Data Assimilation Workshop

French School : LeDimet, Talagrant, Courtier (NWP)
Oregon : Bennett

Least squares, Minimum Variance & Maximum Likelihood Estimates

An Example: two independent measurements of $x$ : $x_1, x_2$

- Combine in an optimal way to get an estimate : Form a linear weighted estimate

$$x_a = \underbrace{a_1 x_1 + a_2 x_2}_{\text{unknwon coefficients}} \qquad (1)$$

- Denote the true value of $x$ as $x_t$

$$\varepsilon_1 = x_1 - x_t \; ; \quad \varepsilon_2 = x_2 - x_t$$

Assumptions :

- unbiased, Gaussian, random errors

$$\langle \varepsilon_1 \rangle = \langle \varepsilon_2 \rangle = 0$$

- assume that $x_a$, unbiased :

$$\langle \varepsilon_a \rangle = \langle x_a - x_t \rangle = 0$$

It is easy to show that

$$a_1 + a_2 = 1 \qquad (2)$$

- No correlation between measurement error

$$\langle \varepsilon_1 \varepsilon_2 \rangle = 0$$

but
$$\langle \varepsilon_1^2 \rangle = \sigma_1^2$$
$$\langle \varepsilon_2^2 \rangle = \sigma_2^2$$

$$\therefore \quad \langle \varepsilon_a^2 \rangle = \langle (x_a - x_t)^2 \rangle = a_1^2 \langle \varepsilon_1^2 \rangle + a_2^2 \langle \varepsilon_2^2 \rangle$$
$$= \sigma_a^2$$

$$\sigma_a^2 = a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 \qquad (3)$$

<u>Problem to Solve</u>

Find a minimum variance estimate which minimizes (3) but subject to (2)

Use the method of Lagrange multiplier

$$\mathcal{L} = \sigma_a^2 + \lambda (a_1 + a_2 - 1)$$

↑
unknown Lagrange multiplier

At the extrema of $\mathcal{L}$

$$\frac{\partial \mathcal{L}}{\partial a_1} = 2 a_1 \sigma_1^2 + \lambda = 0$$

$$\frac{\partial \mathcal{L}}{\partial a_2} = 2 a_2 \sigma_2^2 + \lambda = 0$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = a_1 + a_2 - 1 = 0$$

Solve for $a_1$ and $a_2$

$$X_a = \frac{\sigma_1^{-2}}{(\sigma_1^2 + \sigma_2^2)} X_1 + \frac{\sigma_2^{-2}}{(\sigma_1^2 + \sigma_2^2)} X_2 \qquad (4)$$

$$\frac{1}{\sigma_a^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \qquad (5)$$

Generalizing for $M$ observations, $X_n$, with error variances $\sigma_n^2$

$$X_a = \frac{\sum_{n=1}^{M} \sigma_n^{-2} X_n}{\sum_{n=1}^{M} \sigma_2^{-2}}$$

$$\sigma_a^2 = \left[ \sum_{n=1}^{M} \sigma_n^{-2} \right]$$

Four Dimensional ~~Data A~~ Variational Data Assimilation subject to the strong constraint.

(Talagrand and Courtier, 1987: QJRMS, 113, 1311-1328)

Notation:

(1) $\underline{X}$ ROMS state-vector $\equiv (u, v, T, S, \zeta)$
composed of all the ocean grid points values of ROMS

(2) NL ROMS
$$\frac{d\underline{X}}{dt} = \mathcal{N}(\underline{X}) + \underset{\substack{\uparrow \\ \text{external} \\ \text{forcing}}}{f(t)} \qquad (6)$$

Subject to $\underline{X}(0)$ and $\underline{X}_{\Omega}(t)$ ($\Omega$ denotes the boundary)

(3) Observations $\underline{Y}_i$ at observations $t_i$ with observation error covariance $\underline{\underline{O}}$ (= Matrix)

(4) Model equivalent at obs points

$$\underline{\underline{H}}_i\, \underline{X}(t_i) \equiv \underline{\underline{H}}_i\, X_i$$

(5) Unbiased background state $\underline{X}_b$ with background error covariance $\underline{\underline{B}}$. Typically $\underline{X}_b$ will be climatology or previous forecast.

