_n\mathbb{I})\delta\theta_n = d\mathcal{G}(\theta_n)^T\Sigma_\nu^{-1}(y - \mathcal{G}(\theta_n)). \tag{32}$$

Here $\mathbb{I}$ is the identity matrix on $\mathbb{R}^{N_\theta}$ and λ_n is the (non-negative) damping factor, often chosen adaptively. Because of the damping matrix $\lambda_n\mathbb{I}$, the LMA is found to be more robust than the Gauss–Newton Algorithm and exhibits linear (or even superlinear) convergence in practice. The use of LMA for inverse problems is discussed in [16].

The ExKI procedure solves the optimization problem for (2b) by a different preconditioned gradient descent procedure, defined by the update

$$\left(d\mathcal{G}^T(\theta_n)\Sigma_\nu^{-1}d\mathcal{G}(\theta_n) + (C_n + \Sigma_\omega)^{-1}\right)\delta\theta_n = d\mathcal{G}^T(\theta_n)\Sigma_\nu^{-1}\left(y - \mathcal{G}(\theta_n)\right). \tag{33}$$

This may be viewed as a generalization of the LMA in which the adaptive damping term is now a matrix $C_n + \Sigma_\omega$ and the adaptation is automated through the covariance updates; furthermore this matrix is lower bounded (in the sense of quadratic forms) by Σ_ω, regardless of the adaptation through the covariance, ensuring some damping of the Gauss-Newton approximate Hessian. We may expect that the UKI and EKI, which approximate the linearization $d\mathcal{G}$ in the ExKI to benefit from this generalized LMA. Connections between the LMA and EKI were first systematically explored in [15] and more recently in [68].

3.3. UKI: Unscented Approximation and Averaging

Here we explain that the unscented transform may be viewed as smoothing the energy landscape of UKI, in comparison with ExKI; this helps to explain the improved behaviour of UKI over ExKI on rough landscapes, such as those we will show in what follows when performing parameter estimation for chaotic differential equations.

Comparing with the ExKI, the UKI further smooths the gradient and the landscape of the nonlinear inverse function $\mathcal{G}(\theta)$. We first introduce a useful averaging property. [5]

Lemma 1. *Let θ denote Gaussian random vector $\theta \sim \mathcal{N}(m, C) \in \mathbb{R}^{N_\theta}$. For any nonlinear function $\mathcal{G} : \mathbb{R}^{N_\theta} \to \mathbb{R}^{N_y}$, we define the associated averaged function $\mathcal{FG} : \mathbb{R}^{N_\theta} \times \mathbb{R}^{N_\theta \times N_\theta}_{\succeq 0} \to \mathbb{R}^{N_y}$ and averaged gradient function $\mathcal{F}d\mathcal{G} : \mathbb{R}^{N_\theta} \times \mathbb{R}^{N_\theta \times N_\theta}_{\succeq 0} \to \mathbb{R}^{N_y \times N_\theta}$ as follows:*

$$\mathcal{FG}(m, C) := \mathbb{E}[\mathcal{G}(\theta)] \qquad \mathcal{F}d\mathcal{G}(m, C) := \mathrm{Cov}[\mathcal{G}(\theta), \theta] \cdot C^{-1}. \tag{34}$$

Then we have $\dfrac{\partial \mathcal{FG}(m, C)}{\partial m} = \mathcal{F}d\mathcal{G}(m, C)$.

Proof. The proof is in Appendix A. □

Note that in the linear case $\mathcal{FG}(m, C) = \mathcal{G}(m)$ and $\mathcal{F}d\mathcal{G}(m, C) = \mathcal{G}$; the averaged derivative is exact. This averaging procedure is useful to understand the conceptual GAA precisely because (34) may be used to express $\mathrm{Cov}[\mathcal{G}(\theta), \theta]$, which appears in the conceptual GAA, in terms of the averaged derivative $\mathcal{F}d\mathcal{G}(m, C)$. In order to use this idea in the context of the UKI it is useful to understand related averaging operations when the modified unscented transform is employed to approximate Gaussian expectations. To this end we define, using (18)–(20),

$$\begin{aligned} \mathcal{F}_u\mathcal{G}_n &:= \widehat{y}_n, \\ \mathcal{F}_u d\mathcal{G}_n &:= \widehat{C}_n^{\theta p T}\widehat{C}_n^{-1}, \end{aligned} \tag{35}$$

noting that $\mathcal{F}_u\mathcal{G}_n$ and $\mathcal{F}_u d\mathcal{G}_n$ then correspond to approximation of (34) at step n of the algorithm, using the modified unscented transform from Definition 1.

Proposition 2. *The UKI algorithm* (18)–(20) *may be written in the following form:*

[5] In what follows, the suffix $\succeq 0$ denotes positive semi-definite matrix and $\frac{\partial}{\partial m}$ denotes gradient with respect to m.

- *Prediction step :*

$$\begin{aligned}\widehat{m}_{n+1} =& r + \alpha(m_n - r),\\ \widehat{C}_{n+1} =& \alpha^2 C_n + \Sigma_\omega.\end{aligned} \tag{36}$$

- *Analysis step :*

$$\begin{aligned}&\widehat{y}_{n+1} = \mathcal{F}_u \mathcal{G}_{n+1},\\ &\widehat{C}^{\theta p}_{n+1} = \widehat{C}_{n+1} \mathcal{F}_u d\mathcal{G}^T_{n+1},\\ &\widehat{C}^{pp}_{n+1} = \mathcal{F}_u d\mathcal{G}_{n+1} \widehat{C}_{n+1} \mathcal{F}_u d\mathcal{G}^T_{n+1} + \Sigma_\nu + \widetilde{\Sigma}_{\nu,n+1},\\ &m_{n+1} = \widehat{m}_{n+1} + \widehat{C}^{\theta p}_{n+1} (\widehat{C}^{pp}_{n+1})^{-1} \big(y - \widehat{y}_{n+1}\big),\\ &C_{n+1} = \widehat{C}_{n+1} - \widehat{C}^{\theta p}_{n+1} (\widehat{C}^{pp}_{n+1})^{-1} \widehat{C}^{\theta p\ T}_{n+1}.\end{aligned} \tag{37}$$

Here $\widetilde{\Sigma}_{\nu,n+1} \succeq 0$. *Furthermore,* $\|\widetilde{\Sigma}_{\nu,n+1}\| = \mathcal{O}(\|\widehat{C}_{n+1}\|^2)$ *and* $\widetilde{\Sigma}_{\nu,n+1} = 0$ *when* $\mathcal{G}$ *is linear.*

Proof. The proof is in Appendix A. □

Remark 8. *Comparison of the original UKI algorithm* (18)–(20) *with its rewritten form* (36)-(37) *demonstrates that, in the regime where the covariance is small, or the forward model is linear, the UKI algorithm behaves like the ExKI algorithm* (14)-(15) *but with the nonlinear function* $\mathcal{G}$ *and its associated gradient* $d\mathcal{G}$ *having been averaged according to unscented approximations of the averaging operations defined in Lemma 1. From the preceding subsection it follows that the UKI is also related to a modified LMA applied to an averaged objective function. Note that, by using the unscented approximation of the averaging procedure defined in Lemma 1, we essentially remove the averaging of* $\mathcal{G}$ *and retain it only on* $d\mathcal{G}$. *Averaging of the gradient* $d\mathcal{G}$ *alone will be demonstrated to have an important positive effect on parameter estimation for chaotic dynamical systems in Subsections 5.8, 5.9 and 5.10.*

3.4. Continuous Time Limit

To derive a continuous time limit we set $\alpha = 1 - \alpha_h h$, $\Sigma_\omega \mapsto h\Sigma_\omega$ and $\Sigma_\nu \mapsto h^{-1}\Sigma_\nu$. The algorithm defined by Equations (8) to (10) then has the form of a first order accurate (in h) approximation of the dynamical system

$$\dot{m} = -\alpha_h(m - r) + C^{\theta p} \Sigma_\nu^{-1} \big(y - \mathbb{E}\mathcal{G}(\theta)\big), \tag{38a}$$

$$\dot{C} = -2\alpha_h C + \Sigma_\omega - C^{\theta p} {\Sigma_\nu}^{-1} C^{\theta p T}, \tag{38b}$$

where $\theta \sim \mathcal{N}(m, C)$, expectation $\mathbb{E}$ is with respect to this distribution and

$$C^{\theta p} = \mathbb{E}\Big(\big(\theta - m\big) \otimes \big(\mathcal{G}(\theta) - \mathbb{E}\mathcal{G}(\theta)\big)\Big).$$

This continuous time dynamical system may be used as the basis for practical algorithms by discretizing in time, for example using forward Euler with an adaptive time-step as in [56], and applying the same ideas used in the ExKF, UKI or EKI to approximate the expectations.

The steady state m_∞, C_∞ of the differential equations (38) are implicitly defined in a somewhat complicated fashion. However, any such steady state always has non-singular covariance as we now state and prove.

Lemma 2. *For any steady state* (m_∞, C_∞) *of equation* (38)*, the steady covariance* C_∞ *is non-singular.*

Proof. The proof is in Appendix A. □

4. Variants on the Basic Algorithm

4.1. Enforcing Constraints

Kalman inversion requires solving forward problems at every iteration. Failure of the forward problem to deliver physically meaningful solutions can lead to failure of the inverse problem. Adding constraints to the parameters (for example, dissipation is non-negative) significantly improves the robustness of Kalman inversion. Within the EKI there is a natural way to impose constraints, using the fact that each iteration of the algorithm may be interpreted as solving a set of coupled quadratic optimization problems, with coupling arising from empirical covariances. These optimization problems are readily appended with convex constraints, such as box (inequality) constraints [77]; see also [15, 18]. The UKI does not have this optimization interpretation and so we adopt a different approach to enforcing box constraints.

In this paper there are occasions where we impose element-wise box constraints of the form

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

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

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

The inverse problem is then reformulated as

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

and the UKI methods and variants are employed with $\mathcal{G} \mapsto \mathcal{G} \circ \varphi$.

4.2. Unscented Kalman Sampler

Consider the following stochastic dynamical system, in which W is a standard unit Brownian motion in $\mathbb{R}^{N_\theta}$:

$$\dot{\theta} = C^{\theta p}\Sigma_\eta^{-1}\big(y - \mathcal{G}(\theta)\big) - C\Sigma_0^{-1}(\theta - r_0) + \sqrt{2C}\dot{W}. \tag{39a}$$

Here

$$C^{\theta p} = \mathbb{E}\Big(\big(\theta - m\big) \otimes \big(\mathcal{G}(\theta) - \mathbb{E}\mathcal{G}(\theta)\big)\Big) \tag{40}$$

and all expectations are computed under the law of θ, with respect to which the mean and covariance are denoted m and C respectively. This Itò-McKean diffusion process is approximated by an interacting particle system, and the law of θ approximated using the resulting empirical Gaussian approximation, leading to the EKS [19].

Here we invoke a different Gaussian approximation to approximate the law of θ, based on the unscented transform. First consider the following evolution equations for the mean and covariance of the Gaussian approximation to the law of θ :

$$\dot{m} = C^{\theta p}\Sigma_\eta^{-1}\big(y - \mathbb{E}\mathcal{G}(\theta)\big) - C\Sigma_0^{-1}(m - r_0), \tag{41a}$$

$$\dot{C} = -2C^{\theta p}\Sigma_\eta^{-1}C^{\theta p^T} - 2C\Sigma_0^{-1}C + 2C. \tag{41b}$$

The matrix $C^{\theta p}$ is again given by (40) but now expectation $\mathbb{E}$ is with respect to the distribution $\mathcal{N}(m, C)$. The UKS is defined by approximating the expectations in this system by use of an unscented transform.

In the case where $\mathcal{G}$ is linear and the solution is initialized at a Gaussian then the system (41) is self-consistent in that $\mathcal{N}(m, C)$ is the distribution of the Itò-McKean diffusion (39) governing θ; furthermore the analysis in [19] show that then the system converges to the posterior distribution (3) and the analysis in [78] shows that, when initialized at a non-Gaussian, the Gaussian dynamics is an attractor. It is thus natural to use numerical simulations of (41) to generate approximate samples from the posterior distribution. Illustrative examples are presented in Appendix B.

5. Numerical Results

In this section we present numerical results for Kalman-based inversion using the proposed stochastic dynamical system equation (6).

5.1. Choice of Hyperparameters

We make choices of Σ_ω, Σ_ν and r guided by the discussion in Remark 2. However, for general nonlinear problems C_* is not explicitly defined. Thus we modify the prescription given in (30) and instead choose $r = r_0$ the desired prior mean and

$$\Sigma_\nu = 2\Sigma_\eta \tag{42a}$$

$$\Sigma_\omega = \big(2 - \alpha^2\big)\gamma\mathbb{I} \tag{42b}$$

for some $\gamma > 0$. When the observational noise is absent or negligible, for over-determined problems, we take $\alpha = 1$. To avoid overfitting in the presence of noise, for under-determined problems, we choose $\alpha \in (0, 1)$. In general, cross-validation should be invoked to determine an optimal choice of α. In this paper, we have simply used the values $0.0, 0.5, 0.9$ for illustrative purposes. To be concrete we initialize with $m_0 = r_0$ and $C_0 = \gamma\mathbb{I}$. Specific choices of r_0 and γ will differ between examples and will be spelled out in each example.

5.2. Classes of Problems Studied

For all applications, we focus mainly on the UKI; some comparisons between the UKI and EKI are also presented and computational difficulties inherent in the rough misfit landscape experienced by ExKI for chaotic dynamical systems are demonstrably overcome by deploying the UKI. The applications cover a wide range of problems. They include three categories:

1. Noiseless linear problems, where over-determined, under-determined and well-determined systems are considered.
   - Linear 2-parameter model problem: this problem serves as a proof-of-concept example, which demonstrates the convergence of the mean and the covariance matrix discussed in Subsection 3.1.
   - Hilbert matrix problem: this problem demonstrates the ability of the UKI to solve ill-conditioned inverse problems. It also serves as an example where the EKI suffers from divergence as it is iterated, but UKI behaves well.
2. Noisy field recovery problems, in which we add 0%, 1% and 5% Gaussian random noise to the observation, as follows:

$$y_{obs} = y_{ref} + \epsilon \odot \xi, \quad \xi \sim \mathcal{N}(0, \mathbb{I}), \tag{43}$$

   where $y_{ref} = \mathcal{G}(\theta_{ref})$, $\epsilon = 0\%y_{ref}, 1\%y_{ref},$ and $5\%y_{ref}$, and $\odot$ denotes element-wise multiplication. It is important to distinguish between the added Gaussian random noise appearing in

the data and the observation error model $\eta \sim \mathcal{N}(0, \Sigma_\eta)$ used in the development of the inversion algorithm; in essence we assume imperfect knowledge of the noise model.[6] Comparison of UKI and EKI is presented. EKI is shown to suffer from finite ensemble size effects, and in some cases diverges; in contrast, UKI behaves well. This category of inversion for fields also serves to demonstrate the value of the Tikhonov regularization parameter $\alpha \in (0, 1)$ in the prevention of overfitting. We consider three examples, now listed.

- Darcy flow problem: to find permeability parameters in subsurface flow from measurements of pressure (or piezometric head).
- Damage detection problem: determining the damage field in an elastic body from displacement observations on the surface of the structure.
- Navier-Stokes problem: we study a two dimensional incompressible fluid, using the vorticity-streamfunction formulation, and recover the initial vorticity from noisy observations of the vorticity field at later times.

3. Chaotic problems, in which the parameters are learned from the time-averaged statistics. For these problems, we demonstrate that choosing $\alpha = 1$ is satisfactory, relying on the implicit regularization inherent in the approximate LMA interpretation of ExKI and UKI, as discussed in Subsection 3.2. The three examples considered are now listed.
   - Lorenz63 model problem: we present a discussion of why adjoint based methods including ExKI, fail; we then demonstrate that the UKI succeeds. We attribute the success of the UKI to the averaging effect induced by the unscented transform and discussed in Subsection 3.3.
   - Multiscale Lorenz96 problem: we study a scale-separated setting, in which the closure for the fast dynamics is learned from time-averaged statistics.
   - Idealized general circulation model problem: this is a 3D Navier-Stokes problem with a hydrostatic assumption, and simple parameterized subgrid-scale model; we learn the parameters of the subgrid-scale model from time-averaged data. This problem demonstrates the potential of applying the UKI for large scale chaotic inverse problems.

5.3. Linear 2-Parameter Model Problem

Consider the 2-parameter linear inverse problem

$$y = G\theta + \eta.$$

We explore the following three scenarios

- non-singular (well-determined) system (NS)

$$y = \begin{bmatrix} 3 \\ 7 \end{bmatrix} \quad G = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad \theta_{ref} = \begin{bmatrix} 1 \\ 1 \end{bmatrix};$$

- over-determined system (OD)

$$y = \begin{bmatrix} 3 \\ 7 \\ 10 \end{bmatrix} \quad G = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \quad \theta_{ref} = \begin{bmatrix} 1/3 \\ 17/12 \end{bmatrix};$$

[6] See section 7.1 of [79] for an example with a similar set-up; see also discussion around equation (55) in [80] where the additive Gaussian noise used in the data is carefully constructed to scale relative to the truth underlying it.

- under-determined system (UD)

$$y = \begin{bmatrix} 3 \end{bmatrix} \quad G = \begin{bmatrix} 1 & 2 \end{bmatrix} \quad \theta_{ref} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \ c \in \mathbb{R}.$$

We define

$$\theta_{ref} = \arg\min_{\theta} \frac{1}{2} \|\Sigma_\eta^{-\frac{1}{2}}(y - G\theta)\|^2,$$

and note that, in the UD case, θ_{ref} comprises a one-parameter family of possible solutions. We also note that $y = G\theta_{ref}$ for NS; and $y = G\theta_{ref}(c^\dagger)$ for UD, with $c^\dagger = 0$; but for OD $y \neq G\theta_{ref}$.

