The assimilation of data into atmospheric models, and the use of linearizations optimised for finite perturbations.

Tim Payne

Durham Symposium on Numerical Analysis of Multiscale Problems, July 9th 2010

Local linear approximation to multiscale function

Suppose we have a function $\mathbf{f}$ for which we seek a local linear approximation

$$\mathbf{f}(\mathbf{x}+\boldsymbol{\delta}) \approx \mathbf{F}(\mathbf{x}) + \mathbf{T}(\mathbf{x})\boldsymbol{\delta}$$

where the function $\mathbf{f}$ exhibits small and large scales, but the increments $\boldsymbol{\delta}$ are on the large scale.

More generally, suppose we know the pdf of $\boldsymbol{\delta}$, and we seek the 'best local linear approximation' in the sense of minimising the expectation over $\boldsymbol{\delta}$ of

$$\mathcal{I} = E\{[\mathbf{f}(\mathbf{x}+\boldsymbol{\delta}) - \mathbf{F}(\mathbf{x}) - \mathbf{T}(\mathbf{x})\boldsymbol{\delta}]^T A[\mathbf{f}(\mathbf{x}+\boldsymbol{\delta}) - \mathbf{F}(\mathbf{x}) - \mathbf{T}(\mathbf{x})\boldsymbol{\delta}]\}$$

for a given symmetric positive definite matrix A.

Solution to optimal linearisation problem, 'Opt-M'

Since

$$\frac{\partial \mathcal{I}}{\partial \mathbf{F}(\mathbf{x})} = 2AE[\mathbf{f}(\mathbf{x}+\boldsymbol{\delta}) - \mathbf{F}(\mathbf{x}) - \mathbf{T}(\mathbf{x})\boldsymbol{\delta}]$$

and

$$\frac{\partial \mathcal{I}}{\partial \mathbf{T}(\mathbf{x})} = 2AE\{[\mathbf{f}(\mathbf{x}+\boldsymbol{\delta}) - \mathbf{F}(\mathbf{x})]\boldsymbol{\delta}^T\} - 2A\mathbf{T}(\mathbf{x})E[\boldsymbol{\delta}\boldsymbol{\delta}^T]$$

the solution is independent of A and given by

$$\begin{aligned}\mathbf{T}(\mathbf{x}) &= \{E[\mathbf{f}(\mathbf{x}+\boldsymbol{\delta})\boldsymbol{\delta}^T] - E[\mathbf{f}(\mathbf{x}+\boldsymbol{\delta})]E[\boldsymbol{\delta}]^T\}\{E[\boldsymbol{\delta}\boldsymbol{\delta}^T] - E[\boldsymbol{\delta}][\boldsymbol{\delta}]^T\}^{-1} \\ \mathbf{F}(\mathbf{x}) &= E[\mathbf{f}(\mathbf{x}+\boldsymbol{\delta}) - \mathbf{T}(\mathbf{x})\boldsymbol{\delta}]\end{aligned}$$

Properties of 'Opt-M'

$$\mathbf{T}(\mathbf{x}) = \{E[(\mathbf{f}(\mathbf{x}+\boldsymbol{\delta})\boldsymbol{\delta}^T] - E[\mathbf{f}(\mathbf{x}+\boldsymbol{\delta})]E[\boldsymbol{\delta}]^T\}\{E[\boldsymbol{\delta}\boldsymbol{\delta}^T] - E[\boldsymbol{\delta}]E[\boldsymbol{\delta}]^T\}^{-1}$$

$$\mathbf{F}(\mathbf{x}) = E[\mathbf{f}(\mathbf{x}+\boldsymbol{\delta}) - \mathbf{T}(\mathbf{x})\boldsymbol{\delta}]$$

(1) In general 'optimal' linearisation state $\mathbf{F}(\mathbf{x})$ is not the original function $\mathbf{f}(\mathbf{x})$ and PF operator $\mathbf{T}(\mathbf{x})$ not the tangent linear $\mathbf{f}'(\mathbf{x})$

(2) Differentiability of $\mathbf{f}$ irrelevant

(3) Unlike the usual regularisation approach $\mathbf{T}(\mathbf{x})$ is not the derivative of $\mathbf{F}$

(4) If $\mathbf{f}$ is differentiable then $\mathbf{F}(\mathbf{x}) \to \mathbf{f}(\mathbf{x})$ and $\mathbf{T}(\mathbf{x}) \to \mathbf{f}'(\mathbf{x})$ as distribution $\boldsymbol{\delta} \to$ Dirac delta function

Application to cloud function

Seek a rational basis on which to 'regularise' in 4D-Var the Smith cloud scheme of 1990, where cloud fraction C is expressed as a function of

$$Q_N = \frac{q_T - q_{sat}(T_L, p)}{(1 - RH_c)q_{sat}(T_L, p)}$$

Cloud fraction and its derivative w.r.t. Q_N are

$$C = \begin{cases} 0 \;\; for \; Q_N \le -1 \\ \frac{1}{2}(1+Q_N)^2 \;\; for \; -1 \le Q_N \le 0 \\ 1-\frac{1}{2}(1-Q_N)^2 \;\; for \; 0 \le Q_N \le 1 \\ 1 \; for \; 1 \le Q_N \end{cases} \quad C' = \begin{cases} 0 \;\; for \; Q_N \le -1 \\ (1+Q_N) \;\; for \; -1 \le Q_N \le 0 \\ (1-Q_N) \;\; for \; 0 \le Q_N \le 1 \\ 0 \; for \; 1 \le Q_N \end{cases}$$

How should one regularise this function?

Met Office had used regularisation $C = \frac{1}{2}[1 + \tanh(2Q_N)]$ with derivative $C' = \mathrm{sech}^2(2Q_N)$

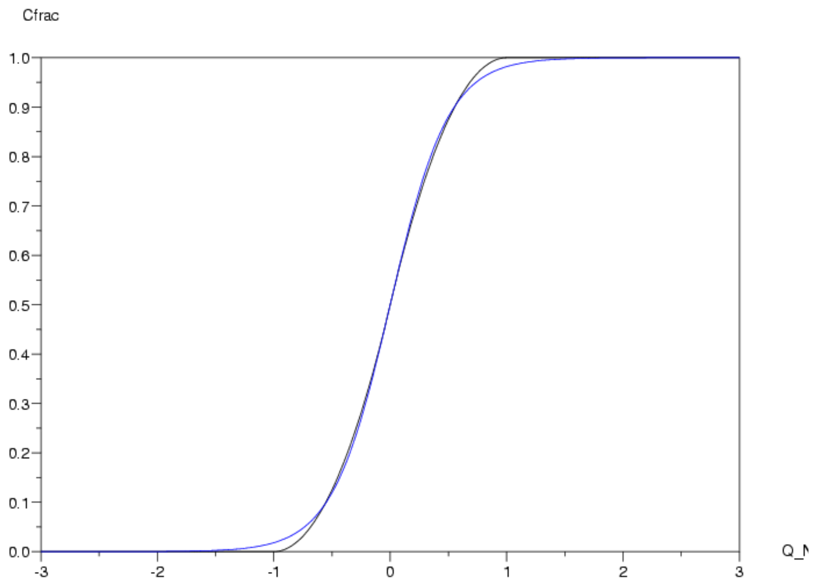
Smith Cloud Scheme with Standard Regularisation