(6) Desired estimate $\underline{x}_a$

Important result from Rabier and Courtier (1992, QJRMS, 118, 649-672)

The minimum variance estimate $\underline{x}_a$ has an error covariance

$$\langle (\underline{x}_a - x_t)(\underline{x}_a - x_t)^T \rangle = \left( \underline{\underline{B}}^{-1} + \sum_{i=1}^{N} \underline{\underline{H}}_i^T \underline{\underline{O}}^{-1} \underline{\underline{H}}_i \right)^{-1} \quad (7)$$

This obtained by ~~[illegible]~~ minimizing

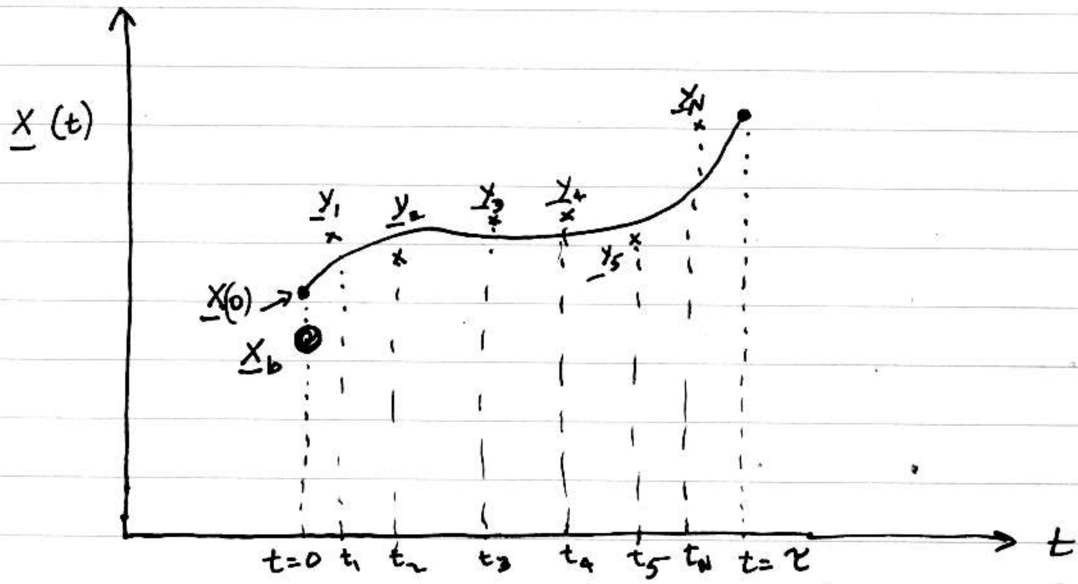
Cost function (Penalty functional)

$$J(\underline{x}) = \frac{1}{2}(\underline{x} - \underline{x}_b)^T \underline{\underline{B}}^{-1} (\underline{x} - \underline{x}_b) + \sum_{i=1}^{N} \frac{1}{2} (\underline{\underline{H}}_i \underline{x}_i - \underline{y}_i)^T \underline{\underline{O}}^{-1} (\underline{\underline{H}}_i \underline{x}_i - \underline{y}_i) \quad (8)$$

subject to the constraint that $\underline{x}(t)$ satisfy (6) exactly.
(The model dynamics are imposed as a "strong constraint"
Sasaki (1970, MWR, 98, 875-883)

Problem:

- Find $\underline{x}(t)$ that minimizes (8) and satisfies (6).
- Solution of (6) depend only on $\underline{x}(0)$, $\underline{x}_\Omega(t)$ and $\underline{f}(t)$
  " control variables "
- For now, assume that $\underline{x}_\Omega(t)$ and $\underline{f}(t)$ are perfectly known, so the control variable i $\underline{x}(0)$




Example:

Assume for simplicity that we have ~~a~~ single observation $\underline{y}$ at time $t=\tau$, and that $\underline{x}_b$ is evaluated at $t=0$.

Then, the cost function

$$J(\underline{x}) = \overbrace{\frac{1}{2}\left(\underline{x}(0)-\underline{x}_b\right)^T \underline{\underline{B}}^{-1}\left(\underline{x}(0)-\underline{x}_b\right)}^{J_b} + \underbrace{\frac{1}{2}\left(\underline{\underline{H}}\,\underline{x}(\tau)-\underline{y}\right)^T \underline{\underline{O}}^{-1}\left(\underline{\underline{H}}\,\underline{x}(\tau)-\underline{y}\right)}_{J_o} \qquad (9)$$

$$J = J_b + J_o$$

Problem :

Minimize (9) subject to (6)

Construct a Lagrange function :

$$\mathcal{L} = J(\underline{x}) + \sum_{i=1}^{M} \underline{\lambda}_i^T \left( \frac{d\underline{x}_i}{dt} - N(\underline{x}_i) - \underline{f}_i \right) \qquad (10)$$

where $t = [0, \tau] \equiv [0, M\Delta t]$

$$\underline{x}_i \equiv \underline{x}(i\Delta t)$$
$$\underline{f}_i \equiv \underline{f}(i\Delta t)$$

and

$\underline{\lambda}(t_i) \equiv \lambda_i \equiv \underline{\lambda}(i\Delta t)$ are unknown Lagrange multiplier