We choose $r_0 = 0$ and $\gamma = 0.5^2$ and hence also initialize the UKI at $\theta_0 \sim \mathcal{N}(0, 0.5^2\mathbb{I})$. We choose the observation error noise to be $\eta \sim \mathcal{N}(0, 0.1^2\mathbb{I})$ although, note, there is no noise in the data we use; we thus set $\alpha = 1$. The convergence of the parameter vector m_n is depicted in Fig. 1. In all scenarios, the estimated parameter vectors converge to the desired values exponentially fast. For UD, θ_{ref} is not unique and the converged mean vector m_∞ depends on the initial conditions for the algorithm. For this specific case[7], $\{m_n\}$ converges to $\theta_{ref} = [0.6 \quad 1.2]^T$, and the corresponding c is -0.2; this is not equal to the true $c^\dagger = 0$ as the data contains no information about it. The convergence of the covariance matrices $\{C_n\}$ to C_∞ is depicted in Fig. 2, with NS and OD on the left and UD on the right. In the cases NS and OD, the estimated covariance matrices converge to the desired values (the steady state of equation (26)). In the case UD, the covariance matrices $\{C_n\}$ diverge to $+\infty$ (see Remark 4); nonetheless, this divergence of the covariance matrix does not affect the convergence of the parameter vector. Discussion concerning the use of UKI with $\alpha \in (0, 1)$ for the under-determined system is presented in Appendix C, illustrating the benefit of regularization in this setting.

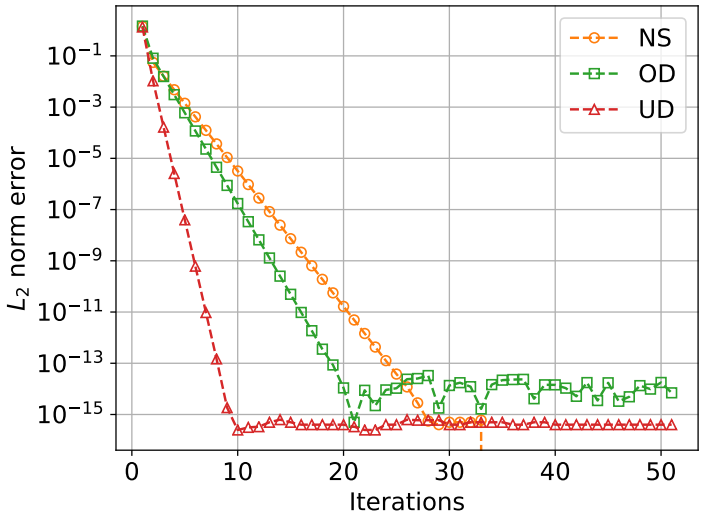


Figure 1: L_2 error $\|m_n - \theta_{ref}\|_2$ of the linear 2-parameter model problem. NS: non-singular system, OD: over-determined system, UD: under-determined system.

[7]Since $\theta_0 \sim \mathcal{N}(m_0, 0.5^2\mathbb{I})$, and since $m_n - m_0 \in \text{Range}(G^T)$, we deduce that $c = \frac{[2 \quad -1] \cdot m_0 - 1}{5}$.

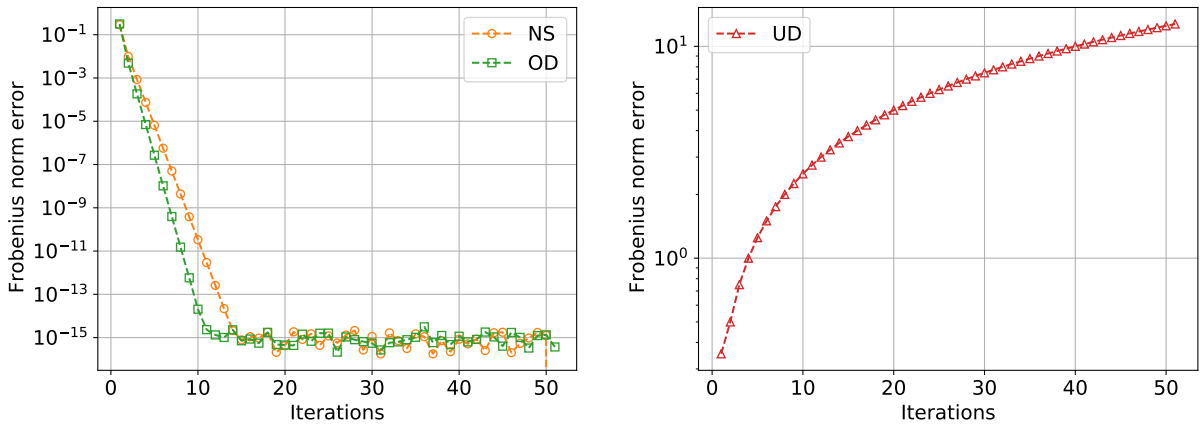


Figure 2: Frobenius norm $\|C_n - C_\infty\|_F$ (left) for non-singular (NS) and over-determined (OD) systems, and $\|C_n\|_F$ (right) for the under-determined (UD) system of the linear 2-parameter model problem.

5.4. Hilbert Matrix Problem

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

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

The condition number of G grows as $\mathcal{O}\left((1+\sqrt{2})^{4N_\theta}/\sqrt{N_\theta}\right)$ [81]. We consider the inverse problem

$$y = G\theta + \eta.$$

We use a true solution $\theta_{ref} = \mathbb{1}$ and data $y = G\mathbb{1}$. The conditioning of G makes determination of θ from y difficult. Traditional linear solvers fail for such a problem.[8]

We consider two scenarios: $N_\theta = 10$ and $N_\theta = 100$. Both UKI and EKI are applied, and we set $r_0 = 0$ and $\gamma = 0.5^2$. Thus $\theta_0 \sim \mathcal{N}(0, 0.5^2\mathbb{I})$. In the inversion algorithm, we take the observation error to be $\eta \sim \mathcal{N}(0, 0.1^2\mathbb{I})$; but the data itself contains no noise, and we again set $\alpha = 1$. For the EKI, the sample sizes are set to $J = 2N_\theta + 1$ and $J = 100N_\theta + 1$. The convergence of the parameter vector m_n is depicted in Fig. 3. The UKI converges, but the convergence rate depends on the condition number of G, slowing as it grows. The EKI converges to certain accuracy as fast as the UKI and then diverges. This demonstrates the importance of early stopping of the EKI, as shown in [14], or the use of adaptive regularization as in [15, 17]. Furthermore, it demonstrates that the UKI does not suffer from these issues and automatically handles the ill-posedness of the inverse problem.

5.5. Darcy Flow Problem

Consider the Darcy flow equation on the two-dimensional spatial domain $D = [0,1]^2$, which describes the pressure field $p(x)$ in a porous medium defined by a positive permeability field $a(x,\theta)$:

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

[8] $G\backslash y$ in Julia leads to an L_2 error of 4250.142 for $N_\theta = 100$.

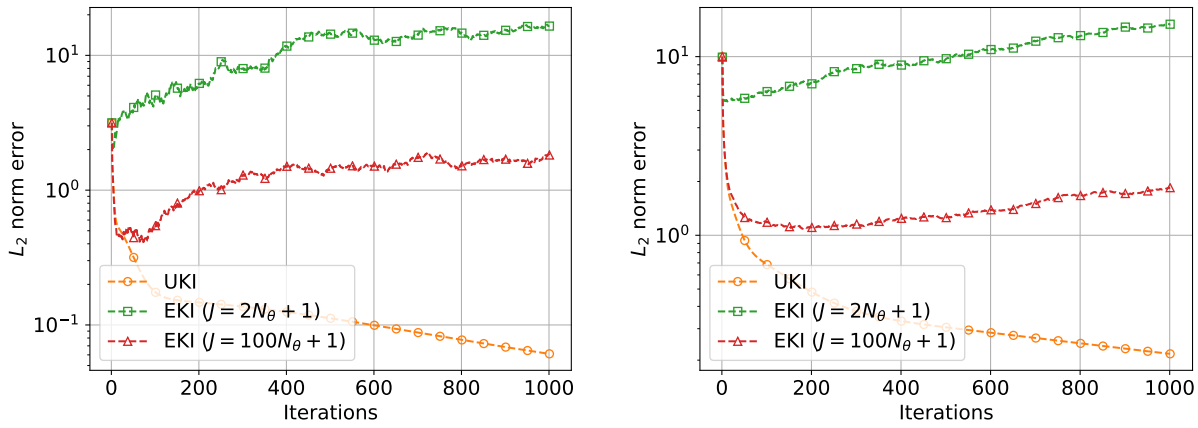


Figure 3: L_2 error $\|m_n - \theta_{ref}\|_2$ of the Hilbert inverse problem with $N_\theta = 10$ (left) and $N_\theta = 100$ (right).

For simplicity, Dirichlet boundary conditions on the pressure are applied on ∂D. The fluid source field f is defined as

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

We place a prior on the permeability field $a(x, \theta)$ by assuming that $\log a(x, \theta)$ is a centred Gaussian with covariance

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

here $-\Delta$ denotes the Laplacian on D subject to homogeneous Neumann boundary conditions on the space of spatial-mean zero functions, $\tau > 0$ denotes the inverse length scale of the random field and $d > 0$ determines its regularity ($\tau = 3$ and $d = 2$ in the present study). See [18, 82, 19, 83] for examples. The parameter θ represents the countable set of coefficients in the Karhunen-Loève (KL) expansion of the Gaussian random field:

$$\log a(x, \theta) = \sum_{l \in K} \theta_{(l)} \sqrt{\lambda_l} \psi_l(x), \tag{44}$$

where $K = \mathbb{Z}^{0+} \times \mathbb{Z}^{0+} \setminus \{0, 0\}$, and the eigenpairs are of the form

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

and $\theta_{(l)} \sim \mathcal{N}(0, 1)$ i.i.d. The KL expansion equation (44) can be rewritten as a sum over $\mathbb{Z}^{0+}$ rather than a lattice:

$$\log a(x, \theta) = \sum_{k \in \mathbb{Z}^{0+}} \theta_{(k)} \sqrt{\lambda_k} \psi_k(x), \tag{45}$$

where the eigenvalues λ_k are in descending order. In practice, we truncate this sum 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.

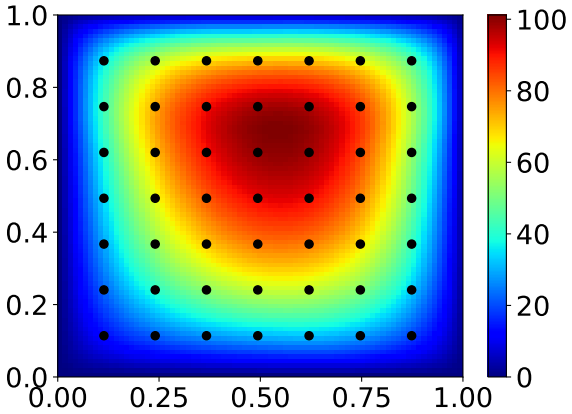


Figure 4: The pressure field of the Darcy flow problem and the 49 equidistant pointwise measurements (black dots).

For the inverse problem, the observation y_{ref} consists of pointwise measurements of the pressure value $p(x)$ at 49 equidistant points in the domain (See Fig. 4). We generate a truth random field $\log a_{ref}(x)$ with $\theta \sim \mathcal{N}(0, \mathbb{I})$ in $\mathbb{R}^{256}$ (i.e. we use the first 256 KL modes) to construct the observation y_{ref}; different levels of noise are added to make data y_{obs} as explained in (43). Using this data, we consider two incomplete parameterization scenarios: solving for the first 32 KL modes ($N_\theta = 32$) and for the first 8 KL modes ($N_\theta = 8$). EKI and UKI are both applied. We take $r_0 = 0$ and $\gamma = 1$ so that $\theta_0 \sim \mathcal{N}(0, \mathbb{I})$. The observation error satisfies $\eta \sim \mathcal{N}(0, \mathbb{I})$. For the EKI, the ensemble size is set to be $J = 100$, which is larger than the number of $\sigma-$points used in UKI ($2N_\theta + 1$).

For the $N_\theta = 32$ case, the convergence of the log-permeability fields $\log a(x, m_n)$ and the optimization errors (2) at each iteration for different noise levels are depicted in Fig. 5; the top row shows the relative L_2 errors in the estimate of $\log a$ and the bottom row shows the optimization errors (data-misfit), left to right corresponds to different noise levels in the data. Without explicit regularization ($\alpha = 1.0$), both UKI and EKI suffer from overfitting for noisy scenarios: the optimization errors keep decreasing, but the parameter errors show the "U-shape" characteristic of overfitting. Adding regularization ($\alpha = 0.5$) relieves the overfitting. The estimated log-permeability fields $\log a(x, m_n)$ at the 50th iteration and the truth random field are depicted in Fig. 6. Both UKI and EKI deliver similar results and these estimated log-permeability fields capture main features of the truth random field.

For the $N_\theta = 8$ case, the convergence of the log-permeability fields $\log a(x, m_n)$ and the optimization errors at each iteration for different noise levels are depicted in Fig. 7. Even without explicit regularization ($\alpha = 1.0$), none of these Kalman inversions suffer from overfitting. Both UKI and EKI lead to similar parameter errors and optimization errors. The estimated log-permeability fields $\log a(x, m_n)$ at the 50th iteration for different noise levels, obtained by the UKI and the truth random field, are depicted in Fig. 8. Comparing with the $N_\theta = 32$ case, all Kalman inversions with $N_\theta = 8$ perform better for the 5% noise scenario. This indicates the possibility of regularizing the inverse problem by reducing the parameter dimensionality.

Finally we observe the smoothness, as a function of the iteration number, of the UKI in comparison to EKI. This may be seen in all the experiments, undertaken in the Darcy flow example.

5.6. Damage Detection Problem

Consider a thin linear elastic arch-like plate, which is fixed on the bottom edges. Tension traction is applied on the top edge, and the distributed load is $p = (2, -20)$. The equations of

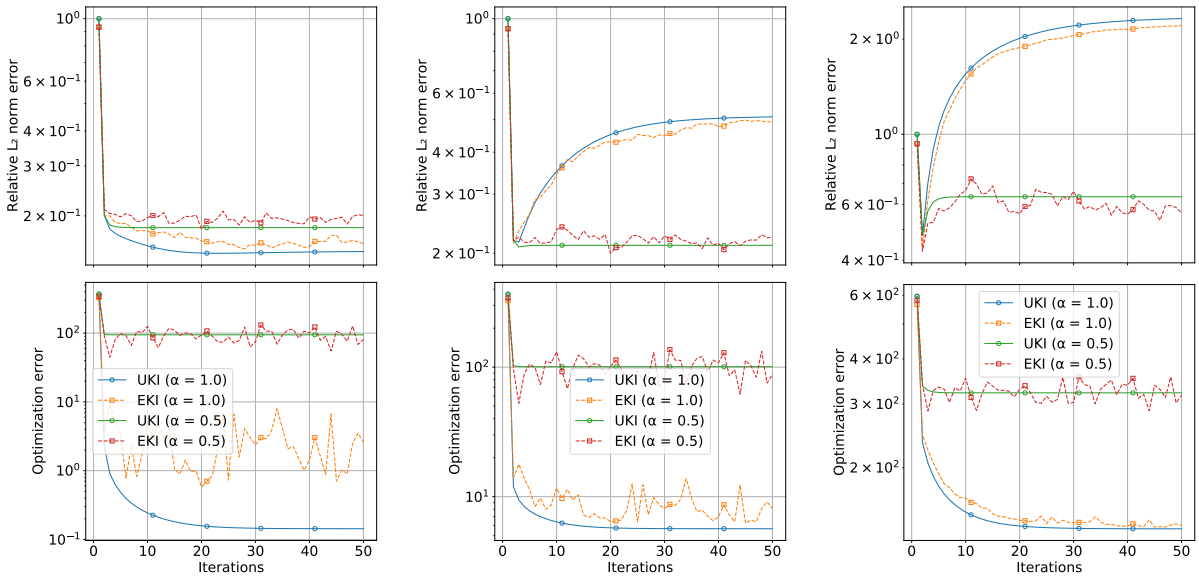


Figure 5: Relative error $\frac{\|\log a(x,m_n) - \log a_{ref}(x)\|_2}{\|\log a_{ref}(x)\|_2}$ (top) and the optimization error $\frac{1}{2}\|\Sigma_\eta^{-\frac{1}{2}}(y_{obs} - \widehat{y}_n)\|^2$ (bottom) of the Darcy problem ($N_\theta = 32$) with different noise levels: noiseless (left), 1% error (middle), and 5% error (right).

linear elastostatics with plane stress assumptions are expressed in terms of the (Cauchy) stress tensor σ and take the form

$$\begin{aligned} \nabla\sigma + b &= 0 \text{ in } \Omega, \\ u &= \bar{u} \text{ on } \Gamma_u, \\ \sigma \cdot n &= \bar{t} \text{ on } \Gamma_t. \end{aligned}$$

Here u is the displacement vector, $b = 0$ is the body force vector, $\Omega \in \mathbb{R}^2$ is the bounded domain occupied by the plate (see Fig. 9) and Γ_u and Γ_t are the displacement boundary and traction boundary respectively; thus $\Gamma_u \bigcup \Gamma_t = \partial\Omega$. The strain tensor is

$$\varepsilon_{mn} = \frac{1}{2}\Big(\frac{\partial u_n}{\partial x_m} + \frac{\partial u_m}{\partial x_n}\Big). \tag{46}$$

The linear constitutive relation between strain and stress is written as

$$\sigma_{ij} = \mathsf{C}_{ijmn}(E,\nu)\varepsilon_{mn}, \tag{47}$$

here C_{ijmn} are the constitutive tensor components, which depends on the Young's modulus E and Poisson's ratio ν; throughout this study, we fix $\nu = 0.4$ and focus on learning spatially-dependent damage information present in E. The damage is assumed to be isotropic elasticity-based damage with

$$E(x,\theta) = \big(1 - \omega(x,\theta)\big)E_0.$$

Throughout this study, we fix $E_0 = 1000$, and $\omega(x,\theta)$ is the scalar-valued damage variable, which varies between zero (no damage) to one (complete damage). The truth damage field (See Fig. 9-left)

