ITERATE AVERAGING AND FILTERING ALGORITHMS FOR LINEAR INVERSE PROBLEMS

FELIX G. JONES AND GIDEON SIMPSON

Abstract. It has been proposed that classical filtering methods, like the Kalman filter and 3DVAR, can be used to solve linear statistical inverse problems. In the work of Igelsias, Lin, Lu, & Stuart (2017), [3], error estimates were obtained for this approach. By optimally tuning a free parameter in the filters, the authors were able to show that the mean squared error can be minimized.

In the present work, we prove that by (i) considering the problem in a weaker, weighted, space and (ii) applying simple iterate averaging of the filter output, 3DVAR will converge in mean square, unconditionally on the parameter. Without iterate averaging, 3DVAR cannot converge by running additional iterations with a given, fixed, choice of parameter. We also establish that the Kalman filter’s performance cannot be improved through iterate averaging. We illustrate our results with numerical experiments that suggest our convergence rates are sharp.

1. Introduction

The focus of this work is on the inverse problem

\[ y = Au^\dagger + \eta, \]

where given the noisy observation \( y \) of \( Au^\dagger \), we wish to infer \( u^\dagger \). In our setting, \( A : X \to Y \) is a compact operator between Hilbert spaces and \( \eta \sim N(0, \gamma^2 I) \) is white noise, modelling measurement error. This problem is well known to be ill-posed in the infinite dimensional setting, as \( A \) has an unbounded inverse.

In the work of [3], the authors considered two classical filtering algorithms, the Kalman filter and 3DVAR, with the goal of using them to solve (1.1). As discussed in [3], the filtering methodology for (1.1) requires the introduction, conceptually, of the artificial dynamical system

\[
\begin{align*}
(1.2a) & \quad u_n = u_{n-1}, \\
(1.2b) & \quad y_n = Au_n + \eta_n.
\end{align*}
\]

Here, at algorithmic step \( n \), \( u_n \) is the quantity of interest, and \( y_n \) is the noisy observation. Having ascribed a notion of time to the problem, we can then apply a filter. This provides a mechanism for estimating \( u^\dagger \) in (1.1) in an online setting, where a sequence of i.i.d. observations, \( \{y_n\} \), is available. This corresponds to “Data Model 1” of [3].
Amongst the key results of [3], reviewed in detail below, is that under sufficiently strong assumptions, the Kalman filter will recover the truth in mean square, unconditionally on the choice of parameters in the filter. Under somewhat weaker assumptions, the error will only be bounded, though through minimax selection of the parameter, an optimal error can be achieved for a given number of iterations.

3DVAR is a simplification of Kalman that has is demonstrated to have, at best, bounded error, though, again, through minimax parameter tuning, it performs comparably to Kalman. Kalman is more expensive than 3DVAR, as it requires updating an entire covariance operator at each iteration. For finite dimensional approximations, this may require costly matrix-matrix multiplications.

Here, by working in a weaker, weighted, norm and averaging the iterates, we are able to establish that 3DVAR will unconditionally converge in mean square for all admissible filter parameters. Further, we show that this simple iterate averaging cannot improve the performance of the Kalman filter.

1.1. Filtering Algorithms. The Kalman filter is a probabilistic filter that estimates a Gaussian distribution, \( N(m_n, C_n) \), for \( u^\dagger \) at each iterate. Given a starting mean and covariance, \( m_0 \) and \( C_0 \), the updates are as follows:

\[
(1.3a) \quad m_n = K_n y_n + (I - K_n A) m_{n-1},
\]

\[
(1.3b) \quad C_n = (I - K_n A) C_{n-1},
\]

\[
(1.3c) \quad K_n = C_{n-1} A^* (AC_{n-1} A^* + \gamma^2 I)^{-1}.
\]

Here, \( K_n \) is the so-called “Kalman gain.” \( m_n \) is a point estimate of \( u^\dagger \).

While Kalman is a probabilistic filter, 3DVAR is not. It is obtained by applying Kalman with a fixed covariance operator. \( C_n = \frac{\alpha}{2} \Sigma \) for some fixed operator \( \Sigma \):

\[
(1.4a) \quad u_n = A^* y_n + (I - \Gamma A) u_{n-1},
\]

\[
(1.4b) \quad \Gamma = (A^* A + \alpha \Sigma^{-1})^{-1} A^*.
\]

Note that 3DVAR corresponds to an infinite dimensional AR(1) process. Our aim is to build on the framework and methodology of [3].

1.2. Key Assumptions and Prior Results. In [3], the following assumptions were made to obtain results. We retain these assumptions for our results.

Assumption 1.

1. \( C_0 = \frac{\alpha}{2} \Sigma \) with \( \text{Ran}(\Sigma^{\frac{1}{2}}) \subset \text{Dom}(A) \), \( \alpha > 0 \), and \( \Sigma \) a self-adjoint positive definite trace class operator with \( \Sigma^{-1} \) densely defined.
2. \( \Sigma \) induces a Hilbert scale and there exists constants \( C > 1, \nu > 0 \) such that \( A \) induces an equivalent norm:

\[
(1.5) \quad C^{-1} \| x \|_\nu \leq \| Ax \| \leq C \| x \|_\nu, \quad \| \bullet \|_\nu = \| \Sigma^{\frac{1}{2}} \bullet \|.
\]

3. The initial error is sufficiently smooth,

\[
(1.6) \quad m_0 - u^\dagger \in \text{Dom}(\Sigma^{-\frac{s}{2}}), \quad 0 \leq s \leq a + 2,
\]

where we replace \( m_0 \) with \( u_0 \) in the case of 3DVAR in the above expression.

