PAPER

Ensemble Kalman methods for inverse problems

View the article online for updates and enhancements.

Related content

- A regularizing iterative ensemble Kalman method for PDE-constrained inverse problems
  Marco A Iglesias

- An IRBS for data inversion in reservoir modeling
  Marco A Iglesias and Clint Dawson

- MAP estimators and their consistency in Bayesian nonparametric inverse problems
  M Dashti, K J H Law, A M Stuart et al.

Recent citations

- Turbulence Modeling in the Age of Data
  Karthik Duraisamy *et al*

- Data-Augmented Modeling of Intracranial Pressure
  Jian-Xun Wang *et al*

- An ensemble Kalman filter approach based on operator splitting for solving nonlinear Hammerstein type ill-posed operator equations
  Xiao-Mei Yang and Zhi-Liang Deng

IOP ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Ensemble Kalman methods for inverse problems

Marco A Iglesias, Kody J H Law and Andrew M Stuart

Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK

E-mail: M.A.Iglesias-Hernandez@warwick.ac.uk, K.J.H.Law@warwick.ac.uk and A.M.Stuart@warwick.ac.uk

Abstract

The ensemble Kalman filter (EnKF) was introduced by Evensen in 1994 (Evensen 1994 *J. Geophys. Res.* **99** 10143–62) as a novel method for *data assimilation*: state estimation for noisily observed time-dependent problems. Since that time it has had enormous impact in many application domains because of its robustness and ease of implementation, and numerical evidence of its accuracy. In this paper we propose the application of an iterative ensemble Kalman method for the solution of a wide class of *inverse problems*. In this context we show that the estimate of the unknown function that we obtain with the ensemble Kalman method lies in a subspace $\mathcal{A}$ spanned by the initial ensemble. Hence the resulting error may be bounded above by the error found from the best approximation in this subspace. We provide numerical experiments which compare the error incurred by the ensemble Kalman method for inverse problems with the error of the best approximation in $\mathcal{A}$, and with variants on traditional least-squares approaches, restricted to the subspace $\mathcal{A}$. In so doing we demonstrate that the ensemble Kalman method for inverse problems provides a derivative-free optimization method with comparable accuracy to that achieved by traditional least-squares approaches. Furthermore, we also demonstrate that the accuracy is of the same order of magnitude as that achieved by the best approximation. Three examples are used to demonstrate these assertions: inversion of a compact linear operator; inversion of piezometric head to determine hydraulic conductivity in a Darcy model of groundwater flow; and inversion of Eulerian velocity measurements at positive times to determine the initial condition in an incompressible fluid.

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

1. Introduction

Since its introduction in [11], the ensemble Kalman filter (EnKF) has had enormous impact on applications of data assimilation to state and parameter estimation, and in particular in oceanography [13], reservoir modelling [1] and weather forecasting [18]; the books [12, 22, 30] give further details and references to applications in these fields. Multiple variants

of EnKF for state and parameter estimation in dynamic systems are available in the literature [12, 22, 1]. In essence, all those techniques use an ensemble of states and parameters that is sequentially updated by means of the Kalman formula which blends the model and data available at a given time.

Motivated by ensemble Kalman-based approaches, in this paper we propose the application of an iterative ensemble Kalman method for the solution of inverse problems of finding u given observations of the form

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

where $\mathcal{G} : X \rightarrow Y$ is the *forward response operator* mapping the unknown u to the response/observation space. X and Y are Hilbert spaces, $\eta \in Y$ is a noise and $y \in Y$ the observed data. We assume that η is an unknown realization of a mean zero random variable whose covariance Γ is known to us. We are particularly interested in the case where $\mathcal{G}$ is the forward response that arises from physical systems described by the solution of a PDE system. It is important to note, that in the abstract formulation of the inverse problem (1), both static and dynamic problems are considered in the same manner. For dynamic problems, the left-hand side of (1) corresponds to all available observations which are, in turn, collected during a fixed time window contained in the time interval used for the underlying PDE formulation. Inverse problems of the type describe above are often ill-posed and their solution requires some sort of regularization [10]. For the present work, regularization is introduced by incorporating prior knowledge of u in the form of a finite dimensional (and hence compact) set $\mathcal{A}$ where the solution to (1) is sought (see [20] chapter 2). The definition of the space $\mathcal{A}$ will be key in the formulation and properties of the ensemble method proposed for the solution of the inverse problem.

In order to solve the inverse problem described above, artificial dynamics based on state augmentation are constructed. The state augmentation approach, typical for joint state and parameter estimation in the context of EnKF [2], can be applied in our abstract framework by constructing the space $Z = X \times Y$, and the mapping $\Xi : Z \rightarrow Z$ by

$$\Xi(z) = \begin{pmatrix} u \\ \mathcal{G}(u) \end{pmatrix} \quad \text{for} \quad z = \begin{pmatrix} u \\ p \end{pmatrix} \in Z.$$

We define artificial dynamics by

$$z_{n+1} = \Xi(z_n). \tag{2}$$

Let us assume that data related to the artificial dynamics has the form

$$y_{n+1} = Hz_{n+1} + \eta_{n+1}, \tag{3}$$

where the projection operator $H : Z \rightarrow Y$ is defined by $H = (0, I)$ and $\{\eta_n\}_{n \in \mathbb{Z}^+}$ is an i.i.d. sequence with $\eta_1 \sim N(0, \Gamma)$ and Γ defined above. In this paper we propose the application of the EnKF approach for state and parameter estimation for the artificial dynamical system (2). EnKF uses an ensemble of particles that, at each iteration, is updated by combining the model (2) with observational data via the standard Kalman update formula. In order to generate the data $\{y_n\}_{n \in \mathbb{Z}^+}$ required for the EnKF that we will apply to (2), (3) we perturb the single instance of the given observed data y from (1) by independent realizations from the Gaussian random variable $N(0, \Gamma)$. We reemphasize that the iteration index n in (2), (3) is an artificial time; in the case of a dynamic inverse problem, real time is contained in the abstract formulation of $\mathcal{G}$ and is not related to n.

While the objective of standard EnKF approaches is to approximate, via an ensemble, statistical properties of a distribution conditioned to observations [12], here the objective is to study a deterministic iterative scheme that aims at approximating the solution of the

inverse problem (1) in the set $\mathcal{A}$. We employ randomization of the single instance of the data y given by (1) purely as a method to move around the space $\mathcal{A}$ in order to find improved approximations. More precisely, we construct an ensemble of interacting particles $\{z_n^{(j)}\}_{j=1}^J$ from which an estimate of the unknown is defined by

$$u_n \equiv \frac{1}{J} \sum_{j=1}^J u_n^{(j+1)} = \frac{1}{J} \sum_{j=1}^J H^\perp z_n^{(j)} \tag{4}$$

where $H^\perp : Z \to X$ is the projection operator defined by $H^\perp = (I, 0)$. We will show that, for all $n \in \mathbb{N}$, the ensemble $\{u_n^j\}_{j=1}^J$ remains in the set $\mathcal{A}$ and, by means of numerical examples we investigate the properties of u_n at approximating, in the compact set $\mathcal{A}$, the true unknown $u^\dagger \in X$ which underlies the data. That is we assume that the data y is given by

$$y = \mathcal{G}(u^\dagger) + \eta^\dagger \tag{5}$$

for some noise $\eta^\dagger \in Y$.

It is worth pointing out that iterative ensemble methods have been previously developed for the solution of inverse problems [6, 26]. These methods use ensemble approximations of the derivatives appearing in Newton-type methods applied to the minimization of Tikhonov regularized least-squares functionals. In contrast, we propose an ensemble Kalman method as a derivative-free iterative regularization methodology for solving inverse problems in the abstract framework defined above. The purpose of the subsequent analysis is threefold: (i) to demonstrate that the novel non-standard perspective of the iterative ensemble Kalman method is a generic tool for solving inverse problems; (ii) to provide some basic analysis of the properties of this algorithm for inversion; (iii) to demonstrate numerically that the method can be effective on a wide range of applications.

In section 2 we introduce the iterative ensemble Kalman method for the solution to inverse problems. The space $\mathcal{A}$ where a regularized solution of the inverse problem is sought is defined as the linear subspace generated by the initial ensemble members used for the iterative scheme. These, for the applications under consideration, can be generated from prior knowledge available in terms of a prior probability measure. The well-posedness of the algorithm is ensured by theorem 2.1 where we prove that the method produces an approximation which lies in the subspace $\mathcal{A}$. In other words, the approximation provided by the algorithm lies in the subspace spanned by the initial ensemble members, a fact observed for a specific sequential implementation of the EnKF in [26]. It is well known that the analysis step of the EnKF preserves the subspace spanned by the ensemble [14]; we show that for the artificial dynamics (2), (3) the prediction step also preserves the subspace spanned by the ensemble leading to theorem 2.1. In corollary 2.3 we use theorem 2.1 to give a lower bound on the achievable approximation error of the ensemble Kalman algorithm. We then describe two algorithms which we will use to evaluate the ensemble Kalman methodology. The first is the least squares problem restricted to the subspace $\mathcal{A}$. The second is the best approximation of the truth in the subspace $\mathcal{A}$. The best approximation is, of course, not an implementable method as the truth is not known; however it provides an important lower bound on the achievable error of the ensemble Kalman method for synthetic experiments and hence has a key conceptual role. At the end of section 2 we discuss the links between the ensemble Kalman algorithm and the Tikhonov–Phillips regularized least squares solutions for the case of forward linear operators.

