Data Assimilation
- A09. Innovation Statistics -

Shunji Kotsuki

Environmental Prediction Science Laboratory
Center for Environmental Remote Sensing (CEReS), Chiba University
(shunji.kotsuki@chiba-u.jp)

Environmental
Prediction
Science
Laboratory

DA Lectures A (Basic Course)

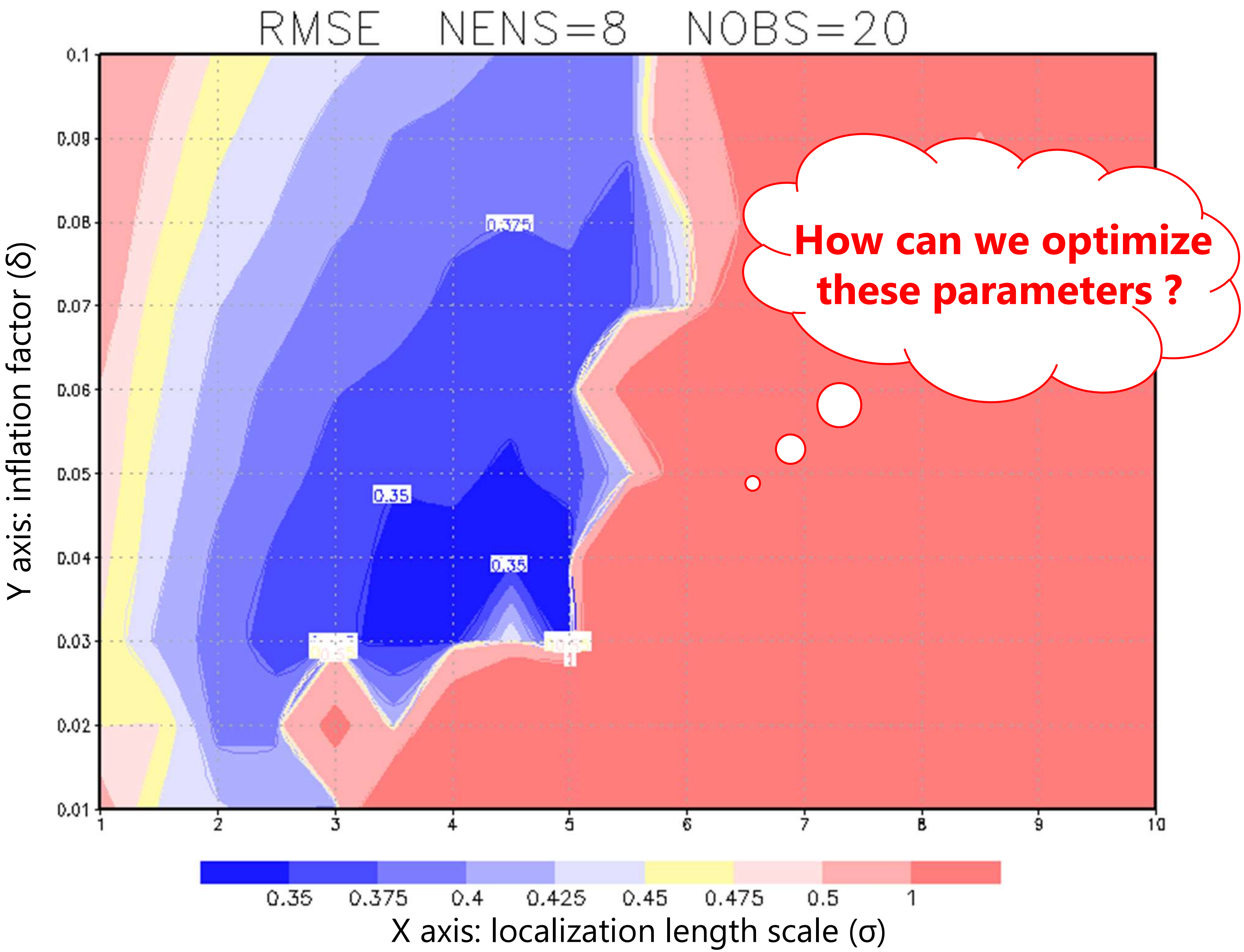
Environmental Prediction Science Laboratory

- (1) Introduction and NWP
- (2) Deterministic Chaos and Lorenz-96 model
- (3) A toy model and Bayesian estimation
- (4) Kalman Filter (KF)
- (5) 3D Variational Method (3DVAR)
- (6) Ensemble Kalman Filter (PO method)
- (7) Serial Ens. Square Root Filter (Serial EnSRF)
- (8) Local Ens. Transform Kalman Filter (LETKF)
- (9) Innovation Statistics
- (10) Adaptive Inflations
- (11) 4D Variational Method (4DVAR)

Today's Goal

- **Lecture: innovation statistics**
  - to introduce innovation statistics
  - to understand adaptive inflation
- **Training: Lorenz 96**
  - to implement adaptive inflation into L96

Motivation



RMSE NENS=8 NOBS=20
Y axis: inflation factor (δ)
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
1
2
3
4
5
6
7
8
9
10
0.375
0.35
0.35
0.35
0.375
0.4
0.425
0.45
0.475
0.5
1
X axis: localization length scale (σ)
How can we optimize these parameters ?

Innovation statistics

$$\mathbf{d}^{o-b} = \mathbf{y}^o - \mathbf{H}\mathbf{x}^b$$

b: background

$$\mathbf{d}^{o-a} = \mathbf{y}^o - \mathbf{H}\mathbf{x}^a$$

a: analysis

$$\mathbf{d}^{a-b} = \mathbf{H}\mathbf{x}^a - \mathbf{H}\mathbf{x}^b$$

o: observation

Desroziers' innovation statistics (Desroziers et al. 2005)

$$\left\langle \mathbf{d}^{o-b}(\mathbf{d}^{o-b})^T \right\rangle = \mathbf{HBH}^T + \mathbf{R}$$

$$\left\langle \mathbf{d}^{a-b}(\mathbf{d}^{o-b})^T \right\rangle \approx \mathbf{HBH}^T$$

$$\left\langle \mathbf{d}^{o-a}(\mathbf{d}^{o-b})^T \right\rangle \approx \mathbf{R}$$

$$\left\langle \mathbf{d}^{a-b}(\mathbf{d}^{o-a})^T \right\rangle \approx \mathbf{HAH}^T$$

Innovation Statistics (Derivations)

