COMMUN. MATH. SCI.
Vol. 15, No. 7, pp. 1867–1896

FILTER BASED METHODS FOR STATISTICAL LINEAR INVERSE PROBLEMS*

MARCO A. IGLESIAS†, KUI LIN‡, SHUAI LU§, AND ANDREW M. STUART¶

Abstract. Ill-posed inverse problems are ubiquitous in applications. Understanding of algorithms for their solution has been greatly enhanced by a deep understanding of the linear inverse problem. In the applied communities ensemble-based filtering methods have recently been used to solve inverse problems by introducing an artificial dynamical system. This opens up the possibility of using a range of other filtering methods, such as 3DVAR and Kalman based methods, to solve inverse problems, again by introducing an artificial dynamical system. The aim of this paper is to analyze such methods in the context of the linear inverse problem.

Statistical linear inverse problems are studied in the sense that the observational noise is assumed to be derived via realization of a Gaussian random variable. We investigate the asymptotic behavior of filter based methods for these inverse problems. Rigorous convergence rates are established for 3DVAR and for the Kalman filters, including minimax rates in some instances. Blowup of 3DVAR and a variant of its basic form is also presented, and optimality of the Kalman filter is discussed. These analyses reveal a close connection between (iterated) regularization schemes in deterministic inverse problems and filter based methods in data assimilation. Numerical experiments are presented to illustrate the theory.

Keywords. Kalman filter; 3DVAR; statistical inverse problems; artificial dynamics.

AMS subject classifications. 93E11; 65J22; 47A52.

1. Introduction

In many geophysical applications, in particular in the petroleum industry and in hydrology, distributed parameter estimation problems are often solved by means of ensemble Kalman filters [19]. The basic methodology is to introduce an artificial dynamical system, to supplement this with observations, and to apply the ensemble Kalman filter. The methodology is described in a basic, abstract form, applicable to a general, possibly nonlinear, inverse problem in [11]. In this basic form of the algorithm regularization is present due to dynamical preservation of a subspace spanned by the ensemble during the iteration. The paper [12] gives further insight into the development of regularization for these ensemble Kalman inversion methods, drawing on links with the Levenberg–Marquardt scheme [8]. In this paper our aim is to further the study of filters for the solution of inverse problems, going beyond the ensemble Kalman filter to encompass the study of other filters such as 3DVAR and the Kalman filter itself; see [15] for an overview of these filtering methods. A key issue will be the implementation of regularization with the aim of deriving optimal error estimates.

We focus on the linear inverse problem

$$y = Au^\dagger + \eta, \tag{1.1}$$

where A is a compact operator acting between Hilbert spaces X and Y. The exact solution is denoted by $u^\dagger \in X$ and η is a noise polluting the observations. We will consider

*Received: June 12, 2016; accepted: May 6, 2017. Communicated by Guillaume Bal.
†School of Mathematical Sciences, University of Nottingham, UK (Marco.Iglesias@nottingham.ac.uk).
‡School of Mathematical Sciences, Fudan University, P.R. China (link10@fudan.edu.cn).
§Corresponding author, School of Mathematical Sciences, Fudan University, P.R. China (slu@fudan.edu.cn).
¶Computing and Mathematical Sciences, California Institute of Technology, USA (astuart@caltech.edu).

two situations: **Data Model** 1 where multiple observations are made in the form (1.1); and **Data Model** 2 where a single observation is made. For modelling purposes we will assume that the noise η is generated by the Gaussian $\mathcal{N}(0,\gamma^2 I)$, independently in the case of multiple observations. In each case we create a sequence $\{y_n\}_{n\geq 1}$; for Data Model 1 the elements of this sequence are i.i.d. $\mathcal{N}(Au^\dagger,\gamma^2 I)$ whilst for Data Model 2 they are $y_n \equiv y$, with y a single draw from $\mathcal{N}(Au^\dagger,\gamma^2 I)$. The case where multiple independent observations are made is not uncommon in applications (for example in electrical impedance tomography (EIT, [4]) and, although we do not pursue it here, our methodology also opens up the possibility of considering multiple instances with correlated observational noise, by means of similar filtering-based techniques.

The artificial, partially observed linear dynamical system that underlies our methodology is as follows:

$$\begin{aligned} u_n &= u_{n-1}, \\ y_n &= Au_n + \eta_n. \end{aligned} \tag{1.2}$$

In *deriving* the filters we apply to this dynamical system, it is assumed that the $\{\eta_n\}_{n\geq 1}$ are i.i.d. from $\mathcal{N}(0,\gamma^2 I)$. Note, however, that whilst the data sequence $\{y_n\}_{n\geq 1}$ we use in Data Model 1 is of this form, the assumption is not compatible with Data Model 2; thus for Data Model 2 we have a form of *model error* or *model mis-specification* [15].

By studying the application of filtering methods to the solution of the linear inverse problem our aim is to open up the possibility of employing the filtering methodology to (static) inverse problems of the form (1.1), and nonlinear generalizations. We confine our analysis to the linear setting as experience has shown that a deep understanding of this case is helpful both because there are many linear inverse problems which arise in applications, and because knowledge of the linear case guide methodologies for the more general nonlinear problem [6]. The last few decades have seen a comprehensive development of the theory of linear inverse problems, both classically and statistically; see [2,6] and the references therein. Consider the Tikhonov-Phillips regularization method

$$\operatorname{argmin}_u \Big(\frac{1}{2\gamma^2}\|y-Au\|_Y^2 + \frac{\alpha}{2}\|u-u_0\|_E^2 \Big).$$

This can be reformulated from a probabilistic perspective as the MAP estimator for Bayesian inversion given a Gaussian smoothness prior, with mean u_0 and Cameron–Martin space E compactly embedded into X, and a Gaussian noise model as defined above; this connection is eludicated in [5, 13]. We note that from the point of view of Tikhonov-Phillips regularization only the parameter $\alpha\gamma^2$ is relevant, but that each of α and γ have separate interpretations in the overarching Bayesian picture, the first as a scaling of the prior precision and the second as observational noise variance. In this paper we deepen the connection between the Bayesian methodology and classical methods.

The recent paper [11] opens up the prospect for a statistical explanation of iterated regularization methods in the form of

$$u_n = u_{n-1} + K_n(y - Au_{n-1})$$

with a general Kalman gain operator K_n. In this paper, we establish the connection between iterated regularization methods (c.f. [6,9]) and filter based methods [15] with respect to an artificial dynamic system. More precisely, for a linear inverse problem, we verify that the iterated Tikhonov regularization method

$$u_n = u_{n-1} + A^*(AA^* + \alpha I)^{-1}(y - Au_{n-1}) \tag{1.3}$$

is closely related to filtering methods such as 3DVAR and the Kalman filter when applied to the partially observed linear dynamical system (1.2). The similarity between both schemes provides a probabilistic interpretation of iterated regularization methods, and allows the possibility of quantifying uncertainty via the variance. On the other hand, we will employ techniques from the convergence analysis arising in regularization theories [6] to shed light on the convergence of filter based methods, especially when the linear observation operator is ill-posed. We do not employ Hilbert scales [7, 10, 17, 18] in our analysis as this would lead to a different focus, but their study in this problem would also be of interest.

The paper is organized as follows. We first introduce filter based methods for the artificial dynamics (1.2) in Section 2. Section 3 describes some general useful formulae which are relevant to all the filters we study, and lists our main assumptions on the inverse problem of interest. In Sections 4 and 5 respectively, detailed asymptotic analyses are given for the Kalman filter method and 3DVAR, for both data models. The final Section 6 presents numerical illustrations confirming the theoretical predictions.

2. Filters for the artificial dynamics

2.1. Filter definitions. Recall the artificial dynamics (1.2), where the observation operator A also defines the inverse problem (1.1), and $\{\eta_n\}_{n\geq 1}$ is an i.i.d. sequence with $\eta_1 \sim \mathcal{N}(0,\gamma^2 I)$. The aim of filters is to estimate u_n given the data $\{y_j\}_{j=1}^n$. In particular, probabilistic filtering aims to estimate the probability distribution of the conditional random variable $u_n|\{y_j\}_{j=1}^n$.

If we assume that $u_0 \sim \mathcal{N}(m_0, C_0)$ then the desired conditional random variable is Gaussian, because of the linearity inherent in system (1.2), together with the assumed Gaussian structure of the noise sequence $\{\eta_n\}_{n\geq 1}$. Furthermore the independence of the elements of the noise sequence means that the Gaussian can be updated sequentially in a Markovian fashion. If we denote by m_n the mean, and by C_n the covariance, then we obtain the Kalman filter updates for these two quantities:

$$K_n = C_{n-1}A^*\left(AC_{n-1}A^* + \gamma^2 I\right)^{-1} \tag{2.1a}$$

$$m_n = m_{n-1} + K_n(y_n - Am_{n-1}) \tag{2.1b}$$

$$C_n = (I - K_n A)C_{n-1}. \tag{2.1c}$$

The operator K_n is known as the *Kalman gain* and the inverse of the covariance, the *precision operator* C_n^{-1}, may be shown to satisfy

$$C_n^{-1} = C_{n-1}^{-1} + \frac{1}{\gamma^2}A^*A. \tag{2.2}$$

All of these facts concerning the Kalman filter may be found in Chapter 4 of [15]. Expression (2.2) requires careful justification in infinite dimensions, and this is provided in [1] in certain settings. However we will only use equation (2.2) as a quick method for deriving useful formulae, not expressed in terms of precision operators, which can be justified directly under the assumptions we make.

A simplification of the Kalman filter method is the 3DVAR algorithm [15] which is not, strictly speaking, a probabilistic filter because it does not attempt to accurately track covariance information. Instead the covariance is fixed in time at

$$C_{n-1} = \frac{\gamma^2}{\alpha}\Sigma_0 \tag{2.3}$$

for some fixed positive and self-adjoint operator Σ_0. The parameter α is a scaling constant the inverse of which measures the relative size of the fixed covariance of the filter relative to that of the data. Imposing this simplification on Equations (2.1a), (2.1b) gives

$$K_n \equiv \mathcal{K} := \Sigma_0 A^* (A\Sigma_0 A^* + \alpha I)^{-1} \tag{2.4a}$$

$$\zeta_n = \zeta_{n-1} + \mathcal{K}(y_n - A\zeta_{n-1}). \tag{2.4b}$$

It is also helpful to define, from equation (2.1c),

$$\mathcal{C} \equiv \frac{\gamma^2}{\alpha}(I - \mathcal{K}A)\Sigma_0. \tag{2.5}$$

Notice [6, 16] that the update form (2.4b) looks like a stationary iterated Tikhonov method (1.3) with A replaced by $A\Sigma_0^{\frac{1}{2}}$.

Throughout the paper (K_n, m_n, C_n) stands for Kalman gain, updated mean and updated covariance for the Kalman filter method and $(\mathcal{K}, \zeta_n, \mathcal{C})$ is the related sequence of quantities for 3DVAR.

2.2. Asymptotic behaviour of filters. We will view the filters as methods for reconstructing the truth $u^\dagger$; in particular we will study the proximity of m_n (for the Kalman filter) and ζ_n (for 3DVAR) to $u^\dagger$ for various large n asymptotics. Although the assumption in the *derivation* of the filters is that y_n is an i.i.d. sequence of the form $\mathcal{N}(Au^\dagger, \gamma^2 I)$, we will not always assume that the data available is of this form; to be precise Data Model 1 is compatible with this assumption whilst Data Model 2 is not.

Recall that Data Model 1 refers to the situation where the data used in the Kalman and 3DVAR filters has the form $y_n = Au^\dagger + \eta_n$, where the η_n are i.i.d. $\mathcal{N}(0, \gamma^2 I)$. Given such a data sequence we can generate an auxiliary element

$$\bar{y} = \frac{1}{n}\sum_{j=1}^{n} y_j = Au^\dagger + \frac{1}{n}\sum_{j=1}^{n} \eta_j$$

with $\bar{\eta} = \frac{1}{n}\sum_{j=1}^{n} \eta_j$ and $\bar{\eta} \sim \mathcal{N}(0, \frac{\gamma^2}{n} I)$. The law of large numbers and central limit theorem thus allows us to consider an inverse problem of the form (1.1) with noise level reduced by a factor of $\sqrt{n}$. In particular, when n increases to infinity, the noise and uncertainty in the auxiliary element are removed and one can thus obtain a (generalized) Moore-Penrose inversion by application of a least squares method. We will study, in the sequel, whether the Kalman or 3DVAR filters are able to automatically exploit the decreased uncertainty inherent in an i.i.d. data set of this form.

For Data Model 2 we simply assume that the data used in the filters is of the form $y_n \equiv y$ where y is given by equation (1.1) with $\eta \sim \mathcal{N}(0, \gamma^2 I)$. From the discussion in the preceding paragraph, we clearly expect less accurate reconstruction in this case. For this data model we may view 3DVAR as a stationary iterated Tikhonov regularization, whilst the Kalman filter is an alternative iterated non-stationary regularization scheme, since K_n is updated in each step. In addition, the statistical perspective not only allows us to obtain an estimator (the mean (2.1b) or (2.4b)), but also in the case of the Kalman filter method, to quantify the uncertainty (via the covariance (2.1c)). This uncertainty quantification perspective provides additional motivation for the filtering approaches considered herein.

