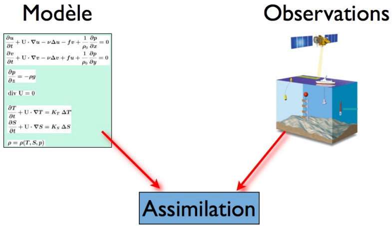
An introduction to data assimilation

Eric Blayo
University of Grenoble and INRIA

Data assimilation, the science of compromises

Context characterizing a (complex) system and/or forecasting its evolution, given several heterogeneous and uncertain sources of information



Widely used for geophysical fluids (meteorology, oceanography, atmospheric chemistry. . .), but also in other numerous domains (e.g. glaciology, nuclear energy, medicine, agriculture planning. . .)

Closely linked to *inverse methods, control theory, estimation theory, filtering. . .*

Data assimilation, the science of compromises

Numerous possible aims:

- Forecast: estimation of the present state (initial condition)
- Model tuning: parameter estimation
- Inverse modeling: estimation of parameter fields
- Data analysis: re-analysis (model = interpolation operator)
- OSSE: optimization of observing systems
- ...

Data assimilation, the science of compromises

Its application to Earth sciences generally raises a number of difficulties, some of them being rather specific:

- non linearities
- huge dimensions
- poor knowledge of error statistics
- non reproducibility (each experiment is unique)
- operational forecast (computations must be performed in a limited time)

Objectives for these two lectures

- introduce data assimilation from several points of view
- give an overview of the main methods
- detail the basic ones and highlight their pros and cons
- introduce some current research problems

Objectives for these two lectures

- introduce data assimilation from several points of view
- give an overview of the main methods
- detail the basic ones and highlight their pros and cons
- introduce some current research problems

Outline

1. Data assimilation for dummies: a simple model problem
2. Generalization: linear estimation theory, variational and sequential approaches
3. Variational algorithms - Adjoint techniques
4. Reduced order Kalman filters
5. Some current research tracks

Some references

1. Blayo E. and M. Nodet, 2012: Introduction à l'assimilation de données variationnelle. *Lecture notes for UJF Master course on data assimilation*.
   https://team.inria.fr/moise/files/2012/03/Methodes-Inverses-Var-M2-math-2009.pdf
2. Bocquet M., 2014 : Introduction aux principes et méthodes de l'assimilation de données en géophysique. *Lecture notes for ENSTA - ENPC data assimilation course*.
   http://cerea.enpc.fr/HomePages/bocquet/Doc/assim-mb.pdf
3. Bocquet M., E. Cosme and E. Blayo (Eds.), 2014: Advanced Data Assimilation for Geosciences. *Oxford University Press*.
4. Bouttier F. and P. Courtier, 1999: Data assimilation, concepts and methods. *Meteorological training course lecture series ECMWF*, European Center for Medium range Weather Forecast, Reading, UK.
   http://www.ecmwf.int/newsevents/training/rcourse_notes/DATA_ASSIMILATION/ASSIM_CONCEPTS/Assim_concepts21.html
5. Cohn S., 1997: An introduction to estimation theory. *Journal of the Meteorological Society of Japan*, **75**, 257-288.
6. Daley R., 1993: Atmospheric data analysis. *Cambridge University Press*.
7. Evensen G., 2009: Data assimilation, the ensemble Kalman filter. *Springer*.
8. Kalnay E., 2003: Atmospheric modeling, data assimilation and predictability. *Cambridge University Press*.
9. Lahoz W., B. Khattatov and R. Menard (Eds.), 2010: Data assimilation. *Springer*.
10. Rodgers C., 2000: Inverse methods for atmospheric sounding. *World Scientific, Series on Atmospheric Oceanic and Planetary Physics*.
11. Tarantola A., 2005: Inverse problem theory and methods for model parameter estimation. *SIAM*.
    http://www.ipgp.fr/~tarantola/Files/Professional/Books/InverseProblemTheory.pdf

A simple but fundamental example

Model problem: least squares approach

Two different available measurements of a single quantity. Which estimation for its true value ? $\longrightarrow$ least squares approach

Model problem: least squares approach

Two different available measurements of a single quantity. Which estimation for its true value ? $\longrightarrow$ least squares approach

Example 2 obs $y_1 = 19°$C and $y_2 = 21°$C of the (unknown) present temperature x.

- Let $J(x) = \frac{1}{2}\left[(x - y_1)^2 + (x - y_2)^2\right]$
- $\text{Min}_x \, J(x) \quad \longrightarrow \hat{x} = \dfrac{y_1 + y_2}{2} = 20°\text{C}$

Model problem: least squares approach

Observation operator If $\neq$ units: $y_1 = 66.2°\text{F}$ and $y_2 = 69.8°\text{F}$

- Let $H(x) = \frac{9}{5}x + 32$
- Let $J(x) = \frac{1}{2}\left[(H(x) - y_1)^2 + (H(x) - y_2)^2\right]$
- $\text{Min}_x\ J(x) \quad \longrightarrow \hat{x} = 20^\circ\text{C}$

Model problem: least squares approach

Observation operator If $\neq$ units: $y_1 = 66.2°\text{F}$ and $y_2 = 69.8°\text{F}$

- Let $H(x) = \frac{9}{5}x + 32$
- Let $J(x) = \frac{1}{2}\left[(H(x) - y_1)^2 + (H(x) - y_2)^2\right]$
- $\text{Min}_x\, J(x) \quad \longrightarrow \hat{x} = 20^\circ\text{C}$

Drawback # 1: *if observation units are inhomogeneous*
$y_1 = 66.2°\text{F}$ and $y_2 = 21°\text{C}$

- $J(x) = \frac{1}{2}\left[(H(x) - y_1)^2 + (x - y_2)^2\right] \quad \longrightarrow \hat{x} = 19.47^\circ\text{C}$!!

Model problem: least squares approach

Observation operator If $\neq$ units: $y_1 = 66.2°$F and $y_2 = 69.8°$F

- Let $H(x) = \dfrac{9}{5}x + 32$
- Let $J(x) = \dfrac{1}{2}\left[(H(x) - y_1)^2 + (H(x) - y_2)^2\right]$
- $\text{Min}_x\, J(x) \quad \longrightarrow \hat{x} = 20^\circ\text{C}$

Drawback # 1: *if observation units are inhomogeneous*

$y_1 = 66.2°$F and $y_2 = 21°$C

- $J(x) = \dfrac{1}{2}\left[(H(x) - y_1)^2 + (x - y_2)^2\right] \quad \longrightarrow \hat{x} = 19.47^\circ\text{C !!}$

Drawback # 2: *if observation accuracies are inhomogeneous*

If y_1 is twice more accurate than y_2, one should obtain $\hat{x} = \dfrac{2y_1 + y_2}{3} = 19.67^\circ\text{C}$

$$\longrightarrow J \text{ should be } J(x) = \frac{1}{2}\left[\left(\frac{x - y_1}{1/2}\right)^2 + \left(\frac{x - y_2}{1}\right)^2\right]$$

Model problem: statistical approach

Reformulation in a **probabilistic framework**:

- the goal is to estimate a scalar value x
- y_i is a realization of a random variable Y_i
- One is looking for an estimator (i.e. a r.v.) $\hat{X}$ that is
  - **linear**: $\hat{X} = \alpha_1 Y_1 + \alpha_2 Y_2$ (in order to be simple)
  - **unbiased**: $E(\hat{X}) = x$ (it seems reasonable)
  - **of minimal variance**: $\text{Var}(\hat{X})$ minimum (optimal accuracy)

$\longrightarrow$ BLUE (Best Linear Unbiased Estimator)

Model problem: statistical approach

Let $Y_i = x + \varepsilon_i$ with

Hypotheses

- $E(\varepsilon_i) = 0 \quad (i = 1, 2)$ unbiased measurement devices
- $\text{Var}(\varepsilon_i) = \sigma_i^2 \quad (i = 1, 2)$ known accuracies
- $\text{Cov}(\varepsilon_1, \varepsilon_2) = 0$ independent measurement errors

Reminder: covariance of two random variables

Let X and Y two random variables.