Innovation Statistics

Definition

$\langle \bullet \rangle$: Statistical Expectations

$$\boldsymbol{\varepsilon}^o = \mathbf{y}^o - \mathbf{y}^t, \qquad \boldsymbol{\varepsilon}^b = \mathbf{x}^b - \mathbf{x}^t, \qquad \boldsymbol{\varepsilon}^a = \mathbf{x}^a - \mathbf{x}^t$$

$$\mathbf{R} = \langle \boldsymbol{\varepsilon}^o (\boldsymbol{\varepsilon}^o)^T \rangle, \qquad \mathbf{B} = \mathbf{P}^b = \left\langle \boldsymbol{\varepsilon}^b (\boldsymbol{\varepsilon}^b)^T \right\rangle, \qquad \mathbf{A} = \mathbf{P}^a = \langle \boldsymbol{\varepsilon}^a (\boldsymbol{\varepsilon}^a)^T \rangle$$

Derivation from Kalman Gain

$\widetilde{\mathbf{B}}, \widetilde{\mathbf{R}}, \widetilde{\mathbf{A}}, \widetilde{\mathbf{K}}$ matrices used in DA (usually imperfect)

$$\mathbf{x}^a = \mathbf{x}^b + \widetilde{\mathbf{K}}\mathbf{d}^{o-b}, \qquad \widetilde{\mathbf{A}} = \left[\mathbf{I} - \widetilde{\mathbf{K}}\mathbf{H}\right]\widetilde{\mathbf{B}}$$

$$\mathbf{K} = \widetilde{\mathbf{B}}\mathbf{H}^T\left[\mathbf{H}\widetilde{\mathbf{B}}\mathbf{H}^T + \widetilde{\mathbf{R}}\right]^{-1} = \widetilde{\mathbf{A}}\mathbf{H}^T\widetilde{\mathbf{R}}^{-1}$$

$$\left\langle \mathbf{d}^{o-b} (\mathbf{d}^{o-b})^T \right\rangle = \mathbf{HBH}^T + \mathbf{R} \qquad \left\langle \mathbf{d}^{a-b} (\mathbf{d}^{o-b})^T \right\rangle \approx \mathbf{HBH}^T$$

$$\left\langle \mathbf{d}^{o-a} (\mathbf{d}^{o-b})^T \right\rangle \approx \mathbf{R} \qquad \left\langle \mathbf{d}^{a-b} (\mathbf{d}^{o-a})^T \right\rangle \approx \mathbf{HAH}^T$$

Derivations

$$\left\langle \mathbf{H}\boldsymbol{\varepsilon}^b (\boldsymbol{\varepsilon}^o)^T \right\rangle = \left\langle (\boldsymbol{\varepsilon}^o)^T \mathbf{H}\boldsymbol{\varepsilon}^b \right\rangle = 0$$

Environmental Prediction Science Laboratory

$$\left\langle \mathbf{d}^{o-b} (\mathbf{d}^{o-b})^T \right\rangle = \mathbf{HBH}^T + \mathbf{R}$$

$$\mathbf{d}^{o-b} = \mathbf{y}^o - \mathbf{H}\mathbf{x}^b = \boldsymbol{\varepsilon}^o - \mathbf{H}\boldsymbol{\varepsilon}^b$$

$$\left\langle \mathbf{d}^{o-b} (\mathbf{d}^{o-b})^T \right\rangle = \left\langle \left(\boldsymbol{\varepsilon}^o - \mathbf{H}\boldsymbol{\varepsilon}^b\right)\left(\boldsymbol{\varepsilon}^o - \mathbf{H}\boldsymbol{\varepsilon}^b\right)^T \right\rangle = \mathbf{R} + \mathbf{HBH}^T$$

$$\left\langle \mathbf{d}^{a-b} (\mathbf{d}^{o-b})^T \right\rangle \approx \mathbf{HBH}^T$$

$$\mathbf{d}^{a-b} = \mathbf{H}\mathbf{x}^a - \mathbf{H}\mathbf{x}^b = \mathbf{H}\tilde{\mathbf{K}}\mathbf{d}^{o-b}$$

$$\begin{aligned}\left\langle \mathbf{d}^{a-b} (\mathbf{d}^{o-b})^T \right\rangle &= \left\langle \mathbf{H}\tilde{\mathbf{K}}(\mathbf{d}^{o-b})\left(\mathbf{d}^{o-b}\right)^T \right\rangle = \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T\left(\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}}\right)^{-1}(\mathbf{HBH}^T + \mathbf{R}) \\ &= \mathbf{HB}^e\mathbf{H}^T \approx \mathbf{HBH}^T\end{aligned}$$

$$\left\langle \mathbf{d}^{o-a} (\mathbf{d}^{o-b})^T \right\rangle \approx \mathbf{R}$$

$$\begin{aligned}\mathbf{d}^{o-a} &= \mathbf{d}^{o-b} + \mathbf{d}^{b-a} \\ &= \mathbf{d}^{o-b} - \mathbf{H}\tilde{\mathbf{K}}\mathbf{d}^{o-b} \\ &= \left(\mathbf{I} - \mathbf{H}\tilde{\mathbf{K}}\right)\mathbf{d}^{o-b} \\ &= \tilde{\mathbf{R}}\left(\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}}\right)^{-1}\mathbf{d}^{o-b}\end{aligned}$$

$$\begin{aligned}&\left\langle \mathbf{d}^{o-a} (\mathbf{d}^{o-b})^T \right\rangle \\ &= \tilde{\mathbf{R}}\left(\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}}\right)^{-1}\left\langle \mathbf{d}^{o-b}\left(\mathbf{d}^{o-b}\right)^T \right\rangle \\ &= \tilde{\mathbf{R}}\left(\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}}\right)^{-1}(\mathbf{HBH}^T + \mathbf{R}) \\ &= \mathbf{R}^e \approx \mathbf{R}\end{aligned}$$

$$\mathbf{R}^e + \mathbf{HB}^e\mathbf{H}^T = \tilde{\mathbf{R}}\left(\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}}\right)^{-1}(\mathbf{HBH}^T + \mathbf{R}) + \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T\left(\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}}\right)^{-1}(\mathbf{HBH}^T + \mathbf{R}) = \mathbf{R} + \mathbf{HBH}^T$$