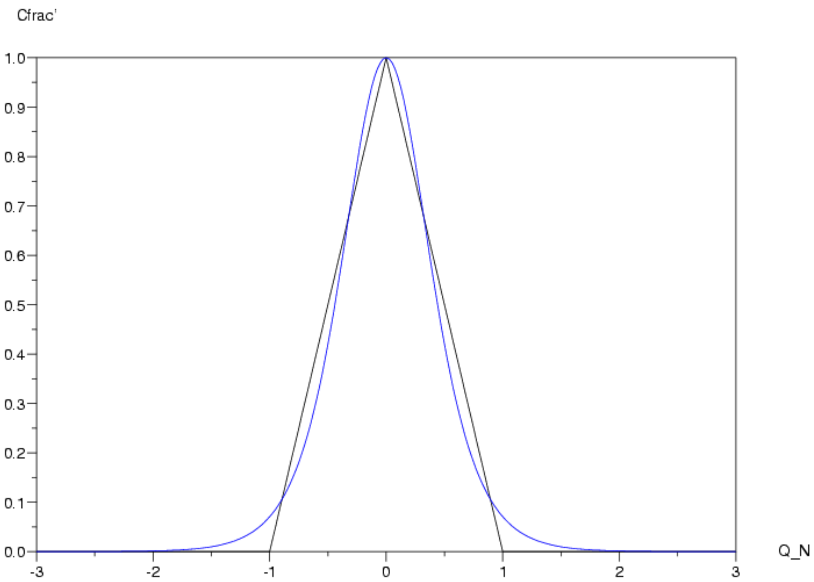
$$C = \begin{cases} 0 \quad for\ Q_N \le -1 \\ \tfrac{1}{2}(1+Q_N)^2 \quad for\ -1 \le Q_N \le 0 \\ 1-\tfrac{1}{2}(1-Q_N)^2 \quad for\ 0 \le Q_N \le 1 \\ 1\ for\ 1 \le Q_N \end{cases}$$

$$C' = \begin{cases} 0 \quad for\ Q_N \le -1 \\ (1+Q_N)\ for\ -1 \le Q_N \le 0 \\ (1-Q_N)\ for\ 0 \le Q_N \le 1 \\ 0\ for\ 1 \le Q_N \end{cases}$$

Standard regularisation $C = \tfrac{1}{2}[1+\tanh(2Q_N)]$

$$C' = \mathrm{sech}^2(2Q_N)$$





Smith Cloud Scheme with Optimal Regularisation

Optimal regularisation: $C(Q_N+\delta Q_N) \approx F(Q_N)+T(Q_N)\delta Q_N$

If $E(\delta Q_N) = 0$ then $F(Q_N)=E[C(Q_N+\delta Q_N)]$ and $T(Q_N)=\dfrac{E[C(Q_N+\delta Q_N)\delta Q_N]}{E[\delta Q_N{}^2]}$

$Eg.\ p=1$ Gaussian

$\delta Q \sim N_1(0,\sigma^2)$

$$\text{pdf } f(x) = \frac{1}{2\sigma}\exp\left(-\frac{|x|}{\sigma}\right)$$

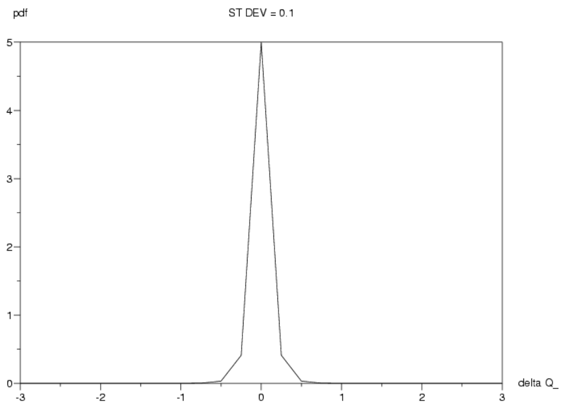
Opt reg LS:

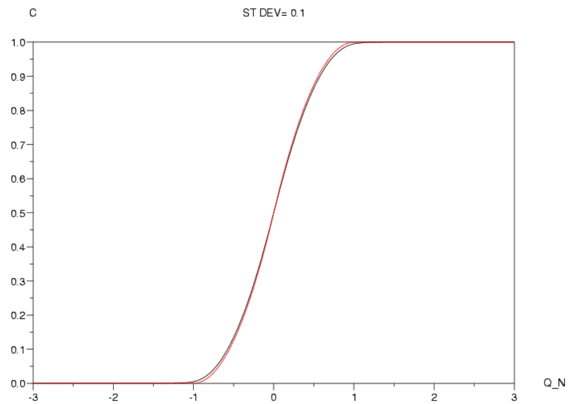
$E\{C(Q+\delta Q)\}$

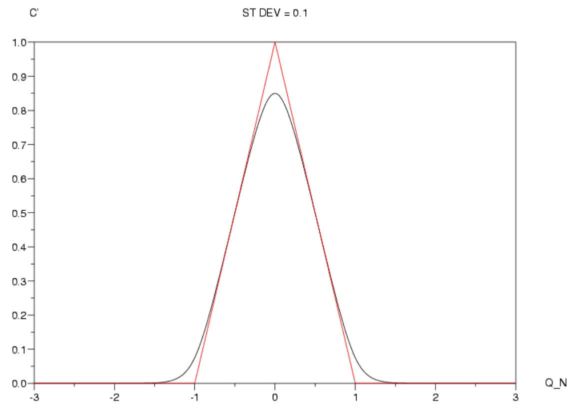
$$=\frac{1}{2\sigma}\int_{-\infty}^{\infty} C(Q+x)\exp(-\tfrac{|x|}{\sigma})dx$$

Opt reg PF: $\dfrac{E\{C(Q+\delta Q)\delta Q\}}{\sigma^2}$

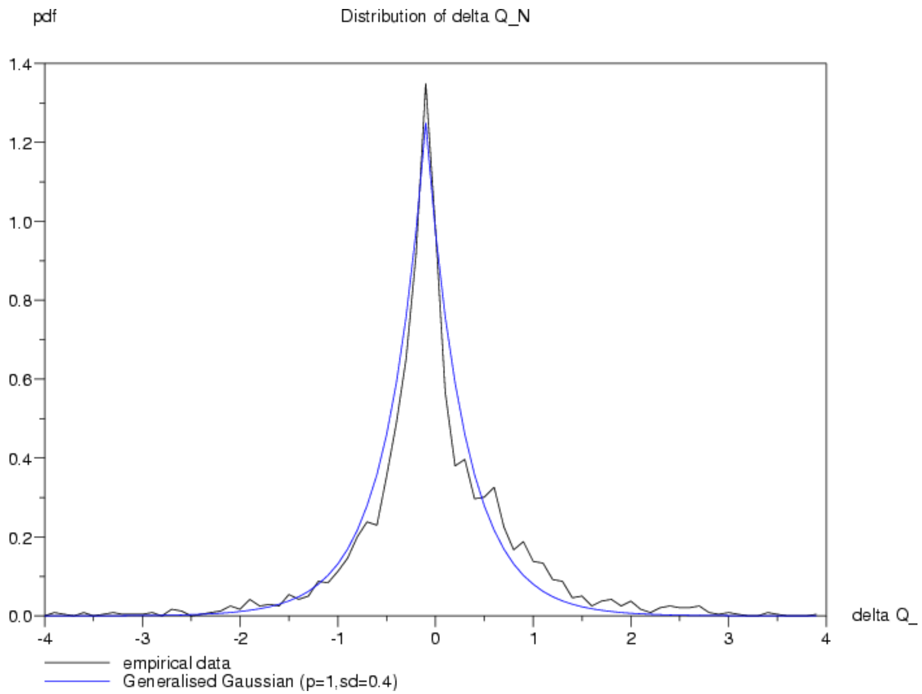
$$=\frac{1}{2\sigma^3}\int_{-\infty}^{\infty} C(Q+x)x\exp(-\tfrac{|x|}{\sigma})dx$$