- Covariance: $$\begin{aligned} \text{Cov}(X,Y) &= E\left[(X-E(X))\,(Y-E(Y))\right] \\ &= E(XY)-E(X)E(Y) \\ \\ \text{Cov}(X,X) &= \text{Var}(X) \end{aligned}$$

- Linear correlation coefficient: $\rho(X,Y) = \dfrac{\text{Cov}(X,Y)}{\sigma_X\,\sigma_Y}$

- Property: X and Y independent $\Longrightarrow \text{Cov}(X,Y)=0$

  The reciprocal is generally false.

Model problem: statistical approach

Let $Y_i = x + \varepsilon_i$ with

Hypotheses

- $E(\varepsilon_i) = 0 \qquad (i = 1, 2)$ — unbiased measurement devices
- $\text{Var}(\varepsilon_i) = \sigma_i^2 \qquad (i = 1, 2)$ — known accuracies
- $\text{Cov}(\varepsilon_1, \varepsilon_2) = 0$ — independent measurement errors

Model problem: statistical approach

Let $Y_i = x + \varepsilon_i$ with

Hypotheses

- $E(\varepsilon_i) = 0 \quad (i = 1, 2)$ — unbiased measurement devices
- $\text{Var}(\varepsilon_i) = \sigma_i^2 \quad (i = 1, 2)$ — known accuracies
- $\text{Cov}(\varepsilon_1, \varepsilon_2) = 0$ — independent measurement errors

Then, since $\hat{X} = \alpha_1 Y_1 + \alpha_2 Y_2 = (\alpha_1 + \alpha_2)x + \alpha_1\varepsilon_1 + \alpha_2\varepsilon_2$:

- $E(\hat{X}) = (\alpha_1 + \alpha_2)x + \alpha_1 E(\varepsilon_1) + \alpha_2 E(\varepsilon_2) \quad \Longrightarrow \alpha_1 + \alpha_2 = 1$

Model problem: statistical approach

Let $Y_i = x + \varepsilon_i$ with

Hypotheses

- $E(\varepsilon_i) = 0 \quad (i = 1, 2)$ — unbiased measurement devices
- $\text{Var}(\varepsilon_i) = \sigma_i^2 \quad (i = 1, 2)$ — known accuracies
- $\text{Cov}(\varepsilon_1, \varepsilon_2) = 0$ — independent measurement errors

Then, since $\hat{X} = \alpha_1 Y_1 + \alpha_2 Y_2 = (\alpha_1 + \alpha_2)x + \alpha_1\varepsilon_1 + \alpha_2\varepsilon_2$:

- $E(\hat{X}) = (\alpha_1 + \alpha_2)x + \alpha_1 E(\varepsilon_1) + \alpha_2 E(\varepsilon_2) \quad \Longrightarrow \alpha_1 + \alpha_2 = 1$
- $\text{Var}(\hat{X}) = E\left[(\hat{X} - x)^2\right] = E\left[(\alpha_1\varepsilon_1 + \alpha_2\varepsilon_2)^2\right] = \alpha_1^2\sigma_1^2 + (1-\alpha_1)^2\sigma_2^2$

$$\frac{\partial}{\partial \alpha_1} = 0 \Longrightarrow \alpha_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

Model problem: statistical approach

In summary:

BLUE

$$\hat{X} = \frac{\dfrac{1}{\sigma_1^2}\,Y_1 + \dfrac{1}{\sigma_2^2}\,Y_2}{\dfrac{1}{\sigma_1^2} + \dfrac{1}{\sigma_2^2}}$$

Its accuracy: $\left[\mathrm{Var}(\hat{X})\right]^{-1} = \dfrac{1}{\sigma_1^2} + \dfrac{1}{\sigma_2^2}$ accuracies are added

go to general case

Model problem: statistical approach

In summary:

BLUE

$$\hat{X} = \frac{\dfrac{1}{\sigma_1^2}\,Y_1 + \dfrac{1}{\sigma_2^2}\,Y_2}{\dfrac{1}{\sigma_1^2} + \dfrac{1}{\sigma_2^2}}$$

Its accuracy: $\left[\mathrm{Var}(\hat{X})\right]^{-1} = \dfrac{1}{\sigma_1^2} + \dfrac{1}{\sigma_2^2}$ accuracies are added

go to general case

Remarks:

- The hypothesis $\mathrm{Cov}(\varepsilon_1, \varepsilon_2) = 0$ is not compulsory at all.
  $\mathrm{Cov}(\varepsilon_1, \varepsilon_2) = c \longrightarrow \alpha_i = \frac{\sigma_j^2 - c}{\sigma_1^2 + \sigma_2^2 - 2c}$

- Statistical hypotheses on the two first moments of $\varepsilon_1, \varepsilon_2$ lead to statistical results on the two first moments of $\hat{X}$.

Model problem: statistical approach

Variational equivalence

This is equivalent to the problem:

$$\text{Minimize } J(x) = \frac{1}{2}\left[\frac{(x-y_1)^2}{\sigma_1^2} + \frac{(x-y_2)^2}{\sigma_2^2}\right]$$

Model problem: statistical approach

Variational equivalence

This is equivalent to the problem:

$$\text{Minimize } J(x) = \frac{1}{2}\left[\frac{(x-y_1)^2}{\sigma_1^2} + \frac{(x-y_2)^2}{\sigma_2^2}\right]$$

Remarks:

- This answers the previous problems of sensitivity to inhomogeneous units and insensitivity to inhomogeneous accuracies
- This gives a rationale for choosing the norm for defining J
- $$\underbrace{J''(\hat{x})}_{\text{convexity}} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} = \underbrace{[\mathrm{Var}(\hat{x})]^{-1}}_{\text{accuracy}}$$

Model problem

Alternative formulation: background + observation

If one considers that y_1 is a prior (or *background*) estimate x_b for x, and $y_2 = y$ is an independent observation, then:

$$J(x) = \underbrace{\frac{1}{2}\,\frac{(x - x_b)^2}{\sigma_b^2}}_{J_b} + \underbrace{\frac{1}{2}\,\frac{(x - y)^2}{\sigma_o^2}}_{J_o}$$

and

$$\hat{x} = \frac{\dfrac{1}{\sigma_b^2}\,x_b + \dfrac{1}{\sigma_o^2}\,y}{\dfrac{1}{\sigma_b^2} + \dfrac{1}{\sigma_o^2}} = x_b + \underbrace{\frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}}_{\text{gain}}\;\underbrace{(y - x_b)}_{\text{innovation}}$$

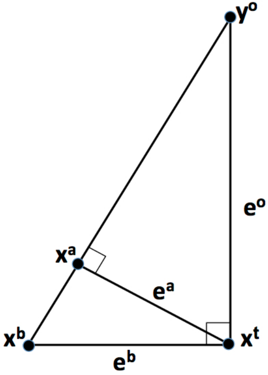
Model problem

Interpretation

If the background error and the observation error are uncorrelated: $E(e^o e^b) = 0$, then one can show that the estimation error and the innovation are uncorrelated:

$$E(e^a(Y - X_b)) = 0$$

$\rightarrow$ orthogonal projection for the scalar product $< Z_1, Z_2 >= E(Z_1 Z_2)$



Model problem: Bayesian approach

One can also consider x as a realization of a r.v. X, and be interested in the pdf $p(X|Y)$.

Reminder: Bayes theorem

Let A and B two events.

- **Conditional probability:** $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

  Example:

  $$P(\text{heart card} \mid \text{red card}) = \frac{1}{2} = \frac{P(\text{heart card} \ \cap \ \text{red card})}{P(\text{red card})} = \frac{8/32}{16/32}$$

- **Bayes theorem:** $P(A|B) = \dfrac{P(B|A)\,P(A)}{P(B)}$

Thus, if X and Y are two random variables:

$$P(X = x \,|\, Y = y) = \frac{P(Y = y \,|\, X = x)\,P(X = x)}{P(Y = y)}$$

Model problem: Bayesian approach

One can also consider x as a realization of a r.v. X, and be interested in the pdf $p(X|Y)$.