Under this first set of assumptions, Iglesias et al. established
Theorem 1.1 (Theorem 4.1 of [3]). The Kalman filter admits the mean square error bound
\[ E[\| m_n - u^\dagger \|^2] \lesssim \left( \frac{n}{\alpha} \right)^{-\frac{\pi}{2\alpha}} + \frac{\gamma^2}{\alpha} \text{Tr} \Sigma \]

and

Theorem 1.2 (Theorem 5.1 of [3]). 3DVAR admits the mean square error bound
\[ E[\| u_n - u^\dagger \|^2] \lesssim \left( \frac{n}{\alpha} \right)^{-\frac{\pi}{2\alpha}} + \frac{\gamma^2}{\alpha} \text{Tr} \log n. \]

At fixed values of \( \alpha \), Theorems 1.1 and 1.2 preclude convergence, and, in the case of 3DVAR, the error may even grow. However, there are two free parameters: the number of iterations \( n \) and the regularization parameter \( \alpha \). Indeed, \( \alpha \) within a Bayesian framework, \( \alpha \) can be interpreted as the strength of a prior relative to a likelihood. By tuning these parameters one can either:

- Select \( n \) so as to minimize the error at a given \( \alpha \);
- Select \( \alpha \) so as the minimize the error for a given \( n \).

This is accomplished in the usual way by minimizing the upper bounds on the error over \( \alpha \) and \( n \). It suggests that the error can be made arbitrarily small. However, in both expressions, there is an unknown constant. If the error at the given, optimal choice of \( n \) and \( \alpha \) is inadequate, one must choose a different value of \( \alpha \) and rerun the algorithm with this new choice. A benefit of the present work is that, by using iterate averaging, the error of 3DVAR can always be reduced by computing additional iterates, without adjusting \( \alpha \) and discarding previously computed iterations.

Somewhat stronger results were obtained in [3] under a simultaneous diagonalization assumption.

Assumption 2.

- (1) \( \Sigma \) and \( A^*A \) simultaneously diagonalize with respective eigenvalues \( \sigma_i \) and \( a_i^2 \), and these eigenvalues satisfy
  \[ \sigma_i = i^{1-2\epsilon}, \quad a_i = i^{-p}, \quad \epsilon > 0, \quad p > 0. \]

- (2) \( m_0 = 0 \) (or \( u_0 \) in 3DVAR) and \( u^\dagger \) satisfies, for \( \beta \in (0,1+2\epsilon+2p] \),
  \[ \sum_{k=1}^{\infty} k^{2\beta} |u_k^\dagger|^2 < \infty. \]

Under the diagonalization assumption, one has

Theorem 1.3 (Theorem 4.2 of [3]). Under Assumption 2 for the Kalman filter,
\[ E[\| m_n - u^\dagger \|^2] \lesssim \left( \frac{n}{\alpha} \right)^{-\frac{2\beta}{1+2\epsilon+2p}} + \gamma^2 n^{-\frac{2\epsilon}{1+2\epsilon+2p}} \alpha^{-\frac{1+2p}{1+2\epsilon+2p}} \]
and

Theorem 1.4 (Theorem 5.2 of [3]). Under Assumption 2 for 3DVAR,
\[ E[\| u_n - u^\dagger \|^2] \lesssim \left( \frac{n}{\alpha} \right)^{-\frac{2\beta}{1+2\epsilon+2p}} + C\gamma^2 \alpha^{-\frac{1+2p}{1+2\epsilon+2p}}. \]

Now, the Kalman filter will converge at any choice of parameter, while 3DVAR has at worst a bounded error. Again, \( \alpha \) can be tuned so as to obtain the minimax convergence rate.
1.3. Main Results. The main results of this paper are contained in the following theorems.

First, we have the elementary result that 3DVAR, without averaging, cannot converge at fixed parameter choices:

**Theorem 1.5.** Under Assumption 1 in dimension one, if \( u_n \) generated by 3DVAR, then

\[
E[\|u_n - u^\dagger\|^2] \geq \gamma^2 K^2.
\]

As the method cannot converge in dimension one, it has no hope of converging in higher dimensions.

By time averaging,

\[
\bar{u}_n = \frac{1}{n} \sum_{k=1}^{n} u_k = \frac{1}{n} u_n + \frac{n-1}{n} \bar{u}_{n-1},
\]

under some additional assumptions, we can obtain convergence independently of the choice of \( \alpha \):

**Theorem 1.6.** Under Assumption 4, fix \( t \in [0, \nu] \) and \( \tau \in [0, 1] \), and, having set these indices, assume that \( \Sigma_{t+1-\tau \nu} \) is trace class. Then

\[
E[\|\bar{u}_n - u^\dagger\|^2] \lesssim \left( \frac{n}{\alpha} \right)^{-\frac{s}{1+\nu}} \|z_0\|^2 + \frac{\gamma^2}{\alpha} \text{Tr}(\Sigma_{t+1-\tau \nu}) \left( \frac{n}{\alpha} \right)^{-\tau \nu}
\]

where \( z_0 \) is the solution to

\[
\Sigma^{-\frac{1}{2}} (u_0 - u^\dagger) = (B^* B)^{\frac{1}{1+\nu}} z_0.
\]

Consequently, we will have unconditional mean squared convergence of the iterate averaged value, \( \bar{u}_n \), provided:

- We study the problem in a sufficiently weak weighted space \( (t > 0) \) and/or have sufficiently smooth data \( (s > 0) \);
- \( \Sigma \) has a sufficiently well behaved spectrum allowing \( \tau \nu > 0 \). Note that taking \( \tau \nu = t/(1 + \nu) \) will not require additional assumptions on \( \Sigma \), but will require \( t > 0 \) for convergence.

We emphasize that iterate averaging is a post-processing step, requiring no modification of the underlying 3DVAR iteration.