Consider the first-variation of $\mathcal{L}$ associated with changes $\delta\underline{x}(0)$ in $\underline{x}(0)$, $\delta\underline{x}(t)$ in $\underline{x}(t)$ and $\mathcal{L}\underline{\lambda}(t)$ in $\underline{\lambda}(t)$:

$$\delta\mathcal{L} \simeq \delta\underline{x}^T(0)\,\underline{\underline{B}}^{-1}(\underline{x}(0)-\underline{x}_b) + \delta\underline{x}^T(\tau)\,\underline{\underline{H}}^T\underline{\underline{O}}^{-1}(\underline{\underline{H}}\,\underline{x}(t)-\underline{y})$$
$$\text{(A)} \quad + \sum_{i=1}^{M} \underline{\lambda}_i^T \left( \frac{d\,\delta\underline{x}_i}{dt} - \left(\frac{\partial N}{\partial x}\right)_i \delta\underline{x}_i \right) \qquad (11)$$
$$+ \sum_{i=1}^{M} \delta\lambda_i^T \left( \frac{d\underline{x}_i}{dt} - N(x) - f_i \right)$$

Term (A) can be

$$\sum_{i=1}^{M} \frac{d}{dt}\left(\underline{\lambda}_i^T \delta \underline{x}_i\right) - \delta \underline{x}_i^T \frac{d\underline{\lambda}_i}{dt} - \delta \underline{x}_i^T \left(\frac{\partial N^{\rho}}{\partial \underline{x}}\right)_i^T \underline{\lambda}_i =$$

$$\underbrace{\sum_{i=1}^{M} \frac{d}{dt}\left(\underline{\lambda}_i^T \delta \underline{x}_i\right)}_{} + \underbrace{\sum_{i=1}^{M} \delta x_i^T - \frac{d\underline{\lambda}_i}{dt} - \left(\frac{\partial N^{\rho}}{d\underline{x}}\right)_i^T \lambda_i}_{\text{important equation}} \quad (12)$$

TL ROMS: $\frac{d\,\delta x_i}{dt} - \left(\frac{\partial N^{\rho}}{d\underline{x}}\right) \delta \underline{x}_i = 0$ (13)

Tangent linear model

AD ROMS: $-\frac{d\lambda_i}{dt} - \left(\frac{\partial N}{\partial \underline{x}}\right)_i^T \lambda_i = 0$ (14)

Adjoint model

Therefore:

$$\sum_{i=1}^{M} \frac{d}{dt}\left(\underline{\lambda}_i^T \delta \underline{x}_i\right) \equiv \int_0^{\tau} \frac{d}{dt}\left(\underline{\lambda}^T \delta \underline{x}\right) dt \equiv \delta \underline{x}^T(\tau)\, \underline{\lambda}(\tau) - \delta \underline{x}^T(0)\, \underline{\lambda}(0)$$

So (11) becomes

$$\delta \mathcal{L} \simeq \delta^T \underline{x}(0)\, \underline{\underline{B}}^{-1} \left(\underline{x}(0) - \underline{x}_b\right) + \delta \underline{x}^T(\tau)\, \lambda(\tau) - \delta^T \underline{x}(0)\, \underline{\lambda}(0)$$

$$+ \sum_{i=1}^{M} \delta \underline{x}_i^T \left(-\frac{d\underline{\lambda}_i}{dt} - \left(\frac{\partial N^{\rho}}{\partial \underline{x}}\right)^T \underline{\lambda}_i + \delta_{i,M}\, \underline{\underline{H}}^T \underline{\underline{O}}^{-1} \left(\underline{\underline{H}}\, \underline{x}_M - \underline{y}\right)\right)$$

$$+ \sum_{i=1}^{M} \delta \lambda_i^T \left(\frac{d\underline{x}_i}{dt} - N^{\rho}(\underline{x}) - \underline{f}_i\right)$$

Kronecker delta

At the extrema of $\mathcal{L}$, we require

$$\frac{\partial \mathcal{L}}{\partial \underline{x}(0)} = 0 \; ; \quad \frac{\partial \mathcal{L}}{\partial \underline{x}_i} = 0 \; ; \quad \frac{\partial \mathcal{L}}{\partial \lambda_i} = 0 \; ; \quad \frac{\partial \mathcal{L}}{\partial \underline{x}(\tau)} = 0$$

$$\lim_{\delta \lambda_i \to} \left( \frac{\delta \mathcal{L}}{\delta \underline{\lambda}_i} \right) = \left( \frac{\partial \mathcal{L}}{\partial \underline{\lambda}_i} \right) = \underbrace{\left( \frac{d \underline{x}_i}{dt} - N(\underline{x}_i) - f_i \right)}_{\text{NL ROMS}} = 0 \qquad (15)$$

$$\lim_{\delta \underline{x}_i \to 0} \left( \frac{\delta \mathcal{L}}{\delta \underline{x}_i} \right) = \left( \frac{\partial \mathcal{L}}{\partial \underline{x}_i} \right) = \underbrace{\left( - \frac{d \lambda_i}{dt} - \left( \frac{\partial N}{d \underline{x}} \right)_i^T \underline{\lambda}_i + \delta_{i,M} \underline{\underline{H}}^T \underline{\underline{O}}^{-1} \left( \underline{\underline{H}} \underline{x}_M - \underline{y} \right) \right)}_{\text{AD ROMS forced by } - \delta_{i,M} \underline{\underline{H}}^T \underline{\underline{O}}^{-1} \left( \underline{\underline{H}} \underline{x}_M - \underline{y} \right)} = 0 \qquad (16)$$