Several optimality criteria

- **minimum variance**: $\hat{X}_{MV}$ such that the spread around it is minimal
  $\longrightarrow \hat{X}_{MV} = E(X|Y)$
- **maximum a posteriori**: most probable value of X given Y
  $\longrightarrow \hat{X}_{MAP}$ such that $\frac{\partial p(X|Y)}{\partial X} = 0$
- **maximum likelihood**: $\hat{X}_{ML}$ that maximizes $p(Y|X)$

- Based on the Bayes rule:
  $$P(X = x \,|\, Y = y) = \frac{P(Y = y \,|\, X = x)\, P(X = x)}{P(Y = y)}$$
- requires additional hypotheses on prior pdf for X and for $Y|X$
- In the Gaussian case, these estimations coincide with the BLUE

Model problem: Bayesian approach

Back to our example: observations y_1 and y_2 of an unknown value x.

The simplest approach: maximum likelihood (no prior on X)

Hypotheses:

- $Y_i \hookrightarrow \mathcal{N}(X, \sigma_i^2)$ — unbiased, known accuracies + known pdf
- $\text{Cov}(Y_1, Y_2) = 0$ — independent measurement errors

Likelihood function: $\mathcal{L}(x) = dP(Y_1 = y_1 \text{ and } Y_2 = y_2 \,|\, X = x)$

One is looking for $\hat{x}_{ML} = \text{Argmax}\ \mathcal{L}(x)$ — maximum likelihood estimation

Model problem: Bayesian approach

$$\mathcal{L}(x) = \prod_{i=1}^{2} dP(Y_i = y_i \,|\, X = x) = \prod_{i=1}^{2} \frac{1}{\sqrt{2\pi}\,\sigma_i}\, e^{-\frac{(y_i - x)^2}{2\sigma_i^2}}$$

$$\begin{aligned} \text{Argmax } \mathcal{L}(x) &= \text{Argmin } \left(-\ln \mathcal{L}(x)\right) \\ &= \text{Argmin } \frac{1}{2}\left[\frac{(x-y_1)^2}{\sigma_1^2} + \frac{(x-y_2)^2}{\sigma_2^2}\right] \end{aligned}$$

Hence $\hat{x}_{ML} = \dfrac{\dfrac{1}{\sigma_1^2}\, y_1 + \dfrac{1}{\sigma_2^2}\, y_2}{\dfrac{1}{\sigma_1^2} + \dfrac{1}{\sigma_2^2}}$ BLUE again

(because of Gaussian hypothesis)

Model problem: synthesis

Data assimilation methods are often split into two families: variational methods and statistical methods.

- Variational methods: minimization of a cost function (least squares approach)
- Statistical methods: algebraic computation of the BLUE (with hypotheses on the first two moments), or approximation of pdfs (with hypotheses on the pdfs) and computation of the MAP estimator
- There are strong links between those approaches, depending on the case (linear ? Gaussian ?)

Model problem: synthesis

Data assimilation methods are often split into two families: variational methods and statistical methods.

- Variational methods: minimization of a cost function (least squares approach)
- Statistical methods: algebraic computation of the BLUE (with hypotheses on the first two moments), or approximation of pdfs (with hypotheses on the pdfs) and computation of the MAP estimator
- There are strong links between those approaches, depending on the case (linear ? Gaussian ?)

Theorem

If you have understood this previous stuff, you have (almost) understood everything on data assimilation.

Generalization: variational approach

Generalization: arbitrary number of unknowns and observations

To be estimated: $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbf{R}^n$

Observations: $\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} \in \mathbf{R}^p$

Observation operator: $\mathbf{y} \equiv H(\mathbf{x})$, with $H : \mathbf{R}^n \longrightarrow \mathbf{R}^p$

Generalization: arbitrary number of unknowns and observations

A simple example of observation operator

$$\text{If } \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \quad \text{and} \quad \mathbf{y} = \begin{pmatrix} \text{an observation of } \frac{x_1+x_2}{2} \\ \text{an observation of } x_4 \end{pmatrix}$$

$$\text{then} \quad H(\mathbf{x}) = \mathbf{H}\mathbf{x} \quad \text{with } \mathbf{H} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Generalization: arbitrary number of unknowns and observations

To be estimated: $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbf{R}^n$

Observations: $\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} \in \mathbf{R}^p$

Observation operator: $\mathbf{y} \equiv H(\mathbf{x})$, with $H : \mathbf{R}^n \longrightarrow \mathbf{R}^p$

Cost function: $J(\mathbf{x}) = \frac{1}{2} \|H(\mathbf{x}) - \mathbf{y}\|^2$ with $\|.\|$ to be chosen.

Reminder: norms and scalar products

$$\mathbf{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} \in \mathbf{R}^n$$

- Euclidian norm: $\|\mathbf{u}\|^2 = \mathbf{u}^T\mathbf{u} = \sum_{i=1}^{n} u_i^2$

  Associated scalar product: $(\mathbf{u},\mathbf{v}) = \mathbf{u}^T\mathbf{v} = \sum_{i=1}^{n} u_i v_i$

- Generalized norm: let $\mathbf{M}$ a symmetric positive definite matrix

  $\mathbf{M}$-norm: $\|\mathbf{u}\|^2_{\mathbf{M}} = \mathbf{u}^T\mathbf{M}\,\mathbf{u} = \sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij}\, u_i u_j$

  Associated scalar product: $(\mathbf{u},\mathbf{v})_{\mathbf{M}} = \mathbf{u}^T\mathbf{M}\,\mathbf{v} = \sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij}\, u_i v_j$

Generalization: arbitrary number of unknowns and observations

To be estimated: $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbf{R}^n$

Observations: $\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} \in \mathbf{R}^p$

Observation operator: $\mathbf{y} \equiv H(\mathbf{x})$, with $H : \mathbf{R}^n \longrightarrow \mathbf{R}^p$

Cost function: $J(\mathbf{x}) = \frac{1}{2} \|H(\mathbf{x}) - \mathbf{y}\|^2$ with $\|.\|$ to be chosen.

Generalization: arbitrary number of unknowns and observations

To be estimated: $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbf{R}^n$

Observations: $\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} \in \mathbf{R}^p$

Observation operator: $\mathbf{y} \equiv H(\mathbf{x})$, with $H : \mathbf{R}^n \longrightarrow \mathbf{R}^p$

Cost function: $J(\mathbf{x}) = \frac{1}{2} \|H(\mathbf{x}) - \mathbf{y}\|^2$ with $\|.\|$ to be chosen.

Remark

(Intuitive) necessary (but not sufficient) condition for the existence of a unique minimum:

$$p \geq n$$

Formalism "background value + new observations"

$$\mathbf{z} = \begin{pmatrix} \mathbf{x}_b \\ \mathbf{y} \end{pmatrix} \begin{matrix} \longleftarrow \text{background} \\ \longleftarrow \text{new observations} \end{matrix}$$

The cost function becomes:

$$J(\mathbf{x}) = \underbrace{\frac{1}{2}\,\|\mathbf{x} - \mathbf{x}_b\|_b^2}_{J_b} + \underbrace{\frac{1}{2}\,\|H(\mathbf{x}) - \mathbf{y}\|_o^2}_{J_o}$$

Formalism "background value + new observations"

$$\mathbf{z} = \begin{pmatrix} \mathbf{x}_b \\ \mathbf{y} \end{pmatrix} \begin{matrix} \longleftarrow \text{background} \\ \longleftarrow \text{new observations} \end{matrix}$$

The cost function becomes:

$$J(\mathbf{x}) = \underbrace{\frac{1}{2} \|\mathbf{x} - \mathbf{x}_b\|_b^2}_{J_b} + \underbrace{\frac{1}{2} \|H(\mathbf{x}) - \mathbf{y}\|_o^2}_{J_o}$$

The necessary condition for the existence of a unique minimum ($p \geq n$) is automatically fulfilled.

