Introduction to data assimilation

Eric Blayo Arthur Vidard

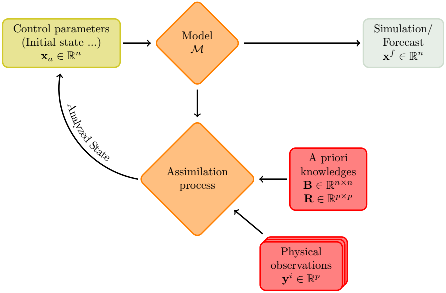
Peyresq, $25^{th} - 26^{th}$ of April 2015

Peyresq Summer School 2015

Data assimilation

Principle

- Forecast is produced by integration of a model from an initial state



Information sources:

- Model $\Rightarrow$ system state
  - wind
  - pressure
  - ...
- Observations
- a priori knowledge
  - First Guess and background error statistics
  - Observation error statistics
  - ...

- Data Assimilation combines in a consistent manner all the available information (model, observations, a priori knowledge) in order to retrieve the 'optimal' initial state.

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 some input 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 the 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

Outline

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

Model problem: least squares approach

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

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

Model problem: least squares approach

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

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

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

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

Model problem: least squares approach

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

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

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

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

Drawback # 2: *if observation accuracies are inhomogeneous*

If y_1 is twice more accurate than y_2, one should obtain $\hat{x} = \dfrac{2y_1 + y_2}{3} = 19.67°C \qquad \longrightarrow J$

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**: $\mathrm{Var}(\hat{X})$ minimum (optimal accuracy)

$\longrightarrow$ BLUE (Best Linear Unbiased Estimator)

Model problem: statistical approach

Let $Y_i = x + \varepsilon_i$ with

Hypotheses

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

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

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

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

(variance minimum)

Model problem: statistical approach

In summary:

BLUE

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

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

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

Outline

Definitions and notations:

italic capitals : non linear operators (M, H, ...);
bold capitals: linear operator or matrices ($\mathbf{M}$, $\mathbf{R}$, ...);
bold small caps: vectors ($\mathbf{x}^b$, $\mathbf{y}^o$, ...)

- Assimilation window: time window over which the data will be considered all at once
- First guess or background: prior estimation of the control vector ($\mathbf{x}^b$)
- Analysis: estimation of the control vector after data assimilation ($\mathbf{x}^a$)
- Increment: correction to the control vector ($\mathbf{x}^a - \mathbf{x}^b$)
- Innovation vector: misfit to the observation ($\mathbf{d} = \mathbf{y}^o - H(\mathbf{x})$)

In this part, we sometime consider a time evolving system described by a set of non linear PDE (aka the model M)

$$\left\{ \begin{array}{ll} \mathbf{x}_{t_i} = M_{t_{i-1},t_i}(\mathbf{x}_{t_{i-1}}), & i = 1, N \\ \mathbf{x}_0 = \mathbf{x}^b & \end{array} \right.$$

Origin: Bayes theory

assuming the system state and the observation are random variable we want to maximise the probability of $\mathbf{x}^a$ knowing the observations. , i.e.:

$$P(X = \mathbf{x}|Y = \mathbf{y}^o) = P_{X|Y}(\mathbf{x}|\mathbf{y}^o)$$

Origin: Bayes theory

assuming the system state and the observation are random variable we want to maximise the probability of $\mathbf{x}^a$ knowing the observations. , i.e.:

$$P(X = \mathbf{x} | Y = \mathbf{y}^o) = P_{X|Y}(\mathbf{x}|\mathbf{y}^o)$$

thanks to the Bayes theorem, we know:

$$P_{X|Y}(\mathbf{x}|\mathbf{y}^o) = \frac{P_{Y|X}(\mathbf{y}^o|\mathbf{x})P_X(\mathbf{x})}{P_Y(\mathbf{y}^o)}$$

- $P_{Y|X}(\mathbf{y}^o|\mathbf{x})$: likelihood function
- $P_X(\mathbf{x})$: prior distribution
- $P_{X|Y}(\mathbf{x}|\mathbf{y}^o)$: posterior distribution

The marginal $P_Y(\mathbf{y}^o)$ does not depend on the choice of $\mathbf{x}$, so maximising $P_{X|Y}(\mathbf{x}|\mathbf{y}^o)$ is equivalent to maximising $P_{Y|X}(\mathbf{y}^o|\mathbf{x})P_X(\mathbf{x})$.