$$\lim_{\delta \underline{x}(0) \to 0} \left( \frac{\delta \mathcal{L}}{\delta \underline{x}(0)} \right) = \left( \frac{\partial \mathcal{L}}{\partial \underline{x}(0)} \right) = \underline{\underline{B}}^{-1} \left( \underline{x}(0) - \underline{x}_b \right) - \underline{\lambda}(0) = 0 \qquad (17)$$

$$\lim_{\delta \underline{x}(\tau) \to 0} \left( \frac{\delta \mathcal{L}}{\delta \underline{x}(\tau)} \right) = \left( \frac{\partial \mathcal{L}}{\partial \underline{x}(\tau)} \right) = \underline{\lambda}(\tau) = 0 \qquad (18)$$

Consider now

$$\delta J(\underline{x}) \simeq \delta \underline{x}^T(0) \, \underline{\underline{B}}^{-1} \left( \underline{x}(0) - \underline{x}_b \right) + \delta \underline{x}^T(\tau) \, \underline{\underline{H}}^T \underline{\underline{O}}^{-1} \left( \underline{\underline{H}} \, \underline{x}(\tau) - \underline{y} \right)$$

Solutions of (13) can be written as

$$\delta \underline{x}(\tau) = \underline{\underline{R}}(0, \tau) \, \delta \underline{x}(0)$$

Similarly, solutions of AD ROMS (16) can be written as:

$$\underline{\lambda}(0) = \underline{\underline{R}}^T(\tau,0)\,\underline{\lambda}(\tau)$$

Therefore

$$\delta J(\underline{x}) \simeq \delta\underline{x}^T(0)\,\underline{\underline{B}}^{-1}\left(\underline{x}(0)-\underline{x}_b\right) + \delta\underline{x}^T(0)\,\underline{\underline{R}}^T(\tau,0)\,\underline{\underline{H}}^T\underline{\underline{O}}^{-1}\left(\underline{\underline{H}}\,\underline{x}(\tau)-\underline{y}\right)$$

$$\frac{\partial J}{\partial x(0)} = \underline{\underline{B}}^{-1}\left(\underline{x}(0)-\underline{x}_b\right) + \underbrace{\underline{\underline{R}}^T(\tau,0)\,\underline{\underline{H}}^T\underline{\underline{O}}^{-1}\left(\underline{\underline{H}}\,x(\tau)-\underline{y}\right)}_{-\lambda(0)}$$

$$\boxed{\frac{\partial \mathcal{L}}{\partial \underline{x}(0)} = \frac{\partial J}{\partial \underline{x}(0)}}$$

So what do we need to do?

(1) Solve (15)

$$\left.\begin{array}{ll} \dfrac{d\underline{x}}{dt} = N(\underline{x}) + \underline{f}(t) & \text{NL ROMS} \\ \text{subject to } \underline{x}(0),\ \underline{x}_{\partial\Omega}(t) & \text{for } t=[0,\tau] \end{array}\right\} \rightarrow J(\underline{x})$$

(2) Solve (1b)

$$-\frac{d\lambda}{dt} = \left(\frac{\partial N}{\partial \underline{x}}\right)^T \lambda - \delta_{i,M} \underline{\underline{H}}^T \underline{\underline{O}}^{-1} (\underline{\underline{H}}\, \underline{x}_M - \underline{y})$$

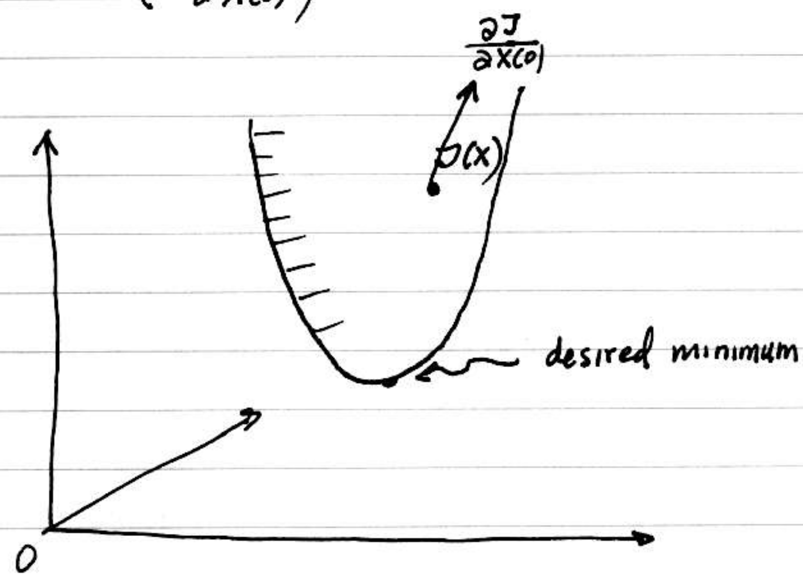
subject to $\underline{\lambda}(\tau) = 0$ and $\underline{\lambda}_\Omega(t)$ for $t = [\tau, 0]$