If the problem is time dependent

- Observations are distributed in time: $\mathbf{y} = \mathbf{y}(t)$
- The observation cost function becomes:

$$J_o(\mathbf{x}) = \frac{1}{2} \sum_{i=0}^{N} \| H_i(\mathbf{x}(t_i)) - \mathbf{y}(t_i) \|_o^2$$

If the problem is time dependent

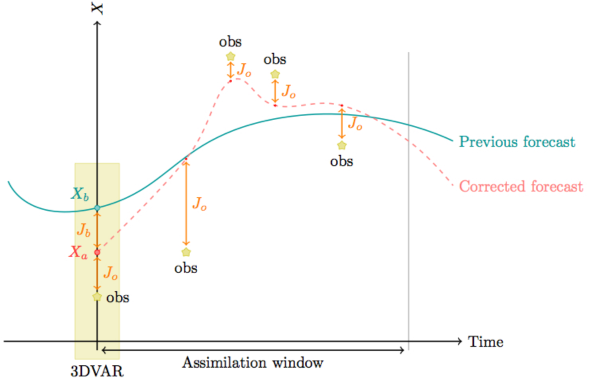
- Observations are distributed in time: $\mathbf{y} = \mathbf{y}(t)$
- The observation cost function becomes:

$$J_o(\mathbf{x}) = \frac{1}{2} \sum_{i=0}^{N} \|H_i(\mathbf{x}(t_i)) - \mathbf{y}(t_i)\|_o^2$$

- There is a model describing the evolution of $\mathbf{x}$: $\dfrac{d\mathbf{x}}{dt} = M(\mathbf{x})$ with $\mathbf{x}(t=0) = \mathbf{x}_0$. Then J is often no longer minimized w.r.t. $\mathbf{x}$, but w.r.t. $\mathbf{x}_0$ only, or to some other parameters.

$$J_o(\mathbf{x}_0) = \frac{1}{2} \sum_{i=0}^{N} \|H_i(\mathbf{x}(t_i)) - \mathbf{y}(t_i)\|_o^2 = \frac{1}{2} \sum_{i=0}^{N} \|H_i(M_{0\to t_i}(\mathbf{x}_0)) - \mathbf{y}(t_i)\|_o^2$$

If the problem is time dependent



$$J(\mathbf{x}_0) = \underbrace{\frac{1}{2}\,\|\mathbf{x}_0 - \mathbf{x}_0^b\|_b^2}_{\text{background term } J_b} + \underbrace{\frac{1}{2}\sum_{i=0}^{N}\|H_i(\mathbf{x}(t_i)) - \mathbf{y}(t_i)\|_o^2}_{\text{observation term } J_o}$$

Uniqueness of the minimum ?

$$J(\mathbf{x}_0) = J_b(\mathbf{x}_0) + J_o(\mathbf{x}_0) = \frac{1}{2}\,\|\mathbf{x}_0 - \mathbf{x}_b\|_b^2 + \frac{1}{2}\,\sum_{i=0}^{N} \|H_i(M_{0\rightarrow t_i}(\mathbf{x}_0)) - \mathbf{y}(t_i)\|_o^2$$

- If H and M are linear then J_o is quadratic.

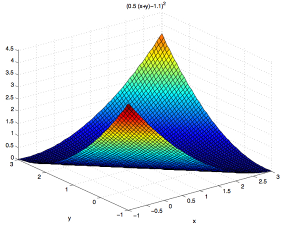
Uniqueness of the minimum ?

$$J(\mathbf{x}_0) = J_b(\mathbf{x}_0) + J_o(\mathbf{x}_0) = \frac{1}{2}\|\mathbf{x}_0 - \mathbf{x}_b\|_b^2 + \frac{1}{2}\sum_{i=0}^{N}\|H_i(M_{0\to t_i}(\mathbf{x}_0)) - \mathbf{y}(t_i)\|_o^2$$

- If H and M are linear then J_o is quadratic.
- However it generally does not have a unique minimum, since the number of observations is generally less than the size of $\mathbf{x}_0$ (the problem is underdetermined: $p < n$).

Example: let $(x_1^t, x_2^t) = (1, 1)$ and $y = 1.1$ an observation of $\frac{1}{2}(x_1 + x_2)$.

$$J_o(x_1, x_2) = \frac{1}{2}\left(\frac{x_1 + x_2}{2} - 1.1\right)^2$$



Uniqueness of the minimum ?

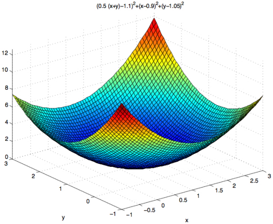
$$J(\mathbf{x}_0) = J_b(\mathbf{x}_0) + J_o(\mathbf{x}_0) = \frac{1}{2}\,\|\mathbf{x}_0 - \mathbf{x}_b\|_b^2 + \frac{1}{2}\sum_{i=0}^{N}\|H_i(M_{0\to t_i}(\mathbf{x}_0)) - \mathbf{y}(t_i)\|_o^2$$

- If H and M are linear then J_o is quadratic.
- However it generally does not have a unique minimum, since the number of observations is generally less than the size of $\mathbf{x}_0$ (the problem is underdetermined).
- Adding J_b makes the problem of minimizing $J = J_o + J_b$ well posed.

Example: let $(x_1^t, x_2^t) = (1, 1)$ and $y = 1.1$ an observation of $\frac{1}{2}(x_1 + x_2)$. Let $(x_1^b, x_2^b) = (0.9, 1.05)$

$$J(x_1, x_2) = \underbrace{\frac{1}{2}\left(\frac{x_1 + x_2}{2} - 1.1\right)^2}_{J_o} + \underbrace{\frac{1}{2}\left[(x_1 - 0.9)^2 + (x_2 - 1.05)^2\right]}_{J_b}$$

$$\longrightarrow (x_1^*, x_2^*) = (0.94166..., 1.09166...)$$



Uniqueness of the minimum ?

$$J(\mathbf{x}_0) = J_b(\mathbf{x}_0) + J_o(\mathbf{x}_0) = \frac{1}{2}\,\|\mathbf{x}_0 - \mathbf{x}_b\|_b^2 + \frac{1}{2}\,\sum_{i=0}^{N}\|H_i(M_{0\to t_i}(\mathbf{x}_0)) - \mathbf{y}(t_i)\|_o^2$$

- If H and/or M are nonlinear then J_o is no longer quadratic.

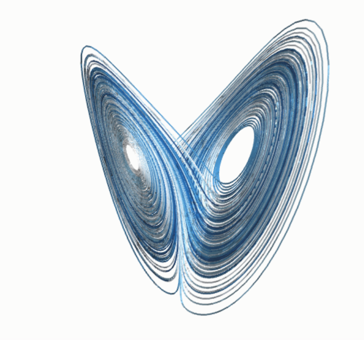
Uniqueness of the minimum ?

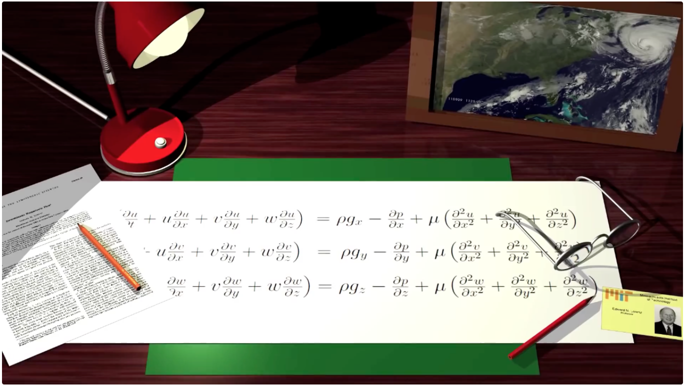
$$J(\mathbf{x}_0) = J_b(\mathbf{x}_0) + J_o(\mathbf{x}_0) = \frac{1}{2}\,\|\mathbf{x}_0 - \mathbf{x}_b\|_b^2 + \frac{1}{2}\sum_{i=0}^{N}\|H_i(M_{0\to t_i}(\mathbf{x}_0)) - \mathbf{y}(t_i)\|_o^2$$

- If H and/or M are nonlinear then J_o is no longer quadratic.

