$l_p$-regularization for Ensemble Kalman Inversion

Yoonsang Lee

Department of Mathematics
Dartmouth College

SIMDA seminar

October 27, 2020
Inverse problems

Goal: estimate a variable of interest, such as state variables or a set of parameters that constitute a forward model (or a measurement operator) from *noisy, imperfect observation or measurement data.*

**Examples**

- deblurring and denoising in image processing,
- recovery of permeability in subsurface flow using pressure fields,
- training a neural network in machine learning
- estimating sea ice thickness from measurement data
- and more.
Inverse problems

Mathematical formulation

Find \( u \in \mathbb{R}^N \) from measurement data \( y \in \mathbb{R}^m \) where \( u \) and \( y \) are related as follows

\[
y = G(u) + \eta.
\]  

**G : \( \mathbb{R}^N \rightarrow \mathbb{R}^m \)**: a forward model that can be nonlinear and computationally expensive to solve, for example, solving a PDE problem.

**\( \eta \)** is a measurement error. This error is unknown in general, but we assume that it is drawn from a known probability distribution, a Gaussian distribution with mean zero and a known covariance \( \Gamma \).
Inverse problems

Optimization problem
The unknown variable $u$ is estimated by solving an optimization problem

$$\arg\min_{u \in \mathbb{R}^N} \mathcal{R}(u) + \frac{1}{2} \| y - G(u) \|^2_{\Gamma}.$$  \hfill (2)

- $\| \cdot \|_{\Gamma}$: norm induced from the inner product using the inverse of the covariance matrix $\Gamma$, that is $\| a \|^2_{\Gamma} = \langle a, \Gamma^{-1} a \rangle$ for the standard inner product $\langle \cdot, \cdot \rangle$ in $\mathbb{R}^m$.
- $\mathcal{R}(u)$: regularizer, for example, $\| u \|_1$ or $\| u \|_2$. 
Ensemble Kalman Inversion (EKI)

- An optimization method for a nonlinear measurement operator $G(u)$.
- First appeared in oil industry.
- Mathematical formulation (Iglesias et al. ’13) and analysis (Schillings et al. ’17)

Key characteristics

- Derivative-free method that lies between deterministic and probabilistic approaches for inverse problems.
- Iterative use of the Kalman update of ensemble-based Kalman filters.
- Straightforward parallelization.
Ensemble Kalman Inversion (EKI)

One-step Kalman update
From a Gaussian prior $\mathcal{N}(u_{prior}, \Gamma_{prior})$ and a linear measurement $y = Hu$ with a Gaussian error, the posterior distribution is also Gaussian with mean $u_{post}$

$$u_{post} = u_{prior} + K(y - u_{prior})$$

where $K$ is the Kalman gain matrix

$$K = \Gamma_{prior} H^T (H \Gamma_{prior} H^T + \sigma_o^2 I_n)^{-1}.$$  

The posterior covariance matrix $\Gamma_{post}$ is given by

$$\Gamma_{post} = (I - KH)\Gamma_{prior}.$$  

Ensemble Kalman Inversion (EKI)

One-step ensemble-based Kalman update
An ensemble-based method uses a set of samples (an ensemble) to estimate the mean and covariance. Several variants are available based on how to get the posterior ensemble

- Ensemble Kalman Filter (perturbed observations)
- Ensemble Transform Kalman Filter
- Ensemble Adjustment Kalman Filter

To handle a nonlinear measurement operator, use the idea of a trivial dynamics for an augmented variable \((u, G(u))\).
Ensemble Kalman Inversion (EKI)

Algorithm An initial ensemble of size $K$, $\{u_0^{(k)}\}_{k=1}^K$ from prior information, is given. For $n = 1, 2, \ldots$

1. Prediction step using the trivial dynamics:
   (a) Apply the forward model $G$ to each ensemble member
   
   $$g_n^{(k)} := G(u_{n-1}^{(k)})$$

   (b) From the set of the predictions $\{g_n^{(k)}\}_{k=1}^K$, calculate the mean and covariances

   $$\bar{g}_n = \frac{1}{K} \sum_{k=1}^{K} g_n^{(k)},$$

   $$C_{n}^{ug} = \frac{1}{K} \sum_{k=1}^{K} (u_{n}^{(k)} - \bar{u}_n) \otimes (g_n^{(k)} - \bar{g}_n),$$

   $$C_{n}^{gg} = \frac{1}{K} \sum_{k=1}^{K} (g_n^{(k)} - \bar{g}_n) \otimes (g_n^{(k)} - \bar{g}_n),$$

   where $\bar{u}_n$ is the mean of $\{u_{n}^{(k)}\}$, i.e., $\frac{1}{K} \sum_{k=1}^{K} u_{n}^{(k)}$. 
**Ensemble Kalman Inversion (EKI)**

**Algorithm**

2 Analysis step:

(a) Update each ensemble member $u_n^{(k)}$ using the Kalman update

$$u_{n+1}^{(k)} = u_n^{(k)} + C_n^{ug} (C_n^{gg} + \Sigma)^{-1} (y_n^{(k)} - g_n^{(k)}), \quad (6)$$

where $y_{n+1}^{(k)} = y + \zeta_{n+1}^{(k)}$ is a perturbed measurement using Gaussian noise $\zeta_{n+1}^{(k)}$ with mean zero and covariance $\Gamma$.

