A New Look at the Ensemble Kalman Filter: Optimization and Duality Perspectives

C.G. Krishnanunni, J. Wittmer, H. Nguyen, and T. Bui-Thanh

Probabilistic and High Order Inference, Computation, Estimation, and Simulation (Pho-Ices)
Department of Aerospace Engineering and Engineering Mechanics
The Oden Institute for Computational Engineering and Sciences
The University of Texas at Austin

EnKF Workshop 22, May 30 to June 2, 2022

Acknowledgements

We appreciate the support from:

- **DOE**
- **NSF**
- **UT-Portugal**

Behind the Scence

Left to right: A. Myers, C.G. Krishnanunni, J. Wittmer, H. Nguyen

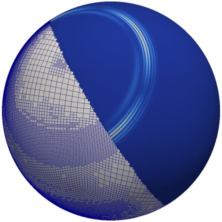
Seismic wave propagation

With C. Burstedde, O. Ghattas, J. R. Martin, G. Stadler, and L. Wilcox

$$\frac{\partial \boldsymbol{E}}{\partial t} = \frac{1}{2}\left(\nabla \boldsymbol{v} + \nabla^T \boldsymbol{v}\right),$$

$$\rho \frac{\partial \boldsymbol{v}}{\partial t} = \nabla \cdot (\mathsf{C}\boldsymbol{E}) + \boldsymbol{f}$$

Strain-velocity formulation



Animated by Greg Abram

- I: fourth-order identity tensor,
- $\boldsymbol{I}$: second-order identity tensor,
- $\boldsymbol{f}$: external volumetric forces,
- C: four-order material tensor.
- $\boldsymbol{E}$: strain tensor,
- $\boldsymbol{v}$: velocity vector,
- ρ: density,
- $\boldsymbol{e}_i$: ith unit vector,

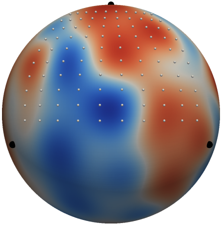
Inverse problem statement

- Earth surface velocity at given locations is recorded
- Infer the wave velocities $c_s = \sqrt{\mu/\rho}$ and $c_p = \sqrt{(\lambda + 2\mu)/\rho}$

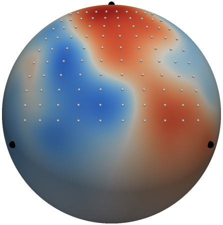
An example of global seismic inversion

with Burstedde, C., Ghattas, O., Martin, J., Stadler, G., and Wilcox

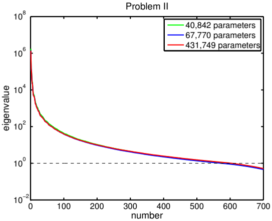
- **inversion field:** c_p in acoustic wave equation
- **prior mean:** PREM (radially symmetric model)
- **"truth" model:** S20RTS (Ritsema et al.), (laterally heterogeneous)
- Piecewise-trilinear on same mesh as forward/adjoint 3rd order dG fields
- dimensions: 1.07 million parameters, 630 million field unknowns
- Final time: $T = 1000s$ with 2400 time steps
- A single forward solve takes 1 minute on 64K Jaguar cores



"truth", sources (black)

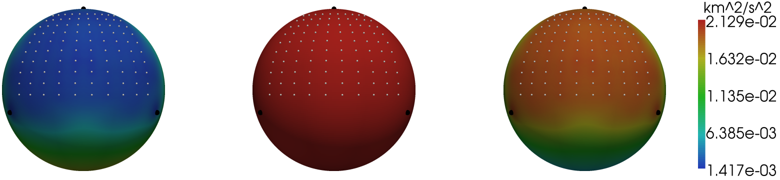


MAP, receivers (white)



Hessian eigenvalues

Uncertainty quantification



$$\mathfrak{C} \approx \mathfrak{C}_0 - \mathfrak{C}_0^{1/2} \boldsymbol{V}_r \boldsymbol{D}_r \boldsymbol{V}_r^* \mathfrak{C}_0^{1/2}$$

Details in:

- Bui-Thanh, T., Burstedde, C., Ghattas, O., Martin, J., Stadler, G., and Wilcox, L.C.,*Extreme-scale UQ for Bayesian inverse problems governed by PDEs*, ACM/IEEE Supercomputing SC12, **Gordon Bell Prize Finalist**, 2012.
- Bui-Thanh, T., Ghattas, O., Martin, J., and Stadler, G., *A computational framework for infinite-dimensional Bayesian inverse problems. Part I: The linearized case*, **SIAM Journal on Scientific Computing**, 35(6), pp. A2494–A2523, 2013.

