Variational approach to data assimilation: optimization aspects and adjoint method

Eric Blayo
Univ. Grenoble Alpes and Inria

Objectives

- introduce (once again) the several points of view for data assimilation
- introduce data assimilation as an optimization problem
- discuss the different forms of the objective functions
- discuss their properties w.r.t. optimization
- introduce the adjoint technique for the computation of the gradient

Link with statistical methods: cf lectures by E. Cosme

Variational data assimilation algorithms, tangent and adjoint codes: cf lectures by A. Vidard

Outline

Model problem: least squares approach

Two pieces of information on a single quantity. Which estimation for its true value ? $\longrightarrow$ least squares approach

Model problem: least squares approach

Two pieces of information on a single quantity. Which estimation for its true value ? $\longrightarrow$ least squares approach

Example a prior value $x^b = 19°\text{C}$ and an observation $y = 21°\text{C}$ of the (unknown) present temperature x.

- Let $J(x) = \frac{1}{2}\left[(x - x^b)^2 + (x - y)^2\right]$
- $\text{Min}_x \; J(x)$

Model problem: least squares approach

Two pieces of information on a single quantity. Which estimation for its true value ? $\longrightarrow$ least squares approach

Example a prior value $x^b = 19°$C and an observation $y = 21°$C of the (unknown) present temperature x.

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

Model problem: least squares approach

Two pieces of information on a single quantity. Which estimation for its true value ? $\longrightarrow$ least squares approach

Example a prior value $x^b = 19°\text{C}$ and an observation $y = 21°\text{C}$ of the (unknown) present temperature x.

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

If $\neq$ units: $x^b = 66.2°\text{F}$ and $y = 69.8°\text{F}$

- Let $H(x) = \dfrac{9}{5}x + 32$ **observation operator**

Model problem: least squares approach

Two pieces of information on a single quantity. Which estimation for its true value ? $\longrightarrow$ least squares approach

Example a prior value $x^b = 19°\mathrm{C}$ and an observation $y = 21°\mathrm{C}$ of the (unknown) present temperature x.

- Let $J(x) = \frac{1}{2}\left[(x - x^b)^2 + (x - y)^2\right]$
- $\mathrm{Min}_x\ J(x) \quad \longrightarrow\ x^a = \dfrac{x^b + y}{2} = 20°\mathrm{C}$

If $\neq$ units: $x^b = 66.2°\mathrm{F}$ and $y = 69.8°\mathrm{F}$

- Let $H(x) = \dfrac{9}{5}x + 32$ **observation operator**

  Let $J(x) = \dfrac{1}{2}\left[(H(x) - x^b)^2 + (H(x) - y)^2\right]$

  $\mathrm{Min}_x\ J(x) \quad \longrightarrow x^a = 20^\circ\mathrm{C}$

Model problem: least squares approach

Drawback # 1: if observation units are inhomogeneous

$x^b = 19°\text{C}$ and $y = 69.8°\text{F}$

- $J(x) = \dfrac{1}{2}\left[(x - x^b)^2 + (H(x) - y)^2\right]$

Model problem: least squares approach

Drawback # 1: if observation units are inhomogeneous

$x^b = 19°\mathrm{C}$ and $y = 69.8°\mathrm{F}$

- $J(x) = \frac{1}{2}\left[(x - x^b)^2 + (H(x) - y)^2\right] \quad \longrightarrow x^a = 20.53°\mathrm{C}$

Model problem: least squares approach

Drawback # 1: if observation units are inhomogeneous

$x^b = 19^\circ\text{C}$ and $y = 69.8^\circ\text{F}$

- $J(x) = \dfrac{1}{2}\left[(x - x^b)^2 + (H(x) - y)^2\right] \quad \longrightarrow x^a = 20.53^\circ\text{C}$

$\longrightarrow$ *adding apples and oranges !!*

Model problem: least squares approach

Drawback # 1: if observation units are inhomogeneous

$x^b = 19°\text{C}$ and $y = 69.8°\text{F}$

- $J(x) = \dfrac{1}{2}\left[(x - x^b)^2 + (H(x) - y)^2\right] \quad \longrightarrow x^a = 20.53°\text{C}$

  $\longrightarrow$ *adding apples and oranges !!*

Drawback # 2: if observation accuracies are inhomogeneous

If x^b is twice more accurate than y, one should obtain $x^a = \dfrac{2x^b + y}{3} = 19.67°\text{C}$

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

Model problem: linear statistical approach

Reformulation in a **probabilistic framework**:

