The Fundamental Subspaces of Ensemble Kalman Inversion

Elizabeth Qian, *Christopher Beattie*

Abstract

Ensemble Kalman Inversion (EKI) methods are a family of iterative methods for solving weighted least-squares problems of the form

$$\min_{\mathbf{v}\in\mathbb{R}^d}(\mathbf{y}-\mathbf{H}(\mathbf{v}))^\top\mathbf{\Sigma}^{-1}(\mathbf{y}-\mathbf{H}(\mathbf{v})) = \min_{\mathbf{v}\in\mathbb{R}^d}\|\mathbf{y}-\mathbf{H}(\mathbf{v})\|_{\mathbf{\Sigma}^{-1}}^2, \tag{1}$$

where $\mathbf{\Sigma}\in\mathbb{R}^{n\times n}$ is symmetric positive definite, and $\mathbf{H}:\mathbb{R}^d\to\mathbb{R}^n$. Such problems arise in many settings, including in *inverse problems* in which $\mathbf{v}\in\mathbb{R}^d$ represents an unknown parameter or state of a system of interest which must be inferred from observed data $\mathbf{y}\in\mathbb{R}^n$. Inverse problems arise in many disciplines across science, engineering, and medicine, including earth, atmospheric, and ocean modeling, medical imaging, robotics and autonomy, and more. In large-scale scientific and engineering applications, solving (1) using standard gradient-based optimization methods can be prohibitively expensive due to the high cost of evaluating derivatives or adjoints of the forward operator $\mathbf{H}$. In contrast, EKI methods can be implemented in an *adjoint-/derivative-free* way. This makes EKI an attractive alternative to gradient-based methods for solving (1) in large-scale inverse problems.

We introduce a basic version of EKI from [4] in Algorithm 1, noting that other EKI methods can be viewed as variations on this theme. In Algorithm 1, we use E and cov to denote the *empirical* (sample) mean and covariance operators, respectively: given $J\in\mathbb{N}$ samples $\{\mathbf{a}^{(j)}\}_{j=1}^J$ and $\{\mathbf{b}^{(j)}\}_{j=1}^J$, we define $\mathsf{E}[\mathbf{a}^{(1:J)}]=\frac{1}{J}\sum_{j=1}^J\mathbf{a}^{(j)}$, and

$$\mathsf{cov}[\mathbf{a}^{(1:J)},\mathbf{b}^{(1:J)}]=\frac{1}{J-1}\sum_{j=1}^J(\mathbf{a}^{(j)}-\mathsf{E}[\mathbf{a}^{(1:J)}])(\mathbf{b}^{(j)}-\mathsf{E}[\mathbf{b}^{(1:J)}])^\top,$$

and $\mathsf{cov}[\mathbf{a}^{(1:J)}]=\mathsf{cov}[\mathbf{a}^{(1:J)},\mathbf{a}^{(1:J)}]$. Algorithm 1 prescribes the evolution of an ensemble of J particles, $\{\mathbf{v}_i^{(1)},\ldots,\mathbf{v}_i^{(J)}\}$, initialized at $i=0$ in some way, e.g., by drawing from a suitable prior distribution, and subsequently updated for $i=1,2,3$, etc. We emphasize that Algorithm 1 does not require the evaluation of adjoints or derivatives of $\mathbf{H}$. Those familiar with ensemble Kalman *filtering* methods will recognize familiar elements in Algorithm 1. Indeed, one way to obtain Algorithm 1 is to apply the ensemble Kalman filter to a system whose dynamics are given by the identity map in the "forecast" step of the filter. The connection to the ensemble Kalman filter also motivates the perturbation of the observations by random noise in Step 7; these perturbations ensure unbiased estimates of the filtering statistics in the linear Gaussian setting.

There is a very rich literature developing both EKI methods and accompanying theory (see [3] for an extensive survey). Variants of the basic method include the incorporation of a Tikhonov regularization term into the least-squares objective function, the enforcement of constraints in the optimization, or hierarchical, multilevel, and parallel versions of the algorithm. Beyond the successful use of EKI for solving diverse inverse problems in the physical sciences, e.g., in geophysical and biological contexts, EKI has also been used as an optimizer for training machine learning models. In particular, the use of EKI for training neural networks has motivated the development of EKI variants based on ideas used for gradient-based training of neural networks, including dropout,

---

Algorithm 1 Basic Ensemble Kalman Inversion (EKI)

0: **Input:** forward operator $\mathbf{H} : \mathbb{R}^d \to \mathbb{R}^n$, initial ensemble $\{\mathbf{v}_0^{(1)}, \ldots, \mathbf{v}_0^{(J)}\} \subset \mathbb{R}^d$, observations $\mathbf{y} \in \mathbb{R}^n$, observation error covariance $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$

1: **for** $i = 0, 1, 2, \ldots,$ **do**

2:　　Compute observation-space ensemble: $\mathbf{h}_i^{(j)} = \mathbf{H}\big(\mathbf{v}_i^{(j)}\big)$, $j = 1, 2, \ldots, J$.