Outline

1. Optimization-based Ensemble Kalman Inversion

2. EnKF from Duality

PDE-constrained Inverse problem: Summary

Given observations of the form $y = \mathbb{G}(u) + \eta$

$$-\nabla \cdot (e^u \nabla w) \;=\; 0 \text{ in } \Omega$$

$$-e^u \nabla w \cdot \mathbf{n} \;=\; Bi\, w \text{ in } \partial\Omega \setminus \Gamma_R$$

$$-e^u \nabla w \cdot \mathbf{n} \;=\; -1 \text{ on } \Gamma_R$$

Infer u by solving

$$\min_u \underbrace{\frac{1}{2}\left\|y - \mathbb{G}(u)\right\|^2_{\Gamma^{-1}}}_{\mathcal{J}} + \frac{\beta}{2}\left\|u - u_0\right\|^2_{\mathcal{C}^{-1}}$$

EnKF: A Recap

State estimation for dynamical system

$$
\begin{aligned}
u_{n+1} &= \mathbb{M}\left(u_n\right) \\
y_{n+1} &= H u_{n+1} + \eta_{n+1}
\end{aligned}
$$

EnKF: A Recap

State estimation for dynamical system

$$u_{n+1} = \mathbb{M}\left(u_n\right)$$
$$y_{n+1} = H u_{n+1} + \eta_{n+1}$$

Algorithm (Kalman Update)

Ensure: *Initial ensemble* $\{u_n\}_{n \in 1,\ldots,N}$

1: *Forecast* $\hat{u}_{n+1} = \mathbb{M}(u_n)$
2: *Compute Empirical mean and covariance,* $\bar{u}_{n+1}, P_{n+1}$
3: *Compute Kalman gain* $K_{n+1} = P_{n+1} H^T (H P_{n+1} H^T + \Gamma)^{-1}$
4: *analysis* $u_{n+1} = \hat{u}_{n+1} + K_{n+1}(y_{n+1} - H\hat{u}_{n+1})$

where $\bar{u}_n = \frac{1}{J}\sum_{j=1}^{J} u_n^j, \quad P_n = \frac{1}{J}\sum_{j=1}^{J} u_n^j (u_n^j)^T - \bar{u}_n(\bar{u}_n)^T$

"Standard" Ensemble Kalman Inversion (EnKI): A Recap

Iglesias, Law, and Stuart 13

The joint state-parameter estimation in the context of EnKF proceeds by defining a mapping, artificial dynamics, and observation operator as follows:

$$\mathbb{M}(z) = \begin{Bmatrix} u \\ \mathbb{G}(u) \end{Bmatrix}$$

$$z_{n+1} = \begin{Bmatrix} u_{n+1} \\ p_{n+1} \end{Bmatrix} = \mathbb{M}(z_n) = \begin{Bmatrix} u_n \\ \mathbb{G}(u_n) \end{Bmatrix}$$

$$y_{n+1} = y + \eta_{n+1}$$

$$H = [0, I]$$

- EnKF uses an ensemble of particles, each updated at every iteration by the Kalman update formula
- $\{y_n\}$ is generated by perturbing true observed data y by a Gaussian random variable

Results for Standard EnKI

- 1D nonlinear elliptic inverse example with 32 parameters
- Initial samples are drawn from $\mathcal{N}\left(0, 0.01^2 I\right)$, and noise $\mathcal{N}\left(0, 0.5^2 I\right)$

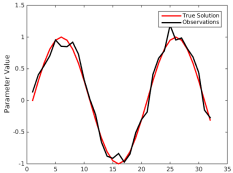


Figure: True Solution

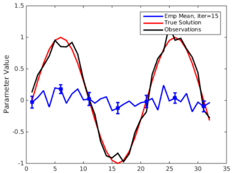


Figure: 15 iter

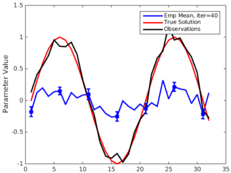


Figure: 40 iter

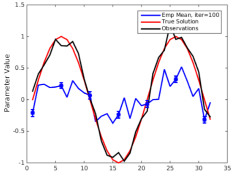


Figure: 100 iter

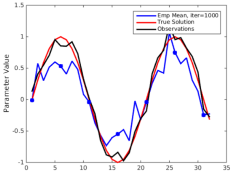


Figure: 1000 iter

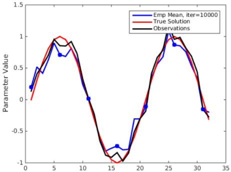


Figure: 10000 iter

Standard dynamics and Gradient Descent dynamics

Standard Dynamics:

$$\begin{Bmatrix} u_{n+1} \\ p_{n+1} \end{Bmatrix} = \begin{Bmatrix} u_n \\ \mathbb{G}(u_n) \end{Bmatrix}$$