- the goal is to find an estimator X^a of the true unknown value x
- x^b and y are realizations of random variables X^b and Y

Model problem: linear statistical approach

Reformulation in a **probabilistic framework**:

- the goal is to find an estimator X^a of the true unknown value x
- x^b and y are realizations of random variables X^b and Y

Let $X^b = x + \varepsilon^b$ and $Y = x + \varepsilon^o$ with

Hypotheses

- $E(\varepsilon^b) = E(\varepsilon^o) = 0$ unbiased background and measurement device
- $\mathrm{Var}(\varepsilon^b) = \sigma_b^2 \qquad \mathrm{Var}(\varepsilon^o) = \sigma_o^2$ known accuracies
- $\mathrm{Cov}(\varepsilon^b, \varepsilon^o) = 0$ independent errors

Model problem: linear statistical approach

One is looking for an estimator (i.e. a r.v.) X^a that is

- **linear**: $X^a = \alpha_b X^b + \alpha_o Y$ (in order to be simple)
- **unbiased**: $E(X^a) = x$ (it seems reasonable)
- **of minimal variance**: $\mathrm{Var}(X^a)$ minimum (optimal accuracy)

$\longrightarrow$ BLUE (Best Linear Unbiased Estimator)

Model problem: linear statistical approach

One is looking for an estimator (i.e. a r.v.) X^a that is

- **linear**: $X^a = \alpha_b X^b + \alpha_o Y$ (in order to be simple)
- **unbiased**: $E(X^a) = x$ (it seems reasonable)
- **of minimal variance**: $\text{Var}(X^a)$ minimum (optimal accuracy)

$\longrightarrow$ BLUE (Best Linear Unbiased Estimator)

Since $X^a = \alpha_b X^b + \alpha_o Y = (\alpha_b + \alpha_o)x + \alpha_b \varepsilon^b + \alpha_o \varepsilon^o$:

- $E(X^a) = (\alpha_b + \alpha_o)x + \alpha_b \underbrace{E(\varepsilon^b)}_{=0} + \alpha_o \underbrace{E(\varepsilon^o)}_{=0} \quad \Longrightarrow \alpha_b + \alpha_o = 1$

Model problem: linear statistical approach

One is looking for an estimator (i.e. a r.v.) X^a that is

- **linear**: $X^a = \alpha_b X^b + \alpha_o Y$ (in order to be simple)
- **unbiased**: $E(X^a) = x$ (it seems reasonable)
- **of minimal variance**: $\text{Var}(X^a)$ minimum (optimal accuracy)

$\longrightarrow$ BLUE (Best Linear Unbiased Estimator)

Since $X^a = \alpha_b X^b + \alpha_o Y = (\alpha_b + \alpha_o)x + \alpha_b \varepsilon^b + \alpha_o \varepsilon^o$:

- $E(X^a) = (\alpha_b + \alpha_o)x + \alpha_b \underbrace{E(\varepsilon^b)}_{=0} + \alpha_o \underbrace{E(\varepsilon^o)}_{=0} \quad \Longrightarrow \alpha_b + \alpha_o = 1$
- $\text{Var}(X^a) = E\left[(X^a - x)^2\right] = E\left[(\alpha_b \varepsilon^b + \alpha_o \varepsilon^o)^2\right] = \alpha_b^2 \sigma_b^2 + (1 - \alpha_b)^2 \sigma_o^2$

$$\frac{\partial}{\partial \alpha_b} = 0 \quad \Longrightarrow \quad \alpha_b = \frac{\sigma_o^2}{\sigma_b^2 + \sigma_o^2}$$

Model problem: linear statistical approach

BLUE

$$X^a = \frac{\dfrac{1}{\sigma_b^2}\, X^b + \dfrac{1}{\sigma_o^2}\, Y}{\dfrac{1}{\sigma_b^2} + \dfrac{1}{\sigma_o^2}}$$

Model problem: linear statistical approach

BLUE

$$X^a = \frac{\dfrac{1}{\sigma_b^2} X^b + \dfrac{1}{\sigma_o^2} Y}{\dfrac{1}{\sigma_b^2} + \dfrac{1}{\sigma_o^2}} = X^b + \underbrace{\frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}}_{\text{gain}} \underbrace{(Y - X^b)}_{\text{innovation}}$$

Model problem: linear statistical approach

BLUE

$$X^a = \frac{\dfrac{1}{\sigma_b^2} X^b + \dfrac{1}{\sigma_o^2} Y}{\dfrac{1}{\sigma_b^2} + \dfrac{1}{\sigma_o^2}} = X^b + \underbrace{\frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}}_{\text{gain}} \underbrace{(Y - X^b)}_{\text{innovation}}$$

