Chapter

A Review of the EnKF for Parameter Estimation

Neil K. Chada

Abstract

The ensemble Kalman filter is a well-known and celebrated data assimilation algorithm. It is of particular relevance as it used for high-dimensional problems, by updating an ensemble of particles through a sample mean and covariance matrices. In this chapter we present a relatively recent topic which is the application of the EnKF to inverse problems, known as ensemble Kalman Inversion (EKI). EKI is used for parameter estimation, which can be viewed as a black-box optimizer for PDE-constrained inverse problems. We present in this chapter a review of the discussed methodology, while presenting emerging and new areas of research, where numerical experiments are provided on numerous interesting models arising in geosciences and numerical weather prediction.

Keywords: ensemble Kalman filter, Kalman filter, inverse problems, parameter estimation, data assimilation, optimization

1. Introduction

Inverse problems [1–3] are a class of mathematical problems which have gained significant attention of recent. Simply put, inverse problems are concerned with the recovery of some parameter of interest from noisy unstructured data. Mathematically we can express an inverse problem as the recovery of $u \in \mathcal{X}$ from noisy measurements of data $y \in \mathcal{Y}$, expressed as

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

where $\mathcal{G} : \mathcal{X} \to \mathcal{Y}$ is the forward operator, and $\eta \sim \mathcal{N}(0, \Gamma)$ is some form of additive Gaussian noise. Specifically $\mathcal{N}(0, \Gamma)$ denotes a normal distribution with mean 0 and variance Γ. Commonly the covariance can be taken to be some form of the identity, i.e. $\Gamma = \gamma^2 I$, where $\gamma \in \mathbb{R}$ is some constant and I is the identity. Inverse problems are of high interest due to the amount of relevant problems that arise in wide variety of applications, most notably geophysical sciences, medical imaging and numerical weather prediction [4–6]. The classical approach to solving inverse problems, which is the theme of this chapter, is to construct a least-squares functional, and the solution is represented as a minimizer of some functional of the form

$$u^* := \arg\min_{u \in \mathcal{X}} \frac{1}{2} \|y - \mathcal{G}(u)\|_\Gamma^2 + \lambda R(u), \tag{2}$$

where $\lambda > 0$ is a regularization parameter and $R(u)$ is some regularization term, usually required to prevent the overfitting of the data. A common example is Tikhonov regularization, i.e. $R(u) = \frac{1}{2}\|u\|^2$. Traditional methods for solving (1) include optimization schemes such as the Gauss–Newton method, or Levenburg–Marquardt method which require derivative information of $\mathcal{G}$, which can prove costly and cumbersome. Therefore a motivation for solving inverse problems is to provide gradient-free optimizers which can reduce this computational burden, while attaining a good level of accuracy. The methodology that we motivate, which alleviates these issues, is that of ensemble Kalman inversion (EKI). EKI can be viewed as the application of the ensemble Kalman filter (EnKF) to inverse problems, which is a natural way to solve inverse problems given the connections between data assimilation and inverse problems. The EnKF is a Monte-Carlo version of the celebrated Kalman filter, which is more favorable in high-dimensions. It operates by updating an ensemble of particles through sample mean and covariances. In particular we will take the viewpoint of EKI which acts as PDE-constrained derivative-free optimizer. Therefore EKI can be viewed as a black-box solver where no derivative information is required. Since this method was proposed for inverse problems, it has seen wide applications to various engineering-based applications, as well as developments related to both theory and methodology. In this chapter we discuss some of these keys concepts and insights, while briefly mentioning particular directions with EKI.

The general outline of these chapter is as follows. In Section 2 we provide the necessary background material, which covers the basics of EKI with some intuition and motivation We will discuss the algorithm in both the usual discrete-time setting, but also the continuous-time setting. This will lead onto Section 3 where we discuss one recent direction which is that of regularization theory, and its application to EKI. Furthermore we will also discuss how EKI can be extended to the notion of sampling in statistics within Section 4. Other, less-developed, directions are provided in Section 5. Numerical experiments are provided in basic settings in Section 6 on a number of basic differential equations, before providing some future remarks and a conclusion in Section 7.

2. EKI: background material

In this section we provide the background material related to the understanding and intiution of EKI. This will begin with a discussion on the ensemble Kalman filter, and how it connections with EKI. We will then present EKI in its vanilla form, which is a discrete-time optimizer, before discussing its connections with various existing methods. Finally we will extend the original formulation to the setting of continuous-time where we aim to provide a gradient flow structure of the resulting equations.

2.1 Kalman filtering

The ensemble Kalman filter (EnKF), is a popular methodology based on the celebrated Kalman filter (KF), which was originally developed by Rudolph Kalman in the 1960s [7, 8]. The Kalman filters initial aim was to solve a recursive estimation problem

from dynamics processes and systems. Specifically the KF aims to merge data with model, or signal, dynamics where both equations have the form

$$u_{n+1} = \Psi(u_n) + \xi_n, \qquad \{\xi_n\}_{n\in\mathbb{Z}^+} \sim \mathcal{N}(0, \Sigma), \tag{3}$$

$$y_{n+1} = H(u_{n+1}) + \eta_{n+1}, \qquad \{\eta_{n+1}\}_{n\in\mathbb{Z}^+} \sim \mathcal{N}(0, \Gamma). \tag{4}$$

Here $\{u_n\}_{n\in\mathbb{Z}^+}$ is our signal which is updated through a forward operator $\Psi : \mathbb{R}^m \to \mathbb{R}^m$, which when combined with noise, provides the update u_{n+1}. Our data is denoted as y_{n+1} which is produced by sending our updated signal through the operator $H : \mathbb{R}^m \to \mathbb{R}^{\overline{m}}$, where $\overline{m} > m$, which is known as observational operator. Our initial conditions for the system are given as $u_0 \sim \mathcal{N}(m_0, \mathcal{C}_0)$. This area of recursive estimation, in this setup, became to be known as data assimilation [9, 10].

In particular in the linear and Gaussian setting, where the dynamics and noise are Gaussian, the KF updates state using the first two moments, which we know are the mean and covariance. Assume that the state-space dimension is $d \in R^+$, then the cost of the KF has complexity $\mathcal{O}(d^2)$. For high-dimensional examples this can be an issue, therefore an algorithm that was developed to alleviate this is the EnKF, a Monte Carlo version, proposed by Evensen [11, 12].

The EnKF operates by replacing the true covariance by a sample covariance and mean and updates an ensemble of particles $u_n^{(j)}$, with $1 \le j \le J$ particles, using these moments combined with information from the data. The EnKF can be split into a two-step procedure, which is the prediction step

$$\begin{aligned} \hat{u}_{n+1}^{(j)} &= \Psi\left(u_n^{(j)}\right) + \xi_n^{(j)}, \quad \hat{m}_{n+1} = \frac{1}{J}\sum_{j=1}^{J} u_{n+1}^{(j)}, \\ \hat{C}_{n+1} &= \frac{1}{J-1}\sum_{j=1}^{J}\left(u_{n+1}^{(j)} - \hat{m}_{n+1}\right)\left(u_{n+1}^{(j)} - \hat{m}_{n+1}\right)^T, \end{aligned} \tag{5}$$

and update step

$$\begin{aligned} K_{n+1} &= \hat{C}_{n+1}H^T\left(H\hat{C}_{n+1}H^T + \Gamma\right), \\ u_{n+1}^{(j)} &= \left(I - K_{j+1}H\right)\hat{u}_{n+1}^{(j)} + K_{n+1}y_{n+1}^{(j)}, \\ y_{n+1}^{(j)} &= y_{n+1} + \eta_{n+1}^{(j)}, \end{aligned} \tag{6}$$