3:　　Compute empirical covariances: $\mathsf{cov}[\mathbf{v}_i^{(1:J)}, \mathbf{h}_i^{(1:J)}]$ and $\mathsf{cov}[\mathbf{h}_i^{(1:J)}]$

4:　　Compute Kalman gain: $\mathbf{K}_i = \mathsf{cov}[\mathbf{v}_i^{(1:J)}, \mathbf{h}_i^{(1:J)}] \cdot \big(\mathsf{cov}[\mathbf{h}_i^{(1:J)}] + \boldsymbol{\Sigma}\big)^{-1}$

5:　　Sample $\boldsymbol{\varepsilon}_i^{(j)}$ i.i.d. from $\mathcal{N}(0, \boldsymbol{\Sigma})$ for $j = 1, 2, \ldots, J$.

6:　　Perturb observations: set $\mathbf{y}_i^{(j)} = \mathbf{y} + \boldsymbol{\varepsilon}_i^{(j)}$ for $j = 1, 2, \ldots, J$.

7:　　Compute particle update: $\mathbf{v}_{i+1}^{(j)} = \mathbf{v}_i^{(j)} + \mathbf{K}_i(\mathbf{y}_i^{(j)} - \mathbf{H}\mathbf{v}_i^{(j)})$ for $j = 1, 2, \ldots, J$.

8:　　**if** converged **then**

9:　　　**return** current ensemble mean, $\mathsf{E}[\mathbf{v}_{i+1}^{(1:J)}]$

*This is **Stochastic EKI**. For **Deterministic EKI**, skip 5-6 and assign $\mathbf{y}_i^{(j)} = \mathbf{y}$ in 7.*

---

data subsampling (also called '(mini-)batching'), adaptive step sizes, and convergence acceleration with Nesterov momentum.

Theoretical analyses of EKI convergence behavior have mostly considered linear observation operators $\mathbf{H} \in \mathbb{R}^{n \times d}$, for which the standard norm-minimizing solution of (1) is given by

$$\mathbf{v}^* = (\mathbf{H}^\top \boldsymbol{\Sigma}^{-1} \mathbf{H})^\dagger \mathbf{H}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} \equiv \mathbf{H}^+ \mathbf{y}, \tag{2}$$

where "†" denotes the usual Moore-Penrose pseudoinverse and we have introduced the weighted pseudoinverse, $\mathbf{H}^+ = (\mathbf{H}^\top \boldsymbol{\Sigma}^{-1} \mathbf{H})^\dagger \mathbf{H}^\top \boldsymbol{\Sigma}^{-1}$. Previous analyses of linear EKI have largely considered mean-field limits (equivalent to an infinitely large ensemble) [3] or continuous-time limits of the EKI iteration, in which the deterministic iteration becomes a system of coupled ordinary differential equations (ODEs) [2, 6] and the stochastic iteration becomes a system of coupled stochastic differential equations (SDEs) [1]. These continuous-time analyses have shown that the EKI ensemble covariance collapses at a rate inversely proportional to time [6, 1, 2], meaning that the residual of the EKI iteration (with respect to the final solution) converges at a $1/\sqrt{i}$ rate. These analyses have also characterized EKI solutions either by assuming $\mathbf{H}$ is one-to-one or by assuming the ensemble covariance is full rank. In particular, the works [6] show that if $\mathbf{H}$ is one-to-one, then EKI converges to the pre-image of the data restricted to the span of the ensemble [6, 1]. On the other hand, the work [2] shows that if the ensemble covariance is full rank (and $\mathbf{H}$ may be low-rank), then EKI converges to the (non-standard) minimizer of (1) closest to the initial ensemble mean in the norm induced by the initial ensemble covariance. The characterization of EKI solutions in the general case where both $\mathbf{H}$ and the ensemble covariance may be low-rank, and the relationship between EKI solutions and the standard minimum-norm least-squares solution (2), are open questions addressed in this work.

