Exploring the Sparsity of Error Covariance in Data Assimilation

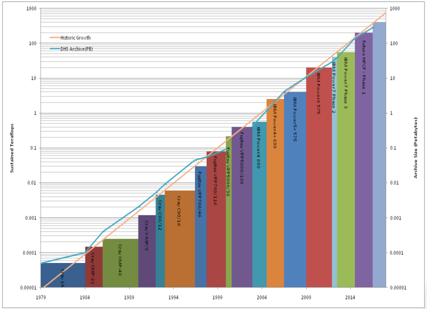
Wei Kang
U.S. Naval Postgraduate School

Liang Xu
U.S. Naval Research Laboratory

3rd AFOSR WORKSHOP ON
COMPUTATIONAL ISSUES IN NONLINEAR CONTROL

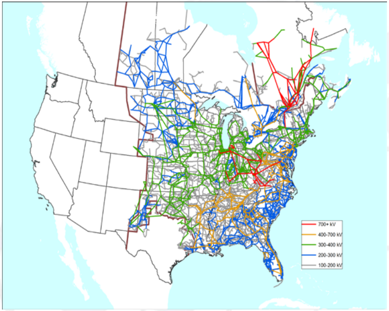
High-Dimensional Systems

Numerical Weather Prediction - ECMWF Global Model



8×10^7 model variables are updated every 6-12 hours using 1.3×10^7 observations

Power System - Eastern Interconnection



A simplified model has more than 25K buses, 28K lines, 8K transformers, $1,000$ generators.

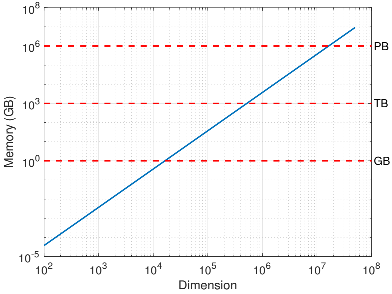
High-Dimensional Systems

Challenges

The scalability of algorithms is limited by several factors: computational load, I/O overhead and required memory size, degree of parallelism, and power consumption.

Covariance Matrix

dimension vs RAM size



Quantitative Change
Becomes Qualitative

Data Assimilation for NWP

Variational methods

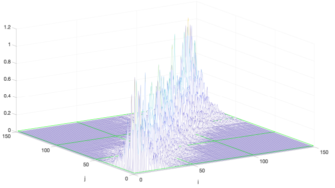
- Motivated by optimal control theory (Sasaki 1958)
- 3D-Var
- 4D-Var

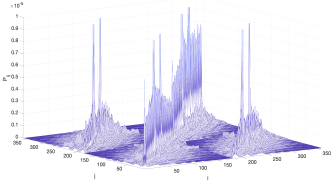
Bayesian estimation methods

- Kalman filter gain based on ensembles
- EnKF
- LETKF

Goal: Characterize the shape of error covariance

- P is approximately **sparse**.
- The decay rate seems to be faster than **exponential**.
- Additional **constraints** should be deduced from observation model





Controllability Gramian and KF

System model

$$\dot{x}(t) = Ax(t) + Bw(t)$$
$$y(t) = Cx(t) + Dv(t)$$

$x \in \mathbb{R}^n$ – state variable

$y \in \mathbb{R}^m$ – observation variable

w, v – zero mean Gaussian white noise with identity covariance

$Q = BB^T$ – covariance of model uncertainty

$R = DD^T$ – covariance of observation noise

The Kalman-Bucy filter

$$\dot{\hat{x}}(t) = A\hat{x}(t) + K(t)(y(t) - C\hat{x}(t))$$
$$K(t) = P(t)C^T R^{-1}$$
$$\dot{P}(t) = AP(t) + P(t)A^T + Q - K(t)RK(t)^T$$