(3) compute $\frac{\partial J}{\partial X(0)} = \underline{\underline{B}}^{-1}(\underline{X}(0) - \underline{X}_b) - \underline{\lambda}(0)$

$$= \frac{\partial J_b}{\partial \underline{x}(0)} + \frac{\partial J_o}{\partial \underline{X}(0)}$$

→ AD ROMS forced by $-\delta_{i,M} \underline{\underline{H}}^T \underline{\underline{O}}^{-1} (\underline{\underline{H}}\, \underline{X}_M - \underline{Y})$

yield $\left(\frac{\partial J_o}{\partial X(0)}\right)$



The basic strong constraint 4-dimensional variational data assimilation algorithm (# ifdef S4DVAR)

(0) Choose an $\underline{x}(0)$ (e.g. $\underline{x}(0) = \underline{x}_b$)

(1) Integrate NL ROMS, $t = [0, \tau]$, and compute $J$.

(2) Integrate AD ROMS, $t = [\tau, 0]$, forced by

$$-\underline{\underline{H}}^T \underline{\underline{O}}^{-1} (\underline{\underline{H}}\,\underline{x}(\tau) - \underline{y}) \text{ to yield } \left(\frac{\partial J}{\partial \underline{x}(0)}\right) = -\underline{\lambda}(0)$$

Step (3)

(3) Use a descent algorithm to determine a "down gradient" correction to $\underline{x}(0)$ that yield a smaller value of $J(\underline{x})$ } the most critical step

Conjugate gradient Descent:

Two central components to this

(1) Step size determination: How far down gradient do we go?

(2) Preconditioning (i.e. modify the shape of the cost function to increase the efficiency of the algorithm).

Using the basic idea of conjugate gradients method we can expand step (3) as:

(i) Let $\underline{x}_n(0)$ be the current NLROMS initial condition for iteration $n$.

(ii) The new initial condition is given by

$$\underline{x}_{n+1}(0) = \underline{x}_n(0) + \alpha \underline{d}_n \qquad (19)$$

where

$\underline{d}_n$ : descent direction

$\alpha$ : step size, scalar, positive

(iii) The $\underline{d}_n$ are determined by the sequence:

$$\underline{d}_n = - \left(\frac{\partial J}{\partial x(0)}\right)_n + \gamma_{n-1} \underline{d}_{n-1} \qquad (20)$$

where

$$\gamma_{n-1} = \frac{\left(\left(\frac{\partial J}{\partial x(0)}\right)_n - c\left(\frac{\partial J}{\partial \underline{x}(0)}\right)\right)^T \left(\frac{\partial J}{\partial \underline{x}(0)}\right)_n}{\left(\frac{\partial J}{\partial x(0)}\right)_{n-1}^T \left(\frac{\partial J}{\partial \underline{x}(0)}\right)_{n-1}} \qquad (21)$$

(iv) The sequence of search directions is constructed so that each new search direction is orthogonal (conjugate) to all previous search directions

(v) Two variants of this basic algorithms current employed in ROMS

$c = 0$ Fletcher-Reeves algorithm.

$c = 1$ Polack-Rivere algorithm

(vi) In practice, the conjugate nature of the search direction breaks down and becomes less efficient.

→ "restart" using stepest descent $\gamma = 0$ every so often

## Line Searching

- Once we have $\underline{d}_n$, how far down gradient we go? i.e. how big is $\alpha$?

- Let's assume that everything (i.e. NL ROMS) is linear
→ we can estimate an optimal $\alpha$ from some trial $\alpha$

(a) choose arbitrary step size $\alpha$, and compute the new $\underline{x}(0)$

$$\underline{x}_n(0) = \underline{x}_{n-1}(0) + \underbrace{\alpha \underline{d}_n}_{\delta x(0)}$$