In this paper our primary focus is the asymptotic large n behavior of the Kalman filter method and 3DVAR. More precisely, we are interested in the accurate recovery of

the true state $u^\dagger$ when the noise variance vanishes, i.e. $\gamma^2 \to 0$ for Data Models 1 and 2, or as $n \to \infty$ for Data Model 1 (by the law of large numbers/central limit theorem discussion above).

To highlight the difficulties inherent in ill-posed inverse problems in this regard, we note the following which is a straightforward consequence of Theorem 4.10 in [15] when specialized to linear problems. Here, and in what follows, $\|\cdot\|$ denotes both the norm on X and the operator norm from X into itself.

PROPOSITION 2.1. *Consider the 3DVAR algorithm with $\{y_n\}_{n\geq 1}$. Assume that there exists a constant L such that $\|I - \mathcal{K}A\| \leq L < 1$ and that $\|\mathcal{K}\| < \infty$. Then, for Data Model 2, it yields*

$$\limsup_{n\to\infty} \|\zeta_n - u^\dagger\| \leq \frac{\|\mathcal{K}\|\|\eta\|}{1-L}.$$

Note however, that if the observation operator A is compact or the inversion is ill-posed, the assumption $L < 1$ in the preceding proposition cannot hold. More precisely, the operator $I - \mathcal{K}A$ is no longer contractive since the spectrum of the operator $\mathcal{K}A$ clusters at the origin. Our focus in the remainder of the paper will be on such ill-posed inverse problems.

3. Main assumptions and general properties of filters

3.1. Assumptions. Recall that $\|\cdot\|$ denotes both the norm on X and the operator norm from X into itself. Throughout C will denote a generic constant, independent of the key limiting quantities γ and n. Our main assumption is:

ASSUMPTION 3.1. *For both the Kalman filter and the 3DVAR filter, we assume*

(i) *the variance $C_0 = \frac{\gamma^2}{\alpha}\Sigma_0$ and $\mathcal{R}(\Sigma_0^{1/2}) \subset \mathcal{D}(A)$, where α is a positive constant and Σ_0 is positive self-adjoint, and Σ_0^{-1} is a densely defined unbounded self-adjoint strictly positive operator;*

(ii) *the forward operator A satisfies*

$$C^{-1}\|\Sigma_0^{\frac{a}{2}} x\| \leq \|Ax\| \leq C\|\Sigma_0^{\frac{a}{2}} x\| \tag{3.1}$$

on X for some constants $a > 0$ and $1 \leq C < \infty$;

(iii) *the initial mean satisfies $m_0 - u^\dagger \in \mathcal{D}\left(\Sigma_0^{-\frac{s}{2}}\right)$ (or $\zeta_0 - u^\dagger \in \mathcal{D}\left(\Sigma_0^{-\frac{s}{2}}\right)$) with $0 \leq s \leq a+2$;*

(iv) *the operator Σ_0 in item (i) is trace class on X.*

We briefly comment on these items. Item (i) allows a well defined operator

$$B_0 := A\Sigma_0^{1/2} \tag{3.2}$$

which is essential in carrying out our analysis. Item (ii) is often called the *link condition* and it connects both operators A and Σ_0 (or C_0). The third item (iii) is the *source condition* (regularity) of the true solution [6]. The final item (iv) makes C_0 a well-defined covariance operator on X [3].

Item (ii) in the preceding assumption is automatically satisfied if A^*A and Σ_0 have the same eigenfunctions and certain decaying singular values. Item (iii) can then be expressed in this eigenbasis. When studying the Kalman filter we will, in some instances,

employ the following specific form of items (ii), (iii). Comparison of Assumptions 3.1 and 3.2 we see that they are identical if $a = \frac{2p}{1+2\epsilon}$ and $s = \frac{2\beta}{1+2\epsilon}$.

ASSUMPTION 3.2. *Let the variance $C_0 = \frac{\gamma^2}{\alpha}\Sigma_0$. The operators Σ_0 and A^*A have the same eigenfunctions $\{e_i\}$ with their eigenvalues $\{\lambda_i\}$ and $\{\kappa_i^2\}$ satisfying*

$$\lambda_i = i^{-1-2\epsilon}, \qquad C^{-1}i^{-p} \le \kappa_i \le Ci^{-p}$$

for some $\epsilon > 0$, $p > 0$ and $C \ge 1$. Furthermore, by choosing the initial mean $m_0 = 0$, the true solution $u^\dagger$ with its coordinates $\{u^{\dagger,i}\}$ in the basis $\{e_i\}$ obeys $\sum_{i=1}^{\infty}(u^{\dagger,i})^2 i^{2\beta} < \infty$ and $0 < \beta \le 1 + 2\epsilon + 2p$.

3.2. Filter properties. We start by deriving properties of the Kalman filter method under Data Model 1. Other cases can be simply derived from minor variants of this setting. Recall from (2.1b)

$$m_n = (I - K_n A)m_{n-1} + K_n y_n$$

and note that

$$u^\dagger = (I - K_n A)u^\dagger + K_n A u^\dagger.$$

Under Data Model 1 we have $y_n = Au^\dagger + \eta_n$ and hence the total error $e_n := m_n - u^\dagger$ satisfies

$$\begin{aligned} e_n &= (I - K_n A)e_{n-1} + K_n \eta_n \\ &= \prod_{j=1}^{n}(I - K_j A)e_0 + \sum_{j=1}^{n-1}\left(\prod_{i=n-j}^{n-1}(I - K_{i+1}A)\right)K_{n-j}\eta_{n-j} + K_n\eta_n \\ &:= J_1 + J_2. \end{aligned} \tag{3.3}$$

Here

$$\begin{aligned} J_1 &= \prod_{j=1}^{n}(I - K_j A)e_0 \quad \text{and} \\ J_2 &= \sum_{j=1}^{n-1}\left(\prod_{i=n-j}^{n-1}(I - K_{i+1}A)\right)K_{n-j}\eta_{n-j} + K_n\eta_n. \end{aligned}$$

To establish a rigorous convergence analysis, the mean squared error (MSE) $\mathbb{E}\|m_n - u^\dagger\|^2$ is of particular interest. Since $u^\dagger$ is deterministic and each η_n is i.i.d Gaussian we obtain a bias-variance decomposition of the MSE:

$$\mathbb{E}\|m_n - u^\dagger\|^2 = \|J_1\|^2 + \mathbb{E}\|J_2\|^2. \tag{3.4}$$

To proceed further, both terms J_1 and J_2 need to be calibrated more carefully.

We consider the operator $I - K_n A$ which appears in both terms. By equation (2.1c), we obtain

$$C_n = (I - K_n A)C_{n-1} = \prod_{j=1}^{n}(I - K_j A)C_0,$$

which is equivalent to

$$\prod_{j=1}^{n}(I-K_jA)=C_nC_0^{-1}. \tag{3.5}$$

Notice that equation (2.2) yields

$$C_n^{-1}=C_{n-1}^{-1}+\frac{1}{\gamma^2}A^*A=C_0^{-1}+\frac{n}{\gamma^2}A^*A. \tag{3.6}$$

By equations (3.5) and (3.6) we obtain

$$\begin{aligned}\prod_{j=1}^{n}(I-K_jA)&=C_nC_0^{-1}=(C_0^{-1}+nA^*A/\gamma^2)^{-1}C_0^{-1}\\&=C_0^{\frac{1}{2}}\gamma^2(\gamma^2I+nC_0^{\frac{1}{2}}A^*AC_0^{\frac{1}{2}})^{-1}C_0^{-\frac{1}{2}}.\end{aligned} \tag{3.7}$$

We will use this expression (which is well-defined in view of Assumption 3.1 (i)) and the labelled equations preceding it in this subsection, frequently in what follows.

4. Asymptotic analysis of the Kalman filter

In this section we investigate the asymptotic behaviour of the Kalman filter (2.1a)-(2.1c), under Assumption 3.1. In particular, we are interested in whether we can reproduce the minimax convergence rate. This minimax rate is achieved by adopting Assumption 3.1 in the diagonal form of Assumption 3.2.

4.1. Kalman filter and data model 1. We present the main results in current subsection.

THEOREM 4.1. *Let Assumption 3.1 hold. Then the Kalman filter method (2.1a)-(2.1c) yields a bias-variance decomposition of the MSE*

$$\mathbb{E}\|m_n-u^\dagger\|^2\le C\left(\frac{\alpha}{n}\right)^{\frac{s}{a+1}}+\frac{\gamma^2}{\alpha}\mathrm{tr}(\Sigma_0)$$

for the Data Model 1. Setting $\alpha=N^{\frac{s}{s+a+1}}$ and stopping the dynamic iteration at $n=N$ then gives

$$\mathbb{E}\|m_N-u^\dagger\|^2\le\left(C+\gamma^2\mathrm{tr}(\Sigma_0)\right)N^{-\frac{s}{s+a+1}}. \tag{4.1}$$

THEOREM 4.2. *Let Assumption 3.2 hold. Then the Kalman filter method (2.1a)-(2.1c) yields a bias-variance decomposition of the MSE*

$$\mathbb{E}\|m_n-u^\dagger\|^2\le C\left(\frac{\alpha}{n}\right)^{\frac{2\beta}{1+2\epsilon+2p}}+\gamma^2n^{-\frac{2\epsilon}{1+2\epsilon+2p}}\alpha^{-\frac{1+2p}{1+2\epsilon+2p}}$$

for the Data Model 1. Setting $\alpha=N^{\frac{2(\beta-\epsilon)}{1+2\beta+2p}}$ and stopping the dynamic iteration at $n=N$ then gives the following minimax convergence rate:

$$\mathbb{E}\|m_N-u^\dagger\|^2\le CN^{-\frac{2\beta}{1+2\beta+2p}}.$$

REMARK 4.1.

(1) Under the Assumption 3.2 we prove unconditional convergence of the Kalman filter method for any fixed $\alpha>0$ and $\gamma>0$, noticing that both the bias and variance vanish when n goes to infinity. The key ingredient which leads to this unconditional convergence, in comparison with Assumption 3.1, is that the rate of decay of the eigenvalues of the variance operator Σ_0 is made explicit under Assumption 3.2; this is to be contrasted with the weaker assumption $\mathrm{tr}(\Sigma_0)<\infty$ made in Assumption 3.1 (iv).

(2) By choosing α depending on the update step N, again with fixed γ, both Theorems 4.1 and 4.2 yield convergence rates in the MSE sense. Indeed in the second case, where we use Assumption 3.2, the minimax rate of $N^{-\frac{2\beta}{1+2\beta+2p}}$ is achieved. This minimax rate may also be achieved from the Bayesian posterior distribution with appropriate tuning of the prior in terms of the (effective) noise size $\sqrt{N}$ [14]; the tuning of the prior is identical to the tuning of the initial condition for the covariance C_0, via choice of α. ♢

Proof of Theorems 4.1 and 4.2 is straightforward by means of a bias-variance decomposition. Let Assumption 3.1 (i) hold, noting that then $B_0 := A\Sigma_0^{1/2}$ is well-defined. We thus obtain, by equation (3.7),

$$\prod_{i=1}^{n}(I-K_iA)=\Sigma_0^{\frac{1}{2}}\alpha(\alpha I+nB_0^*B_0)^{-1}\Sigma_0^{-\frac{1}{2}}=\Sigma_0^{\frac{1}{2}}r_{1,\frac{\alpha}{n}}(B_0^*B_0)\Sigma_0^{-\frac{1}{2}}, \tag{4.2}$$

where

$$r_{1,\frac{\alpha}{n}}(\lambda):=\frac{\frac{\alpha}{n}}{\frac{\alpha}{n}+\lambda}=\frac{\alpha}{\alpha+n\lambda}. \tag{4.3}$$

The operator-valued function $r_{1,\frac{\alpha}{n}}$ (4.3) has been verified to be powerful in the convergence analysis of deterministic regularization schemes; see [16, Ch.2]. In that context the following inequality is useful:

$$|\lambda^t r_{1,\frac{\alpha}{n}}(\lambda)|\le\left(\frac{\alpha}{n}\right)^t,\qquad \lambda\in(0,\|B_0^*B_0\|],\quad 0\le t\le 1. \tag{4.4}$$

Following these ideas we obtain the next two lemmas describing the bias and variance error bounds. We leave the proofs of both lemmas to the Appendix. Theorems 4.1 and 4.2 are consequences, by choosing the parameter α appropriately.