In this work, we provide a new analysis of EKI for linear observation operators $\mathbf{H} \in \mathbb{R}^{n \times d}$ which directly considers the discrete iteration for a finite ensemble, relying principally on linear algebra as an analysis tool. Our analysis yields new results relating EKI solutions to the standard minimum-norm least-squares solution (2), together with a new and natural interpretation of EKI convergence behavior in terms of 'fundamental subspaces of EKI', analogous to the four fundamental subspaces characterizing Strang's 'fundamental theorem of linear algebra' [7], which we now review.

Strang's four 'fundamental subspaces of linear algebra' arise from dividing observation space $\mathbb{R}^n$

and state space $\mathbb{R}^d$ into two *sub*spaces each, one subspace associated with 'observable' directions and a complementary subspace associated with 'unobservable' directions [7]. That is, in observation space $\mathbb{R}^n$, the two fundamental subspaces are:

1. $\mathsf{Ran}(\mathbf{H})$ (denoting the range of $\mathbf{H}$), and

2. $\mathsf{Ker}(\mathbf{H}^\top \mathbf{\Sigma}^{-1})$ (denoting the null space of $\mathbf{H}^\top \mathbf{\Sigma}^{-1}$), the $\mathbf{\Sigma}^{-1}$-orthogonal complement to $\mathsf{Ran}(\mathbf{H})$.

In state space $\mathbb{R}^d$, the two fundamental subspaces are:

1. $\mathsf{Ran}(\mathbf{H}^\top)$, and

2. $\mathsf{Ker}(\mathbf{H})$, the orthogonal complement to $\mathsf{Ran}(\mathbf{H}^\top)$ with respect to the Euclidean norm.

The standard minimum-norm solution (2) to the linear least-squares problem (1) can be understood in terms of these fundamental subspaces as follows (see [5, Figure 1]): in observation space $\mathbb{R}^n$, the closest that $\mathbf{Hv}$ can come to $\mathbf{y} \in \mathbb{R}^n$ with respect to the $\mathbf{\Sigma}^{-1}$-norm is the $\mathbf{\Sigma}^{-1}$-orthogonal projection of $\mathbf{y}$ onto the observable space $\mathsf{Ran}(\mathbf{H})$, which then has a zero component in the (unobservable) subspace $\mathsf{Ker}(\mathbf{H}^\top \mathbf{\Sigma}^{-1})$. In state space $\mathbb{R}^d$, directions in $\mathsf{Ker}(\mathbf{H})$ are unobservable because they are mapped by $\mathbf{H}$ to zero and thus do not influence the minimand of (1). If $\mathsf{Ker}(\mathbf{H})$ is non-trivial, multiple minimizers of (1) exist. The unique norm-minimizing solution (2) lies in the observable space $\mathsf{Ran}(\mathbf{H}^\top)$ and has a zero component in the unobservable space $\mathsf{Ker}(\mathbf{H})$.

Our analysis reveals that EKI solutions to the weighted least squares problem admit a similar interpretation in terms of fundamental subspaces of EKI. However, the EKI fundamental subspaces arise first from dividing the state and observation spaces into directions that are 'populated' by particles, lying in the range of ensemble covariance $\mathbf{\Gamma}_i$ (it is well-known [1, 2, 4, 6] that $\mathsf{Ran}(\mathbf{\Gamma}_i)$ is invariant for all i), and 'unpopulated' directions lying in a complementary subspace. The populated subspace can then be further divided into two subspaces associated with observable and unobservable directions. There are therefore three subspaces in each of the observation and state spaces. In observation space $\mathbb{R}^n$, the three fundamental subspaces of EKI are associated with three complementary oblique projection operators, $\mathcal{P}, \mathcal{Q}, \mathcal{N} \in \mathbb{R}^{n \times n}$. These projections are defined via a spectral analysis of the iteration map which governs the evolution of the *data misfit*, $\mathbf{Hv}_i^{(j)} - \mathbf{y}$, so that the range of each projector is an invariant subspace under the misfit iteration map. The three fundamental subspaces of observation space $\mathbb{R}^n$ are then

1. $\mathsf{Ran}(\mathcal{P}) \equiv \mathsf{Ran}(\mathbf{H}\mathbf{\Gamma}_i)$, associated with observable populated directions,

2. $\mathsf{Ran}(\mathcal{Q}) \equiv \mathbf{H}\,\mathsf{Ker}(\mathbf{\Gamma}_i \mathbf{H}^\top \mathbf{\Sigma}^{-1} \mathbf{H})$, associated with observable but *un*populated directions, and

3. $\mathsf{Ran}(\mathcal{N}) \equiv \mathsf{Ker}(\mathbf{H}^\top \mathbf{\Sigma}^{-1})$, associated with unobservable directions.