where K_{n+1} represents the Kalman gain matrix and $\xi_n^{(j)}$ and $\eta_{n+1}^{(j)}$ are i.i.d. Gaussian noise. In the EnKF context our prediction step defines a sample mean and covariance from our signal. From this in the analysis step we define our Kalman gain through our sample covariance, which updates our signal, which is given by $u_{n+1}^{(j)}$. This is aided by aiming to minimize the discrepancy of the data $y_{n+1}^{(j)}$ and the quantity $H(u)$. To better understand this discrepancy, there is an alternative approach of looking at the EnKF is through a variational approach, where we consider the follow cost function

$$I_n(u) := \frac{1}{2}\left|y_{n+1}^{(j)} - H(u)\right|_{\Gamma}^2 + \frac{1}{2}\left|u - \hat{u}_{n+1}^{(j)}\right|_{\hat{C}_{n+1}}^2, \tag{7}$$

for which we aim to minimize, which is defined as the updated mean

$$\hat{m}_{n+1} = \arg\min_u I_n(u). \tag{8}$$

This minimization procedure relies on the updated covariance $\hat{C}_{n+1}$ which is dependent entirely on $\hat{u}^{(j)}$. As described in the prediction step and update step of filtering, a mapping is presented between distributions. As we related the distributions in the filtering setting, for each step, we can do so similarly for the EnKF, i.e.

$$\left\{u_n^{(j)}\right\}_{j=1}^J \mapsto \left\{u_{n+1}^{(j)}\right\}_{j=1}^J, \qquad \left\{u_{n+1}^{(j)}\right\}_{j=1}^J \mapsto \left\{\hat{u}_{n+1}^{(j)}\right\}_{j=1}^J. \tag{9}$$

With the EnKF, compared to KF, the computational complexity associated with it is $\mathcal{O}(Jd)$, where one usually assumes $J < d$, therefore implying the reduction in cost.

2.2 EnKF applied to inverse problems

Since the formulation of the EnKF, there has been a huge interest from practitioners in various applicable disciplines. Most notably this has been within numerical weather prediction, geophysical sciences and signal processing related to state estimation. In this chapter our focus is on the application of the EnKF to inverse problems, namely to solve (1). We now introduce this application which is known as ensemble Kalman inversion (EKI), which was introduced by Iglesias et al., motivated from Li et al., [13] as a derivative-free optimizer for PDE-constrained inverse problems.

As with the EnKF, we are concerned with updating an ensemble of particles, for which now we modify notation with n now denoting the iteration count. Given an initial ensemble $\left\{u_0^{(j)}\right\}$, our aim is to learn a true underlying unknown $u^\dagger$. To do so, as done with the EnKF, we first define our sample mean and covariance matrices

$$\overline{u}_n^{(j)} = \frac{1}{J}\sum_{j=1}^J u_n^{(j)}, \quad \overline{u}_n^{(j)} = \frac{1}{J}\sum_{j=1}^J G\left(u_n^{(j)}\right),$$

$$C_n^{uu} = \frac{1}{J-1}\sum_{j=1}^J \left(u_n^{(j)} - \overline{u}\right)\left(u_n^{(j)} - \overline{u}\right)^T, \quad C_n^{up} = \frac{1}{J-1}\sum \left(u_n^{(j)} - \overline{u}\right)\left(G\left(u_n^{(j)}\right) - \overline{G}\right)^T. \tag{10}$$

which we can through the update equation

$$u_{n+1}^{(j)} = u_n^{(j)} + hC^{up}(hC^{pp} + \Gamma)^{-1}\left(y_n^{(j)} - \mathcal{G}\left(u_n^{(j)}\right)\right), \tag{11}$$

$$y_n^{(j)} = y + \eta_n^{(j)}, \tag{12}$$

where y represents our true data and $h > 0$ denotes a step size related to the level of discretization. **Figure 1** provides a pictorial description of the EnKF, which has been described above.

The update equation of EKI (11) is of interest as it coincides with the update formula for Tikhonov regularization for linear statistical inverse problems. Namely if we consider $R(u) = \frac{1}{2}\|u\|_{C_0}^2$, then the update formula, in the linear $\mathcal{G}(\cdot) = \mathcal{G}$ and Gaussian setting is given as

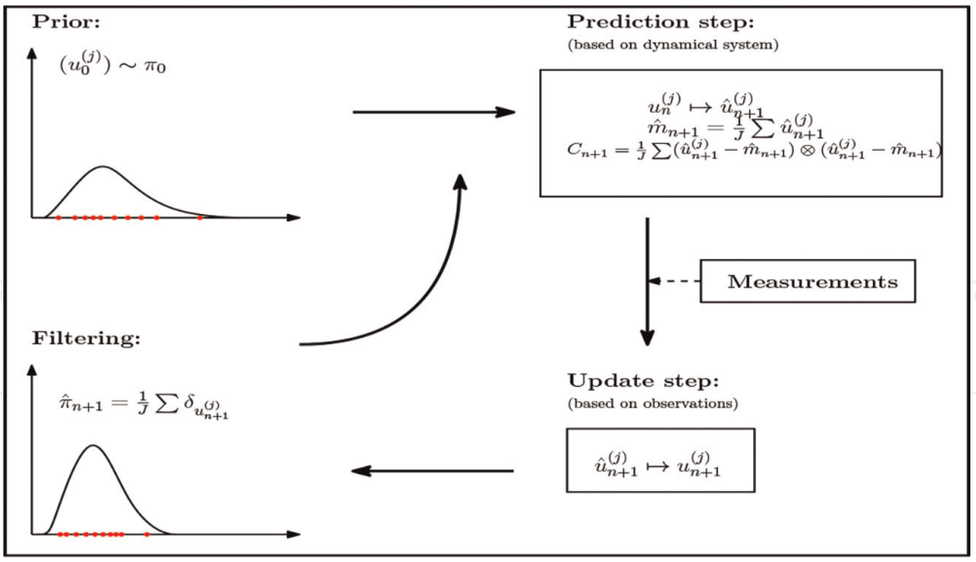

Figure 1.
Dynamics of the ensemble Kalman filter, split into the prediction and update steps.

$$u_{TP} = \overline{u} + C\mathcal{G}^*(\mathcal{G}C\mathcal{G}^* + \Gamma)^{-1}(y - \mathcal{G}\overline{u}), \tag{13}$$

where $\mathcal{G}^*$ denotes the derivative of the operator $\mathcal{G}$. This connection is of relevance and was discussed in [14], where it was shown that taking the limit as $J \to \infty$, it was shown that $u \to u_{TP}$. This is of interest as the minimizing the regularized functional (13) is equivalent to the following maximization procedure in statistics

$$u := \arg\max_{u \in \mathcal{X}} \mathbb{P}(u|y). \tag{14}$$

known as the MAP formulation, where $\mathbb{P}(u|y) = \mathbb{P}(y|u)\mathbb{P}(u)$ denotes the posterior distribution. This connection is discussed in [15]. Therefore this provides some insight into EKI and its connection with other known existing methodologies in inverse problems. An important entity to discuss is a property that EKI inherits, which is the *subspace property*. It is given by the following lemma.

Lemma 1.1 Let $\mathcal{A}$ be the linear span of the initial ensemble $\left\{u_0^{(j)}\right\}_{j=1}^{J}$, then we that $\left\{blacku_n^{(j)}\right\}_{j=1}^{J} \in \mathcal{A}$ for all $n \in \mathbb{N}$.

The essence of the subspace property states that the updated ensemble of particles is spanned by the initial ensemble. This is important, because it provides a justification on the performance, whether the initial ensemble is a good choice or not. Therefore it can act as an advantage or a disadvantage.