LEMMA 4.1 (Bias for Kalman filter). *Let Assumption 3.1 (i)-(iii) hold. Then the Kalman filter method (2.1a)-(2.1c) yields*

$$\|J_1\|^2\le C\left(\frac{\alpha}{n}\right)^{\frac{s}{\alpha+1}}.$$

Furthermore, if Assumption 3.2 is valid, the bias obeys

$$\|J_1\|^2\le C\left(\frac{\alpha}{n}\right)^{\frac{2\beta}{1+2\epsilon+2p}}.$$

LEMMA 4.2 (Variance for Kalman filter – Data Model 1). *Let Assumption 3.1 (i), (iv) hold and $\{\eta_n\}$ in system (1.2) be an i.i.d sequence with $\eta_1\sim\mathcal{N}(0,\gamma^2 I)$. Then the Kalman filter method (2.1a)-(2.1c) yields*

$$\mathbb{E}\|J_2\|^2\le\frac{\gamma^2}{\alpha}\mathrm{tr}(\Sigma_0).$$

Furthermore, if Assumption 3.2 is valid, the variance obeys

$$\mathbb{E}\|J_2\|^2 \le \gamma^2 n^{-\frac{2\epsilon}{1+2\epsilon+2p}} \alpha^{-\frac{1+2p}{1+2\epsilon+2p}}.$$

4.2. Kalman filter and data model 2. The key difference between Data Model 2 and Data Model 1 is that, in the case 2, the noises η_n appearing in the expression for the term J_2 are identical, rather than i.i.d. mean zero as in case 1. This results in a reduced rate of convergence in case 2 over case 1, as seen in the following two theorems:

THEOREM 4.3. *Let Assumption 3.1 hold. Then the Kalman filter method (2.1a)-(2.1c) yields a bias-variance decomposition of the MSE*

$$\mathbb{E}\|m_n - u^\dagger\|^2 \le C\left(\frac{\alpha}{n}\right)^{\frac{s}{a+1}} + \frac{n\gamma^2}{\alpha}\mathrm{tr}(\Sigma_0)$$

for the Data Model 2. Fix $\alpha = 1$ and assume that the noise variance $\gamma^2 = N^{-\frac{a+s+1}{a+1}}$. If the dynamic iteration is stopped at $n = N$ then the following convergence rate is valid:

$$\mathbb{E}\|m_N - u^\dagger\|^2 \le (C + \mathrm{tr}(\Sigma_0))\, N^{-\frac{s}{a+1}}. \tag{4.5}$$

THEOREM 4.4. *Let Assumption 3.2 hold. Then the Kalman filter method (2.1a)-(2.1c) yields a bias-variance decomposition of the MSE*

$$\mathbb{E}\|m_n - u^\dagger\|^2 \le C\left(\frac{\alpha}{n}\right)^{\frac{2\beta}{1+2\epsilon+2p}} + \gamma^2\left(\frac{n}{\alpha}\right)^{\frac{1+2p}{1+2\epsilon+2p}}$$

for the Data Model 2. Fix $\alpha = 1$ and assume that the noise variance $\gamma^2 = N^{-\frac{1+2\beta+2p}{1+2\epsilon+2p}}$. If the dynamic iteration is stopped at $n = N$ then the following convergence rate is valid:

$$\mathbb{E}\|m_N - u^\dagger\|^2 \le C N^{-\frac{2\beta}{1+2\epsilon+2p}}.$$

Both convergence rates in Theorems 4.3 and 4.4 are of the same order since the variance has been tuned to scale in the same way as the bias by choosing to stop the dynamic iteration at N, depending on $\gamma \ll 1$, appropriately. In comparison with the convergence rates in Theorems 4.1 and 4.2, the ones in this section under Data Model 2 require small noise γ; those in the preceding subsection do not because multiple observations, with additive independent noise, are made of $Au^\dagger$. Proof of the two preceding theorems is straightforward: the J_1 terms is analyzed as in the preceding subsection and the J_2 term must be carefully analyzed under the assumptions of Data Model 2. The key new result is stated in the following lemma, whose proof may be found in the Appendix.

LEMMA 4.3 (Variance for Kalman filter method – Data Model 2). *Let Assumption 3.1 hold and each observation $y_n \equiv y$ be fixed. Then the Kalman filter method (2.1a)-(2.1c) yields*

$$\mathbb{E}\|J_2\|^2 \le \frac{n\gamma^2}{\alpha}\mathrm{tr}(\Sigma_0).$$

Furthermore, if Assumption 3.2 is valid, the variance obeys

$$\mathbb{E}\|J_2\|^2 \le \gamma^2\left(\frac{n}{\alpha}\right)^{\frac{1+2p}{1+2\epsilon+2p}}.$$

5. Asymptotic analysis of 3DVAR

5.1. Classical 3DVAR. The mean of the 3DVAR algorithm is given by (2.4b) and has the form

$$\zeta_n = (I - \mathcal{K}A)\zeta_{n-1} + \mathcal{K}y_n.$$

Furthermore

$$u^\dagger = (I - \mathcal{K}A)u^\dagger + \mathcal{K}Au^\dagger.$$

If we define $\varepsilon_n = \zeta_n - u^\dagger$ then we obtain, since $y_n = Au^\dagger + \eta_n$,

$$\begin{aligned}\varepsilon_n &= (I - \mathcal{K}A)\varepsilon_{n-1} + \mathcal{K}\eta_n \\ &= (I - \mathcal{K}A)^n \varepsilon_0 + \sum_{j=0}^{n-1} (I - \mathcal{K}A)^j \mathcal{K}\eta_{n-j}\end{aligned}$$

with $\varepsilon_0 := \zeta_0 - u^\dagger$. We further derive

$$(I - \mathcal{K}A)^n = \Sigma_0^{\frac{1}{2}} (\alpha(\alpha I + B_0^* B_0)^{-1})^n \Sigma_0^{-\frac{1}{2}} = \Sigma_0^{\frac{1}{2}} r_{n,\alpha}(B_0^* B_0) \Sigma_0^{-\frac{1}{2}},$$

by inserting the definition of the Kalman gain (2.4a), and by assuming Assumption 3.1 (i). The operator-valued function $r_{n,\alpha}(\cdot)$ is defined

$$r_{n,\alpha}(\lambda) := \left(\frac{\alpha}{\alpha + \lambda}\right)^n.$$

Similarly to the analysis of the Kalman filter, we derive

$$\begin{aligned}\varepsilon_n &= (I - \mathcal{K}A)^n \varepsilon_0 + \sum_{j=0}^{n-1} (I - \mathcal{K}A)^j \mathcal{K}\eta_{n-j} \\ &= \Sigma_0^{\frac{1}{2}} r_{n,\alpha}(B_0^* B_0) \Sigma_0^{-\frac{1}{2}} \varepsilon_0 + \sum_{j=0}^{n-1} (I - \mathcal{K}A)^j \mathcal{K}\eta_{n-j} \\ &:= I_1 + I_2. \end{aligned} \tag{5.1}$$

Thus, the MSE takes the bias-variance decomposition form

$$\mathbb{E}\|\zeta_n - u^\dagger\|^2 = \mathbb{E}\|\varepsilon_n\|^2 = \|I_1\|^2 + \mathbb{E}\|I_2\|^2.$$

This leads to the following two theorems:

THEOREM 5.1. *Let Assumption 3.1 hold. Then 3DVAR filter (2.4a)-(2.4b) yields a bias-variance decomposition of the MSE*

$$\mathbb{E}\|\zeta_n - u^\dagger\|^2 \le C\left(\frac{\alpha}{n}\right)^{\frac{s}{a+1}} + C\frac{\gamma^2 \ln n}{\alpha} \mathrm{tr}(\Sigma_0)$$

for the Data Model 1. *Setting* $\alpha = N^{\frac{s}{s+a+1}}$ *and stopping the dynamic iteration at* $n = N$ *then gives*

$$\mathbb{E}\|\zeta_n - u^\dagger\|^2 \le C\left(1 + \gamma^2 \mathrm{tr}(\Sigma_0)\right) N^{-\frac{s}{s+a+1}} \ln N. \tag{5.2}$$

THEOREM 5.2. *Let Assumption 3.2 hold. Then 3DVAR filter (2.4a)-(2.4b) yields a bias-variance decomposition of the MSE*

$$\mathbb{E}\|\zeta_n - u^\dagger\|^2 \le C\left(\frac{\alpha}{n}\right)^{\frac{2\beta}{1+2\epsilon+2p}} + C\gamma^2\alpha^{-\frac{1+2p}{1+2\epsilon+2p}}$$

for the Data Model 1. *Setting* $\alpha = N^{\frac{2\beta}{1+2\epsilon+2\beta+2p}}$ *and stopping the dynamic iteration at* $n = N$ *then gives*

$$\mathbb{E}\|\zeta_N - u^\dagger\|^2 \le C\left(1+\gamma^2\right) N^{-\frac{2\beta}{1+2\epsilon+2\beta+2p}} \ln N.$$

REMARK 5.1. The decay rate at the end of the preceding Theorem is the same as that in Theorem 5.1 if $a = \frac{2p}{1+2\epsilon}$ and $s = \frac{2\beta}{1+2\epsilon}$. This is the setting in which Assumptions 3.1 and 3.2 are identical.

The preceding two theorems show that, under Data Model 1 and for any fixed α, the (bound on the) MSE of the 3DVAR filter blows up logarithmically as $n \to \infty$ under Assumption 3.1, and is asymptotically bounded for Assumption 3.2. In contrast, for the Kalman filter method the MSE is asymptotically bounded or unconditionally converges in n under the same assumptions – see Theorems 4.1 and 4.2.

With optimal choice of α in terms of the stopping time of the dynamic iteration at $n = N$, comparison of the convergence rates in Theorems 4.1 and 5.1 (or Theorems 4.2 and 5.2) shows that the Kalman filter outperforms 3DVAR, but only by a logarithmic factor (or a Hölder factor). For simplicity we only analyze and discuss Data Model 1 for 3DVAR filter under the additional Assumption 3.2; as for Data Model 2 one can derive consequences similar to those in the preceding section in an analogous manner.

We emphasize that in the case of deterministic iterated Tikhonov regularization it is possible to use infinite dynamic observations to overcome the restriction $s \le a+2$ in Assumption 3.1, for 3DVAR in the case of a constant regularization parameter $\alpha = 1$; this is related to the Lardy's method [9].

We now study Data Model 2 and 3DVAR. We consider only Assumption 3.1; however the reader may readily extend the analysis to include Assumption 3.2. In the case of Data Model 2, both the Kalman and 3DVAR filters have the same error bounds:

THEOREM 5.3. *Let Assumption 3.1 hold. Then the 3DVAR algorithm (2.4a)-(2.4b) yields a bias-variance decomposition of the MSE*

$$\mathbb{E}\|\zeta_n - u^\dagger\|^2 \le C\left(\frac{\alpha}{n}\right)^{\frac{s}{a+1}} + \frac{n\gamma^2}{\alpha}\mathrm{tr}(\Sigma_0)$$

for the Data Model 2. *Fix* $\alpha = 1$ *and assume that the noise variance* $\gamma^2 = N^{-\frac{a+s+1}{a+1}}$. *If the dynamic iteration is stopped at* $n = N$ *then the following convergence rate is valid:*

$$\mathbb{E}\|\zeta_n - u^\dagger\|^2 \le (C + \mathrm{tr}(\Sigma_0)) N^{-\frac{s}{a+1}}. \tag{5.3}$$

The preceding three theorems can be proved by the bias-variance decomposition and application of the following three lemmas, whose proofs are left to the Appendix. However the proof of Theorem 5.2 is not as straightforward as the others and we present details in the Appendix.

LEMMA 5.1 (Bias for 3DVAR). *Let Assumption 3.1* (*i*)-(*iii*) *hold. Then 3DVAR (2.4) yields*

$$\|I_1\|^2 \le C\left(\frac{\alpha}{n}\right)^{\frac{s}{a+1}}.$$

Furthermore, if Assumption 3.2 is valid, the bias obeys

$$\|I_1\|^2 \leq C\left(\frac{\alpha}{n}\right)^{\frac{2\beta}{1+2\epsilon+2p}}.$$

LEMMA 5.2 (Variance for 3DVAR - Data Model 1). *Let Assumption 3.1 (i), (iv) hold and each noise η_n in system (1.2) i.i.d. generated by $\mathcal{N}(0,\gamma^2 I)$. Then 3DVAR (2.4) yields*

$$\mathbb{E}\|I_2\|^2 \leq C\frac{\gamma^2 \ln n}{\alpha}\mathrm{tr}(\Sigma_0)$$

for Data Model 1. Furthermore, if Assumption 3.2 is valid, the variance obeys

$$\mathbb{E}\|I_2\|^2 \leq C\gamma^2\alpha^{-\frac{1+2p}{1+2\epsilon+2p}}$$

and simultaneously

$$\mathbb{E}\|I_2\|^2 \leq C\frac{\gamma^2 \ln n}{\alpha}.$$

LEMMA 5.3 (Variance for 3DVAR - Data Model 2). *Let Assumption 3.1 hold and each observation $y_n \equiv y$ be fixed. Then 3DVAR (2.4) yields*

$$\mathbb{E}\|I_2\|^2 \leq \frac{n\gamma^2}{\alpha}\mathrm{tr}(\Sigma_0)$$

for Data Model 2.

