Hybrid (Variational / Kalman) ensemble methods for state estimation in NS Systems

Thomas Bewley, Joseph Cessna

Flow Control and Coordinated Robotics Labs, UC San Diego

Outline

1. **Problem Formulation**
2. **Propagation of uncertainty**
   - complete answer
   - (Extended) Kalman Filter
   - Ensemble Kalman Filter
   - Ensemble Kalman Smoother
3. **Variational methods [aka 4Dvar, MHE]**
   - Gradient derivation in mixed CT/DT setting
4. **Ensemble 3D & 4D Variational Assimilation [En3DVar, En4DVar]**
5. **Hybrid Ensemble Smoother [HEnS]**
6. Some preliminary results and interpretation

Problem Formulation

Continuous-time ODE system, discrete-time measurements:

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{f}(\mathbf{x}(t), \mathbf{w}(t)), \quad \mathbf{y}_k = H\mathbf{x}_k + \mathbf{v}_k, \quad \mathbf{x}_k = \mathbf{x}(t_k) = \mathbf{x}(k\Delta t).$$

Zero-mean, uncorrelated noise with autocorrelation modeled:

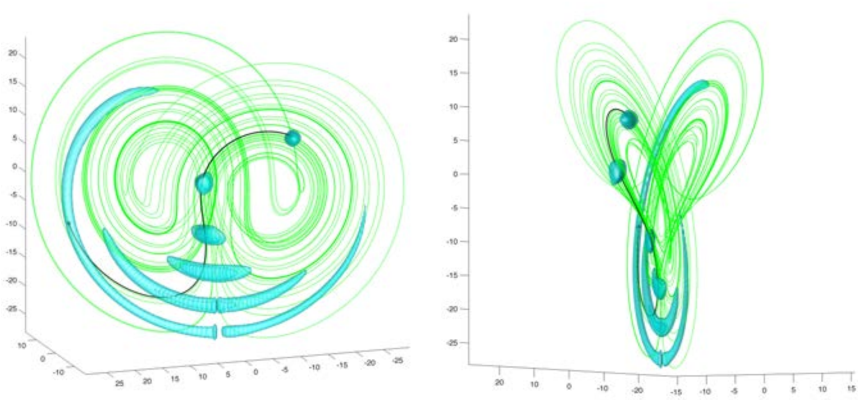
$$E\{\mathbf{v}_k\} = 0, \quad E\{\mathbf{w}(t)\} = 0, \quad R_{\mathbf{v}}(j;k) = E\{\,\mathbf{v}_{k+j}\,\mathbf{v}_k^T\,\} = R\,\delta_{j0},$$

$$R_{\mathbf{w}}(\tau;t) = E\{\,\mathbf{w}(t+\tau)\,\mathbf{w}^T(t)\,\} = Q\,\delta^\sigma(\tau), \quad \delta^\sigma(\tau) = \frac{e^{-\tau^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}}.$$

Covariance of “white” state disturbances $R > 0$. Spectral density of “essentially white” (i.e. $0 < \sigma \ll 1$) measurement noise $Q \geq 0$.

Propagation of uncertainty [complete answer]

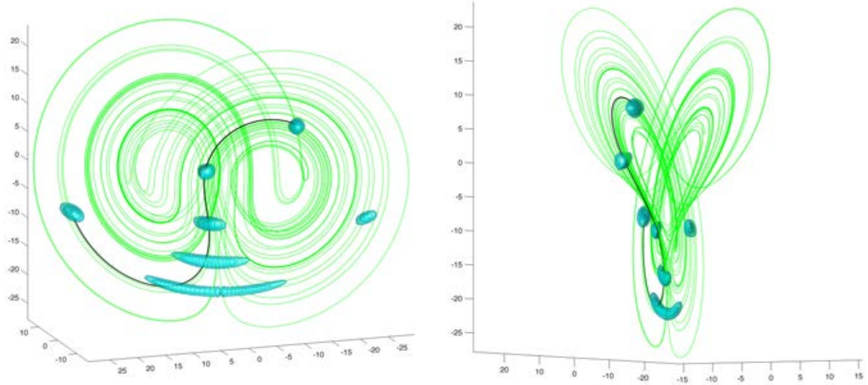
Track entire probability density function (PDF) $\mathcal{P}(\mathbf{x}, t)$ in phase space.



Evolution of uncertainty in Lorenz system without measurements.