2.3 Continuous-time formulation

The original representation of EKI, as shown in (11), is a discrete-time iterative scheme similar to other optimization methods. However it is of interest to understand EKI in a continuos-time setting, which was considered by Schillings et al. [16, 17]. This is primarily for two reasons; (i) firstly that one can understand more easily how the dynamics of (11) and (12) behaves, and secondly (ii) it provides

new numerical schemes for EKI, which is specific in the continuous-time setting. In order to derive such equations, as usual we require to take the step-size to zero, i.e. $h \to 0$. Once we do this, we have the following set of stochastic differential equations

$$\frac{du^{(j)}}{dt} = C^{uw}(u)\Gamma^{-1}\left(y - \mathcal{G}\left(u^{(j)}\right)\right) + C^{uw}(u)\sqrt{\Gamma^{-1}}\frac{dW^{(j)}}{dt}, \tag{15}$$

with $W^{(j)}$ denoting independent cylindrical Brownian motions. By substituting the form of the covariance operator, we see

$$\frac{du^{(j)}}{dt} = \frac{1}{J}\sum_{k=1}^{J}\left\langle \mathcal{G}\left(u^{(k)}\right) - \overline{\mathcal{G}}, y - \mathcal{G}\left(u^{(j)}\right) + \sqrt{\Gamma}\frac{dW^{(j)}}{dt}\right\rangle_{\Gamma}\left(u^{(k)} - \overline{u}\right). \tag{16}$$

For this we take our forward operator $\mathcal{G}(\cdot) = A\cdot$ to be bounded and linear. Using this notion and by substituting our linear operator A in (16) we have the following diffusion limit

$$\frac{du^{(j)}}{dt} = \frac{1}{J}\sum_{k=1}^{J}\left\langle A\left(u^{(k)} - \overline{u}\right), y - Au^{(j)}\right\rangle_{\Gamma}\left(u^{(k)} - \overline{u}\right). \tag{17}$$

By defining the empirical covariance operator

$$C(u) = \frac{1}{J-1}\sum_{k=1}^{J}\left(u^{(k)} - \overline{u}\right)\otimes\left(u^{(k)} - \overline{u}\right), \tag{18}$$

and taking $\Gamma = 0$ we can express (17) as

$$\begin{aligned} \frac{du^{(j)}}{dt} &= -C(u)D_u\Phi\left(u^{(j)}; y\right), \\ \Phi(u; y) &= \frac{1}{2}\|\Gamma^{-1/2}(y - Au)\|^2. \end{aligned} \tag{19}$$

Thus we note that each particle performs a preconditioned gradient descent for $\Phi(\cdot; y)$ where all the gradient descents are preconditioned through the covariance $C(u)$. Since our covariance operator $C(u)$ is semi-positive definite we have that

$$\frac{d}{dt}\Phi(u(t); y) = \frac{d}{dt}\,\frac{1}{2}\|\Gamma^{-1/2}(y - Au)\|^2 \leq 0. \tag{20}$$

In the context of EKI this is of interest as it is a first result providing some indication of the dynamics, which was not achievable through the discrete-time update formula (11). Indeed given the gradient flow structure, we are able to see that the EKI abides by a usual optimization function, with the dynamics following the direction of the negative gradient, or in other-words towards to minimizer of Φ. Since the continuous-time formulation was derived, there has been different works deriving further analysis, most notably with recent success on the nonlinear setting, and other well-known results. This can be found in [18].

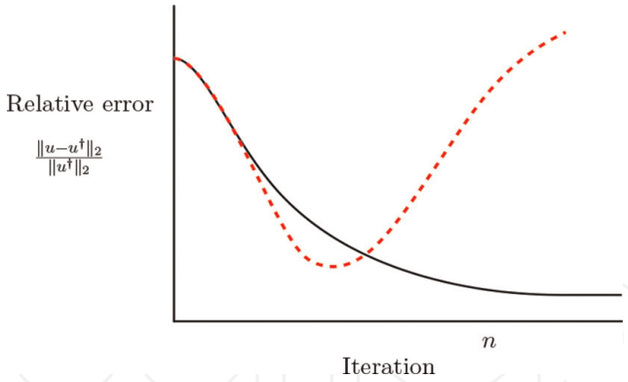

Figure 2.
The figure presents two simulations of EKI as the iterations increase. The black curve represents what we aim to achieve, however in certain situations the data is commonly overfitted. Therefore this can cause a divergence in the relative error, as shown by the dashed red curve.

3. Regularization

In this section we discuss the role of regularization in EKI. We will begin with an introduction into iterative regularization schemes, that have been used before discussing Tikhonov regularization, L_p and particular adaptive choices.

As briefly discussed regularization is an important tool in optimization, and inverse problems aimed at preventing the over-fitting, or influence, of the data. We refer the reader to various pieces of literature that give a concise overview on this [19, 20]. The over-fitting of data can cause issues in inverse problems, such as the divergence of the error, therefore careful consideration is needed to prevent this. A cartoon representation of this is given in **Figure 2**.

To initiate this chapter, there are two main forms of regularization one can apply for inverse problems. The first is related to *iterative regularization*, where the regularization is included within the iterative scheme. This can be included directly such as the form

$$u_{n+1}^{(j)} = u_n^{(j)} + hC^{up}(hC^{pp} + \alpha_n\Gamma)^{-1}\left(y_n^{(j)} - \mathcal{G}\left(u_n^{(j)}\right)\right), \tag{21}$$

or in the presence of a discrepancy principle of the form

$$\|\Gamma^{-1}\left(y - \overline{u}_n^{(j)}\right)\|^2 \leq \vartheta\eta, \qquad \vartheta \in (0, 1), \tag{22}$$

which controls the error between the updated ensemble and the true unknown. The discrepancy principle acts as a stopping rule if the error becomes big, and the the modified update formula contains a sequence of numbers $\{\alpha_n\}_{n\in\mathbb{N}}$ aimed at also preventing the overfitting of the data. This sequence is chosen in such a way that is related to a discrepancy principle. Specifically for EKI this has been considered in numerous work by Iglesias et al. [21, 22].

However more recent work has considered regularization through the least-squares functional (LSF) (2). For EKI the first known form to consider this, is Tikhonov regularization which has the penalty form of $R(u) = \frac{1}{2}\|u\|_{C_0}^2$. This form of regularization is a natural choice, as it very well-known and understood but can view viewed as a Gaussian form of regularization, which smoothes the problems. In the context of EKI this makes sense, as commonly one assumes Gaussian dynamics. The

work of Chada et al. [23] first developed this extension, which was done by modifying (1) to the following

$$\begin{aligned} y &= G(u) + \eta_1, \\ u &= \eta_2, \end{aligned} \tag{23}$$

where $\eta_1 \sim \mathcal{N}(0, \Gamma)$,$\eta_2 \sim \mathcal{N}\left(0, \lambda^{-1}C_0\right)$..
Now we introduce z,η and the mapping $\mathcal{F} : \mathcal{X} \times \mathcal{X} \mapsto \mathcal{Y} \times \mathcal{X}$ as follows:

$$z = \begin{bmatrix} y \\ 0 \end{bmatrix}, \quad F(u) = \begin{bmatrix} G(u) \\ u \end{bmatrix}, \quad \eta = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix}, \tag{24}$$

and

$$\eta \sim N(0, \Sigma), \quad \Sigma = \begin{bmatrix} \Gamma & 0 \\ 0 & \lambda^{-1}C_0 \end{bmatrix}. \tag{25}$$

Therefore our inverse problem is now reformulated at

$$z = \mathcal{F}(u) + \eta. \tag{26}$$

now from this we can modify EKI to include the above setup, for which we refer to it as Tikhonov ensemble Kalman inversion (TEKI), which takes the following form

$$u_{n+1}^{(j)} = u_n^{(j)} + hB^{up}(hB^{pp} + \Gamma)^{-1}\left(z_n^{(j)} - \mathcal{F}\left(u_n^{(j)}\right)\right), \tag{27}$$

where we have now modified covariance matrices B^{up},B^{pp}. From this inclusion, the authors of [23] were able to show that analytically, the subspace property still holds, while other such results as observability and controllability and the ensemble collapse. More importantly through the numerical simulations, it was shown that one can prevent the over-fitting phenomenon.