5.2. Variant of 3DVAR for data model 1. Recall that the 3DVAR update form (2.4b) looks like a stationary iterated Tikhonov method (1.3) [6, 16], with A replaced by $A\Sigma_0^{\frac{1}{2}}$ and with a fixed parameter α. The non-stationary iterated Tikhonov regularization, with varying α, has been proven to be powerful in deterministic inverse problems [9]. We generalize this method to the 3DVAR setting. Furthermore we demonstrate that, for Data Model 1, it is possible to see blow-up phenomena with this algorithm.

Starting from the classical form of the 3DVAR filter as given in equations (2.4a), (2.4b), and (2.5) we propose a variant method in which α varies with n. We obtain

$$\mathcal{K}_n := \Sigma_0 A^*\left(A\Sigma_0 A^* + \alpha_n I\right)^{-1} \tag{5.4a}$$

$$v_n := v_{n-1} + \mathcal{K}_n(y_n - Av_{n-1}) \tag{5.4b}$$

$$\mathcal{C}_n := \frac{\gamma^2}{\alpha_n}(I - \mathcal{K}_n A)\Sigma_0. \tag{5.4c}$$

If we define $\epsilon_n = v_n - u^\dagger$ then we obtain, analogously to the derivation of equations (3.3) and (5.1), the bias-variance decomposition as follows:

$$\begin{aligned}\epsilon_n &= (I - \mathcal{K}_n A)\epsilon_{n-1} + \mathcal{K}_n\eta_n \\ &= \prod_{j=1}^{n}(I - \mathcal{K}_j A)\epsilon_0 + \sum_{j=1}^{n-1}\left(\prod_{i=n-j}^{n-1}(I - \mathcal{K}_{i+1}A)\right)\mathcal{K}_{n-j}\eta_{n-j} + \mathcal{K}_n\eta_n \\ &:= \mathcal{I}_1 + \mathcal{I}_2\end{aligned}$$

with

$$\epsilon_0 = v_0 - u^\dagger;$$
$$\mathcal{I}_1 = \prod_{j=1}^{n} (I - \mathcal{K}_j A)\epsilon_0;$$
$$\mathcal{I}_2 = \sum_{j=1}^{n-1} \left(\prod_{i=n-j}^{n-1} (I - \mathcal{K}_{i+1} A) \right) \mathcal{K}_{n-j}\eta_{n-j} + \mathcal{K}_n \eta_n.$$

By calibrating both terms $\mathcal{I}_1$, $\mathcal{I}_2$ carefully, we obtain the following blow-up result, proved in the Appendix. Although the result only provides an upper bound, numerical evidence does indeed show blow-up in this regime.

THEOREM 5.4. *Let Assumption 3.1 hold and let α_n be a sequence satisfying $\frac{1}{\alpha_n} \leq \tilde{c}\sigma_{n-1}$, with constant $\tilde{c}$, for $\sigma_n := \sum_{j=1}^{n} \frac{1}{\alpha_j}$. Then the variant EnKF method (5.4a)-(5.4c) yields a bias-variance decomposition of the MSE*

$$\mathbb{E}\|v_n - u^\dagger\|^2 \leq C\left(\sigma_n^{-\frac{s}{a+1}} + \gamma^2 \mathrm{tr}(\Sigma_0)\sigma_n\right)$$

for Data Model 1. In particular, the geometric sequence $\alpha_n := \alpha q^{n-1}$ with $\alpha > 0$ and $0 < q < 1$ yields

$$\mathbb{E}\|v_n - u^\dagger\|^2 \leq C\left(q^{\frac{s}{a+1}n} + \gamma^2 \mathrm{tr}(\Sigma_0) q^{-n}\right).$$

6. Numerical illustrations

In this section we provide numerical results which display the capabilities of 3DVAR and Kalman filter for solving linear inverse problems of the type described in Section 1. In addition, we verify numerically some of the theoretical results from Section 4 and Section 5.

6.1. Set-Up. We consider the two-dimensional domain $\Omega = [0,1]^2$ and define the operators

$$A = (-\triangle)^{-1}, \qquad \Sigma_0 = A^2 \tag{6.1}$$

with

$$D(\triangle) = \left\{ v \in H^2(\Omega) \mid \quad \nabla v \cdot \mathbf{n} = 0 \quad \text{on } \partial\Omega, \quad \int_\Omega v = 0 \right\}.$$

With this domain $-\triangle$ is positive, self-adjoint and invertible. Note that inequality (3.1) from Assumption 3.1 is satisfied with $a = 1$ and $C = 1$. In the following subsections we produce synthetic data from a true function $u^\dagger$ that we generate as a two-dimensional random field drawn from the Gaussian measure $N(0,\Sigma)$ with covariance

$$\Sigma = \left(-\triangle + \frac{1}{10} I\right)^{-(2s+1)} \tag{6.2}$$

with domain of definition $D(\Sigma) = D(\triangle)$ and where $s > 0$ is selected as described below. The shift of $-\triangle$ by a constant introduces a correlation length into the true function. Furthermore, for simplicity we consider $m_0 = 0$ and note [5] that the given selection of

$u^\dagger$ yields $u^\dagger \in \mathcal{H}^t$ for all $0<t<2s$. Therefore, for discussion of the present experiments we simply assume that $u^\dagger - m_0 \in \mathcal{D}(\Sigma_0^{-s/2}) = \mathcal{H}^{2s}$. Consequently, in order to satisfy Assumption 3.1 (iii) we need s such that $0<s\leq a+2=3$. Note that operator Σ_0 in definition (6.1) satisfies Assumption 3.1 (iv).

The numerical generation of $u^\dagger$ is carried out by means of the Karhunen–Loeve decomposition of $u^\dagger$ in terms of the eigenfunctions of Σ which, from the definition of $D(\Sigma)$, are cosine functions. We recall that for the application of the Kalman filter and 3DVAR with Data Model 1 we need to generate N instances of synthetic data $\{y_n\}_{n=1}^N$ where N is the maximum number of iterations of the scheme. Below we discuss the selection of such N. The aforementioned synthetic data are generated by means of $y_n \equiv Au^\dagger + \eta_n$ with $\eta_n \sim N(0,\gamma^2 I)$ and γ specified below. For Data Model 2 we produce synthetic data simply by setting $y_n \equiv y = Au^\dagger + \eta$ with $\eta \sim N(0,\gamma^2 I)$. In order to avoid inverse crimes [13], all synthetic data used in our experiments are generated on a finer grid (120×120 cells) than the one (of 60×60 cells) used for the inversion. We use splines to interpolate synthetic data on the coarser grid that we use for the application of the filters.

For both 3DVAR and Kalman filter with Data Model 1, we fix the number of iterations N for the scheme and consider the selection $\alpha = N^{\frac{s}{s+1+a}}$ stated in Theorem 4.1 and Theorem 5.1. Provided that the filters are stopped according to $n=N$, these theorems ensure the convergence rates in (4.1) and (5.2) that we verify numerically in the following subsection. For Data Model 2, Theorem 4.3 and Theorem 5.3 suggest that convergence rates (4.5) and (5.3) are satisfied for $\alpha=1$ and provided that $\gamma^2 = N^{-\frac{a+s+1}{a+1}}$. The latter equality provides an expression for the number of iterations $N=N(\gamma)$ that we may view as a stopping criterion, extended from the deterministic setting for these filtering algorithms applied to Data Model 2. In Algorithm 6.1 we summarize the Kalman filter and 3DVAR schemes applied to both data models.

ALGORITHM 6.1. *Kalman Filter/3DVAR (Data Model 1/Data Model 2)*
Let

$$N = \begin{cases} \text{fix integer selected a priori} & \text{for Data Model 1} \\ \text{round}(\gamma^{-\frac{2(a+1)}{a+s+1}}) & \text{for Data Model 2} \end{cases}$$

and

$$\alpha = \begin{cases} N^{\frac{s}{s+1+a}} & \text{for Data Model 1} \\ 1 & \text{for Data Model 2} \end{cases}$$

For $n=1,\ldots,N$, update m_{n-1} and C_{n-1} as follows

$$m_n = m_{n-1} + K_n(y_n - Am_{n-1})$$

$$C_n = \begin{cases} (I-K_nA)C_{n-1}, & \text{for Kalman Filter} \\ C_{n-1}, & \text{for 3DVAR} \end{cases}$$

where

$$K_n = C_{n-1}A^*\left(AC_{n-1}A^* + \gamma^2 I\right)^{-1}.$$

and where we recall that

$$y_n = \begin{cases} Au^\dagger + \eta_n & \text{for Data Model 1} \\ y & \text{for Data Model 2} \end{cases}$$

Note that for Data Model 1 we need at least N independent instances of data.

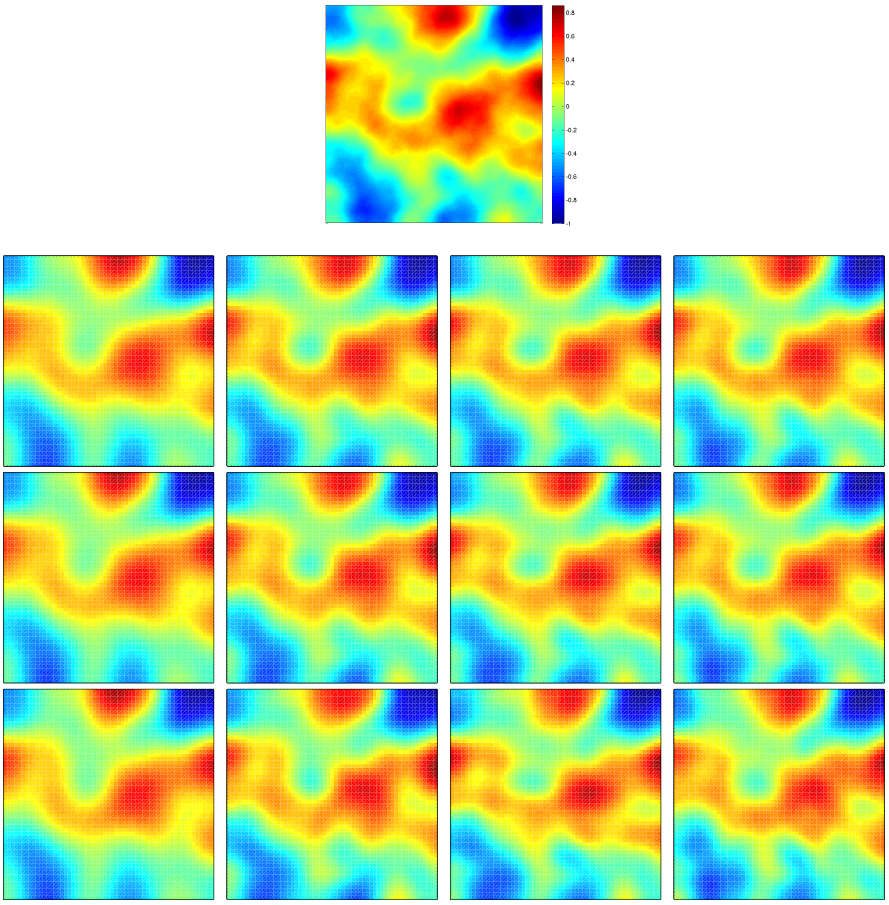

FIG. 6.1. *Top: truth* $u^\dagger$. *Top-middle, bottom-middle and bottom: Estimates obtained with 3DVAR and Data Model 1 at iterations (from left to right) 1, 10, 20, 30) for noise levels of* 1% *(top-middle),* 2.5% *(bottom-middle) and* 5% *(bottom) .*

6.2. Using Kalman filter and 3DVAR for solving linear inverse problems. In this subsection we demonstrate how the filters under consideration in the iterated framework described in Algorithm 6.1 can be used, with Data Model 1 and Data Model 2, to solve the linear inverse problem presented in Section 1. Let us consider the truth $u^\dagger$ displayed in Figure 6.1 (top) generated, as described in Subsection 6.1, from a Gaussian measure with covariance (6.2) and $s=1$. Synthetic data are generated as described above with three different choices of γ that yield noise levels of approximately 1%, 2.5%, and 5% of the norm of the noise free data (i.e. $Au^\dagger$).

We apply Algorithm 6.1 to this synthetic data generated both for the application of Data Model 1 and Data Model 2. For Data Model 1 we consider a selection of $N=25$. Algorithm 6.1 states that the these schemes should be stopped at the level $n=N$. However, in order to observe the performance of these schemes, in these experiments we allowed for a few more iterations. In the left column of Figure 6.2 we display the plots of the error w.r.t the truth of the estimator m_n as a function of the iterations, i.e.

$$E(m_n) \equiv \|m_n - \mathcal{P}u^\dagger\|$$

where $\mathcal{P}u^\dagger$ denotes the interpolation of $u^\dagger$ on the coarse grid used for the inversion. Note that the error w.r.t. the truth of the estimates produced by both schemes decreases monotonically. Interestingly, the value at the final iteration displayed in these figures is approximately the same for all these experiments independently of the noise level. Moreover, the stability of the scheme does not seem to depend on the early termination of the scheme. In addition, we note that both Kalman filter and 3DVAR exhibit very similar performance in terms of reducing the error w.r.t the truth. However, for larger noise levels we observe small fluctuations in the error obtained with 3DVAR.