Propagation of uncertainty [complete answer]

Track entire probability density function (PDF) $\mathcal{P}(\mathbf{x}, t)$ in phase space.



Evolution of uncertainty in Lorenz system with measurements.

Propagation of uncertainty [complete answer]

Track entire probability density function (PDF) $\mathcal{P}(\mathbf{x}, t)$ in phase space.

1. March PDF between measurements according to Fokker-Planck:

$$\frac{\partial \mathcal{P}}{\partial t} = -\frac{\partial (f_i\, \mathcal{P})}{\partial x_i} + q_{ij}\frac{\partial^2 \mathcal{P}}{\partial x_i \partial x_j}$$

where f_i are components of RHS $\mathbf{f}(\bar{\mathbf{x}}(t), 0)$ of ODE, and q_{ij} are components of spectral density Q of state disturbances $\mathbf{w}(t)$.

2. Update PDF at measurement times according to Bayes rule:

$$\mathcal{P}(\mathbf{x})_{k|k} = \frac{\mathcal{P}(\mathbf{m}|\mathbf{x})\, \mathcal{P}(\mathbf{x})_{k|k-1}}{\mathcal{P}(\mathbf{m})}$$

[Notation: $\mathcal{P}_{k|\ell}$ means the value of $\mathcal{P}$ at timestep k given information through timestep ℓ.]

Propagation of uncertainty [(Extended) Kalman Filter]

Don't track entire PDF. Instead, track "football of uncertainty": that is, the state estimate $\bar{\mathbf{x}}(t) = E[x]$ and covariance $P = E\big[\,(\mathbf{x}-\bar{\mathbf{x}})(\mathbf{x}-\bar{\mathbf{x}})^T\big]$.

1. March $\bar{\mathbf{x}}(t)$ and P between measurements according to

$$\frac{d\bar{\mathbf{x}}(t)}{dt} = \mathbf{f}(\bar{\mathbf{x}}(t), 0), \quad \frac{dP}{dt} = AP + PA^T + Q, \tag{1}$$

where A is the linearization of $\mathbf{f}(\mathbf{x}(t), 0)$ in the vicinity of $\bar{\mathbf{x}}(t)$.

2. Update $\bar{\mathbf{x}}(t)$ and P at measurements according to

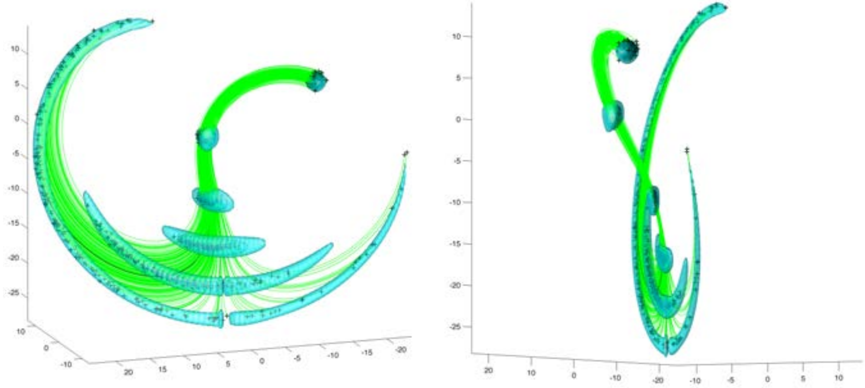
$$\bar{\mathbf{x}}_{k|k} = \bar{\mathbf{x}}_{k|k-1} + K\,(\,\mathbf{y}_k - H\,\bar{\mathbf{x}}_{k|k-1}), \tag{2a}$$

$$P_{k|k} = (I - KH)\,P_{k|k-1}(I - KH)^T + K\,R\,K^T, \tag{2b}$$

$$K = P_{k|k-1}H^T\big[H P_{k|k-1}H^T + R\big]^{-1}. \tag{2c}$$

Propagation of uncertainty [Ensemble Kalman Filter]

Lagrangian approach in phase space: track ‘ensemble members’ $\hat{\mathbf{x}}^j(t)$.



Approximation of uncertainty with candidate trajectories.