Actual Distribution and Application using Real Data



$$T(Q) = \frac{T_1(Q) - C_1(Q)E[\delta]}{Var(\delta)}$$

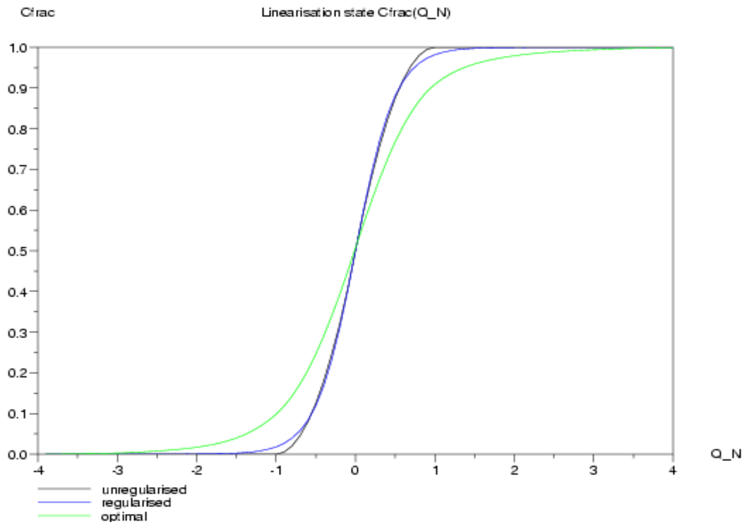
$$C(Q) = C_1(Q) - T(Q)E[\delta]$$

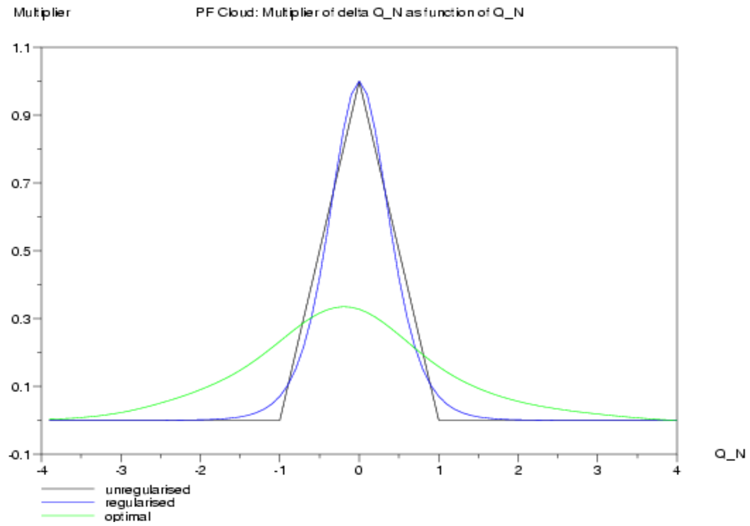
where

$$C_1(Q) = \int_{-\infty}^{\infty} f(\delta)C(Q+\delta)d\delta \approx \sum_{\iota=1}^{N} f(\delta_i)C(Q+\delta_\iota)\Delta\delta_\iota$$

$$T_1(Q) = \int_{-\infty}^{\infty} \delta f(\delta)C(Q+\delta)d\delta \approx \sum_{\iota=1}^{N} \delta_\iota f(\delta_i)C(Q+\delta_\iota)\Delta\delta_\iota$$

Results with Real Data





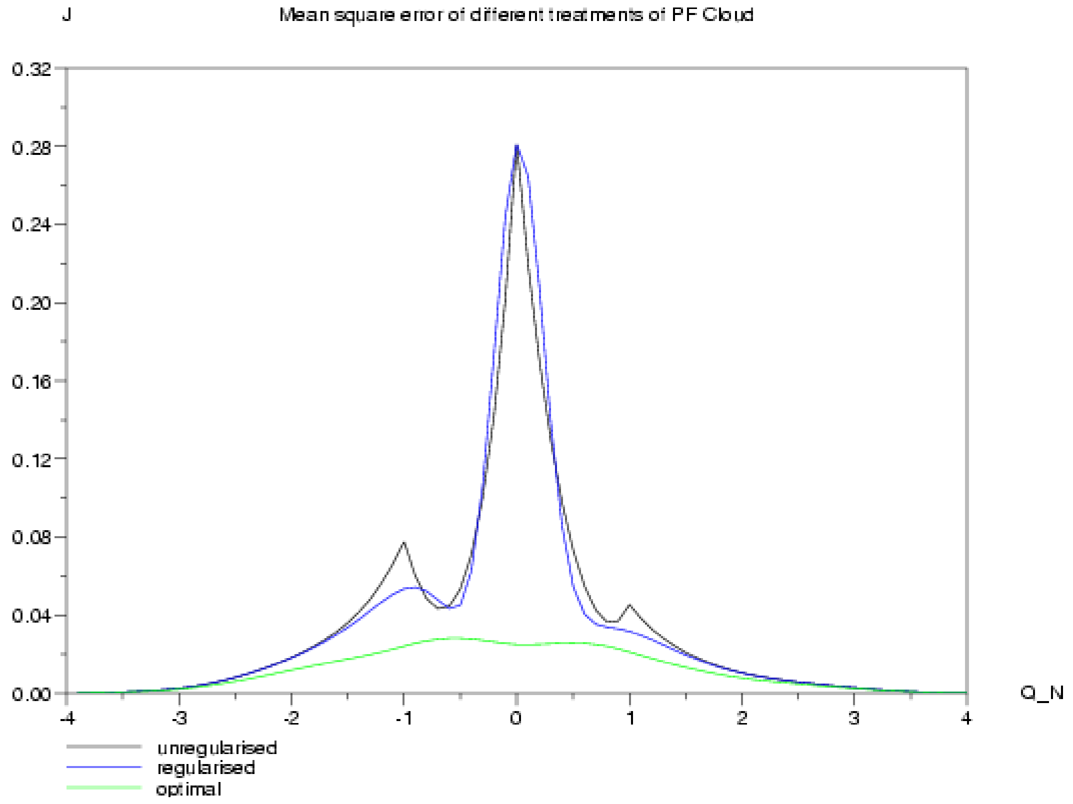
Suboptimality of TL Models

Recall that J (with A=I) is simply the mean square error in the approximation

$$\mathbf{f}(\mathbf{x}+\boldsymbol{\delta}) \approx \mathbf{F}(\mathbf{x}) + \mathbf{T}(\mathbf{x})\boldsymbol{\delta}$$

Hence have direct means of assessing how suboptimal different regularisations (including TL) are

Eg if $\mathbf{f}$ is quadratic and $\boldsymbol{\delta}$ distributed as standard (p=2) Gaussian then $\frac{J_{TL}}{J_{Opt\ Reg}} = 3/2$



Remarks on "Opt-M"

Whereas the tangent-linear opeator is a function only of f, Opt-M is a function of f **and** the pdf of the increments.