Derivations (cont'd)

$$\left\langle \mathbf{d}^{a-b}(\mathbf{d}^{o-a})^T \right\rangle \approx \mathbf{HAH}^T$$

$$\mathbf{d}^{o-a} = \widetilde{\mathbf{R}}\left(\mathbf{H}\widetilde{\mathbf{B}}\mathbf{H}^T + \widetilde{\mathbf{R}}\right)^{-1}\mathbf{d}^{o-b}$$

$$\begin{aligned}
\left\langle \mathbf{d}^{a-b}(\mathbf{d}^{o-a})^T \right\rangle &= \left\langle \mathbf{H}\widetilde{\mathbf{K}}\mathbf{d}^{o-b}\left(\widetilde{\mathbf{R}}\left(\mathbf{H}\widetilde{\mathbf{B}}\mathbf{H}^T + \widetilde{\mathbf{R}}\right)^{-1}\mathbf{d}^{o-b}\right)^T \right\rangle \\
&= \mathbf{H}\widetilde{\mathbf{K}}(\mathbf{HBH}^T + \mathbf{R})\left(\mathbf{H}\widetilde{\mathbf{B}}\mathbf{H}^T + \widetilde{\mathbf{R}}\right)^{-1}\widetilde{\mathbf{R}} \\
&= \mathbf{H}\widetilde{\mathbf{A}}\mathbf{H}^T\widetilde{\mathbf{R}}^{-1}(\mathbf{HBH}^T + \mathbf{R})^{-1}\left(\mathbf{H}\widetilde{\mathbf{B}}\mathbf{H}^T + \widetilde{\mathbf{R}}\right)^{-1}\widetilde{\mathbf{R}} \\
&= \mathbf{HA}^e\mathbf{H}^T \approx \mathbf{HAH}^T
\end{aligned}$$

Innovation Statistics

Innovation Statistics

$$\langle \mathbf{d}^{o-b}(\mathbf{d}^{o-b})^T \rangle = \mathbf{HBH}^T + \mathbf{R}$$

$$\langle \mathbf{d}^{o-a}(\mathbf{d}^{o-b})^T \rangle = \tilde{\mathbf{R}}\left(\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}}\right)^{-1}(\mathbf{HBH}^T + \mathbf{R}) = \mathbf{R}^e$$

$$\langle \mathbf{d}^{a-b}(\mathbf{d}^{o-b})^T \rangle = \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T\left(\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}}\right)^{-1}(\mathbf{HBH}^T + \mathbf{R}) = \mathbf{HB}^e\mathbf{H}^T$$

$$\langle \mathbf{d}^{a-b}(\mathbf{d}^{o-a})^T \rangle = \mathbf{HB}^e\mathbf{H}^T\left(\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T\right)^{-1}\mathbf{H}\tilde{\mathbf{A}}\mathbf{H}^T = \mathbf{HA}^e\mathbf{H}^T$$

Their Relationship

$$\mathbf{R}^e + \mathbf{HB}^e\mathbf{H}^T = \mathbf{R} + \mathbf{HBH}^T$$

NOTE: We need to consider DA uses imperfect B & R in the real-world applications

$\mathbf{R}, \mathbf{B}, \mathbf{A}$: Truth

$\tilde{\mathbf{R}}, \tilde{\mathbf{B}}, \tilde{\mathbf{A}}$: DA (maybe imperfect)

$\mathbf{R}^e, \mathbf{B}^e, \mathbf{A}^e$: Estimated

o, R: obs

b, B: background

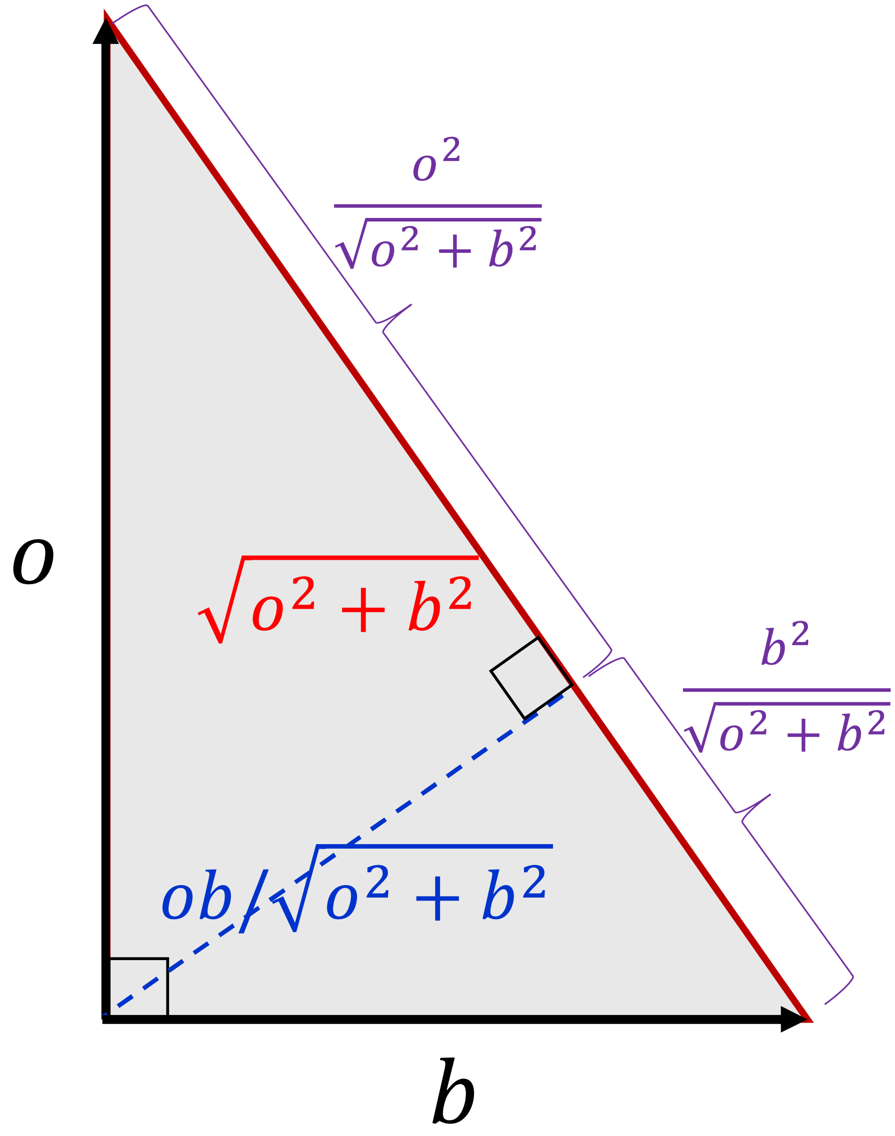
a, A: analysis

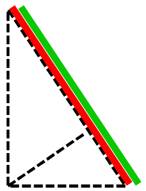
Geometric Interpretations of Data Assimilation