(b) The corresponding change $\delta J$ in $J$ is

$$\delta J = \delta \underline{x}^T(0)\, \underline{\underline{B}}^{-1} (\underline{x}(0) - x_b) + \delta \underline{x}^T(0)\, \underline{\underline{H}}^T \underline{\underline{O}}^{-1} (\underline{\underline{H}}\, x(\tau) - \underline{y})$$
$$+ \tfrac{1}{2} \delta \underline{x}^T(0)\, \underline{\underline{B}}^{-1} \delta \underline{x}(0) + \tfrac{1}{2} \delta x^T(\tau)\, \underline{\underline{H}}^T O^{-1} \delta \underline{x}(\tau)$$

c) If the system is linear, then any othe choice $\alpha_r$ of step size will yield a $\delta J$ given by:

$$\delta J_r = \frac{\alpha_r}{\alpha}\, \delta \underline{x}^T(0)\, \underline{\underline{B}}^{-1}(\underline{x}(0)-\underline{x}_b) + \frac{\alpha_r}{\alpha}\, \delta \underline{x}^T(\tau)\, \underline{\underline{H}}^T \underline{\underline{O}}^{-1}(\underline{\underline{H}}\,\underline{x}(\tau)-\underline{y})$$

$$+ \frac{\alpha_r^2}{\alpha^2}\cdot\frac{1}{2}\, \delta \underline{x}^T(0)\, \underline{\underline{B}}^{-1}\, \delta \underline{x}(0) + \frac{\alpha_r^2}{\alpha^2}\, \delta \underline{x}^T(\tau)\, \underline{\underline{H}}^T \underline{\underline{O}}^{-1} \underline{\underline{H}}\, \delta \underline{x}(\tau)$$

d) The optimal choice of $\alpha_r$, is that which produces the biggest change in $\delta J_r$ that lies on the surface $J(\underline{x})$

$$\frac{\partial\, \delta J_r}{\partial \alpha_r} = 0$$

$$\Rightarrow \quad \alpha_r = -\alpha\, \frac{\left[\delta \underline{x}^T(0)\, \underline{\underline{B}}^{-1}(\underline{x}(0)-\underline{x}_b) + \delta \underline{x}^T(\tau)\, \underline{\underline{H}}^T \underline{\underline{O}}^{-1}(\underline{\underline{H}}\,\underline{x}(\tau)-\underline{y})\right]}{\left[\delta \underline{x}^T(0)\, \underline{\underline{B}}^{-1}\, \delta \underline{x}(0) + \delta \underline{x}^T(\tau)\, \underline{\underline{H}}^T \underline{\underline{O}}^{-1} \underline{\underline{H}}\, \delta \underline{x}(\tau)\right]} \qquad (22)$$

Long & Thacker, 1989, DAO, 13, 413-440

(i)

(ii)

(iii) Based on $\delta \underline{x}(0)$ and $\delta \underline{x}($ at obs points, compute $\alpha_r$ using (22)

(iv) compute the new TL ROMS initial conditions

$$\underline{x}_n(0) = \underline{x}_{n-1}(0) + \alpha_r \underline{d}_n$$

The full ROMS S4DVAR procedure is:

(0) Choose an $\underline{X}_0(0)$ (e.g. $\underline{X}_0(0) = \underline{X}_b$)

(1) Integrate NL ROMS, $t = [0, \tau]$ and compute $J$ from all available obs in the interval

(2) Integrate AD ROMS, $t = [\tau, 0]$, forced by all available obs in the interval $\rightarrow \left(\partial J_0 / \partial \underline{X}(0)\right)$

(3) Compute $\partial J / \partial X(0) = \underline{\underline{B}}^{-1}\left(\underline{X}(0) - \underline{X}_b\right) + \left(\partial J_0 / \partial \underline{X}(0)\right)$

(4) Compute new descent direction:

$$\underline{d}_n = -\left(\frac{\partial J}{\partial \underline{X}(0)}\right) + \gamma_{n-1}\, \underline{d}_{n-1}$$

where $\gamma_0 = 0$ (i.e. steepest descent, 1st time)
and $\gamma_n = 0$ for $\mathrm{mod}(n, N) = 0$, $N \sim 5$
otherwise use (21)

5) Choose as trial step size $\alpha$ (usually choose $\alpha_r$ from previous iteration) and construct a trial initial condition

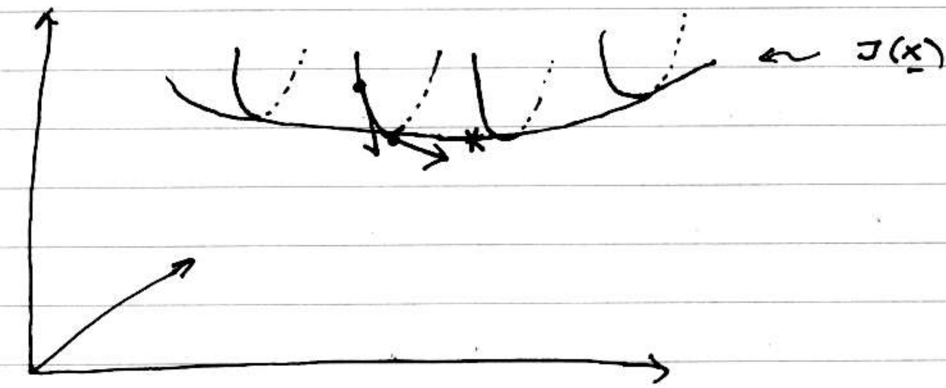
$$\underline{X}_n(0) = \underline{X}_{n-1}(0) + \alpha\, \underline{d}_n$$