  Example: the Lorenz system (1963)

$$\begin{cases} \dfrac{dx}{dt} = \alpha(y - x) \\ \dfrac{dy}{dt} = \beta x - y - xz \\ \dfrac{dz}{dt} = -\gamma z + xy \end{cases}$$





http://www.chaos-math.org

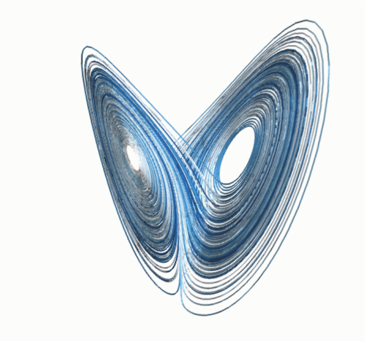
Uniqueness of the minimum ?

$$J(\mathbf{x}_0) = J_b(\mathbf{x}_0) + J_o(\mathbf{x}_0) = \frac{1}{2}\,\|\mathbf{x}_0 - \mathbf{x}_b\|_b^2 + \frac{1}{2}\sum_{i=0}^{N}\|H_i(M_{0\to t_i}(\mathbf{x}_0)) - \mathbf{y}(t_i)\|_o^2$$

- If H and/or M are nonlinear then J_o is no longer quadratic.

  Example: the Lorenz system (1963)

$$\begin{cases} \dfrac{dx}{dt} = \alpha(y - x) \\ \dfrac{dy}{dt} = \beta x - y - xz \\ \dfrac{dz}{dt} = -\gamma z + xy \end{cases}$$

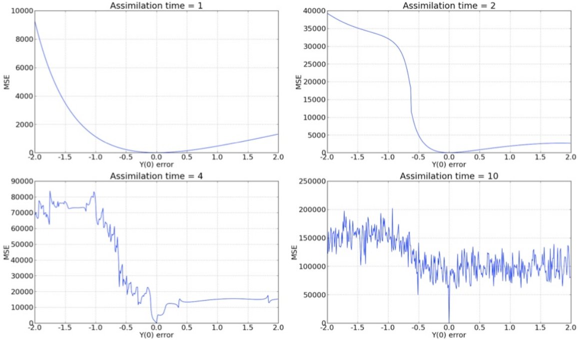


$$J_o(y_0) = \frac{1}{2}\sum_{i=0}^{N}\left(x(t_i) - x_{obs}(t_i)\right)^2\,dt$$

Uniqueness of the minimum ?

$$J(\mathbf{x}_0) = J_b(\mathbf{x}_0) + J_o(\mathbf{x}_0) = \frac{1}{2}\,\|\mathbf{x}_0 - \mathbf{x}_b\|_b^2 + \frac{1}{2}\sum_{i=0}^{N}\|H_i(M_{0\rightarrow t_i}(\mathbf{x}_0)) - \mathbf{y}(t_i)\|_o^2$$

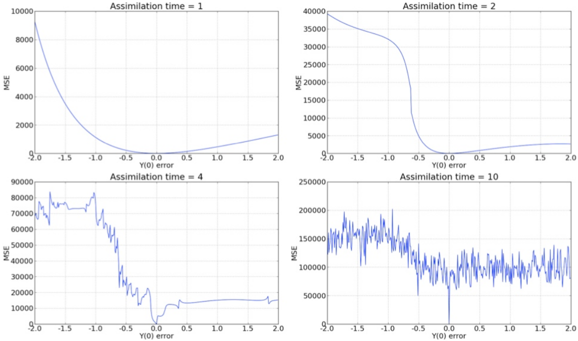
- If H and/or M are nonlinear then J_o is no longer quadratic.



Uniqueness of the minimum ?

$$J(\mathbf{x}_0) = J_b(\mathbf{x}_0) + J_o(\mathbf{x}_0) = \frac{1}{2}\,\|\mathbf{x}_0 - \mathbf{x}_b\|_b^2 + \frac{1}{2}\sum_{i=0}^{N} \|H_i(M_{0\rightarrow t_i}(\mathbf{x}_0)) - \mathbf{y}(t_i)\|_o^2$$

- If H and/or M are nonlinear then J_o is no longer quadratic.



- Adding J_b makes it "more quadratic" (J_b is a regularization term), but $J = J_o + J_b$ may however have several local minima.

A fundamental remark before going into minimization aspects

Once J is defined (i.e. once all the ingredients are chosen: control variables, norms, observations...), the problem is entirely defined. Hence its solution.

The "physical" (i.e. the most important) part of data assimilation lies in the definition of J.

The rest of the job, i.e. minimizing J, is "only" technical work.

Minimum of a quadratic function in finite dimension

Theorem: Generalized (or Moore-Penrose) inverse

Let $\mathbf{M}$ a $p \times n$ matrix, with rank n, and $\mathbf{b} \in \mathbf{R}^p$. *(hence $p \geq n$)*

Let $J(\mathbf{x}) = \|\mathbf{M}\mathbf{x} - \mathbf{b}\|^2 = (\mathbf{M}\mathbf{x} - \mathbf{b})^T(\mathbf{M}\mathbf{x} - \mathbf{b})$.

J is minimum for $\hat{\mathbf{x}} = \mathbf{M}^+\mathbf{b}$, where $\mathbf{M}^+ = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T$ *(generalized, or Moore-Penrose, inverse)*.

Minimum of a quadratic function in finite dimension

Theorem: Generalized (or Moore-Penrose) inverse

Let $\mathbf{M}$ a $p \times n$ matrix, with rank n, and $\mathbf{b} \in \mathbf{R}^p$. *(hence $p \geq n$)*

Let $J(\mathbf{x}) = \|\mathbf{Mx} - \mathbf{b}\|^2 = (\mathbf{Mx} - \mathbf{b})^T(\mathbf{Mx} - \mathbf{b})$.

J is minimum for $\hat{\mathbf{x}} = \mathbf{M}^+\mathbf{b}$, where $\mathbf{M}^+ = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T$ *(generalized, or Moore-Penrose, inverse)*.

Corollary: with a generalized norm

Let $\mathbf{N}$ a $p \times p$ symmetric definite positive matrix.

Let $J_1(\mathbf{x}) = \|\mathbf{Mx} - \mathbf{b}\|_N^2 = (\mathbf{Mx} - \mathbf{b})^T\mathbf{N}\,(\mathbf{Mx} - \mathbf{b})$.

J_1 is minimum for $\hat{\mathbf{x}} = (\mathbf{M}^T\mathbf{N}\mathbf{M})^{-1}\mathbf{M}^T\mathbf{N}\,\mathbf{b}$.

Link with data assimilation

In the case of a linear, time independent, data assimilation problem:

$$J_o(\mathbf{x}) = \frac{1}{2} \|\mathbf{Hx} - \mathbf{y}\|_o^2 = \frac{1}{2} (\mathbf{Hx} - \mathbf{y})^T \mathbf{R}^{-1} (\mathbf{Hx} - \mathbf{y})$$

Optimal estimation in the linear case: J_o only

$$\min_{\mathbf{x} \in \mathbf{R}^n} J_o(\mathbf{x}) \quad \longrightarrow \quad \hat{\mathbf{x}} = (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{y}$$

go to statistical approach

Link with data assimilation

With the formalism "background value + new observations":

$$
\begin{aligned}
J(\mathbf{x}) &= J_b(\mathbf{x}) + J_o(\mathbf{x}) \\
&= \frac{1}{2}\,\|\mathbf{x}-\mathbf{x}_b\|_b^2 \quad + \quad \frac{1}{2}\,\|H(\mathbf{x})-\mathbf{y}\|_o^2 \\
&= \frac{1}{2}\,(\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + \frac{1}{2}\,(\mathbf{H}\mathbf{x}-\mathbf{y})^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{x}-\mathbf{y}) \\
&= (\mathbf{M}\mathbf{x}-\mathbf{b})^T\mathbf{N}\,(\mathbf{M}\mathbf{x}-\mathbf{b}) = \|\mathbf{M}\mathbf{x}-\mathbf{b}\|_{\mathbf{N}}^2
\end{aligned}
$$

$$
\text{with } \mathbf{M} = \begin{pmatrix} \mathbf{I}_n \\ \mathbf{H} \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} \mathbf{x}_b \\ \mathbf{y} \end{pmatrix} \quad \mathbf{N} = \begin{pmatrix} \mathbf{B}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}^{-1} \end{pmatrix}
$$