Gradient Descent Dynamics:

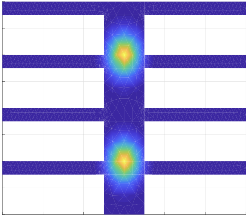
$$\begin{Bmatrix} u_{n+1} \\ p_{n+1} \end{Bmatrix} = \begin{Bmatrix} u_n - \alpha F(a) \\ \mathbb{G}(u_n - \alpha F(a)) \end{Bmatrix}$$

Method	F	a
GD-OpEnKF	$\nabla \mathcal{J}$	$\{u\}_{i=1}^N$
GDmean-OpEnKF	$\nabla \mathcal{J}$	$\bar{u}$
H-OpEnKF	$(\nabla^2 \mathcal{J})^{-1} \nabla \mathcal{J}$	$\{u\}_{i=1}^N$
Hmean-OpEnKF	$(\nabla^2 \mathcal{J})^{-1} \nabla \mathcal{J}$	$\bar{u}$
GDa-OpEnKF	$\sum_{i=1}^N w_i (u_i - \bar{u})$	$\{u_i\}_{i=1}^N$
GDmax-OpEnKF	$\sum_{i=1}^N w_{max} (u_i - \bar{u})$	$\{u_i\}_{i=1}^N$

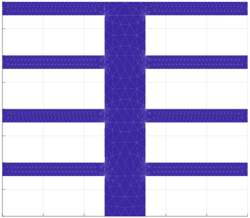
Table: Variations of OpEnKF

Optimized-based EnKI (OpEnKI)

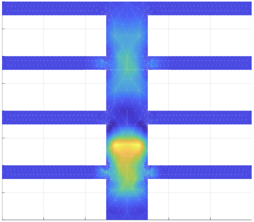
Nonlinear inversion in two spatial dimension: 50 particles, 1.3K parameters



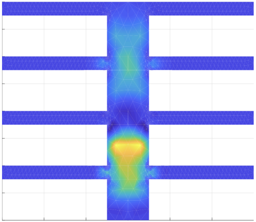
(a) True parameter



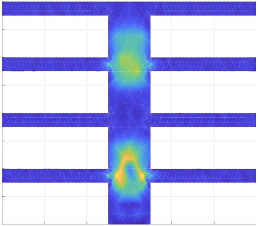
(b) EnKI 11K



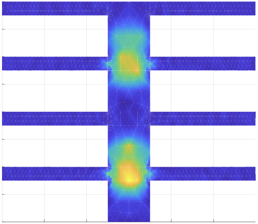
(c) GD-OpEnKI 100



(d) GDmean-OpEnKI 100



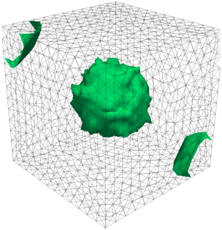
(e) H-OpEnKI 100



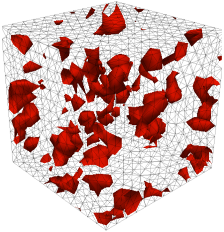
(f) Hmean-OpEnKI 100

Optimized-based EnKI (OpEnKI)

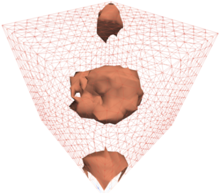
Nonlinear inversion in three spatial dimension: 100 particles, 12K parameters



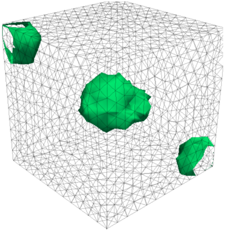
(g) True parameter



(h) EnKI 55



(i) GDa-OpEnKI 50



(j) H-OpEnKI 15

Optimized-based EnKI (OpEnKI)

Convergence in the small time limit

Theorem

Suppose $\mathbb{G}(u) = \mathcal{A}u$. OpEnKI can be considered as an Euler-Maruyama discretization of of an SDE. Ignoring the noise, and in the limit of zero time step, there holds

$$\frac{du^j}{dt} = \underbrace{-C(u)\nabla\mathcal{J}(u^j)}_{\text{"Gauss-Newton" step}} - \underbrace{\alpha C(u)\tilde{H}_{GN}\nabla\mathcal{J}(\bar{u})}_{\text{"Gradient" step}}$$

Two observations

1. The gradient step of the mean particle is introduced by OpEnKF

Optimized-based EnKI (OpEnKI)

Convergence in the small time limit

Theorem

Suppose $\mathbb{G}(u) = \mathcal{A}u$. OpEnKI can be considered as an Euler-Maruyama discretization of of an SDE. Ignoring the noise, and in the limit of zero time step, there holds

$$\frac{du^j}{dt} = \underbrace{-C(u)\nabla\mathcal{J}(u^j)}_{\text{"Gauss-Newton" step}} \underbrace{-\alpha C(u)\tilde{H}_{GN}\nabla\mathcal{J}(\bar{u})}_{\text{"Gradient" step}}$$

Two observations

1. The gradient step of the mean particle is introduced by OpEnKF
2. While the subspaces spanned by EnKF particles remains the same, those of OpEnKF changes from one step to another.