Under a modified version of Assumption 2,

**Assumption 2.**

1. \( \Sigma \) and \( A^* A \) simultaneously diagonalize with respective eigenvalues \( \sigma_i \) and \( a_i^2 \), and these eigenvalues satisfy

\[
\sigma_i \asymp i^{-1-2\epsilon}, \quad a_i \asymp i^{-p}, \quad \epsilon > 0, \quad p > 0.
\]

2. Fixing a \( \|\cdot\|_t \)-norm in which to study convergence, assume the data satisfies

\[
\sum_{k=1}^{\infty} k^{2\beta} |u_{0,k} - u_k^\dagger|^2 < \infty.
\]
**Theorem 1.7.** Under Assumption 2 and having fixed $t \geq 0$, assume $\tau_b, \tau_v \in [0,1]$ satisfy

\[ \tau_b \leq \frac{t(1 + 2\epsilon) + 2\beta}{2(1 + 2\epsilon + 2p)} \equiv \bar{\tau}_b \]

\[ \tau_v < \frac{t(1 + 2\epsilon) + 2\epsilon}{1 + 2\epsilon + 2p} \equiv \bar{\tau}_v \]

then

\[ \mathbb{E}[\|\bar{u}_n - u^\dagger\|_2^2] \lesssim \left(\frac{n}{\alpha}\right)^{-2\tau_b} + \frac{\gamma^2}{\alpha} \left(\frac{n}{\alpha}\right)^{-\tau_v} \]

In contrast to iterate averaged 3DVAR, there is no gain to iterate averaging for Kalman:

**Theorem 1.8.** For the scalar Kalman filter, take $C_0 = \frac{\gamma^2}{\alpha} \sigma > 0$. Then the bias and variance of the iterate-averaged mean, $\bar{m}_n$ satisfy the inequalities

\[ |\mathbb{E}[\bar{m}_n] - u^\dagger| \geq |\mathbb{E}[m_n] - u^\dagger|, \]

\[ \text{Var}(\bar{m}_n) \geq \text{Var}(m_n). \]

1.4. **Outline.** The structure of this paper is as follows. In Section 2 we review certain background results needed for our main results. Section 3 examines the scalar case, and it includes proofs of Theorems 1.5 and 1.8. We prove Theorems 1.6 and 1.7 in Section 4. Numerical examples are given in Section 5. We conclude with a brief discussion in Section 6.

**Acknowledgements:** The authors thank A.M. Stuart for suggesting an investigation of this problem. This work was supported by US National Science Foundation Grant DMS-1818716. The content of this work originally appeared in [4] as a part of F.G. Jones’s PhD dissertation. Work reported here was run on hardware supported by Drexel’s University Research Computing Facility.

2. **Preliminary Results**

In this section, we establish some identities and estimates that will be crucial to proving our main results. Much of our analysis relies on spectral calculus involving the following rational functions:

\[ r_{n,\alpha}(\lambda) = \left(\frac{\alpha}{\alpha + \lambda}\right)^n, \]

\[ q_{n,\alpha}(\lambda) = \frac{1}{\lambda} \left\{ 1 - \left(\frac{\alpha}{\alpha + \lambda}\right)^n \right\} = \lambda^{-1}(1 - r_{n,\alpha}(\lambda)). \]

Throughout, $\alpha > 0$ and $n \in \mathbb{N}$. These are related by the identity

\[ \sum_{k=1}^{m} r_{k,\alpha}(\lambda) = \alpha q_{m,\alpha}(\lambda). \]

The following estimates are established in [2] and [5], particularly Section 2.2 of the latter reference:
Lemma 2.1. For $\lambda \in [0, \Lambda]$ and $n \in \mathbb{N}$,

\begin{equation}
0 < r_{n, \alpha}(\lambda) \leq \frac{\alpha}{\alpha + n\lambda} \leq 1,
\end{equation}

\begin{equation}
\lambda^p r_{n, \alpha}(\lambda) \leq \left\{ \begin{array}{ll}
\left(\frac{\alpha + n\lambda}{n}\right)^p, & p \in [0, n], \\
\Lambda^p, & p > n.
\end{array} \right.
\end{equation}

Lemma 2.2. For $\lambda \in [0, \Lambda]$, $n \in \mathbb{N}$,

\begin{equation}
\lambda^p q_{n, \alpha}(\lambda) \leq \left\{ \begin{array}{ll}
\left(\frac{n}{\alpha + n\lambda}\right)^{1-p}, & p \in [0, 1], \\
\Lambda^{p-1}, & p > 1,
\end{array} \right.
\end{equation}

\begin{equation}
\lambda^p q_{n, \alpha}(\lambda) \leq \lambda^{p-1}.
\end{equation}

Next, we recall the following result on Hilbert scales,

Proposition 2.3. With $B = A\Sigma_{\hat{\nu}}^\hat{\nu}$, there exists a constant $D > 1$, such that for $|\theta| \leq 1$,

\begin{equation}
D^{-1}||x||_{\theta(1+\nu)} \leq ||(B^*B)^{\frac{1}{2}}x|| \leq D||x||_{\theta(1+\nu)}
\end{equation}

and $\text{Ran}((B^*B)^{\frac{1}{2}}) = \text{Dom}(\Sigma_0^{-\frac{1}{2}})$.

This result, based on a duality argument, is proven in Lemma 4.1 of [3]. See, also, Section 8.4 of [1], particularly Corollary 8.22.

We also have a few useful identities for the filters which we state without proof.

Lemma 2.4. For the Kalman filter, the mean and covariance operators and the Kalman gains satisfy the identities

\begin{align*}
m_n &= (\gamma^2 n^{-1}C_0^{-1} + A^*A)^{-1}(A^*\bar{y}_n + \gamma^2 n^{-1}C_0^{-1}m_0) \\
C_n^{-1} &= C_{n-1}^{-1} + \gamma^{-2}A^*A = C_{n-1}^{-1} + \gamma^{-2}nA^*A \\
K_n &= (\gamma^2 C_{n-1}^{-1} + A^*A)^{-1}A^* = (\gamma^2 C_{n-1}^{-1} + nA^*A)^{-1}A^* = \gamma^{-2}C_nA^*.
\end{align*}

Lemma 2.5. For 3DVAR,

\begin{equation}
\bar{u}_n = \sum_{k=0}^{n-1} \frac{n-k}{n} (I - KA)^k \bar{y}_{n-k} + \sum_{k=0}^{n-1} \frac{1}{n} (I - KA)^k (I - KA)u_0.
\end{equation}

Corollary 2.6. Letting $v_n = u_n - u^\dagger$, $\bar{v}_n = \frac{1}{n} \sum_{k=1}^{n} v_k$,

\begin{equation}
\bar{v}_n = \sum_{k=0}^{n-1} \frac{n-k}{n} (I - KA)^k \bar{y}_{n-k} + \sum_{k=0}^{n-1} \frac{1}{n} (I - KA)^k (I - KA)v_0.
\end{equation}

Remark 2.7. As this is a linear problem, it will be sufficient to study the behavior of $\bar{v}_n$ to infer convergence of $\bar{u}_n$ to $u^\dagger$.