Since this work a number of useful extensions have been considered, such as its understanding in the continuous-case, as well as the new variants in the discrete-time setting [24]. Two recent developments on this have been firstly on the extension to L_p regularization [25, 26], which is to motivate reconstructing edges or lines, where the LSF is modified to

$$\Phi(u; y) := \frac{1}{2}\|y - \mathcal{G}(u)\|_\Gamma^2 + \lambda\|u\|_p, \quad p \geq 1. \tag{28}$$

Finally another direction is related to producing adaptive strategies for TEKI. Adaptive regularization schemes are of importance, as choosing a correct choice of the regularization parameter $\lambda > 0$ can have a big impact on the reconstruction. Therefore thinking adaptively allows one to evolve the parameter over the iteration count, now denoted as λ_n. The work of Weissmann et al. [27] provides these developments in an adaptive fashion.

4. Ensemble Kalman sampling

Although the EKI has been introduced through the application of the EnKF to inverse problems and hence sequential sampling method, the trending viewpoint of

EKI lies in optimization. So far, we have seen its motivation from the gradient flow structure in the continuous-time formulation in Section 2.3 and the representation as SDE. For applying EKI as a consistent sampling method, we would instead of taking the limit $t \to \infty$ rather consider the limit $t \to 1$. For linear forward models EKI is consistent with the posterior distribution, however, it is known to be not consistent with the Bayesian perspective in the nonlinear setting [28].

Building up on this fact, the motivation behind the ensemble Kalman sampler [29] is to modify the time-dynamical system of EKI in a way such that the limiting distribution for $t \to \infty$ corresponds to the posterior distribution. We will start the discussion with an introductory example.

Example 1.1 Let π_* be a pdf of the form $\pi_*(x) \propto \exp(-\Phi(u))$ with $\Phi(u) = \frac{1}{2}\|y - G(u)\|_\Gamma^2 + \|u\|_C^2$, i.e. π_* corresponds to the posterior pdf under Gaussian prior assumption $\pi_0 = \mathcal{N}(0, C)$. We consider the Langevin diffusion given by

$$du_t = \nabla_u \log \pi_*(u_t)dt + \sqrt{2}dW_t, \quad u_0 \sim \pi_0, \tag{29}$$

where $(W_t)_{t\geq 0}$ denotes a Brownian motion in $\mathcal{X} = \mathbb{R}^{n_u}$. The evolution of the distribution ρ_t of the state u_t can then be described through the Fokker–Planck equation

$$\partial\rho_t = \nabla \cdot (\rho_t \nabla \log \pi_*) + \Delta\rho_t, \quad \rho_0 = \pi_0, \tag{30}$$

where under certain assumptions on Φ the underlying Markov process $(u_t)_{t\geq 0}$ is ergodic and its unique invariant distribution is given by π_* [30]. Taking the Fokker–Planck eq. (30) into account the convergence to equilibrium can be described through the Kullback–Leibler (KL) divergence $\mathrm{KL} = \int_{\mathcal{X}} q_1(x) \log\left(\frac{q_1(x)}{q_2(x)}\right) \mathrm{d}x$ [31]. Assuming a $\log$-Sobolev inequality (e.g. satisfied for $\log$-concave π_*), it follows that

$$\mathrm{KL}(\rho_t|\pi_*) \leq \exp(-\lambda t)\mathrm{KL}(\rho_0|\pi_*) \tag{31}$$

for some $\lambda > 0$ [32].

4.1 Interacting Langevin sampler

The interacting Langevin sampler has been introduced, motivated by the preconditioned gradient descent method, as interacting particle system represented by the coupled system of SDEs

$$du_t^{(j)} = C(u_t)\nabla_u \log \pi_*\left(u_t^{(j)}\right)dt + \sqrt{2C(u_t)}dW_t, \quad j = 1, \ldots, J, \tag{32}$$

initialized through an i.i.d. sample $u_0^{(j)} \sim \pi_0$. The idea of preconditioning with $C(u_t)$ instead of a fixed preconditioning matrix $C \in \mathbb{R}^{n_u \times n_u}$ is motivated through the corresponding mean-field limit. In the large particle limit, the corresponding SDE is given as

$$du_t = C(\rho_t)\nabla_u \log \pi_*(u_t)dt + \sqrt{2C(\rho_t)}dW_t, \quad u_0 \sim \pi_0, \tag{33}$$

where the macroscopic mean and covariance operator are defined as

$$m(\rho) = \int_{\mathcal{X}} x\rho(x)dx, \quad C(\rho) = \int_{\mathcal{X}} (x - m(\rho)) \otimes (x - m(\rho))dx. \tag{34}$$

This connects the interacting Langevin system to its origin Langevin diffusion (29). Hence, in the long-time limit the preconditioning matrix will formally be given by the covariance operator corresponding to the stationary distribution (assuming it exists).

The resulting modified Fokker–Planck equation is given by

$$\partial\rho_t = \nabla \cdot (\rho_t C(\rho_t)\nabla \log \pi_*) + \mathrm{Tr}\big(C(\rho_t)D^2\rho_t\big), \quad \rho_0 = \pi_0. \tag{35}$$

Assuming that $C(\rho_t) \geq \alpha \mathrm{Id}$ and the target distribution of the form $\pi_*(u) \propto \exp(-\Phi(u))$, $\Phi(u) = \frac{1}{2}\|y - \mathcal{G}(u)\|_\Gamma^2 + \lambda\|u\|_{C_0}^2$, to be log-concave, the solution ρ_t of (35) converges exponentially fast to equilibrium

$$\mathrm{KL}(\rho_t|\pi_*) \leq \exp(-\lambda t)\mathrm{KL}(\rho_0|\pi_*), \tag{36}$$

for some $\lambda > 0$ [29], Proposition 3.1. Furthermore, through the preconditioning with the sample covariance the resulting scheme remains invariant under affine transformations [33].

4.2 Ensemble Kalman sampler

One of the attractive features of the EnKF as well as of EKI is its derivative-free implementation. The basis of the ensemble Kalman sampler (EKS) is to build a modified interacting Langevin sampler avoiding to compute derivatives. Let $\pi_*(u) \propto \exp\left(-\frac{1}{2}\|y - \mathcal{G}(u)\|_\Gamma^2 - \|u\|_{C_0}^2\right)$, then the interacting Langevin system is given by

$$\begin{aligned} du_t^{(j)} &= -C(u_t)\mathrm{D}\mathcal{G}\left(u_t^{(j)}\right)^{\mathrm{T}}\Gamma^{-1}\left(G\left(u_t^{(j)}\right) - y\right) - C(u_t)C_0^{-1}u_t^{(j)}dt + \sqrt{2C(u_t)}dW_t, \\ j &= 1, \ldots, \\ J. \end{aligned} \tag{37}$$

Motivated by the approximation $C^{uw}(u) \approx C(u)\mathrm{D}\mathcal{G}\left(u^{(j)}\right)^{\mathrm{T}}$ the EKS is then formulated as the solution of the system of coupled SDEs

$$du_t^{(j)} = -C^{uw}(u_t)\Gamma^{-1}\left(\mathcal{G}\left(u_t^{(j)}\right) - y\right) - C(u_t)C_0^{-1}u_t^{(j)}dt + \sqrt{2C(u_t)}dW_t, \quad j = 1, \ldots, J. \tag{38}$$