Outline

Duality view of Kalman Filter

Linear forward + Gausian prior + Gaussian noise

One-Step Kalman Filter: Dual view 1

$$\min_{\mathbf{u},\mathbf{v}} \mathcal{J}(\mathbf{u};\mathbf{v};\mathbf{u}_0,\mathbf{d}) := \frac{1}{2}\|\mathbf{d}-\mathbf{v}\|^2_{\mathbf{\Gamma}^{-1}} + \frac{1}{2}\|\mathbf{u}-\mathbf{u}_0\|^2_{\mathcal{C}^{-1}}, \quad s.t. \quad \mathcal{A}\mathbf{u} = \mathbf{v}$$

Duality view of Kalman Filter

Linear forward + Gausian prior + Gaussian noise

One-Step Kalman Filter: Dual view 1

$$\min_{\mathbf{u},\mathbf{v}} \mathcal{J}(\mathbf{u};\mathbf{v};\mathbf{u}_0,\mathbf{d}) := \frac{1}{2}\|\mathbf{d}-\mathbf{v}\|^2_{\mathbf{\Gamma}^{-1}} + \frac{1}{2}\|\mathbf{u}-\mathbf{u}_0\|^2_{\mathcal{C}^{-1}}, \quad s.t. \quad \mathcal{A}\mathbf{u} = \mathbf{v}$$

One-Step Kalman Filter: Dual view 2

$$\mathcal{L}(\mathbf{u},\mathbf{v},\boldsymbol{\lambda}) = \frac{1}{2}\|\mathbf{d}-\mathbf{v}\|^2_{\mathbf{\Gamma}^{-1}} + \frac{1}{2}\|\mathbf{u}-\mathbf{u}_0\|^2_{\mathcal{C}^{-1}} + \boldsymbol{\lambda}^T\left(A\mathbf{u}-\mathbf{v}\right).$$

Duality view of Kalman Filter

Linear forward + Gausian prior + Gaussian noise

One-Step Kalman Filter: Dual view 1

$$\min_{\mathbf{u},\mathbf{v}} \mathcal{J}(\mathbf{u};\mathbf{v};\mathbf{u}_0,\mathbf{d}) := \frac{1}{2}\|\mathbf{d}-\mathbf{v}\|^2_{\mathbf{\Gamma}^{-1}} + \frac{1}{2}\|\mathbf{u}-\mathbf{u}_0\|^2_{\mathcal{C}^{-1}}, \quad s.t. \quad \mathcal{A}\mathbf{u} = \mathbf{v}$$

One-Step Kalman Filter: Dual view 2

$$\mathcal{L}(\mathbf{u},\mathbf{v},\boldsymbol{\lambda}) = \frac{1}{2}\|\mathbf{d}-\mathbf{v}\|^2_{\mathbf{\Gamma}^{-1}} + \frac{1}{2}\|\mathbf{u}-\mathbf{u}_0\|^2_{\mathcal{C}^{-1}} + \boldsymbol{\lambda}^T\left(A\mathbf{u}-\mathbf{v}\right).$$

One-Step Kalman Filter: Dual view 3

$$\mathcal{D}(\boldsymbol{\lambda}) := \inf_{\mathbf{u},\mathbf{v}} \mathcal{L}(\mathbf{u},\mathbf{v},\boldsymbol{\lambda}) = -\frac{1}{2}\boldsymbol{\lambda}^T\left(\mathbf{\Gamma}+\mathcal{A}\mathcal{C}\mathcal{A}^T\right)\boldsymbol{\lambda} + \boldsymbol{\lambda}^T\left(\mathcal{A}\mathbf{u}_0-\mathbf{d}\right)$$

Duality view of Kalman Filter

Linear forward + Gausian prior + Gaussian noise

One-Step Kalman Filter: Dual view 1

$$\min_{\mathbf{u},\mathbf{v}} \mathcal{J}(\mathbf{u};\mathbf{v};\mathbf{u}_0,\mathbf{d}) := \frac{1}{2}\|\mathbf{d}-\mathbf{v}\|^2_{\mathbf{\Gamma}^{-1}} + \frac{1}{2}\|\mathbf{u}-\mathbf{u}_0\|^2_{\mathcal{C}^{-1}}, \quad s.t. \quad \mathcal{A}\mathbf{u} = \mathbf{v}$$