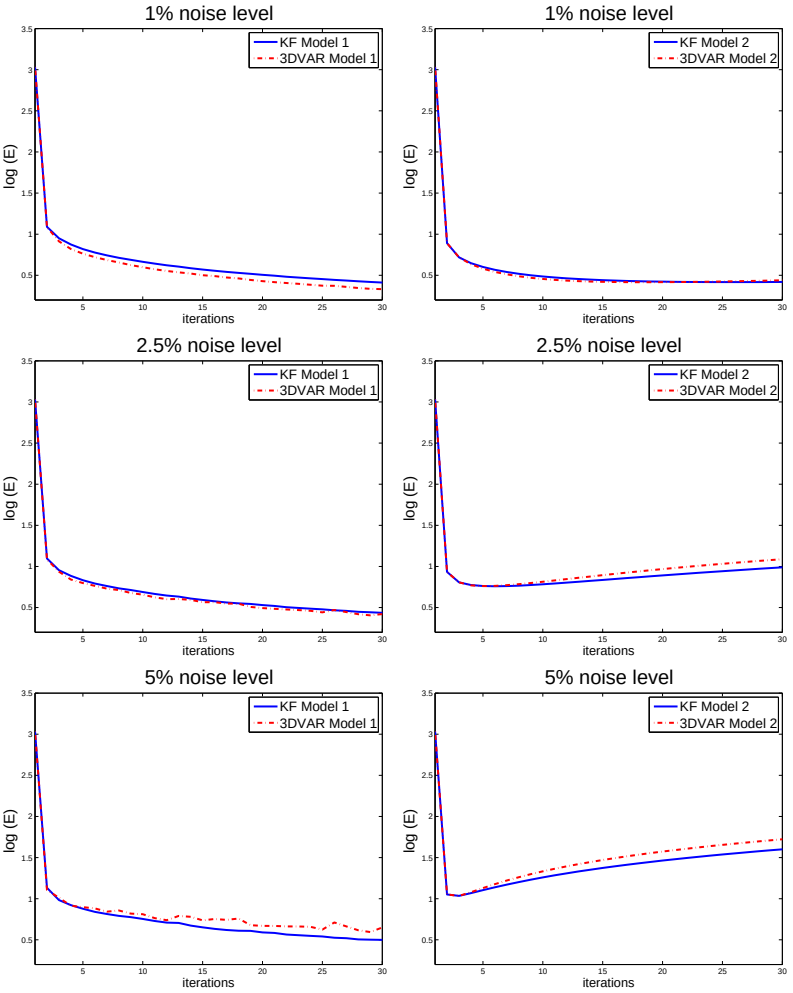

FIG. 6.2. *log Error with respect to the truth vs iterations of 3DVAR and Kalman filter applied to Data Model 1 (left) and Data Model 2 (right) for different noise levels.*

In Figure 6.1 we display the estimates m_n obtained with 3DVAR with Data Model 1 at iterations $n = 1, 10, 20, 30$ for noise levels (determined by the choice of γ) of 1% (top-middle), 2.5% (middle-bottom) and 5% (bottom). We can visually appreciate from Figure 6.1 that the estimates obtained at the final iteration $n = 30$ are indeed similar one to another even though the one in the bottom row was computed by inverting data five times noisier than the one in the first row. Similar estimates (not shown) were

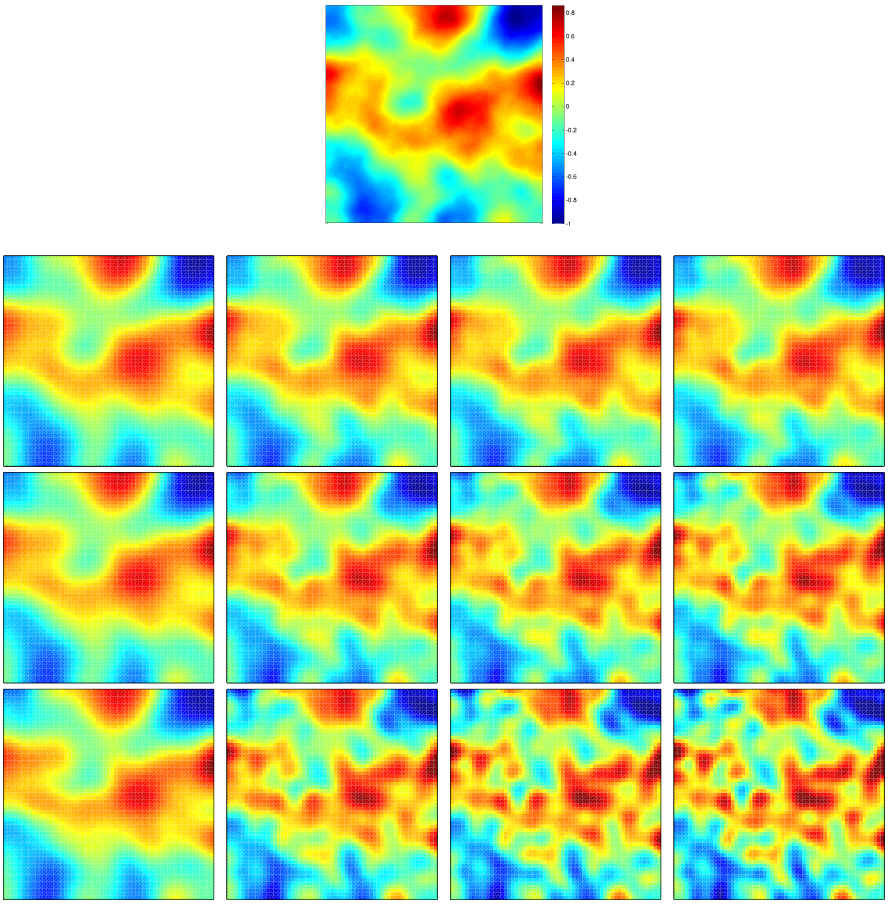

FIG. 6.3. *Top: truth* $u^\dagger$. *Top-middle, bottom-middle and bottom: Estimates obtained with 3DVAR and Data Model 2 at iterations (from left to right) 1, 10, 20, 30) for noise levels of* 1% *(top-middle),* 2.5% *(bottom-middle) and* 5% *(bottom) .*

obtained with the Kalman filter for Data Model 1.

For Data Model 2, the selection of γ corresponding to noise levels of 1%, 2.5%, and 5% yields a maximum number of iterations $N = 20, 6, 3$ respectively. Clearly, for Data Model 2, smaller observational noise results in schemes that can be iterated longer, presumably achieving more accurate estimates. Similarly to Data Model 1, we are required to stop the algorithm at $n = N$. However, for the purpose of this study we iterate until $n = 30$. In the right column of Figure 6.2 we display the plots of the error w.r.t the truth of m_n. In contrast to Data Model 1, we observe that the error w.r.t the truth increases for $n > N$. In other words, the Kalman filter and 3DVAR, when applied to Data Model 2, need to be stopped at $n = N$ in order to stabilize the scheme and obtain accurate estimates of the truth. Moreover, stopping the scheme at $n = N$ results in estimates whose accuracy increases with smaller noise levels. Clearly, in the small noise limit, both data models tend to exhibit similar behaviour. In Figure 6.3 we display m_n obtained from 3DVAR applied to Data Model 2 at iterations $n = 1, 10, 20, 30$ for data with noise levels of 1% (top-middle), 2.5% (middle-bottom) and 5% (bottom). Similar estimates (not shown) were obtained with the Kalman Filter for Data Model 2.

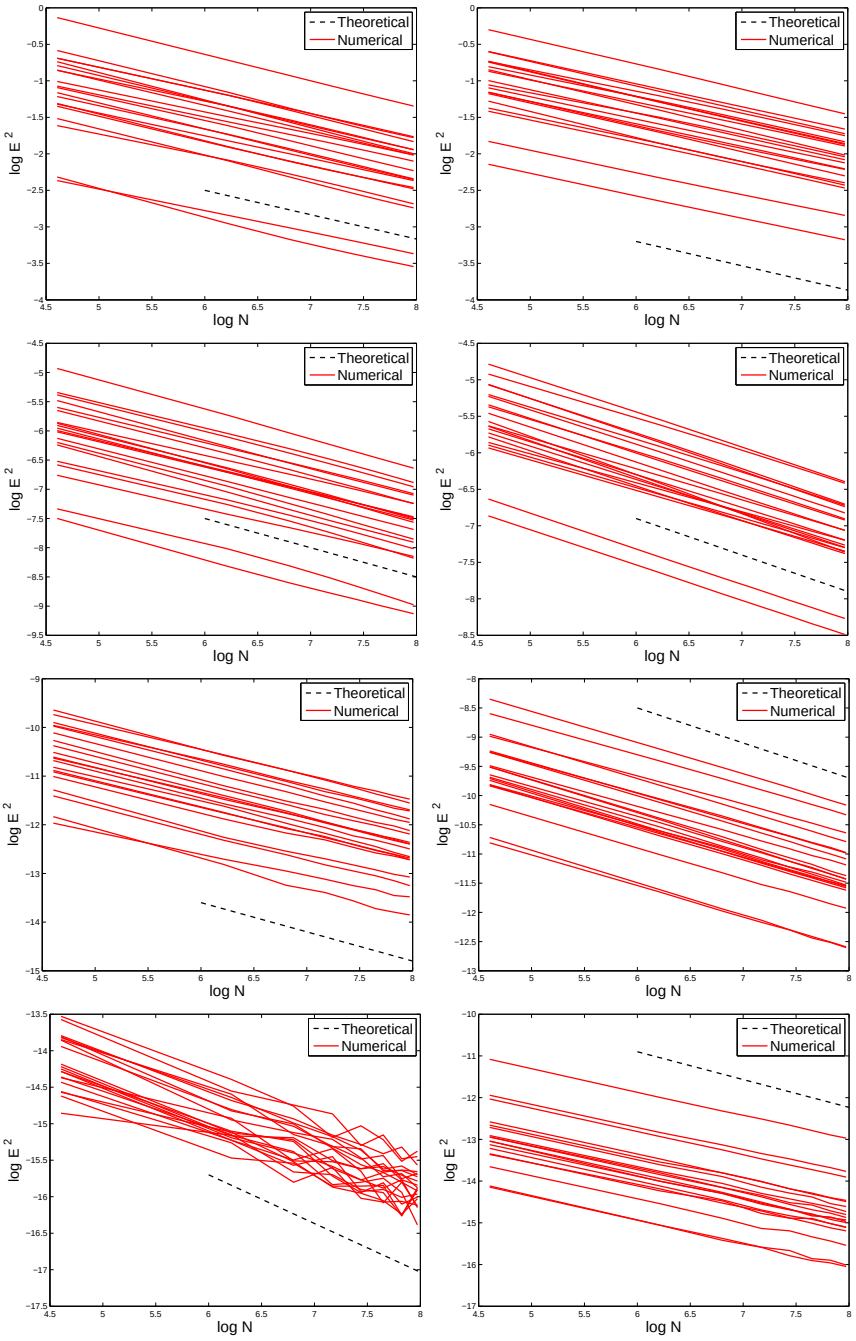

FIG. 6.4. *Convergence rates for 3DVAR (Left) and Kalman Filter (Right) with Data Model 1 and synthetic data generated from 20 different truths with regularity H^{2s} with s (from top to bottom) 1,2,3 and 4.*