For the analysis of 3DVAR, the essential decomposition is into a bias and a variance term. From Corollary 2.6 these are

\begin{equation}
\bar{f}_n^{\text{bias}} = \sum_{k=0}^{n-1} \frac{1}{n} (I - KA)^k (I - KA)v_0,
\end{equation}

\begin{equation}
\bar{f}_n^{\text{var}} = \sum_{k=0}^{n-1} \frac{n-k}{n} (I - KA)^k \bar{y}_{n-k}.
\end{equation}
The bias and variance can be expressed in the more useful forms using \( q_{n,\alpha} \):

**Lemma 2.8.**

\[
\begin{align*}
\bar{I}^{\text{bias}}_n &= \frac{\alpha}{n} \sum q_{n,\alpha} (B^*B) \Sigma^{1/2} v_0, \\
\bar{I}^{\text{var}}_n &= \frac{1}{n} \sum_{j=1}^n \Sigma^{1/2} q_{n-j+1,\alpha} (B^*B) B^* \eta_j.
\end{align*}
\]

**Proof.** First, observe that

\[ I - KA = \Sigma^{1/2} (\alpha I + B^*B) \Sigma^{-1/2}. \]

Using this in (2.8) together with spectral calculus applied to positive self-adjoint compact operator \( B^*B \), along with (2.3),

\[
\bar{I}^{\text{bias}}_n = \frac{1}{n} \sum_{k=0}^{n-1} \Sigma^{1/2} \alpha^k (\alpha I + B^*B)^{-k+1} \Sigma_0^{-1/2} v_0 = \frac{\alpha}{n} \sum q_{n,\alpha} (B^*B) \Sigma^{-1/2} v_0.
\]

Applying the same computations to (2.9), we have,

\[
\bar{I}^{\text{var}}_n = \frac{1}{n} \sum_{j=1}^n \{ \sum_{k=0}^{n-j-1} \Sigma^{1/2} \alpha^k (B^*B) B^* \eta_{n-k} \} \eta_j = \frac{1}{n} \sum_{j=1}^n \Sigma^{1/2} q_{n-j+1,\alpha} (B^*B) B^* \eta_j.
\]

\[ \square \]

3. **Analysis of the Scalar Problem**

Before proceeding to the general, infinite-dimensional case, it is instructive to consider the scalar problem, where \( X = Y = \mathbb{R} \) and \( A, \Sigma, \) and \( K \) are now scalars.

This setting will also allow us to establish the limitations of both 3DVAR and the Kalman filter alluded to in the introduction. The scalar problem also serves as a building block in the case that it is possible to simultaneously diagonalize operators \( A \) and \( \Sigma \) in the general case.

3.1. **3DVAR.** Operator \( \Sigma \) is now just the scalar constant, the regularization remains \( \alpha > 0 \), and the 3DVAR gain \( K \) defined in (1.4) is now the scalar.

First, we have prove Theorem 1.5, which asserts that the 3DVAR iteration cannot converge in mean square:

**Proof.** Since \( y_n \sim \mathcal{N}(Au^\dagger, \gamma^2) \), we write \( y_n = Au^\dagger + \eta_n \) for \( \eta_n \sim \mathcal{N}(0, \gamma^2) \). By (1.4),

\[
\begin{align*}
\eta_n &= K\eta_n + KA u_n - (K - K) u_{n-1} - u^\dagger \\
&= K\eta_n + (K - K) (u_{n-1} - u^\dagger).
\end{align*}
\]

Consequently,

\[
\begin{align*}
\mathbb{E}[|u_n - u^\dagger|^2] &= \mathbb{E}[|K\eta_n|^2] + \mathbb{E}[|(K - K)(u_{n-1} - u^\dagger)|^2] \\
&\geq \mathbb{E}[|K\eta_n|^2] = \kappa^2 \gamma^2.
\end{align*}
\]

\[ \square \]
Next, studying the bias and variance of the time averaged problem, given by (2.8) and (2.9), we prove

**Theorem 3.1.** For scalar time averaged 3DVAR, for \( \tau_b, \tau_v \in [0, 1] \)

\[
E[|\bar{u}_n - u|^2] \leq (A^2 \Sigma)^{-2\tau_v} |v_0|^2 \left( \frac{n}{\alpha} \right)^{-2\tau_b} + \frac{\Sigma \gamma^2}{\alpha} (A^2 \Sigma)^{-\tau_v} \left( \frac{n}{\alpha} \right)^{-\tau_v}.
\]

Thus, we have unconditional convergence for any choice for \( \alpha > 0 \), something that we do not have for 3DVAR without any iterate averaging. Indeed, Theorem 1.5 tells us that for any fixed set of parameters, we would always have a finite error, regardless of \( n \). The rate of convergence is greatest when \( \tau_b \geq 1/2 \) and \( \tau_v = 1 \).