One-Step Kalman Filter: Dual view 2

$$\mathcal{L}(\mathbf{u},\mathbf{v},\boldsymbol{\lambda}) = \frac{1}{2}\|\mathbf{d}-\mathbf{v}\|^2_{\mathbf{\Gamma}^{-1}} + \frac{1}{2}\|\mathbf{u}-\mathbf{u}_0\|^2_{\mathcal{C}^{-1}} + \boldsymbol{\lambda}^T\left(\mathcal{A}\mathbf{u}-\mathbf{v}\right).$$

One-Step Kalman Filter: Dual view 3

$$\mathcal{D}(\boldsymbol{\lambda}) := \inf_{\mathbf{u},\mathbf{v}} \mathcal{L}(\mathbf{u},\mathbf{v},\boldsymbol{\lambda}) = -\frac{1}{2}\boldsymbol{\lambda}^T\left(\mathbf{\Gamma}+\mathcal{A}\mathcal{C}\mathcal{A}^T\right)\boldsymbol{\lambda} + \boldsymbol{\lambda}^T\left(\mathcal{A}\mathbf{u}_0-\mathbf{d}\right)$$

One-Step Kalman Filter: dual view solution

$$\boldsymbol{\lambda}^* = \arg\max_{\boldsymbol{\lambda}} \mathcal{D}(\boldsymbol{\lambda}) = \left(\mathbf{\Gamma}+\mathcal{A}\mathcal{C}\mathcal{A}^T\right)^{-1}\left(\mathcal{A}\mathbf{u}_0-\mathbf{d}\right)$$

$$\mathbf{u}^* = \mathbf{u}_0 - \mathcal{C}\mathcal{A}^T\boldsymbol{\lambda}^* = \mathbf{u}_0 + \mathcal{C}\mathcal{A}^T\left(\mathbf{\Gamma}+\mathcal{A}\mathcal{C}\mathcal{A}^T\right)^{-1}\left(\mathbf{d}-\mathcal{A}\mathbf{u}_0\right)$$

Randomized Dual

Linear forward + Gausian prior + Gaussian noise

Define

$$\boldsymbol{\sigma} \sim \pi_{\boldsymbol{\sigma}} := \mathcal{N}\left(0, \boldsymbol{\Gamma}\right), \qquad \boldsymbol{\delta} \sim \pi_{\boldsymbol{\delta}} := \mathcal{N}\left(0, \mathcal{C}\right)$$

$$\boldsymbol{\xi} := \left[\boldsymbol{\sigma}, \boldsymbol{\delta}\right], \qquad \pi\left(\boldsymbol{\sigma}, \boldsymbol{\delta}\right) := \pi_{\boldsymbol{\sigma}} \times \pi_{\boldsymbol{\delta}}$$

$\left[\boldsymbol{\delta}_1, \boldsymbol{\delta}_2 ... \boldsymbol{\delta}_N\right]$ and $\left[\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2 ... \boldsymbol{\sigma}_N\right]$ be ensemble pairs from $\pi\left(\boldsymbol{\sigma}, \boldsymbol{\delta}\right)$

Randomized Dual

$\mathscr{D}\left(\boldsymbol{\lambda}\right) = \mathbb{E}_\pi\left[\tilde{\mathcal{D}}\left(\boldsymbol{\lambda}; \mathbf{u}_0, \mathbf{d}, \boldsymbol{\xi}\right)\right]$, where

$$\tilde{\mathcal{D}}\left(\boldsymbol{\lambda}; \mathbf{u}_0, \mathbf{d}, \boldsymbol{\xi}\right) := -\frac{1}{2}\boldsymbol{\lambda}^T\left(\boldsymbol{\Gamma} + \mathcal{A}\boldsymbol{\delta}\boldsymbol{\delta}^T\mathcal{A}^T\right)\boldsymbol{\lambda} + \boldsymbol{\lambda}^T\left(\mathcal{A}\left(\mathbf{u}_0 + \boldsymbol{\delta}\right) - \mathbf{d} - \boldsymbol{\sigma}\right).$$

Randomized Dual

Linear forward + Gausian prior + Gaussian noise

Define

$$\boldsymbol{\sigma} \sim \pi_{\boldsymbol{\sigma}} := \mathcal{N}(0, \boldsymbol{\Gamma}), \qquad \boldsymbol{\delta} \sim \pi_{\boldsymbol{\delta}} := \mathcal{N}(0, \mathcal{C})$$

$$\boldsymbol{\xi} := [\boldsymbol{\sigma}, \boldsymbol{\delta}], \qquad \pi(\boldsymbol{\sigma}, \boldsymbol{\delta}) := \pi_{\boldsymbol{\sigma}} \times \pi_{\boldsymbol{\delta}}$$