Propagation of uncertainty [Ensemble Kalman Filter]

Individual ensemble members $\hat{\mathbf{x}}^j(t)$ excited by disturbances $\mathbf{w}^j(t)$ and $\mathbf{v}_k^j$ with statistics consistent with $\mathbf{w}(t)$ and $\mathbf{v}_k$:

$$\frac{d\hat{\mathbf{x}}^j(t)}{dt} = f(\hat{\mathbf{x}}^j(t), \mathbf{w}^j(t)), \quad \mathbf{d}_k^j = \mathbf{y}_k + \mathbf{v}_k^j,$$

$$\hat{\mathbf{x}}^j_{k|k} = \hat{\mathbf{x}}^j_{k|k-1} + P^e_{k|k-1} H^T (H P^e_{k|k-1} H^T + R)^{-1} (\mathbf{d}_k^j - H \hat{\mathbf{x}}^j_{k|k-1})$$

$$\text{with} \quad \hat{\mathbf{x}}^j_{k|k-1} = \hat{\mathbf{x}}^j(t_k) \text{ before update}, \quad \hat{\mathbf{x}}^j_{k|k} = \hat{\mathbf{x}}^j(t_k) \text{ after update},$$

where the (low-rank) ensemble approximation of the covariance P is

$$P^e = \frac{(\delta\mathbf{X})\,(\delta\mathbf{X})^T}{N-1}, \quad \delta\mathbf{X} = \begin{bmatrix} \delta\hat{\mathbf{x}}^1 & \delta\hat{\mathbf{x}}^2 & \cdots & \delta\hat{\mathbf{x}}^N \end{bmatrix},$$

$$\delta\hat{\mathbf{x}}^j = \hat{\mathbf{x}}^j - \bar{\mathbf{x}}, \quad \bar{\mathbf{x}} = \frac{1}{N}\sum \hat{\mathbf{x}}^j.$$

Random excitations essential when distribution is undersampled!

Update of uncertainty [Ensemble Kalman Smoother]

Update past estimates (at t_k) based on future observations (up to t_ℓ, with $\ell > k$) via:

$$\hat{\mathbf{x}}^j_{k|\ell} = \hat{\mathbf{x}}^j_{k|\ell-1} + S^e_{\ell-1} H^T (H P^e_{\ell|\ell-1} H^T + R)^{-1} (\mathbf{d}^j_\ell - H \hat{\mathbf{x}}^j_{\ell|\ell-1}),$$

where

$$S^e_{\ell-1} = \frac{(\delta \widehat{X}_{k|\ell-1})\,(\delta \widehat{X}_{\ell|\ell-1})^T}{N-1}.$$

Taking $\ell = k$, EnKS reduces to EnKF.

Variational methods [aka 4Dvar, MHE]

Consider window $t \in (\,0, T\,]$ with K measurement times t_k where

$$0 < t_1 < t_2 < \ldots < t_K = T.$$

Take

$$\frac{d\tilde{\mathbf{x}}(t)}{dt} = \mathbf{f}(\tilde{\mathbf{x}}(t), 0), \quad \tilde{\mathbf{x}}_0 = \mathbf{u}.$$

Goal: minimize cost function:

$$J(\mathbf{u}) = \frac{1}{2}\,(\,\mathbf{u} - \bar{\mathbf{x}}_{0|0})^T\,P_{0|0}^{-1}\,(\,\mathbf{u} - \bar{\mathbf{x}}_{0|0}) + \frac{1}{2}\sum_{k=1}^{K}\big(\,\mathbf{y}_k - H\,\tilde{\mathbf{x}}_k\,\big)^T\,R^{-1}\,\big(\,\mathbf{y}_k - H\,\tilde{\mathbf{x}}_k\,\big).$$

Weakness: doesn't quantify new uncertainties, $P_{0|K}$ and $P_{K|K}$.

Gradient derivation in mixed CT/DT setting

Perturbation of system:

$$\frac{d\tilde{\mathbf{x}}'(t)}{dt} = A\big(\tilde{\mathbf{x}}(t)\big)\,\tilde{\mathbf{x}}'(t) \quad \text{with} \quad \tilde{\mathbf{x}}_0' = \mathbf{u}'$$
$$\Rightarrow \quad \mathcal{L}\,\tilde{\mathbf{x}}' = 0 \quad \text{where} \quad \mathcal{L} = \frac{d}{dt} - A\big(\tilde{\mathbf{x}}(t)\big)$$

Perturbation of cost function:

$$J'(\mathbf{u}') = (\mathbf{u} - \bar{\mathbf{x}}_{0|0})^T\, P_{0|0}^{-1}\,\mathbf{u}' - \sum_{k=1}^{K} \big[H^T\, R^{-1}\,(\mathbf{y}_k - H\tilde{\mathbf{x}}_k)\big]^T \tilde{\mathbf{x}}_k'$$