Controllability Gramian and KF

System model

$$\dot{x}(t) = Ax(t) + Bw(t)$$
$$y(t) = Cx(t) + Dv(t)$$

$x \in \mathbb{R}^n$ – state variable

$y \in \mathbb{R}^m$ – observation variable

w, v – zero mean Gaussian white noise with identity covariance

$Q = BB^T$ – covariance of model uncertainty

$R = DD^T$ – covariance of observation noise

The Kalman-Bucy filter

$$\dot{\hat{x}}(t) = A\hat{x}(t) + K(t)(y(t) - C\hat{x}(t))$$
$$K(t) = P(t)C^T R^{-1}$$
$$\dot{P}(t) = AP(t) + P(t)A^T + Q - K(t)RK(t)^T$$

Controllability Gramian and KF

System model

$$\dot{x}(t) = Ax(t) + Bw(t)$$
$$y(t) = Cx(t) + Dv(t)$$

$x \in \mathbb{R}^n$ – state variable

$y \in \mathbb{R}^m$ – observation variable

w, v – zero mean Gaussian white noise with identity covariance

$Q = BB^T$ – covariance of model uncertainty

$R = DD^T$ – covariance of observation noise

The Kalman-Bucy filter

$$\dot{\hat{x}}(t) = A\hat{x}(t) + K(t)Dv(t)$$
$$K(t) = P(t)C^T R^{-1}$$
$$\dot{P}(t) = AP(t) + P(t)A^T + Q - K(t)RK(t)^T$$

Controllability Gramian and KF

Two controllability Gramians

For the model: $\dot{x}(t) = Ax(t) + Bw(t)$

$$\boldsymbol{G^C(t) = \int_0^t e^{A(t-\tau)} Q e^{A^T(t-\tau)} d\tau}$$

For the filter: $\dot{\hat{x}}(t) = A\hat{x}(t) + K(t)Dv(t)$

$$\boldsymbol{\hat{G}^C(t) = \int_0^t e^{A(t-\tau)} K(\tau) R K^T(\tau) e^{A^T(t-\tau)} d\tau}$$

Controllability Gramian and KF

Proposition: In the Kalman-Bucy filter,

$$P(t) = e^{At}P(0)e^{A^T t} + G^C(t) - \hat{G}^C(t)$$

and

$$0 \leq P(t) \leq e^{At}P(0)e^{A^T t} + G^C(t) \quad (\text{upper bound})$$

Remark A more controllable model uncertainty input tends to enlarge error; a more controllable KF results in larger correction, thus smaller error.

Decay Rate

Theorem (A. Iserles 2000). Let $E = e^A$, where $A \in \mathbb{R}^{n\times n}$ is a banded matrix of bandwidth $s \geq 1$. Let

$$\rho = \max_{0\leq i,j\leq n} \{|A_{i,j}|\},$$

then

$$|E_{i,j}| \leq h(\rho, |i-j|/s) \left(e^{|i-j|/s} - \sum_{m=0}^{|i-j|-1} \frac{(|i-j|/s)^m}{m!} \right), \quad |i-j| > L,$$

for some integer $L \leq n$, where

$$h(\alpha, x) = \frac{\alpha^x}{x^x}, \quad \alpha > 0, x > 0.$$

Decay Rate

Theorem. Suppose that A, the matrix of a linear system, is banded with a bandwidth $s \geq 1$. Let

$$\rho = \max_{0 \leq i,j \leq n} \{|A_{i,j}|\}.$$

Suppose Q is also banded. Then

$$|(G^C)_{ij}| \leq Mh(\rho e^2 T, (|j-i| - L)/s), \quad \text{if } |j-i| > 2L$$

for some constant $M > 0$ and some integer $L > 0$.

Remark. $e^{-x^2} < h(\alpha, x) < e^{-|x|}.$

Component-based Computation

Proposition. Consider the dual system

$$\begin{aligned} \dot{z}(\tau) &= -A^T z(\tau), & z &\in \mathbb{R}^n \\ y^z(\tau) &= B^T z(\tau), & y^z &\in \mathbb{R}^n \end{aligned}$$