In the TL approach the original function is used for the base state and its derivative for the linear model; neither is optimal unless the function is close to linear within a couple of standard deviations of the base state.

Non-differentiability of f is not of itself a problem as can easily approximate by a smooth function arbitrarily C^0 close to f. The issue is how rapidly the derivative changes compared with likely size of increments.

The results apply equally well to discontinuous functions, and at the other extreme, it shows that taking the tangent-linear is generally suboptimal even for smooth functions.

Data Assimilation - Formulation

State of atmosphere $\mathbf{x}$ defined by (at least) 7 variables

$(u, v, w, p, q, \rho, \theta)$ at each gridpoint

Before any observations have been assimilated all we know is that the true state of the atmosphere $\mathbf{x}$ is 'close' to some 'background' state $\mathbf{x_b}$ which is the short period forecast from the latest best estimate of the atmosphere 6 hours earlier,

$$\mathbf{x} \sim N(\mathbf{x_b}, B) \text{ where } B = E[(\mathbf{x} - \mathbf{x_b})(\mathbf{x} - \mathbf{x_b})^T]$$

$$\text{so } p(\mathbf{x}) \propto \exp\{-\tfrac{1}{2}(\mathbf{x} - \mathbf{x_b})^T B^{-1} (\mathbf{x} - \mathbf{x_b})\}$$

Let $\mathbf{y}$ denote a vector of observations, and H be the observation operator which maps $\mathbf{x}$ to the vector of observations implied by $\mathbf{x}$

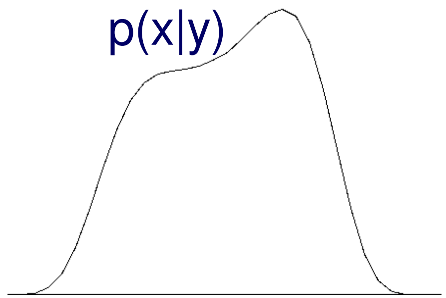
$$y \sim N(H\mathbf{x}, R) \text{ where } R = E[(H\mathbf{x} - \mathbf{y})(H\mathbf{x} - \mathbf{y})^T]$$

$$\text{so } p(\mathbf{y} \mid \mathbf{x}) \propto \exp\{-\tfrac{1}{2}(H\mathbf{x} - \mathbf{y})^T R^{-1} (H\mathbf{x} - \mathbf{y})\}$$

Best estimate of state of atmosphere

The *most likely* state of the atmosphere given the background *and* information in the observations is given by the maximum of the posterior pdf $\mathrm{p}(\mathbf{x} \mid \mathbf{y})$

By Bayes theorem $\mathrm{p}(\mathbf{x}, \mathbf{y}) = \mathrm{p}(\mathbf{x} \mid \mathbf{y})\mathrm{p}(\mathbf{y}) = \mathrm{p}(\mathbf{y} \mid \mathbf{x})\mathrm{p}(\mathbf{x})$

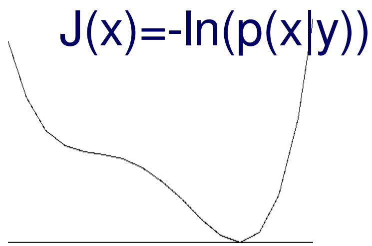


$$\mathrm{p}(\mathbf{x} \mid \mathbf{y}) = \frac{\mathrm{p}(\mathbf{y} \mid \mathbf{x})\mathrm{p}(\mathbf{x})}{\mathrm{p}(\mathbf{y})}$$

$$\propto \exp\{-\tfrac{1}{2}(\mathbf{x}-\mathbf{x_b})^T B^{-1}(\mathbf{x}-\mathbf{x_b}) - \tfrac{1}{2}(\mathrm{H}\mathbf{x}-\mathbf{y})^T R^{-1}(\mathrm{H}\mathbf{x}-\mathbf{y})\}$$

ie, we seek x which maximises the RHS of this expression and therefore which minimises

$$J(\mathbf{x}) = \tfrac{1}{2}(\mathbf{x}-\mathbf{x_b})^T B^{-1}(\mathbf{x}-\mathbf{x_b}) + \tfrac{1}{2}(\mathrm{H}\mathbf{x}-\mathbf{y})^T R^{-1}(\mathrm{H}\mathbf{x}-\mathbf{y})$$



Data assimilation as large optimization problem

In reality observations are distributed over a time window, so denoting

$\mathbf{y}^{\mathbf{i}}$ as vector of observations at time i,

$M_0^{\,i}$ is nonlinear model evolution from timestep 0 to i

H^i is observation operator at time i

B, R_i are covariance matrices

Best estimate of state of atmosphere $\mathbf{x}$ obtained by minimising cost function

$$J(\mathbf{x}) = \tfrac{1}{2}(\mathbf{x}-\mathbf{x}_\mathbf{b})^T B^{-1}(\mathbf{x}-\mathbf{x}_\mathbf{b}) + \tfrac{1}{2}\sum_i [\mathbf{y}^i - H^i M_0^{\,i}\mathbf{x}]^T R_i^{-1}[\mathbf{y}^i - H^i M_0^{\,i}\mathbf{x}]$$

State of atmosphere $\mathbf{x}$ defined by 7 variables $(\mathrm{u}, \mathrm{v}, \mathrm{w}, \mathrm{p}, \mathrm{q}, \rho, \theta)$ at each of $432 * 325 * 50 = 7$ million gridpoints $\rightarrow$ $\mathbf{x}$ has 50 million elements

H^i may be nonlinear (eg radiative transfer equation for satellite brightness temperatures) and the forecast model $M_0^{\,i}$ is certainly nonlinear