Section 3 contains numerical experiments which illustrate the ideas in this paper on a linear inverse problem. The forward operator is a compact operator found from inverting the negative Laplacian plus identity with homogeneous boundary conditions. In section 4 we numerically study the groundwater flow inverse problem of determining hydraulic conductivity from

piezometric head measurement in an elliptic Darcy flow model. Section 5 contains numerical results concerning the problem of determining the initial condition for the velocity in a Navier–Stokes model of an incompressible fluid; the observed data are pointwise (Eulerian) measurements of the velocity field. Conclusion and final remarks are presented in section 6.

The numerical results in this paper all demonstrate that the iterative ensemble Kalman method for inversion is a derivative-free regularized optimization technique which produces numerical results similar in accuracy to those found from least-squares based methods in the same subspace $\mathcal{A}$. Furthermore, the three examples serve to illustrate the point that the method offers considerable flexibility through the choice of initial ensemble, and hence the subspace $\mathcal{A}$ in which it produces an approximation. In particular, for the linear and Darcy inverse problems, we make two choices of initial ensemble: (i) draws from a prior Gaussian measure and (ii) the Karhunen–Loéve basis functions of the centred Gaussian measure found by shifting the prior by its mean. For the Navier–Stokes inverse problem the initial ensemble is also chosen to comprise randomly drawn functions on the attractor of the dynamical system.

2. An iterative ensemble Kalman method for inverse problems

2.1. Preliminaries

In the following, we use $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, etc as the inner-product and norm on both X and Y, and it will be clear from the context which space is intended. Let $B^{-1} : D(B^{-1}) \subset X \to X$ be a densely-defined unbounded self-adjoint operator with compact resolvent. Let us denote by $\{\lambda_j\}_{j=1}^{\infty}$ and $\{\phi_j\}_{j=1}^{\infty}$ the corresponding eigenvalues and eigenfunctions of B^{-1}. From standard theory it follows that

$$D(B^{-1}) = \left\{ u \in X \,\Big|\, \sum_{j=1}^{\infty} u_j^2 \lambda_j^2 < \infty \right\} \tag{6}$$

and that B^{-1} has the following spectral representation: $B^{-1}u = \sum_{j=1}^{\infty} \lambda_j u_j \phi_j$. We can additionally define the set

$$D(B^{-1/2}) = \left\{ u \in X \,\Big|\, \sum_{j=1}^{\infty} u_j^2 \lambda_j < \infty \right\} \tag{7}$$

and consider the densely-defined operator $B^{-1/2} : D(B^{-1/2}) \to X$ such that $B^{-1/2}u = \sum_{j=1}^{\infty} \lambda_j^{1/2} u_j \phi_j$. We also find it useful to define $\| \cdot \|_B \equiv \|B^{-1/2} \cdot \|$.

2.2. The initial ensemble

The ensemble Kalman method uses an ensemble of particles $\{z_n^{(j)}\}_{j=1}^{J}$ which, at the n iteration level, is updated by combining the artificial dynamics (2) with artificial data y_n obtained from perturbing our original data (1) in order to obtain a new ensemble $\{z_{n+1}^{(j)}\}_{j=1}^{J}$ from which the estimate of the unknown (4) is computed. The scheme requires an initial (or first guess) ensemble of particles $\{z_0^{(j)}\}_{j=1}^{J}$ which will be iteratively updated with the EnKF method described below. The ensemble $\{z_0^{(j)}\}_{j=1}^{J}$ can be defined by constructing an ensemble $\{\psi^{(j)}\}_{j=1}^{J}$ in X. Once $\{\psi^{(j)}\}_{j=1}^{J}$ is specified, we can simply define

$$z_0^{(j)} = \begin{pmatrix} \psi^{(j)} \\ \mathcal{G}(\psi^{(j)}) \end{pmatrix},$$

and so our first guess of the iterative scheme u_0 (see expression (4)) is simply the mean of the initial ensemble in the space of the unknown. The construction of the initial ensemble $\{\psi^{(j)}\}_{j=1}^{J}$

is in turn related to the definition of the space $\mathcal{A}$ where the solution to the inverse problem is sought. Clearly, $\psi^{(j)}$ must belong to the compact set $\mathcal{A}$ which regularizes the inverse problem by incorporating prior knowledge. For the applications described in sections 3, 4 and 5, we assume that prior knowledge is available in terms of a prior probability measure that we denote by μ_0. Given this prior distribution, we construct the initial ensemble $\{\psi^{(j)}\}_{j=1}^{J}$ defined as $\psi^{(j)} \sim \mu_0$ i.i.d. for some $J < \infty$. Then, for consistency we define

$$\mathcal{A} = \mathrm{span}\{\psi^{(j)}\}_{j=1}^{J} \tag{8}$$

comprised of the initial ensemble. Note that if μ_0 is Gaussian $N(\bar{u}, C)$, we may additionally consider $\mathcal{A}$ with defined $\psi^{(j)} = \bar{u} + \sqrt{\lambda_j}\phi_j$ where (λ_j, ϕ_j) denote eigenvalue/eigenvector pairs of C, in descending order by eigenvalue—this is the Karhunen–Loéve basis.

For the experiments of sections 3 and 4, Gaussian priors are considered. For the Navier–Stokes example of section 5, the prior will be the empirical measure supported on the attractor. For the latter case an empirical covariance will be used to construct the KL basis. In summary, the proposed approach for solving inverse problems will be tested with an initial prior ensemble generated from (a) $\psi^{(j)} \sim \mu_0$ i.i.d. and all algorithms using this choice of $\mathcal{A}$ will bear the subscript R for random; (b) $\psi^{(j)} = \bar{u} + \sqrt{\lambda_j}\phi_j$, $j \leqslant J$, and all algorithms using this method will bear the subscript KL for Karhunen–Loéve.

We emphasize that the initial ensemble $\{\psi^{(j)}\}_{j=1}^{J}$ and therefore the definition of $\mathcal{A}$ is a design parameter aiming at incorporating prior knowledge relevant to the application. For example, in the case not considered here, where no underlying prior probability distribution is prescribed, the initial ensemble can be defined in terms of a truncated basis $\{\psi^{(j)}\}_{j=1}^{J}$ of X. Regardless of how the initial ensemble is chosen, the proposed approach will then find a solution to the inverse problem in the subspace $\mathcal{A}$ defined by (8).

2.3. Iterative ensemble Kalman method for inverse problems

The iterative algorithm that we propose for solution of the inverse problem (1) is the following:

Algorithm 1. *Iterative ensemble method for inverse problems.*
   Let $\{z_0^{(j)}\}_{j=1}^{J}$ be the initial ensemble.
   For $n = 1, \ldots$

(1) **Prediction step.** *Propagate, under the artificial dynamics (2), the ensemble of particles*

$$\widehat{z}_{n+1}^{(j)} = \Xi(z_n^{(j)}). \tag{9}$$

From this ensemble we define a sample mean and covariance as follows:

$$\bar{z}_{n+1} = \frac{1}{J} \sum_{j=1}^{J} \widehat{z}_{n+1}^{(j)} \tag{10}$$

$$C_{n+1} = \frac{1}{J} \sum_{j=1}^{J} \widehat{z}_{n+1}^{(j)} (\widehat{z}_{n+1}^{(j)})^T - \bar{z}_{n+1}\bar{z}_{n+1}^T. \tag{11}$$

(2) **Analysis step.** *Define the Kalman gain K_n by*

$$K_{n+1} = C_{n+1}H^*(HC_{n+1}H^* + \Gamma)^{-1}, \tag{12}$$

where H^* is the adjoint operator of H. Update each ensemble member as follows

$$z_{n+1}^{(j)} = \widehat{z}_{n+1}^{(j)} + K_{n+1}\big(y_{n+1}^{(j)} - H\widehat{z}_{n+1}^{(j)}\big) \tag{13}$$

$$= (I - K_{n+1}H)\widehat{z}_{n+1}^{(j)} + K_{n+1}y_{n+1}^{(j)}, \tag{14}$$

where

$$y_{n+1}^{(j)} = y + \eta_{n+1}^{(j)}. \tag{15}$$

and the $\eta_{n+1}^{(j)}$ are an i.i.d collections of vectors indexed by (j, n) with $\eta_1^{(1)} \sim N(0, \Gamma)$.