(b) Compute the mean of the ensemble as an estimate for the solution

$$\bar{u}_{n+1} = \frac{1}{K} \sum_{k=1}^{K} u_n^{(k)} \quad (7)$$
Ensemble Kalman Inversion (EKI)

In this talk, we focus on the ensemble Kalman filter update by Evensen with a constant learning rate.

Possible Variants

- Ensemble square-root (ensemble transform or adjustment) filter updates.
- Adaptive inflation (related to learning rate).
- Localization
Ensemble Kalman Inversion (EKI)

Regularizations in EKI

- Restriction of an ensemble to a compact set
- An iterative regularization that approximates the Levenberg-Marquardt scheme.

These approaches still suffer from overfitting.

Tikhonov EKI (Chada et al. '20) uses an augmented measurement system to impose $l_2$ regularization.
$l_p$ regularization for EKI

L., arXiv:2009.03470.

- Implements $l_p$, $0 < p \leq 1$, regularization; recovery with sparsity.
- Key idea: transformation of a variable.
- The transformation in $l_p$-regularized EKI is explicit and straightforward to calculate for $p \leq 1$.
- A transformation between $l_1$ and $l_2$ regularizations (Wang et al. ’17). A transformation between the Laplace and the Gaussian distributions in the context of Bayesian inference.
$l_p$ regularization for EKI

Transformation from $l_p$ to $l_2$

For $x \in \mathbb{R}$,
\[
\psi(x) = \text{sgn}(x) |x|^\frac{p}{2}, \quad x \in \mathbb{R}.
\] (8)

For $u$ in $\mathbb{R}^N$, a nonlinear map $\Psi : \mathbb{R}^N \rightarrow \mathbb{R}^N$ applies $\psi$ to each component of $u = (u_1, u_2, ..., u_N)$,
\[
\Psi(u) = (\psi(u_1), \psi(u_2), ..., \psi(u_N)).
\] (9)

For $v = \Psi(u)$, it can be checked that for each $i = 1, 2, ..., N$,
\[
|v_i|^2 = |\psi(u_i)|^2 = |u_i|^p,
\]
and thus we have the following norm relation
\[
\|v\|_2^2 = \|u\|_p^p.
\] (10)
$l_p$ regularization for EKI

**Transformation from $l_p$ to $l_2$**
The transformation from $u$ to $v = \Psi(u)$ converts the $l_p$-regularized optimization problem in $u$

$$
\arg\min_{u \in \mathcal{X}} \frac{\lambda}{2} \|u\|^p_p + \frac{1}{2} \|y - G(u)\|^2_F,
$$

(11)

...to a $l_2$ regularized problem in $v$,

$$
\arg\min_{v \in \mathbb{R}^N} \frac{\lambda}{2} \|v\|^2_2 + \frac{1}{2} \|y - \tilde{G}(v)\|^2_F,
$$

(12)

where $\tilde{G}$ is the pullback of $G$ by $\Xi := \Psi^{-1}$

$$
\tilde{G} = G \circ \Xi.
$$

(13)
$l_p$ regularization for EKI

Theorem
For an objective function $J(u) : \mathbb{R}^N \to \mathbb{R}$, if $u^*$ is a local minimizer of $J(u)$, $\Psi(u^*)$ is also a local minimizer of $\tilde{J}(v) = J \circ \Xi(v)$.
Similarly, if $v^*$ is a local minimizer of $\tilde{J}(v)$, then $\Xi(v^*)$ is also a local minimizer of $J(u) = \tilde{J} \circ \Psi(u)$. 
Algorithm An initial ensemble of size $K$, $\{v_0^{(k)}\}_{k=1}^K$, is given. For $n = 1, 2, \ldots$,

1. Prediction step using the forward model:
   (a) Apply the augmented forward model $F$ to each ensemble member
   
   $$f_n^{(k)} := F(v_n^{(k)}) = (\tilde{G}(v_n^{(k)}), v_n^{(k)})$$  \hspace{1cm} (14)

   (b) From the set of the predictions $\{f_n^{(k)}\}_{k=1}^K$, calculate the mean and covariances

   $$\bar{f}_n = \frac{1}{K} \sum_{k=1}^K f_n^{(k)},$$  \hspace{1cm} (15)

   $$C_{n}^{vf} = \frac{1}{K} \sum_{k=1}^K (v_n^{(k)} - \bar{v}_n) \otimes (f_n^{(k)} - \bar{f}_n),$$

   $$C_{n}^{ff} = \frac{1}{K} \sum_{k=1}^K (f_n^{(k)} - \bar{f}_n) \otimes (f_n^{(k)} - \bar{f}_n)$$  \hspace{1cm} (16)

   where $\bar{v}_n$ is the ensemble mean of $\{v_n^{(k)}\}$, i.e., $\frac{1}{K} \sum_{k=1}^K v_n^{(k)}$. 
2. Analysis step:

(a) Update each ensemble member $v_n^{(k)}$ using the Kalman update

$$v_{n+1}^{(k)} = v_n^{(k)} + C_n^v (C_n^{ff} + \Sigma)^{-1} (z_{n+1}^{(k)} - f_n^{(k)}),$$

where $z_{n+1}^{(k)} = z + \zeta_{n+1}^{(k)}$ is a perturbed measurement using Gaussian noise $\zeta_{n+1}^{(k)}$ with mean zero and covariance $\Sigma$.