4D-Var

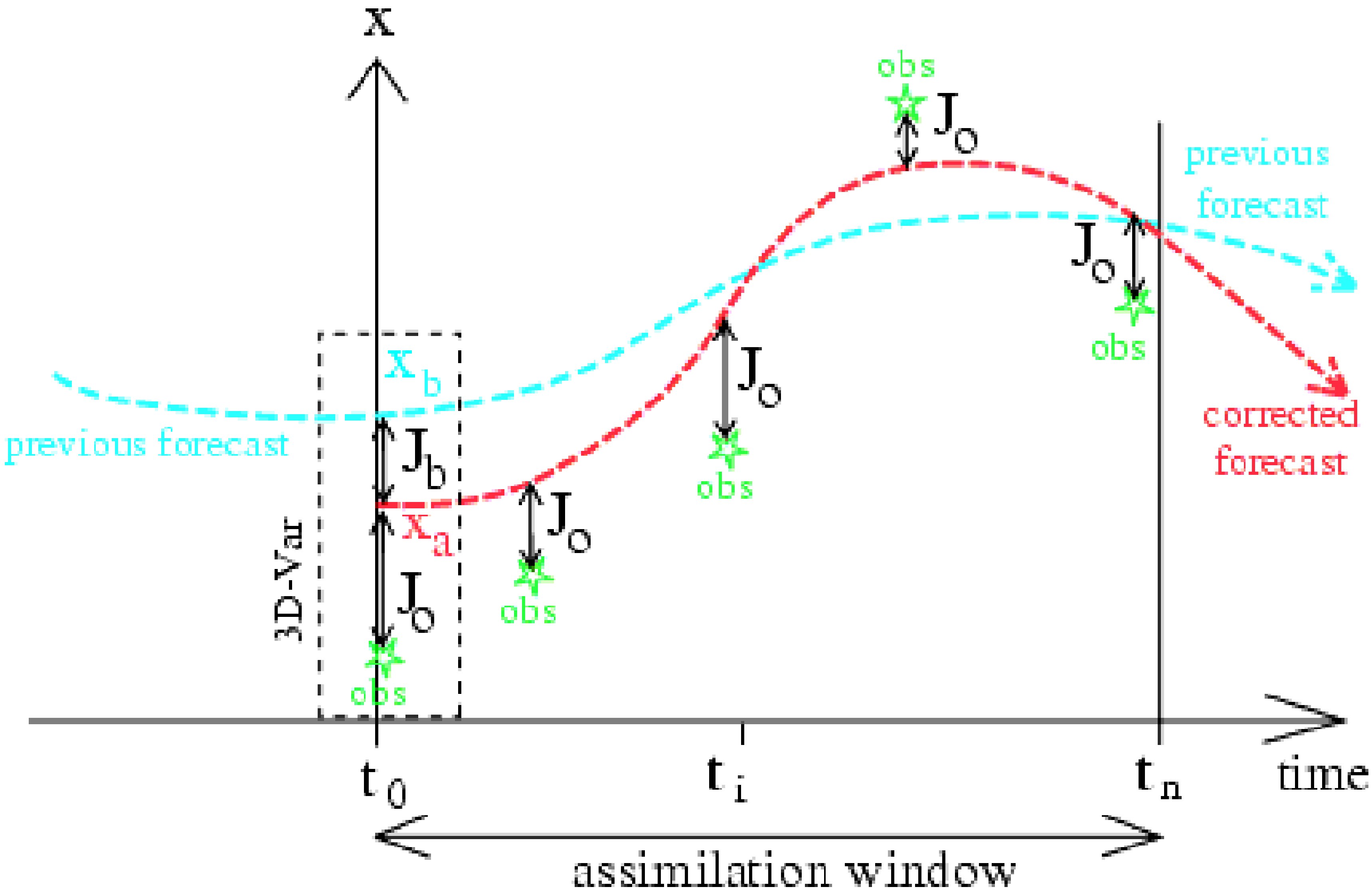


Figure courtesy ECMWF

Incremental Method

Recall

$$J(\mathbf{x}) = \tfrac{1}{2}(\mathbf{x}-\mathbf{x_b})^T B^{-1}(\mathbf{x}-\mathbf{x_b}) + \tfrac{1}{2}\sum_i [\mathbf{y}^i - H^i M_0{}^i \mathbf{x}]^T R_i^{-1}[\mathbf{y}^i - H^i M_0{}^i \mathbf{x}]$$

Let S be a "simplification" operator (eg projection to lower resolution)

Set $\delta\mathbf{x} = S(\mathbf{x}-\mathbf{x_b})$

Let $\mathbf{M_0}^{\mathbf{i}}$ denote a simplified linearized version of the forecast model $M_0{}^i$ in the same space as the range of S (eg the tangent linear of a low dimensional form of $M_0{}^i$ with physical processes omitted) and approximate

$M_0{}^i\mathbf{x}$ as $M_0{}^i\mathbf{x} \approx M_0{}^i\mathbf{x_b} + S^{-1}\mathbf{M_0}^{\mathbf{i}}\delta\mathbf{x}$

Incremental Method

Substituting

$$M_0^{\,i}\mathbf{x} \approx M_0^{\,i}\mathbf{x_b} + S^{-1}\mathbf{M_0}^{\mathbf{i}}\delta\mathbf{x}$$

into J gives

$$J(\delta\mathbf{x}) = \tfrac{1}{2}\delta\mathbf{x}^T S^{-T} B^{-1} S^{-1}\delta\mathbf{x} +$$
$$\tfrac{1}{2}\sum_i [\mathbf{y}^i - H^i(M_0^{\,i}\mathbf{x_b} + S^{-1}\mathbf{M_0}^{\mathbf{i}}\delta\mathbf{x})]^T R_i^{-1}[\mathbf{y}^i - H^i(M_0^{\,i}\mathbf{x_b} + S^{-1}\mathbf{M_0}^{\mathbf{i}}\delta\mathbf{x})]$$

which is quadratic in $\delta\mathbf{x}$ apart from the effects of the weakly non-linear observation operator H^i

Much cheaper as at lower resolution and easier to implement as physical processes omitted. To capture the effects of full resolution and the missing physics this is iterated in an outer loop:

$$\mathbf{x_b} \leftarrow \mathbf{x_b} + S^{-1}\delta\mathbf{x}$$

Incremental Method

Non - incremental J is

$$J(\mathbf{x}) = \tfrac{1}{2}(\mathbf{x}-\mathbf{x_b})^T B^{-1}(\mathbf{x}-\mathbf{x_b}) + \tfrac{1}{2}\sum_i [\mathbf{y}^i - H^i M_0{}^i\mathbf{x}]^T R_i^{-1}[\mathbf{y}^i - H^i M_0{}^i\mathbf{x}]$$

Minimise instead using $\delta\mathbf{x} = S(\mathbf{x}-\mathbf{x_b})$

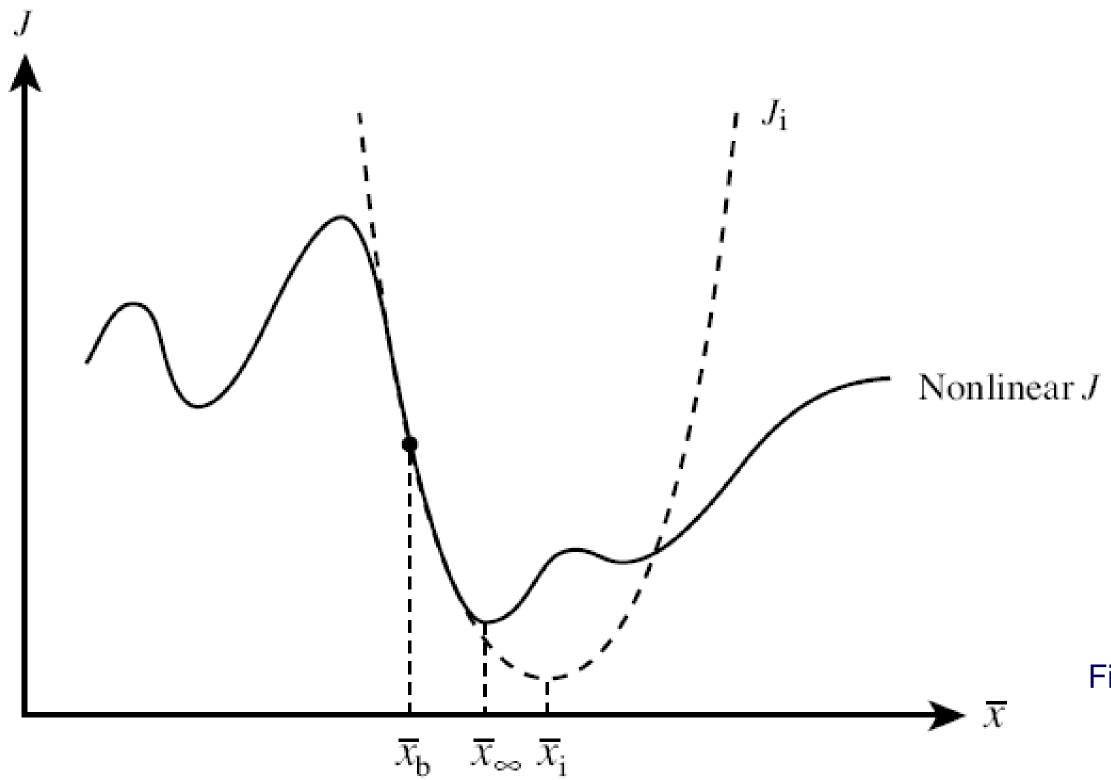


Figure courtesy ECMWF

How to choose linear model in 4D-Var?