(3) Compute the mean of the parameter update

$$u_{n+1} \equiv \frac{1}{J} \sum_{j=1}^{J} H^{\perp} z_{n+1}^{(j)} = \frac{1}{J} \sum_{j=1}^{J} u_{n+1}^{(j)} \tag{16}$$

and check for convergence (see discussion below).

Each iteration of the ensemble Kalman algorithm breaks into two parts, a *prediction step* and an *analysis step*. The prediction step maps the current ensemble of particles into the data space, and thus introduces information about the forward model. The analysis step makes comparisons of the mapped ensemble, in the data space, with the data, or with versions of the data perturbed with noise; it is at this stage that the ensemble is modified in an attempt to better match the data. As noted earlier, we generate artificial data (3) consistent with the artificial dynamics (2) by perturbing the observed data (1). More precisely, data is perturbed according to (15) with noise consistent with the distribution assumed on the noise η in the inverse problem (1). Perturbing the noise in the standard EnKF methods is typically used to properly capture statistical properties. However, in the present application, we are merely interested in a deterministic estimation of the inverse problem. Nonetheless, numerical results (not shown) without perturbing the noise (e.g. with $\eta_{n+1}^{(j)} = 0$ in (15)) gave rise to less accurate solutions than the ones obtained when the data was perturbed according to (15). The added noise presumably helps the algorithm explore the approximation space and hence find a better approximation within it.

The proper termination of iterative regularization techniques [23] is essential for the regularization of ill-posed inverse problems. The discrepancy principle, for example, provides a stopping criterion that will ensure the convergence and regularizing properties of iterative techniques such as the Landweber iteration, Levenberg–Marquardt and the iteratively regularized Gauss–Newton [23]. A complete analysis of the convergence and regularizing properties of the iterative ensemble Kalman method is beyond the scope of this paper. Nonetheless, numerical experiments suggest that the discrepancy principle can be a useful stopping criterion for the ensemble Kalman algorithm; in particular these experiments indicate that the criterion ensures the stable computation of an approximation to the inverse problem (1). Concretely, according to the discrepancy principle, the ensemble Kalman method is terminated for the first n such that $||y - \mathcal{G}(u_n)||_{\Gamma} \leqslant \tau ||\eta^{\dagger}||_{\Gamma}$ for some $\tau > 1$ and where $\eta^{\dagger}$ is the noise associated to the data (5).

We show in the next section that the entire ensemble at each step of the algorithm lies in the set $\mathcal{A}$ defined by (8) and that, hence, the parameter estimate (16) lies in the set $\mathcal{A}$.

2.4. Properties of the iterative ensemble method

We wish to show that each estimate u_{n+1} of the ensemble Kalman method, given by formula (16), is a linear of combinations of the initial ensemble; this linear combination has nonlinear weights reflecting the observed data. From the definition (8) this observation implies that $u_{n+1} \in \mathcal{A}$. First, note that all the vectors and operators involved have block structure inherited from the structure of the space $Z = X \times Y$. For example we have

$$\widehat{z}_{n+1}^{(j)} = \begin{pmatrix} \widehat{u}_{n+1}^{(j)} \\ \widehat{p}_{n+1}^{(j)} \end{pmatrix} = \begin{pmatrix} u_n^{(j)} \\ \mathcal{G}(u_n^{(j)}) \end{pmatrix}, \quad z_n^{(j)} = \begin{pmatrix} u_n^{(j)} \\ p_n^{(j)} \end{pmatrix}.$$

We also have

$$\bar{z}_n = \begin{pmatrix} \bar{u}_n \\ \bar{p}_n \end{pmatrix}, \quad C_n = \begin{pmatrix} C_n^{uu} & C_n^{up} \\ (C_n^{up})^T & C_n^{pp} \end{pmatrix}.$$

The vectors $\bar{u}_n$ and $\bar{p}_n$ are given by

$$\bar{u}_n = \frac{1}{J} \sum_{j=1}^{J} \widehat{u}_n^{(j)} = \frac{1}{J} \sum_{j=1}^{J} u_{n-1}^{(j)} \tag{17}$$

$$\bar{p}_n = \frac{1}{J} \sum_{j=1}^{J} \widehat{p}_n^{(j)} = \frac{1}{J} \sum_{j=1}^{J} \mathcal{G}(u_{n-1}^{(j)}). \tag{18}$$

The blocks within C_n are given by

$$C_n^{uu} = \frac{1}{J} \sum_{j=1}^{J} \widehat{u}_n^{(j)} (\widehat{u}_n^{(j)})^T - \bar{u}_n \bar{u}_n^T, \tag{19}$$

$$C_n^{up} = \frac{1}{J} \sum_{j=1}^{J} \widehat{u}^{(j)} (\widehat{p}^{(j)})^T - \bar{u}_n \bar{p}_n^T, \tag{20}$$

$$C_n^{pp} = \frac{1}{J} \sum_{j=1}^{J} \widehat{p}^{(j)} (\widehat{p}^{(j)})^T - \bar{p}_n \bar{p}_n^T. \tag{21}$$

As we indicated before, it is well known that the analysis step of the EnKF provides an updated ensemble which is in the linear span of the forecast ensemble [14]. In the context of our iterative method for inverse problems the forecast ensemble itself is in the linear span of the preceding analysis ensemble, when projected into the parameter coordinate. Combining these two observations shows that the ensemble parameter estimate lies in the linear span of the initial ensemble.

Theorem 2.1. *For every* $(n, j) \in \mathbb{N} \times \{1, \ldots, J\}$ *we have* $u_{n+1}^{(j)} \in \mathcal{A}$ *and hence* $u_{n+1} \in \mathcal{A}$ *for all* $n \in \mathbb{N}$.

Proof. This is a straightforward induction, using the properties of the update formulae. Clearly the statement is true for $n = 0$. Assume that it is true for n. The operators K_n have particular structure inherited from the form of H. In concrete, we have

$$K_n = \begin{pmatrix} C_n^{up}(C_n^{pp} + \Gamma)^{-1} \\ C_n^{pp}(C_n^{pp} + \Gamma)^{-1} \end{pmatrix}.$$

Note that

$$C_n^{up} = \frac{1}{J} \sum_{j=1}^{J} \widehat{u}_n^{(j)} (\widetilde{p}_n^{(j)})^T, \qquad C_n^{pp} = \frac{1}{J} \sum_{j=1}^{J} \widehat{p}_n^{(j)} (\widetilde{p}_n^{(j)})^T \tag{22}$$

where

$$\widetilde{p}_n^{(j)} = \widehat{p}_n^{(j)} - \frac{1}{J} \sum_{\ell=1}^{J} \widehat{p}_n^{(\ell)}. \tag{23}$$

Recall that, from the structure of the map Ξ, we have

$$\widehat{u}_{n+1}^{(j)} = u_n^{(j)}, \quad \widehat{p}_{n+1}^{(j)} = \mathcal{G}(u_n^{(j)}) \tag{24}$$

and from the definition of H we get that

$$K_n H = \begin{pmatrix} 0 & C_n^{up}(C_n^{pp} + \Gamma)^{-1} \\ 0 & C_n^{pp}(C_n^{pp} + \Gamma)^{-1} \end{pmatrix}.$$

Using these facts in (16) we deduce that the update equations are

$$u_{n+1}^{(j)} = u_n^{(j)} + C_{n+1}^{up}(C_{n+1}^{pp} + \Gamma)^{-1}(y_{n+1}^{(j)} - \mathcal{G}(u_n^{(j)})) \tag{25}$$

$$p_{n+1}^{(j)} = \mathcal{G}(u_n^{(j)}) + C_{n+1}^{pp}(C_{n+1}^{pp} + \Gamma)^{-1}(y_{n+1}^{(j)} - \mathcal{G}(u_n^{(j)})). \tag{26}$$

If we define

$$d_{n+1}^{(j)} = (C_{n+1}^{pp} + \Gamma)^{-1}\left(y_{n+1}^{(j)} - \mathcal{G}(u_n^{(j)})\right)$$

then the update formula (25) for the unknown u may be written as

$$u_{n+1}^{(j)} = u_n^{(j)} + \frac{1}{J}\sum_{k=1}^{J}\langle \widetilde{p}_{n+1}^{(k)}, d_{n+1}^{(j)}\rangle \widehat{u}_{n+1}^{(k)} \tag{27}$$

$$= u_n^{(j)} + \frac{1}{J}\sum_{k=1}^{J}\langle \widetilde{p}_{n+1}^{(k)}, d_{n+1}^{(j)}\rangle u_n^{(k)}. \tag{28}$$

In view of the inductive hypothesis at step n, this demonstrates that $u_{n+1}^{(j)} \in \mathcal{A}$ for all $j \in \{1, \ldots, J\}$. Then, from (16) and (8) it follows that $u_{n+1} \in \mathcal{A}$ which finalizes the proof. $\square$