To obtain the result, we make use of the bias variance decomposition and expressions (2.10) and (2.11). In the scalar case, \( B^*B = B^2 = \Sigma A^2 \), so that

\[
(3.1) \quad |\bar{\imath}_{\text{bias}}|^2 = \left( \frac{\alpha}{n} \right)^2 \sum_j q_{j,\alpha} (\Sigma A^2)^2 |v_0|^2.
\]

Applying (2.7) to this expression, we immediately obtain

**Proposition 3.2.** For \( 0 \leq \tau_b \leq 1 \),

\[
(3.2) \quad |\bar{\imath}_{\text{bias}}|^2 \leq (A^2 \Sigma)^{-2\tau_v} |v_0|^2 \left( \frac{n}{\alpha} \right)^{-2\tau_b}.
\]

For the variance, we have the result

**Proposition 3.3.** Let \( \tau_v \in [0, 1] \),

\[
(3.3) \quad E[|\bar{\imath}_{\text{var}}|^2] \leq \frac{\Sigma \gamma^2}{\alpha} (A^2 \Sigma)^{-\tau_v} \left( \frac{n}{\alpha} \right)^{-\tau_v}.
\]

**Proof.** For the scalar case of (2.11),

\[
E[|\bar{\imath}_{\text{var}}|^2] = \frac{\gamma^2 (A\Sigma)^2}{n^2} \sum_{j=1}^n q_{j,\alpha} (\Sigma A^2)^2.
\]

Then, using Lemma 2.2,

\[
E[|\bar{\imath}_{\text{var}}|^2] = \frac{\gamma^2 (A\Sigma)^2}{n^2} \sum_{j=1}^n q_{j,\alpha} (A^2 \Sigma)^2
\]

\[
= \frac{\gamma^2 \Sigma}{n^2} (A^2 \Sigma)^{1-(1+\tau_v)} \sum_{j=1}^n \left( \frac{j}{\alpha} \right)^{2(1-\frac{1+\tau_v}{2})}
\]

\[
\leq \frac{\Sigma \gamma^2}{n^2} (A^2 \Sigma)^{-\tau_v} \sum_{j=1}^n \left( \frac{j}{\alpha} \right)^{2(1-\frac{1+\tau_v}{2})}
\]

\[
\leq \frac{\Sigma \gamma^2}{n^2} (A^2 \Sigma)^{-\tau_v} \left( \frac{n}{\alpha} \right)^{1-\tau_v} = \frac{\Sigma \gamma^2}{\alpha} (A^2 \Sigma)^{-\tau_v} \left( \frac{n}{\alpha} \right)^{-\tau_v}.
\]

\[\square\]

**Proof of Theorem 3.1.** The result then follows immediately by combining the two preceding propositions.

\[\square\]
3.2. **Kalman Filter.** Next, we provide a proof of Theorem 1.8, showing there is no improvement in mean squared convergence of Kalman under iterate averaging.

**Proof.** Using Lemma 2.4, for the $k$-the estimate of the mean,

$$m_k = \left( \frac{\alpha}{\Sigma_k} + a^2 \right)^{-1} \left( A\bar{y}_k + \frac{\alpha}{\Sigma_k} m_0 \right)$$

$$= \left( \frac{\alpha}{\Sigma_k} + A^2 \right)^{-1} \left( A^2 u^\dagger + A\bar{y}_k + \frac{\alpha}{\Sigma_k} m_0 \right)$$

$$= \left( 1 + \frac{\alpha}{A^2 \Sigma_k} \right)^{-1} u^\dagger + \left( 1 + \frac{A^2 \Sigma_k}{\alpha} \right)^{-1} m_0 + \left( A + \frac{\alpha}{A \Sigma_k} \right)^{-1} \eta_k.$$

and without averaging,

$$E[m_n] - u^\dagger = \left( 1 + \frac{A^2 \Sigma n}{\alpha} \right)^{-1} (m_0 - u^\dagger),$$

$$\text{Var}(m_n) = \left( A + \frac{\alpha}{A \Sigma n} \right)^{-2} \frac{\gamma^2}{n}.$$

Then, with averaging, for the bias,

$$E[\bar{m}_n] - u^\dagger = \frac{1}{n} \sum_{k=1}^n \left( 1 + \frac{A^2 \Sigma_k}{\alpha} \right)^{-1} (m_0 - u^\dagger),$$

and

$$|E[\bar{m}_n] - u^\dagger|^2 = \left| \frac{1}{n} \sum_{k=1}^n \left( 1 + \frac{A^2 \Sigma_k}{\alpha} \right)^{-1} \right|^2 |m_0 - u^\dagger|^2$$

$$\geq \left| \frac{1}{n} \sum_{k=1}^n \left( 1 + \frac{A^2 \Sigma_n}{\alpha} \right)^{-1} \right|^2 |m_0 - u^\dagger|^2 = |E[m_n] - u^\dagger|^2.$$

For the variance, first note

$$\bar{m}_n - E[\bar{m}_n] = \frac{1}{n} \sum_{k=1}^n \left( A + \frac{\alpha}{A \Sigma_k} \right)^{-1} \eta_k = \frac{1}{n} \sum_{k=1}^n \left( A + \frac{\alpha}{A \Sigma_k} \right)^{-1} \left\{ \sum_{j=1}^k \eta_j \right\}$$

$$= \frac{1}{n} \sum_{j=1}^n \eta_j \left\{ \sum_{k=j}^n \left( A + \frac{\alpha}{A \Sigma_k} \right)^{-1} \right\}.$$ 

Then, by dropping all but the $k = n$-th term in the inner sum,

$$\text{Var}(\bar{m}_n) = \frac{1}{n^2} \sum_{j=1}^n \gamma^2 \left\{ \sum_{k=j}^n \left( A + \frac{\alpha}{A \Sigma_k} \right)^{-1} \right\}^2$$

$$\geq \frac{1}{n^2} \sum_{j=1}^n \gamma^2 \left( A + \frac{\alpha}{A \Sigma n} \right)^{-2}$$

$$= \text{Var}(m_n) \quad \square$$

4. **Analysis of the Infinite Dimensional Problem**

We return to the bias and variance of 3DVAR in the general, potentially infinite dimensional, setting and obtain estimates on the terms.
4.1. General Case. Here, we prove Theorem 1.6.

**Proposition 4.1.** Under Assumption 1, with \( t \in [0, \nu] \),

\[
\| \bar{I}_{\text{bias}} \|_2^2 \lesssim \left( \frac{n}{\alpha} \right)^{\frac{s+t}{\frac{s}{\alpha} + t + \frac{1}{\nu}}} \| z_0 \|^2
\]

where \( z_0 \) solves (1.13).

**Remark 4.2.** The fastest possible decay available for the squared bias in Proposition 4.1 is \( \mathcal{O}(n^{-2}) \) when \( s = \nu + 2 \) and \( t = \nu \).