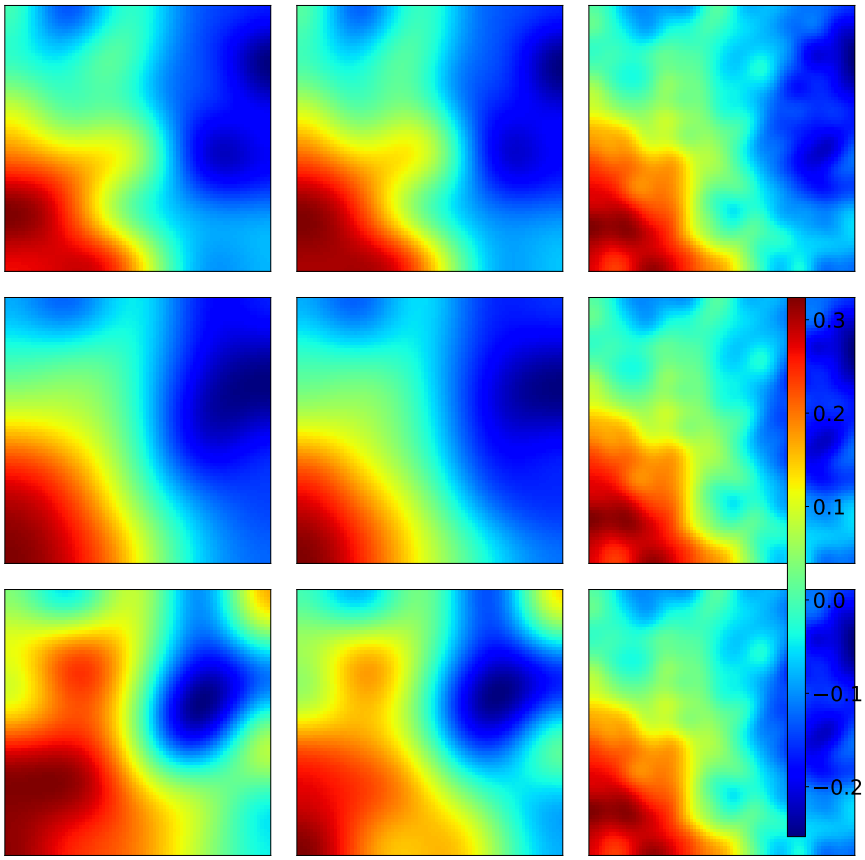


Figure 6: Log-permeability fields $\log a(x, m_n)$ with $N_\theta = 32$ obtained by UKI, EKI, and the truth (left to right) for different noise levels: noiseless $\alpha = 1$ (top), 1% noise $\alpha = 0.5$ (middle), 5% noise $\alpha = 0.5$ (bottom).

is

$$\omega_{ref}(x) = a_1 e^{-\frac{1}{2}(x-x_1)\Sigma_1^{-1}(x-x_1)} + a_2 e^{-\frac{1}{2}(x-x_2)\Sigma_2^{-1}(x-x_2)} + a_3 e^{-\frac{1}{2}(x-x_3)\Sigma_3^{-1}(x-x_3)},$$
$$a_1 = 0.8,\ a_2 = 0.6,\ a_3 = 0.5,$$
$$x_1 = \begin{bmatrix}50\\50\end{bmatrix},\ x_2 = \begin{bmatrix}250\\160\end{bmatrix},\ x_3 = \begin{bmatrix}380\\100\end{bmatrix},\ \Sigma_1 = \begin{bmatrix}200 & 0\\0 & 200\end{bmatrix},\ \Sigma_2 = \begin{bmatrix}800 & 0\\0 & 400\end{bmatrix},\ \Sigma_3 = \begin{bmatrix}100 & 0\\0 & 400\end{bmatrix},$$

and may be seen to exhibit three flaws. Noise is added to the observations on the boundary as in (43). The forward equation is solved by the finite element method with 384 quadratic quadrilateral elements (1649 nodes) using the `NNFEM` library [30, 31].

For the inverse problem, the damage field is parameterized as follows

$$\omega(\theta(x)) = 0.9\frac{1 - e^{-\theta(x)}}{1 + 9e^{-\theta(x)}} \in (-0.1, 0.9),$$

here $\theta(x)$ field is discretized and represented by 24 quadratic quadrilateral elements ($N_\theta = 125$)[9]. The observations are x_1 and x_2 displacements measured at 46 ($N_y = 92$) locations on the surface

[9]It is worth mentioning that increasing the parameter dimensionality by refining the parameter mesh exacerbates the ill-posedness and, therefore, deteriorates the performance of both Kalman inversions.

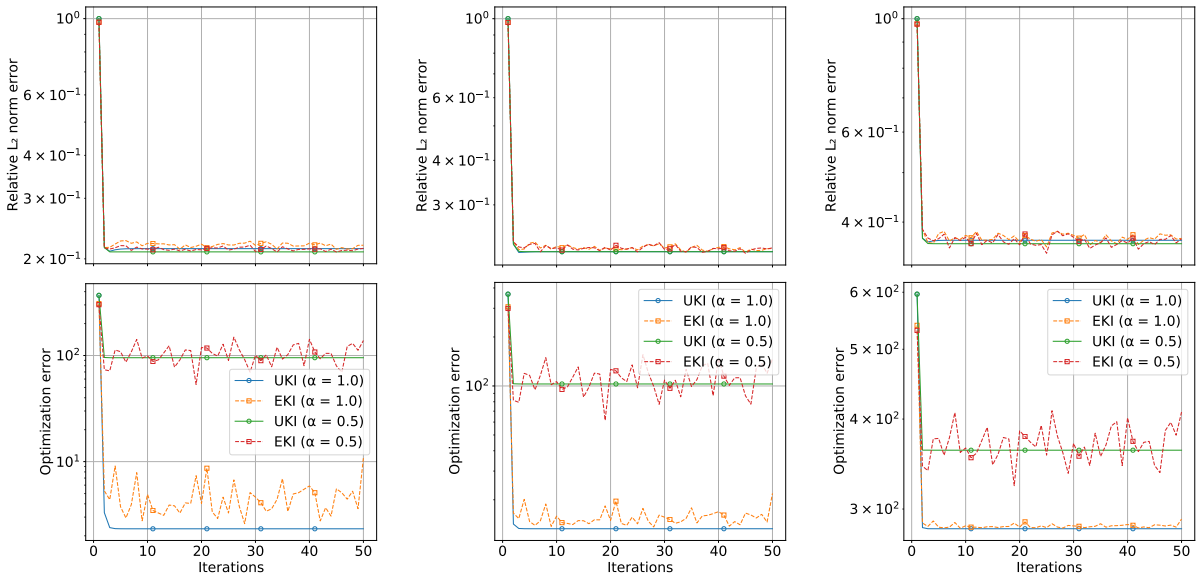


Figure 7: Relative error $\frac{\|\log a(x,m_n) - \log a_{ref}(x)\|_2}{\|\log a_{ref}(x)\|_2}$ (top) and the optimization error $\frac{1}{2}\|\Sigma_\eta^{-\frac{1}{2}}(y_{obs} - \widehat{y}_n)\|^2$ (bottom) of the Darcy problem ($N_\theta = 8$) with different noise levels: noiseless (left), 1% error (middle), and 5% error (right).

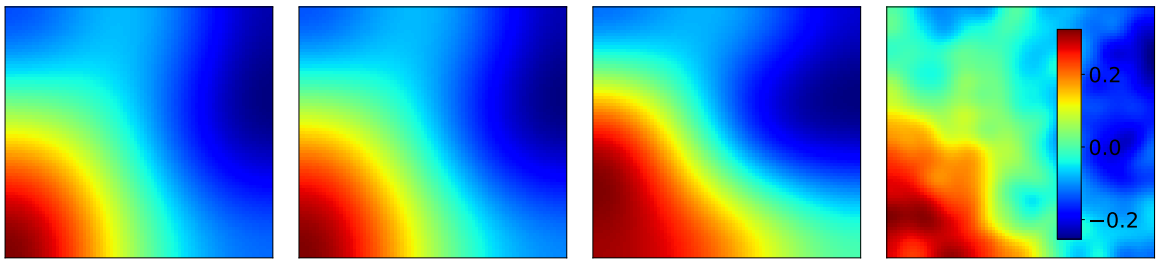


Figure 8: Log-permeability fields $\log a(x, m_n)$ with $N_\theta = 8$ obtained by the UKI and the truth (right) for different noise levels: noiseless $\alpha = 1$ (left), 1% noise $\alpha = 1$ (middle-left), 5% noise $\alpha = 1$ (middle-right).

boundaries (see Fig. 9-right). We set $r_0 = 0$ and $\gamma = 1$ and UKI and EKI are both applied, initialized with $\theta_0 \sim \mathcal{N}(0, \mathbb{I})$. The observation error model used in the algorithm is $\eta \sim \mathcal{N}(0, 0.1^2\mathbb{I})$. For this problem the prior information $\omega(\theta = 0) = 0$ corresponds to an undamaged plate, and is expected to be reasonable for most of the domain; therefore the choice $\alpha = 0.5$ is considered. For the EKI, the ensemble size is set to be $J = 500$, which is larger than the number of $\sigma-$points used in UKI $(2N_\theta + 1)$.

The convergence of the damage field $\omega(\theta(x, m_n))$ and the optimization errors at each iteration are depicted in Fig. 10; the organization of the information is the same as in the Darcy flow example. In the noiseless scenario, the EKI exhibits divergence without regularization ($\alpha = 1.0$) due to the ill-posedness, however, the UKI converges [10]. For noisy scenarios, the effect of overfitting is significant. At 1% noise level, setting $\alpha = 0.5$ eliminates overfitting; however at 5% noise level, setting $\alpha = 0.5$ does not eliminate overfitting. Therefore, the results obtained with $\alpha = 0.0$ are also

[10] We will see the same phenomenon in Subsection 5.7.

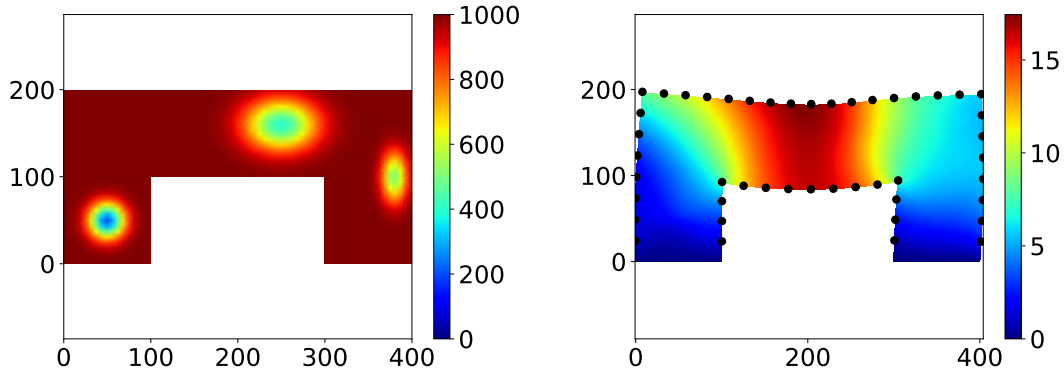


Figure 9: The damaged Young's modulus (left) and the displacement magnitude field (right) with 46 measurement locations on the surface of the boundaries (black dots).

reported for the 5% noise scenario. The estimated damaged Young's modulus fields $E(x, \theta)$ and the truth are depicted in Fig. 11. Both Kalman inversion methods perform comparably, and these three flaw areas are captured; however at 5% noise level noticeable bias is visible in the flaws to the left and right of the domain. As in the Darcy flow case, the convergence histories of the UKI are smoother than for the EKI.

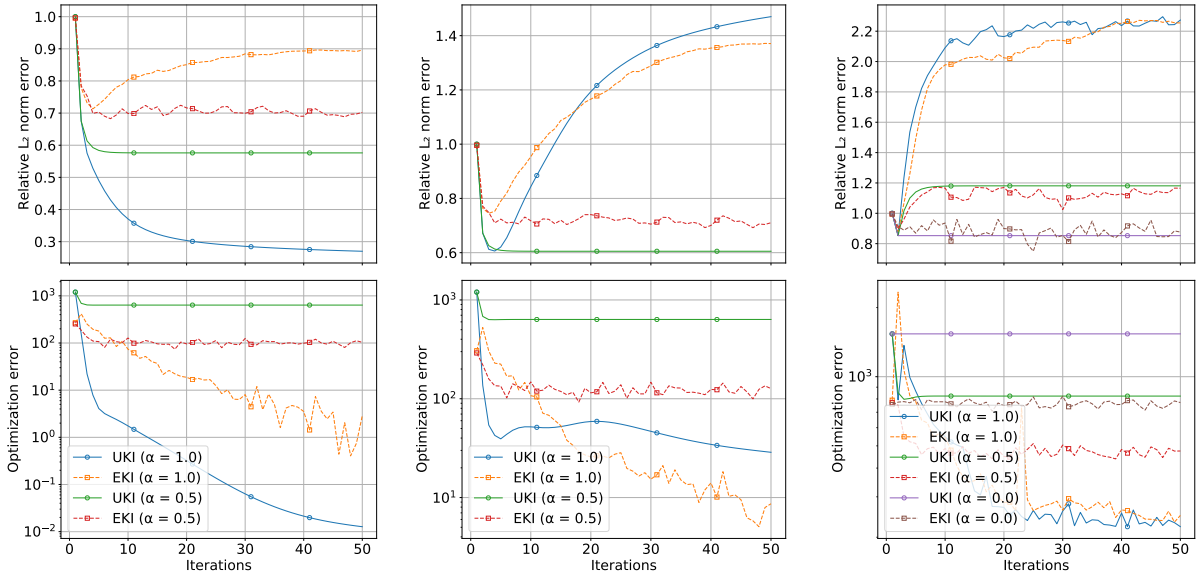


Figure 10: Relative error $\frac{\|\omega(\theta(x, m_n)) - \omega_{ref}\|_2}{\|\omega_{ref}\|_2}$ (top) and the optimization error $\frac{1}{2}\|\Sigma_\eta^{-\frac{1}{2}}(y_{obs} - \widehat{y}_n)\|^2$ (bottom) of the damage detection problem with different noise levels: noiseless (left), 1% error (middle), and 5% error (right).

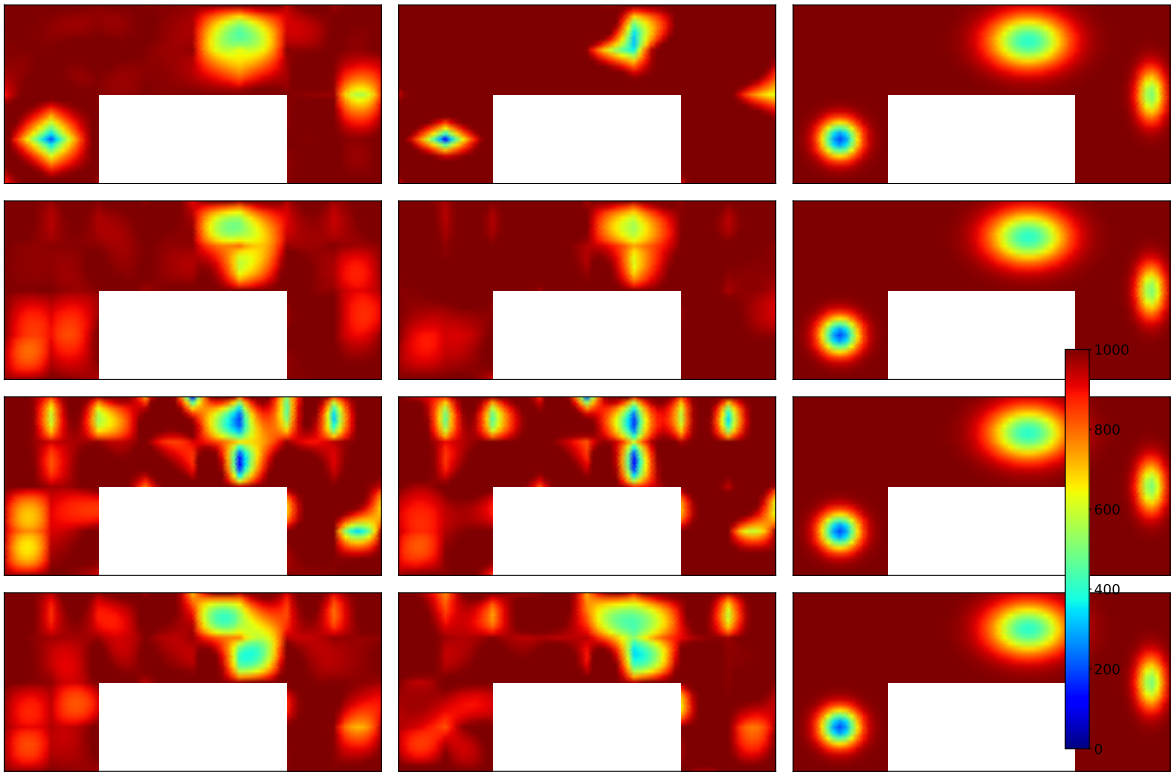


Figure 11: Damaged Young's modulus fields $(1-\omega(x,m_n))E_0$ obtained by UKI, EKI, and the truth (left to right) at different noise levels: noiseless $\alpha = 1$, 1% noise $\alpha = 0.5$, 5% noise $\alpha = 0.5$, and 5% noise $\alpha = 0$ (top to bottom).