For simplicity, following discussion assumes a scalar problem

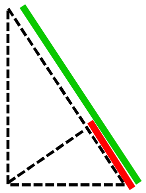
Pythagorean theorem

Environmental Prediction Science Laboratory

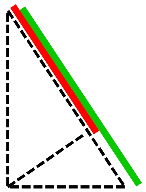




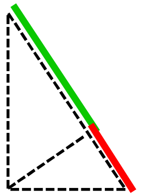
$$\sqrt{o^2+b^2}\sqrt{o^2+b^2} = o^2+b^2$$



$$\frac{b^2}{\sqrt{o^2+b^2}}\sqrt{o^2+b^2} = b^2$$

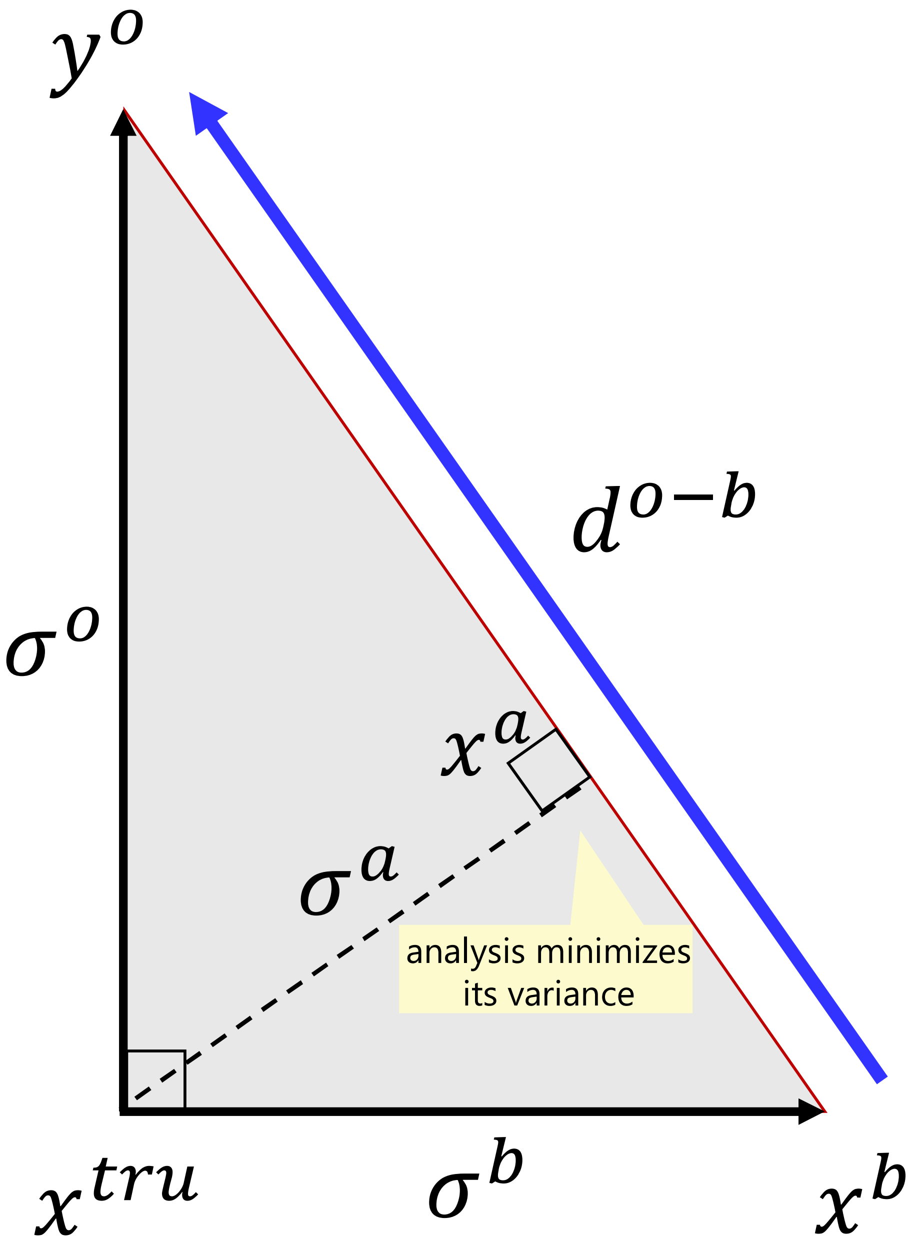


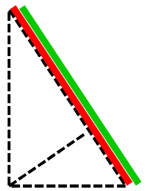
$$\frac{a^2}{\sqrt{o^2+b^2}}\sqrt{o^2+b^2} = o^2$$



$$\frac{b^2}{\sqrt{o^2+b^2}}\frac{o^2}{\sqrt{o^2+b^2}} = \frac{o^2 b^2}{o^2+b^2}$$

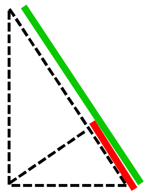
Innovation Statistics





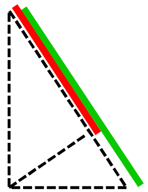
(1) OMB-OMB Statistics $\mathbf{HBH}^T + \mathbf{R}$