The controllability Gramian, $G^C(t)$, equals the observability Gramian of the dual system at t. More specifically, if $z_i(\tau)$ and $z_j(\tau)$ are trajectories of the dual system satisfying $z_i(t) = e_i$ and $z_j(t) = e_j$, then

$$(G^C(t))_{ij} = \int_0^t z_i^T(\tau) B B^T z_j(\tau) d\tau,$$

Component-based computation

Decay rate - fitting the curve

$$|(G^C)_{ij}| \leq Mh(\alpha, (|j - i| + \beta)/\gamma)$$

the value of parameters are computed as follows

$$\beta = \frac{\alpha\gamma}{e}$$

$$(M, \alpha, \gamma) = \arg\min_{M,\alpha,\gamma} \sum_{i\in\mathcal{I}, j\in\mathcal{J}} \Big(Mh(\alpha, \alpha/e + (|j - i|)/\gamma) - (G^C)_{ij} \Big)^2$$

A summary - the shape of error covariance

- **Upper bound**

$$P(t) \le e^{At}P(0)e^{A^Tt} + G^C(t)$$

- **Decay rate**
  Suppose A and Q are banded matrices, $h(\alpha, x) = \frac{\alpha^x}{x^x}$, then

$$(G^C)_{ij} \le Mh(\alpha, (|j-i|+\beta)/\gamma), \qquad M, \alpha, \beta, \gamma \text{ are parameters}$$

- **Covariance constrained by observation**

$$\begin{aligned} (P_k^+)_{ii} &\le (R_k)_{ii}, && \text{if } y_k = (x_k)_i \\ |(P_k^+)_{ij}| &< \sqrt{(R_k)_{ii}M_P}, && 1 \le j \le n, \ \ j \ne i \end{aligned}$$

Remark. $e^{-x^2} < h(\alpha, x) < e^{-|x|}$.

The Model

$$A = \begin{bmatrix} -2.4689 & -0.7236 & 0.1301 & -0.2498 & -0.2450 & \cdots & \cdots \\ 0.7227 & -2.2126 & 0.7261 & 0.6476 & 0.6578 & \cdots & \cdots \\ 0.4485 & -0.0722 & -2.7618 & -0.1565 & -0.7478 & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}_{150\times 150} \quad \begin{pmatrix} \text{random value} \\ \text{banded matrix} \\ \text{bandwidth} = 8 \end{pmatrix}$$

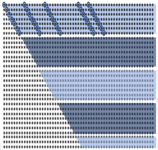
$$B = \begin{bmatrix} 1.00 & 1.00 & 0.50 & 0.33 & 0.25 & 0 & 0 & 0 & \cdots & \cdots \\ 1.00 & 1.00 & 1.00 & 0.50 & 0.33 & 0.25 & 0 & 0 & \cdots & \cdots \\ 0.50 & 1.00 & 1.00 & 1.00 & 0.50 & 0.33 & 0.25 & 0 & \cdots & \cdots \\ 0.33 & 0.50 & 1.00 & 1.00 & 1.00 & 0.50 & 0.33 & 0.25 & \cdots & \cdots \\ 0.25 & 0.33 & 0.50 & 1.00 & 1.00 & 1.00 & 0.50 & 0.33 & \cdots & \cdots \\ 0 & 0.25 & 0.33 & 0.50 & 1.00 & 1.00 & 1.00 & 0.50 & \cdots & \cdots \\ 0 & 0 & 0.25 & 0.33 & 0.50 & 1.00 & 1.00 & 1.00 & \cdots & \cdots \\ 0 & 0 & 0 & 0.25 & 0.33 & 0.50 & 1.00 & 1.00 & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

$$y = \begin{bmatrix} x_1 & x_{50} & x_{100} & x_{150} \end{bmatrix} + Dv(t), \quad D = \begin{bmatrix} 0.0102 & 0.0098 & 0.0044 & 0.0110 \\ -0.0083 & -0.0088 & 0.0028 & -0.0014 \\ -0.0097 & 0.0034 & -0.0103 & -0.0035 \\ 0.0047 & -0.0080 & 0.0067 & -0.0102 \end{bmatrix}$$

Example

Covariance constrained by observation

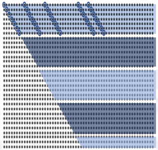
$(P_k^+)_{ii} \leq (R_k)_{ii}, \;\; |(P_k^+)_{ij}| < \sqrt{(R_k)_{ii} M_P}$