Origin: Bayes theory

If we assume that

$$\mathbf{y}^o = H(\mathbf{x}) + \varepsilon \text{ and } \mathbf{x} = \mathbf{x}^b + \nu$$

with

$$\varepsilon \sim N(0, \mathbf{R}) \text{ and } \nu \sim N(0, \mathbf{B}),\ \varepsilon \text{ and } \nu \text{ uncorrelated}$$

therefore

$$\mathbf{y}^o|\mathbf{x} \sim N(H(\mathbf{x}), \mathbf{R}) \text{ and } \mathbf{x} \sim N(\mathbf{x}^b, \mathbf{B})$$

Origin: Bayes theory

If we assume that

$$\mathbf{y}^o = H(\mathbf{x}) + \varepsilon \text{ and } \mathbf{x} = \mathbf{x}^b + \nu$$

with

$$\varepsilon \sim N(0, \mathbf{R}) \text{ and } \nu \sim N(0, \mathbf{B}),\ \varepsilon \text{ and } \nu \text{ uncorrelated}$$

therefore

$$\mathbf{y}^o|\mathbf{x} \sim N(H(\mathbf{x}), \mathbf{R}) \text{ and } \mathbf{x} \sim N(\mathbf{x}^b, \mathbf{B})$$

then

$$P_{Y|X}(\mathbf{y}^o|\mathbf{x}) = (2\pi^m|R|)^{1/2} exp\left[-\frac{1}{2}(\mathbf{y}^o - H(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}^o - H(\mathbf{x}))\right]$$

and

$$P_X(\mathbf{x}) = (2\pi^n|B|)^{1/2} exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b)\right]$$

finally

$$P_{X|Y}(\mathbf{x}|\mathbf{y}^o) \propto exp\left[-\frac{1}{2}(\mathbf{y}^o - H(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}^o - H(\mathbf{x}))\right] exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b)\right]$$

3D-Var

by taking minus the log we obtains the so-called 3D-Var cost function:

$$J(\mathbf{x}) = -log\left(exp\left[-\frac{1}{2}(\mathbf{y}^o - H(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}^o - H(\mathbf{x}))\right] exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{x}^b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^b)\right]\right)$$

$$= \frac{1}{2}(\mathbf{y}^o - H(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}^o - H(\mathbf{x})) + \frac{1}{2}(\mathbf{x}-\mathbf{x}^b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^b)$$

and maximizing $P_{X|Y}(\mathbf{x}|\mathbf{y}^o)$ is equivalent to minimizing $J(\mathbf{x})$

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

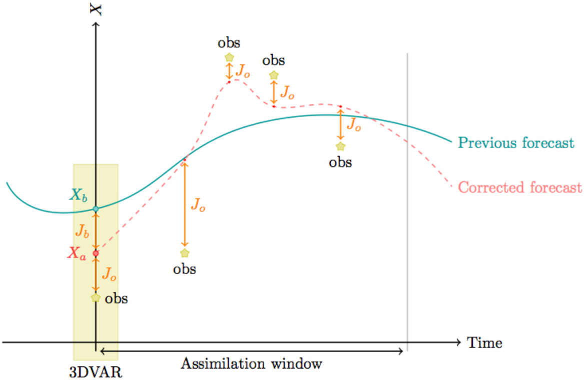
- Observations are distributed in time: $\mathbf{y} = \mathbf{y}(t)$
- The observation cost function becomes:

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

- There is a model describing the evolution of $\mathbf{x}$: $\mathbf{x}_{t_i} = M_{t_{i-1},t_i}(\mathbf{x}_{t_{i-1}})$ 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?

Linear case

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

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

Uniqueness of the minimum?

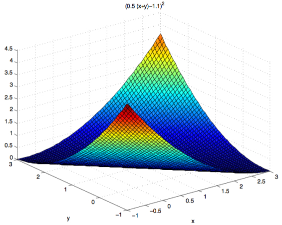
Linear case

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

- 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?