Remark 2.2. The proof demonstrates that the solution at step n is simply a linear combination of the original samples; however the coefficients in the linear combination depend nonlinearly on the process. Note also that the update formula (26) for the system response may be written as

$$p_{n+1}^{(j)} = \mathcal{G}(u_n^{(j)}) + \frac{1}{J}\sum_{k=1}^{J}\langle \widetilde{p}_{n+1}^{(k)}, d_{n+1}^{(j)}\rangle \widehat{p}_{n+1}^{(k)} \tag{29}$$

$$= \mathcal{G}(u_n^{(j)}) + \frac{1}{J}\sum_{k=1}^{J}\langle \widetilde{p}_{n+1}^{(k)}, d_{n+1}^{(j)}\rangle \mathcal{G}(u_n^{(k)}). \tag{30}$$

We have the following lower bound for the accuracy of the estimates produced by the ensemble Kalman algorithm:

Corollary 2.3. *The error between the estimate of u at time n and the truth $u^{\dagger}$ satisfies*

$$\|u_n - u^{\dagger}\| \geqslant \inf_{v \in \mathcal{A}} \|v - u^{\dagger}\|.$$

2.5. Evaluating the performance of the iterative Kalman method for inverse problems

With the iterative EnKF-based algorithm previously described we aim at finding solutions u_{EnKF} of the inverse problem (1) in the subspace $\mathcal{A}$. We now wish to evaluate the performance of u_{EnKF} at recovering the truth $u^{\dagger}$ defined in (5). Central to this task is the misfit functional

$$\Phi(u) = \|y - \mathcal{G}(u)\|_{\Gamma}^2, \tag{31}$$

which should be minimized in some sense. As we indicated earlier, the inversion of $\mathcal{G}$ is ill-posed and so minimization of Φ over the whole space X is not possible. As for the Ensemble Kalman method, we choose to regularize the least squares methods by incorporating prior

knowledge via minimization of Φ over the compact set $\mathcal{A} \in X$. Then, we compute the *least squares* solution

$$u_{\mathrm{LS}} = \mathrm{argmin}_{u \in \mathcal{A}} ||y - \mathcal{G}(u)||_{\Gamma}^2 \qquad (32)$$

or generalizations to include truncated Newton-CG iterative methods [16] as well as Tikhonov–Phillips regularization [33], based on $\mu_0 = N(\bar{u}, C)$. In other words, we consider $u_{\mathrm{TP}} = \mathrm{argmin}_{u \in \mathcal{A}} ||y - \mathcal{G}(u)||_{\Gamma}^2 + ||u - \bar{u}||_C^2$.

We recall our definition to the true solution $u^\dagger$ (see equation (5)) to the inverse problem (1). Given that both the ensemble Kalman method and the least-squares solution (32) provide, in the subspace $\mathcal{A}$, some approximation to the truth, it is then natural to consider the *best approximation* to the truth in $\mathcal{A}$. In other words,

$$u_{\mathrm{BA}} = \mathrm{argmin}_{u \in \mathcal{A}} ||u - u^\dagger||_{\Gamma}^2. \qquad (33)$$

Best approximation properties of ensemble methods is discussed in a different context in [31].

In the experiments below we compare the accuracy in the approximations obtained with the ensemble Kalman method with respect to the least square solution and the best approximation. The latter, however, is done only for the sake of assessing the performance of the technique with synthetic experiments for which the truth is known.

In typical applications of EnKF-based methods the amount of observational data is usually much smaller than the dimensions of the space X for the unknown. Therefore, the matrix computations for the construction of the Kalman gain matrix (12) are often negligible compared to the computational cost of evaluating, for each ensemble member, the forward model (2) at each iteration of the scheme. In this case, the computational cost of the ensemble Kalman algorithm is dominated by the size of the ensemble multiplied by the total number of iterations. While the computational cost of EnKF is quite standard, the cost of solving the least-squares problem depends on the particular implementation used for the application under consideration. Since the optimality of the implementation of least-squares problems is beyond the scope of our work, we do not assess the computational efficiency of the ensemble method with respect to the optimization methods used for the solution of the least-squares problem.

2.6. Connection between regularized least-squares and the iterative ensemble Kalman method for linear inverse problems

It is instructive to consider the case where $\mathcal{G}(u) = Gu$ for some linear operator $G : X \to Y$ as this will enable us to make links between the iterative ensemble Kalman method and the standard regularized least squares problems. Let C^{-1} be an operator like the one defined in section 2.1 with $D(C^{-1/2})$ defined analogously to (7). Consider the Tikhonov–Phillips regularized functional

$$I(u) = \|y - Gu\|_{\Gamma}^2 + \|u - \bar{u}\|_C^2. \qquad (34)$$

If $\|y - G \cdot \|_{\Gamma}^2$ is continuous on $D(C^{-1/2})$ then $I(u)$ is weakly lower semicontinuous in X and hence the unique minimizer $u_{\mathrm{TP}} = \mathrm{argmin}_{u \in D(C^{-1/2})} I(u)$ is attained in $D(C^{-1/2})$; indeed minimizing sequences converge strongly in $D(C^{-1/2})$ along a subsequence. The existence of a minimizer follows from the standard theory of calculus of variations, see theorem 1, part 2, in [9]. Uniqueness follows from the quadratic nature of I. The fact that minimizing sequences converge strongly uses an argument from theorem II.2.1 in [24], as detailed in theorem 2.7 of [7]. We note that minimization of I given by (34) has a statistical interpretation as the maximum *a posteriori* (MAP) estimator for Bayesian solution to the inverse problem with Gaussian prior $N(\bar{u}, C)$ on u [21]. Furthermore the minimizer of (34) can be computed

explicitly under a variety of different assumptions on the linear operators C, G and Γ. In all cases the formal expression for the solution can then be written [25, 28]

$$u_{\text{TP}} = \bar{u} + CG^*(GCG^* + \Gamma)^{-1}(y - G\bar{u}). \tag{35}$$

The purpose of this subsection is now to show the connection between the iterative Kalman method that we introduce in this paper and the regularized least-squares solution (35) for the solution of linear inverse problems that arise from linear forward operators. Recall the set $\mathcal{A}$ defined by (8). We consider the case where $\psi^{(j)} \sim \mu_0$. Let $\mu_0 = N(\bar{u}, C)$ and consider $n = 0$ in the algorithm of the previous section. Let us define the prior ensemble mean and covariance

$$m_J \equiv \frac{1}{J} \sum_{j=1}^{J} u_0^{(j)}, \qquad C_J \equiv \frac{1}{J-1} \sum_{j=1}^{J} \left(u_1^{(j)} - m_J\right)\left(u_1^{(j)} - m_J\right)^T. \tag{36}$$

From (22)–(24) for $n = 0$ and $\mathcal{G}(u) = Gu$ we have

$$
\begin{aligned}
C_1^{up} &= \frac{1}{J} \sum_{j=1}^{J} \widehat{u}_0^{(j)} \big(\widetilde{p}_0^{(j)}\big)^T = \frac{1}{J} \sum_{j=1}^{J} u_0^{(j)} \left(Gu_0^{(j)} - \frac{1}{J} \sum_{\ell=1}^{J} Gu_0^{(l)}\right)^T = \frac{1}{J} \sum_{j=1}^{J} u_0^{(j)} \big(Gu_0^{(j)} - Gm_J\big)^T \\
&= \frac{1}{J} \sum_{j=1}^{J} \big(u_0^{(j)} - m_J\big)\big(Gu_0^{(j)} - Gm_J\big)^T = \frac{1}{J} \sum_{j=1}^{J} \big(u_0^{(j)} - m_J\big)\big(u_0^{(j)} - m_J\big)^T G^* \\
&= \left[\frac{J-1}{J}\right] C_J G^*.
\end{aligned}
\tag{37}
$$

Similarly,

$$
\begin{aligned}
C_1^{pp} &= \frac{1}{J} \sum_{j=1}^{J} \widehat{p}_0^{(j)} \big(\widetilde{p}_0^{(j)}\big)^T = \frac{1}{J} \sum_{j=1}^{J} Gu_0^{(j)} \left(Gu_0^{(j)} - \frac{1}{J} \sum_{\ell=1}^{J} Gu_0^{(l)}\right)^T = \frac{1}{J} \sum_{j=1}^{J} Gu_0^{(j)} \big(Gu_0^{(j)} - Gm_J\big)^T \\
&= \frac{1}{J} \sum_{j=1}^{J} G\big(u_0^{(j)} - m_J\big)\big(u_0^{(j)} - m_J\big)^T G^* = \left[\frac{J-1}{J}\right] GC_J G^*.
\end{aligned}
\tag{38}
$$

Therefore, from (25), (16) and (15) it follows that the u_1 estimate of the iterative method is

$$
\begin{aligned}
u_1 &\equiv \frac{1}{J} \sum_{j=1}^{J} u_1^{(j)} = \frac{1}{J} \sum_{j=1}^{J} u_0^{(j)} + C_1^{up}\big(C_1^{pp} + \Gamma\big)^{-1} \left(\frac{1}{J} \sum_{j=0}^{J} \big(y_1^{(j)} - Gu_0^{(j)}\big)\right) \\
&= m_J + C_1^{up}\big(C_1^{pp} + \Gamma\big)^{-1} \Big(y + \frac{1}{J} \sum_{j=0}^{J} \eta_1^{(j)} - Gm_J\Big).
\end{aligned}
\tag{39}
$$