We note that the approximation $C^{uw}(u) \approx C(u)\mathrm{D}\mathcal{G}\left(u^{(j)}\right)^{\mathrm{T}}$ is exact for linear forward models and hence, the EKS coincides with the interacting Langevin sampler in the linear setting. However, for nonlinear forward models the approximation of derivatives is only accurate in case the particles are close to each other. Since in the application of EKS the particles are aiming to represent a distribution, the particles are not expected to be close to each other. This fact suggests to formulate a localized version of the preconditioning sample covariance matrix, incorporating more weights on particles close to each other, but reducing the weight between particles far away. Therefore, we define the distance-dependent weights between particle $u_t^{(j)}$ and $u_t^{(i)}$

$$w_t^{ji} = \frac{\exp\left(-\frac{1}{2\gamma}\|u_t^{(j)} - u_t^{(i)}\|_D^2\right)}{\sum_{l=1}^{J}\exp\left(-\frac{1}{2\gamma}\|u_t^{(j)} - u_t^{(l)}\|_D^2\right)}, \tag{39}$$

for scaling parameters $\gamma > 0$ and symmetric positive-definite matrix $D \in \mathbb{R}^{n_u \times n_u}$. The localized (mixed) sample covariance matrix around particle $u_t^{(j)}$ is the defined as

$$\begin{aligned} C\left(u_t^{(j)}\right) &= \sum_{i=1}^{J} w_t^{ji}\left(u_t^{(i)} - \overline{u}_t^{(j)}\right) \otimes \left(u_t^{(i)} - \overline{u}_t^{(j)}\right), \\ C^{uw}\left(u_t^{(j)}\right) &= \sum_{i=1}^{J} w_t^{ji}\left(u_t^{(i)} - \overline{u}_t^{(j)}\right) \otimes \left(\mathcal{G}\left(u_t^{(i)}\right) - \overline{\mathcal{G}}_t^{(j)}\right), \end{aligned} \tag{40}$$

with localized mean

$$\overline{u}_t^{(j)} = \sum_{i=1}^{J} w_t^{ji} u_t^{(i)}, \quad \overline{\mathcal{G}}_t^{(j)} = \sum_{i=1}^{J} w_t^{ji}\mathcal{G}\left(u_t^{(i)}\right). \tag{41}$$

The localized EKS then reads as

$$du_t^{(j)} = -C^{uw}\left(u_t^{(j)}\right)\Gamma^{-1}\left(\mathcal{G}\left(u_t^{(j)}\right) - y\right) - C\left(u_t^{(j)}\right)C_0^{-1}u_t^{(j)}dt + \sqrt{2C\left(u_t^{(j)}\right)}dW_t, \quad j = 1, \ldots, J. \tag{42}$$

While the original EKS shows promising results for nearly Gaussian target distribution, the considered localized variant helps to extend the scope to multimodal target distributions [34]. Other such related work has aimed to provide further understandings of the EKS. This has included the derivation of providing mean field limits and, but also providing various generalizations [33, 35].

5. Other directions

As we have discussed some of the more recent developments in EKI, we now focus on other, more smaller, extensions. In this section we will discuss these each in turn, which will include machine learning, understanding EKI in the context of nonlinear inverse problems, and finally applications related to engineering such as geophysical sciences.

5.1 Applications in machine learning

The developments of machine learning methodologies has seen a significant increase in the last decade, which have been produced to solve problems related to health-care, imaging, and decision processes. In particular much of the these developments has been to due the advancements in optimizaion theory. As a result, ensemble Kalman methods can be viewed as a natural class of algorithms to be directly applied, as they are derivative-free optimizers.

The first work aimed at characterizing this connection was [36] which demonstrated this. The authors motivated EKI as a replacement to SGD where they initially

applied it to supervising learning problems. Given a dataset $\left\{x_j, y_j\right\}_{j=1}^{N}$ assumed to be i.i.d. samples from a particular distribution, then given the Monte Carlo approximation one has the minimization procedure

$$\begin{aligned} &\arg\min_u \Phi_s(u;x,y), \\ &\Phi_s(u;x,y) = \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}\left(\mathcal{G}\left(u|x_j\right), y_j\right) + \frac{\lambda}{2}\|u\|_{C_0}^2, \end{aligned} \tag{43}$$

where $\mathcal{L} : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}^+$, is some positive-definite function. In other words, one is trying to learn x_j from the labeled data y_j. Supervised learning is used for common ML applications such as image classification and natural language processing. Another related application is that of semi-supervised learning, which aims to learn x_j from some of the data y_j where do not have access to all of it. This modified the least squares functional given in (43).

Another interesting direction has been the inclusion of EKI for training and learning neural networks [37]. This builds upon the previous work discussed, but with a number of modifications. In particular what the authors show is that they are able to prove convergence of EKI to the minimizer of a strongly convex function. They apply their modified methodology to a nonlinear regression problem of the form

$$F(\theta) = A\theta + \varepsilon \sin(B\theta), \tag{44}$$

where θ is the parameter of interest and $F(\theta)$ is the objective functional of interest. This was also extended to the likes of image classification problems, specifically the well-known MNIST handwritten data set.

A final and more recent direction of EKI and ML, was the work of Guth et al. [38], which provided a way of solving the forward problem, within EKI.

5.2 Extensions to nonlinear convergence analysis

A major challenge with EKI, and the EnKF in general, is establishing convergence analysis and properties in the nonlinear setting. As it is well known in the linear and Gaussian setting, as the the number of particles $N \to \infty$, the EnKF coincides with the KBF. However in the nonlinear setting it is has been challenging to derive any such results rigorously. Some ongoing and recent work has aimed to bridge the connections between EKI and nonlinear dynamics. The first paper that provided some form of analysis was the work of Chada et al. [24] which considered a specific form of EKI, in the discrete-time setting.

Namely the update formula is modified to

$$\begin{aligned} m_{n+1} &= m_n + C_n^{pp}\left(C_n^{up} + h_n^{-1}\Gamma\right)^{-1}(z - H(m_n)), \\ C_{n+1} &= C_n^{uu} - C_n^{up}\left(C_n^{pp} + h_n^{-1}\Gamma\right)^{-1}C_n^{pu} + \alpha_n^2\Sigma, \end{aligned} \tag{45}$$

where we adopt an ensemble square root filter formulation, which is known to perform better. As well as this we also include covariance inflation (i.e. inflation factor of α_n), and an adaptive step-size h_n motivated from stochastic optimization to allow an acceleration for the convergence. However the other underlying contribution, as

eluded to, is that given this update form we are able to prove convergence towards both local and global minimizers. In other words for the later, we have the following result

$$\lambda_c \|m_N - u^*\|^2 \leq \ell(m_N) - \ell(u^*) \leq \frac{D}{N^\alpha}, \tag{46}$$

which the above result establishes polynomial convergence. We note from the above equation, that λ_c is a convexity constant, ℓ is the associated loss function, D is some constant, u^* is the global minimizer and α is some term, which we refer to [24], for further details.

As one can notice, this convergence analysis was considered for the discrete-time setting, so a natural extension from this is to the continuous-time framework. The work of Blomker et al. [18] provide a first convergence analysis in this direction. However given both these works, a full understanding in the nonlinear setting has not been achieved, where considerable work is still required. Thus these papers provide a first step in doing so, for both settings.

5.3 Engineering applications

As a final direction to discuss in detail, which is very much related to the theme of this book, are applications in particular engineering applications. The advantage of these ensemble Kalman methods, is that they can be viewed as a black box-solver, therefore it is highly applicable. One particular application has been geophysical sciences, related to recovering quantities of interest which are below the surface, or subsurface. Examples include the inverse problem of electrical resistivity tomography (ERT), shown below (**Figure 3**).

ERT is concerned with recovering, or characterizing sub-surface materials in terms of their electrical properties, which are recorded through electrodes. It operates very similarly to electrical impedance tomography (EIT), expect the difference being that it is subsurface. This has been also considered for learning permeability of subsurface flow in a range of different settings which can be found in the following papers [39, 40].