Linear case

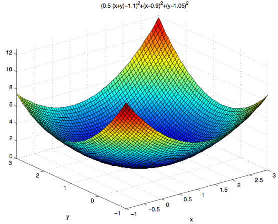
$$J(\mathbf{x}_0) = J_b(\mathbf{x}_0) + J_o(\mathbf{x}_0) = \frac{1}{2}\|\mathbf{x}_0 - \mathbf{x}_b\|_b^2 + \frac{1}{2}\sum_{i=0}^{N}\|H_i(M_{0\to t_i}(\mathbf{x}_0)) - \mathbf{y}(t_i)\|_o^2$$

- 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 $(x_1^b, x_2^b) = (0.9, 1.05)$

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

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



Uniqueness of the minimum?

Non linear case

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

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

Uniqueness of the minimum?

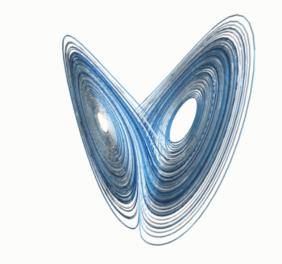
Non linear case

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

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

Example: the Lorenz system (1963)

$$\begin{cases} \dfrac{\mathrm{d}x}{\mathrm{d}t} = \alpha(y - x) \\ \dfrac{\mathrm{d}y}{\mathrm{d}t} = \beta x - y - xz \\ \dfrac{\mathrm{d}z}{\mathrm{d}t} = -\gamma z + xy \end{cases}$$



Uniqueness of the minimum?

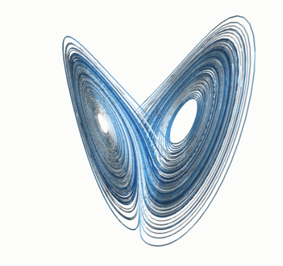
Non linear case

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

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

Example: the Lorenz system (1963)

$$\begin{cases} \dfrac{\mathrm{d}x}{\mathrm{d}t} = \alpha(y-x) \\ \dfrac{\mathrm{d}y}{\mathrm{d}t} = \beta x - y - xz \\ \dfrac{\mathrm{d}z}{\mathrm{d}t} = -\gamma z + xy \end{cases}$$



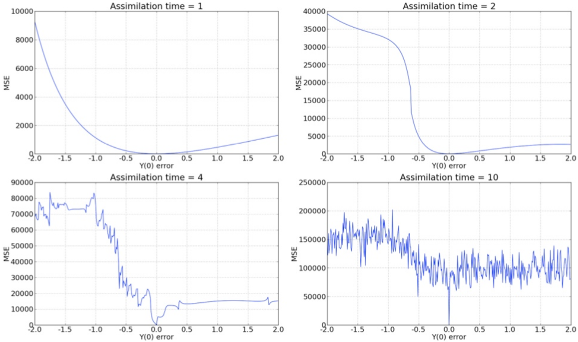
$$J_o(y_0) = \frac{1}{2}\sum_{i=0}^{N}(x(t_i) - x_{\text{obs}}(t_i))^2 \; dt$$

Uniqueness of the minimum?

Non linear case

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

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

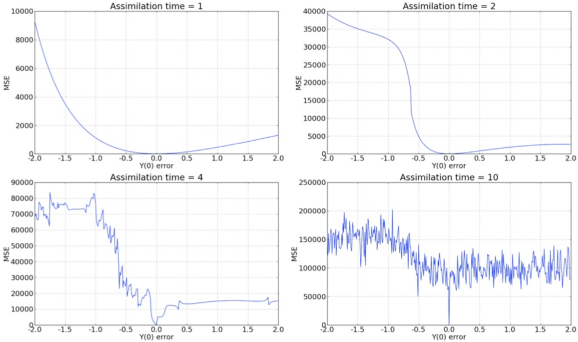


Uniqueness of the minimum?

Non linear case

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

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



- 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.

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).

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:

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

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":

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

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

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

Outline

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 $\mathbf{Y}$

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**:

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

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
  - $\mathrm{Cov}(\varepsilon) = E(\varepsilon\varepsilon^T) = \mathbf{R}$ known accuracies and covariances

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
  - $\mathrm{Cov}(\varepsilon) = E(\varepsilon\varepsilon^T) = \mathbf{R}$ — known accuracies and covariances

BLUE:

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

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

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
  - $\mathrm{Cov}(\varepsilon) = E(\varepsilon\varepsilon^T) = \mathbf{R}$ — known accuracies and covariances

BLUE:

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

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

Remark:

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

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

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 $\mathbf{AH} = \mathbf{I}_n$