Thus, the estimate u_1 provides an approximation to the regularized least-squares problem (35) that converges in the limit of $J \to \infty$ (i.e. with infinite number of ensemble members). Indeed, notice that, almost surely as $J \to \infty$

$$\bar{u}_J \to \bar{u}, \quad C_J \to C, \quad C_1^{up} \to CG^*, \quad C_1^{pp} \to GCG^*, \quad \frac{1}{J} \sum_{j=0}^{J} \eta_1^{(j)} \to 0. \tag{40}$$

Therefore, as $J \to \infty$

$$u_1 \to u_{\text{TP}} = \bar{u} + CG^*(GCG^* + \Gamma)^{-1}(y - G\bar{u}). \tag{41}$$

This link between the ensemble Kalman method and the regularized least square-problems for the linear inverse problem opens the possibility of solving nonlinear inverse problems by iterating the Kalman filter as we proposed in the subsection 2.3. However, it is important to remark that, in contrast to the least-squares approach, the implementation of the ensemble Kalman method does not require the derivative of the forward operator.

3. Elliptic equation

As a simple pedagogical example, we consider the ill-posed inverse problem of recovering the right-hand side of an elliptic equation in one spatial dimension given noisy observation of the solution. This explicitly solvable linear model will allow us to elucidate the performance of the ensemble Kalman method as an iterative regularization method. The results show that the ensemble Kalman method performs comparably to Tikhonov–Phillips regularized least squares for this problem. The lower bound BA is always significantly better, but of course unimplementable in practice, since the truth is unknown.

3.1. Setting

Consider the one dimensional elliptic equation

$$-\frac{\mathrm{d}^2 p}{\mathrm{d}x^2} + p = u$$
$$u(0) = u(\pi) = 0.$$

Thus $G = A^{-1}$ where $A = (-\frac{\mathrm{d}^2}{\mathrm{d}x^2} + 1)$ and $D(A) = H^2(I) \cap H_0^1(I)$ with $I = (0, \pi)$. We are interested in the inverse problem of recovering u from noisy observations of p:

$$y = p + \eta$$
$$= A^{-1}u + \eta.$$

For simplicity we assume that the noise is white: $\eta \sim N(0, \gamma^2 I)$. We consider Tikhonov–Phillips regularization of the form $\|u\|_C^2$, where $C = \beta(A - I)^{-1}$. The problem may be solved explicitly in the Fourier sine basis, and the coefficients of the Tikhonov–Phillips regularized least squares solution, in this basis, u_k may be expressed in terms of the coefficients of the data in the same basis, y_k:

$$\left[\left(\frac{1}{\gamma(1+k^2)} \right)^2 + \beta^{-1}k^2 \right] u_k = \frac{y_k}{\gamma^2(1+k^2)}, \qquad k = 1, \ldots, \infty. \qquad (42)$$

This demonstrates explicitly the regularization which is present for wavenumbers k such that $k^6 \geqslant \mathcal{O}(\beta\gamma^{-2})$. We now present some detailed numerical experiments which will place the ensemble Kalman algorithm in the context of an iterative regularization scheme. This will provide a roadmap for understanding the nonlinear inverse problem applications we explore in subsequent sections.

3.2. Numerical results

Throughout this subsection we choose $\beta = 10$ and $\gamma = 0.01$. We choose a truth $u^\dagger \sim N(0, C)$ and simulate data from the model: $y = A^{-1}u^\dagger + \eta^\dagger$, where $\eta^\dagger \sim \mathcal{N}(0, \Gamma)$. Recall from section 2 that the space $\mathcal{A} = \mathrm{span}\{\psi^{(j)}\}_{j=1}^J$ will be chosen either based on draws from $N(0, C)$ (with subscript R) or from the Karhunen–Loéve basis for $N(0, C)$ (with subscript KL). The experiments obtained from the iterative ensemble Kalman method with the set $\mathcal{A}$ chosen according to the two scenarios described above are denoted by EnKF$_R$ and EnKF$_{KL}$, respectively. In figure 1 (left) we display the relative error with respect to the truth of EnKF$_R$ and EnKF$_{KL}$ (i.e. $\|u_{EnKF} - u^\dagger\|_{L^2(I)}/\|u^\dagger\|_{L^2(I)}$). The data misfit $\|y - G(u_{EnKF})\|$ is shown in figure 1 (right). The black dotted line corresponds to the value of the noise level, i.e. $\|y - G(u^\dagger)\| = \|\eta^\dagger\|$. Note that the error of EnKF$_{KL}$ decreases for some iterations before reaching its minimum, while EnKF$_R$ reaches its minimum after a small number of iterations (often only one), and then increases. In both cases the error reaches its minimum at an iteration
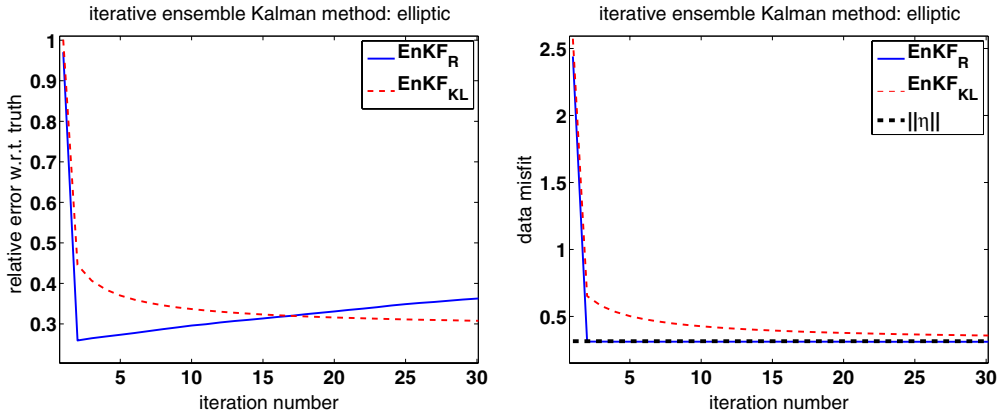
Figure 1. Performance of $EnKF_R$ and $EnKF_{KL}$. Left: relative error with respect to the truth. Right: data misfit.

step such that the value of the associated data misfit is approximately the noise level defined above. In particular, for $EnKF_R$ we observe that once $||y^\dagger - G(u_{EnKF})||_\Gamma$ is below the noise, the error in the estimate increases. This behaviour has been often reported when some optimization techniques are applied for the solution of inverse ill-posed problems [29]. It is clear that, once the data misfit is at, or below, the noise level, the given choice of $\mathcal{A}$ does not provide sufficient regularization of the problem. Indeed this suggests that an early termination based on the discrepancy principle [16] may furnish the ensemble Kalman algorithm with the regularizing properties needed for this choice of $\mathcal{A}$. It is worth mentioning that the aforementioned increase in the error after the data misfit reaches the noise level was observed in additional experiments (not shown) where the data was generated with different noise levels.

It is clear from figure 1 that the selection of $\mathcal{A}$ with elements from the first elements of the KL basis alleviates the ill-posedness of the inverse problem. More precisely, those first elements of the KL basis corresponds to the largest eigenvalues of the covariance operator C. Therefore, the subspace $\mathcal{A}$ where the solution of the ill-posed problem is sought does not contain the directions associated with smaller eigenvalues which are, in turn, responsible for the lack of stability of the inversion. Then, in contrast to the $EnKF_R$, the estimate generated $EnKF_{KL}$ is a linear combination of eigenfunctions associated with larger eigenvalues of C.