**Proof.** We make use of bias term from Lemma 2.8, allowing us to write

\[
\| \bar{I}_{\text{bias}} \|_2^2 = \left\| \frac{\alpha}{n} \Sigma^{t+1} q_{n,\alpha}(B^* B) \Sigma^{-\frac{1}{2}} v_0 \right\|^2.
\]

Next, we make use of (1.5) and argue as in the Appendix of [3], applying Proposition 2.3. Since, by assumption, \( v_0 \in \text{Dom}(\Sigma^{\frac{s}{2}}) \), \( \Sigma^{-\frac{1}{2}} v_0 \in \text{Dom}(\Sigma^{-\frac{s}{2}}) \). By Proposition 2.3, letting \( \theta = (s - 1)/(1 + \nu) \), \( \Sigma^{-\frac{1}{2}} v_0 \in \text{Ran}((B^* B)^{\frac{s-1}{2(1+\nu)}}) \) allows us to conclude the existence of \( z_0 \). Therefore,

\[
\| \bar{I}_{\text{bias}} \|_2^2 = \left\| \frac{\alpha}{n} \Sigma^{t+1} q_{n,\alpha}(B^* B)(B^* B)^{\frac{s-1}{2(1+\nu)}} z_0 \right\|^2.
\]

Next, using Proposition 2.3 again, now with \( \theta = (1 + t)/(1 + \nu) \),

\[
\| \bar{I}_{\text{bias}} \|_2^2 \lesssim \left( \frac{n}{\alpha} \right)^{\frac{s+t}{\frac{s}{\alpha} + t + \frac{1}{\nu}}} \left\| \Sigma^{t+1} q_{n,\alpha}(B^* B)(B^* B)^{\frac{s-1}{2(1+\nu)}} z_0 \right\|^2.
\]

The last inequality holds since, \( s \leq \nu + 2 \) and \( t \leq \nu \), so that \( 0 \leq s + t \leq s + \nu \leq 2\nu + 2 \) allowing for the application of Lemma 2.2. \( \square \)

**Proposition 4.3.** Under Assumption 1, for \( t \geq 0, \tau_\nu \in [0, 1] \), and for this choice of \( \tau_\nu \) and \( t \), assume \( \Sigma^{(1+t) - \tau_\nu(1+\nu)} \) is trace class. Then

\[
\mathbb{E}[\| \bar{I}_{\text{var}} \|_2^2] \lesssim \left( \frac{n}{\alpha} \right)^{\frac{s}{\alpha} + \frac{1}{\nu}} \text{Tr}((\Sigma^{t+1} - \tau_\nu(1+\nu))) \left( \frac{n}{\alpha} \right)^{-\tau_\nu}.
\]

**Remark 4.4.** The fastest possible decay in the variance will be \( \mathcal{O}(n^{-1}) \) when \( \tau_\nu = 1 \) and \( t \) is sufficiently large such that \( \Sigma^{t-\nu} \) is trace class. However, the bias term requires \( t \leq \nu \). This requires the identity operator to be trace class which will not hold in infinite dimensions.
Proof. We begin with equation (2.11) and using that for any bounded operator $T$ and positive self adjoint trace class operator $C$,

\[ |\text{Tr}(CT)| \leq \|T\|\|\text{Tr}C\|, \]

\[
\mathbb{E}[\|\tilde{I}_{n}\|_{2}^{2}] = \frac{1}{n^{2}} \sum_{j=1}^{n} \mathbb{E}[\|\Sigma_{\frac{j-1}{2}} q_{n-j+1,\alpha}(B^{*}B)B^{*}e_{j}\|_{2}^{2}]
\]

\[
= \frac{\gamma^{2}}{n^{2}} \sum_{j=1}^{n} \text{Tr}\left(\Sigma_{\frac{j-1}{2}} q_{j,\alpha}(B^{*}B)(B^{*}B)q_{j,\alpha}(B^{*}B)\Sigma_{\frac{j-1}{2}}\right)
\]

\[
= \frac{\gamma^{2}}{n^{2}} \sum_{j=1}^{n} \text{Tr}\left(\Sigma_{\frac{j-1}{2}} (B^{*}B)q_{j,\alpha}(B^{*}B)(B^{*}B)\right)^{2}
\]

\[
\leq \frac{\gamma^{2}}{n^{2}} \sum_{j=1}^{n} \|\Sigma_{\frac{j-1}{2}} (B^{*}B)q_{j,\alpha}(B^{*}B)(B^{*}B)\|^{2} \text{Tr}(\Sigma_{\frac{j-1}{2}} (B^{*}B)q_{j,\alpha}(B^{*}B)(B^{*}B))
\]

Using Proposition 2.3 with $\theta = \tau_{\nu}$ and Lemma 2.2,

\[
\|\Sigma_{\frac{j-1}{2}} (B^{*}B)q_{j,\alpha}(B^{*}B)(B^{*}B)\| \lesssim \|B^{*}B\|^{\frac{1}{2}} q_{j,\alpha}(B^{*}B)\|B^{*}B\|,
\]

\[
\leq \sup_{\lambda \in [0,\|B^{*}B\|]} \lambda^{\frac{1+2\tau_{\nu}}{2}} q_{j,\alpha}(\lambda) \lesssim \left(\frac{j}{\alpha}\right)^{1-\frac{1+2\tau_{\nu}}{2}}
\]

Therefore,

\[
\mathbb{E}[\|\tilde{I}_{n}\|_{2}^{2}] \leq \frac{\gamma^{2}}{n^{2}} \text{Tr}(\Sigma_{\frac{j-1}{2}} (B^{*}B)q_{j,\alpha}(B^{*}B)(B^{*}B)) \sum_{j=1}^{n} \left(\frac{j}{\alpha}\right)^{1-\tau_{\nu}} \lesssim \frac{\gamma^{2}}{\alpha} \text{Tr}(\Sigma_{\frac{j-1}{2}} (B^{*}B)q_{j,\alpha}(B^{*}B)(B^{*}B)) \left(\frac{n}{\alpha}\right)^{-\tau_{\nu}}
\]

\[\square\]