Goal: rewrite as follows to identify gradient $\nabla J(\mathbf{u})$:

$$J'(\mathbf{u}') = \langle\langle \nabla J(\mathbf{u})\,,\,\mathbf{u}'\,\rangle\rangle \triangleq \big[\nabla J(\mathbf{u})\,\big]^T\,\mathbf{u}'$$

Gradient derivation in mixed CT/DT setting

Duality pairing, adjoint identity defined piecewise on $t \in (t_{k-1}, t_k)$:

$$\langle \mathbf{r}^{(k)}, \tilde{\mathbf{x}}' \rangle \triangleq \int_{t_{k-1}}^{t_k} (\mathbf{r}^{(k)})^T \tilde{\mathbf{x}}' \, dt \quad \Rightarrow \quad \langle \mathbf{r}^{(k)}, \mathcal{L}\, \tilde{\mathbf{x}}' \rangle = \langle \mathcal{L}^* \mathbf{r}^{(k)}, \tilde{\mathbf{x}}' \rangle + b^{(k)}.$$

$$\Rightarrow \quad \mathcal{L}^* \mathbf{r}^{(k)} = -\frac{d\mathbf{r}^{(k)}(t)}{dt} - A\big(\tilde{\mathbf{x}}(t)\big)^T \mathbf{r}^{(k)}(t),$$

$$b^{(k)} = (\mathbf{r}_k^{(k)})^T \tilde{\mathbf{x}}'_k - (\mathbf{r}_{k-1}^{(k)})^T \tilde{\mathbf{x}}'_{k-1} \quad \text{where} \quad \mathbf{r}_k^{(k)} \triangleq \mathbf{r}^{(k)}(t_k).$$

Perturbation of cost function [K'th step - peel off last term of sum]:

$$J'(\mathbf{u}') = (\mathbf{u} - \bar{\mathbf{x}}_{0|0})^T P_{0|0}^{-1} \mathbf{u}' - J'_K - \sum_{k=1}^{K-1} [H^T R^{-1} (\mathbf{y}_k - H\tilde{\mathbf{x}}_k)]^T \tilde{\mathbf{x}}'_k,$$

$$J'_K = \big[H^T R^{-1} (\mathbf{y}_K - H \tilde{\mathbf{x}}_K)\big]^T \tilde{\mathbf{x}}'_K$$

Definition of adjoint [K'th step - march backwards from $t_K = T$]:

$$\mathcal{L}^* \mathbf{r}^{(K)} = 0, \quad \mathbf{r}_K^{(K)} = H^T R^{-1} (\mathbf{y}_K - H \tilde{\mathbf{x}}_K)$$

Gradient derivation in mixed CT/DT setting

Substituting into adjoint identity [K'th step]:

$$b^{(K)} = 0 \quad \Rightarrow \quad (\mathbf{r}_{K}^{(K)})^T \, \tilde{\mathbf{x}}'_{K} - (\mathbf{r}_{K-1}^{(K)})^T \, \tilde{\mathbf{x}}'_{K-1} = 0 \quad \Rightarrow$$

$$\left[H^T R^{-1} (\mathbf{y}_{K} - H \, \tilde{\mathbf{x}}_{K})\right]^T \tilde{\mathbf{x}}'_{K} = (\mathbf{r}_{K-1}^{(K)})^T \, \tilde{\mathbf{x}}'_{K-1} \quad \Rightarrow \quad J'_{K} = (\mathbf{r}_{K-1}^{(K)})^T \, \tilde{\mathbf{x}}'_{K-1}$$

Perturbation of cost function [$(K-1)$'th step - peel off next term]:

$$J'(\mathbf{u}') = (\, \mathbf{u} - \bar{\mathbf{x}}_{0|0})^T \, P_{0|0}^{-1} \, \mathbf{u}' - J'_{K-1} - \sum_{k=1}^{K-2} [H^T R^{-1} (\mathbf{y}_k - H \tilde{\mathbf{x}}_k)]^T \, \tilde{\mathbf{x}}'_k,$$

$$J'_{K-1} = \left[\, H^T R^{-1} (\, \mathbf{y}_{K-1} - H \, \tilde{\mathbf{x}}_{K-1} \,) + \mathbf{r}_{K-1}^{(K)} \,\right]^T \tilde{\mathbf{x}}'_{K-1}$$