In state space $\mathbb{R}^d$, the three fundamental subspaces of EKI are also associated with three complementary oblique projection operators, $\mathbb{P}, \mathbb{Q}, \mathbb{N} \in \mathbb{R}^{n \times n}$. These projections are defined via a spectral analysis of the iteration map which governs the evolution of the *least squares residual*, $\mathbf{v}_i^{(j)} - \mathbf{v}^*$. The range of each projector is again an invariant subspace under the residual iteration map. The three fundamental subspaces of state space $\mathbb{R}^d$ are then

1. $\mathsf{Ran}(\mathbb{P}) \subset \mathsf{Ran}(\mathbf{\Gamma}_i)$, associated with observable populated directions (but generally *not* simply the intersection of $\mathsf{Ran}(\mathbf{\Gamma}_i)$ with $\mathsf{Ran}(\mathbf{H}^\top)$),

2. $\mathsf{Ran}(\mathbb{Q}) \subset \mathsf{Ran}(\mathbf{H}^\top)$, associated with observable *un*populated directions, and

3. $\mathsf{Ran}(\mathbb{N})$, associated with unobservable directions.

The fundamental subspaces of EKI are depicted in [5, Figure 2], and an interactive three-dimensional visualization is available at `https://elizqian.github.io/eki-fundamental-subspaces/`.

3

We show that EKI misfits [residuals] converge to zero at a $1/\sqrt{i}$ rate in the fundamental subspace associated with observable and populated directions, $\mathsf{Ran}(\mathcal{P})$ [$\mathsf{Ran}(\mathbb{P})$], and remain constant in the fundamental subspaces associated with observable unpopulated directions, $\mathsf{Ran}(\mathcal{Q})$ [$\mathsf{Ran}(\mathbb{Q})$]. The misfits [residuals] also remain constant in the unobservable directions, $\mathsf{Ran}(\mathcal{N})$ [$\mathsf{Ran}(\mathbb{N})$]. Numerical experiments illustrating these results may be found in [5, Figure 3]. Our results verify for the discrete iteration and finite ensemble case the $1/\sqrt{i}$ convergence rate previously shown in continuous time or infinite ensemble limits, and provide the first results describing the relationship between EKI solutions and the standard minimum-norm least squares solution (2).

Our analysis sheds light on several directions of interest for future work connecting EKI with classical iterative methods. Because we have shown that the convergence behavior of deterministic EKI can be characterized in terms of an evolving spectral problem that has invariant subspaces that are independent of iteration index, this allows for straightforward EKI acceleration strategies analogous to overrelaxation schemes in classical stationary iterative methods. Other potential directions of interest could exploit the well-known connection between the extended Kalman filter and Gauss-Newton methods to establish further connections between EKI and classical methods.

References

[1] Dirk Blömker, Claudia Schillings, Philipp Wacker, and Simon Weissmann. Well posedness and convergence analysis of the ensemble Kalman inversion. *Inverse Problems*, 35(8):085007, 2019.

[2] Leon Bungert and Philipp Wacker. Complete deterministic dynamics and spectral decomposition of the linear ensemble Kalman inversion. *SIAM/ASA Journal on Uncertainty Quantification*, 11(1):320–357, 2023.

[3] Edoardo Calvello, Sebastian Reich, and Andrew M Stuart. Ensemble Kalman methods: A mean field perspective. *arXiv preprint arXiv:2209.11371*, 2022.

[4] Marco A. Iglesias, Kody J. H. Law, and Andrew M. Stuart. Ensemble Kalman methods for inverse problems. *Inverse Problems*, 29:045001, 2013.

[5] Elizabeth Qian and Christopher Beattie. The fundamental subspaces of ensemble Kalman inversion. *arXiv:2409.08862*, 2024.

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

[7] Gilbert Strang. The fundamental theorem of linear algebra. *The American Mathematical Monthly*, 100(9):848–855, 1993.