$$\langle d^{o-b} d^{o-b} \rangle = (\sigma^b)^2 + (\sigma^o)^2$$



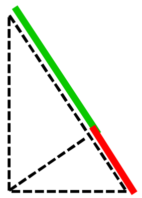
(2) AMB-OMB Statistics $\mathbf{HBH}^T$

$$\langle d^{a-b} d^{o-b} \rangle = (\sigma^b)^2$$



(3) OMA-OMB Statistics $\mathbf{R}$

$$\langle d^{o-a} d^{o-b} \rangle = (\sigma^o)^2$$

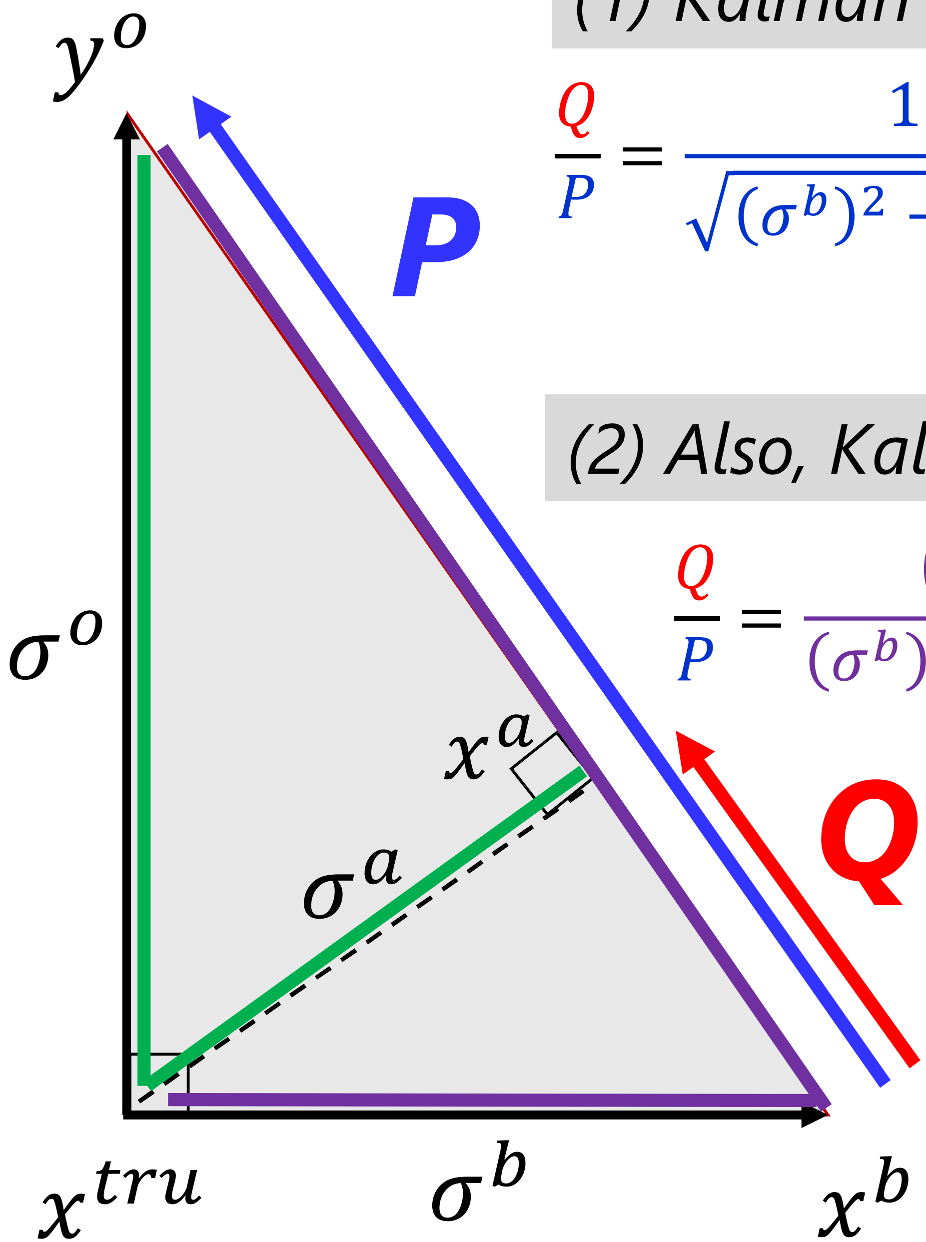


(4) AMB-OMA Statistics $\mathbf{HAH}^T$

$$\langle d^{a-b} d^{o-a} \rangle = (\sigma^a)^2$$

cross terms correspond to right angle: $\langle \mathbf{H}\boldsymbol{\varepsilon}^b (\boldsymbol{\varepsilon}^o)^T \rangle = \langle (\boldsymbol{\varepsilon}^o)^T \mathbf{H}\boldsymbol{\varepsilon}^b \rangle = 0$

Kalman Gain



(1) Kalman gain gives ***weights***

$$\frac{Q}{P} = \frac{1}{\sqrt{(\sigma^b)^2 + (\sigma^o)^2}} \frac{(\sigma^b)^2}{\sqrt{(\sigma^b)^2 + (\sigma^o)^2}} = \frac{(\sigma^b)^2}{(\sigma^b)^2 + (\sigma^o)^2}$$

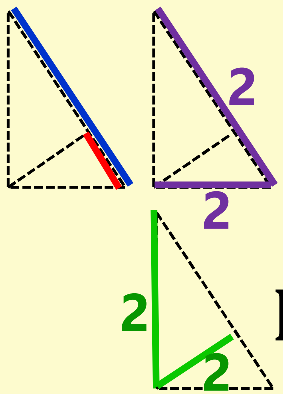
i.e., $\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}$

(2) Also, Kalman gain is

$$\frac{Q}{P} = \frac{(\sigma^b)^2}{(\sigma^b)^2 + (\sigma^o)^2} = \frac{(\sigma^a)^2}{(\sigma^o)^2}$$