For the two scenarios described earlier, we now compare the performance of the ensemble Kalman method with respect to the Tikhonov–Phillips regularized least-squares (LS) and the best approximation BA methods. Let us consider first the 'R' random scenario where $\mathcal{A}$ is the linear subspace generated from random draws from μ_0. Since the estimate depends on the choice of $\mathcal{A}$, we compare $EnKF_R$, LS_R and BA_R on 100 different $\mathcal{A}$'s corresponding to 100 different prior ensembles. Furthermore, since we noted that $EnKF_R$ increases after the first iterations, in this example we consider only the first iteration of $EnKF_R$. The relative errors of the three estimators for different $\mathcal{A}$'s are displayed in figure 2. Note that the ensemble Kalman method is comparable to the least-squares method in terms of accuracy, but does not involve derivatives of the forward operator. Both the ensemble Kalman method and the least-squares are less accurate than the best approximation, of course, but produce errors of similar order of magnitude. In the first column of table 1 we display the values of the aforementioned estimators averaged over the sets $\mathcal{A}$ (generated from the prior). For the case where $\mathcal{A}$ is generated from the KL basis, the comparison of the $EnKF_{KL}$, LS_{KL} and BA_{KL} is straightforward and the values
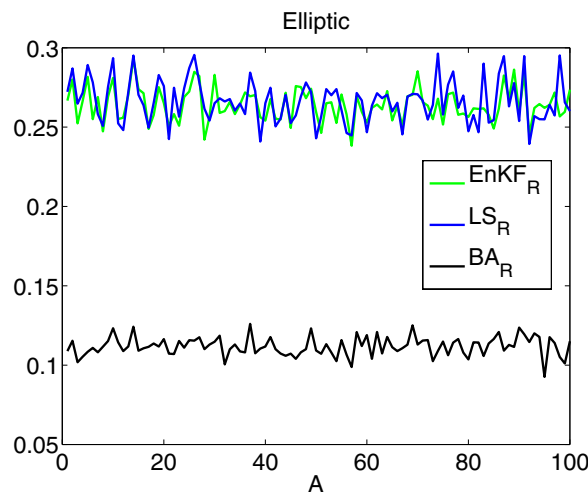
Figure 2. Comparison over different subspaces $\mathcal{A}$, generated from draws from the prior, of the relative errors of one iteration of EnKF$_R$ versus (Tikhonov regularized) least squares (LS$_R$) and the best approximation (BA$_R$).

Table 1. Relative errors with respect to the truth for the experiments in section 3 (elliptic), section 4 (groundwater) and section 5 (NSE).

Method	Elliptic	Groundwater	NSE
EnKF$_R$ (averaged over $\mathcal{A}$)	0.257	0.597	0.661
LS$_R$ (averaged over $\mathcal{A}$)	0.264	0.581	0.591
BA$_R$ (averaged over $\mathcal{A}$)	0.111	0.367	0.499
EnKF$_{KL}$ (final iteration $n = 30$)	0.270	0.591	0.650
LS$_{KL}$	0.250	0.569	0.500
BA$_{KL}$	0.070	0.278	0.439

are also displayed in the first column of table 1. Note that for EnKF$_{KL}$ we consider the estimate obtained from at the last iteration. These results again show that EnKF and LS are similar in terms of accuracy, and produce errors of similar order of magnitude to BA.

4. Groundwater flow

Next, we will investigate an inverse problem arising in groundwater modelling. Typically, the calibration of subsurface flow models consists of estimating geologic properties whose model predictions 'best fit' the measurements of flow-related quantities. In particular, herein we will consider the estimation of the conductivity of an aquifer from hydraulic head measurements. Similarly to the previous section, we find comparable performance of EnKF and LS, with the latter here regularized in a Newton-CG fashion. Again, BA is included for comparison.

4.1. Setting

We consider groundwater flow in a two-dimensional confined aquifer whose physical domain is $\Omega = [0, 6] \times [0, 6]$. The hydraulic conductivity is denoted by K. The flow in the aquifer, is described in terms of the piezometric head $h(x)$ ($x \in \Omega$) which, in the steady-state is governed

by the following equation [3]

$$-\nabla \cdot e^u \nabla h = f \qquad \text{in } \Omega \tag{43}$$

where $u \equiv \log K$ and f is defined by

$$f(x_1, x_2) = \begin{cases} 0 & \text{if} \quad 0 < x_2 \leqslant 4, \\ 137 & \text{if} \quad 4 < x_2 < 5, \\ 274 & \text{if} \quad 5 \leqslant x_2 < 6. \end{cases} \tag{44}$$

We consider the following boundary conditions

$$h(x, 0) = 100, \qquad \frac{\partial h}{\partial x}(6, y) = 0, \qquad -e^u \frac{\partial h}{\partial x}(0, y) = 500, \qquad \frac{\partial h}{\partial y}(x, 6) = 0. \tag{45}$$

For the physical interpretation of the source term and boundary conditions (44)–(45), we refer the reader to [5] where a similar model was used as a benchmark for inverse modelling in groundwater flow. A similar model was also studied in [15, 19] also in the context of parameter identification. We will be interested in the inverse problem of recovering the hydraulic conductivity, or more precisely its logarithm u, from noisy pointwise measurements of the piezometric head h. This is a model for the situation in groundwater applications where observations of head are used to infer the conductivity of the aquifer.

4.2. Numerical results

We let $\mathcal{G}(u) = \{h(x_k)\}_{k \in \mathbb{K}}$ where $\mathbb{K}$ is some finite set of points in Ω with cardinality κ. In particular, $\mathbb{K}$ is given by the configuration of $N = 100$ observation wells displayed in figure 5 (top right). We introduce the prior Gaussian $\mu_0 \sim \mathcal{N}(\overline{u}, C)$, let $u^\dagger \sim \mu_0$, and simulate data $y = \mathcal{G}(u^\dagger) + \eta^\dagger$, where $\eta^\dagger \sim \mathcal{N}(0, \Gamma)$. We let $C = \beta L^{-\alpha}$ with $L \equiv -\Delta$ defined on $D(L) = \{v \in H^2(\Omega) | \nabla v \cdot \mathbf{n} = 0, \text{ on } \partial\Omega, \int_\Omega v = 0\}$. Additionally, we define $\Gamma = \gamma^2 I$ and we choose $\alpha = 1.3$, $\beta = 0.5$, $\overline{u} = 4$ and $\gamma = 7$ fixed. We reiterate from section 2 that the space $\mathcal{A} = \text{span}\{\psi^{(j)}\}_{j=1}^J$ will be chosen based on either draws from the prior μ_0, with subscript R, or on the Karhunen–Loéve basis, with subscript KL.

The forward model (43)–(45) is discretized with cell-centred finite differences [32]. For the approximation of the LS problem, we implemented the Newton-CG method of [16]. We conduct experiments analogous to those of the previous section and the results are shown in figures 3, 4, 5 and the second column of table 1. These results are very similar to those shown in the previous section for the linear elliptic problem, demonstrating the robustness of the observations made in that section for the solution of inverse problems in general, using the ensemble Kalman methodology herein.

5. Navier–Stokes equation

In this section, we consider an inverse problem in fluid dynamics, which is relevant to data assimilation applications in oceanography and meteorology. In particular, we examine the problem of recovering the initial condition of the Navier–Stokes equation (NSE), given noisy pointwise observations of the velocity field at later times. We will investigate a regime in which the combination of viscosity, time-interval, and truncation of the forward model is such that the exponential ill-posedness of the inverse problem is alleviated.
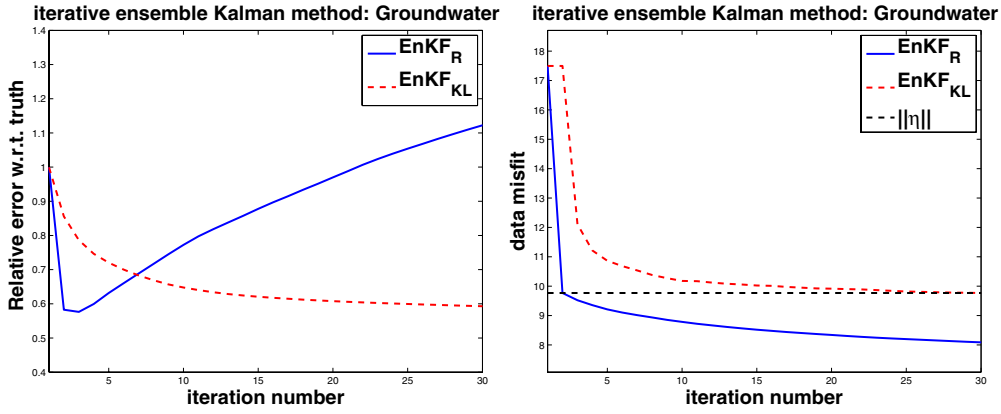
Figure 3. Performance of EnKF$_R$ and EnKF$_{KL}$. Left: relative error with respect to the truth. Right: data misfit.
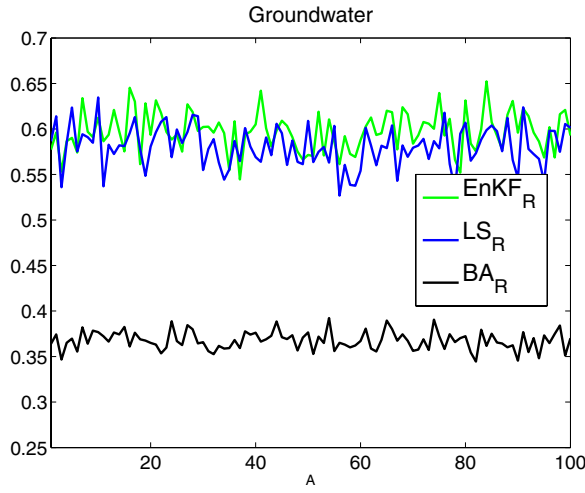

Figure 4. Comparison over different subspaces $\mathcal{A}$, generated from draws from the prior, of the relative errors of one iteration of EnKF$_R$ versus (Newton-CG regularized) least squares (LS$_R$) and the best approximation (BA$_R$).