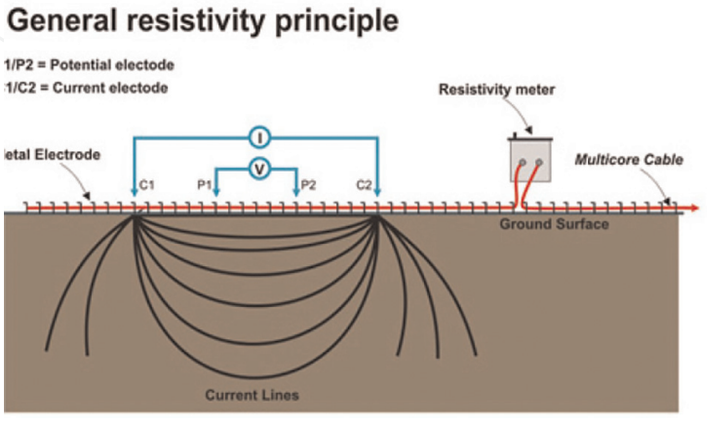

Figure 3.
Image depicting electrical resistivity tomography, where the the electric currents are recorded at the electrodes of the subsurface material.

Another interesting direction is related to walls, specifically quantifying uncertainty in thermo-physical properties of walls. This work was conducted by Iglesias et al. [41, 42]. Specifically the application is the inverse problem of recovering the thermodynamic property or temperature. Similar work related to the methodology used here has been used in resin transfer modeling [43], based on problems of moving boundaries. This is a difficult problem to model, however it provides a first step in doing so. Aside from these applications other particular applications include mineral exploration scattering problems, numerical climate models and others [44–46]. It is worth mentioning that, as of now, there is no official online software package for EKI in general. This is currently being developed, but we emphasize to the reader that the methodology presented, with the examples later, are not related to well known softwares that are available in `Matlab` or `Python`.

As a side remark, there are more directions beyond what is discussed above. Some others, without going into details, include developing hierarchical approaches, incorporating constrained optimization, and connections with data assimilation strategies [47–51].

6. Numerical experiments

In this section we provide some numerical experiments highlighting the performance of ensemble Kalman methods for inverse problems. Specifically we will consider EKI as discussed in Section 2. We will compare EKI with its regularized version of TEKI. Both these methodologies will be tested on on two motivating inverse problems arising in geophysical and atmospheric sciences, i.e. a Darcy flow partial differential equation and the Navier–Stokes Equation.

In order to assess a comparison, we will present three different figures. (i) The first being a reconstruction at the end of the iterative scheme; (ii) the error between the approximate solution and the ground truth, and (iii) the data misfit. The equations associated with each are given as.

- *Reconstruction through EKI*: $\frac{1}{J}\sum_{j=1}^{J} u_n^{(j)}$.
- *Relative error*: $\frac{\|u^\dagger - u\|_{L^2}^2}{\|u^\dagger\|_{L^2}}$.
- *Data misfit*: $\|\Gamma^{-1/2}\big(y - \mathcal{G}(u^\dagger)\|^2$.

6.1 Darcy flow

Our first model problem is an elliptic partial differential equation (PDE), which has numerous applications. Specifically one of them is subsurface flow in a porous medium. The forward problem is concerned with solving for the pressure $p \in H_0^1(\Omega)$, given the permeability $\kappa \in L^\infty(\Omega)$ and source function $f \in L^\infty(\Omega)$, where the PDE is given as

$$-\nabla \cdot (\kappa \nabla p) = f, \quad \in \Omega, \tag{47}$$

$$p = 0, \quad \text{on } \Omega. \tag{48}$$

such that we have prescribed Dirichlet boundary conditions, and $\Omega = [0, 1]^2 \subset \mathbb{R}^d$, for $d = 2$, is a Lipschitz domain. The inverse problem associated to solving p from (47) is the recovery of the permeability $\kappa \in L^\infty(\Omega)$, from noisy measurements of p, i.e.

$$y = \mathcal{G}(\kappa) + \eta, \quad \eta \sim \mathcal{N}(0, \Gamma), \tag{49}$$

where recalling that $\mathcal{G}(\kappa) = p$. We consider 64 equidistance observations within the domain, and on the boundary. To numerically solve (47) we employ a centered-finite difference method with a mesh size of $h = 1/100$. For our noisy observations we consider $\Gamma = \gamma I$, where $\gamma = 0.01$. We will use and compare EKI and TEKI, with an ensemble size of $J = 50$ for both methods. We will run both iterative schemes for $n = 24$ iterations. For our initial ensemble $\{u_0\}_{j=1}^{J}$ we consider modeling it as a Gaussian random field, i.e. $u \sim N(0, C)$, which can be done via the Karhunen-Loève expansion

$$u = \sum_{k \in \mathbb{Z}^+} \sqrt{\lambda_k} \phi_k \xi_k, \quad \xi_k \sim N(0, 1), \tag{50}$$

where (λ_k, ϕ_k) are the associated eigenvalues and eigenvectors of the covariance operator C. There are numerous choice of covariance functions one can take, however a popular choice is the Matérn covariance function, which provides much flexibility for modeling. For full details on various covariance functions, or operators, we refer to reader to [52]. The true unknown of interest is taken to be also a Gaussian random field, but one that is smoother than that of that of the initial ensemble.

Our first set of experiments are provided in **Figure 4** which shows the truth, the reconstruction from using EKI, and that of using TEKI. As we can observe, is it clear that both methodologies work well at learning the true unknown function. However it is clear that the TEKI induces a smoother reconstruction, which arises from the regularization. However, what is interesting is that if we analyze **Figure 5**, we notice

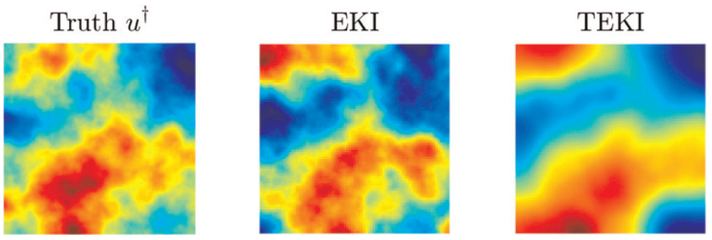

Figure 4.
Reconstruction plots for the Darcy flow PDE example. Left: Truth. Middle: EKI reconstruction. Right: TEKI reconstruction.

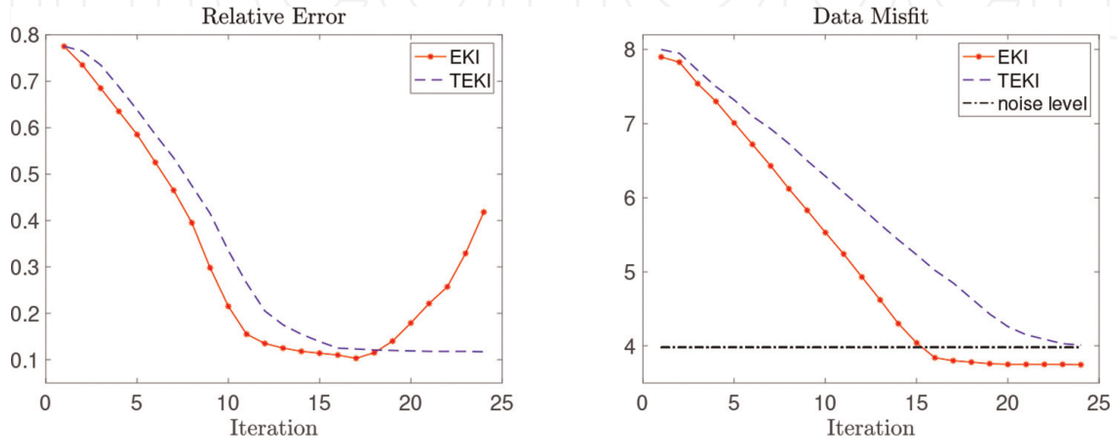

Figure 5.
Relative errors and data misfits for the Darcy flow PDE example. We compare EKI with TEKI.