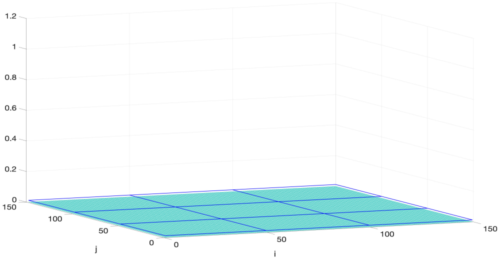


Example

Covariance constrained by observation

$$(P_k^+)_{ii} \leq (R_k)_{ii}, \quad |(P_k^+)_{ij}| < \sqrt{(R_k)_{ii} M_P}$$

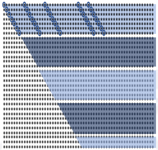


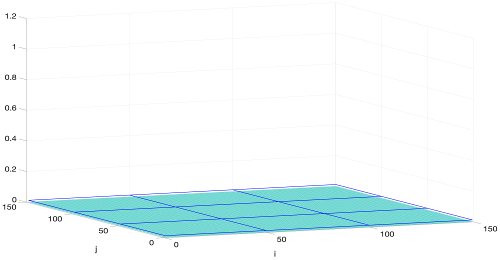


Example

G^C - **diagonal**

$P(t) \leq e^{At}P(0)e^{A^Tt} + G^C(t)$

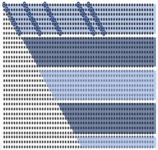


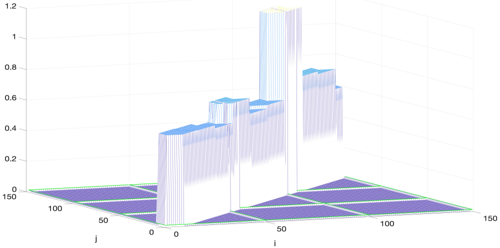


Example

G^C - **diagonal**

$P(t) \leq e^{At}P(0)e^{A^Tt} + G^C(t)$





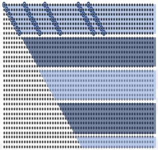
Example

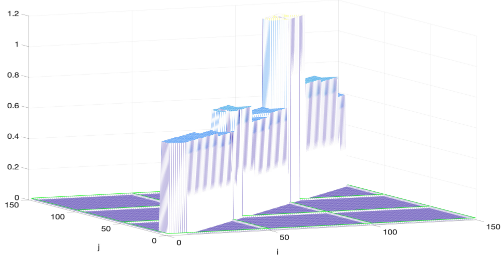
Decay rate

$$(G^C)_{ij} \leq Mh(\alpha, (|j-i| + \beta)/\gamma)$$

$$\beta = \frac{\alpha\gamma}{e}$$

$$(M, \alpha, \gamma) = \arg\min_{M,\alpha,\gamma} \sum_{i\in\mathcal{I}, j\in\mathcal{J}} \left(Mh(\alpha, \alpha/e + (|j-i|)/\gamma) - (G^C)_{ij} \right)^2$$



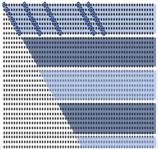


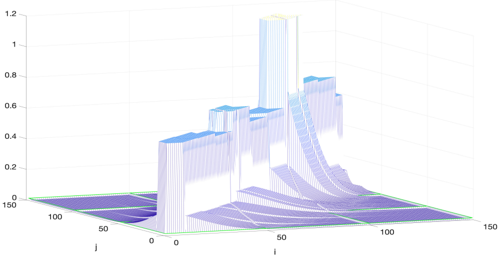
Decay rate

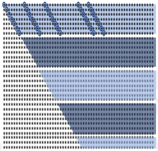
$$(G^C)_{ij} \le Mh(\alpha, (|j-i|+\beta)/\gamma)$$

$$\beta = \frac{\alpha\gamma}{e}$$

$$(M,\alpha,\gamma) = \arg\min_{M,\alpha,\gamma} \sum_{i\in\mathcal{I}, j\in\mathcal{J}} \left(Mh(\alpha, \alpha/e + (|j-i|)/\gamma) - (G^C)_{ij}\right)^2$$