$[\boldsymbol{\delta}_1, \boldsymbol{\delta}_2 ... \boldsymbol{\delta}_N]$ and $[\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2 ... \boldsymbol{\sigma}_N]$ be ensemble pairs from $\pi(\boldsymbol{\sigma}, \boldsymbol{\delta})$

Randomized Dual

$\mathscr{D}(\boldsymbol{\lambda}) = \mathbb{E}_\pi \left[\tilde{\mathcal{D}}(\boldsymbol{\lambda}; \mathbf{u}_0, \mathbf{d}, \boldsymbol{\xi}) \right]$, where

$$\tilde{\mathcal{D}}(\boldsymbol{\lambda}; \mathbf{u}_0, \mathbf{d}, \boldsymbol{\xi}) := -\frac{1}{2} \boldsymbol{\lambda}^T \left(\boldsymbol{\Gamma} + \mathcal{A} \boldsymbol{\delta} \boldsymbol{\delta}^T \mathcal{A}^T \right) \boldsymbol{\lambda} + \boldsymbol{\lambda}^T \left(\mathcal{A}(\mathbf{u}_0 + \boldsymbol{\delta}) - \mathbf{d} - \boldsymbol{\sigma} \right).$$

Monte Carlo approximation of $\mathscr{D}(\boldsymbol{\lambda})$ is:

$$\mathscr{D}_N := -\frac{1}{2N} \sum_{i=1}^{N} \boldsymbol{\lambda}^T \left(\boldsymbol{\Gamma} + \mathcal{A} \boldsymbol{\delta}_i \boldsymbol{\delta}_i^T \mathcal{A}^T \right) \boldsymbol{\lambda} + \boldsymbol{\lambda}^T \left(\mathcal{A}(\mathbf{u}_0 + \boldsymbol{\delta}_i) - \mathbf{d} - \boldsymbol{\sigma}_i \right).$$

Randomized Dual

Linear forward + Gausian prior + Gaussian noise

The optimal solution for the dual problem reads

$$\overline{\boldsymbol{\lambda}}^* := \arg\max_{\boldsymbol{\lambda}} \mathscr{D}_N = \left(\boldsymbol{\Gamma} + \mathcal{A}\Omega\Omega^T\mathcal{A}^T\right)^{-1}\left(\mathcal{A}\left(\mathbf{u}_0 + \overline{\boldsymbol{\delta}}\right) - (\mathbf{d} + \overline{\boldsymbol{\sigma}})\right).$$

Randomized Dual

Linear forward + Gausian prior + Gaussian noise

The optimal solution for the dual problem reads

$$\overline{\boldsymbol{\lambda}}^* := \arg\max_{\boldsymbol{\lambda}} \mathscr{D}_N = \left(\boldsymbol{\Gamma} + \mathcal{A}\Omega\Omega^T\mathcal{A}^T\right)^{-1}\left(\mathcal{A}\left(\mathbf{u}_0 + \overline{\boldsymbol{\delta}}\right) - \left(\mathbf{d} + \overline{\boldsymbol{\sigma}}\right)\right).$$

The induced Monte Carlo optimal solution for the primal problem reads

$$\hat{\mathbf{u}}^* := \mathbf{u}_0 - \mathcal{C}\mathcal{A}^T\overline{\boldsymbol{\lambda}}^* = \mathbf{u}_0 + \mathcal{C}\mathcal{A}^T\left(\boldsymbol{\Gamma} + \mathcal{A}\Omega\Omega^T\mathcal{A}^T\right)^{-1}\left(\mathbf{d} + \overline{\boldsymbol{\sigma}} - \mathcal{A}\left(\mathbf{u}_0 + \overline{\boldsymbol{\delta}}\right)\right).$$

Randomized Dual

Linear forward + Gausian prior + Gaussian noise

The optimal solution for the dual problem reads

$$\overline{\boldsymbol{\lambda}}^* := \arg\max_{\boldsymbol{\lambda}} \mathscr{D}_N = \left(\boldsymbol{\Gamma} + \mathcal{A}\Omega\Omega^T\mathcal{A}^T\right)^{-1}\left(\mathcal{A}\left(\mathbf{u}_0 + \overline{\boldsymbol{\delta}}\right) - \left(\mathbf{d} + \overline{\boldsymbol{\sigma}}\right)\right).$$