5.7. Navier-Stokes Problem

We consider the 2D Navier-Stokes equation on a periodic domain $D = [0, 2\pi] \times [0, 2\pi]$:

$$\begin{aligned}
&\frac{\partial v}{\partial t} + (v \cdot \nabla)v + \nabla p - \nu \Delta v = 0, \\
&\nabla \cdot v = 0, \\
&\frac{1}{4\pi^2} \int v = v_b;
\end{aligned}$$

here v and p denote the velocity vector and the pressure, $\nu = 0.01$ denotes the dynamic viscosity, and $v_b = (2\pi, 2\pi)$ denotes the non-zero mean background velocity. The forward problem is rewritten in the vorticity-stream $(\omega - \psi)$ formulation:

$$\begin{aligned}
&\frac{\partial \omega}{\partial t} - (v \cdot \nabla)\omega - \nu \Delta \omega = 0, \\
&\omega = -\nabla \psi \qquad \frac{1}{4\pi^2} \int \psi = 0, \\
&v = \Big(\frac{\partial \psi}{\partial x_2}, -\frac{\partial \psi}{\partial x_1} \Big) + v_b,
\end{aligned}$$

and solved by the pseudo-spectral method [84] on a 128×128 grid. To eliminate aliasing error, the Orszag 2/3-Rule [85] is applied and, therefore there are 85^2 Fourier modes (padding with zeros). Time-integration is performed using the Crank–Nicolson method with $\Delta T = 2.5 \times 10^{-4}$.

We study the problem of recovering the initial vorticity field from measurements at positive times. We parameterize it as $\omega_0(x,\theta)$, defined by parameters $\theta \in \mathbb{R}^{N_\theta}$, and modeled *a priori* as a Gaussian field with covariance operator $\mathsf{C} = \Delta^{-2}$, subject to periodic boundary conditions, on the space of spatial-mean zero functions. The KL expansion of the initial vorticity field is given by

$$\omega_0(x,\theta) = \sum_{l\in K} \theta^c_{(l)}\sqrt{\lambda_l}\psi^c_l + \theta^s_{(l)}\sqrt{\lambda_l}\psi^s_l, \tag{48}$$

where $K = \{(k_x,k_y)|k_x+k_y>0 \text{ or } (k_x+k_y=0 \text{ and } k_x>0)\}$, and the eigenpairs are of the form

$$\psi^c_l(x) = \frac{\cos(l\cdot x)}{\sqrt{2}\pi} \quad \psi^s_l(x) = \frac{\sin(l\cdot x)}{\sqrt{2}\pi} \quad \lambda_l = \frac{1}{|l|^4},$$

and $\theta^c_{(l)}, \theta^s_{(l)} \sim \mathcal{N}(0, 2\pi^2)$ i.i.d. The KL expansion equation (48) can be rewritten as a sum over $\mathbb{Z}^{0+}$ rather than a lattice:

$$\omega_0(x,\theta) = \sum_{k\in\mathbb{Z}^{0+}} \theta_{(k)}\sqrt{\lambda_k}\psi_k(x), \tag{49}$$

where the eigenvalues λ_k are in descending order.

For the inverse problem, we recover the initial condition, specifically the initial vorticity field of the Navier-Stokes equation, given pointwise observations y_{ref} of the vorticity field at 16 equidistant points ($N_y = 32$) at $T = 0.25$ and $T = 0.5$ (See Fig. 12). The observations y_{obs} are defined as in (43). The initial vorticity field $\omega_{0,ref}$ is generated with all 85^2 Fourier modes, and the first $N_\theta = 100$ KL modes of equation (49) are recovered. We take $r_0 = 0$ and $\gamma = 10$. Both UKI and EKI are applied with $\theta_0 \sim \mathcal{N}(0, 10\mathbb{I})$ and the observation error assumed for inversion purposes is $\eta \sim \mathcal{N}(0,\mathbb{I})$. For the EKI, the ensemble size is set to be $J = 201$, which equals the number of $\sigma-$points in UKI $(2N_\theta+1)$.

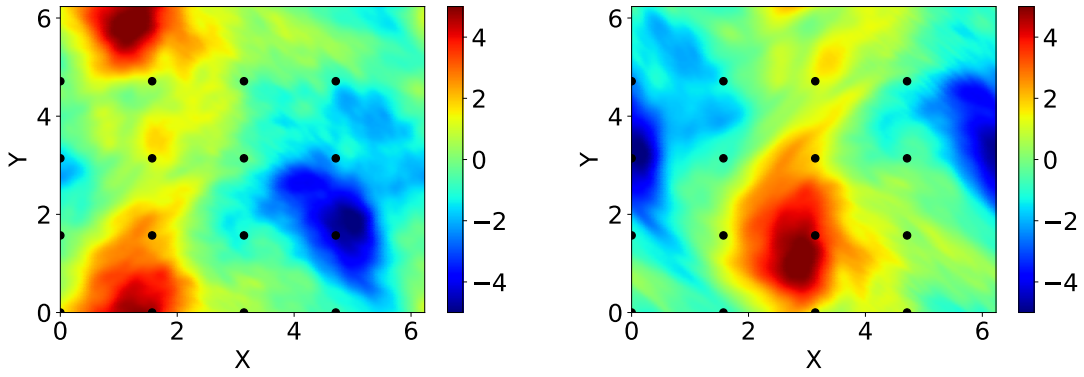


Figure 12: The vorticity fields of the Navier-Stokes problem and the 16 equidistant pointwise measurements (black dots) at two observation times ($T = 0.25$ and $T = 0.5$).

The convergence of the initial vorticity field $\omega_0(x, m_n)$ and the optimization errors for different noise levels at each iteration are depicted in Fig. 13; the organization of the figure is the same as in the Darcy case. In all scenarios, the UKI outperforms EKI. Moreover, without regularization ($\alpha = 1.0$), EKI exhibits slight divergence. We find this inverse problem is not sensitive with the added Gaussian random noises, and the behavior of any Kalman inversion at different noise levels are almost indistinguishable. The estimated initial vorticity fields $\omega_0(x, m_n)$ at the 50th iteration for different noise levels obtained by the Kalman inversions and the truth random field are depicted in Fig. 14. Both Kalman inversions capture main features of the truth random initial field, but not the detailed small features, due to the irreversibility of the diffusion process ($\nu = 0.01$).

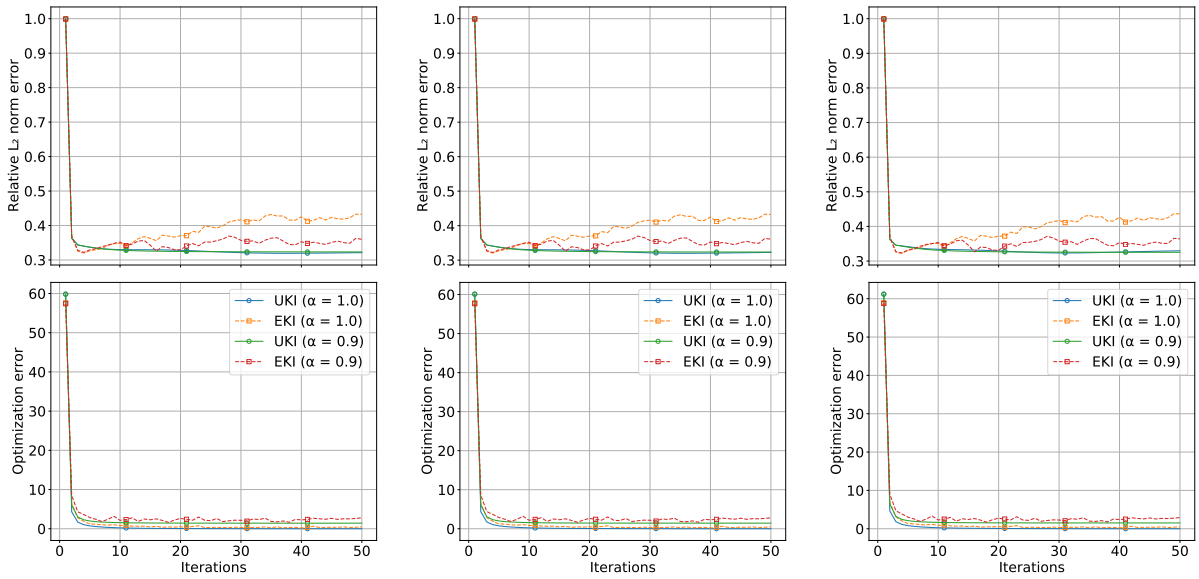


Figure 13: Relative error $\frac{\|\omega_0(x,m_n)-\omega_{0,ref}\|_2}{\|\omega_{0,ref}\|_2}$ (top) and the optimization error $\frac{1}{2}\|\Sigma_\eta^{-\frac{1}{2}}(y_{obs}-\widehat{y}_n)\|^2$ (bottom) of the Navier-Stokes problem with different noise levels: noiseless (left), 1% error (middle), and 5% error (right).

5.8. Lorenz63 Model Problem

Consider the Lorenz63 system, a simplified mathematical model for atmospheric convection [86]:

$$\begin{aligned}\frac{dx_1}{dt} &= \sigma(x_2 - x_1),\\ \frac{dx_2}{dt} &= x_1(r - x_3) - x_2,\\ \frac{dx_3}{dt} &= x_1x_2 - \beta x_3;\end{aligned}$$

the system is parameterized by $\sigma, r, \beta \in \mathbb{R}_+$. The observation consists of the time-average of the various moments over time windows of size $T = 20$, with an initial spin-up period $T = 30$ to eliminate the influence of the initial condition; if $f : \mathbb{R}^3 \mapsto \mathbb{R}$ computes a moment, then we define

$$\overline{f(x)} = \int_{30}^{50} f\big(x(t)\big)dt.$$

The truth observation is computed with parameters $(\sigma,\ r,\ \beta) = (10,\ 28,\ 8/3)$ over a time window of size $T = 200$, also with an initial spin-up period $T = 30$. Since the data is generated on an interval 10 times as long as the observation window, we may appeal to the central limit theorem [87] to argue that the observation error caused by time-averaging using only 20 time units is approximately Gaussian. We split the observation time-series into 10 windows of size $T = 20$ and compute covariance of the observation error η following [53]. We set $r_0 = 5.0\mathbb{1}$ and $\gamma = 1$. The UKI is initialized with $\theta_0 \sim \mathcal{N}(5.0\mathbb{1}, \mathbb{I})$, and α is set to 1.

We start with the following one-parameter inverse problem with fixed $\sigma = 10$ and $\beta = 8/3$:

$$y = \mathcal{G}(r) + \eta \quad \text{with} \quad y = \overline{x_3}. \tag{50}$$

The landscape of $\mathcal{G}$ and sensitivity of $\mathcal{G}(\cdot)$ with respect to the input for observations, derived from chaotic problems such as equation (50), are widely studied [54, 55]. For our specific set-up, we

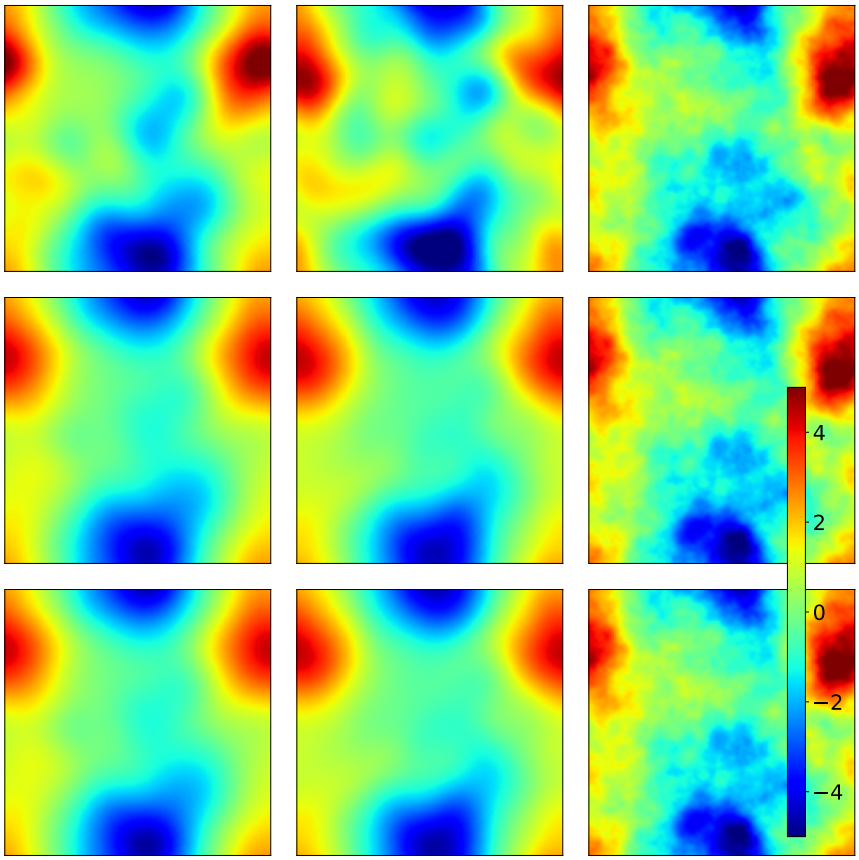


Figure 14: Initial vorticity fields $\omega_0(x, m_n)$ recovered by UKI, EKI, and the truth (left to right) for different noise levels: noiseless $\alpha = 1$, 1% noise $\alpha = 0.9$, 5% noise $\alpha = 0.9$ (top to bottom).

demonstrate this in Fig. 15. The function $\mathcal{G}$ is characterized by a sudden change at $r \approx 22$ and the landscape is highly oscillatory for $r > 22$; furthermore the sensitivity $d\mathcal{G}(r)$ computed with the discrete adjoint method blows up:

$$|d\mathcal{G}(r)| \propto \mathcal{O}(e^{\lambda T}),$$

with value of the exponent λ consistent with the first global Lyapunov exponent [54, 88]. This illustrates the challenges inherent in parameter estimation and sensitivity analyses for chaotic systems. In particular, the ExKI method suffers from the large derivatives of $\mathcal{G}$.

The UKI is applied, and the estimated r and the associated 3-σ confidence intervals at each iteration are depicted in Fig. 16. The confidence intervals give an indication of the evolving covariance C_n. The estimation of r at the 20th iteration is $r \sim \mathcal{N}(28.03, 0.22)$. Based on Lemma 1, it is natural to study the landscape of the averaged function $\mathcal{F}\mathcal{G}$ and its associated gradient $\mathcal{F}d\mathcal{G}$, with the standard deviation $\sigma_r = \sqrt{0.22}$ fixed; this gives an indication of the energy landscape as perceived by the UKI. In particular we have:

$$\mathcal{F}\mathcal{G}(r) = \int \mathcal{G}(x) \frac{1}{\sqrt{2\pi}\sigma_r} e^{-\frac{(x-r)^2}{2\sigma_r^2}} dx, \qquad \mathcal{F}d\mathcal{G}(r) = \frac{\int (x-r)(\mathcal{G}(x) - \mathcal{G}(r)) \frac{1}{\sqrt{2\pi}\sigma_r} e^{-\frac{(x-r)^2}{2\sigma_r^2}} dx}{\int (x-r)^2 \frac{1}{\sqrt{2\pi}\sigma_r} e^{-\frac{(x-r)^2}{2\sigma_r^2}} dx}.$$

These functions are depicted in Fig. 17, which should be compared with Fig. 15. We see that

$\mathcal{F}\mathcal{G}$ is smooth (except the transition point), and $\mathcal{F}d\mathcal{G}$ does not suffer from blow-up in the way $d\mathcal{G}$ does; furthermore $\mathcal{F}d\mathcal{G}$ represents the true gradient $\frac{d\mathcal{G}}{dr} \approx 0.96$ well. This explains why the adjoint/gradient-based methods, including ExKI, fail, but the UKI succeeds for this chaotic inverse problem.

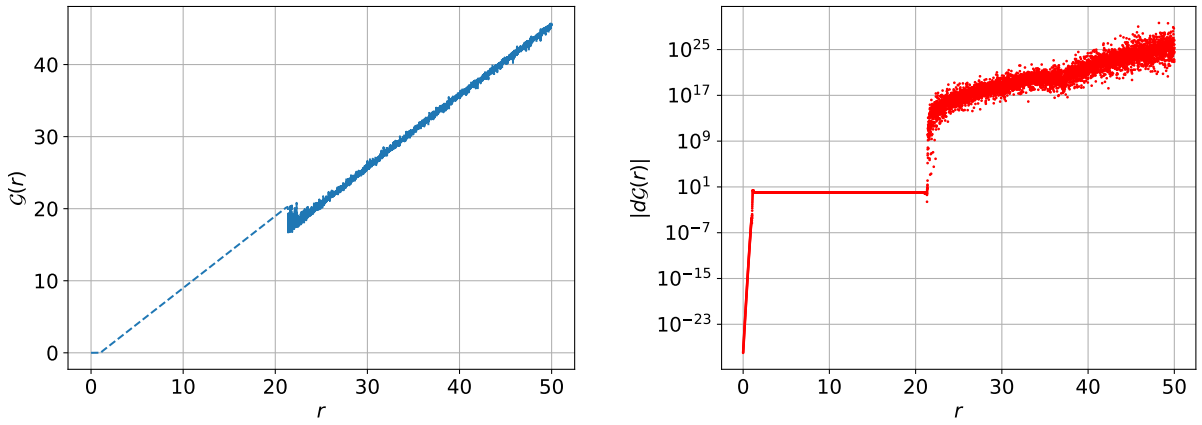


Figure 15: Landscape (left) and sensitivity (right) of $\mathcal{G}$ in the 1-parameter Lorenz63 inverse problem equation (50)

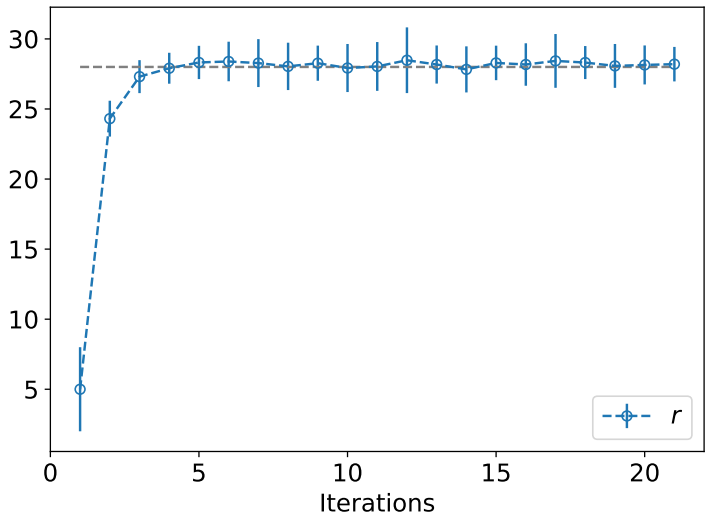


Figure 16: Convergence of the 1-parameter Lorenz63 inverse problem with UKI ($\alpha = 1.0$); the true parameter value is represented by the dashed grey line.