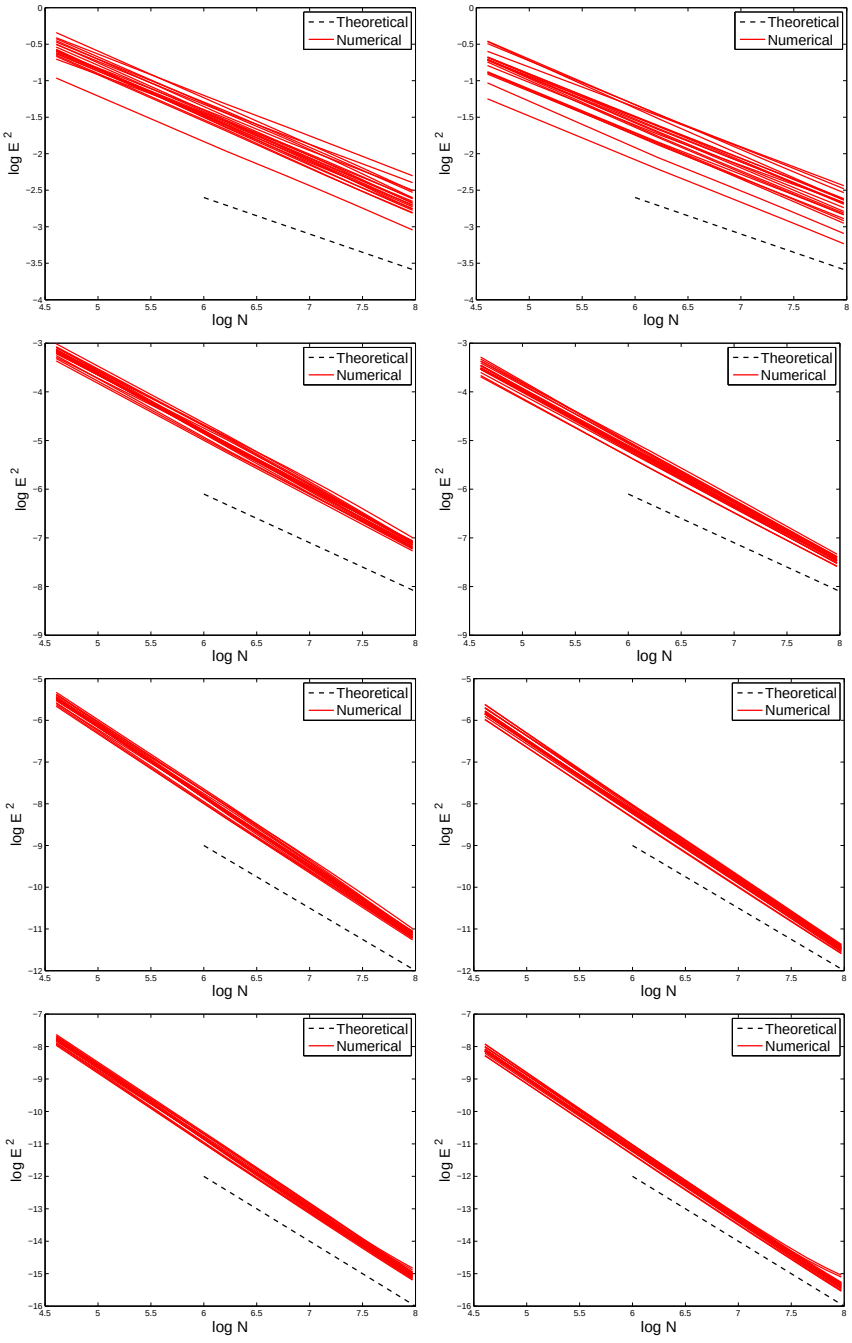

FIG. 6.5. *Convergence rates for 3DVAR (left) and Kalman Filter (right) with Data Model 2 and synthetic data generated from 20 different truths with regularity H^{2s} with s (from top to bottom) 1,2,3 and 4.*

It is clear indeed, that for Data Model 1, the application of Kalman filter and 3DVAR results in more accurate and stable estimates when the noise level in the data is not sufficiently small. The results from this subsection show that the reduction in the variance of the noise due to the law of large numbers and central limit theorem effect in Data Model 1 produces more accurate algorithms. Although Data Model 1 requires multiple instances of the data, in some applications such as in EIT [4], the data collection can be repeated multiple times in order to obtain these data.

6.3. Numerical verification of convergence rates. In this subsection we test the convergence rates from Theorems 4.1, 5.1, 4.3, and 5.3. For the verification of each of these rates we let $\Sigma := \Sigma_s$ denote the covariance from expression (6.2), and we consider 20 experiments corresponding to different truths $u^\dagger$ generated from $N(0,\Sigma_s)$ with the four selections of $s=1,2,3,4$. Note that Assumption 3.1 (iii) is satisfied only for $s \leq 3$. Again, for Data Model 1 we generate synthetic data for each of the truths and for as many iterations used for the application of both schemes. Inverse crimes are avoided as described in Subsection 6.1.

The verification of Theorem 4.1 and Theorem 5.1 by means of Algorithm 6.1 in the case of Data Model 1 is straightforward. For each of the set of synthetic data associated to each of the 20 truths $u^\dagger$ previously mentioned, we fix $\gamma = 5 \times 10^{-4}$. For each N (with $N = \{100,\ldots,3000\}$) we run Algorithm 6.1, stop the schemes at $n=N$ and record the value of $\|\zeta_N - u^\dagger\|^2$. In the right (resp. left) Figure 6.4 we display a plot of $\|\zeta_N - u^\dagger\|^2$ vs $\log N$ for the Kalman filter (resp. 3DVAR) for each of the set of 20 experiments associated to different truths (red solid lines) generated as described above with (from top to bottom) $s=1,2,3,4$. From Theorem 4.1 we note that the corresponding slopes of the convergence rates should be approximately given by $-\frac{s}{s+1+a}$. For Theorem 5.1 there is an additional term of $\log N$, but this is of course negligible compared to the algebraic decay and we ignore it for the purposes of this discussion. For comparison, a line (black dotted) with slope $-\frac{s}{s+1+a}$ is displayed in Figure 6.4.

We now verify the convergence rates of Theorem 4.3 and Theorem 5.3. Note first that in Algorithm 6.1 for Data Model 2 we define N in terms of the given small noise γ, in order to obtain convergence. However, for the purpose of the verification of the aforementioned convergence rates we define γ in terms of N by means of the same expressions. In other words, for each N ($N = \{100,\ldots,3000\}$) we produce synthetic data (or each of the 20 truths) with $\eta \sim N(0,\gamma^2)$ and $\gamma = N^{-\frac{a+s+1}{2(a+1)}}$. We then run Algorithm 6.1 and stop the schemes at $n=N$. In the right (resp. left) Figure 6.5 we display a plot of $\|\zeta_N - u^\dagger\|^2$ vs $\log N$ for the Kalman filter (resp. 3DVAR) for each of the set of 20 experiments associated to different truths (red solid lines) generated as before with (from top to bottom) $s=1,2,3,4$. We again include a line (black dotted) with slope of $-\frac{s}{a+1}$ which is the asymptotic behavior predicted by Theorems 4.3 and 5.3.

We can clearly appreciate that, for s satisfying Assumption 3.1 (iii) (i.e. $0 < s \leq 3$), the numerical convergence rates fit very well the ones predicted by the theory. Note that the higher the regularity of the truth (i.e. the larger the s), the smaller the error. w.r.t the truth in the estimates. We note that for $s=4$, the aforementioned assumption is violated and, in the case of Data Model 1, the slopes of the numerical convergence rates are slightly smaller than the theoretical ones. In this case ($s=4$) there are also fluctuations of the error w.r.t. the truth obtained with 3DVAR. These fluctuations may be associated with the fact that since for the error w.r.t. the truth is very small for sufficiently large iterations and for Data Model 1 the noise level is fixed a priori (recall $\gamma = 5 \times 10^{-4}$). However, for the Kalman filter these fluctuations are not so evident; presumably updating the covariance has a stabilizing effect. For Data Model 2, as N

increases, the corresponding γ decreases and so these fluctuations in the error are non existent.

7. Conclusions

- We have presented filter based algorithms for the linear inverse problem, based on introduction of an artificial dynamic. This results in methods which are closely related to iterated Tikhonov-type regularization. Two data scenarios are considered, one (Data Model 1) involving multiple realizations of the data, with independent noise; the other (Data Model 2) involving a single realization of the data; both are relevant in applications.
- We present theoretical results demonstrating convergence of the algorithms in the two data scenarios. For multiple realizations of the noisy the convergence is induced by the inherent averaging present in the iterated method, and the link to the law of large numbers and central limit theorem. For the single instance of data the small observational noise limit must be considered.
- For both Data Model 1 and Data Model 2 the Kalman Filter and 3DVAR produced very similar results for relatively small N ($N<100$). In practice it is clear that 3DVAR is preferable as the Kalman filter requires covariance updates which may be impractical for large scale models. However, updating the covariance in the Kalman filter seems to have an stabilizing effect in the error w.r.t the truth.
- For Data Model 1 the level of accuracy of the estimator is independent of the noise level. Moreover, the stability of the scheme is not conditioned to the early termination of the scheme. In contrast, for Data Model 2 we need to stop at $n=N$ to avoid an increase in the error w.r.t the truth. Again, this illustrates that, whenever multiple instances of the data are available, Data Model 1 offers a more stable and accurate framework for solving the inverse problems under consideration.
- The theoretical results from this work are verified numerically whenever the assumptions of the theory are satisfied.

Appendix A.

Proof. **(Proof of Lemma 4.1.)** The proof follows the classic arguments on Tikhonov regularization with Hilbert scales, c.f. [6, Ch8.4]. We recall inequality (3.1) in Assumption 3.1 (ii) and rewrite it in the form

$$c_1\|\Sigma_0^{\frac{a+1}{2}}x\| \le \|A\Sigma_0^{\frac{1}{2}}x\| \le c_2\|\Sigma_0^{\frac{a+1}{2}}x\|.$$

Notice that the definition of B_0 in expression (3.2) gives, since X is a Hilbert space, and using Assumption 3.1 (i) to ensure that B_0 and B_0^* are well-defined bounded linear operators,

$$\|A\Sigma_0^{\frac{1}{2}}x\| = \|B_0x\| = \|(B_0^*B_0)^{\frac{1}{2}}x\|.$$

Combining the two preceding displays we obtain

$$c_1\|\Sigma_0^{\frac{a+1}{2}}x\| \le \|(B_0^*B_0)^{\frac{1}{2}}x\| \le c_2\|\Sigma_0^{\frac{a+1}{2}}x\|$$

and a duality argument yield

$$c_2^{-1}\|\Sigma_0^{-\frac{a+1}{2}}x\| \le \|(B_0^*B_0)^{-\frac{1}{2}}x\| \le c_1^{-1}\|\Sigma_0^{-\frac{a+1}{2}}x\|$$

for any $x \in \mathcal{R}(\Sigma_0^{\frac{a+1}{2}})$. Let $\theta \in [-1,1]$. Then the inequality of Heinz [6, Ch.8.4, pp. 213] and an additional duality argument gives

$$c_1 \|\Sigma_0^{\frac{\theta(a+1)}{2}} x\| \le \|(B_0^* B_0)^{\frac{\theta}{2}} x\| \le c_2 \|\Sigma_0^{\frac{\theta(a+1)}{2}} x\|, \tag{A.1}$$

which yields $\mathcal{R}\left((B_0^* B_0)^{\frac{\theta}{2}}\right) = \mathcal{D}\left(\Sigma_0^{-\frac{\theta(a+1)}{2}}\right)$.

Let Assumption 3.1 (*iii*) be valid with $m_0 - u^\dagger \in \mathcal{D}\left(\Sigma_0^{-\frac{s}{2}}\right)$, and define $z^\dagger := \Sigma_0^{-\frac{1}{2}}(m_0 - u^\dagger) \in \mathcal{D}\left(\Sigma_0^{-\frac{s-1}{2}}\right)$. Since $s-1 \in [-1, a+1]$, and consequently $\theta = \frac{s-1}{a+1} \in (-1,1]$, we obtain from inequality (A.1) that $z^\dagger \in R((B_0^* B_0)^{\frac{s-1}{2(a+1)}})$. Furthermore there exists a $v \in X$ such that

$$z^\dagger = (B_0^* B_0)^{\frac{s-1}{2(a+1)}} v.$$

Noting that $\frac{1}{a+1} \in (0,1)$, and employing inequality (A.1) with $\theta = \frac{1}{a+1}$, together with equation (4.2) and inequality (4.4), we have

$$\begin{aligned}
\|J_1\|^2 &= \|\Sigma_0^{\frac{1}{2}} r_{1,\frac{\alpha}{n}}(B_0^* B_0) \Sigma_0^{-\frac{1}{2}}(m_0 - u^\dagger)\|^2 \\
&= \|(B_0^* B_0)^{\frac{1}{2(a+1)}} r_{1,\frac{\alpha}{n}}(B_0^* B_0) z^\dagger\|^2 \\
&= \|(B_0^* B_0)^{\frac{1}{2(a+1)}} r_{1,\frac{\alpha}{n}}(B_0^* B_0)(B_0^* B_0)^{\frac{s-1}{2(a+1)}} v\|^2 \\
&= \|(B_0^* B_0)^{\frac{s}{2(a+1)}} r_{1,\frac{\alpha}{n}}(B_0^* B_0) v\|^2 \\
&\le \left(\frac{\alpha}{n}\right)^{\frac{s}{a+1}} \|v\|^2.
\end{aligned}$$

In case of Assumption 3.2 we insert $a = \frac{2p}{1+2\epsilon}$ and $s = \frac{2\beta}{1+2\epsilon}$. ∎