i.e., $\mathbf{K} = \mathbf{A}\mathbf{H}^T\mathbf{R}^{-1}$

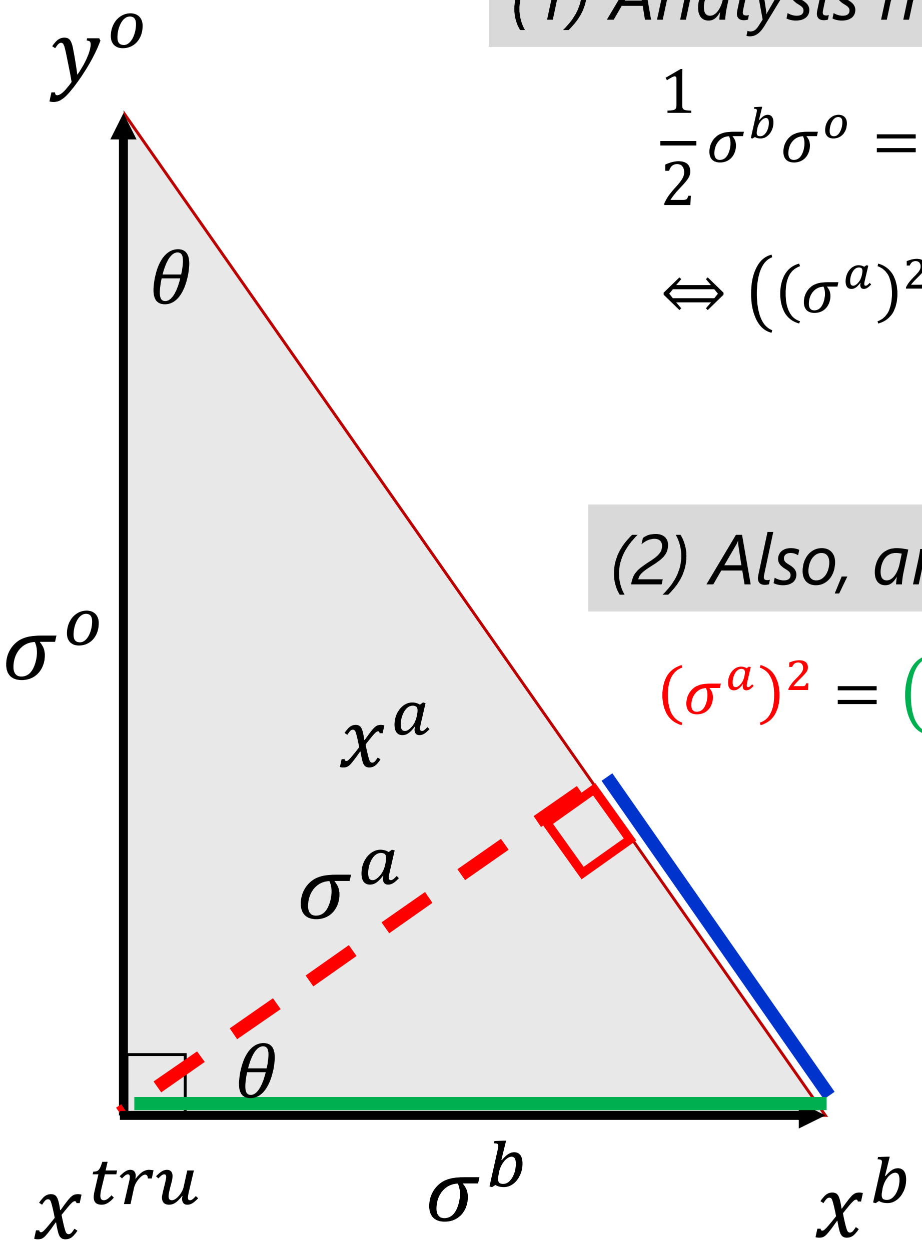
Namely, Kalman gain is uncertainty of posterior error cov. w.r.t. obs. error cov.



$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}$

$\mathbf{K} = \mathbf{A}\mathbf{H}^T\mathbf{R}^{-1}$

Analysis Error Covariance



(1) Analysis minimizes variance of estimation

$$\frac{1}{2}\sigma^b\sigma^o = \frac{1}{2}\sigma^a\sqrt{(\sigma^b)^2 + (\sigma^o)^2}$$

$$\Leftrightarrow \left((\sigma^a)^2\right)^{-1} = \left((\sigma^b)^2\right)^{-1} + \left((\sigma^o)^2\right)^{-1}$$