Next, we consider a three-parameter inverse problem, using the ideas in Subsection 4.1. Let $\theta = (\theta_{(1)}, \theta_{(2)}, \theta_{(3)})$ and let $(\sigma, r, \beta) = (|\theta_{(1)}|, |\theta_{(2)}|, |\theta_{(3)}|)$. The map $\mathcal{G}(\theta)$ is found by computing time-averages of all three components of x, as described above, for given input parameter θ. The use of the modulus helps ensure solution trajectories, which do not blow-up. We have

$$y = \mathcal{G}(\theta) + \eta \quad \text{with} \quad y = (\overline{x_1}, \overline{x_2}, \overline{x_3}, \overline{x_2 x_3}, \overline{x_3 x_1}, \overline{x_1 x_2}). \tag{51}$$

All other aspects of the setup are the same as the aforementioned one-parameter inverse problem. The estimated parameters and associated 3-σ confidence intervals for each component at each

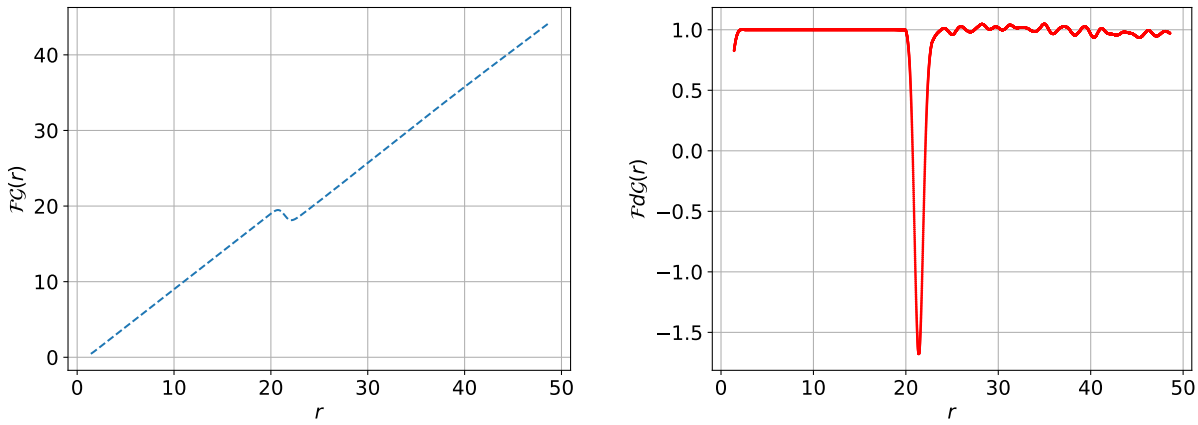


Figure 17: Landscape (left) and sensitivity (right) of $\mathcal{FG}$ in the 1-parameter Lorenz63 inverse problem equation (50) smoothed and viewed by UKI.

iteration are depicted in Fig. 18. The estimation of the parameters at the 20th iteration is

$$\begin{bmatrix}\sigma\\ r\\ \beta\end{bmatrix} \sim \mathcal{N}\Big(\begin{bmatrix}10.28\\ 27.90\\ 2.63\end{bmatrix}, \begin{bmatrix}0.119 & 0.0105 & -0.0037\\ 0.0105 & 0.009 & 0.0015\\ -0.0037 & 0.0015 & 0.0022\end{bmatrix} \Big).$$

For both scenarios, the UKI converges efficiently, thanks to the linear (or superlinear) convergence rate of the LMA and the averaging property. Although the covariance of the iteration does not represent the Bayesian posterior uncertainty, it does indicate the sensitivities inherent in the estimation problem, and in particular that estimation of β involves the largest sensitivities.

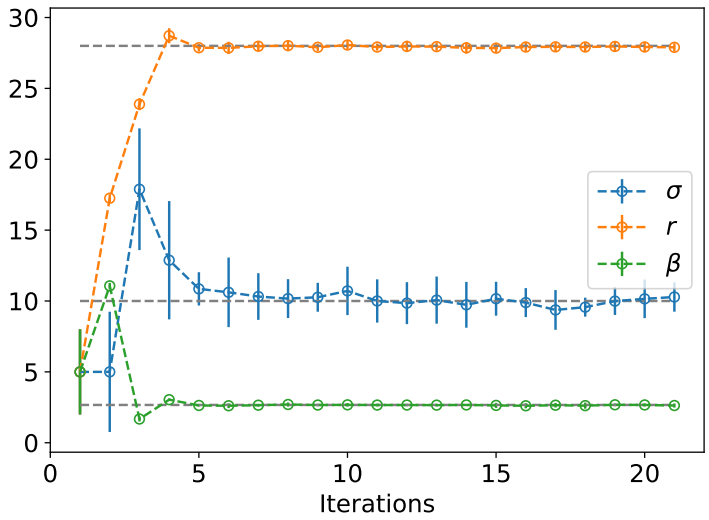


Figure 18: Convergence of the 3-parameter Lorenz63 inverse problem with UKI ($\alpha = 1.0$); true parameter values are represented by dashed grey lines.

5.9. Multiscale Lorenz96 Problem

Consider the multi-scale Lorenz96 system, a simplified mathematical model for the midlatitude atmosphere [89], with K slow variables $X^{(k)}$ which are each coupled with J fast variables $Y^{(j,k)}$, given by:

$$\begin{aligned}\frac{dX^{(k)}}{dt} &= -X^{(k-1)}(X^{(k-2)} - X^{(k+1)}) - X^{(k)} + F - \frac{hc}{b}\sum_{j=1}^{J} Y^{(j,k)},\\ \frac{dY^{(j,k)}}{dt} &= -cbY^{(j+1,k)}(Y^{(j+2,k)} - Y^{(j-1,k)}) - cY^{(j,k)} + \frac{hc}{b}X^{k}.\end{aligned} \tag{52}$$

To close the system, it is appended with the cyclic boundary conditions $X^{(k+K)} = X^{(k)}$, $Y^{(j,k+K)} = Y^{(j,k)}$ and $Y^{(j+J,k)} = Y^{(j,k+1)}$. The time scale separation is parameterized by the coefficient c and the large-scales are subjected to external forcing F. We choose here as parameters $K = 8$, $J = 32$, $F = 20$, $c = b = 10$ and $h = 1$ as in [90, 91, 92, 93]. As time-integrator, we use the 4th-order Runge Kutta method with $\Delta T = 5 \times 10^{-3}$.

Our goal is to learn the closure model $\psi(X)$ of the fast dynamics for a reduced model of the form

$$\frac{dX^{(k)}}{dt} = -X^{(k-1)}(X^{(k-2)} - X^{(k+1)}) - X^{(k)} + F + \psi(X^{(k)}).$$

The closure model $\psi : D \subset \mathbb{R} \mapsto \mathbb{R}$ is parameterized by the finite element method with cubic Hermite polynomials. The domain is set to be $D = [-20, 20]$ and decomposed into 5 elements and, therefore, $N_\theta = 12$.

For the inverse problem, the observations consist of the time-average of the first and second moments of $X^{(1)}, X^{(2)}, X^{(3)}$, and $X^{(4)}$ over a time window of size $T = 1000$ and, therefore $N_y = 14$. The truth observation y_{ref} is computed with the multiscale chaotic system equation (52) with a random initial condition $X^{(k)} \sim \mathcal{N}(0, 1)$ and $Y^{(j,k)} \sim \mathcal{N}(0, 0.01^2)$. And 1%, 2%, and 5% Gaussian random noises are added to the observation following equation (43).

We set $r_0 = 0$ and $\gamma = 1$; the UKI is thus initialized with $\theta_0 \sim \mathcal{N}(0, \mathbb{I})$. The observation error is set to be $\eta = \mathcal{N}(0, \mathrm{diag}\{0.05^2 y_{obs} \odot y_{obs}\})$, and we take $\alpha = 1$, since the system is over-determined. Moreover, these simulations start with another random initialization of $X^{(k)} \sim \mathcal{N}(0, 1)$. The learned closure models at the 20th iteration are reported in Fig. 19. The estimated empirical probability density functions of the slow variables are reported in Fig. 20. For all scenarios, although the learned closure models show non-trivial variability with respect to those published in [91, 92] at the left most extreme of D, the predicted probability density functions match well with the reference, obtained from a full multiscale simulation. It is worth mentioning this problem is not sensitive with respect to the added Gaussian random noise.

5.10. Idealized General Circulation Model

Finally, we consider an idealized general circulation model. The model is based on the 3D Navier-Stokes equations, making the hydrostatic and shallow-atmosphere approximations common in atmospheric modeling. Specifically, we test UKI on the well-known Held-Suarez test case [94], in which a detailed radiative transfer model is replaced by Newtonian relaxation of temperatures toward a prescribed "radiative equilibrium" $T_{\mathrm{eq}}(\phi, p)$ that varies with latitude ϕ and pressure p. Specifically, the thermodynamic equation for temperature T

$$\frac{\partial T}{\partial t} + \cdots = Q$$

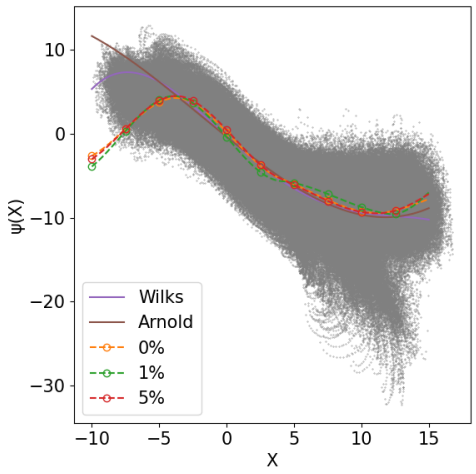


Figure 19: Closure terms $\psi(X)$ for the multi-scale Lorenz96 system obtained from the truth (grey dots) and polynomial data-fitting by Wilks [91] and Arnold [92], compared with what is learned using the UKI approach ($\alpha = 1$) with different noise levels.

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

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

with relaxation coefficient (inverse relaxation time)

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

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

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

is the equilibrium temperature profile ($p_0 = 10^5$ Pa is a reference surface pressure and $\kappa = 2/7$ is the adiabatic exponent). Default parameters 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 produce an Earth-like zonal-mean circulation, albeit without moisture or precipitation. A single jet is generated with maximum strength of roughly 30 m s^{-1} near 45° latitude (Fig. 21).

Our inverse problem is constructed to learn parameters in the Newtonian relaxation term Q:

$$(k_a,\ k_s,\ \Delta T_y,\ \Delta\theta_z).$$

We do so in the presence of the following constraints:

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

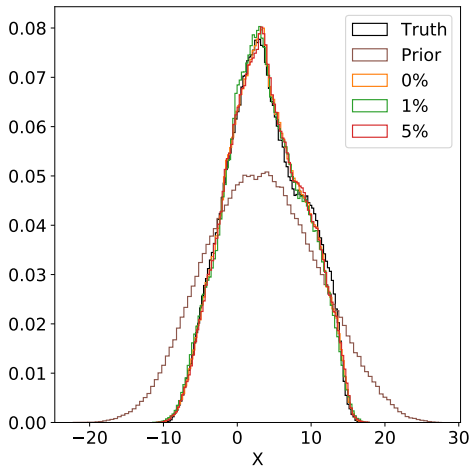


Figure 20: Empirical probability density functions of the slow variables $X^{(k)}$ obtained from the full multi-scale Lorenz96 system (Truth), the initial closure model (Prior), and the closure models learned by the UKI ($\alpha = 1$) at different noise levels.

Conceptually, the setting is identical to that for the Lorenz63 example. We use the same overline notation to denote averaging, which here in addition to the time average in the Lorenz model also includes an zonal average over longitude (because the model is statistically symmetric under rotations around the planet's spin axis). To incorporate the imposition of the constraints, the inverse problem is formulated as follows (see Subsection 4.1 for details):

$$y = \mathcal{G}(\theta) + \eta \quad \text{with} \quad \mathcal{G}(\theta) = \overline{T}(\phi, \sigma) \tag{53}$$

with the parameter transformation

$$\theta : (k_a, k_s, \Delta T_y, \Delta\theta_z) = \Big(\frac{1}{1+|\theta_{(1)}|},\ \frac{1}{1+|\theta_{(1)}|} + \frac{1}{1+|\theta_{(2)}|},\ |\theta_{(3)}|, |\theta_{(4)}|\Big). \tag{54}$$

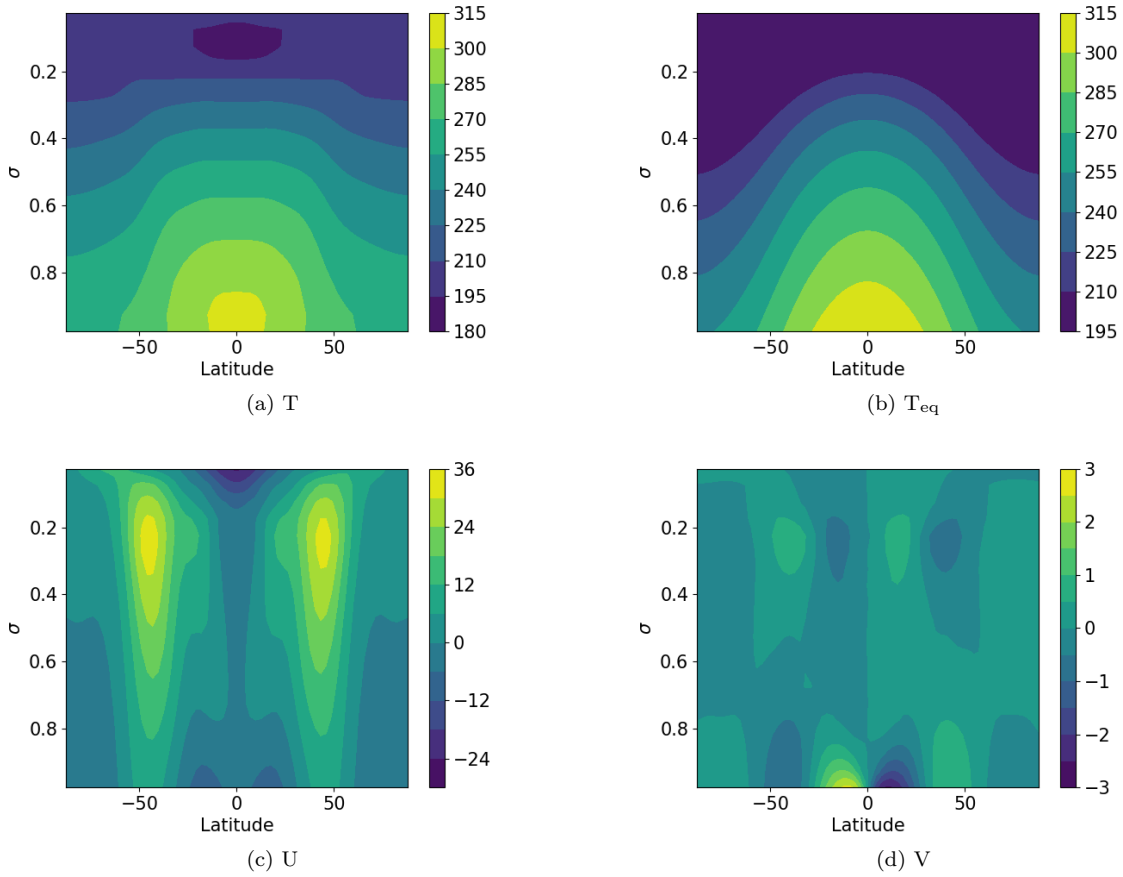
The observation mapping is defined by mapping from the unknown θ to the 200-day zonal mean of the temperature as a function of latitude (ϕ) and height (σ), after an initial spin-up of 200 days. The truth observation is the 1000-day zonal mean of the temperature (see Fig. 21-a), after an initial spin-up 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. As for the Lorenz63 setting, the central limit theorem may be invoked to model the observation error from internal variability.

To perform the inversion, we set $r_0 = [2 \text{ day},\ 2 \text{ day},\ 20 \text{ K},\ 20 \text{ K}]^T$ and $\gamma = 1$. Thus UKI is initialized with $\theta_0 \sim \mathcal{N}\Big(r_0,\ \mathbb{I}\Big)$. Within the algorithm, we assume that the observation error satisfies $\eta \sim \mathcal{N}(0 \text{ K}, 3^2\mathbb{I} \text{ K}^2)$. Because the problem is over-parameterized, we set $\alpha = 1$. The estimated parameters and associated 3-σ confidence intervals for each component at each iteration are depicted in Fig. 22. The estimation of model parameters at the 20th iteration are

$$\begin{bmatrix} k_a \\ k_s \\ \Delta T_y \\ \Delta\theta_z \end{bmatrix} \sim \mathcal{N}\Bigg(\begin{bmatrix} 0.0243 \text{ day}^{-1} \\ 0.243 \text{ day}^{-1} \\ 60.2 \text{ K} \\ 9.91 \text{ K} \end{bmatrix},\ \begin{bmatrix} 2.6\times10^{-7} \text{ day}^{-2} & 4.6\times10^{-6} \text{ day}^{-2} & 1.1\times10^{-5} \text{ day}^{-1}\text{K} & 2.4\times10^{-6} \text{ day}^{-1}\text{K} \\ 4.6\times10^{-6} \text{ day}^{-2} & 5.3\times10^{-4} \text{ day}^{-2} & 5.5\times10^{-3} \text{ day}^{-1}\text{K} & -7.1\times10^{-4} \text{ day}^{-1}\text{K} \\ 1.1\times10^{-5} \text{ day}^{-1}\text{K} & 5.5\times10^{-3} \text{ day}^{-1}\text{K} & 2.3\times10^{-1} \text{ K}^2 & 4.7\times10^{-2} \text{ K}^2 \\ 2.4\times10^{-6} \text{ day}^{-1}\text{K} & -7.1\times10^{-4} \text{ day}^{-1}\text{K} & 4.7\times10^{-2} \text{ K}^2 & 1.8\times10^{-1} \text{ K}^2 \end{bmatrix} \Bigg).$$

The reported covariances of the iteration do not reflect the true Bayesian posterior uncertainty, but they do capture the relative sensitivities in the problem. UKI converges to the true parameters in

fewer than 10 iterations with 9 σ-points, demonstrating the potential of applying UKI for large-scale inverse problems.