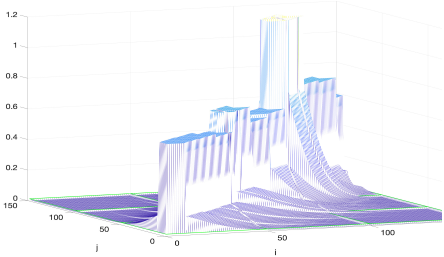


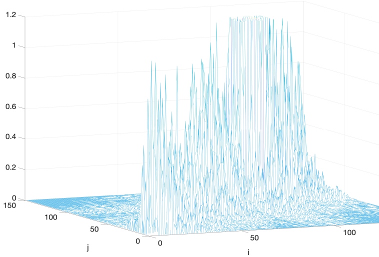
Decay rate

$$(G^C)_{ij} \leq Mh(\alpha, (|j-i| + \beta)/\gamma)$$

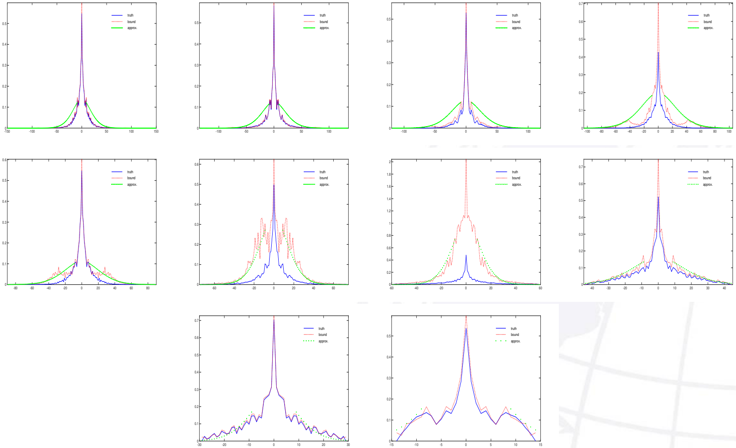
$$\beta = \frac{\alpha\gamma}{e}$$

$$(M, \alpha, \gamma) = \arg \min_{M,\alpha,\gamma} \sum_{i\in\mathcal{I}, j\in\mathcal{J}} \Big(Mh(\alpha, \alpha/e + (|j-i|)/\gamma) - (G^C)_{ij}\Big)^2$$





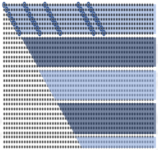
The shape of covariance matrix

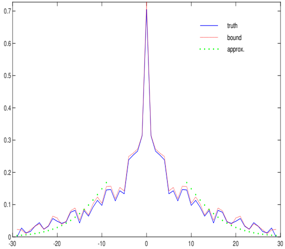


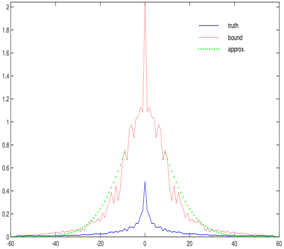
NAVAL POSTGRADUATE SCHOOL

Example

The shape of covariance matrix



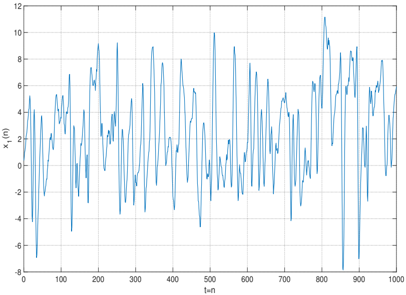




Lorenz-96 Model

$$\begin{aligned} \frac{dx_i}{dt} &= (x_{i+1} - x_{i-2})x_{i-1} - x_i + F + Bw, \quad i = 1, 2, \cdots, n, \\ y &= \begin{bmatrix} x_1 & x_3 & x_5 & \cdots \end{bmatrix}^T + Dv \end{aligned}$$

- $n = 40, \quad \Delta t = 0.05, \quad F = 8.$
- $Q = BB^T$ is banded $s = 2$, $R = DD^T$ is diagonal.
- $T = 1000$ time steps.