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

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

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

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

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

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

This is again the general BLUE framework, with

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

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 $\left[\mathrm{Cov}(\hat{\mathbf{X}})\right]^{-1} = \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$ 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 $\mathrm{Cov}(\hat{\mathbf{X}}) = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}$

go to variational approach

Variational approach in the linear case

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

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

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

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

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

Same remarks as previously

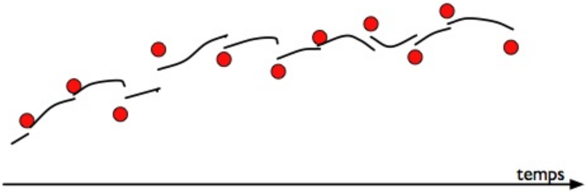
- The statistical approach rationalizes the choice of the norms for J_o and J_b in the variational approach.
- $\underbrace{\left[\mathrm{Cov}(\hat{\mathbf{X}})\right]^{-1}}_{\text{accuracy}} = \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} = \underbrace{\mathrm{Hess}(J)}_{\text{convexity}}$

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)

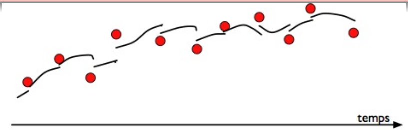
Initialization:
$$\begin{array}{rcll} \mathbf{x}^a(t_0) & = & \mathbf{x}_0 & \textit{approximate initial state} \\ \mathbf{P}^a(t_0) & = & \mathbf{P}_0 & \textit{error covariance matrix} \end{array}$$

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*.



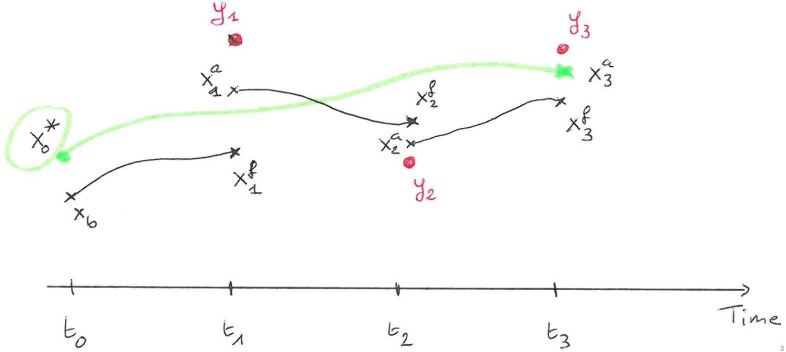
If the problem is time dependent

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

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

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)

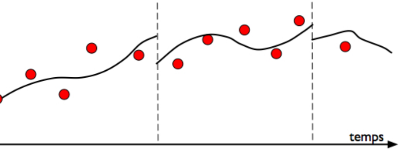
- hypotheses on the pdfs: Bayesian approach (pdf) + ML or MAP estimator

In summary

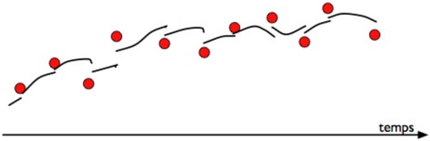
The statistical approach gives a rationale for the choice of the norms, and gives an estimation of the uncertainty.

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

Further generalisation

Considering an imperfect model (e.g. $\mathbf{x}_t = M_{t-1,t}(\mathbf{x}_t) + \eta_t$ with $\eta_t \sim N(0, \mathbf{Q}_t)$ leads to

$$J(\mathbf{x}_0, \ldots, \mathbf{x}_N) = \frac{1}{2}\sum_{t=1}^{N}(\mathbf{y}_t^o - H_t(\mathbf{x}_t))^T \mathbf{R}_t^{-1}(\mathbf{y}_t^o - H_t(\mathbf{x}_t)) + \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b)$$

$$+ \sum_{t=1}^{N}(\mathbf{x}_t - M_{t-1,t}(\mathbf{x}_{t-1}))^T \mathbf{Q}_t^{-1}(\mathbf{x}_t - M_{t-1,t}(\mathbf{x}_{t-1}))$$

If we note the model state trajectory:

$$\mathbf{x} = (\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_{N-1}, \mathbf{x}_N)^T$$