Proof. **(Proof of Lemma 4.2.)** Notice that

$$\begin{aligned}
K_n &= C_{n-1} A^* (A C_{n-1} A^* + \gamma^2 I)^{-1} \\
&= C_{n-1}^{\frac{1}{2}} C_{n-1}^{\frac{1}{2}} A^* (A C_{n-1}^{\frac{1}{2}} C_{n-1}^{\frac{1}{2}} A^* + \gamma^2 I)^{-1} \\
&= C_{n-1}^{\frac{1}{2}} (C_{n-1}^{\frac{1}{2}} A^* A C_{n-1}^{\frac{1}{2}} + \gamma^2 I)^{-1} C_{n-1}^{\frac{1}{2}} A^*
\end{aligned}$$

and

$$C_n = (I - K_n A) C_{n-1} = \gamma^2 C_{n-1}^{\frac{1}{2}} (C_{n-1}^{\frac{1}{2}} A^* A C_{n-1}^{\frac{1}{2}} + \gamma^2 I)^{-1} C_{n-1}^{\frac{1}{2}}.$$

Thus we obtain

$$C_n A^* = \gamma^2 K_n. \tag{A.2}$$

By virtue of equations (3.5) and (A.2) we derive

$$\begin{aligned}
J_2 &= \sum_{j=1}^{n-1} \left(\prod_{i=n-j}^{n-1} (I - K_{i+1} A) \right) K_{n-j} \eta_{n-j} + K_n \eta_n \\
&= \sum_{j=0}^{n-1} \left(C_n C_{n-j}^{-1} K_{n-j} \right) \eta_{n-j}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{j=0}^{n-1} \left(C_n A^*/\gamma^2\right) \eta_{n-j} \\
&= \sum_{j=0}^{n-1} \left(\left(C_0^{-1} + n\frac{A^*A}{\gamma^2} \right)^{-1} A^*/\gamma^2 \right) \eta_{n-j}.
\end{aligned}
$$

We denote $F := \left(C_0^{-1} + n\frac{A^*A}{\gamma^2} \right)^{-1} A^*/\gamma^2$ and obtain

$$
\mathbb{E}\|J_2\|^2 = \sum_{j=0}^{n-1} \mathbb{E}\|F\eta_{n-j}\|^2 = n\gamma^2 \mathrm{tr}(FF^*).
$$

By the definition of F and Assumption 3.1 (i), (iv) we obtain

$$
F = \left(C_0^{-1} + n\frac{A^*A}{\gamma^2} \right)^{-1} A^*/\gamma^2 = \Sigma_0^{\frac{1}{2}} (\alpha I + nB_0^*B_0)^{-1} B_0^*
$$

and consequently derive

$$
\begin{aligned}
\mathrm{tr}(FF^*) &= \mathrm{tr}\left(\left(\Sigma_0^{\frac{1}{2}} (\alpha I + nB_0^*B_0)^{-1} B_0^* \right) \left(B_0 (\alpha I + nB_0^*B_0)^{-1} \Sigma_0^{\frac{1}{2}} \right) \right) \\
&= \mathrm{tr}((\alpha I + nB_0^*B_0)^{-2} B_0^*B_0\Sigma_0) \\
&\le \|(\alpha I + nB_0^*B_0)^{-2} B_0^*B_0\| \mathrm{tr}(\Sigma_0) \\
&= \frac{1}{\alpha^2} \|r_{2,\frac{\alpha}{n}}(B_0^*B_0) B_0^*B_0\| \mathrm{tr}(\Sigma_0) \\
&\le \frac{1}{\alpha^2}\frac{\alpha}{n}\mathrm{tr}(\Sigma_0) = \frac{1}{\alpha n}\mathrm{tr}(\Sigma_0).
\end{aligned}
$$

With the operator-valued function $r_{2,\frac{\alpha}{n}}(\lambda) := \left(\frac{\frac{\alpha}{n}}{\frac{\alpha}{n}+\lambda} \right)^2 = \left(\frac{\alpha}{\alpha+n\lambda} \right)^2$. Such an observation then yields

$$
\mathbb{E}\|J_2\|^2 = n\gamma^2 \mathrm{tr}(FF^*) \le \frac{\gamma^2}{\alpha}\mathrm{tr}(\Sigma_0).
$$

Concerning Assumption 3.2, we further estimate, by exploiting [14, Lemma 8.2],

$$
\begin{aligned}
\mathrm{tr}(FF^*) &= \frac{1}{\alpha^2} \sum_{i=1}^{\infty} \frac{i^{-(1+2\epsilon+2p)-(1+2\epsilon)}}{\left(1 + \frac{n}{\alpha} i^{-(1+2\epsilon+2p)}\right)^2} \\
&= \frac{1}{n} \sum_{i=1}^{\infty} \frac{\frac{n}{\alpha^2} i^{-4\epsilon-2p-2}}{\left(1 + \frac{n}{\alpha} i^{-2\epsilon-2p-1}\right)^2} \\
&\asymp \frac{1}{n\alpha} \left(\frac{n}{\alpha} \right)^{-\frac{2\epsilon}{1+2\epsilon+2p}}
\end{aligned}
$$

and

$$
\mathbb{E}\|J_2\|^2 \asymp \gamma^2 n^{-\frac{2\epsilon}{1+2\epsilon+2p}} \alpha^{-\frac{1+2p}{1+2\epsilon+2p}}.
$$

∎

Proof. **(Proof of Lemma 4.3.)** Notice that for Data Model 2, we derive

$$\begin{aligned} J_2 &= \sum_{j=1}^{n-1} \left(\prod_{i=n-j}^{n-1} (I - K_{i+1}A) \right) K_{n-j}\eta_{n-j} + K_n\eta_n \\ &= \sum_{j=0}^{n-1} \left(\left(C_0^{-1} + n\frac{A^*A}{\gamma^2} \right)^{-1} A^*/\gamma^2 \right) \eta_{n-j} \\ &= nF\eta \end{aligned}$$

which yields

$$\mathbb{E}\|J_2\|^2 = n^2\gamma^2 \mathrm{tr}(FF^*).$$

The remainder of the proof follows Lemma 4.2. ∎

Proof. **(Proof of Theorem 5.2.)** As for the other theorems, the proof rests, of course, on the bias variance decomposition, and then use of Lemmas 5.1 and 5.2. This yields

$$\mathbb{E}\|\zeta_n - u^\dagger\|^2 \le C\left(\frac{\alpha}{n}\right)^{\frac{2\beta}{1+2\epsilon+2p}} + C\gamma^2\alpha^{-\frac{1+2p}{1+2\epsilon+2p}}$$

and simultaneously

$$\mathbb{E}\|\zeta_n - u^\dagger\|^2 \le C\left(\frac{\alpha}{n}\right)^{\frac{2\beta}{1+2\epsilon+2p}} + C\frac{\gamma^2 \ln n}{\alpha}.$$

Choosing $\alpha = N^{\frac{2\beta}{1+2\beta+2p}}$ for the former inequality and $\alpha = N^{\frac{2\beta}{1+2\epsilon+2\beta+2p}}$ for the latter inequality, we conclude that, by stopping the iteration when $n = N$,

$$\mathbb{E}\|\zeta_N - u^\dagger\|^2 \le CN^{-\frac{2p}{1+2\epsilon+2\beta+2p}} \ln N.$$

∎

Proof. **(Proof of Lemma 5.1.)** Analogously to the proof of Lemma 4.1, it may be shown that

$$\begin{aligned} \|I_1\|^2 &= \|\Sigma_0^{\frac{1}{2}} r_{n,\alpha}(B_0^*B_0)\Sigma_0^{-\frac{1}{2}}(u^\dagger - m_0)\|^2 \\ &\le \|(B_0^*B_0)^{\frac{1}{2(a+1)}} r_{n,\alpha}(B_0^*B_0) z^\dagger\|^2 \\ &= \|(B_0^*B_0)^{\frac{1}{2(a+1)}} r_{n,\alpha}(B_0^*B_0)(B_0^*B_0)^{\frac{s-1}{2(a+1)}} v\|^2 \\ &= \|(B_0^*B_0)^{\frac{s}{2(a+1)}} r_{n,\alpha}(B_0^*B_0) v\|^2 \\ &\le \left(\frac{\alpha}{n}\right)^{\frac{s}{a+1}} \|v\|^2. \end{aligned}$$

The final inequality follows from the asymptotic behavior of $r_{n,\alpha}(\lambda)$, established, for example, in [16, Ch.2, pp. 63]. In the case of Assumption 3.2 we insert $a = \frac{2p}{1+2\epsilon}$ and $s = \frac{2\beta}{1+2\epsilon}$. ∎

Proof. **(Proof of Lemma 5.2.)** We denote $F_j = (I - \mathcal{K}A)^j\mathcal{K}$ and obtain

$$I_2 = \sum_{j=0}^{n-1} F_j\eta_{n-j}.$$

Furthermore, we derive

$$\mathbb{E}\|I_2\|^2 = \sum_{j=0}^{n-1} \mathbb{E}\|F_j \eta_{n-j}\|^2 = \gamma^2 \sum_{j=0}^{n-1} \mathrm{tr}(F_j F_j^*)$$

and

$$\begin{aligned}
\sum_{j=0}^{n-1} \mathrm{tr}\left(F_j F_j^*\right) &= \sum_{j=0}^{n-1} \mathrm{tr}\left((I-\mathcal{K}A)^j \mathcal{K}\mathcal{K}^* \left((I-\mathcal{K}A)^*\right)^j\right) \\
&= \frac{1}{\alpha^2} \sum_{j=0}^{n-1} \mathrm{tr}\left(\Sigma_0^{\frac{1}{2}} r_{2j+2,\alpha}(B_0^* B_0) B_0^* B_0 \Sigma_0^{\frac{1}{2}}\right) \\
&\leq \frac{1}{\alpha^2} \sum_{j=0}^{n-1} \|r_{2j+2,\alpha}(B_0^* B_0) B_0^* B_0\| \mathrm{tr}(\Sigma_0) \\
&\leq \frac{\mathrm{tr}(\Sigma_0)}{\alpha^2} \sum_{j=0}^{n-1} \frac{\alpha}{2j+2} \\
&\asymp C \ln(n) \frac{\mathrm{tr}(\Sigma_0)}{\alpha}.
\end{aligned}$$

Thus,

$$\mathbb{E}\|I_2\|^2 \leq C \frac{\ln(n)\gamma^2}{\alpha} \mathrm{tr}(\Sigma_0).$$

For Assumption 3.2 we need to estimate $\mathrm{tr}(F_j F_j^*)$ carefully. We substitute the given decay rate of the different eigenvalues, to obtain

$$\begin{aligned}
\mathrm{tr}(F_j F_j^*) &= \frac{1}{\alpha^2} \mathrm{tr}\left(\Sigma_0^{\frac{1}{2}} r_{2j+2,\alpha}(B_0^* B_0) B_0^* B_0 \Sigma_0^{\frac{1}{2}}\right) \\
&= \frac{1}{\alpha^2} \sum_{i=1}^{\infty} \frac{i^{-2-4\epsilon-2p}}{\left(1+\frac{1}{\alpha} i^{-1-2\epsilon-2p}\right)^{2j+2}} \\
&\leq \frac{1}{\alpha^2} \sum_{i=1}^{\infty} \frac{i^{-2-4\epsilon-2p}}{\left(1+\frac{j+1}{\alpha} i^{-1-2\epsilon-2p}\right)^{2}}.
\end{aligned}$$

By arguments similar to those used in the proof of Lemma 4.2 (or [14, Lemma 8.2]), we further estimate

$$\frac{1}{\alpha^2} \sum_{i=1}^{\infty} \frac{i^{-2-4\epsilon-2p}}{\left(1+\frac{j+1}{\alpha} i^{-1-2\epsilon-2p}\right)^{2}} \asymp \left(\frac{1}{j+1}\right)^{1+\frac{2\epsilon}{1+2\epsilon+2p}} \alpha^{-\frac{1+2p}{1+2\epsilon+2p}}$$

and

$$\mathbb{E}\|I_2\| \leq C\gamma^2 \alpha^{-\frac{1+2p}{1+2\epsilon+2p}} \sum_{j=0}^{n-1} \left(\frac{1}{j+1}\right)^{1+\frac{2\epsilon}{1+2\epsilon+2p}}$$

where the summation term in the right-hand side is bounded.