Definition of adjoint [$(K-1)$'th step - continue backwards march]:

$$\mathcal{L}^* \, \mathbf{r}^{(K-1)} = 0, \quad \mathbf{r}_{K-1}^{(K-1)} = H^T R^{-1} (\mathbf{y}_{K-1} - H \, \tilde{\mathbf{x}}_{K-1}) + \mathbf{r}_{K-1}^{(K)}$$

$$\Rightarrow \quad \dots \quad \Rightarrow \quad J'_{K-1} = (\mathbf{r}_{K-2}^{(K-1)})^T \, \tilde{\mathbf{x}}'_{K-2}$$

Gradient derivation in mixed CT/DT setting

Continuing all the way back to the first step gives

$$J'(\mathbf{u}') = (\,\mathbf{u} - \bar{\mathbf{x}}_{0|0})^T P_{0|0}^{-1}\, \mathbf{u}' - (\mathbf{r}_0^{(1)})^T\, \tilde{\mathbf{x}}'_0$$

with

$$\frac{d\mathbf{r}^{(K)}(t)}{dt} = -A\big(\tilde{\mathbf{x}}(t)\big)^T \mathbf{r}^{(K)}(t), \quad \mathbf{r}_K^{(K)} = \ 0\ + H^T R^{-1}(\mathbf{y}_K - H\,\tilde{\mathbf{x}}_K)$$

$$\vdots$$

$$\frac{d\mathbf{r}^{(1)}(t)}{dt} = -A\big(\tilde{\mathbf{x}}(t)\big)^T \mathbf{r}^{(1)}(t), \qquad \mathbf{r}_1^{(1)} = \mathbf{r}_1^{(2)} + H^T R^{-1}(\mathbf{y}_1 - H\,\tilde{\mathbf{x}}_1)$$

Piecewise continuous march over $T \to 0$, update w/ measurements $\mathbf{y}_k$:

$$\frac{d\mathbf{r}(t)}{dt} = -A\big(\tilde{\mathbf{x}}(t)\big)^T \mathbf{r}(t) \quad \text{with} \quad \mathbf{r}_k \leftarrow \mathbf{r}_k + H^T\, R^{-1}\,(\,\mathbf{y}_k - H\,\tilde{\mathbf{x}}_k\,).$$

Resulting in identification of gradient $\nabla J(\mathbf{u})$:

$$J'(\mathbf{u}') = \Big[\, P_{0|0}^{-1}\,(\,\mathbf{u} - \bar{\mathbf{x}}_{0|0}) - \mathbf{r}_0 \Big]^T \mathbf{u}' \quad \Rightarrow \quad \nabla J(\mathbf{u}) = P_{0|0}^{-1}\,(\,\mathbf{u} - \bar{\mathbf{x}}_{0|0}) - \mathbf{r}_0\,.$$

Ensemble 3D Variational Assimilation [En3DVar]

The ensemble members after the measurement update, $\mathbf{u}^j = \hat{\mathbf{x}}^j_{k|k}$ for $j = 1, \ldots, N$, are now defined by *independent* minimization of the "component cost functions" $J_j(\mathbf{u}^j)$ where

$$J_j(\mathbf{u}^j) = \frac{1}{2}\,(\mathbf{u}^j - \hat{\mathbf{x}}^j_{k|k-1})^T\,(P^e_{k|k-1})^{-1}\,(\mathbf{u}^j - \hat{\mathbf{x}}^j_{k|k-1}) + \frac{1}{2}\,(\mathbf{d}^j_k - H\mathbf{u}^j)^T\,R^{-1}\,(\mathbf{d}^j_k - H\mathbf{u}^j).$$

with, as in EnKF,

$$\mathbf{d}^j_k = \mathbf{y}_k + \mathbf{v}^j_k.$$

Theorem 1. *In the limit of $N \to \infty$, the En3DVar update is equivalent to the corresponding Kalman filter update.*

Ensemble 3D Variational Assimilation [En3DVar]

Proof. Note that each $J_j(\mathbf{u}^j)$ is convex in $\mathbf{u}^j = \hat{\mathbf{x}}^j_{k|k}$, with gradient

$$\nabla J_j = (P^e_{k|k-1})^{-1}\,(\mathbf{u}^j - \hat{\mathbf{x}}^j_{k|k-1}) - H^T\,R^{-1}\,(\mathbf{d}^j_k - H\,\mathbf{u}^j).$$