(a) T (b) T_{eq}

(c) U (d) V

Figure 21: Zonal mean profile of temperature (a), radiative equilibrium temperature (b), zonal wind velocity (c), and meridional wind velocity (d), all from a 1000-day average. The horizontal coordinate is latitude and the vertical coordinate is the nondimensional σ coordinate of the model.

6. Conclusion

The unscented Kalman inversion is attractive for at least four main reasons: (i) it is black-box and derivative-free; (ii) it is robust for chaotic inverse problems and noisy observations; (iii) it provides sensitivity information; (iv) the method is embarrassingly parallel. It is well-adapted to parameter estimation problems for large complex models given as a black box. Our numerical results demonstrate its theoretical properties and its applicability; in particular it is demonstrated to outperform the EKI on large scale problems in which the number of unknown parameters is small. Because the methodology constitutes a novel approach to parameter estimation, there are many avenues for future research, including applications of the method, methodological improvements and extensions, and theoretical analysis.

Acknowledgments. This work was supported by the generosity of Eric and Wendy Schmidt by recommendation of the Schmidt Futures program and by the National Science Foundation (NSF, award AGS-1835860). A.M.S. was also supported by the Office of Naval Research (award N00014-17-1-2079).

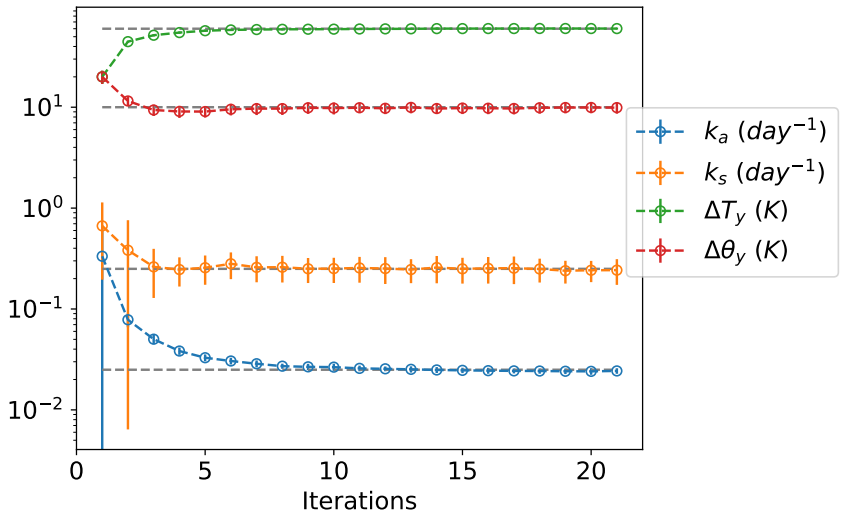


Figure 22: Convergence of the idealized general circulation model inverse problem with UKI ($\alpha = 1.0$). The true parameter values are represented by dashed grey lines.

Appendix A. Proof of Theorems

Proof of Proposition 1. An affine transformation is an invertible mapping from R^{N_θ} to R^{N_θ} of the form ${}^*x = Ax + b$. When we apply the following affine transformation

$$ {}^*m_n = Am_n + b \qquad {}^*C_n = AC_nA^T \quad \text{with} \quad {}^*r = Ar + b \qquad {}^*\Sigma_\omega = A\Sigma_\omega A^T, $$

keep y_n and Σ_ν unchanged, and define ${}^*\mathcal{G}(\theta) = \mathcal{G}\big(A^{-1}(\theta - b)\big)$. We prove

$$ {}^*m_{n+1} = Am_{n+1} + b \qquad {}^*C_{n+1} = AC_{n+1}A^T. \tag{A.1} $$

Equation (8) leads to

$$ {}^*\widehat{m}_{n+1} = {}^*r + \alpha({}^*m_n - {}^*r) = A\widehat{m}_{n+1} + b \qquad {}^*\widehat{C}_{n+1} = \alpha^2 C_n^* + \Sigma_\omega^* = A\widehat{C}_{n+1}A^T. \tag{A.2} $$

Therefore, the distribution of ${}^*\theta_{n+1}|Y_n \sim \mathcal{N}({}^*\widehat{m}_{n+1}{}^*\widehat{C}_{n+1})$ is the same as ${}^*A\theta_n + b|Y_n$ and equation (9) becomes

$$ {}^*\widehat{y}_{n+1} = \widehat{y}_{n+1} \qquad {}^*\widehat{C}^{\theta p}_{n+1} = A\widehat{C}^{\theta p}_{n+1} \qquad {}^*\widehat{C}^{pp}_{n+1} = \widehat{C}^{pp}_{n+1}. \tag{A.3} $$

Finally, equation (10) leads to

$$ \begin{aligned} {}^*m_{n+1} &= {}^*\widehat{m}_{n+1} + {}^*\widehat{C}^{\theta p}_{n+1}({}^*\widehat{C}^{pp}_{n+1})^{-1}(y_{n+1} - {}^*\widehat{y}_{n+1}) = Am_{n+1} + b, \\ {}^*C_{n+1} &= {}^*\widehat{C}_{n+1} - {}^*\widehat{C}^{\theta p}_{n+1}({}^*\widehat{C}^{pp}_{n+1})^{-1}{}^*\widehat{C}^{\theta p}_{n+1}{}^T = AC_{n+1}A^T. \end{aligned} \tag{A.4} $$

□

Proof of Theorem 1. In this proof we let $\mathcal{B}$ denote the Banach space of matrices in $\mathbb{R}^{N_\theta \times N_\theta}$ equipped with the operator norm induced by the Euclidean norm on $\mathbb{R}^{N_\theta}$. Furthermore we let $\mathcal{L}$ denote the Banach space of bounded linear operators from $\mathcal{B}$ into itself, equipped with the standard induced operator norm. For simplicity we consider the case $r = 0$; a change of origin may be used to extend to the case $r \neq 0$. We first prove that the precision operators converge: $\lim_{n\to\infty} C_n^{-1} = C_\infty^{-1}$; we then

study behaviour of the mean $\{m_n\}$. For both the precision and the mean we first study $\alpha \in (0,1)$ and then $\alpha = 1$. In what follows it is useful to note [72][Theorem 4.1] that the mean and covariance update equations (25) can be rewritten as

$$\begin{aligned} C_{n+1}^{-1} &= G^T \Sigma_\nu^{-1} G + (\alpha^2 C_n + \Sigma_\omega)^{-1}, \\ C_{n+1}^{-1} m_{n+1} &= G^T \Sigma_\nu^{-1} y + (\alpha^2 C_n + \Sigma_\omega)^{-1} \alpha m_n; \end{aligned} \tag{A.5}$$

furthermore the iteration for the covariance remains in the cone of positive semi-definite matrices [72][Theorem 4.1]. Since $\Sigma_\omega \succ 0$, the sequence $\{C_n^{-1}\}$ is bounded:

$$G^T \Sigma_\nu^{-1} G \preceq C_n^{-1} \preceq G^T \Sigma_\nu^{-1} G + \Sigma_\omega^{-1}, \quad \forall n \in \mathbb{Z}_+. \tag{A.6}$$

Introducing $\widetilde{C}_n^{-1} := \sqrt{\Sigma_\omega} C_n^{-1} \sqrt{\Sigma_\omega}$, we may rewrite the covariance update equation (A.5) in the form

$$\widetilde{C}_{n+1}^{-1} = \sqrt{\Sigma_\omega} G^T \Sigma_\nu^{-1} G \sqrt{\Sigma_\omega} + \left(\alpha^2 \widetilde{C}_n + \mathbb{I}\right)^{-1}. \tag{A.7}$$

We define mapping

$$f(X;\alpha) = \sqrt{\Sigma_\omega} G^T \Sigma_\nu^{-1} G \sqrt{\Sigma_\omega} + \left(\alpha^2 X^{-1} + \mathbb{I}\right)^{-1} \tag{A.8}$$

then $\widetilde{C}_{n+1}^{-1} = f\big(\widetilde{C}_n^{-1}; \alpha\big)$. This iteration is well-defined for $\widetilde{C}_n$ in $\mathcal{B}$ satisfying (A.6) and hence for the iteration (A.5).

We first consider $\alpha \in (0,1)$. Then equation (A.7) leads to

$$\widetilde{C}_{n+1} \preceq \alpha^2 \widetilde{C}_n + \mathbb{I} \preceq \frac{1-\alpha^{2n+2}}{1-\alpha^2}\mathbb{I} + \alpha^{2n+2}\widetilde{C}_0 \preceq \frac{1}{1-\alpha^2}\mathbb{I} + \alpha^{2n+2}\widetilde{C}_0, \tag{A.9}$$

and hence, for n is sufficiently large, we have $0 < \epsilon_0 < 1 - \alpha$, such that

$$\widetilde{C}_{n+1}^{-1} \succeq (1 - \alpha^2 - \epsilon_0)\mathbb{I}. \tag{A.10}$$

Let $\mathcal{M} \subset \mathcal{B}$ denote the set of matrices $B \in \mathcal{B}$ satisfying $B \succeq (1-\alpha^2-\epsilon_0)\mathbb{I}$. Then $\mathcal{M}$ is absorbing and forward invariant under f. Thus to show the existence of a globally exponentially attracting steady state it suffices to show that $f(\cdot;\alpha)$ is a contraction on $\mathcal{M}$. The derivative of $f(\cdot;\alpha) : \mathcal{M} \mapsto \mathcal{M}$ is the element $Df(X;\alpha) \in \mathcal{L}$ defined by its action on $\Delta X \in \mathcal{B}$ as follows:

$$Df(X;\alpha)\Delta X = \alpha^2 (X + \alpha^2 \mathbb{I})^{-1} \Delta X (X + \alpha^2 \mathbb{I})^{-1}. \tag{A.11}$$

Thus

$$\begin{aligned} \|Df(X;\alpha)\Delta X\| =& \alpha^2 \left\|(X + \alpha^2\mathbb{I})^{-1}\Delta X (X + \alpha^2\mathbb{I})^{-1}\right\|. \\ \leq& \frac{\alpha^2}{(1-\epsilon_0)^2} \|\Delta X\|. \end{aligned}$$

Therefore, since $\alpha \in (0, 1-\epsilon_0)$,

$$\sup_{X \in \mathcal{M}} \|Df(X;\alpha)\|_{\mathcal{L}} < 1$$

and f is a contraction map on $\mathcal{M}$. This establishes the exponential convergence of $\{\widetilde{C}_n^{-1}\}$. Finally, the sequence $\{C_n^{-1}\}$ converges exponentially fast to C_∞^{-1}, the non-singular fixed point of equation (A.5); Equation (A.6) indicates that C_∞^{-1} is indeed non-singular.

When $\alpha = 1$ define mapping $f(X) = f(X;1)$ so that

$$\widetilde{C}_{n+1}^{-1} = f\big(\widetilde{C}_n^{-1}\big).$$

The derivative $Df(X) \in \mathcal{L}$ is defined by its action on $\Delta X \in \mathcal{B}$ as follows:

$$Df(X)\Delta X = (\mathbb{I} + X)^{-1}\Delta X(I + X)^{-1}. \tag{A.12}$$

Thus, using the lower bound from (A.6) and $\text{Range}(G^T) = \mathbb{R}^{N_\theta}$,

$$\begin{aligned}\|Df(X)\Delta X\| \leq& \left\|(\mathbb{I} + X)^{-1}\right\|^2 \|\Delta X\| \\ \leq& \left\|\left(\mathbb{I} + \sqrt{\Sigma_\omega}G^T\Sigma_\nu^{-1}G\sqrt{\Sigma_\omega}\right)^{-1}\right\|^2 \|\Delta X\| \\ \leq& (1+\epsilon_1)^{-2}\|\Delta X\|,\end{aligned} \tag{A.13}$$

where $\epsilon_1 > 0$. Therefore, f is a contraction map on the whole of $\mathcal{B}$ and the sequence $\{C_n^{-1}\}$ converges. This completes the proof of exponential convergence of $\{C_n^{-1}\}$ to a limit; the sequence $\{C_n^{-1}\}$ converges to C_∞^{-1}, the fixed point of equation (A.5), viewed as a mapping on precision matrices. That $C_\infty \succ 0$ follows from (A.6). Because the convergence is global, the result also indicates the uniqueness of the steady state of equation (26).

We now prove that the mean $\{m_n\}$ converges exponentially fast to m_∞. Using (A.5) the update equation (25) of m_n can be rewritten as

$$m_{n+1} = \alpha(\mathbb{I} - C_{n+1}G^T\Sigma_\nu^{-1}G)m_n + C_{n+1}G^T\Sigma_\nu^{-1}y. \tag{A.14}$$

Thus convergence to m_∞ satisfying

$$m_\infty = \alpha(\mathbb{I} - C_\infty G^T\Sigma_\nu^{-1}G)m_\infty + C_\infty G^T\Sigma_\nu^{-1}y \tag{A.15}$$

is determined by the spectral radius of $\alpha(\mathbb{I} - C_{n+1}G^T\Sigma_\nu^{-1}G)$. If $\alpha \in (0,1)$, using equation (A.6), it follows that

$$\rho(\alpha\mathbb{I} - \alpha C_{n+1}G^T\Sigma_\nu^{-1}G) \leq \alpha < 1$$

and $\{m_n\}$ converges exponentially fast to m_∞. If $\alpha = 1$ then we use the fact that $B := G^T\Sigma_\nu^{-1}G$ is symmetric and that $B \succ 0$. From this it follows 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 (A.6) we deduce that

$$\begin{aligned}\rho\Big(\mathbb{I} - \sqrt{B}C_{n+1}\sqrt{B}\Big) &\leq 1 - \rho\Big(\sqrt{B}\big(B + \Sigma_\omega^{-1}\big)^{-1}\sqrt{B}\Big) \\ &= 1 - \epsilon_2,\end{aligned}$$

for some $\epsilon_2 \in (0,1)$. Since the spectral radius of $I - C_{n+1}B$ is less than one, there is a norm on $\mathbb{R}^{N_\theta}$ in which the operator norm on $I - C_{n+1}B$ is less than one and exponential convergence follows. When $n \mapsto \infty$, equation (A.15) can be rewritten as

$$\begin{aligned}0 &= C_\infty\Big(G^T\Sigma_\nu^{-1}(y - Gm_\infty) + (1-\alpha)(G^T\Sigma_\nu^{-1}G - C_\infty^{-1})m_\infty\Big) \\ &= C_\infty\Big(G^T\Sigma_\nu^{-1}(y - Gm_\infty) - (1-\alpha)\widehat{C}_\infty^{-1}m_\infty\Big).\end{aligned}$$

And hence m_∞ is the minimizer of equation (27).