Its accuracy: $[\mathrm{Var}(X^a)]^{-1} = \dfrac{1}{\sigma_b^2} + \dfrac{1}{\sigma_o^2}$ accuracies are added

Model problem: linear statistical approach

BLUE

$$X^a = \frac{\dfrac{1}{\sigma_b^2} X^b + \dfrac{1}{\sigma_o^2} Y}{\dfrac{1}{\sigma_b^2} + \dfrac{1}{\sigma_o^2}} = X^b + \underbrace{\frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}}_{\text{gain}} \underbrace{(Y - X^b)}_{\text{innovation}}$$

Its accuracy: $[\mathrm{Var}(X^a)]^{-1} = \dfrac{1}{\sigma_b^2} + \dfrac{1}{\sigma_o^2}$ accuracies are added

Remark: Hypotheses on the two first moments of $\varepsilon^b, \varepsilon^o$ lead to results on the two first moments of X^a.

Model problem: linear statistical approach

Variational equivalence

This is equivalent to the problem:

$$\text{Minimize } J(x) = \frac{1}{2}\left[\frac{(x-x^b)^2}{\sigma_b^2} + \frac{(x-y)^2}{\sigma_o^2}\right]$$

Model problem: linear statistical approach

Variational equivalence

This is equivalent to the problem:

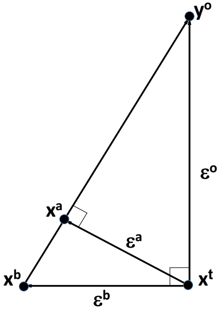
$$\text{Minimize } J(x) = \frac{1}{2}\left[\frac{(x-x^b)^2}{\sigma_b^2} + \frac{(x-y)^2}{\sigma_o^2}\right]$$

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
- $\underbrace{J''(x^a)}_{\text{convexity}} = \frac{1}{\sigma_b^2} + \frac{1}{\sigma_o^2} = \underbrace{[\mathrm{Var}(x^a)]^{-1}}_{\text{accuracy}}$

Model problem: linear statistical approach

Geometric interpretation $E(\varepsilon^o \varepsilon^b) = 0 \implies E(\varepsilon^a (Y - X_b)) = 0$



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

Model problem: Bayesian approach

- x: a realization of a random variable X. What is the pdf $p(X|Y)$?
- Based on the Bayes rule:

$$P(X = x \mid Y = y) = \frac{\overbrace{P(Y = y \mid X = x)}^{\textit{likelihood}}\ \overbrace{P(X = x)}^{\textit{prior}}}{\underbrace{P(Y = y)}_{\textit{normalisation factor}}}$$

Model problem: Bayesian approach

- x: a realization of a random variable X. What is the pdf $p(X|Y)$?
- Based on the Bayes rule:

$$P(X=x\,|\,Y=y)=\frac{\overbrace{P(Y=y\,|\,X=x)}^{\textit{likelihood}}\;\overbrace{P(X=x)}^{\textit{prior}}}{\underbrace{P(Y=y)}_{\textit{normalisation factor}}}$$

- Back to our example:
  - Background $X^b \rightsquigarrow \mathcal{N}(19,\sigma_b^2)$
  - Observation $y=21°\mathrm{C}$, and $Y=X+\varepsilon^o$ with $\varepsilon^o \rightsquigarrow \mathcal{N}(0,\sigma_o^2)$

Model problem: Bayesian approach

- Background $X^b \rightsquigarrow \mathcal{N}(19, \sigma_b^2)$
- Observation $y = 21°\mathrm{C}$, and $Y = X + \varepsilon^o$ with $\varepsilon^o \rightsquigarrow \mathcal{N}(0, \sigma_o^2)$

$$P(X = x \,|\, Y = 21) = \frac{P(Y = 21 \,|\, X = x)\, P(X = x)}{P(Y = y)}$$

Model problem: Bayesian approach

- Background $X^b \rightsquigarrow \mathcal{N}(19, \sigma_b^2)$
- Observation $y = 21°$C, and $Y = X + \varepsilon^o$ with $\varepsilon^o \rightsquigarrow \mathcal{N}(0, \sigma_o^2)$

$$P(X = x \,|\, Y = 21) = \frac{P(Y = 21 \,|\, X = x)\, P(X = x)}{P(Y = y)}$$

- Prior: $P(X = x) = P(X^b = x) = \dfrac{1}{\sqrt{2\pi}\,\sigma_b} \exp\left(\dfrac{(x-19)^2}{2\sigma_b^2}\right)$
- Likelihood:

$$\begin{aligned} p(Y = 21 \,|\, X = x) &= p(\varepsilon^o = 21 - X \,|\, X = x) \\ &= p(\varepsilon^o = 21 - x) \quad \varepsilon^o \text{ is assumed independent from } X \\ &= \frac{1}{\sqrt{2\pi}\,\sigma_o} \exp\left(-\frac{(21-x)^2}{2\,\sigma_o^2}\right) \end{aligned}$$