that the relative error tends to diverge at the end with EKI, and this is due to the overfitting of data. A motivation behind TEKI is to alleviate this. This can be seen vividly as it tends to decrease, and for the data misfit, it remains within the noise level, which is given as

$$\text{noise level} = \|(y - G(u^\dagger))\| = \|\eta^\dagger\|. \tag{51}$$

6.2 Navier: stokes equation

Our final test problem is a well-known PDE model arising in numerical weather prediction which is the Navier–Stokes equation (NSE). We consider a 2D NSE defined on a torus $\mathbb{T}^2 = [0, 1]^2$ with periodic boundary conditions. The aim to estimate the velocity $v{:=}[0, \infty) \times \mathbb{T}^2 \to \mathbb{R}^2$ defined as a vector field from the scalar pressure field $p{:=}[0, \infty) \times \mathbb{T}^2 \to \mathbb{R}^2$. The NSE is given as

$$\partial_t v + (v \cdot \nabla)v + \nabla p - \nu \Delta v = f, \quad [0, \infty) \times \mathbb{T}^2, \tag{52}$$

$$\nabla \cdot v = 0, \quad [0, \infty) \times \mathbb{T}^2, \tag{53}$$

$$v = u, \quad \{0\} \times \mathbb{T}^2, \tag{54}$$

with initial condition (54) and zero flux (53). From (52) $f \in [0, \infty) \times \mathbb{T}$ corresponds to a volume forcing, ν is the associated viscosity of the fluid. For the NSE equation we consider a spectral Fourier solver for (52). The PDE is more challenging

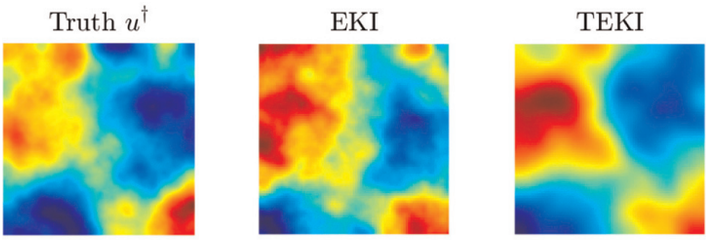

Figure 6.
Reconstruction plots for the NSE PDE example. Left: Truth. Middle: EKI reconstruction. Right: TEKI reconstruction.

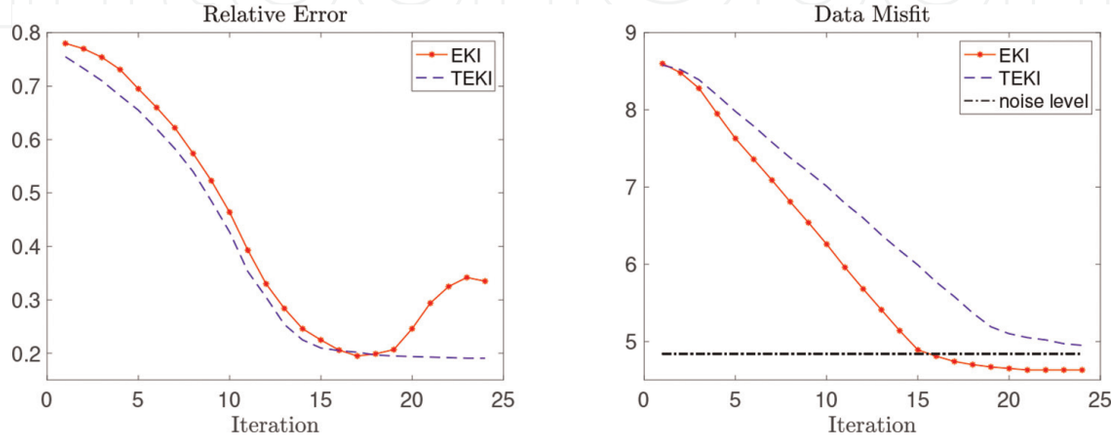

Figure 7.
Relative errors and data misfits for the NSE PDE example. We compare EKI with TEKI.

to invert than the previous example, therefore we take 100 point-wise observations. The setup is largely the same as the previous example, where we take an initial condition based on a Gaussian random field through the KL expansion (50). We will aim to recover the true underlying function $u^\dagger$ using both EKI and TEKI. The results are obtained from the experiments are presented in **Figures 6** and 7. A similar phenomenon shows, where the reconstructions work well, however there is an additional smoothness induced through the regularization in TEKI. Similarly, as we see with the relative errors and data misfit the overfitting of the data in the end for EKI. We note that this can be avoided depending on the prior form, its hyperparameters, the observations, and the noise. However we specify particular choices to demonstrate it can occur.

7. Conclusion

The ensemble Kalman filter (EnKF) is a simplistic, easy-to-implement and powerful algorithm. This has been particularly the case in numerous data assimilation applications for state estimation, which includes the likes of numerical weather prediction, geosciences and more recently machine learning. A major advantage of the method is that, unlike other filters such as the particle filter, it scales better in high dimensions, and can be significantly cheaper. In this chapter we consider the EnKF and its application to parameter estimation. Such a mathematical procedure also has similar applications to the ones states, where one can exploit such techniques for inverse problems. We provide a review and overview of some of the major contributions in this direction, where the resulting methodology is known as ensemble Kalman inversion (EKI), based largely on the work of Iglesias et al. [13]. We presented various avenues the field of EKI has taken such as regularization, extensions to sampling, and other areas. We demonstrated how EKI can perform on two numerical examples PDE examples.

The EKI methodology is one which builds very naturally from many different fields, which acts a strong motivation. For example being an optimizer, one can naturally apply optimization procedures, but also techniques from data assimilation and uncertainty quantification. As a result, this methodology naturally brings researchers from different fields working towards parameter estimation, and inverse problems. This synergy of areas will hopefully ensure new emerging directions within EKI, from a methodological, theoretical and application perspective.

Acknowledgements

This work was funded by KAUST baseline funding. The author thanks Simon Weissmann for helpful discussions, and for the use of some of the earlier figures, and information on EKS.

Abbreviations

EnKF	Ensemble Kalman filter
EKI	Ensemble Kalman inversion

EKS	Ensemble Kalman sampler
EIT	Electrical impedance tomography
ERT	Electrical restivitiy tomography
MNIST	Modified National Institute of Standards and Technology
KL	Kullback-Leibler
KF	Kalman filter
L96	Lorenz 96 model
LSF	Least squares functional
PDE	Partial differential equation
SDE	Stochastic differential equation

Author details

Neil K. Chada
King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

*Address all correspondence to: neilchada123@gmail.com

References

[1] Kaipio J, Somersalo E. Statistical and Computational Inverse Problems. New York: Springer Verlag; 2004

[2] Stuart AM. Inverse problems: A Bayesian perspective. Acta Numer. 2010; **19**:451-559

[3] Tarantola A. Inverse Problem Theory and Methods for Model Parameter Estimation. Philadelphia: SIAM; 1987

[4] Lorenc AC. Analysis methods for numerical weather prediction. Quarterly Journal of the Royal Meteorological Society. 1986;**112**(474):1177-1194

[5] Majda A, Wang X. Non-linear Dynamics and Statistical Theories for Basic Geophysical Flows. Cambridge University Press; 2006

[6] Oliver D, Reynolds AC. Liu N. 1st ed. Cambridge University Press: Inverse Theory for Petroleum Reservoir Characterization and History Matching; 2008

[7] Bucy RS. Nonlinear filtering theory. IEEE Transactions on Automatic Control. 1965;**10**(198):198