(6) Integrate NL ROMS again and compute optimal step size using (22)

(7) compute the initial condition

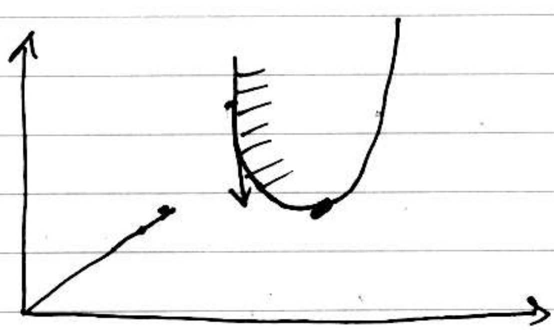
$$\underline{X}_n(0) = \underline{X}_{n-1}(0) + \alpha_r\, \underline{d}_n$$

# Preconditioning of the Gradient Vector




very slow convergence

We can transform its shape via pre-conditioning



How do we do this transformation?

- Consider as an exemple

$$J(\underline{x}) = \frac{1}{2}\, \underline{x}^T \underline{\underline{E}}\, \underline{x}$$

$\underline{\underline{E}}$ : symmetric matrix

- Introduce a new variable $\underline{z} = \underline{\underline{E}}^{1/2} \underline{x}$

$$J(\underline{z}) = \frac{1}{2}\, \underline{z}^T \underline{z}$$

- $\underline{\underline{E}}$ is the second derivative of $J(x)$ with respect $\underline{x}$ (Hessian matrix).

- For $J(\underline{x})$ given by (9), the Hessian matrix

$$J''(\underline{x}) = \underline{\underline{B}}^{-1} + \underline{\underline{H}}^T \underline{\underline{O}}^{-1} \underline{\underline{H}}$$

Since $\underline{x} \sim 10^6$ cannot built it, and take its inverse.

- An important practical consideration:

  - The behavior of $J(\underline{x})$ is dominated by $\underline{\underline{B}}$

  - A good approximation for the Hessian is $\underline{\underline{B}}^{-1}$.

Incremental strong Constraint 4DVAR
(#ifdef IS4DVAR)

- Courtier et al. (1994: QJRMS, 120, 1367-1387)
- Weaver et al. (2003: MWR, 131, 1360-1378)

- In the incremental formulation of 4DVAR

$$\underline{x} = \underline{x}_b + \delta\underline{x} \qquad (24)$$

"close to truth"

- Denote solutions of NL ROMS in (6) as

$$\underline{x}(t_i) = \mathcal{M}(t_i, t_{i-1})\left(\underline{x}(t_{i-1})\right) \qquad (25)$$

where $\mathcal{M}(t_i, t_{i-1})$ represents NL ROMS acting on $\underline{x}(t_{i-1})$ between $t = [t_{i-1}, t_i]$

Combine (24) and (25)

$$\underline{x}(t_i) = \mathcal{M}(t_i, t_{i-1})\left(\underline{x}_b(t_{i-1}) + \delta\underline{x}(t_{i-1})\right)$$

$$= \mathcal{M}(t_i, t_{i-1})\,\underline{x}_b(t_{i-1}) + \mathcal{M}(t_i, t_{i-1})\,\delta\underline{x}(t_{i-1})$$

$$\simeq \mathcal{M}(t_i, t_{i-1})\,\underline{x}_b(t_{i-1}) + \underbrace{\underline{\underline{R}}(t_{i-1}, t_i)}_{\text{TL ROMS propagator}}\,\delta\underline{x}(t_{i-1}) \qquad (26)$$

- Consider again the cost function (8):

$$J(\underline{x}) = \overbrace{\tfrac{1}{2}\left(\underline{x}(0) - \underline{x}_b(0)\right)^T \underline{\underline{B}}^{-1}\left(\underline{x}(0) - \underline{x}_b(0)\right)}^{J_b}$$

$$+ \underbrace{\sum_{i=1}^{N} \tfrac{1}{2}\left(\underline{\underline{H}}_i\,\underline{x}(t_i) - \underline{y}_i\right)^T \underline{\underline{O}}^{-1}\left(\underline{\underline{H}}_i\,\underline{x}(t_i) - \underline{y}_i\right)}_{J_o}$$