The induced Monte Carlo optimal solution for the primal problem reads

$$\hat{\mathbf{u}}^* := \mathbf{u}_0 - \mathcal{C}\mathcal{A}^T\overline{\boldsymbol{\lambda}}^* = \mathbf{u}_0 + \mathcal{C}\mathcal{A}^T\left(\boldsymbol{\Gamma} + \mathcal{A}\Omega\Omega^T\mathcal{A}^T\right)^{-1}\left(\mathbf{d} + \overline{\boldsymbol{\sigma}} - \mathcal{A}\left(\mathbf{u}_0 + \overline{\boldsymbol{\delta}}\right)\right).$$

If we further randomize $\mathbf{u}_0$ and $\mathcal{C}$ with $\mathbf{u}_0 + \overline{\boldsymbol{\delta}}$ and $\Omega\Omega^T$:

$$\overline{\mathbf{u}}^* := \mathbf{u}_0 + \overline{\boldsymbol{\delta}} + \Omega\Omega^T\mathcal{A}^T\left(\boldsymbol{\Gamma} + \mathcal{A}\Omega\Omega^T\mathcal{A}^T\right)^{-1}\left(\mathbf{d} + \overline{\boldsymbol{\sigma}} - \mathcal{A}\left(\mathbf{u}_0 + \overline{\boldsymbol{\delta}}\right)\right),$$

which is exactly the EnKI.

Randomized Dual

Linear forward + Gausian prior + Gaussian noise

The optimal solution for the dual problem reads

$$\overline{\boldsymbol{\lambda}}^* := \arg\max_{\boldsymbol{\lambda}} \mathscr{D}_N = \left(\boldsymbol{\Gamma} + \mathcal{A}\Omega\Omega^T\mathcal{A}^T\right)^{-1}\left(\mathcal{A}\left(\mathbf{u}_0 + \overline{\boldsymbol{\delta}}\right) - \left(\mathbf{d} + \overline{\boldsymbol{\sigma}}\right)\right).$$

The induced Monte Carlo optimal solution for the primal problem reads

$$\hat{\mathbf{u}}^* := \mathbf{u}_0 - \mathcal{C}\mathcal{A}^T\overline{\boldsymbol{\lambda}}^* = \mathbf{u}_0 + \mathcal{C}\mathcal{A}^T\left(\boldsymbol{\Gamma} + \mathcal{A}\Omega\Omega^T\mathcal{A}^T\right)^{-1}\left(\mathbf{d} + \overline{\boldsymbol{\sigma}} - \mathcal{A}\left(\mathbf{u}_0 + \overline{\boldsymbol{\delta}}\right)\right).$$

If we further randomize $\mathbf{u}_0$ and $\mathcal{C}$ with $\mathbf{u}_0 + \overline{\boldsymbol{\delta}}$ and $\Omega\Omega^T$:

$$\overline{\mathbf{u}}^* := \mathbf{u}_0 + \overline{\boldsymbol{\delta}} + \Omega\Omega^T\mathcal{A}^T\left(\boldsymbol{\Gamma} + \mathcal{A}\Omega\Omega^T\mathcal{A}^T\right)^{-1}\left(\mathbf{d} + \overline{\boldsymbol{\sigma}} - \mathcal{A}\left(\mathbf{u}_0 + \overline{\boldsymbol{\delta}}\right)\right),$$

which is exactly the EnKI.

Unrolling the sums

$$\mathbf{u}_i := \mathbf{u}_0 + \boldsymbol{\delta}_i + \Omega\Omega^T\mathcal{A}^T\left(\boldsymbol{\Gamma} + \mathcal{A}\Omega\Omega^T\mathcal{A}^T\right)^{-1}\left(\mathbf{d} + \boldsymbol{\sigma}_i - \mathcal{A}\left(\mathbf{u}_0 + \boldsymbol{\delta}_i\right)\right),$$

which is the well-known EnKF.

Non-Asymptotic Convergence

Linear forward + Gausian prior + Gaussian noise

Theorem (Non-asymptotic error estimator for EnKI)

Let $\boldsymbol{\delta}_i \sim \mathcal{N}(0, \mathcal{C})$, *and* $\boldsymbol{\sigma}_i \sim \mathcal{N}(0, \boldsymbol{\Gamma})$, $i = 1, \ldots, N$. *For any* $0 < \varepsilon < \dfrac{\left\| \boldsymbol{\Gamma} + \mathcal{A}\mathcal{C}\mathcal{A}^T \right\|}{\kappa(\mathcal{B}) \left\|\mathcal{A}\right\|^2 \left\|\mathcal{C}\right\|}$, *there exists a constant* c, *independent of* ε, *such that*

$$\left\| \mathbf{u}^* - \overline{\mathbf{u}}^* \right\| \le c\varepsilon$$

holds with probability at least $1 - 6\exp\left(-\mathcal{O}\left(N\varepsilon^2\right)\right)$.