[8] Kalman RE. A new approach to linear filtering and prediction problems. Transactions ASME (Journal of Basic Engineering). 1960;**82**:35-45

[9] Bain A, Crisan D. Fundamentals of Stochastic Filtering. New York: Springer; 2009

[10] Law KJH, Stuart AM, Zygalakis K. Data assimilation: A mathematical introduction. In: Texts in Applied Mathematics. Springer; 2015

[11] Evensen G. Data Assimilation: The Ensemble Kalman Filter. Springer; 2009

[12] Evensen G. The ensemble Kalman filter: Theoretical formulation and practical implementation. Ocean dynamics. 2003;**53**(4):343-367

[13] Iglesias MA, Law KJH, Stuart AM. Ensemble Kalman methods for inverse problems. Inverse Problems. 2013;**29**

[14] Li G, Reynolds AC. Iterative ensemble Kalman filters for data assimilation. SPE Journal. 2009;**14**:496-505

[15] Lehtinen MS, Paivarinta L, Somersalo E. Linear inverse problems for generalised random variables. Inverse Problems. 1989;**5**(4):599-612

[16] Schillings C, Stuart AM. Analysis of the ensemble Kalman filter for inverse problems. SIAM Journal on Numerical Analysis. 2017;**55**(3):1264-1290

[17] Blomker D, Schillings C, Wacker P, Weissmann S. Well posedness and convergence analysis of the ensemble Kalman inversion. Inverse Problems. 2019;**35**(8):085007

[18] Blomker D, Schillings C, Wacker P, Weissmann S. Continuous time limit of the stochastic ensemble Kalman inversion: Strong convergence analysis. Preprint arXiv:2107.14508. 2021

[19] Benning M, Burger M. Modern regularization methods for inverse problems. Acta Numer. 2018;**27**:1-111

[20] Engl HW, Hanke K, Neubauer A. Regularization of Inverse Problems, Mathematics and its Applications. Vol. 375. Dordrecht: Kluwer Academic Publishers Group; 1996

[21] Iglesias MA. A regularising iterative ensemble Kalman method for PDE-

constrained inverse problems. Inverse Problems. 2016;**32**

[22] Iglesias MA, Yang Y. Adaptive regularisation for ensemble Kalman inversion with applications to non-destructive testing and imaging. arXiv preprint arXiv:2006.14980. 2020

[23] Chada NK, Tong XT, Stuart AM. Tikhonov regularization for ensemble Kalman inversion. SIAM Journal on Numerical Analysis. 2020;**58**(2): 1263-1294

[24] Chada NK, Tong XT. Convergence acceleration of ensemble Kalman inversion in nonlinear settings. Mathematics of Computation. 2022; **91**(335):1247-1280

[25] Lee Y. l_p regularization for ensemble Kalman inversion. SIAM Journal on Scientific Computing. 2021;**43**(5): 3417-3437

[26] Schneider T, Stuart AM, Wu J-L. Imposing sparsity within ensemble Kalman inversion. arXiv preprint, arXiv: 2007.06175. 2020

[27] Weissmann S, Chada NK, Schillings C, Tong XT. Adaptive Tikhonov strategies for stochastic ensemble Kalman inversion. Inverse Problems. 2022;**38**(4):045009

[28] Ernst OG, Sprungk B, Starkloff H-J. Analysis of the ensemble and polynomial chaos Kalman filters in Bayesian inverse problems. SIAM/ASA Journal on Uncertainty Quantification. 2015;**3**: 823-851

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

[30] Pavliotis G. Stochastic processes and applications: Diffusion processes, the Fokker–Planck and Langevin equations. In: Texts in Applied Mathematics. New York: Springer; 2014

[31] Kullback R, Leibler S. On information and sufficiency. Annals of Mathematical Statistics. 1951;**22**:79-86

[32] Markowich PA, Villani C. On the trend to equilibrium for the Fokker Planck equation: An interplay between physics and functional analysis, Physics and Functional Analysis, Matematica Contemporanea (SBM). 19, 1999

[33] Garbuno-Inigo A, Nüsken N, Reich S. Affine invariant interacting Langevin dynamics for Bayesian inference. SIAM Journal on Applied Dynamical Systems. 2020;**19**(3): 1633-1658

[34] Reich S, Weissmann S. Fokker–Planck particle Systems for Bayesian Inference: Computational approaches. SIAM Journal of Uncertainty Quantification. 2021;**9**(2):446-482

[35] Ding Z, Li Q. Ensemble Kalman sampler: Mean-field limit and convergence analysis. SIAM Journal on Mathematical Analysis. 2021;**53**(2): 1546-1578

[36] Kovachki NB, Stuart AM. Ensemble Kalman inversion: A derivative-free technique for machine learning tasks. Inverse Problems. 2019;**35**(9):095005

[37] Haber E, Lucka F, Ruthotto L. Never look back - A modified EnKF method and its application to the training of neural networks without back propagation. arxiv preprint. 1805;**08034**: 2018

[38] Guth PA, Schillings C, Weissmann S. Ensemble Kalman filter for neural

network based one-shot inversion. arXiv preprint. 2020

[39] Tso CM, Iglesias M, Wilkinson P, Kuras O, Chambers J, Binley A. Efficient multiscale imaging of subsurface resistivity with uncertainty quantification using ensemble Kalman inversion. Geophysical Journal International. 2021;**25**(2)

[40] Muir JB, Tsai VC. Geometric and level set tomography using ensemble Kalman inversion. Geophysical Journal International. 2020;**220**:967-980

[41] Iglesias MA, Sawlan Z, Scavino TR, Wood C. Bayesian inferences of the thermal properties of a wall using temperature and heat flux measurements. International Journal of Heat and Mass Transfer. 2018;**116**: 417-431

[42] De Simon L, Iglesias MA, Jones B, Wood C. Quantifying uncertainty in thermal properties of walls by means of Bayesian inversion. Energy and Buildings. 2018;**177**(2):177

[43] Iglesias M, Park M, Tretyakov MV. Bayesian inversion in resin transfer modelling. Inverse Problems. 2018;**34**(10)

[44] Sungkono S, Apriliani E, Saifuddin N, Fajriani F, Srigutomo W. Ensemble Kalman inversion for determining model parameter of self-potential data in the mineral exploration. In: Biswas A, editor. Self-Potential Method: Theoretical Modeling and Applications in Geosciences. Cham: Springer Geophysics. Springer; 2021

[45] Dunbar ORA, Garbuno-Inigo A, Schneider T, Stuart AM. Calibration and uncertainty quantification of convective parameters in an idealized GCM. Journal of Advances in Modeling Earth Systems. 2021

[46] Huang J, Li Z, Wang B. A Bayesian level set method for the shape reconstruction of inverse scattering problems in elasticity. Computers & Mathematics with Applications. 2021; **97**(1):18-27

[47] Albers DJ, Blancquart PA, Levine ME, Seylabi EE, Stuart AM. Ensemble Kalman methods with constraints. Inverse Problems. 2019; **35**(9):095007

[48] Chada NK, Chen Y, Sanz-Alonso D. Iterative ensemble Kalman methods: A unified perspective with some new variants. Foundations of Data Science. 2021;**3**(3):331-369

[49] Chada NK, Iglesias MA, Roininen L, Stuart AM. Parameterizations for ensemble Kalman inversion. Inverse Problems. 2018;**34**

[50] Chada NK, Schillings C, Weissmann S. On the incorporation of box-constraints for ensemble Kalman inversion. Foundations of Data Science. 2019;**1**(4):433-456

[51] Tong XT, Morzfield M. Localized ensemble Kalman inversion. arXiv preprint arXiv:2201.10821. 2022

[52] Lord G, Powell CE, Shardlow T. An Introduction to Computational Stochastic PDEs. Cambridge Texts in Applied Mathematics; 2014