Non-incremental 4D-Var cost function is

$$J(\delta\mathbf{x}) = \frac{1}{2}\sum_{i=1}^{N}(\mathbf{y}_i - \mathcal{H}_i(\mathcal{M}_i(\mathbf{x}_0 + \delta\mathbf{x})))^T R_i^{-1}(\mathbf{y}_i - \mathcal{H}_i(\mathcal{M}_i(\mathbf{x}_0 + \delta\mathbf{x}))) + \frac{1}{2}\delta\mathbf{x}^T B^{-1}\delta\mathbf{x} \quad (1)$$

where $\mathbf{x}_0$ is the background state at time t_0, $\mathcal{M}_i, \mathcal{H}_i$ are respectively the full model evolution to time t_i and the observation operator at time t_i, $\mathbf{y}_i$ is a vector of observations at t_i, and B, R_i are background and observation error covariance matrices.

Suppose for present that observation operators $\mathcal{H}_i$ are linear

Tangent linear approximation is to replace $\mathcal{M}_i(\mathbf{x}_0 + \delta\mathbf{x})$ by $\mathcal{M}_i(\mathbf{x}_0) + \mathcal{M}_i'(\mathbf{x}_0)\delta\mathbf{x}$

Can we do better? - ie, can we find $M(\mathbf{x}_0), \boldsymbol{\ell}(\mathbf{x}_0)$ so that $\boldsymbol{\ell}(\mathbf{x}_0) + M(\mathbf{x}_0)\delta\mathbf{x}$

(a) better approximates $\mathcal{M}_i(\mathbf{x}_0 + \delta\mathbf{x})$ than $\mathcal{M}_i(\mathbf{x}_0) + \mathcal{M}_i'(\mathbf{x}_0)\delta\mathbf{x}$ does, or

(b) 'performs better' in Eqn (1) than $\mathcal{M}_i(\mathbf{x}_0) + \mathcal{M}_i'(\mathbf{x}_0)\delta\mathbf{x}$ does

Opt-A: Optimising linear model for analysis

Supposing there are observations only at one time, so

$$\mathbf{y} = \mathbf{h}(\mathbf{x}_b + \boldsymbol{\epsilon}_b) + \boldsymbol{\epsilon}_o$$

the analysis in incremental 4D-Var is

$$\mathbf{x}_a = \mathbf{x}_b + \boldsymbol{\delta}$$

where $\boldsymbol{\delta}$ is obtained in the inner loop by minimising, for some R, B

$$J(\boldsymbol{\delta}) = \frac{1}{2}\boldsymbol{\delta}^T B^{-1}\boldsymbol{\delta} + \frac{1}{2}(\mathbf{y} - \boldsymbol{\ell}(\mathbf{x}_b) - M(\mathbf{x}_b)\boldsymbol{\delta})^T R^{-1}(\mathbf{y} - \boldsymbol{\ell}(\mathbf{x}_b) - M(\mathbf{x}_b)\boldsymbol{\delta})$$

where $\boldsymbol{\ell}(\mathbf{x}_b) + M(\mathbf{x}_b)\boldsymbol{\delta}$ approximates $\mathbf{h}(\mathbf{x}_b + \boldsymbol{\delta})$

eg in TL have $\boldsymbol{\ell}(\mathbf{x}_b) = \mathbf{h}(\mathbf{x}_b))$ and $M(\mathbf{x}_b) = \mathbf{h}'(\mathbf{x}_b)$

What about instead choosing $M(\mathbf{x}_b), \boldsymbol{\ell}(\mathbf{x}_b)$ to improve analysis directly? Might do this by choosing analysis

(i) to best approximate truth, or

(ii) to best approximate conditional mean

Opt-A: Optimising linear model for analysis

$\mathbf{x}_a$ minimising J on previous slide is

$$\mathbf{x}_a = \mathbf{x}_b + K(\mathbf{x}_b)(\mathbf{y} - \boldsymbol{\ell}(\mathbf{x}_b))$$

where

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

so choosing $M(\mathbf{x}_b), \boldsymbol{\ell}(\mathbf{x}_b)$ is equivalent to choosing $K(\mathbf{x}_b), \boldsymbol{\ell}(\mathbf{x}_b)$.

We first consider two problems where the objective is to choose $K(\mathbf{x}_b), \ell(\mathbf{x}_b)$ in such a way as to minimise the expected error in the analysis.

(Opt-A) Find matrix $K(\mathbf{x}_b)$ and vector $\ell(\mathbf{x}_b)$ which minimise the expected analysis error

$$E[\|\mathbf{x}_a - \mathbf{x}_t\|^2] = E[\|\mathbf{x}_b + K(\mathbf{x}_b)(\mathbf{h}(\mathbf{x}_b + \boldsymbol{\epsilon}_b) + \boldsymbol{\epsilon}_o - \ell(\mathbf{x}_b)) - \mathbf{x}_t\|^2]$$

where the expectation is over $\boldsymbol{\epsilon}_b, \boldsymbol{\epsilon}_o$ and $\|.\|$ denotes some norm.

Solution to Opt-A

For the norm (on analysis error etc) we will use

$$\|\mathbf{x}\|^2 = \mathbf{x}^T A \mathbf{x}$$

where A is some positive definite matrix. In Opt-A we seek to find vector $\boldsymbol{\ell}(\mathbf{x}_b)$ and matrix $K(\mathbf{x}_b)$ which minimises

$$E[(\mathbf{x}_t - \mathbf{x}_a)^T A(\mathbf{x}_t - \mathbf{x}_a)]$$

where

$$\mathbf{x}_a = \mathbf{x}_b + K(\mathbf{h}(\mathbf{x}_b + \boldsymbol{\epsilon}_b) + \boldsymbol{\epsilon}_o - \boldsymbol{\ell})$$

The solution to Reg-A is independent of the matrix A:

$$K = E[\{\mathbf{x}_t - \mathbf{x}_b - E[\mathbf{x}_t - \mathbf{x}_b]\}\mathbf{h}(\mathbf{x}_b + \boldsymbol{\epsilon}_b)^T] \times \\ \{E[(\mathbf{h}(\mathbf{x}_b + \boldsymbol{\epsilon}_b) + \boldsymbol{\epsilon}_o - E[\mathbf{h}(\mathbf{x}_b + \boldsymbol{\epsilon}_b)])(\mathbf{h}(\mathbf{x}_b + \boldsymbol{\epsilon}_b) + \boldsymbol{\epsilon}_o)^T]\}^{-1}$$

$$K\boldsymbol{\ell}(\mathbf{x}_b) = KE[\mathbf{h}(\mathbf{x}_b + \boldsymbol{\epsilon}_b)] - E[\mathbf{x}_t - \mathbf{x}_b]$$

Opt-A': optimising linear model for conditional mean

Another popular single-value choice for the analysis is the conditional mean

$$\mathbf{x}_a^e = E[\mathbf{x}|\mathbf{y}] = \int_{R^n} \mathbf{x} p(\mathbf{x}|\mathbf{y}) d\mathbf{x}$$

as this is also the minimum variance solution. Note that unlike the maximum likelihood estimate the conditional mean is a function of the whole pdf

This suggests the second problem

(Opt-A′) find matrix $K(\mathbf{x}_b)$ and vector $\ell(\mathbf{x}_b)$ which minimise the expected error

$$E[\|\mathbf{x}_a - \mathbf{x}_a^e\|^2] = E[\|\mathbf{x}_b + K(\mathbf{x}_b)(\mathbf{h}(\mathbf{x}_b + \boldsymbol{\epsilon}_b) + \boldsymbol{\epsilon}_o - \ell(\mathbf{x}_b)) - \mathbf{x}_a^e\|^2]$$

where the expectation is again over $\boldsymbol{\epsilon}_b, \boldsymbol{\epsilon}_o$.