□

Proof of Lemma 1.

$$\begin{aligned}
\frac{\partial \mathcal{F}\mathcal{G}(m,C)}{\partial m} &= \frac{\partial \mathbb{E}[\mathcal{G}(\theta)]}{\partial m} \\
&= \int \mathcal{G}(\theta)\frac{1}{\sqrt{(2\pi)^{N_\theta}|C|}}\exp\big(-\frac{1}{2}\|C^{-\frac{1}{2}}(\theta-m)\|^2\big)\big(C^{-1}(\theta-m)\big)^T d\theta \\
&= \int \mathcal{G}(\theta)(\theta-m)^T\frac{1}{\sqrt{(2\pi)^{N_\theta}|C|}}\exp\big(-\frac{1}{2}\|C^{-\frac{1}{2}}(\theta-m)\|^2\big)d\theta\cdot C^{-1} \\
&= \int \big(\mathcal{G}(\theta)-\mathbb{E}\mathcal{G}(\theta)\big)(\theta-m)^T\frac{1}{\sqrt{(2\pi)^{N_\theta}|C|}}\exp\big(-\frac{1}{2}\|C^{-\frac{1}{2}}(\theta-m)\|^2\big)d\theta\cdot C^{-1} \\
&= \mathcal{F}d\mathcal{G}(m,C).
\end{aligned} \tag{A.16}$$

□

Proof of Proposition 2. From equation (35) we have

$$\begin{aligned}
\widehat{y}_{n+1} &= \mathcal{F}_u\mathcal{G}_{n+1}, \\
\widehat{C}^{\theta p}_{n+1} &= \widehat{C}_{n+1}\mathcal{F}_u d\mathcal{G}^T_{n+1}.
\end{aligned} \tag{A.17}$$

In what follow we use the modified unscented transform Definition 1, and specifically its use to derive (19) and (20). First note that

$$\widehat{m}_{n+1} = \widehat{\theta}^0_{n+1}, \quad \widehat{y}_{n+1} = \mathcal{G}(\widehat{\theta}^0_{n+1}) = \widehat{y}^0_{n+1}, \quad \text{and} \quad w = W^c_1 = W^c_2 = \cdots = W^c_{2N_\theta}.$$

Now define the matrices

$$\begin{aligned}
\mathcal{Y}_1 &= [\widehat{y}^1_{n+1}-\widehat{y}_{n+1} \quad \widehat{y}^2_{n+1}-\widehat{y}_{n+1} \quad \cdots \quad \widehat{y}^{N_\theta}_{n+1}-\widehat{y}_{n+1}], \\
\mathcal{Y}_2 &= [\widehat{y}^{N_\theta+1}_{n+1}-\widehat{y}_{n+1} \quad \widehat{y}^{N_\theta+2}_{n+1}-\widehat{y}_{n+1} \quad \cdots \quad \widehat{y}^{2N_\theta}_{n+1}-\widehat{y}_{n+1}], \\
\Theta &= [\widehat{\theta}^1_{n+1}-\widehat{m}_{n+1} \quad \widehat{\theta}^2_{n+1}-\widehat{m}_{n+1} \quad \cdots \quad \widehat{\theta}^{N_\theta}_{n+1}-\widehat{m}_{n+1}].
\end{aligned}$$

Then we have

$$\widehat{C}^{\theta p}_{n+1} = \sum_{j=0}^{2N_\theta} W^c_j(\widehat{\theta}^j_{n+1}-\widehat{m}_{n+1})(\widehat{y}^j_{n+1}-\widehat{y}_{n+1})^T = w\Theta(\mathcal{Y}_1^T-\mathcal{Y}_2^T), \tag{A.18a}$$

$$\widehat{C}^{pp}_{n+1} = \sum_{j=0}^{2N_\theta} W^c_j(\widehat{y}^j_{n+1}-\widehat{y}_{n+1})(\widehat{y}^j_{n+1}-\widehat{y}_{n+1})^T + \Sigma_\nu = w(\mathcal{Y}_1\mathcal{Y}_1^T+\mathcal{Y}_2\mathcal{Y}_2^T)+\Sigma_\nu, \tag{A.18b}$$

$$\widehat{C}_{n+1} = \sum_{j=0}^{2N_\theta} W^c_j(\widehat{\theta}^j_{n+1}-\widehat{m}_{n+1})(\widehat{\theta}^j_{n+1}-\widehat{m}_{n+1})^T = 2w\Theta\Theta^T. \tag{A.18c}$$

Equation (A.18c) follows from the definition of the sigma points (19). Since $\widehat{C}_{n+1} \succeq \Sigma_\omega \succ 0$, the matrix $\Theta \in \mathbb{R}^{N_\theta\times N_\theta}$ is non-singular. Thus we have

$$\begin{aligned}
\mathcal{F}_u d\mathcal{G}_{n+1}\widehat{C}_{n+1}\mathcal{F}_u d\mathcal{G}^T_{n+1} &= \widehat{C}^{\theta p}_{n+1}{}^T\widehat{C}^{-1}_{n+1}\widehat{C}_{n+1}\widehat{C}^{-1}_{n+1}\widehat{C}^{\theta p}_{n+1} \\
&= \widehat{C}^{\theta p}_{n+1}{}^T\widehat{C}^{-1}_{n+1}\widehat{C}^{\theta p}_{n+1} \\
&= w(\mathcal{Y}_1-\mathcal{Y}_2)\Theta^T\Big(2w\Theta\Theta^T\Big)^{-1}\Theta(\mathcal{Y}_1^T-\mathcal{Y}_2^T)w \\
&= \frac{w}{2}(\mathcal{Y}_1\mathcal{Y}_1^T+\mathcal{Y}_2\mathcal{Y}_2^T-\mathcal{Y}_1\mathcal{Y}_2^T-\mathcal{Y}_2\mathcal{Y}_1^T).
\end{aligned} \tag{A.19}$$

Using equation (A.19) in equation (A.18b) yields

$$\widehat{C}^{pp}_{n+1} = \mathcal{F}_u d\mathcal{G}_{n+1} \widehat{C}_{n+1} \mathcal{F}_u d\mathcal{G}^T_{n+1} + \Sigma_\nu + \widetilde{\Sigma}_{\nu,n+1}, \tag{A.20}$$

where

$$\widetilde{\Sigma}_{\nu,n+1} := \frac{w}{2}(\mathcal{Y}_1 + \mathcal{Y}_2)(\mathcal{Y}_1 + \mathcal{Y}_2)^T.$$

We note that $\widetilde{\Sigma}_{\nu,n+1}$ is positive semi-definite. Furthermore, the i-th column of $\mathcal{Y}_1 + \mathcal{Y}_2$ satisfies

$$\begin{aligned} \widehat{y}^i_{n+1} + \widehat{y}^{i+N_\theta}_{n+1} - 2\widehat{y}_{n+1} &= \mathcal{G}(\widehat{m}_{n+1} + c_i[\sqrt{\widehat{C}_{n+1}}]_j) + \mathcal{G}(\widehat{m}_{n+1} - c_i[\sqrt{\widehat{C}_{n+1}}]_j) - 2\mathcal{G}(\widehat{m}_{n+1}) \\ &\approx \frac{d^2\mathcal{G}(\widehat{m}_{n+1})}{d^2\theta} : [\sqrt{\widehat{C}_{n+1}}]_j \otimes [\sqrt{\widehat{C}_{n+1}}]_j. \end{aligned} \tag{A.21}$$

Hence $\widetilde{\Sigma}_{\nu,n+1} = 0$ when $\mathcal{G}$ is linear; otherwise $\|\widetilde{\Sigma}_{\nu,n+1}\| = \mathcal{O}(\|\widehat{C}^2_{n+1}\|)$, a second order term with small covariance $\widehat{C}_{n+1}$. □

Proof of Lemma 2. If the steady state C of equation (38b) is singular, then $\exists v \in R^{N_\theta}$ s.t. $v^T C v = 0$. We have

$$\begin{aligned} \Big(v^T C^{\theta p} u\Big)^2 &= \Big(\mathbb{E}[v^T(\theta - m) \otimes (\mathcal{G}(\theta) - \mathcal{G}(m))u]\Big)^2 \\ &\le \mathbb{E}[v^T(\theta - m) \otimes (\theta - m)v]\mathbb{E}[u^T(\mathcal{G}(\theta) - \mathcal{G}(m)) \otimes (\mathcal{G}(\theta) - \mathcal{G}(m))u] \\ &= 0, \end{aligned}$$

for any $u \in R^{N_y}$. This implies that $v^T C^{\theta p} = 0$, and therefore,

$$-2\alpha_h v^T C v - v^T C^{\theta p} {\Sigma_\nu}^{-1} {C^{\theta p}}^T v = 0,$$

which contradicts the assumption that $\Sigma_\omega \succ 0$. □

Appendix B. Illustrative Examples for UKS

The primary focus of the paper is on using the UKI for optimization purposes. However the basic ingredients of the method, and the dynamical system (41) in particular, can also be used to perform approximate posterior sampling from the measure μ given by (3). In the case where μ is Gaussian, the posterior is exactly captured by the steady state of these equations; when the posterior is not Gaussian, then only an approximation is obtained. To illustrate the UKS, we consider, in Subsection Appendix B.1, application to three linear inverse problems from Subsection 3.1, for which the posterior is Gaussian if the prior is Gaussian; and then give a simple example of application to a non-Gaussian posterior in Subsection Appendix B.2.

The UKS equations (41) can be discretized by the following semi-implicit scheme

$$\begin{aligned} m_{n+1} - m_n &= h\Big(C^{\theta p}\Sigma_\eta^{-1}\big(y - \mathbb{E}\mathcal{G}(\theta)\big) - C\Sigma_0^{-1}(m_{n+1} - r_0)\Big), \\ C_{n+1} - C_n &= h\Big(-2C^{\theta p}\Sigma_\eta^{-1}{C^{\theta p}}^T - 2C_n\Sigma_0^{-1}C_n + 2C_{n+1}\Big), \end{aligned} \tag{B.1}$$

with a fixed time-step. The integrals defining $C^{\theta p}$ and $\mathbb{E}\mathcal{G}(\theta)$ are explicitly approximated by the modified unscented transform (see Definition 1) using the Gaussian $\mathcal{N}(m_n, C_n)$. Integration could also be performed using an adaptive time-step, as in [53]; however more work is needed to develop efficient methods stemming from the UKS as formulated here.

Appendix B.1. Linear 2-parameter Model Problem

The linear 2-parameter model problems discussed in Section 5.3 are used with prior

$$r_0 = 0 \quad \text{and} \quad \Sigma_0 = \mathbb{I}.$$

Therefore, the posterior distribution is $\mu \sim \mathcal{N}(m_{ref}, C_{ref})$, where

$$m_{ref} = \left(\Sigma_0^{-1} + G^T\Sigma_\eta^{-1}G\right)^{-1}\left(G^T\Sigma_\eta^{-1}y + \Sigma_0^{-1}r_0\right) \quad \text{and} \quad C_{ref} = \left(\Sigma_0^{-1} + G^T\Sigma_\eta^{-1}G\right)^{-1}. \tag{B.2}$$

The UKS is initialized with $\theta_0 \sim \mathcal{N}(r_0, \Sigma_0)$. The convergence of the UKS, in terms of the posterior mean and covariance errors for $t \in [0, 10]$ are reported in Fig. B.23. Both mean and covariance converge to the posterior mean and covariance. However, even with the semi-implicit scheme the maximum time step that allows for stable simulation is $h = 5 \times 10^{-5}$.

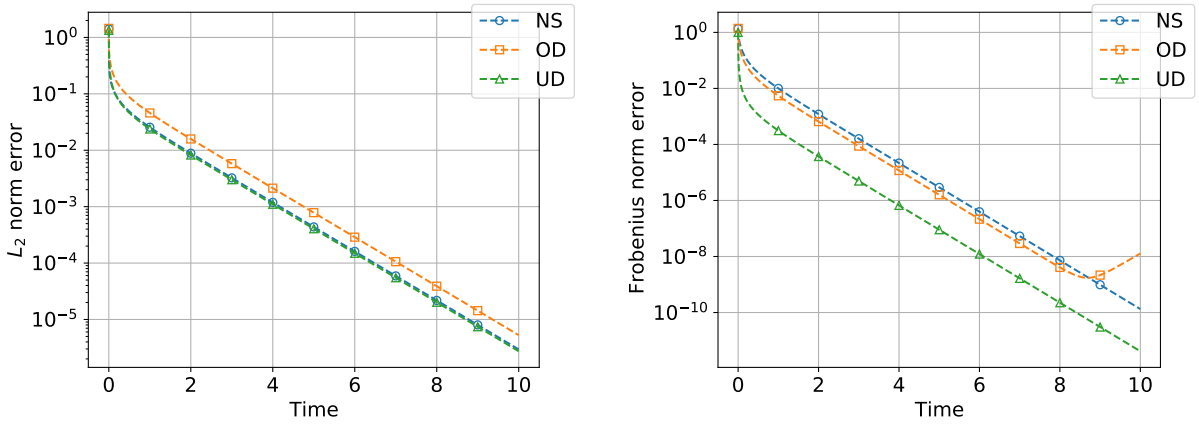


Figure B.23: L_2 error $\|m_n - m_{ref}\|_2$ (left) and Frobenius norm $\|C_n - C_{ref}\|_F$ (right) obtained by UKS for non-singular (NS), over-determined (OD), and under-determined (UD) systems of the linear 2-parameter model problem.

Appendix B.2. Nonlinear 2-Parameter Model Problem

The following Bayesian logistic regression problem is considered,

$$y = \frac{1}{1 + \exp(\theta_{(1)} + \theta_{(2)}x)} + \eta.$$

Here $N_\theta = 2$ and $N_y = 1$, and hence this is an under-determined problem. The prior distribution $\mathcal{N}(r_0, \Sigma_0)$ satisfies

$$r_0 = [1 \quad 1]^T \quad \text{and} \quad \Sigma_0 = \mathbb{I}.$$

The observation data $y_{ref} = 0.08$ is generated at $x = \frac{1}{2}$, with observation error $\eta \sim \mathcal{N}(0, 0.1^2)$ and $\theta_{ref} = [2 \quad 2]^T$.

The UKS is initialized with $\theta_0 \sim \mathcal{N}(r_0, \Sigma_0)$. The posterior distributions obtained by the UKS at $t = 10$ with a time step $h = 5 \times 10^{-5}$ and Markov chain Monte Carlo method (MCMC) with a step size 1.0 and 5×10^6 samples (with a 10^6 sample burn-in period) are presented in Fig. B.24. The estimated posterior distributions are in reasonably good agreement, but of course not as accurate as in the linear setting in the previous subsection, because of a Gaussian approximation being made

to a non-Gaussian distribution. Specifically, the posterior mean and covariance estimated by the UKS are

$$[1.41 \quad 1.20]^T \quad \text{and} \quad \begin{bmatrix} 0.526 & -0.235 \\ -0.235 & 0.884 \end{bmatrix},$$

whilst the posterior mean and covariance estimated by the MCMC are

$$[1.62 \quad 1.31]^T \quad \text{and} \quad \begin{bmatrix} 0.619 & -0.254 \\ -0.254 & 1.00 \end{bmatrix}.$$

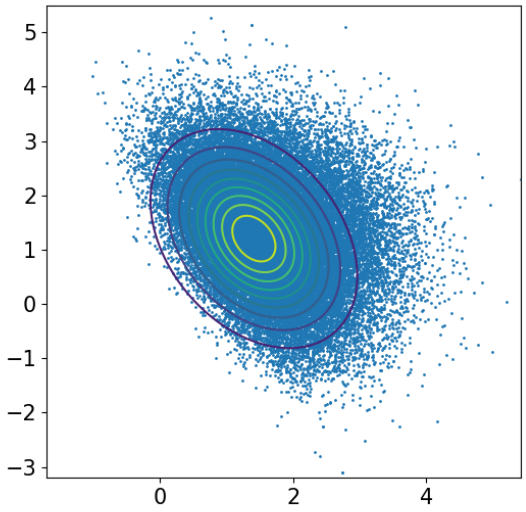


Figure B.24: Contour plot: posterior distributions obtained by UKS at $t = 10$; blue dots: reference posterior distribution obtained by MCMC for the nonlinear 2-parameter model problem. x-axis is for $\theta_{(1)}$ and y-axis is for $\theta_{(2)}$.

Appendix C. Discussion Concerning Regularization ($\alpha \in (0, 1)$) for Under-Determined Linear Systems

Recall the under-determined linear system from Subsection 5.3. In the computations performed there, parameter $\alpha = 1$ and the least squares solution of the inverse problem comprises a one-parameter family of solutions. Nonetheless $\{m_n\}$ converges, but the limit depends on the initial condition. Here we revisit this problem with $\alpha \in (0, 1)$. Theorem 1 shows that the sequence $\{m_n\}$ converges to a uniquely defined limit, which minimizes a regularized least squares problem. Thus, with regularization ($\alpha \in (0, 1)$) the solution does not depend on the initialization of the algorithm. Rather the chosen solution balances the data misfit and a prior regularization term.

In our experiments, we fix $r_0 = 0$ and $\gamma = 0.5^2$. We then employ three different initializations of the UKI:

$$\mathcal{N}(0, 0.5^2\mathbb{I}) \quad \mathcal{N}(\mathbb{1}, 0.5^2\mathbb{I}) \quad \text{and} \quad \mathcal{N}([1\ -1]^T, 0.5^2\mathbb{I}).$$

Thus, in contrast to the experiments in Subsection 5.3, $m_0 \neq r_0$ except in the first case. We employ three different α :

$$\alpha = 0.1,\ 0.5, \text{ and } 0.9.$$

As predicted by Theorem 1, the UKI converges exponentially fast to a point which does not depend on the initial condition, but does depend on the choice of α; for $\alpha = 0.1, 0.5$ and 0.9 respectively we find m_∞ is given by

$$[0.5957 \quad 1.1914]^T \quad [0.5973 \quad 1.1946]^T \quad \text{and} \quad [0.5992 \quad 1.1984]^T.$$

The convergence of the parameter vector $\{m_n\}$ is reported in Fig. C.25. For this test case, smaller α leads to better convergence, and setting $m_0 = r_0$ also leads to improved convergence.

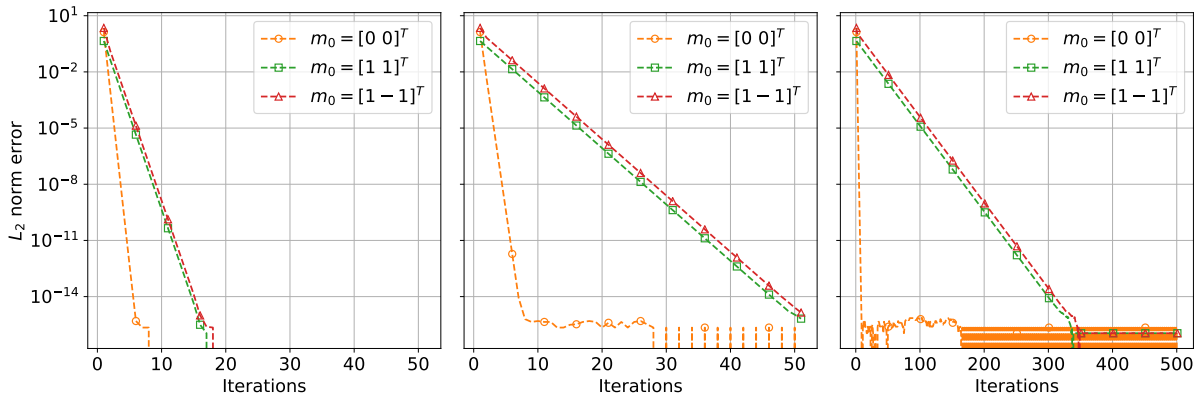


Figure C.25: L_2 error $\|m_n - \theta_{ref}\|_2$ of the under-determined (UD) linear 2-parameter model problem with $\alpha = 0.1$, 0.5 and 0.9 (from left to right).

References

[1] Heinz Werner Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems*, volume 375. Springer Science & Business Media, 1996.

