Projected Data Assimilation

Erik S. Van Vleck
Department of Mathematics
University of Kansas
erikvv@ku.edu
people.ku.edu/~erikvv

SIAM Conference on Dynamical Systems

May 23, 2019

Joint work with John Maclean (Adelaide)

Inverse Problem

Consider a discrete time stochastic model

$$u_{n+1} = F_n(u_n) + \omega_n, \;\; n = 0, 1, \ldots \tag{1}$$

where $u_n \in \mathbb{R}^N$ are the state variables at time n and, e.g., $\omega_n \sim \mathcal{N}(0, \mathbf{\Sigma})$, i.e., drawn from a normal distribution with mean zero and model error covariance $\mathbf{\Sigma}$.

Let the sequence $\{u_0^t, u_1^t, \ldots\}$ denote the unknown "truth."

As each time t_n is reached we collect an observation y_n related to u_n^t via

$$y_n = \mathbf{H} u_n^t + \eta_n, \qquad y_n \in \mathbb{R}^M \tag{2}$$

where $\mathbf{H} : \mathbb{R}^N \to \mathbb{R}^M$, $M \leq N$, is the observation operator, and, e.g., $\eta_n \sim \mathcal{N}(0, \mathbf{R})$.

In general the observation operator can be nonlinear.

Projected Physical Model

Consider a nonlinear evolution equation (solution operator of a model)

$$u_{n+1} = F_n(u_n; \boldsymbol{\alpha}), \ \ n = 0, 1, ..., N$$

where

- u_n are the state variables at time n,
- $\boldsymbol{\alpha}$ are adjustable model parameters, e.g., global in time.

Write $u_n = u_n^{(0)} + \delta_n$.

If we can decompose the time dependent tangent space into slow variables and fast variables, then we write $\delta_n = \Pi_n \delta_n + (I - \Pi_n)\delta_n$.

Rewrite original nonlinear evolution approximately as two subsystems ...

For $k = 0, 1, 2, ...$ (k like Newton iterate)

P System: [$u_n^{(k)}$ and $u_{n+1}^{(k)}$ known $(d_n^+ := \Pi_n \delta_n, \forall n)$]

$$u_{n+1}^{(k)} + d_{n+1}^+ = \Pi_{n+1} F_n(u_n^{(k)} + d_n^+),$$

Update: $u_n^{(k+\frac{1}{2})} = u_n^{(k)} + d_n^+, \;\; n = 0, 1, ..., N-1,$

I-P System: [$u_n^{(k+\frac{1}{2})}$ and $u_{n+1}^{(k+\frac{1}{2})}$ known $(d_n^- := (I - \Pi_n)\delta_n, \forall n)$]

$$u_{n+1}^{(k+\frac{1}{2})} + d_{n+1}^- = (I - \Pi_{n+1}) F_n(u_n^{(k+\frac{1}{2})} + d_n^-),$$

Update: $u_n^{(k+1)} = u_n^{(k+\frac{1}{2})} + d_n^-, \;\; n = 0, 1, ..., N-1.$

Π_n can be $\Pi_n^{(k)}$ or also $\Pi_n^{(k+1/2)}$.

Forming time dependent projections

Discrete QR algorithm for determining Lyapunov exponents, Sacker-Sell spectrum, local in time stability information, etc.:

For $Q_0 \in \mathbb{R}^{N \times p}$ random such that $Q_0^T Q_0 = I$,

$$Q_{n+1} R_n = F_n'(u_n) Q_n \approx \frac{1}{\epsilon}[F_n(u_n + \epsilon Q_n) - F_n(u_n)], \;\; n = 0, 1, \ldots$$

where $Q_{n+1}^T Q_{n+1} = I$ and R_n is upper triangular with positive diagonal elements.

Orthogonal Projections:

$$\Pi_n = Q_n Q_n^T, \;\; I - \Pi_n = I - Q_n Q_n^T$$

Roughly speaking the first subsystem contains the slow variables (positive, zero, and slightly negative Lyapunov exponents) and the second subsystem contains the fast variables (strongly negative).

Importance of observations rich in unstable subspace

- Assimilation in the Unstable Subspace (AUS) [Carrassi, Trevisan, Uboldi '07 Tellus, Carrassi, Ghil, Trevisan, Uboldi '08 Chaos, Trevisan, D'Isidoro, Talagrand '10 Q.J.R. Meteorol. Soc., Palatella, Carrassi, Trevisan '13 J. Phys A, ...]
- Error analysis in DA for hyperbolic system [González-Tokman, Hunt '13 Phys D]
- Adaptive observation operators and unstable subspace [K.J.H. Law, D. Sanz-Alonso, A. Shukla, A.M. Stuart '16 Phys D, ...]
- Filter stability, convergence of covariances matrices in unstable subspace [Bosquet etal. '17 SIAM UQ, Grudzien etal. '18 SIAM UQ, Frank and Zhuk '18 Nonlin, ...]