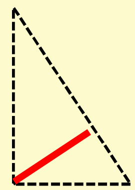
➔ vertical

i.e., $\mathbf{A}^{-1} = \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$

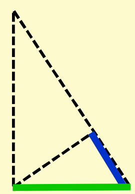
(2) Also, analysis is

$$(\sigma^a)^2 = \left(\sigma^b\right)^2 - \sin^2\theta\left(\sigma^b\right)^2$$

i.e., $\mathbf{A} = (\mathbf{I} - \mathbf{KH})\mathbf{B}$

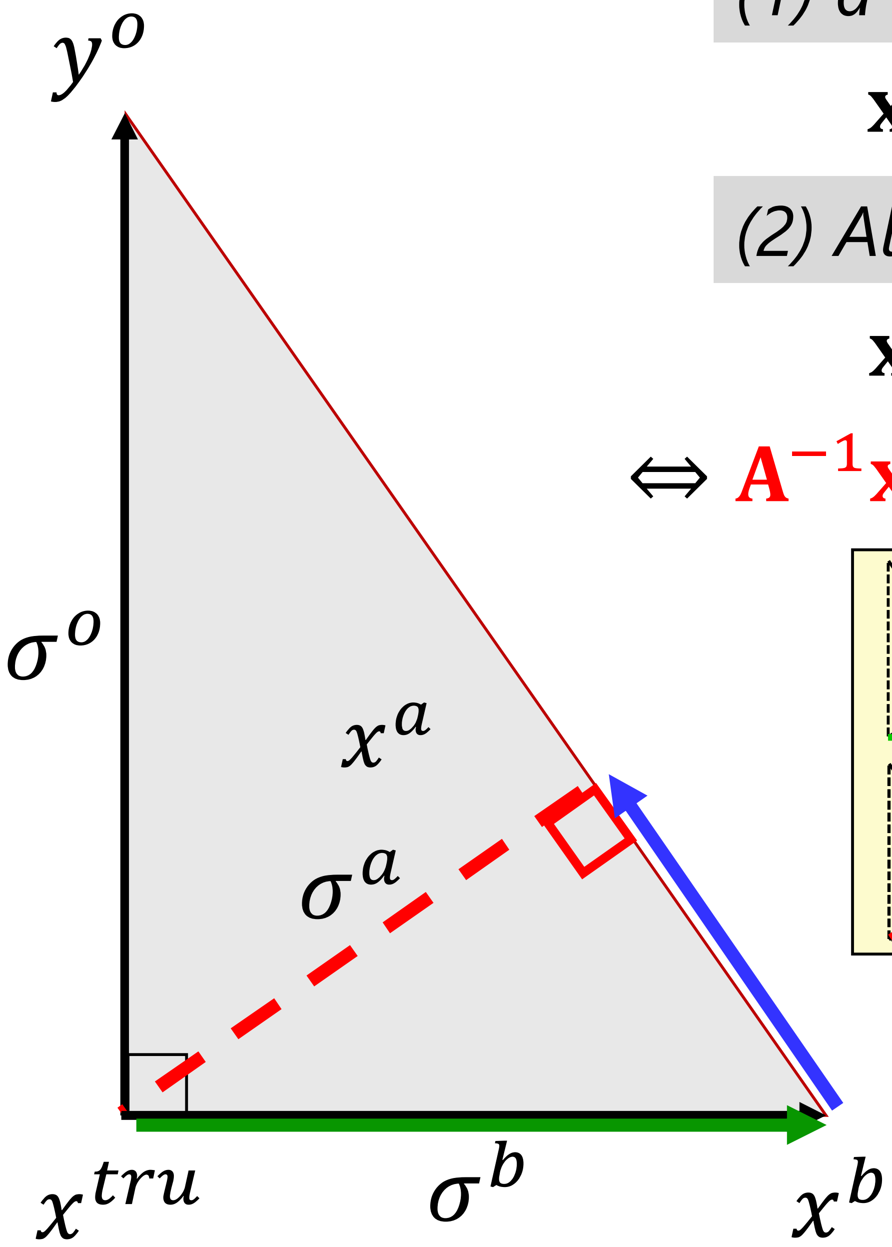


$$\mathbf{A}^{-1} = \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$$



$$\mathbf{A} = (\mathbf{I} - \mathbf{KH})\mathbf{B}$$

Analysis State



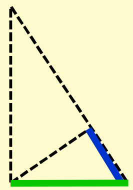
(1) a standard update equation

$$\mathbf{x}_t^a = \mathbf{x}_t^b + \mathbf{K}_t \mathbf{d}_t^{o-b}$$

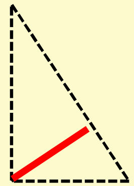
(2) Also, analysis is given by

$$\mathbf{x}_t^a = \mathbf{A}\left[\mathbf{B}^{-1}\mathbf{x}_t^b + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}_t^o\right]$$

$$\Leftrightarrow \mathbf{A}^{-1}\mathbf{x}_t^a = \mathbf{B}^{-1}\mathbf{x}_t^b + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}_t^o$$

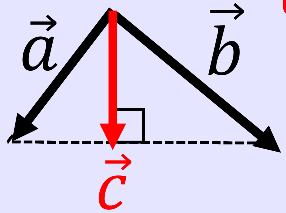


$$\mathbf{x}_t^a = \mathbf{x}_t^b + \mathbf{K}_t \mathbf{d}_t^{o-b}$$



$$\mathbf{A}^{-1}\mathbf{x}_t^a = \mathbf{B}^{-1}\mathbf{x}_t^b + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}_t^o$$

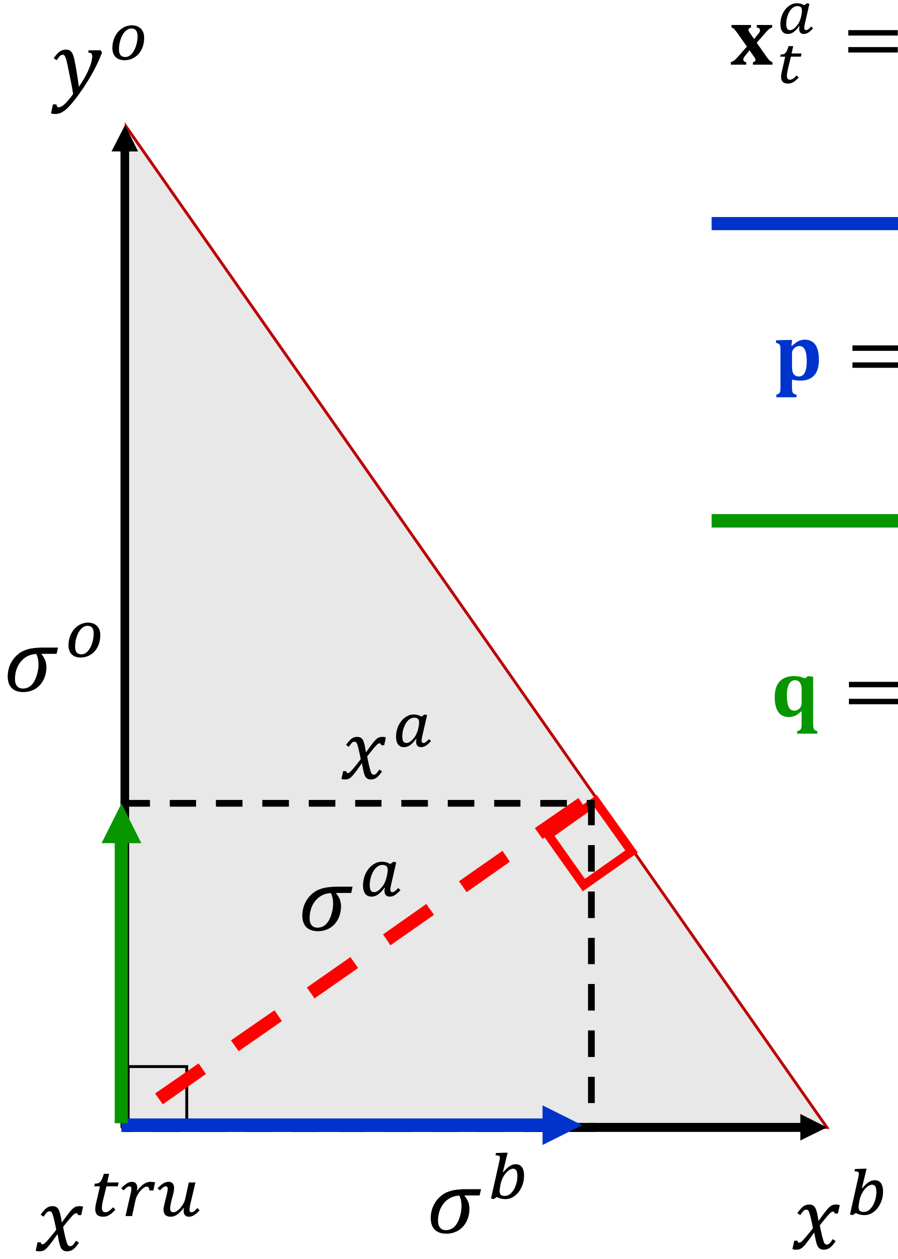
Review



$$\vec{c} = \frac{\beta^2\vec{a}+\alpha^2\vec{b}}{\alpha^2+\beta^2} \Leftrightarrow \frac{\vec{c}}{\gamma^2} = \frac{\vec{a}}{\alpha^2} + \frac{\vec{b}}{\beta^2}$$

$$|\vec{a}| = \alpha,\ |\vec{b}| = \beta,\ |\vec{c}| = \gamma = \frac{\alpha\beta}{\sqrt{\alpha^2+\beta^2}}$$