Link with data assimilation

With the formalism "background value + new observations":

$$
\begin{aligned}
J(\mathbf{x}) &= J_b(\mathbf{x}) + J_o(\mathbf{x}) \\
&= \frac{1}{2}\,\|\mathbf{x}-\mathbf{x}_b\|_b^2 \quad + \quad \frac{1}{2}\,\|H(\mathbf{x})-\mathbf{y}\|_o^2 \\
&= \frac{1}{2}\,(\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + \frac{1}{2}\,(\mathbf{Hx}-\mathbf{y})^T\mathbf{R}^{-1}(\mathbf{Hx}-\mathbf{y}) \\
&= (\mathbf{Mx}-\mathbf{b})^T\mathbf{N}\,(\mathbf{Mx}-\mathbf{b}) = \|\mathbf{Mx}-\mathbf{b}\|_{\mathbf{N}}^2
\end{aligned}
$$

$$
\text{with } \mathbf{M} = \begin{pmatrix} \mathbf{I}_n \\ \mathbf{H} \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} \mathbf{x}_b \\ \mathbf{y} \end{pmatrix} \quad \mathbf{N} = \begin{pmatrix} \mathbf{B}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}^{-1} \end{pmatrix}
$$

Optimal estimation in the linear case: $J_b + J_o$

$$
\hat{\mathbf{x}} = \mathbf{x}_b + \underbrace{(\mathbf{B}^{-1}+\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}}_{\text{gain matrix}} \quad \underbrace{(\mathbf{y}-\mathbf{Hx}_b)}_{\text{innovation vector}}
$$

Remark: The gain matrix also reads $\mathbf{BH}^T(\mathbf{HBH}^T+\mathbf{R})^{-1}$ 
(Sherman-Morrison-Woodbury formula)

go to statistical approach

Remark

Given the size of n and p, it is generally impossible to handle explicitly $\mathbf{H}$, $\mathbf{B}$ and $\mathbf{R}$. So the direct computation of the gain matrix is impossible.

- even in the linear case (for which we have an explicit expression for $\hat{\mathbf{x}}$), the computation of $\hat{\mathbf{x}}$ is performed using an optimization algorithm.

Generalization: statistical approach

Generalization: statistical approach

To be estimated: $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbf{R}^n$ Observations: $\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} \in \mathbf{R}^p$

Observation operator: $\mathbf{y} \equiv H(\mathbf{x})$, with $H : \mathbf{R}^n \longrightarrow \mathbf{R}^p$

Generalization: statistical approach

To be estimated: $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbf{R}^n$ Observations: $\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} \in \mathbf{R}^p$

Observation operator: $\mathbf{y} \equiv H(\mathbf{x})$, with $H : \mathbf{R}^n \longrightarrow \mathbf{R}^p$

Statistical framework:

- $\mathbf{y}$ is a realization of a random vector $\mathbf{Y}$

Reminder: random vectors and covariance matrices

- Random vector: $\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ where each X_i is a random variable

- Covariance matrix:

$$\mathbf{C} = (\mathrm{Cov}(X_i, X_j))_{1 \le i,j \le n} = E\left([\mathbf{X} - E(\mathbf{X})][\mathbf{X} - E(\mathbf{X})]^T\right)$$

on the diagonal: $C_{ii} = \mathrm{Var}(X_i)$

- Property: A covariance matrix is symmetric positive semidefinite. (definite if the r.v. are linearly independent)

Generalization: statistical approach

To be estimated: $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbf{R}^n$ Observations: $\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} \in \mathbf{R}^p$

Observation operator: $\mathbf{y} \equiv H(\mathbf{x})$, with $H : \mathbf{R}^n \longrightarrow \mathbf{R}^p$

Statistical framework:

- $\mathbf{y}$ is a realization of a random vector $\mathbf{Y}$
- One is looking for the BLUE, i.e. a r.v. $\hat{\mathbf{X}}$ that is
  - **linear**: $\hat{\mathbf{X}} = \mathbf{AY}$ with size$(\mathbf{A}) = (n, p)$
  - **unbiased**: $E(\hat{\mathbf{X}}) = \mathbf{x}$
  - **of minimal variance**:

$$\text{Var}(\hat{\mathbf{X}}) = \sum_{i=1}^{n} \text{Var}(\hat{X}_i) = \text{Tr}(\text{Cov}(\hat{\mathbf{X}})) \text{ minimum}$$

Generalization: statistical approach

Hypotheses

- Linear observation operator: $H(\mathbf{x}) = \mathbf{Hx}$
- Let $\mathbf{Y} = \mathbf{Hx} + \varepsilon$ with ε random vector in $\mathbf{R}^p$
  - $E(\varepsilon) = 0$ unbiased measurement devices
  - $\text{Cov}(\varepsilon) = E(\varepsilon\varepsilon^T) = \mathbf{R}$ known accuracies and covariances

Generalization: statistical approach

Hypotheses

- Linear observation operator: $H(\mathbf{x}) = \mathbf{H}\mathbf{x}$
- Let $\mathbf{Y} = \mathbf{H}\mathbf{x} + \varepsilon$ with ε random vector in $\mathbf{R}^p$
  - $E(\varepsilon) = 0$ — unbiased measurement devices
  - $\text{Cov}(\varepsilon) = E(\varepsilon\varepsilon^T) = \mathbf{R}$ — known accuracies and covariances

BLUE:

- **linear**: $\hat{\mathbf{X}} = \mathbf{A}\mathbf{Y}$ with $\mathbf{A}(n, p)$
- **unbiased**: $E(\hat{\mathbf{X}}) = E(\mathbf{A}\mathbf{H}\mathbf{x} + \mathbf{A}\varepsilon) = \mathbf{A}\mathbf{H}\mathbf{x} + \mathbf{A}E(\varepsilon) = \mathbf{A}\mathbf{H}\mathbf{x}$

  So: $E(\hat{\mathbf{X}}) = \mathbf{x} \iff \mathbf{A}\mathbf{H} = \mathbf{I}_n$.

Generalization: statistical approach

Hypotheses

- Linear observation operator: $H(\mathbf{x}) = \mathbf{Hx}$
- Let $\mathbf{Y} = \mathbf{Hx} + \varepsilon$ with ε random vector in $\mathbf{R}^p$
  - $E(\varepsilon) = 0$ unbiased measurement devices
  - $\text{Cov}(\varepsilon) = E(\varepsilon\varepsilon^T) = \mathbf{R}$ known accuracies and covariances

BLUE:

- **linear**: $\hat{\mathbf{X}} = \mathbf{AY}$ with $\mathbf{A}(n,p)$
- **unbiased**: $E(\hat{\mathbf{X}}) = E(\mathbf{AHx} + \mathbf{A}\varepsilon) = \mathbf{AHx} + \mathbf{A}E(\varepsilon) = \mathbf{AHx}$

$$\text{So: } E(\hat{\mathbf{X}}) = \mathbf{x} \iff \mathbf{AH} = \mathbf{I}_n.$$

Remark:

$$\mathbf{AH} = \mathbf{I}_n \Longrightarrow \ker \mathbf{H} = \{\mathbf{0}\} \Longrightarrow \text{rank}(\mathbf{H}) = n$$

Since $\text{size}(\mathbf{H}) = (p, n)$, this implies $n \leq p$ (again !)