Example: AUS

Consider a linear (or linearized) physical model

$$u_{n+1} = A_n u_n + \omega_n, \;\; \omega_n \sim \mathcal{N}(0, \mathbf{\Sigma})$$

together with a linear data model $y_n = \mathbf{H} u_n + \eta_n$, $\eta_n \sim \mathcal{N}(0, \mathbf{R})$.

Projecting the physical model we have

$$\Pi_{n+1} u_{n+1} = \Pi_{n+1} A_n \Pi_n u_n + \Pi_{n+1} \omega_n$$

Using $Q_{n+1} R_n = A_n Q_n$, letting $v_n = \Pi_n u_n$, $x_n = Q_n^T v_n = Q_n^T u_n$,

$$v_{n+1} = Q_{n+1} R_n Q_n^T v_n + \Pi_{n+1} \omega_n, \;\; x_{n+1} = R_n x_n + Q_{n+1}^T \omega_n$$

$[\Pi_{n+1} \omega_n \sim \mathcal{N}(0, \Pi_{n+1} \mathbf{\Sigma} \Pi_{n+1}),\ Q_{n+1}^T \omega_n \sim \mathcal{N}(0, Q_{n+1}^T \mathbf{\Sigma} Q_{n+1})]$

EKF-AUS

For the unprojected physical and data models, the extended Kalman filter calculates the exact posterior $u_n|y_n \sim \mathcal{N}(u_n^a, P_n^a)$, where the **analysis** variables are

$$u_n^a = u_n^f + \mathbf{K}_n(y_n - \mathbf{H}u_n^f) \ , \tag{3}$$

$$P_n^a = (\mathbf{I} - \mathbf{K}_n\mathbf{H})\, P_n^f. \tag{4}$$

The matrix $\mathbf{K}_n$ is the Kalman gain

$$\mathbf{K}_n = P_n^f \mathbf{H}^T \left(\mathbf{H}P_n^f\mathbf{H}^T + \mathbf{R}\right)^{-1} \ . \tag{5}$$

The superscript f is used for **forecast** mean and covariance,

$$u_n^f = \mathbf{A}_{n-1}u_{n-1}^a + \omega_{n-1} \ ,$$

$$P_n^f = \mathbf{A}_{n-1}P_{n-1}^a\mathbf{A}_{n-1}^T + \mathbf{\Sigma} \ .$$

EKF-AUS is obtained by forming $\tilde{u}_n^f = \Pi_n u_n^f$, $\tilde{P}_n^f = \Pi_n P_n^f \Pi_n$, etc..

Example: Projected DA

Techniques for P System:

- Particle filters,
- Residual correction (Pseudo Orbit DA (PDA) [Du & Smith I & II, '14 J. Atmos. Sci. 2014],
- Shadowing refinement [Grebogi, Hammel, Yorke, and Sauer, Phys Rev Lett (1990)]
  - Strong model constraint $G(u) = 0$ where $G(u)_n \equiv u_{n+1} - F_n(u_n; \boldsymbol{\alpha}), n = 0 : N$,
  - Solve BVP with no BCs (more unknowns than constraints),
  - Small Gauss-Newton updates $\delta = -DG(u)^{\dagger} G(u)$,
  - Dimension reduction with projection ($\Pi_n = Q_n Q_n^T$).

Techniques I-P System:

- Techniques such as ETKF, LETKF, No DA, ...

Shadowing Refinement and Parameter Estimation

In [B. de Leeuw, S. Dubinkina, J. Franks, A. Steyer, X. Tu, EVV, "Projected Shadowing Based Data Assimilation," (2018) *SIAM J. Appld. Dyn. Sys.*]
developed an **interval sequential/smoothing technique** where

all observations over each subinterval are employed simultaneously, using

- **P:** Shadowing refinement and parameter estimation
  - $G = G(u, \boldsymbol{\alpha})$, augment parameters as variables but don't add $\dot{\alpha} = 0$ (neutral modes),
  - G-N update of parameters using Sherman-Morrison-Woodbury,
- **I-P:** Insertion synchronization/no DA, enforcing continuity between subintervals in the $I - P$ components.

Projected Data Model ($\Pi_n = Q_n Q_n^T$)

Consider linear observation operator $\mathbf{H}$ with full row rank.