5.1. Setting

We consider the 2D NSE on the torus $\mathbb{T}^2 := [-1, 1) \times [-1, 1)$ with periodic boundary conditions:

$$
\begin{aligned}
\partial_t v - \nu \Delta v + v \cdot \nabla v + \nabla p &= f &&\text{for all } (x, t) \in \mathbb{T}^2 \times (0, \infty), \\
\nabla \cdot v &= 0 &&\text{for all } (x, t) \in \mathbb{T}^2 \times (0, \infty), \\
v &= u &&\text{for all } (x, t) \in \mathbb{T}^2 \times \{0\}.
\end{aligned}
$$

Here $v : \mathbb{T}^2 \times (0, \infty) \to \mathbb{R}^2$ is a time-dependent vector field representing the velocity, $p : \mathbb{T}^2 \times (0, \infty) \to \mathbb{R}$ is a time-dependent scalar field representing the pressure, $f : \mathbb{T}^2 \to \mathbb{R}^2$ is
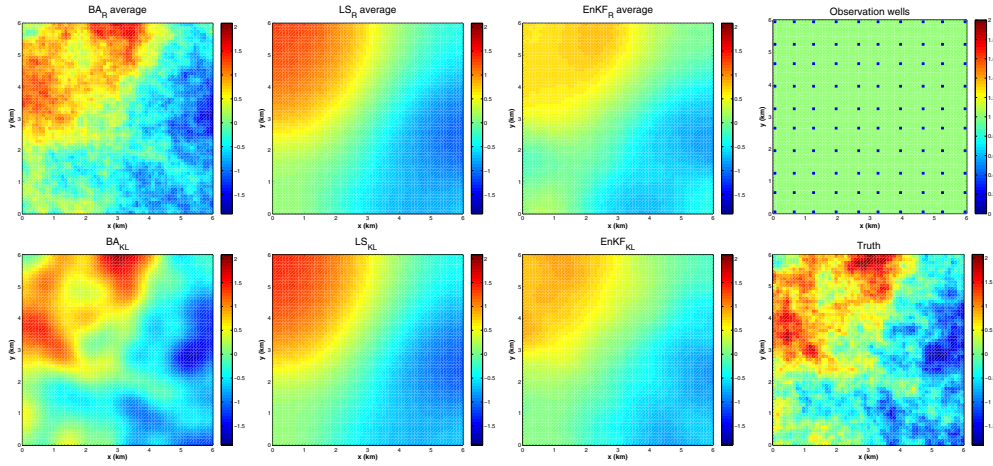
Figure 5. Estimated log K. Top left: BA_R, average. Top middle left: LS_R, average. Top middle right: first iteration of $EnKF_R$, average. Top right: measurement well locations. Bottom left: BA_{KL}. Bottom middle left: LS_{KL}. Bottom middle right: $EnKF_{KL}$. Bottom right: the truth $u^\dagger$.

a vector field representing the forcing (which we assume to be time-independent for simplicity), and ν is the viscosity. We are interested in the inverse problem of determining the initial velocity field u from pointwise measurements of the velocity field at later times. This is a model for the situation in weather forecasting where observations of the atmosphere are used to improve the initial condition used for forecasting.

5.2. Numerical results

We let $t_j = jh$, for $j = 1, \ldots, J$, and $\mathcal{G}(u) = \{v(x_k, t_j)\}_{(j,k) \in \mathbb{K}'}$ where $\mathbb{K}' = \{1, \ldots, J\} \times \mathbb{K}$ and $\mathbb{K}$ is some finite set of points in Ω with cardinality κ. In particular, we take $\mathbb{K}$ to be the set of grid points in physical space implied by the underlying spectral truncation used in the numerical integration (details below). As discussed in section 2, we introduce a prior μ_0, let $u^\dagger \sim \mu_0$, and simulate data $y = \mathcal{G}(u^\dagger) + \eta^\dagger$, where $\eta^\dagger \sim \mathcal{N}(0, \Gamma)$. As in the previous section, we let $\Gamma = \gamma^2 I$. We fix $\gamma = 0.01$ for all our experiments. The prior μ_0 is defined to be the empirical measure supported on the attractor, i.e. it is defined by samples of a trajectory of the forward model after convergence to statistical equilibrium. We use 10^4 time-steps to construct this empirical measure. Once again the space $\mathcal{A} = \text{span}\{\psi^{(j)}\}_{j=1}^J$ is chosen based on J samples from the (now non-Gaussian) prior μ_0, or on a Karhunen–Loéve expansion based on the empirical mean and covariance from the long forward simulation giving rise to μ_0. It is important to note that the truth $u^\dagger$ is included in the long trajectory used to construct μ_0. Therefore, some of the initial ensembles drawn from μ_0 end up containing a snapshot very close to the truth; such results are overly optimistic as this is not the typical situation. It is interesting that all the methods we study here have comparably overly optimistic results for such ensembles.

The forcing in f is taken to be $f = \nabla^\perp \psi$, where $\psi = \cos(\pi k \cdot x)$ and $\nabla^\perp = J\nabla$ with J the canonical skew-symmetric matrix, and $k = (5, 5)$. The method used to approximate the forward model is a modification of a fourth-order Runge–Kutta method, ETD4RK [8], in which the Stokes semi-group is computed exactly by working in the incompressible Fourier basis
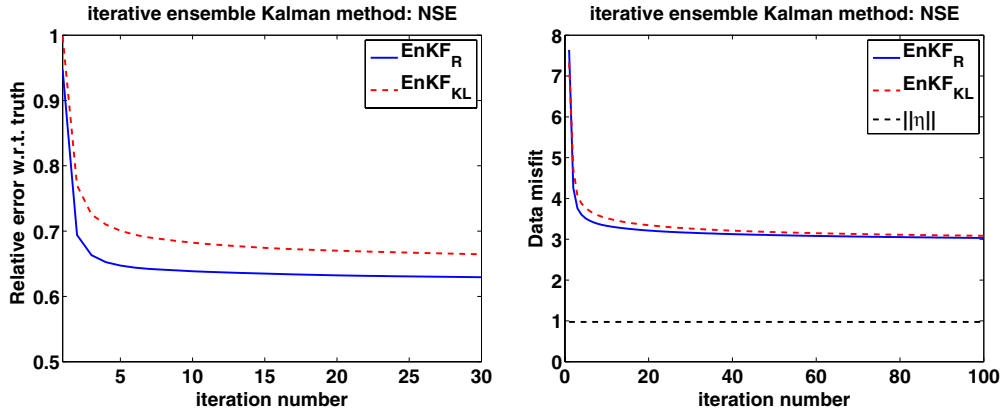
16

Figure 6. Performance of EnKF$_R$ and EnKF$_{KL}$. Left: relative error with respect to the truth. Right: data misfit.

$\{\psi_k(x)\}_{k \in \mathbb{Z}^2 \setminus \{0\}}$, and Duhamel's principle (variation of constants formula) is used to incorporate the nonlinear term. We use a time-step of d$t = 0.005$. Spatially, a Galerkin spectral method [17] is used, in the same basis, and the convolutions arising from products in the nonlinear term are computed via FFTs. We use a double-sized domain in each dimension, padded with zeros, resulting in 64^2 grid-point FFTs, and only half the modes in each direction are retained when transforming back into spectral space again. This prevents aliasing, which is avoided as long as more than one third of the domain in which FFTs are computed consists of such padding with zeros. The dimension of the attractor is determined by the viscosity parameter ν. For the particular forcing used there is an explicit steady state for all $\nu > 0$ and for $\nu \geqslant 0.035$ this solution is stable (see [27], chapter 2 for details). As ν decreases the flow becomes increasingly complex and we focus subsequent studies of the inverse problem on the mildly chaotic regime which arises for $\nu = 0.01$. Regarding observations, we let $h = 4 \times \mathrm{d}t = 0.02$ and take $J = 10$, so that $T = 0.2$. The observations are made at all numerically resolved, and hence observable, wavenumbers in the system; hence $K = 32^2$, because of the padding to avoid aliasing.