[2] Jari Kaipio and Erkki Somersalo. *Statistical and computational inverse problems*, volume 160. Springer Science & Business Media, 2006.

[3] Masoumeh Dashti and Andrew M Stuart. The bayesian approach to inverse problems. *arXiv preprint arXiv:1302.6989*, 2013.

[4] Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. Sequential monte carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436, 2006.

[5] Alexandros Beskos, Ajay Jasra, Ege A Muzaffer, and Andrew M Stuart. Sequential monte carlo methods for bayesian elliptic inverse problems. *Statistics and Computing*, 25(4):727–737, 2015.

[6] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *J. Basic Eng. Mar*, 82(1):35–45, 1960.

[7] Harold Wayne Sorenson. *Kalman filtering: theory and application*. IEEE, 1985.

[8] Andrew H Jazwinski. *Stochastic processes and filtering theory*. Courier Corporation, 2007.

[9] Michael Ghil, S Cohn, John Tavantzis, K Bube, and Eugene Isaacson. Applications of estimation theory to numerical weather prediction. In *Dynamic meteorology: Data assimilation methods*, pages 139–224. Springer, 1981.

[10] Geir Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162, 1994.

[11] Yan Chen and Dean S Oliver. Ensemble randomized maximum likelihood method as an iterative ensemble smoother. *Mathematical Geosciences*, 44(1):1–26, 2012.

[12] Alexandre A Emerick and Albert C Reynolds. Investigation of the sampling performance of ensemble-based methods with a simple reservoir model. *Computational Geosciences*, 17(2):325–350, 2013.

[13] Sebastian Reich. A dynamical systems framework for intermittent data assimilation. *BIT Numerical Mathematics*, 51(1):235–249, 2011.

[14] Marco A Iglesias, Kody JH Law, and Andrew M Stuart. Ensemble kalman methods for inverse problems. *Inverse Problems*, 29(4):045001, 2013.

[15] Marco A Iglesias. A regularizing iterative ensemble kalman method for pde-constrained inverse problems. *Inverse Problems*, 32(2):025002, 2016.

[16] Martin Hanke. A regularizing levenberg-marquardt scheme, with applications to inverse groundwater filtration problems. *Inverse problems*, 13(1):79, 1997.

[17] Marco A Iglesias and Yuchen Yang. Adaptive regularisation for ensemble Kalman inversion. *Inverse Problems*, 2020.

[18] Neil K Chada, Andrew M Stuart, and Xin T Tong. Tikhonov regularization within ensemble kalman inversion. *SIAM Journal on Numerical Analysis*, 58(2):1263–1294, 2020.

[19] Alfredo Garbuno-Inigo, Franca Hoffmann, Wuchen Li, and Andrew M Stuart. Interacting langevin diffusions: Gradient structure and ensemble kalman sampler. *SIAM Journal on Applied Dynamical Systems*, 19(1):412–441, 2020.

[20] Alfredo Garbuno-Inigo, Nikolas Nüsken, and Sebastian Reich. Affine invariant interacting langevin dynamics for bayesian inference. *SIAM Journal on Applied Dynamical Systems*, 19(3):1633–1658, 2020.

[21] Nikolas Nüsken and Sebastian Reich. Note on interacting langevin diffusions: Gradient structure and ensemble kalman sampler by garbuno-inigo, hoffmann, li and stuart. *arXiv preprint arXiv:1908.10890*, 2019.

[22] Kody JH Law and Andrew M Stuart. Evaluating data assimilation algorithms. *Monthly Weather Review*, 140(11):3757–3782, 2012.

[23] Oliver G Ernst, Björn Sprungk, and Hans-Jörg Starkloff. Analysis of the ensemble and polynomial chaos kalman filters in bayesian inverse problems. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):823–851, 2015.

[24] GA Pavliotis, AM Stuart, and U. Vaes. Derivative-free bayesian inversion using multiscale dynamics. *arXiv preprint arXiv:21*, 2021.

[25] Simon J Julier, Jeffrey K Uhlmann, and Hugh F Durrant-Whyte. A new approach for filtering nonlinear systems. In *Proceedings of 1995 American Control Conference-ACC'95*, volume 3, pages 1628–1632. IEEE, 1995.

[26] Eric A Wan and Rudolph Van Der Merwe. The unscented kalman filter for nonlinear estimation. In *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373)*, pages 153–158. Ieee, 2000.

[27] Mrinal K Sen and Paul L Stoffa. *Global optimization methods in geophysical inversion*. Cambridge University Press, 2013.

[28] Tapio Schneider, Shiwei Lan, Andrew Stuart, and Joao Teixeira. Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations. *Geophysical Research Letters*, 44(24):12–396, 2017.

[29] Oliver RA Dunbar, Alfredo Garbuno-Inigo, Tapio Schneider, and Andrew M Stuart. Calibration and uncertainty quantification of convective parameters in an idealized gcm. *arXiv preprint arXiv:2012.13262*, 2020.

[30] Daniel Z Huang, Kailai Xu, Charbel Farhat, and Eric Darve. Learning constitutive relations from indirect observations using deep neural networks. *Journal of Computational Physics*, page 109491, 2020.

[31] Kailai Xu, Daniel Z Huang, and Eric Darve. Learning constitutive relations using symmetric positive definite neural networks. *arXiv preprint arXiv:2004.00265*, 2020.

[32] Philip Avery, Daniel Z Huang, Wanli He, Johanna Ehlers, Armen Derkevorkian, and Charbel Farhat. A computationally tractable framework for nonlinear dynamic multiscale modeling of membrane fabric. *arXiv preprint arXiv:2007.05877*, 2020.

[33] Brian H Russell. *Introduction to seismic inversion methods*. SEG Books, 1988.

[34] Carey Bunks, Fatimetou M Saleck, S Zaleski, and G Chavent. Multiscale seismic waveform inversion. *Geophysics*, 60(5):1457–1473, 1995.

[35] Johannes Töger, Matthew J Zahr, Nicolas Aristokleous, Karin Markenroth Bloch, Marcus Carlsson, and Per-Olof Persson. Blood flow imaging by optimal matching of computational fluid dynamics to 4d-flow data. *Magnetic Resonance in Medicine*, 2020.

[36] Flávio Celso Trigo, Raul Gonzalez-Lima, and Marcelo Britto Passos Amato. Electrical impedance tomography using the extended kalman filter. *IEEE Transactions on Biomedical Engineering*, 51(1):72–81, 2004.

[37] Dan Simon. *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. John Wiley & Sons, 2006.

[38] François Auger, Mickael Hilairet, Josep M Guerrero, Eric Monmasson, Teresa Orlowska-Kowalska, and Seiichiro Katsura. Industrial applications of the kalman filter: A review. *IEEE Transactions on Industrial Electronics*, 60(12):5458–5471, 2013.

[39] Huazhen Fang, Ning Tian, Yebin Wang, MengChu Zhou, and Mulugeta A Haile. Nonlinear bayesian estimation: from kalman filtering to a broader horizon. *IEEE/CAA Journal of Automatica Sinica*, 5(2):401–417, 2018.

[40] Sharad Singhal and Lance Wu. Training multilayer perceptrons with the extended kalman algorithm. In *Advances in neural information processing systems*, pages 133–140, 1989.

[41] Gintaras V Puskorius and Lee A Feldkamp. Decoupled extended kalman filter training of feedforward layered networks. In *IJCNN-91-Seattle International Joint Conference on Neural Networks*, volume 1, pages 771–777. IEEE, 1991.

[42] Zhengyu Huang, Philip Avery, Charbel Farhat, Jason Rabinovitch, Armen Derkevorkian, and Lee D Peterson. Simulation of parachute inflation dynamics using an eulerian computational framework for fluid-structure interfaces evolving in high-speed turbulent flows. In *2018 AIAA Aerospace Sciences Meeting*, page 1540, 2018.

[43] Daniel Z Huang, P-O Persson, and Matthew J Zahr. High-order, linearly stable, partitioned solvers for general multiphysics problems based on implicit–explicit runge–kutta schemes. *Computer Methods in Applied Mechanics and Engineering*, 346:674–706, 2019.

[44] Daniel Z Huang, Philip Avery, Charbel Farhat, Jason Rabinovitch, Armen Derkevorkian, and Lee D Peterson. Modeling, simulation and validation of supersonic parachute inflation dynamics during mars landing. In *AIAA Scitech 2020 Forum*, page 0313, 2020.

[45] Daniel Z Huang, Will Pazner, Per-Olof Persson, and Matthew J Zahr. High-order partitioned spectral deferred correction solvers for multiphysics problems. *Journal of Computational Physics*, page 109441, 2020.

[46] Alistair Adcroft, Whit Anderson, V Balaji, Chris Blanton, Mitchell Bushuk, Carolina O Dufour, John P Dunne, Stephen M Griffies, Robert Hallberg, Matthew J Harrison, et al. The gfdl global ocean and sea ice model om4. 0: Model description and simulation features. *Journal of Advances in Modeling Earth Systems*, 11(10):3167–3211, 2019.

[47] Charles S Peskin. Numerical analysis of blood flow in the heart. *Journal of computational physics*, 25(3):220–252, 1977.

[48] Marsha Berger and Michael Aftosmis. Progress towards a cartesian cut-cell method for viscous compressible flow. In *50th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition*, page 1301, 2012.

[49] Daniel Z Huang, Dante De Santis, and Charbel Farhat. A family of position-and orientation-independent embedded boundary methods for viscous flow and fluid–structure interaction problems. *Journal of Computational Physics*, 365:74–104, 2018.

[50] Daniel Z Huang, Philip Avery, and Charbel Farhat. An embedded boundary approach for resolving the contribution of cable subsystems to fully coupled fluid-structure interaction. *International Journal for Numerical Methods in Engineering*, 2020.

[51] Marsha J Berger, Phillip Colella, et al. Local adaptive mesh refinement for shock hydrodynamics. *Journal of computational Physics*, 82(1):64–84, 1989.

[52] Raunak Borker, Daniel Huang, Sebastian Grimberg, Charbel Farhat, Philip Avery, and Jason Rabinovitch. Mesh adaptation framework for embedded boundary methods for computational fluid dynamics and fluid-structure interaction. *International Journal for Numerical Methods in Fluids*, 90(8):389–424, 2019.

[53] Emmet Cleary, Alfredo Garbuno-Inigo, Shiwei Lan, Tapio Schneider, and Andrew M Stuart. Calibrate, emulate, sample. *arXiv preprint arXiv:2001.03689*, 2020.

[54] Daniel J Lea, Myles R Allen, and Thomas WN Haine. Sensitivity analysis of the climate of a chaotic system. *Tellus A: Dynamic Meteorology and Oceanography*, 52(5):523–532, 2000.

[55] Qiqi Wang, Rui Hu, and Patrick Blonigan. Least squares shadowing sensitivity analysis of chaotic limit cycle oscillations. *Journal of Computational Physics*, 267:210–224, 2014.

[56] Nikola B Kovachki and Andrew M Stuart. Ensemble kalman inversion: a derivative-free technique for machine learning tasks. *Inverse Problems*, 35(9):095005, 2019.

[57] Dean S Oliver, Albert C Reynolds, and Ning Liu. *Inverse theory for petroleum reservoir characterization and history matching*. Cambridge University Press, 2008.

[58] Eric A Wan and Alex T Nelson. Neural dual extended kalman filtering: applications in speech enhancement and monaural blind signal separation. In *Neural Networks for Signal Processing VII. Proceedings of the 1997 IEEE Signal Processing Society Workshop*, pages 466–475. IEEE, 1997.

[59] Alexander G Parlos, Sunil K Menon, and A Atiya. An algorithmic approach to adaptive state filtering using recurrent neural networks. *IEEE Transactions on Neural Networks*, 12(6):1411–1432, 2001.

[60] JH Gove and DY Hollinger. Application of a dual unscented kalman filter for simultaneous state and parameter estimation in problems of surface-atmosphere exchange. *Journal of Geophysical Research: Atmospheres*, 111(D8), 2006.

[61] David J Albers, Matthew Levine, Bruce Gluckman, Henry Ginsberg, George Hripcsak, and Lena Mamykina. Personalized glucose forecasting for type 2 diabetes using data assimilation. *PLoS computational biology*, 13(4):e1005232, 2017.

[62] Kay Bergemann and Sebastian Reich. An ensemble kalman-bucy filter for continuous data assimilation. *Meteorologische Zeitschrift*, 21(3):213–219, 2012.

[63] Claudia Schillings and Andrew M Stuart. Analysis of the ensemble kalman filter for inverse problems. *SIAM Journal on Numerical Analysis*, 55(3):1264–1290, 2017.

[64] Claudia Schillings and Andrew M Stuart. Convergence analysis of ensemble kalman inversion: the linear, noisy case. *Applicable Analysis*, 97(1):107–123, 2018.

[65] Zhiyan Ding and Qin Li. Ensemble kalman inversion: mean-field limit and convergence analysis. *arXiv preprint arXiv:1908.05575*, 2019.

[66] Bradley M Bell and Frederick W Cathey. The iterated kalman filter update as a gauss-newton method. *IEEE Transactions on Automatic Control*, 38(2):294–297, 1993.

[67] Neil K Chada and Xin T Tong. Convergence acceleration of ensemble kalman inversion in nonlinear settings. *arXiv preprint arXiv:1911.02424*, 2019.

[68] Neil K Chada, Yuming Chen, and Daniel Sanz-Alonso. Iterative ensemble Kalman methods: A unified perspective with some new variants. *arXiv preprint arXiv:2010.13299*, 2020.

[69] José A Carrillo, Young-Pil Choi, Claudia Totzeck, and Oliver Tse. An analytical framework for consensus-based global optimization method. *Mathematical Models and Methods in Applied Sciences*, 28(06):1037–1066, 2018.

[70] José A Carrillo, Franca Hoffmann, Andrew M Stuart, and Urbain Vaes. Consensus based sampling. *arXiv*, 2021.

[71] Sebastian Reich and Colin Cotter. *Probabilistic forecasting and Bayesian data assimilation*. Cambridge University Press, 2015.

[72] Kody Law, Andrew Stuart, and Kostas Zygalakis. Data assimilation. *Cham, Switzerland: Springer*, 2015.

[73] Jonathan Goodman and Jonathan Weare. Ensemble samplers with affine invariance. *Communications in applied mathematics and computational science*, 5(1):65–80, 2010.

[74] John A Nelder and Roger Mead. A simplex method for function minimization. *The computer journal*, 7(4):308–313, 1965.

[75] Benedict Leimkuhler, Charles Matthews, and Jonathan Weare. Ensemble preconditioning for markov chain monte carlo simulation. *Statistics and Computing*, 28(2):277–290, 2018.

[76] Simon Julier, Jeffrey Uhlmann, and Hugh F Durrant-Whyte. A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Transactions on automatic control*, 45(3):477–482, 2000.

[77] David J Albers, Paul-Adrien Blancquart, Matthew E Levine, Elnaz Esmaeilzadeh Seylabi, and Andrew Stuart. Ensemble kalman methods with constraints. *Inverse Problems*, 35(9):095007, 2019.

[78] JA Carrillo and U Vaes. Wasserstein stability estimates for covariance-preconditioned fokker–planck equations. *arXiv preprint arXiv:1910.07555*, 2019.

[79] Lassi Roininen, Janne MJ Huttunen, and Sari Lasanen. Whittle-matérn priors for bayesian statistical inversion with applications in electrical impedance tomography. *Inverse Problems & Imaging*, 8(2):561, 2014.

[80] Marco A Iglesias, Kody JH Law, and Andrew M Stuart. Evaluation of gaussian approximations for data assimilation in reservoir models. *Computational Geosciences*, 17(5):851–885, 2013.

[81] Bernhard Beckermann. The condition number of real vandermonde, krylov and positive definite hankel matrices. *Numerische Mathematik*, 85(4):553–577, 2000.

[82] Matthew M Dunlop, Marco A Iglesias, and Andrew M Stuart. Hierarchical bayesian level set inversion. *Statistics and Computing*, 27(6):1555–1584, 2017.

[83] Nicholas H Nelsen and Andrew M Stuart. The random feature model for input-output maps between banach spaces. *arXiv preprint arXiv:2005.10224*, 2020.

[84] Jan S Hesthaven, Sigal Gottlieb, and David Gottlieb. *Spectral methods for time-dependent problems*, volume 21. Cambridge University Press, 2007.

[85] Steven A Orszag and GS Patterson Jr. Numerical simulation of three-dimensional homogeneous isotropic turbulence. *Physical Review Letters*, 28(2):76, 1972.

[86] Edward N Lorenz. Deterministic nonperiodic flow. In *The Theory of Chaotic Attractors*, pages 25–36. Springer, 2004.

[87] Wael Bahsoun, Ian Melbourne, and Marks Ruziboev. Variance continuity for lorenz flows. In *Annales Henri Poincare*, volume 21, pages 1873–1892. Springer, 2020.

[88] Jan Frøyland and Knut H Alfsen. Lyapunov-exponent spectra for the lorenz model. *Physical Review A*, 29(5):2928, 1984.

[89] Edward N Lorenz. Predictability: A problem partly solved. In *Proc. Seminar on predictability*, volume 1, 1996.

[90] Ibrahim Fatkullin and Eric Vanden-Eijnden. A computational strategy for multiscale systems with applications to lorenz 96 model. *Journal of Computational Physics*, 200(2):605–638, 2004.

[91] Daniel S Wilks. Effects of stochastic parametrizations in the lorenz'96 system. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 131(606):389–407, 2005.

[92] HM Arnold, IM Moroz, and TN Palmer. Stochastic parametrizations and model uncertainty in the lorenz'96 system. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1991):20110479, 2013.

[93] Georg A Gottwald and Sebastian Reich. Supervised learning from noisy observations: Combining machine-learning techniques with data assimilation. *arXiv preprint arXiv:2007.07383*, 2020.

[94] Isaac M Held and Max J Suarez. A proposal for the intercomparison of the dynamical cores of atmospheric general circulation models. *Bulletin of the American Meteorological society*, 75(10):1825–1830, 1994.