Model problem: Bayesian approach

- Background $X^b \rightsquigarrow \mathcal{N}(19, \sigma_b^2)$
- Observation $y = 21°\text{C}$, and $Y = X + \varepsilon^o$ with $\varepsilon^o \rightsquigarrow \mathcal{N}(0, \sigma_o^2)$

$$P(X = x \,|\, Y = 21) = \frac{P(Y = 21 \,|\, X = x)\, P(X = x)}{P(Y = y)}$$

- Hence

$$\begin{aligned}
p(X = x)\, p(Y = 21 |\, X = x) \quad &= \quad \frac{1}{\sqrt{2\pi}\sigma_b} \exp\left(- \frac{(x - 19)^2}{2\ \sigma_b^2} \right) \frac{1}{\sqrt{2\pi}\, \sigma_o} \exp\left(- \frac{(21 - x)^2}{2\ \sigma_o^2} \right) \\
&= \quad K\ \exp\left(- \frac{(x - m_a)^2}{2\sigma_a^2} \right) \\
&\quad\ \text{with } m_a = \frac{\frac{1}{\sigma_b^2}\, 19 + \frac{1}{\sigma_o^2}\, 21}{\frac{1}{\sigma_b^2} + \frac{1}{\sigma_o^2}} \quad \text{and} \quad \sigma_a^2 = \left(\frac{1}{\sigma_b^2} + \frac{1}{\sigma_o^2} \right)^{-1}
\end{aligned}$$

Model problem: Bayesian approach

- Background $X^b \rightsquigarrow \mathcal{N}(19, \sigma_b^2)$
- Observation $y = 21°\mathrm{C}$, and $Y = X + \varepsilon^o$ with $\varepsilon^o \rightsquigarrow \mathcal{N}(0, \sigma_o^2)$

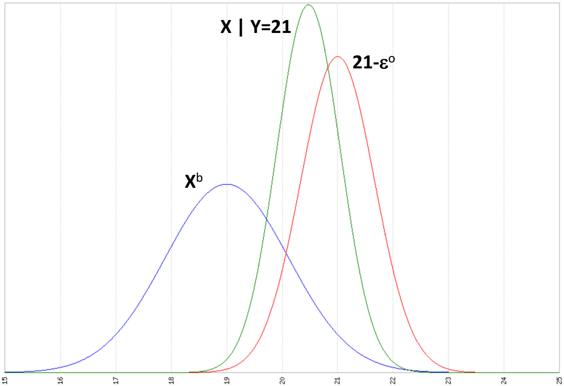
$$P(X = x \,|\, Y = 21) = \frac{P(Y = 21 \,|\, X = x)\, P(X = x)}{P(Y = y)}$$

- Hence

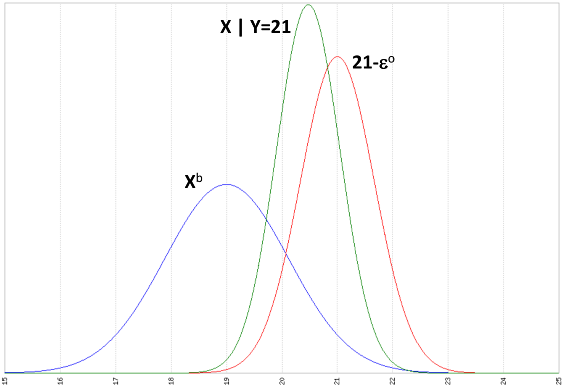
$$\begin{aligned} p(X = x)\, p(Y = 21 |\; X = x) &= \frac{1}{\sqrt{2\pi}\sigma_b} \exp\left(-\frac{(x-19)^2}{2\;\sigma_b^2} \right) \frac{1}{\sqrt{2\pi}\,\sigma_o} \exp\left(-\frac{(21-x)^2}{2\;\sigma_o^2} \right) \\ &= K\;\exp\left(-\frac{(x-m_a)^2}{2\sigma_a^2} \right) \\ &\quad \text{with } m_a = \frac{\frac{1}{\sigma_b^2}\,19 + \frac{1}{\sigma_o^2}\,21}{\frac{1}{\sigma_b^2} + \frac{1}{\sigma_o^2}} \quad \text{and} \quad \sigma_a^2 = \left(\frac{1}{\sigma_b^2} + \frac{1}{\sigma_o^2} \right)^{-1} \end{aligned}$$

$$\longrightarrow\; X \,|\, Y = 21 \;\rightsquigarrow \mathcal{N}(m_a, \sigma_a^2)$$