Observations

1. The larger the ensemble size N, the higher the probability of predicting $\mathbf{u}^*$,

Non-Asymptotic Convergence

Linear forward + Gausian prior + Gaussian noise

Theorem (Non-asymptotic error estimator for EnKI)

Let $\boldsymbol{\delta}_i \sim \mathcal{N}(0, \mathcal{C})$, *and* $\boldsymbol{\sigma}_i \sim \mathcal{N}(0, \boldsymbol{\Gamma})$, $i = 1, \ldots, N$. *For any* $0 < \varepsilon < \dfrac{\left\| \boldsymbol{\Gamma} + \mathcal{A}\mathcal{C}\mathcal{A}^T \right\|}{\kappa(\mathcal{B}) \left\| \mathcal{A} \right\|^2 \left\| \mathcal{C} \right\|}$, *there exists a constant* c, *independent of* ε, *such that*

$$\left\| \mathbf{u}^* - \overline{\mathbf{u}}^* \right\| \leq c\varepsilon$$

holds with probability at least $1 - 6\exp\left(-\mathcal{O}\left(N\varepsilon^2\right)\right)$.

Observations

1. The larger the ensemble size N, the higher the probability of predicting $\mathbf{u}^*$,
2. We need around $N = \mathcal{O}\left(\varepsilon^{-2}\right)$ to have accurate prediction with high probability.

Ensemble Collapse

Linear forward + Gausian prior + Gaussian noise

- $\mathbf{u}_i^k$: ith sample in the kth iteration
- $\boldsymbol{e}_i^k = \mathbf{u}_i^k - \mathbf{u}^*$: difference between the ith sample in the kth iteration and the truth

Theorem (No-noise $\mathbf{d} = \mathcal{A}\mathbf{u}^*$)

The following holds:

1. *If Range* $\left(\mathcal{A}\Omega^0\right) =$ *Range* $(\mathcal{A})$, *then*

$$\mathcal{A}\boldsymbol{e}_i^k \to 0, \text{ as } k \to \infty$$

Ensemble Collapse

Linear forward + Gausian prior + Gaussian noise

- $\mathbf{u}_i^k$: ith sample in the kth iteration
- $\boldsymbol{e}_i^k = \mathbf{u}_i^k - \mathbf{u}^*$: difference between the ith sample in the kth iteration and the truth

Theorem (No-noise $\mathbf{d} = \mathcal{A}\mathbf{u}^*$)

The following holds:

1. *If Range* $\left(\mathcal{A}\Omega^0\right) =$ *Range* $(\mathcal{A})$, *then*

$$\mathcal{A}\boldsymbol{e}_i^k \to 0, \text{ as } k \to \infty$$

2. *If Range* $\left(\mathcal{A}\Omega^0\right) =$ *Range* $(\mathcal{A})$ *and Null* $(\mathcal{A}) = \{0\}$ *then*

$$\boldsymbol{e}_i^k \to 0, \text{ as } k \to \infty$$

Summary

Presented

1. Optimization-based EnKI

Summary

Presented

1. Optimization-based EnKI
2. EnKF as randomization of dual optimization problem

Outlook

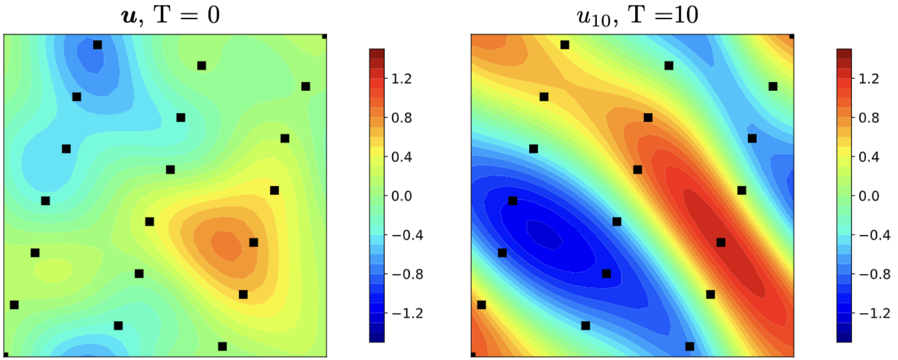
Model-constrained Deep Learning for Inverse Problems

$$\begin{aligned}
\partial_t u(x,t) + v(x,t)\cdot\nabla u(x,t) &= \nu\Delta u(x,t) + f(x), && x\in(0,1)^2, t\in(0,T] \\
\nabla\cdot v(x,t) &= 0, && x\in(0,1)^2, t\in(0,T] \\
u(x,0) &= u_0(x), && x\in(0,1)^2
\end{aligned}$$

Given vorticity u at a few points at final time, infer u_0

Outlook

Model-constrained Deep Learning for Inverse Problems



Outlook

Model-constrained Deep Learning for Inverse Problems