On the other hand, we can also estimate

$$\begin{aligned}\mathrm{tr}(F_jF_j^*) &= \frac{1}{\alpha^2}\mathrm{tr}\left(\Sigma_0^{\frac{1}{2}} r_{2j+2,\alpha}(B_0^*B_0)B_0^*B_0\Sigma_0^{\frac{1}{2}}\right)\\ &= \frac{1}{\alpha^2}\sum_{i=1}^{\infty}\frac{i^{-2-4\epsilon-2p}}{\left(1+\frac{1}{\alpha}i^{-1-2\epsilon-2p}\right)^{2j+2}}\\ &\le \frac{1}{\alpha^2}\sum_{i=1}^{\infty}\frac{i^{-2-4\epsilon-2p}}{\left(1+\frac{2(j+1)}{\alpha}i^{-1-2\epsilon-2p}\right)}\\ &< \frac{1}{2(j+1)\alpha}\sum_{i=1}^{\infty} i^{-1-2\epsilon}\\ &\le C\frac{1}{(j+1)\alpha}\end{aligned}$$

and

$$\mathbb{E}\|I_2\| \le C\frac{\gamma^2\ln n}{\alpha}.$$

□

Proof. **(Proof of Lemma 5.3.)** Since $\eta_n = \eta$ in this case, we derive

$$I_2 = \sum_{j=0}^{n-1}(I-\mathcal{K}A)^j\mathcal{K}\eta$$

and by operator-valued calculation we obtain

$$\begin{aligned}\sum_{j=0}^{n-1}(I-\mathcal{K}A)^j\mathcal{K} &= \frac{1}{\alpha}\Sigma_0^{\frac{1}{2}}\sum_{j=0}^{n-1}\left(\alpha(\alpha I+B_0^*B_0)^{-1}\right)^{j+1}B_0^*\\ &= \Sigma_0^{\frac{1}{2}}q_{n,\alpha}(B_0^*B_0)B_0^*\end{aligned}$$

where $q_{n,\alpha}(\lambda) := \frac{1}{\lambda}\left(1-\frac{\alpha^n}{(\alpha+\lambda)^n}\right)$. Thus we obtain, by the asymptotic behavior of $q_{n,\alpha}(\lambda)$ derived in [16, Ch.2, pp. 64],

$$\begin{aligned}\mathbb{E}\|I_2\|^2 &= \gamma^2\mathrm{tr}\left(\Sigma_0^{\frac{1}{2}}q_{n,\alpha}(B_0^*B_0)B_0^*B_0q_{n,\alpha}(B_0^*B_0)\Sigma_0^{\frac{1}{2}}\right)\\ &\le \gamma^2\|q_{n,\alpha}(B_0^*B_0)(B_0^*B_0)^{\frac{1}{2}}\|^2\mathrm{tr}(\Sigma_0)\\ &\le \frac{n\gamma^2}{\alpha}\mathrm{tr}(\Sigma_0).\end{aligned}$$

□

Proof. **(Proof of Theorem 5.4.)** Similar to the Kalman filter method and 3DVAR, by Assumption 3.1 (i)-(iii), we obtain the bias error estimate

$$\begin{aligned}\|\mathcal{I}_1\| &\le \left\|\Sigma_0^{1/2}\prod_{j=1}^{n}\frac{\alpha_j I}{B_0^*B_0+\alpha_j I}\Sigma_0^{1/2}\epsilon_0\right\|^2\\ &\le \left\|\prod_{j=1}^{n}\left(\frac{\alpha_j I}{B_0^*B_0+\alpha_j I}\right)(B_0^*B_0)^{\frac{s}{2(a+1)}}\upsilon\right\|^2.\end{aligned}$$

Now we need upper bounds of the following operator-valued function

$$f_{n,v}(\lambda) = \lambda^v \prod_{j=1}^{n} \frac{\alpha_j}{\lambda + \alpha_j}, \quad \lambda \in (0,\infty), \quad n > v > 0. \tag{A.3}$$

Define $\sigma_n := \sum_{j=1}^{n} \frac{1}{\alpha_j}$ and assume the sequence $\{\alpha_j\}_{j=1}^{n}$ satisfying

$$\frac{1}{\alpha_n} \leq \tilde{c}\sigma_{n-1} \tag{A.4}$$

with a constant $\tilde{c}$. Then the results in [9] yield

$$\|\mathcal{I}_1\| \leq C\sigma_n^{-\frac{s}{a+1}}, \quad n > 1.$$

It remains to estimate the $\mathcal{I}_2$ term. Define

$$\begin{aligned}\mathcal{F}_j &:= \prod_{i=n-j}^{n-1} (I - \mathcal{K}_{i+1}A)\mathcal{K}_{n-j} \\ &= \Sigma_0^{1/2} \prod_{i=n-j}^{n-1} \left(\frac{\alpha_{i+1} I}{B_0^* B_0 + \alpha_{i+1} I}\right) \frac{1}{B_0^* B_0 + \alpha_{n-j} I} B_0^*.\end{aligned}$$

Then we obtain

$$\mathcal{I}_2 = \sum_{j=1}^{n-1} \mathcal{F}_j \eta_{n-j} + K_n \eta_n$$

which yields

$$\mathbb{E}\|\mathcal{I}_2\|^2 = \gamma^2 \sum_{j=1}^{n-1} \mathrm{tr}(\mathcal{F}_j \mathcal{F}_j^*) + \gamma^2 \mathrm{tr}(K_n K_n^*).$$

Notice that for any $j \geq 1$

$$\mathcal{F}_j \mathcal{F}_j^* = \frac{1}{\alpha_{n-j}^2} \Sigma_0^{1/2} \prod_{i=n-j}^{n-1} \left(\frac{\alpha_{i+1}}{B_0^* B_0 + \alpha_{i+1} I}\right)^2 \left(\frac{\alpha_{n-j}}{B_0^* B_0 + \alpha_{n-j} I}\right)^2 B_0^* B_0 \Sigma_0^{1/2}$$

we derive,

$$\begin{aligned}\mathrm{tr}(\mathcal{F}_j \mathcal{F}_j^*) &\leq \frac{\mathrm{tr}(\Sigma_0)}{\alpha_{n-j}^2} \left\| \prod_{i=n-j}^{n-1} \left(\frac{\alpha_{i+1}}{B_0^* B_0 + \alpha_{i+1} I}\right)^2 \left(\frac{\alpha_{n-j}}{B_0^* B_0 + \alpha_{n-j} I}\right)^2 B_0^* B_0 \right\| \\ &\leq \frac{\mathrm{tr}(\Sigma_0)}{\alpha_{n-j}^2} \left\| \left(\frac{\alpha_{n-j}}{B_0^* B_0 + \alpha_{n-j} I}\right)^2 B_0^* B_0 \right\| \\ &\leq \frac{1}{2\alpha_{n-j}} \mathrm{tr}(\Sigma_0)\end{aligned}$$

and

$$\mathrm{tr}(K_n K_n^*) \leq \frac{1}{2\alpha_n} \mathrm{tr}(\Sigma_0).$$

A rough variance estimate for the variant method is

$$\begin{aligned}\mathbb{E}\|v_n - u^\dagger\|^2 &= \|\mathcal{I}_1\|^2 + \mathbb{E}\|\mathcal{I}_2\|^2 \\ &\le C\left(\sigma_n^{-\frac{s}{a+1}} + \gamma^2 \mathrm{tr}(\Sigma_0)\sigma_n\right).\end{aligned}$$

The first term vanishes but the second term blows up when $n \to \infty$ and $\sigma_n \to \infty$.

To further investigate the blow up, we consider a special geometric sequence $\alpha_n = \alpha q^{n-1}$ with $0 < q < 1$. Thus, we have

$$\sigma_n = \alpha^{-1} q^{1-n} \frac{1-q^n}{1-q} \ge \alpha^{-1} q^{1-n} = q/\alpha_{n+1}$$

and inequality (A.4) is satisfied with $\tilde{c} = 1/q$. Actually, we derive

$$\frac{1}{\alpha_{n-j}} + \sigma_n - \sigma_{n-j} = \frac{1}{\alpha_{n-j}} + \frac{1}{\alpha_{n-j+1}} + \ldots + \frac{1}{\alpha_n} = \alpha^{-1} q^{1-n} \frac{1-q^{j+1}}{1-q} \ge \alpha^{-1} q^{1-n} = q/\alpha_{n+1}.$$

Thus the results in [9] refine, by the asymptotic behavior of (A.3),

$$\begin{aligned}\mathrm{tr}(\mathcal{F}_j \mathcal{F}_j^*) &\le \frac{\mathrm{tr}(\Sigma_0)}{\alpha_{n-j}^2} \left\| \prod_{i=n-j}^{n-1} \left(\frac{\alpha_{i+1}}{B_0^* B_0 + \alpha_{i+1} I}\right)^2 \left(\frac{\alpha_{n-j}}{B_0^* B_0 + \alpha_{n-j} I}\right)^2 B_0^* B_0 \right\| \\ &\le \frac{1}{\alpha_{n-j}^2} \left(\frac{1}{\alpha_{n-j}} + \sigma_n - \sigma_{n-j}\right)^{-1} \mathrm{tr}(\Sigma_0) \\ &= \alpha^{-1} q^{1+j-n} \left(1 + q^{-j} \frac{1-q^j}{1-q}\right)^{-1} \mathrm{tr}(\Sigma_0) \\ &\le \alpha^{-1} q^{1+2j-n} \mathrm{tr}(\Sigma_0)\end{aligned}$$

and we derive

$$\mathbb{E}\|\mathcal{I}_2\|^2 \le \frac{q\gamma^2 \mathrm{tr}(\Sigma_0)}{\alpha} q^{-n} \left(\sum_{j=1}^{n-1} q^{2j} + 1\right).$$

Summing up, for the geometric sequence, we obtain

$$\mathbb{E}\|v_n - u^\dagger\|^2 \le C\left(q^{\frac{s}{a+1}n} + \gamma^2 \mathrm{tr}(\Sigma_0) q^{-n}\right).$$

The second term blows up exponentially when n goes to infinity. ∎

Acknowledgements. Shuai Lu is supported by NSFC No.11522108, Shanghai Municipal Education Commission No.16SG01 and Special Funds for Major State Basic Research Projects of China (2015CB856003). Andrew M. Stuart is funded by EPSRC (under the Programme Grant EQUIP), DARPA (under EQUiPS) and The Office of Naval Research under Award Number N000141712079.

REFERENCES

[1] S. Agapiou, S. Larsson, and A.M. Stuart, *Posterior contraction rates for the Bayesian approach to linear ill-posed inverse problems*, Stochastic Process. Appl., 123(10):3828–3860, 2013.

[2] N. Bissantz, T. Hohage, A. Munk, and F. Ruymgaart, *Convergence rates of general regularization methods for statistical inverse problems and applications*, SIAM J. Numer. Anal., 45:2610–2636, 2007.

[3] V.I. Bogahcev, *Gaussian Measures*, *Mathematical Surveys and Monographs*, American Mathematical Society, 62:433, 1998.

[4] L. Borcea, *Electrical Impedance Tomography*, Inverse Problems Series, 18(6), 2002.

[5] M. Dashti and A.M. Stuart, *The Bayesian Approach to Inverse Problems*, in The Handbook of Uncertainty Quantification, R. Ghanem, D. Higdon and H. Owhadi (Eds.) , Springer, 1–118, 2016.

[6] H.W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Math. Appl., Kluwer Academic Publishers Group, Dordrecht, 375, 1996.

[7] A. Goldenshluger and S.V. Pereverzev, *On adaptive inverse estimation of linear functionals in Hilbert scales*, Bernoulli, 9:783–807, 2003.

[8] M. Hanke, *A regularizing Levenberg–Marquardt scheme, with applications to inverse groundwater filtration problems*, Inverse Problems, 13(1):79–95, 1997.

[9] M. Hanke and C.W. Groetsch, *Nonstationary iterated Tikhonov regularization*, Journal of Optimization Theory and Applications, 98(1):37–53, 1998.

[10] T. Hohage and M. Pricop, *Nonlinear Tikhonov regularization in Hilbert scales for inverse boundary value problems with random noise*, Inv. Probl. Imag., 2:271–290, 2008.

[11] M.A. Iglesias, K.J. Law, and A.M. Stuart, *Ensemble Kalman methods for inverse problems*, Inverse Problems, 29(4):045001, 2013.

[12] M.A. Iglesias, *A regularizing iterative ensemble Kalman method for PDE-constrained inverse problems*, Inverse Problems, 32(2), 2016.

[13] J. Kaipio and E. Somersalo, *Statistical and computational inverse problems*, Appl. Math. Sci., Springer-Verlag, New York, 160, 2005.

[14] B.T. Knapik, A.W. Van Der Vaart, and J.H. Van Zanten, *Bayesian inverse problems with Gaussian priors*, Annal. Stat., 39(5), 2626–2657, 2011.

[15] K.J. Law, A.M. Stuart, and K.C. Zygalakis, *Data Assimilation: A Mathematical Introduction*, Springer, 2015.

[16] Lu S. and S.V. Pereverzev, *Regularization Theory for Ill-Posed Problems. Selected Topics*, Inverse and Ill-posed Problems Series, Walter De Gruyter, Berlin, 58, 2013.

[17] B.A. Mair and F. Ruymgaart, *Statistical inverse estimation in Hilbert scales*, SIAM J. Appl. Math., 56:1424–1444, 1996.

[18] P. Mathé and S.V. Pereverzev, *Optimal discretization of inverse problems in Hilbert scales. Regularization and self-regularization of projection methods*, SIAM J. Numer. Anal., 38:1999–2021, 2001.

[19] D. Oliver, A. Reynolds, and N. Liu, *Inverse Theory for Petroleum Reservoir Characterization and History Matching*, Cambridge Univ Press, 2008.