Proof of Theorem 1.6. The theorem immediately follows from the two preceding propositions. \[\square\]

4.2. Simultaneous Diagonalization. A sharper result is available under the simultaneous diagonalization Assumption 2. Indeed, let us assume that $\Sigma$ and $A^{*}A$ simultaneously diagonalize against the orthonormal set $\{\varphi_{k}\}$, with eigenvalues

\[\Sigma \varphi_{k} = \sigma_{k} \varphi_{k}, \quad A^{*}A \varphi_{k} = a_{k}^{2} \varphi_{k}.\]

The assumptions of (1.14) and (1.15) are equivalent to those of (1.5) and (1.6) under the identifications:

\[\nu(1+2\epsilon) = 2p, \quad s(1+2\epsilon) = 2\beta.\]

Also, observe that, letting

\[\omega = \frac{1+2\epsilon}{1+2\epsilon + 2p},\]

we have the relationship

\[\sigma_{k} \asymp (\sigma_{k} a_{k}^{2})^{\omega}\]

Proposition 4.5. Under Assumption 2, let $\tau_{n} \in [0,1]$ satisfy condition (1.16a),

\[\|\tilde{I}_{n}^{\text{bias}}\|_{2}^{2} \lesssim \left(\frac{n}{\alpha}\right)^{-2\tau_{n}}.\]
Proof. We start with equation (2.10) and then use (4.5) and Lemma 2.2:

\[
\|\hat{f}_n\|_{t}^{2} = \sum_{k=1}^{\infty} \left( \frac{\alpha}{n} \sum_{n,\alpha} (B^*B) \Sigma^{-\frac{1}{2}} v_0, \varphi_k \right)^2 = \sum_{k=1}^{\infty} \left( \frac{\alpha}{n} \sigma_k^2 q_{n,\alpha} (\sigma_k a_k^2) \right)^2 |v_{0,k}|^2 = \left( \frac{\alpha}{n} \right)^2 \sum_{k=1}^{\infty} \sigma_k^2 q_{n,\alpha} (\sigma_k a_k^2)^2 |v_{0,k}|^2 
\]

\[
\geq \left( \frac{\alpha}{n} \right)^2 \sum_{k=1}^{\infty} (\sigma_k a_k^2)^{2-2\tau_0} \left( (\sigma_k a_k^2) \right)^2 |v_{0,k}|^2 
\]

Using (1.16a),

\[
\sum_{k=1}^{\infty} (\sigma_k a_k^2)^{2-2\tau_0} |v_{0,k}|^2 \leq \sum_{k=1}^{\infty} k^{-(1+2\tau_0)} |v_{0,k}|^2 
\]

\[
\leq \sum_{k=1}^{\infty} k^{-(1+2\tau_0)-2\beta} k^{2\beta} |v_{0,k}|^2 
\]

\[
\leq \sum_{k=1}^{\infty} k^{2\beta} |v_{0,k}|^2 < \infty 
\]

we have the result.

\[\square\]

Remark 4.6. Comparing this to the general case, we again see that if the data is sufficiently smooth and/or we study the problem in a sufficiently smooth space (\(\beta\) and/or \(t\) large), we can again obtain \(O(n^{-2})\) convergence of the squared bias.

Proposition 4.7. Under Assumption 2 \(^\star\) having fixed \(t\), for \(\tau_v \in [0,1]\) satisfying (1.16b),

\[
E \left[ \|\hat{f}_{n,\alpha} \|_{t}^{2} \right] \leq \frac{\gamma^2}{\alpha} \left( \frac{n}{\alpha} \right)^{-\tau_v}
\]

Proof. Using (2.11), we begin by writing

\[
E \left[ \|\hat{f}_{n,\alpha} \|_{t}^{2} \right] = \frac{1}{n^2} \sum_{j=1}^{n} E \left[ \Sigma_{t+1}^{j+1} q_{n-j+1,\alpha} (B^*B) \Sigma_{t+1}^{j+1} \right] 
\]

\[
= \frac{\gamma^2}{n^2} \sum_{j=1}^{n} \text{Tr} \left( \Sigma_{t+1}^{j+1} q_{n-j+1,\alpha} (B^*B) q_{n-j+1,\alpha} (B^*B) \Sigma_{t+1}^{j+1} \right) 
\]

\[
= \frac{\gamma^2}{n^2} \sum_{j=1}^{n} \text{Tr} \left( \Sigma_{t+1}^{j+1} (B^*B) q_{j,\alpha} (B^*B)^2 \right) 
\]

Using (2.2) on each term in the sum,

\[
\text{Tr} \left( \Sigma_{t+1}^{j+1} (B^*B) q_{j,\alpha} (B^*B)^2 \right) = \sum_{k=1}^{\infty} \sigma_k^2 q_{j,\alpha} (\sigma_k a_k^2)^2 
\]
Then, using (4.5) and Lemma 2.2
\[ \sigma_{t+2}^2 a_k^2 q_{j,\alpha}(\sigma_k a_k^2)^2 \simeq (\sigma_k a_k^2)^2 q_{j,\alpha}(\sigma_k a_k^2)^2 \]
\[ \simeq (\sigma_k a_k^2)^{\omega(t+1)}(\sigma_k a_k^2)^2 q_{j,\alpha}(\sigma_k a_k^2)^2 \]
\[ \simeq (\sigma_k a_k^2)^{\omega(t+1)-\tau_v}((\sigma_k a_k^2)^{1+\tau_v}/2 q_{j,\alpha}(\sigma_k a_k^2)^2 \]
\[ \leq (\sigma_k a_k^2)^{\omega(t+1)-\tau_v} \left( \frac{j}{\alpha} \right)^{1-\tau_v} \]

Under assumption 1.16
\[ \sum_{k=1}^{\infty} (\sigma_k a_k^2)^{\omega(t+1)-\tau_v} \leq \sum_{k=1}^{\infty} k^{-[(1+2\epsilon)(t+1)-\tau_v(1+2\epsilon+2p)]} < \infty \]

Consequently,
\[ \text{Tr} \left( \Sigma^{t+1}(B^*B)q_{j,\alpha}(B^*B)^2 \right) \lesssim \left( \frac{j}{\alpha} \right)^{1-\tau_v} , \]

and
\[ \frac{\gamma^2}{n^2} \sum_{j=1}^{n} \text{Tr} \left( \Sigma^{t+1}(B^*B)q_{j,\alpha}(B^*B)^2 \right) \lesssim \frac{\gamma^2}{\alpha} \left( \frac{n}{\alpha} \right)^{-\tau_v} \]

Remark 4.8. In contrast to the non-diagonal case, if the problem is studied in a sufficiently weak sense (large enough $t$), one obtains $O(n^{-1})$ convergence of the variance.