BLUE:

- **minimal variance**: $\min \mathrm{Tr}(\mathrm{Cov}(\hat{\mathbf{X}}))$

$$\hat{\mathbf{X}} = \mathbf{AY} = \mathbf{AHx} + \mathbf{A}\varepsilon = \mathbf{x} + \mathbf{A}\varepsilon$$

$$\begin{aligned}\mathrm{Cov}(\hat{\mathbf{X}}) &= E\left([\hat{\mathbf{X}} - E(\hat{\mathbf{X}})][\hat{\mathbf{X}} - E(\hat{\mathbf{X}})]^T\right) \\ &= \mathbf{A}E(\varepsilon\varepsilon^T)\mathbf{A}^T = \mathbf{ARA}^T\end{aligned}$$

Find $\mathbf{A}$ that minimizes $\mathrm{Tr}(\mathbf{ARA}^T)$ under the constraint $\mathbf{AH} = \mathbf{I}_n$

BLUE:

- **minimal variance**: $\min \mathrm{Tr}(\mathrm{Cov}(\hat{\mathbf{X}}))$

$$\hat{\mathbf{X}} = \mathbf{AY} = \mathbf{AHx} + \mathbf{A}\varepsilon = \mathbf{x} + \mathbf{A}\varepsilon$$

$$\begin{aligned}\mathrm{Cov}(\hat{\mathbf{X}}) &= E\left([\hat{\mathbf{X}} - E(\hat{\mathbf{X}})][\hat{\mathbf{X}} - E(\hat{\mathbf{X}})]^T\right) \\ &= \mathbf{A}E(\varepsilon\varepsilon^T)\mathbf{A}^T = \mathbf{ARA}^T\end{aligned}$$

Find $\mathbf{A}$ that minimizes $\mathrm{Tr}(\mathbf{ARA}^T)$ under the constraint $\mathbf{AH} = \mathbf{I}_n$

Gauss-Markov theorem

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

BLUE:

- **minimal variance**: $\min \operatorname{Tr}(\operatorname{Cov}(\hat{\mathbf{X}}))$

$$\hat{\mathbf{X}} = \mathbf{AY} = \mathbf{AHx} + \mathbf{A}\varepsilon = \mathbf{x} + \mathbf{A}\varepsilon$$

$$\begin{aligned} \operatorname{Cov}(\hat{\mathbf{X}}) &= E\left([\hat{\mathbf{X}} - E(\hat{\mathbf{X}})][\hat{\mathbf{X}} - E(\hat{\mathbf{X}})]^T\right) \\ &= \mathbf{A}E(\varepsilon\varepsilon^T)\mathbf{A}^T = \mathbf{ARA}^T \end{aligned}$$

Find $\mathbf{A}$ that minimizes $\operatorname{Tr}(\mathbf{ARA}^T)$ under the constraint $\mathbf{AH} = \mathbf{I}_n$

Gauss-Markov theorem

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

This also leads to $\operatorname{Cov}(\hat{\mathbf{X}}) = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}$

Link with the variational approach

Statistical approach: BLUE

$$\hat{\mathbf{X}} = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{Y} \quad \text{with } \mathrm{Cov}(\hat{\mathbf{X}}) = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}$$

go to variational approach

Link with the variational approach

Statistical approach: BLUE

$$\hat{\mathbf{X}} = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{Y} \quad \text{with } \mathrm{Cov}(\hat{\mathbf{X}}) = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}$$

go to variational approach

Variational approach in the linear case

$$J_o(\mathbf{x}) = \frac{1}{2}\,\|\mathbf{H}\mathbf{x} - \mathbf{y}\|_o^2 = \frac{1}{2}\,(\mathbf{H}\mathbf{x} - \mathbf{y})^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{x} - \mathbf{y})$$

$$\min_{\mathbf{x}\in\mathbf{R}^n} J_o(\mathbf{x}) \quad \longrightarrow \quad \hat{\mathbf{x}} = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\,\mathbf{y}$$

Link with the variational approach

Statistical approach: BLUE

$$\hat{\mathbf{X}} = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{Y} \quad \text{with } \mathrm{Cov}(\hat{\mathbf{X}}) = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}$$

go to variational approach

Variational approach in the linear case

$$J_o(\mathbf{x}) = \frac{1}{2}\|\mathbf{H}\mathbf{x}-\mathbf{y}\|_o^2 = \frac{1}{2}(\mathbf{H}\mathbf{x}-\mathbf{y})^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{x}-\mathbf{y})$$

$$\min_{\mathbf{x}\in\mathbf{R}^n} J_o(\mathbf{x}) \quad \longrightarrow \quad \hat{\mathbf{x}} = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}$$

Remarks

- The statistical approach rationalizes the choice of the norm in the variational approach.
- $$\underbrace{\left[\mathrm{Cov}(\hat{\mathbf{X}})\right]^{-1}}_{\text{accuracy}} = \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} = \underbrace{\mathrm{Hess}(J_o)}_{\text{convexity}}$$

Statistical approach: formalism "background value + new observations"

$$\mathbf{Z} = \begin{pmatrix} \mathbf{X}_b \\ \mathbf{Y} \end{pmatrix} \begin{matrix} \longleftarrow \text{background} \\ \longleftarrow \text{new observations} \end{matrix}$$

Let $\mathbf{X}_b = \mathbf{x} + \varepsilon_b$ and $\mathbf{Y} = \mathbf{H}\mathbf{x} + \varepsilon_o$

Hypotheses:

- $E(\varepsilon_b) = 0$ — unbiased background
- $E(\varepsilon_o) = 0$ — unbiased measurement devices
- $\text{Cov}(\varepsilon_b, \varepsilon_o) = 0$ — independent background and observation errors
- $\text{Cov}(\varepsilon_b) = \mathbf{B}$ et $\text{Cov}(\varepsilon_o) = \mathbf{R}$ — known accuracies and covariances

This is again the general BLUE framework, with

$$\mathbf{Z} = \begin{pmatrix} \mathbf{X}_b \\ \mathbf{Y} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_n \\ \mathbf{H} \end{pmatrix} \mathbf{x} + \begin{pmatrix} \varepsilon_b \\ \varepsilon_o \end{pmatrix} \quad \text{and} \quad \text{Cov}(\varepsilon) = \begin{pmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{pmatrix}$$

Statistical approach: formalism "background value + new observations"

Statistical approach: BLUE

$$\hat{\mathbf{X}} = \mathbf{X}_b + \underbrace{(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}}_{\text{gain matrix}} \underbrace{(\mathbf{Y} - \mathbf{H}\mathbf{X}_b)}_{\text{innovation vector}}$$

with $\left[\text{Cov}(\hat{\mathbf{X}})\right]^{-1} = \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$ accuracies are added

go to model problem

Link with the variational approach

Statistical approach: BLUE

$$\hat{\mathbf{X}} = \mathbf{X}_b + (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{Y} - \mathbf{H}\mathbf{X}_b)$$

with $\text{Cov}(\hat{\mathbf{X}}) = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}$

go to variational approach

Link with the variational approach

Statistical approach: BLUE

$$\hat{\mathbf{X}} = \mathbf{X}_b + (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{Y} - \mathbf{H}\mathbf{X}_b)$$

with $\text{Cov}(\hat{\mathbf{X}}) = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}$

go to variational approach

Variational approach in the linear case

$$\begin{aligned} J(\mathbf{x}) &= \frac{1}{2}\|\mathbf{x} - \mathbf{x}_b\|_b^2 + \frac{1}{2}\|H(\mathbf{x}) - \mathbf{y}\|_o^2 \\ &= \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}(\mathbf{H}\mathbf{x} - \mathbf{y})^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{x} - \mathbf{y}) \end{aligned}$$

$$\min_{\mathbf{x}\in\mathbf{R}^n} J(\mathbf{x}) \longrightarrow \hat{\mathbf{x}} = \mathbf{x}_b + (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\ (\mathbf{y} - \mathbf{H}\mathbf{x}_b)$$

Link with the variational approach

Statistical approach: BLUE

$$\hat{\mathbf{X}} = \mathbf{X}_b + (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{Y} - \mathbf{H}\mathbf{X}_b)$$

with $\mathrm{Cov}(\hat{\mathbf{X}}) = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}$ *go to variational approach*