(b) For the ensemble mean $\overline{v}_n$, the $l_p$ EKI estimate, $u_n$, for the minimizer of the $l_p$ regularization is given by

$$u = \Xi(\overline{v}_n).$$
Numerical test 1: scalar toy problem

Original problem

\[
\argmin_{u \in \mathbb{R}} J(u) = \argmin_{u \in \mathbb{R}} \frac{1}{4} |u|^p + \frac{1}{2} (1 - u)^2. \tag{19}
\]

Using the transformation, we have a $l_2$ regularized problem

\[
\argmin_{v \in \mathbb{R}} \tilde{J}(v) = \argmin_{v \in \mathbb{R}} \frac{1}{4} |v|^2 + \frac{1}{2} (1 - \text{sgn}(v) |v|^{2/p})^2, \tag{20}
\]
Numerical test 1: scalar toy problem

**Figure:** Top: $p = 1$. Bottom: $p = 0.5$.

When $p = 0.5$, $u = v = 0$ is a local minimizer, but not a global one.
Numerical test 1: scalar toy problem

**Figure:** Change of $l_p$EKI estimates, $\xi(\bar{v}_n)$, over iterations
Numerical test 2: compressive sensing

A standard example in image processing.

- $u \in \mathbb{R}^{40}$ is sparse with only four non-zero components out of forty components.
- $G(u) = Au \in \mathbb{R}^{16}$, where $A$ is a random Gaussian matrix.
- Measure error variance: 0.01.
- Ensemble size: 2000

Note that a standard $l_1$ convex method is much faster than $l_p$ EKI for this problem. Our focus is to validate the performance of $l_p$ EKI for $p = 1$ and its result for $p < 1$. 
Numerical test 2: compressive sensing

Reconstruction of $u$

Figure: Reconstruction of sparse signal using $l_p$EKI for $p=2$, $p=1$, and $p=0.2$. The bottom right plot is the reconstruction using the convex $l_1$ minimization method.
Numerical test 2: compressive sensing

Convergence rate

Figure: $l_1$ error of the $l_p$EKI estimate and data misfit.
Numerical test 3: PDE-constrained optimization

A model related to subsurface flow

\[- \nabla \cdot (k(x) \nabla p(x)) = f(x), \quad x = (x_1, x_2) \in (0, 1)^2. \quad (21)\]

Boundary condition

\[
p(x_1, 0) = 100, \quad \frac{\partial p}{\partial x_1}(1, x_2) = 0, \quad -k \frac{\partial p}{\partial x_1}(0, x_2) = 500, \quad \frac{\partial p}{\partial x_2}(x_1, 1) = 0,
\]

and the source term is piecewise constant

\[
f(x_1, x_2) = \begin{cases} 
0 & \text{if } 0 \leq x_2 \leq \frac{4}{6}, \\
137 & \text{if } \frac{4}{6} < x_2 \leq \frac{5}{6}, \\
274 & \text{if } \frac{5}{6} < x_2 \leq 1.
\end{cases}
\]

**Goal**: recovery of the log permeability \( u = \log k(x) \) from partial measurements of the pressure field \( p(x) \).
The log permeability, $\log k$, is represented by 36 components in the cosine basis $\phi_{ij} = \cos(i\pi x_1) \cos(j\pi x_2)$, $i, j = 0, 1, \ldots, 5$,

$$
\log k(x) = \sum_{i,j=0}^{5} u_{ij}\phi_{ij}(x),
$$

where only seven of $\{u_{ij}\}$ are nonzero.

8 × 8 regularly spaced measurement of $p(x)$.

$G(u)$ involves solving the PDE for a given $u = \log k$ whose solution is sampled at the measurement locations.

Measurement error variance: $10^{-5}$.

Ensemble size: 200
Numerical test 3: PDE-constrained optimization

Figure: Left column: the true $u$ and $l_p$EKI estimates for $p = 2$. Right column: log $k$ of the true and $l_p$EKI estimate. Same grey scale.
Numerical test 3: PDE-constrained optimization

**Figure:** Left column: the true $u$ and $l_p$EKI estimates for $p = 1$. Right column: log $k$ of the true and $l_p$EKI estimate. Same grey scale.
Numerical test 3: PDE-constrained optimization

**Figure:** Left column: the true $u$ and $l_p$ EKI estimates for $p = 0.5$. Right column: log $k$ of the true and $l_p$ EKI estimate. Same grey scale.
Numerical test 3: PDE-constrained optimization

Convergence rate

Figure: $l_1$ error of the $l_p$ EKI estimates and data misfit.
Thank you for your attention.