Solution to Opt-A'

Note that Opt-A$'$ has the identical formulation to Opt-A but with the conditional mean $\mathbf{x}_a^e$

$$\mathbf{x}_a^e = E[\mathbf{x}_t|\mathbf{y}] = \int \mathbf{x}p(\mathbf{x}|\mathbf{y})d\mathbf{x}$$

replacing $\mathbf{x}_t$. The solution to Opt-A depended on $\mathbf{x}_t$ only through the terms $E[\mathbf{x}_t]$ and $E[\mathbf{x}_t\mathbf{y}^T]$. However

(i) Since $\mathbf{x}_a^e$ is a function of $\mathbf{y}$ only,

$$E[\mathbf{x}_a^e] = E[E[\mathbf{x}_t|\mathbf{y}]] = E[\mathbf{x}_t]$$

(ii)

$$E[\mathbf{x}_a^e\mathbf{y}^T] = \int\int \mathbf{x}\mathbf{y}^T p(\mathbf{y})p(\mathbf{x}|\mathbf{y})d\mathbf{x}d\mathbf{y}$$

and since by Baye's theorem

$$p(\mathbf{x}|\mathbf{y})p(\mathbf{y}) = p(\mathbf{x},\mathbf{y})$$

it follows that

$$E[\mathbf{x}_a^e\mathbf{y}^T] = \int\int \mathbf{x}\mathbf{y}^T p(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y} = E[\mathbf{x}\mathbf{y}^T]$$

It follows that Opt-A and Opt-A$'$ have exactly the same solution.

Strategies for incremental 4D-Var

Suppose there is only one timestep and we write the non-incremental 4D-Var cost function as

$$J = \frac{1}{2}(\mathbf{y} - \mathbf{h}(\mathbf{x}_b + \delta\mathbf{x}))^T R^{-1}(\mathbf{y} - \mathbf{h}(\mathbf{x}_b + \delta\mathbf{x})) + \frac{1}{2}\delta\mathbf{x}^T B^{-1}\delta\mathbf{x}$$

We now have several strategies for implementing incremental method:

- Standard (tangent-linear): linearise $\mathbf{h}$ by replacing it by $\mathbf{h}(\mathbf{x}_b) + \mathbf{h}'(\mathbf{x}_b)\delta\mathbf{x}$;

- 'Opt-M': linearise $\mathbf{h}$ by replacing it by $\mathbf{F}(\mathbf{x}_b) + \mathbf{T}(\mathbf{x}_b)\boldsymbol{\delta}$ chosen to minimise the expected error in

$$\|\mathbf{h}(\mathbf{x}_b + \boldsymbol{\delta}) - \mathbf{F}(\mathbf{x}_b) - \mathbf{T}(\mathbf{x}_b)\boldsymbol{\delta}\|_A^2$$

- 'Opt-A': linearise $\mathbf{h}$ by choosing it in such a way as to minimise the expected analysis error in

$$\|\mathbf{x}_b + K(\mathbf{x}_b)(\mathbf{h}(\mathbf{x}_b + \boldsymbol{\epsilon}_b) + \boldsymbol{\epsilon}_o - \ell(\mathbf{x}_b)) - \mathbf{x}_t\|_A^2$$

Relation between Opt-M and Opt-A

Relation between Opt-A and Opt-M

What is the relation between
'Opt-M' (optimising for the model)
and
'Opt-A' (optimising for the analysis)?

In every case

$$\mathbf{x}_a = \mathbf{x}_b + K(\mathbf{x}_b)[\mathbf{y} - \boldsymbol{\ell}_0(\mathbf{x}_b)]$$

Supposing for the time being that the background is unbiased so

$$E[\boldsymbol{\epsilon}_b \equiv \mathbf{x}_t - \mathbf{x}_b] = 0$$

TL, Opt-M and Opt-A compared

We have for TL:

$$\boldsymbol{\ell}_0(\mathbf{x}_b) = \mathbf{h}(\mathbf{x}_b)$$
$$K(\mathbf{x}_b) = BM^T(R + MBM^T)^{-1} \; where \; M = \mathbf{h}'(\mathbf{x}_b)$$

for Opt-M:

$$\boldsymbol{\ell}_0(\mathbf{x}_b) = E[\mathbf{h}(\mathbf{x}_b + \boldsymbol{\epsilon}_b)]$$
$$K(\mathbf{x}_b) = BM^T(R + MBM^T)^{-1} \; where \; M = \{E[\mathbf{h}(\mathbf{x}_b + \boldsymbol{\epsilon}_b)\boldsymbol{\epsilon}_b{}^T]\}\{E[\boldsymbol{\epsilon}_b\boldsymbol{\epsilon}_b{}^T]\}^{-1}$$

for Opt-A:

$$\boldsymbol{\ell}_0(\mathbf{x}_b) = E[\mathbf{h}(\mathbf{x}_b + \boldsymbol{\epsilon}_b)]$$
$$K(\mathbf{x}_b) = \{E[\boldsymbol{\epsilon}_b\mathbf{h}(\mathbf{x}_b + \boldsymbol{\epsilon}_b)^T]\}\{E[\mathbf{h}(\mathbf{x}_b + \boldsymbol{\epsilon}_b)\mathbf{h}(\mathbf{x}_b + \boldsymbol{\epsilon}_b)^T] - E[\mathbf{h}(\mathbf{x}_b + \boldsymbol{\epsilon}_b)]E[\mathbf{h}(\mathbf{x}_b + \boldsymbol{\epsilon}_b)]^T + R\}^{-1}$$

Opt-M as high-order approximation to Opt-A

Case (1) $\mathbf{h}$ only slightly nonlinear - suppose

$$\mathbf{h}(\mathbf{x}_b + \delta\mathbf{x}) = \mathbf{a}_0 + A_1\boldsymbol{\delta} + \tilde{A}$$

then

$$\|K_{Opt-A} - K_{TL}\| = O(\tilde{A})$$
$$\|K_{Opt-A} - K_{Opt-M}\| = O(\tilde{A}^2)$$

Ie, Opt-M represents a higher order approximation to Opt-A than TL does.