At the minimum, $\nabla J_j = 0$ for each j, and thus

$$0 = \big[(P^e_{k|k-1})^{-1} + H^T\,R^{-1}\,H\big]\,(\hat{\mathbf{x}}^j_{k|k} - \hat{\mathbf{x}}^j_{k|k-1}) - H^T\,R^{-1}\,(\mathbf{d}^j_k - H\hat{\mathbf{x}}^j_{k|k-1})$$

$$\Rightarrow\;\; (\hat{\mathbf{x}}^j_{k|k} - \hat{\mathbf{x}}^j_{k|k-1}) = \big[\,(P^e_{k|k-1})^{-1} + H^T\,R^{-1}\,H\,\big]^{-1} H^T\,R^{-1}\,(\mathbf{d}^j_k - H\hat{\mathbf{x}}^j_{k|k-1}).$$

Applying the following form of the matrix inversion lemma

$$\big[\,(P^e_{k|k-1})^{-1} + H^T\,R^{-1}\,H\,\big]^{-1}\,H^T\,R^{-1} = P^e_{k|k-1}H^T\,\big[\,H\,P^e_{k|k-1}\,H^T + R\big]^{-1},$$

we obtain [cf. (2c)]

$$\hat{\mathbf{x}}^j_{k|k} = \hat{\mathbf{x}}^j_{k|k-1} + K\,(\mathbf{d}^j_k - H\hat{\mathbf{x}}^j_{k|k-1}), \quad K = P^e_{k|k-1}H^T\big[H\,P^e_{k|k-1}H^T + R\big]^{-1}.$$

Ensemble 3D Variational Assimilation [En3DVar]

Taking the sample mean of the updated ensemble [cf. (2a)],

$$\bar{\mathbf{x}}_{k|k} = \frac{1}{N}\sum_{j=1}^{N} \hat{\mathbf{x}}^j_{k|k} = \frac{1}{N}\sum_{j=1}^{N} \hat{\mathbf{x}}^j_{k|k-1} + K\left(\frac{1}{N}\sum_{j=1}^{N} \mathbf{d}^j_k - H\,\frac{1}{N}\sum_{j=1}^{N} \hat{\mathbf{x}}^j_{k|k-1}\right)$$

$$\xrightarrow[N\to\infty]{} \bar{\mathbf{x}}_{k|k-1} + K\,(\mathbf{y}_k - H\,\bar{\mathbf{x}}_{k|k-1}).$$

Taking the sample covariance of the updated ensemble [cf. (2b)],

$$\begin{aligned} P^e_{k|k} &= \frac{1}{N-1}\sum_{j=1}^{N} (\hat{\mathbf{x}}^j_{k|k} - \bar{\mathbf{x}}_{k|k})\,(\hat{\mathbf{x}}^j_{k|k} - \bar{\mathbf{x}}_{k|k})^T \\ &= (I - KH)\,P^e_{k|k-1}(I - KH)^T + K\,R^e\,K^T + \Phi + \Phi^T, \end{aligned}$$

$$\Phi = \frac{1}{N-1}\sum_{j=1}^{N} \left\{(I - K\,H)\,(\hat{\mathbf{x}}^j_{k|k-1} - \bar{\mathbf{x}}_{k|k-1})(\mathbf{d}^j_k - \mathbf{y}_k)^T\,K^T\right\}.$$

As $N\to\infty$, the spurious correlations between the background error and the measurement noise decreases, and $\Phi\to 0$. □

Ensemble 4D Variational Assimilation [En4DVar]

Over $t \in (0, T]$ with K measurements, the ensemble members after the measurement updates, $\mathbf{u}^j = \hat{\mathbf{x}}^j_{k|k}$ for $j = 1, \dots, N$, are now given by minimization of the component cost functions

$$J_j(\mathbf{u}^j) = \frac{1}{2}(\mathbf{u}^j - \hat{\mathbf{x}}^j_{0|0})^T (P^e_{0|0})^{-1} (\mathbf{u}^j - \hat{\mathbf{x}}^j_{0|0}) + \frac{1}{2} \sum_{k=1}^{K} (\mathbf{d}^j_k - H\,\tilde{\mathbf{x}}^j_k)^T R^{-1} (\mathbf{d}^j_k - H\,\tilde{\mathbf{x}}^j_k)$$

with

$$\frac{d\tilde{\mathbf{x}}^j(t)}{dt} = f(\tilde{\mathbf{x}}^j(t), 0), \quad \tilde{\mathbf{x}}^j_0 = \mathbf{u}^j, \quad \mathbf{d}^j_k = \mathbf{y}_k + \mathbf{v}^j_k.$$