Using (24) - (26)

$$J(\underline{x}) \simeq \tfrac{1}{2}\,\delta\underline{x}^T(0)\,\underline{\underline{B}}^{-1}\,\delta\underline{x}(0)$$

$$+ \tfrac{1}{2}\sum_{i=1}^{N}\left(\underline{\underline{H}}_i\mathcal{M}(t_i,0) + \underline{\underline{H}}_i\,\underline{\underline{R}}(0,t_i)\,\delta\underline{x}(0) - y_i\right)^T \underline{\underline{O}}^{-1}\left(H_i\mathcal{M}(t_i,0)\,\underline{x}_b(0) + R(0,t_i)\,\delta\underline{x}(0) - y_i\right)$$

$$\simeq \tfrac{1}{2}\,\delta\underline{x}^T(0)\,\underline{\underline{B}}^{-1}\,\delta\underline{x}(0) + \tfrac{1}{2}\sum_{i=1}^{N}\left(\underline{\underline{G}}_i\,\delta\underline{x}(0) - d_i\right)^T \underline{\underline{O}}^{-1}\left(\underline{\underline{G}}_i\,\delta\underline{x}(0) - d_i\right) \qquad (27)$$

where $\underline{\underline{G}}_i = \underline{\underline{H}}_i\,\underline{\underline{R}}(0, t_i)$

$$\underline{d}_i = \underline{y}_i - \underline{\underline{H}}_i\,\mathcal{M}(t_i, 0)\,\underline{x}_b(0)$$

$$= \underline{y}_i - \underline{\underline{H}}_i\,\underline{x}_b(t_i)$$

- $J(\underline{x})$ is now quadratic in $\delta\underline{x}(0)$
- Use TLROMS to advance $\delta\underline{x}$ in time

The basic IS4DVAR algorithm

(0) Choose $\underline{x}(0) = \underline{x}_b(0)$

outer loop structure

(1) Integrate NLROMS, $t = [0, \tau]$ and save $\underline{x}(t)$

inner loop structure

(a) Choose $\delta\underline{x}(0)$ (= 0 first time around)

(b) Integrate TLROMS, $t = [0, \tau]$ and evaluate (27)

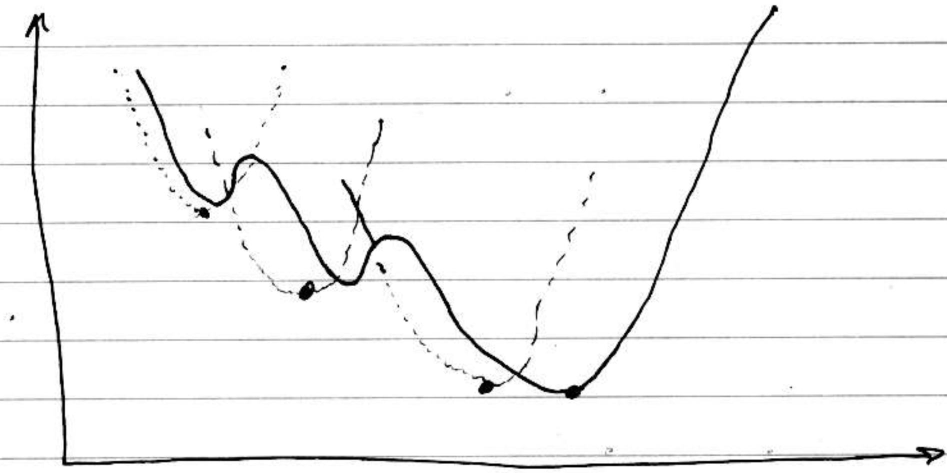
(c) Integrate ADROMS, $t = [\tau, 0]$ forced by

$$-\underline{\underline{H}}_i^T \underline{\underline{O}}^{-1} (\delta\underline{x}_i - \underline{d}_i) \text{ to yield } \left(\frac{\partial J_o}{\partial \delta\underline{x}(0)}\right) = -\underline{\lambda}(0)$$

(d) Use conjugate gradient descent as in S4DVAR to determine the down gradient correction to $\delta\underline{x}(0)$ that yields a smaller $J(\underline{x})$

(2) After several iterations of the inner-loop structure compute a new $\underline{x}(0) = \underline{x}(0) + \delta\underline{x}(0)$

Nonlinear $J(X)$



- A Good approximation of Hessian is $\underline{\underline{B}}^{-1}$

$$\underline{\underline{B}} = \underline{\underline{B}}^{1/2}(\underline{\underline{B}}^{1/2})^T = \underline{\underline{B}}^{1/2}\underline{\underline{B}}^{T/2} \qquad (28)$$

introduce a new variable :

$$\underline{V} = \underline{\underline{B}}^{-1/2}\delta\underline{X} \qquad (29)$$

$$\delta\underline{x} = \underline{\underline{B}}^{1/2}\underline{V}$$

So that :

$$J = \underbrace{\frac{1}{2}\underline{V}^T(0)\underline{V}(0)}_{J_b} + \underbrace{\frac{1}{2}\sum_{i=1}^{N}\left(\underline{\underline{G}}_i\,\delta\underline{x}(0) - \underline{d}_i\right)^T \underline{\underline{O}}^{-1}\left(\underline{\underline{G}}_i\,\delta\underline{x}(0) - \underline{d}_i\right)}_{J_o} \qquad (30)$$

dominates the behavior of $J$