- [Original]

$$y_n = \mathbf{H} u_n + \eta_n, \eta_n \sim \mathcal{N}(0, \mathbf{R})$$

- [Map to phase space]

$$\tilde{y}_n = \mathbf{H}^\dagger y_n = \Pi_H u_n + \mathbf{H}^\dagger \eta_n, \mathbf{H}^\dagger \eta_n \sim \mathcal{N}(0, \mathbf{H}^\dagger \mathbf{R} (\mathbf{H}^\dagger)^T)$$

where $\mathbf{H}^\dagger = \mathbf{H}^T(\mathbf{H}\mathbf{H}^T)^{-1}$, $\Pi_H = \mathbf{H}^T(\mathbf{H}\mathbf{H}^T)^{-1}\mathbf{H} = \mathbf{H}^\dagger \mathbf{H}$,

- [Projected]

$$y_n^p = \Pi_n \mathbf{H}^\dagger y_n \equiv \mathbf{H}_n^p u_n + \gamma_n, \gamma_n \sim \mathcal{N}(0, \mathbf{R}_n^p),$$

where $\mathbf{R}_n^p = \Pi_n \mathbf{H}^\dagger \mathbf{R} (\mathbf{H}^\dagger)^T \Pi_n$ and $\mathbf{H}_n^p = \Pi_n \mathbf{H}^\dagger$,

- [Reduced dimensional projected representation]

$$y_n^q = Q_n^T H^\dagger y_n \equiv \mathbf{H}_n^q u_n + \gamma_n, \gamma_n \sim \mathcal{N}(0, \mathbf{R}_n^q),$$

where $\mathbf{R}_n^q = Q_n \mathbf{H}^\dagger \mathbf{R} (\mathbf{H}^\dagger)^T Q_n^T$ and $\mathbf{H}_n^q = Q_n^T \mathbf{H}^\dagger$.

Properties

Theorem: $p(y^q|u) = p(y^p|u)$ (with the observation error covariances $\mathbf{R}_n^q$ and $\mathbf{R}_n^p$).

If $\mathbf{H}\tilde{Q}_n$ is full rank, then the covariance matrix $\mathbf{R}_n^q$ of y_n^q is invertible and y_n^q, e.g., has a standard normal distribution.

Convergence: Using Bayes' rule and $p(u|y, y^q) = p(u|y)$,

$$p(u|y^q) = p(u|y)\frac{p(y|y^q)}{p(y|u, y^q)}.$$

Examples

- Projected Physical Model (EKF-AUS)
  Kalman gain

$$\mathbf{K}_n = \Pi_n P_n^f \Pi_n \mathbf{H}^T \left[\mathbf{H} \Pi_n P_n^f \Pi_n \mathbf{H}^T + \mathbf{R} \right]^{-1}, \tag{6}$$

Innovation: $y_n - \mathbf{H} u_n^f$.

- Projected Data Model (Proj-EKF)
  Kalman gain

$$\mathbf{K}_n = P_n^f \Pi_{\mathbf{H}} \Pi_n \left[\Pi_n \mathbf{H}^\dagger \left(\mathbf{H} P_n^f \mathbf{H}^T + \mathbf{R} \right) (\mathbf{H}^\dagger)^T \Pi_n \right]^\dagger, \tag{7}$$

Innovation: $y_n^p - \Pi_n \Pi_{\mathbf{H}} u_n^f$.

Example: Projected Data PF

Basic Particle Filter: Suppose that at time $n-1$ we have the posterior distribution (u_{n-1}^i, w_{n-1}^i), $i = 1, ..., L$, supported on u_{n-1}^i with weights w_{n-1}^i ($w_{n-1}^i \geq 0$ and $\sum_{i=1}^L w_{n-1}^i = 1$).

Prediction. Propagate $u_n^i = F_{n-1}(u_{n-1}^i) + \omega_{n-1}$.

Filtering. Update weights $\{w_{n-1}^i\}_{i=1}^L$ by $w_n^i = c\, w_{n-1}^i p(y_n|u_n^i)$.

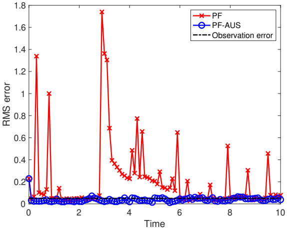
Algorithm 1. [Proj-PF]
(Project data only, and discard the orthogonal component):
Apply a standard DA scheme, e.g., particle filter, using the unprojected forecast model, but replace the standard data model with the projected data y_n^q. The observation operator is replaced by $\mathbf{H}_n^q$, and the covariance matrix of the observations is replaced by $\mathbf{R}_n^q$.

Numerical Results: Linear Problem

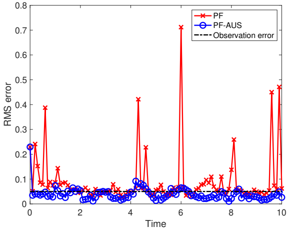
Suppose that forecasts are made with a physical model of the form

$$\frac{du}{dt} = \mathbf{A}u + \sigma \dot{W} \ ,$$

where W is a Wiener process. $\mathbf{A}$ has w eigenvalues with small real part $Re(\lambda_i) \in (0, 0.04)$, the rest have real part $Re(\lambda_i) \leq -100$.