The j'th gradient derivation follows as in the 4Dvar case:

$$\nabla J_j(\mathbf{u}^j) = (P^e_{0|0})^{-1} (\mathbf{u}^j - \hat{\mathbf{x}}^j_{0|0}) - \mathbf{r}^j_0.$$

[An ensemble of states, and an ensemble of adjoints!]

Ensemble 4D Variational Assimilation [En4DVar]

Theorem 2. *In the limit of $N \to \infty$ under the assumption of linear dynamics and Gaussian disturbances, the En4DVar solution converges to the equivalent Kalman smoother.*

Proof. Similar to that of Theorem 1. □

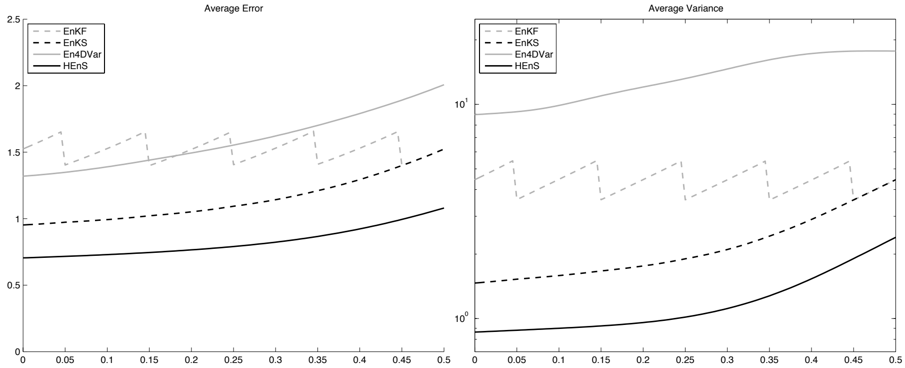
Note: En4DVar formulation is *vector based*, and optimizations are *embarassingly parallel* (after $P^e_{0|0}$ is shared between threads). Thus, En4DVar inherits the numerical tractability of popular ensemble-based and variational-based methods for high-dimensional discretizations of NS systems.

Hybrid Ensemble Smoother [HEnS]

Simply initialize the En4DVar optimization with the EnKS result instead of the background $\hat{\mathbf{x}}^{j}_{0|0}$.

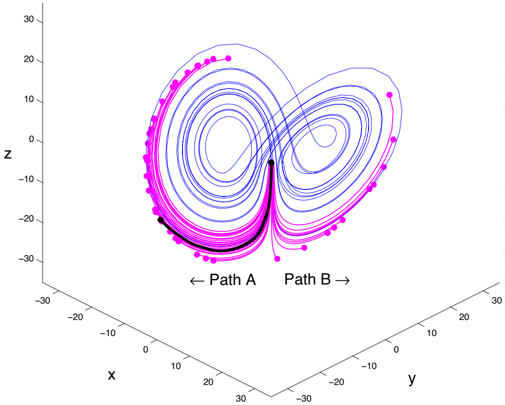
This doesn't affect the converged result in the LQ case, but better initialization typically leads to better result in nonlinear systems, especially when complete convergence of the (local) optimization scheme applied to the variational problem considered is not anticipated.

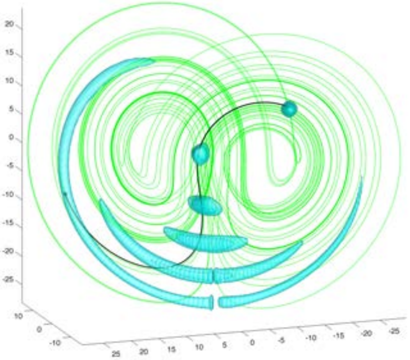
Some results



Improved accuracy (lower error) and precision (lower variance) of HEnS in Lorenz problem over the optimization window, averaged over many optimization windows.

Some interpretation





Variational methods revisit past measurements in light of new data.
Ensemble methods provide a low-rank approximation of covariance P.
This work shows how to put these methods together effectively.