Variational approach in the linear case

$$\begin{aligned} J(\mathbf{x}) &= \frac{1}{2}\|\mathbf{x} - \mathbf{x}_b\|_b^2 + \frac{1}{2}\|H(\mathbf{x}) - \mathbf{y}\|_o^2 \\ &= \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}(\mathbf{H}\mathbf{x} - \mathbf{y})^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{x} - \mathbf{y}) \end{aligned}$$

$$\min_{\mathbf{x}\in\mathbf{R}^n} J(\mathbf{x}) \longrightarrow \hat{\mathbf{x}} = \mathbf{x}_b + (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}_b)$$

Same remarks as previously

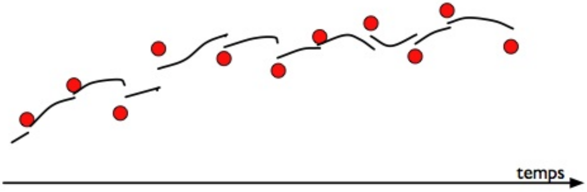
- The statistical approach rationalizes the choice of the norms for J_o and J_b in the variational approach.
- $$\underbrace{\left[\mathrm{Cov}(\hat{\mathbf{X}})\right]^{-1}}_{\text{accuracy}} = \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} = \underbrace{\mathrm{Hess}(J)}_{\text{convexity}}$$

If the problem is time dependent

Dynamical system: $\mathbf{x}^t(t_{k+1}) = \mathbf{M}(t_k, t_{k+1})\mathbf{x}^t(t_k) + \mathbf{e}(t_k)$

- $\mathbf{x}^t(t_k)$ true state at time t_k
- $\mathbf{M}(t_k, t_{k+1})$ model assumed linear between t_k and t_{k+1}
- $\mathbf{e}(t_k)$ model error at time t_k

At every observation time t_k, we have an observation $\mathbf{y}_k$ and a model forecast $\mathbf{x}^f(t_k)$. The BLUE can be applied:



If the problem is time dependent

$$\mathbf{x}^t(t_{k+1}) = \mathbf{M}(t_k, t_{k+1})\,\mathbf{x}^t(t_k) + \mathbf{e}(t_k)$$

Hypotheses

- $\mathbf{e}(t_k)$ is unbiased, with covariance matrix $\mathbf{Q}_k$
- $\mathbf{e}(t_k)$ and $\mathbf{e}(t_l)$ are independent $(k \neq l)$
- Unbiased observation $\mathbf{y}_k$, with error covariance matrix $\mathbf{R}_k$
- $\mathbf{e}(t_k)$ and analysis error $\mathbf{x}^a(t_k) - \mathbf{x}^t(t_k)$ are independent

If the problem is time dependent

Kalman filter (Kalman and Bucy, 1961)

Initialization:

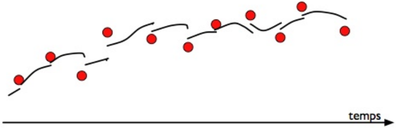
$$\begin{array}{rcll} \mathbf{x}^a(t_0) & = & \mathbf{x}_0 & \textit{approximate initial state} \\ \mathbf{P}^a(t_0) & = & \mathbf{P}_0 & \textit{error covariance matrix} \end{array}$$

Step k: *(prediction - correction, or forecast - analysis)*

$$\begin{array}{rcll} \mathbf{x}^f(t_{k+1}) & = & \mathbf{M}(t_k, t_{k+1})\,\mathbf{x}^a(t_k) & \text{Forecast} \\ \mathbf{P}^f(t_{k+1}) & = & \mathbf{M}(t_k, t_{k+1})\mathbf{P}^a(t_k)\mathbf{M}^T(t_k, t_{k+1}) + \mathbf{Q}_k & \end{array}$$

$$\begin{array}{rcll} \mathbf{x}^a(t_{k+1}) & = & \mathbf{x}^f(t_{k+1}) + \mathbf{K}_{k+1}\left[\mathbf{y}_{k+1} - \mathbf{H}_{k+1}\mathbf{x}^f(t_{k+1})\right] & \text{BLUE} \\ \mathbf{K}_{k+1} & = & \mathbf{P}^f(t_{k+1})\mathbf{H}^T_{k+1}\left[\mathbf{H}_{k+1}\mathbf{P}^f(t_{k+1})\mathbf{H}^T_{k+1} + \mathbf{R}_{k+1}\right]^{-1} & \\ \mathbf{P}^a(t_{k+1}) & = & \mathbf{P}^f(t_{k+1}) - \mathbf{K}_{k+1}\mathbf{H}_{k+1}\mathbf{P}^f(t_{k+1}) & \end{array}$$

where exponents ^{f} and ^{a} stand respectively for *forecast* and *analysis*.



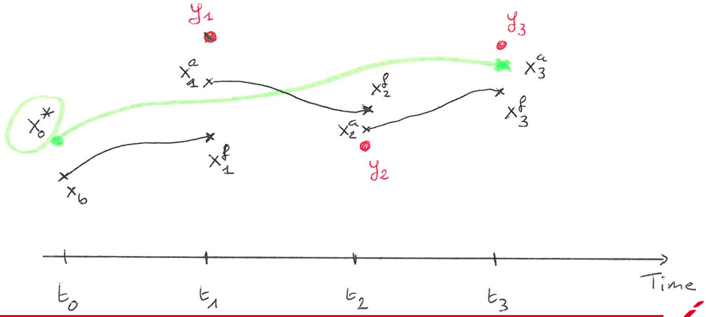
If the problem is time dependent

Equivalence with the variational approach

If $\mathbf{H}_k$ and $\mathbf{M}(t_k, t_{k+1})$ are linear, and if the model is perfect ($\mathbf{e}_k = 0$), then the Kalman filter and the variational method minimizing

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}_0)^T\mathbf{P}_0^{-1}(\mathbf{x}-\mathbf{x}_0) + \frac{1}{2}\sum_{k=0}^{N}(\mathbf{H_k}\mathbf{M}(t_0,t_k)\mathbf{x}-\mathbf{y}_k)^T\mathbf{R}_k^{-1}(\mathbf{H_k}\mathbf{M}(t_0,t_k)\mathbf{x}-\mathbf{y}_k)$$

lead to the same solution at $t = t_N$.



In summary

In summary

variational approach least squares minimization (non dimensional terms)

- no particular hypothesis
- either for stationary or time dependent problems
- If M and H are linear, the cost function is quadratic: a unique solution if $p \geq n$
- Adding a background term ensures this property.
- If things are non linear, the approach is still valid. Possibly several minima

statistical approach

- hypotheses on the first two moments
- time independent + H linear + $p \geq n$: BLUE (first two moments)
- time dependent + M and H linear: Kalman filter (based on the BLUE)

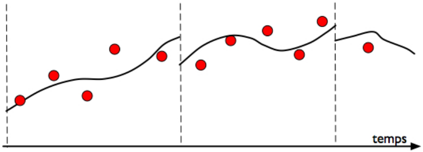
- hypotheses on the pdfs: Bayesian approach (pdf) + ML or MAP estimator

In summary

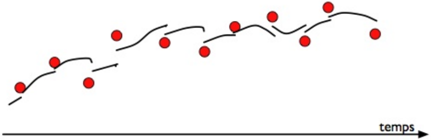
The statistical approach gives a rationale for the choice of the norms, and gives an estimation of the uncertainty.

time independent problems: if H is linear, the variational and the statistical approaches lead to the same solution (provided $\|.\|_b$ is based on $\mathbf{B}^{-1}$ and $\|.\|_o$ is based on $\mathbf{R}^{-1}$)

time dependent problems: if H and M are linear, if the model is perfect, both approaches lead to the same solution at final time.



4D-Var



Kalman filter

Common main methodological difficulties

- Non linearities: J non quadratic / what about Kalman filter ?
- Huge dimensions $[\mathbf{x}] = \mathcal{O}(10^6 - 10^9)$: minimization of J / management of huge matrices
- Poorly known error statistics: choice of the norms / $\mathbf{B}, \mathbf{R}, \mathbf{Q}$
- Scientific computing issues (data management, code efficiency, parallelization...)

$\longrightarrow$ NEXT LECTURE