Analysis State (cont'd)



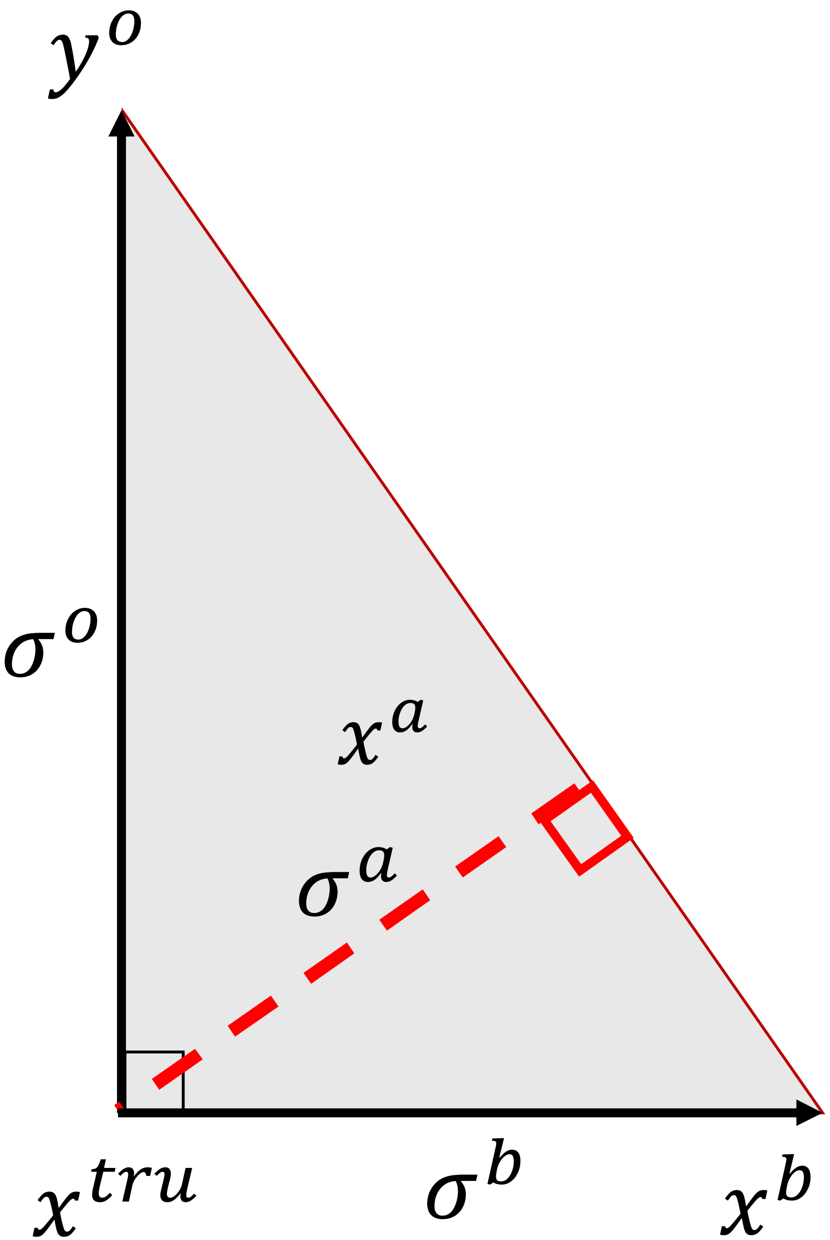
$$\mathbf{x}_t^a = \mathbf{A}\left[\mathbf{B}^{-1}\mathbf{x}_t^b + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}_t^o\right]$$

$$\mathbf{p} = \mathbf{x}_t^b\left(\frac{(\sigma^o)^2}{(\sigma^b)^2 + (\sigma^o)^2}\right)$$

$$\mathbf{q} = \mathbf{y}_t^o\left(\frac{(\sigma^b)^2}{(\sigma^b)^2 + (\sigma^o)^2}\right)$$

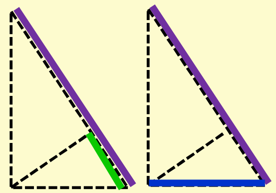
Geometric Interpretations

Environmental Prediction Science Laboratory



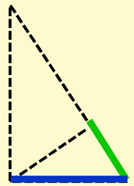
Focusing on the weights gives:

$$\mathbf{K}_t = \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}$$



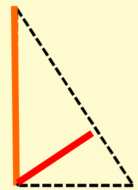
$$\mathbf{A} = \mathbf{B} - \mathbf{K}\mathbf{H}\mathbf{B}$$

$$\mathbf{x}_t^a = \mathbf{x}_t^b + \mathbf{K}_t\mathbf{d}_t^{o-b}$$



Focusing on the AN cov. gives:

$$\mathbf{K} = \mathbf{A}\mathbf{H}^T\mathbf{R}^{-1}$$



$$\mathbf{A}^{-1}\mathbf{x}_t^a = \mathbf{B}^{-1}\mathbf{x}_t^b + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}_t^o$$

To understand DA via two ways

(1) Kalman gain gives the optimal weight

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}$$

From the view point of <u>analysis states</u>

(2) Kalman gain is the analysis error cov. normalized by R

$$\mathbf{K} = \mathbf{A}\mathbf{H}^T\mathbf{R}^{-1}$$

From the view point of <u>analysis uncertainty</u>

To discuss information (or entropy), (2) is more important

For example, DFS (Degrees of Freedom for the Signal) is given by:

$$DFS = \frac{\partial \mathbf{y}^a}{\partial \mathbf{y}^o} = \frac{\partial}{\partial \mathbf{y}^o}(\mathbf{H}\mathbf{x}^a) = \frac{\partial}{\partial \mathbf{y}^o}\left(\mathbf{H}\mathbf{x}^b + \mathbf{H}\mathbf{K}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b\right)\right) = \mathbf{H}\mathbf{K}$$

$$= \mathbf{H}\mathbf{K} = \mathbf{H}\mathbf{A}\mathbf{H}^T\mathbf{R}^{-1}$$

i.e., DFS is obs-space **A** normalized by **R**

DFS=0 : no information from obs, DFS=1 : all information from obs

Thank you for your attention!

Presented by Shunji Kotsuki
(shunji.kotsuki@chiba-u.jp)

Further information is available at
https://kotsuki-lab.com/

A World Beyond Predictions
Environmental
Prediction
Science
Laboratory