N being the number of time steps per assimilation window, the full variational data assimilation scheme can be defined by the minimization of :

cost function

$$J(\mathbf{x}) = \overbrace{\frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b)}^{J^b} + \overbrace{\frac{1}{2}(\mathbf{y} - H(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y} - H(\mathbf{x}))}^{J^o} + \overbrace{\frac{1}{2}F(\mathbf{x})^T \mathbf{Q}^{-1} F(\mathbf{x})}^{J^q}$$

Where **B** and **R** are the background and observation error correlation matrices respectively, **y** is the observation vector, and H the observation operator, and F represents the remaining theoretical knowledge after background information has been accounted for (basically the model).

$$F_i(\mathbf{x}) = \mathbf{x}_i - M_i(\mathbf{x}_{i-1})$$

Further generalisation: weak constraint 4D-Var

cost function

$$J(\mathbf{x}) = \overbrace{\frac{1}{2}(\mathbf{x}-\mathbf{x}^b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^b)}^{J^b} + \overbrace{\frac{1}{2}(\mathbf{y}-H(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}-H(\mathbf{x}))}^{J^o} + \overbrace{\frac{1}{2}F(\mathbf{x})^T\mathbf{Q}^{-1}F(\mathbf{x})}^{J^q}$$

The meaning of this cost function is that we seek for a state trajectory $\mathbf{x}$ that is satisfying the background error statistics (J^b), that is not far from the observation (J^o) and that it follows (weakly) the model equations (J^q)

In practice this algorithm is not doable for large time dependent problems. The size of $\mathbf{x}$ (the size of the state vector times the number of time step of the assimilation window) becomes huge and the definition and handling of $\mathbf{B}$ is more than problematic

Therefore, in practice, additional hypothesis or approximations have to be made in order to implement variational data assimilation.

Back to 4D-Var:

In the so-called 4D-Var algorithm, the assumption is now that the model is perfect, meaning $\mathbf{x}$ is fully determined by the initial condition $\mathbf{x}_0$ (*e.g.* $J^q = 0$). It is also called *strong constraint 4D-Var*, meaning that the model is a strong constrain of the minimization process.

The cost function is then

$$J(\mathbf{x}_0) = J^b(\mathbf{x}_0) + J^o(\mathbf{x}_0) \tag{1}$$

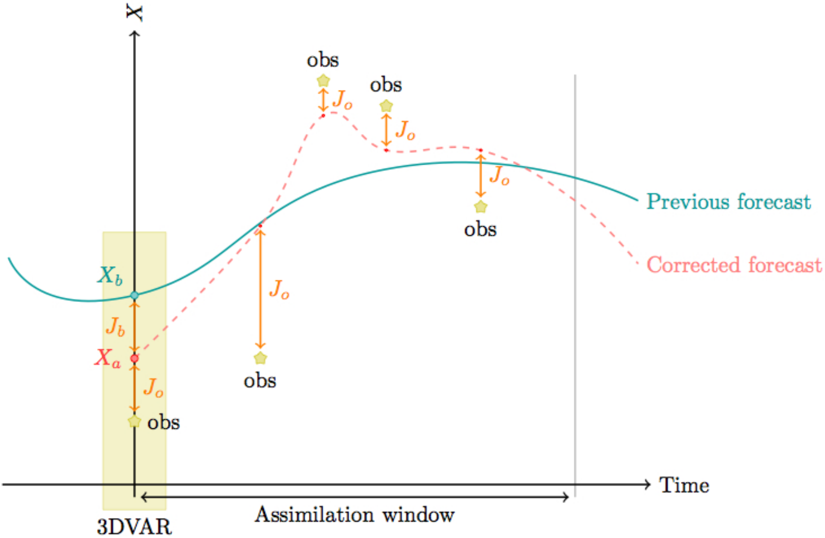
where the background term J^b is the same as before:

$$J^b(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b)$$

The background $\mathbf{x}_0^b$, as $\mathbf{x}_0$, is one possible state vector at initial time $i = 0$.
The observation term J^o is a bit more complex:

$$J^o(\mathbf{x}_0) = \frac{1}{2}\sum_{i=0}^{n}(\mathbf{y}_i^o - H_i(\mathbf{x}_i))^T \mathbf{R}_i^{-1}(\mathbf{y}_i^o - H_i(\mathbf{x}_i))$$

4D-Var:


X
obs
obs
J_o
J_o
J_o
Previous forecast
obs
Corrected forecast
X_b
J_o
J_b
X_a
obs
J_o
obs
Time
Assimilation window
3DVAR