Model problem: Bayesian approach



Model problem: Bayesian approach



Same as the BLUE because of Gaussian hypothesis

Model problem: synthesis

Data assimilation methods are often split into 2-3 families:

- Variational methods: minimization of a cost function (least squares approach)
- Linear statistical approach: computation of the BLUE (with hypotheses on the first two moments)
- Bayesian approach: approximation of pdfs (with hypotheses on the pdfs)

- There are strong links between those approaches, depending on the case (linear, Gaussian...)

Model problem: synthesis

Data assimilation methods are often split into 2-3 families:

- Variational methods: minimization of a cost function (least squares approach)
- Linear statistical approach: computation of the BLUE (with hypotheses on the first two moments)
- Bayesian approach: approximation of pdfs (with hypotheses on the pdfs)

- There are strong links between those approaches, depending on the case (linear, Gaussian...)

Theorem

If you have understood this previous stuff, you have understood a lot on data assimilation.

Outline

Outline

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

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

then $H(\mathbf{x}) = \mathbf{H}\mathbf{x}$ with $\mathbf{H} =$

Generalization: arbitrary number of unknowns and observations

A simple example of observation operator

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

then $\quad H(\mathbf{x}) = \mathbf{Hx} \quad$ 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: $J(\mathbf{x}) = \dfrac{1}{2} \|H(\mathbf{x}) - \mathbf{y}\|^2$ 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: $J(\mathbf{x}) = \dfrac{1}{2} \|H(\mathbf{x}) - \mathbf{y}\|^2$ 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: $J(\mathbf{x}) = \dfrac{1}{2} \|H(\mathbf{x}) - \mathbf{y}\|^2$ with $\|.\|$ to be chosen.

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

$$p \ \geq n$$

Formalism "background value + new observations"

$$\mathbf{Y} = \begin{pmatrix} \mathbf{x}_b \\ \mathbf{y} \end{pmatrix} \begin{array}{l} \longleftarrow \text{background} \\ \longleftarrow \text{new obs} \end{array}$$

The cost function becomes:

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

Formalism "background value + new observations"

$$\mathbf{Y} = \begin{pmatrix} \mathbf{x}_b \\ \mathbf{y} \end{pmatrix} \begin{array}{l} \longleftarrow \text{background} \\ \longleftarrow \text{new obs} \end{array}$$

The cost function becomes:

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

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

Formalism "background value + new observations"

$$\mathbf{Y} = \begin{pmatrix} \mathbf{x}_b \\ \mathbf{y} \end{pmatrix} \begin{array}{l} \longleftarrow \text{background} \\ \longleftarrow \text{new obs} \end{array}$$

The cost function becomes:

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

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

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

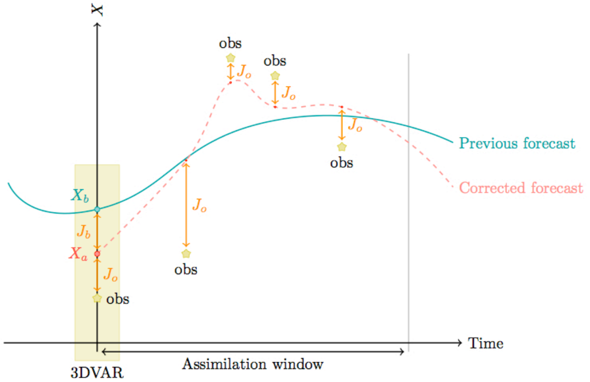
- Observations are distributed in time: $\mathbf{y} = \mathbf{y}(t)$.
- The observation cost function becomes:

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

- 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.

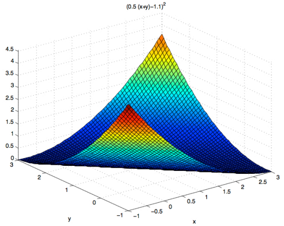
Uniqueness of the minimum ?

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

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

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

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