The numerical results resulting from these experiments are displayed in figures 6, 7, 8 and the third column of table 1. The results are very similar to those of the previous two sections, qualitatively: EnKF and LS type methods perform comparably, in both the case of random and Karhunen–Loéve based initial draws; furthermore the lower bound produced by BA type methods is of similar order of magnitude to the EnKF-based methods although the actual error is, of course, smaller. However, we also see that the behaviour of the iterated EnKF$_R$ is quite different from what we observed in the previous sections since it decreases monotonically. We conjecture that this is because of the mildly ill-posed nature of this problem. Indeed we have repeated the results (not shown) of this section at higher viscosity $\nu = 0.1$, where linear damping in the forward model induces greater ill-posedness, and confirmed that we recover results for the iterated EnKF$_R$ which are similar to those in the previous sections. Finally, we have checked that the behaviour of the error in the iterated EnKF$_R$ is repeatable for Gaussian prior μ_0, and hence is not a result of non-Gaussian ensembles used in the figures.
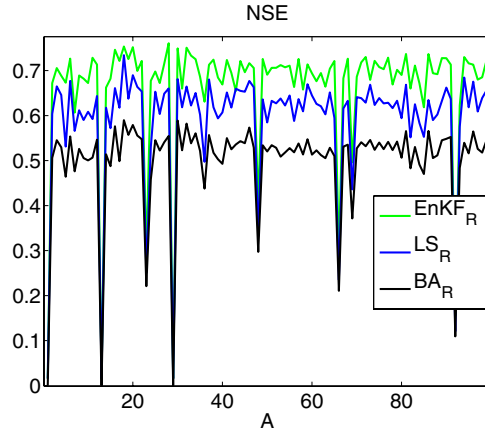
Figure 7. Comparison over different subspaces $\mathcal{A}$, generated from draws of the prior, of the relative errors of one iteration of EnKF$_R$ versus (Tikhonov regularized) least squares (LS$_R$) and the best approximation (BA$_R$).
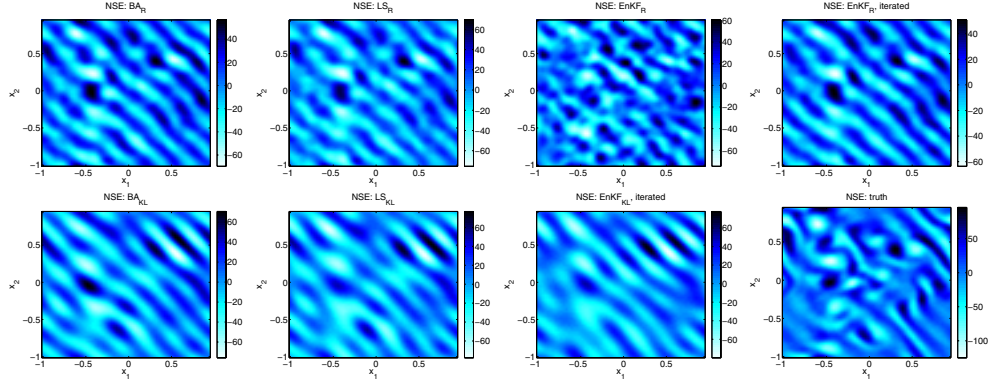

Figure 8. Vorticity of initial condition $\omega(0)$. Top left: BA$_R$, average. Top middle left: LS$_R$, average. Top middle right: first iteration of EnKF$_R$, average. Top right: tenth iteration of EnKF$_R$, average. Bottom left: BA$_{KL}$. Bottom middle left: LS$_{KL}$. Bottom middle right: EnKF$_{KL}$. Bottom right: the truth $u^\dagger$.

6. Conclusions

We have illustrated the use of EnKF as a derivative-free optimization tool for inverse problems, showing that the method computes a nonlinear approximation in the linear span of the initial ensemble. We have also demonstrated comparable accuracy to least squares based methods and shown that, furthermore, the accuracy is of the same order of magnitude as the best approximation within the linear span of the initial ensemble. Further study of the EnKF methodology for inverse problems, and in particular its accuracy with respect to choice of initial ensemble, would be of interest. Furthermore, in this paper we have concentrated purely on the accuracy of state estimation using EnKF. Study of its accuracy in terms of uncertainty quantification will yield further insight. Finally, although we have studied a time-dependent

example (the NSE) we did not use a methodology which exploited the sequential acquisition of the data—we concatenated all the data in space and time. Exposing sequential structure in data can be useful in steady parameter estimation problems as shown in [4]. In future work we will study ideas similar to those in our paper, but exploiting sequential structure.

Acknowledgments

References

[1] Aanonsen S I, Naevdal G, Oliver D S, Reynolds A C and Valles B 2009 The ensemble Kalman filter in reservoir engineering—a review *SPE J.* **14** 393–412
[2] Anderson J L 2001 An ensemble adjustment Kalman filter for data assimilation *Mon. Weather Rev.* **129** 2884–903
[3] Bear J 1972 *Dynamics of Fluids in Porous Media* (New York: Dover)
[4] Calvetti D and Somersalo E 2010 Subjective knowledge or objective belief? *Large-Scale Inverse Problems and Quantification of Uncertainty* ed L Biegler, G Biros, O Ghattas, M Heinkenschloss, D Keyes, B Mallick, Y Marzouk, L Tenorio, B van Bloemen Waanders and K Willcox (New York: Wiley) p 33
[5] Carrera J and Neuman S P Estimation of aquifer parameters under transient and steady state conditions: 3. Application to synthetic and field data *Water Resour. Res.* **22** 228–42
[6] Chen Y and Oliver D 2012 Ensemble randomized maximum likelihood method as an iterative ensemble smoother *Math. Geosci.* **44** 1–26
[7] Cotter S L, Dashti M, Robinson J C and Stuart A M 2009 Bayesian inverse problems for functions and applications to fluid mechanics *Inverse Problems* **25** 115008
[8] Cox S M and Matthews P C 2002 Exponential time differencing for stiff systems *J. Comput. Phys.* **176** 430–55
[9] Dacorogna B 2007 *Direct Methods in the Calculus of Variations* vol 78 (Berlin: Springer)
[10] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* vol 375 (Berlin: Springer)
[11] Evensen G 1994 Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics *J. Geophys. Res.* **99** 10143–62
[12] Evensen G 2009 *Data Assimilation: The Ensemble Kalman Filter* (London: Springer)
[13] Evensen G and Van Leeuwen P J 1996 Assimilation of geosat altimeter data for the Agulhas current using the ensemble Kalman filter with a quasi-geostrophic model *Mon. Weather* **128** 85–96
[14] Evensen G 2003 The ensemble Kalman filter: theoretical formulation and practical implementation *Ocean Dyn.* **53** 343–67
[15] Hanke M 1997 A regularizing Levenberg–Marquardt scheme, with applications to inverse groundwater filtration problems *Inverse Problems* **13** 79–95
[16] Hanke M 1997 Regularizing properties of a truncated Newton-cg algorithm for nonlinear inverse problems *Numer. Funct. Anal. Optim.* **18** 971–93
[17] Hesthaven J S, Gottlieb S and Gottlieb D 2007 *Spectral Methods for Time-Dependent Problems* vol 21 (Cambridge: Cambridge University Press)
[18] Houtekamer P L and Mitchell H L 2001 A sequential ensemble Kalman filter for atmospheric data assimilation *Mon. Weather Rev.* **129** 123–37
[19] Iglesias M A and Dawson C 2009 An iterative representer-based scheme for data inversion in reservoir modeling *Inverse Problems* **25** 035006
[20] Isakov V 2006 *Inverse Problems for Partial Differential Equation* vol 127 (New York: Springer)
[21] Kaipio J P and Somersalo E 2005 *Statistical and Computational Inverse Problems* (Berlin: Springer)
[22] Kalnay E 2003 *Atmospheric Modeling, Data Assimilation, and Predictability* (Cambridge: Cambridge University Press)
[23] Katltenbacher B, Neubauer A and Scherzer O 2008 *Iterative Regularization Methods for Nonlinear Ill-Posed Problems (Radon Series on Computational and Applied Mathematics)* 1st edn (Berlin: de Gruyter & Co)
[24] Kinderlehrer D and Stampacchia G 1980 *An Introduction to Variational Inequalities and Their Applications* vol 88 (New York: Academic)
[25] Lehtinen M S, Paivarinta L and Somersalo E 1999 Linear inverse problems for generalised random variables *Inverse Problems* **5** 599

[26] Li G and Reynolds A C 2009 Iterative ensemble Kalman filters for data assimilation *SPE J.* **14** (3) 496–505 SPE-109808-PA

[27] Majda A and Wang X 2006 *Non-linear Dynamics and Statistical Theories for Basic Geophysical Flows* (Cambridge: Cambridge University Press)

[28] Mandelbaum A 1984 Linear estimators and measurable linear transformations on a Hilbert space *Probab. Theory Relat. Fields* **65** 385–97

[29] Nagy J G and Palmer K M 2003 Steepest descent, cg, and iterative regularization of ill-posed problems *BIT Numer. Math.* **43** 1003–17

[30] Oliver D S, Reynolds A C and Liu N 2008 *Inverse Theory for Petroleum Reservoir Characterization and History Matching* 1st edn (Cambridge: Cambridge University Press) ISBN: 9780521881517

[31] Perianez A, Reich H and Potthast R 2013 Error analysis and adaptive localization for ensemble methods in data assimilation at press

[32] Russell T F and Wheeler M F 1983 Finite element and finite difference methods for continuous flows in porous media *Mathematics of Reservoir Simulation* ed R E Ewing (Philadelphia, PA: SIAM)

[33] Vogel C R 2002 *Computational Methods for Inverse Problems* (Philadelphia PA: SIAM)