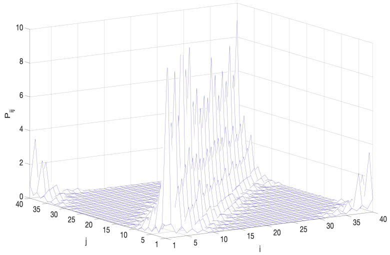


More Examples

Lorenz-96 Model

- Error covariance is computed using UKF .
- 10^5 random initial states uniformly distributed in $x_i \in [-1, 1]$.
- 6×10^4 numerical experimentations.

The maximum value of each entry in P

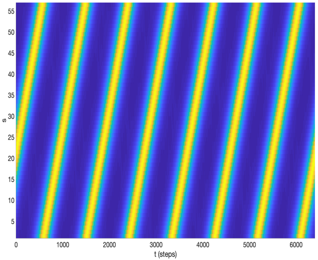


More Examples

The KdV equation

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial s} + \frac{\partial^3 u}{\partial s^3} = 0$$

- ODE model: n=57, 2^{nd}-order finite difference approximation.
- $u|_{t=0} = 3A\,\mathrm{sech}^2(\sqrt{A}(s - s_0)/2)$, $s_0 = 5$, $A = 1$.
- $dt = 0.0156$, $ds = 0.25$, $T = 6400$ filter steps.

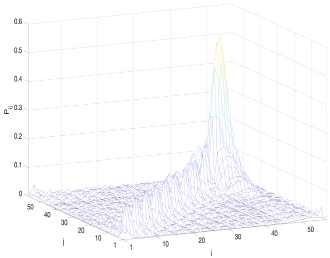


More Examples

The KdV equation

- $n_y = 10$ observation locations.
- $Q = BB^T$ is banded $s = 1$, $R = DD^T$ is diagonal.
- Gaussian model uncertainty and observation noise.
- $3,750$ numerical experimentations.

The maximum value of each entry in P

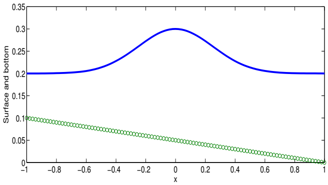


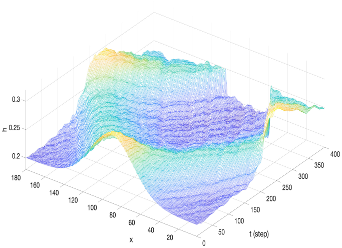
More Examples

The shallow Water Equation

$$\frac{\partial h}{\partial t} + \frac{\partial uh}{\partial x} = 0$$
$$\frac{\partial (uh)}{\partial t} + \frac{\partial (u^2 h + 0.5gh^2)}{\partial x} = 0$$

- Discretization: DG and 3rd-order RK, n=180.
- $h|_{t=0} = 0.1\exp(-8(x-0.5)) + 0.2$, $u|_{t=0} = 0$.
- $dt = 0.00125$, $T = 400$ filter steps.



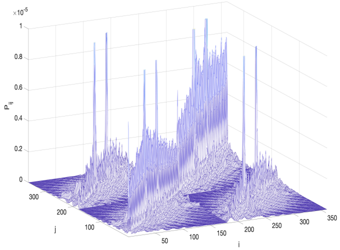


More Examples

The shallow water equation

- $n_y = 18$ observation locations.
- $Q = BB^T$ is banded $s = 2$, $R = DD^T$ is diagonal.
- Gaussian model uncertainty and observation noise.
- 400 numerical experimentations.

The maximum value of each entry in P



Conclusions and future work

- Sparsity pattern is characterized for linear systems using controllability Gramian.

- Error covariance is bounded by controllability Gramians.
- If A is banded, the decay rate of $G^C(t)$ is $h(\alpha, x)$, faster than exponential.
- Covariance satisfies constraints deduced from observation model.
- The computation is component-based using duality and curve-fitting.
- For nonlinear systems, numerical experimentations show almost sparse error covariance.
- Future work: the difference between $G^C(t)$ and $\hat{G}^C(t)$, nonlinear systems, parallel and distributed computation, NWP applications, ...