Proof of Theorem 1.7. This result immediately follows from the previous two propositions.

5. Numerical Experiments

In this section we illustrate our results with some numerical experiments.

5.1. Scalar Examples. As a simple scalar example, let $A = 1$, $\gamma = 0.1$, and $u^\dagger = 0.5$. For 3DVAR, take $u_0 = 0$, $\Sigma = 1$, and $\alpha = 1$, while for Kalman, take $m_0 = 0$ and $C_0$. Running $10^2$ independent trials of each algorithm for $10^4$ iterations, we obtain the results in Figure 1. These demonstrated our predictions from Theorems 1.5, Theorem 3.1, and Theorem 1.8 that 3DVAR can only converge with time averaging, while Kalman will not be improved by time averaging. The confidence bounds are computed using $10^4$ bootstrap samples to produce 95% confidence intervals.

5.2. Simultaneous Diagonalization Example. Next, we consider the case of simultaneous diagonalization, working with functions in $L^2(0, 2\pi; \mathbb{R})$, and
\[ A = (I - \frac{d^2}{dx^2})^{-1} , \quad \Sigma = A^2 , \quad u^\dagger = 0 \]

The $A$ operator is equipped with periodic boundary conditions, allowing us to easily work in Fourier space. As the problem is linear, we can separately consider the bias and the variance. In all examples below we discretize on $N = 2^{12}$ modes, and run for $10^4$ iterations. This corresponds to $p = 2$ and $\epsilon = 1.5$ in Assumption 2.1.
Figure 1. Scalar results for 3DVAR and the Kalman filter. These results are consistent with Theorems 1.5, 1.8, and 3.1; 3DVAR will not converge without time averaging while Kalman will not improve from time averaging. Shaded regions reflect 95% confidence intervals at each $n$.

For the bias, we choose, before truncation, as the initial condition

$$u_0 = \sum_{k=1}^{\infty} k^{-\frac{3}{2} - \beta - \delta} \cos(kx),$$

with $\beta = 1$ and $\delta = 0.01$. Consequently, this function satisfies (1.15) from Assumption 2. The perturbation $\delta$ is introduced so that we can best see the sharpness of our rates. Running the truncated and discretized problem, we obtain the results shown in Figure 2 for the norms $t = 0, 0.5, 1, 2$. As the plots show, we are in good agreement with the maximal rate predicted by Theorem 1.7.

For the variance, we run $10^2$ independent trials of the problem, and then use bootstrapping to estimate 95% confidence intervals. The results, shown in Figure 3 again show good agreement with the maximal rate predicted by Theorem 1.7.

6. Discussion

In this work we have examined the impact of iterate averaging upon the Kalman filter and 3DVAR as tools for solving a statistical inverse problem. We have found that this modest post-processing step ensures that the simpler algorithm, 3DVAR, will converge, unconditionally with respect to $\alpha$, in mean square as the number of iterations $n \to \infty$. In contrast, there is no performance gain when this averaging is applied to the Kalman filter.

Our simulations suggest that our rates, at least in the diagonal case, may be sharp. For the diagonal case, we should expect to see something slower than the Monte Carlo rate of convergence, $O(n^{-1})$ unless working in a sufficiently weak norm (large $t$). In the general case, it would seem that for the infinite dimensional problem, we will never be able to achieve $O(n^{-1})$ convergence for the reasons outlined in Remark 4.4; the operator $\Sigma^{t-\nu}$ would need to be trace class, but $t \leq \nu$ for the bias to converge. The sharpness of the result in the non-diagonalizable case remains to be established.
Figure 2. Decay of the squared bias in our simultaneously diagonalized test problem for different $t$-norms. All are in good agreement with the rates predicted by Theorem 1.7. The constant $\bar{\tau}_v$ reflects the greatest possible decay rate from (1.16a).

On the other hand, in actual applications, the problem will always be finite dimensional in practice, making $O(n^{-1})$ achievable, as it was in the scalar case of Section 3. In a spectral Galerkin formulation, truncating to $N$ modes, and, $\sum_{-\nu}^{\nu}$, will always be finite, though the constant may be large. Thus, in practice, we should expect to see $O(n^{-1})$ convergence, for sufficiently large $n$ and a sufficiently severe dimensional truncation.

REFERENCES

[1] W. E. Heinz, M. Hanke, and A. Neubauer. Regularization of Inverse Problems. Kluwer Academic Publishers, 2000.
[2] M. A. Iglesias. A regularizing iterative ensemble Kalman method for PDE-constrained inverse problems. Inverse Problems, 32:025002, 2016.
[3] M. A. Iglesias, K. Lin, S. Lu, and A. M. Stuart. Filter based methods for statistical linear inverse problems. Communications in Mathematical Sciences, 15(7):1867–1896, 2017.
[4] F. G. E. Jones. High and Infinite-Dimensional Filtering Methods. PhD thesis, Drexel University, 2020.
[5] S. Pereverzev and S. Lu. Regularization Theory for Ill-Posed Problems. De Gruyter, 2013.

Email address: grs53@drexel.edu

Department of Mathematics, Drexel University, Philadelphia, PA, 19103, USA
Figure 3. Decay of the mean squared variance term in our simultaneously diagonalized test problem for different $t$-norms. All are in good agreement with the rates predicted by Theorem 1.7. Shaded regions reflect 95% confidence intervals at each $n$. The constant $\bar{\tau}_v$ reflects the greatest possible decay rate from (1.16b).