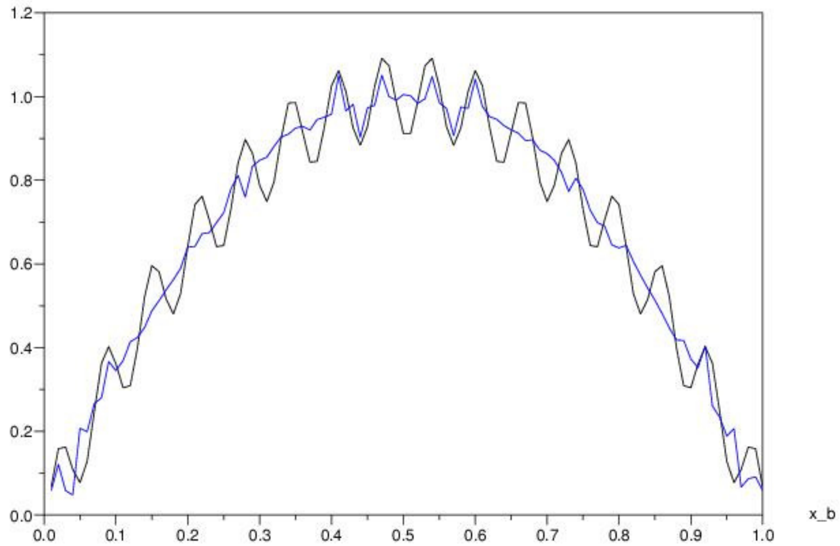
Case (2) $\mathbf{h}$ strongly nonlinear - advantages of Opt-M over TL can be dramatic.

TL versus Opt-M applied to modified logistic map

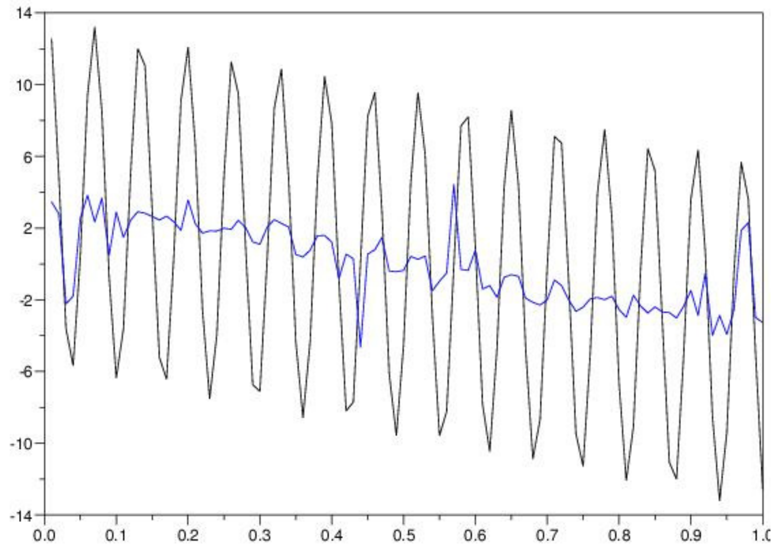
Example

$$h(x) = 4[x](1-[x]) + \gamma sin(K\pi[x]) \ where \ [x] = x, mod\ 1$$
$$g(x) = H(x) = I$$
$$y \sim N(h(x), 0.04)$$
$$x - x_b \sim N(0, 0.034)$$

F(x) using TL (black) and Opt-M (blue)

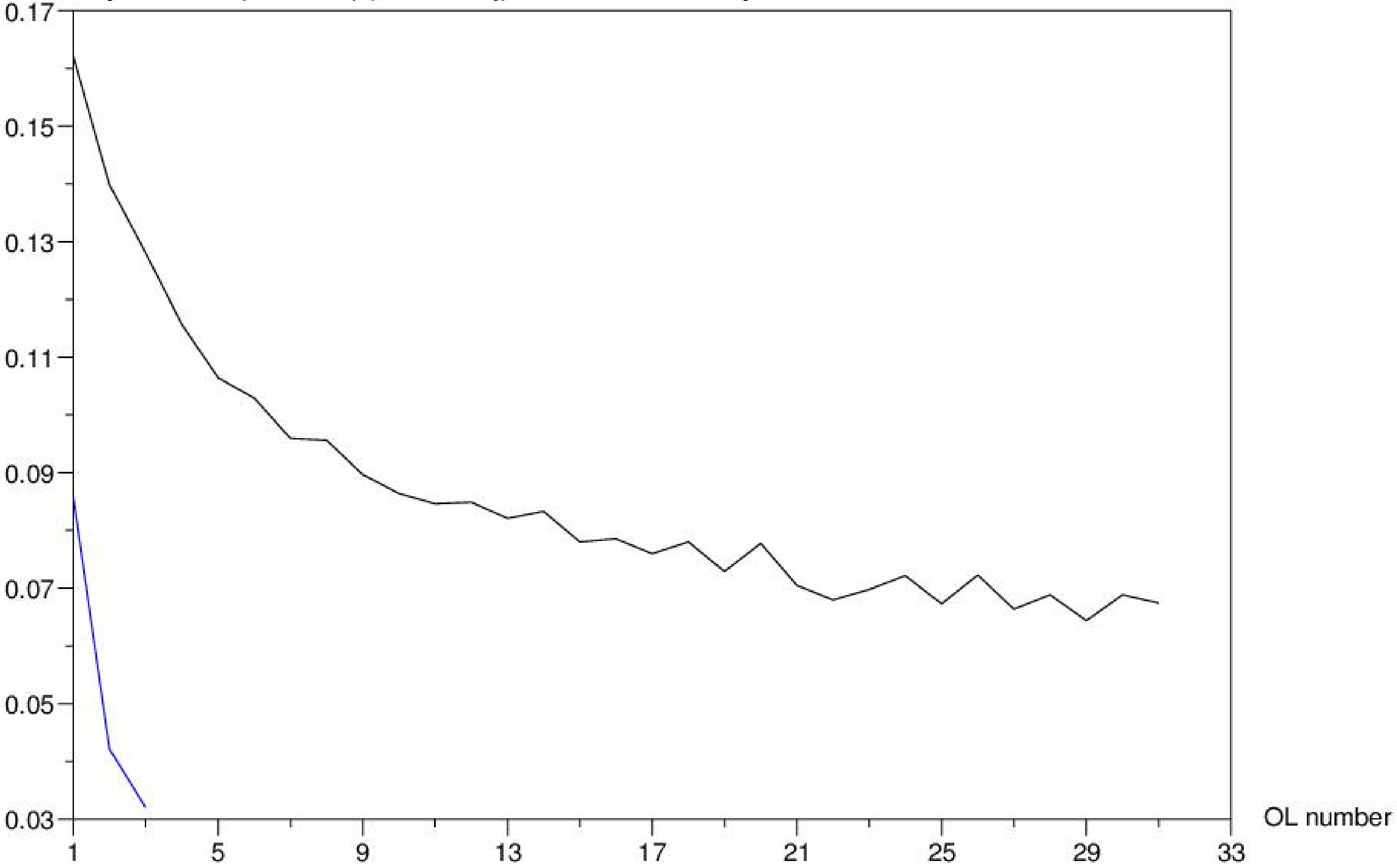


T(x) using TL (black) and Opt-M (blue)



TL versus Opt-M with many outer loops

Mean square background error using tangent-linear (black) and Opt-M (blue), many outer loops



Summary

In any situation where one needs a local linear approximation to a function, where the pdf of the increments is known or can be estimated, Opt-M is superior to a first order Taylor expansion.

Opt-M does not require differentiability of the original function.

In data assimilation locally-linear approximations are widely used. If we optimise the approximation to minimise the expected analysis error this yields an optimal linearisation Opt-A.

Opt-M applied to the model leads to a higher order approximation to Opt-A than does a first-order Taylor expansion, and outperforms it in cycled data assimilation.