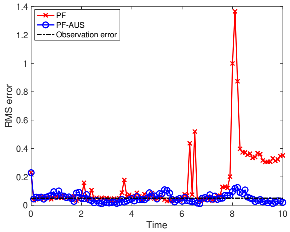


(a) Observations of all 100 variables. Mean RMSE: 0.21 PF, 0.04 Proj-PF.

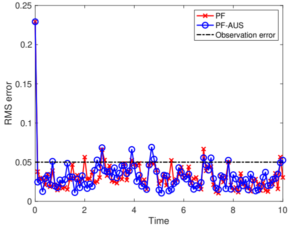


(b) Observations of 50 variables. Mean RMSE: 0.09 PF, 0.04 Proj-PF.

Numerical Results: Linear Problem



(c) Observations of 25 variables. Mean RMSE: 0.15 PF, 0.05 Proj-PF.



(d) Observations of two variables. Mean RMSE: 0.03 PF, 0.03 Proj-PF.

Comparison of PF to Algorithm 1 Proj-PF $(p = 2)$.

Proj-PF performs well independent of number of observed variables.

Numerical Results: Linear Problem

Collect observations every 0.1 time units with 100 observations.

Small measurement error covariance $\mathbf{R} = 0.05^2\mathbf{I}$ and model noise $\sigma = 0.05$.

Both PF algorithms resample if the $ESS := 1/(\sum_{i=1}^{L}(w^i)^2)$ drops below half the number of particles, which is 1000.

On resampling, noise is added to every variable with a standard deviation of 0.02.

Projected Optimal Proposal (Proj-OP-PF)

Algorithm 2. [Proj-OP-PF]

- Employ optimal proposal density $p(u_n^i | u_{n-1}^i, y_n)$ instead of $p(u_n^i | u_{n-1}^i) \sim \mathcal{N}(F_{n-1}(u_{n-1}^i), \mathbf{\Sigma})$.

- Compute the weight update using the projected data model.

We make the following modification to resampling in Proj-OP-PF:

Algorithm 3. [Proj-Resamp]
(Resampling in the Unstable Subspace):

- Generate the usual noise after resampling,
- Multiply this random vector by $\alpha \Pi_n + (1-\alpha) I$, $\alpha \in [0,1]$.

Numerical Results: Projected OP-PF, Lorenz '96

Consider for $i = 1, ..., 40$ and $F = 8$,

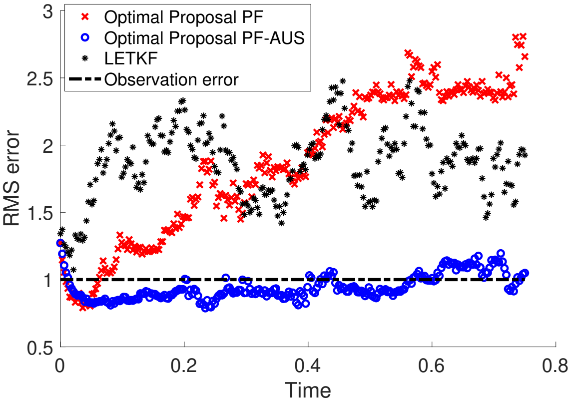
$$\dot{u}_i = (u_{i+1} - u_{i-2})\, u_{i-1} - u_i + F \; . \tag{8}$$

≈ 13 positive Lyapunov exponents, Lyapunov dimension ≈ 28.

We implement Algorithms 2 and 3, Proj-OP-PF/Proj-Resamp, and compare to OP-PF and LETKF.

Localized to 12 variables and with inflation 1.02,
Moderate model noise $\mathbf{\Sigma} = (0.3)^2\mathbf{I}$, large data error $\mathbf{R} = 1\mathbf{I}$,
Tight ensemble spread of S.D. 0.01 with initial RMSE of 1.5.

Parameter regime difficult for the OP-PF.



Compare OP-PF, LETKF, Proj-OP-PF with $p = 2$
Resampling on 30% of steps (99.5% of steps for OP-PF).

Sources of Projections

- AUS type projections,
- Dynamically Orthogonal (DO) projections, see, e.g., the blended particle filter approach of Qi and Majda '15,
- Coherent structure DA, e.g., Maclean, Santitissadeekorn, and Jones '17,
- More generally, projection based dimension reduction techinques for models and data: Reduced Order Modeling (ROM), Inertial manifold, PCA, POD, K-L, ...

In general: Different projections may be utilized for the physical model and for the data model.

What's next ...

Chopping up the unstable subspace into bite sized pieces.

- Working with nested subspaces and filtrations.
- Both parallel and serial techniques.
- In **P** many possible combinations of projections, projected models and data, and algorithmic development.