Variational data assimilation

Amos S. Lawless

Data Assimilation Research Centre
University of Reading

a.s.lawless@reading.ac.uk

@amoslawless

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

A brief recap

Assume that we have

- a prior estimate of the state $\mathbf{x}^b$ with error covariance matrix $\mathbf{B}$
- observations $\mathbf{y}$ with error covariance matrix $\mathbf{R}$

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

Gaussian assumption

If we assume that the errors are Gaussian then the pdf is defined solely by the mean and covariance.

Prior

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\mathbf{B}|^{1/2}} \exp\{-\frac{1}{2}(\mathbf{x}-\mathbf{x}^b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^b)\}$$

Likelihood

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\mathbf{R}|^{1/2}} \exp\{-\frac{1}{2}(H(\mathbf{x})-\boldsymbol{y})^T\mathbf{R}^{-1}(H(\mathbf{x})-\boldsymbol{y})\}$$

Posterior

$$p(\mathbf{x}|\mathbf{y}) \propto \exp\{-\frac{1}{2}\{(\mathbf{x}-\mathbf{x}^b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^b) + (H(\mathbf{x})-\boldsymbol{y})^T\mathbf{R}^{-1}(H(\mathbf{x})-\boldsymbol{y})\}\}$$

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

Variational data assimilation – the idea

In variational data assimilation we seek the solution that maximises the *a posterior* probability $p(\mathbf{x}|\mathbf{y})$.

Since

$$p(\mathbf{x}|\mathbf{y}) \propto \exp\{ -\frac{1}{2}\{(\mathbf{x}-\mathbf{x}^b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^b) + (H(\mathbf{x})-\mathbf{y})^T\mathbf{R}^{-1}(H(\mathbf{x})-\mathbf{y})\} \}$$

we will have the maximum probability when **x** minimises

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

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

We consider two main algorithms

- Three-dimensional variational assimilation (3D-Var)
  - Where we consider 3 space dimensions.
- Four-dimensional variational assimilation (4D-Var)
  - Where we consider 3 space dimensions plus time as the 4^{th} dimension.
  - In this case we can consider the observation operator to include the dynamical model.

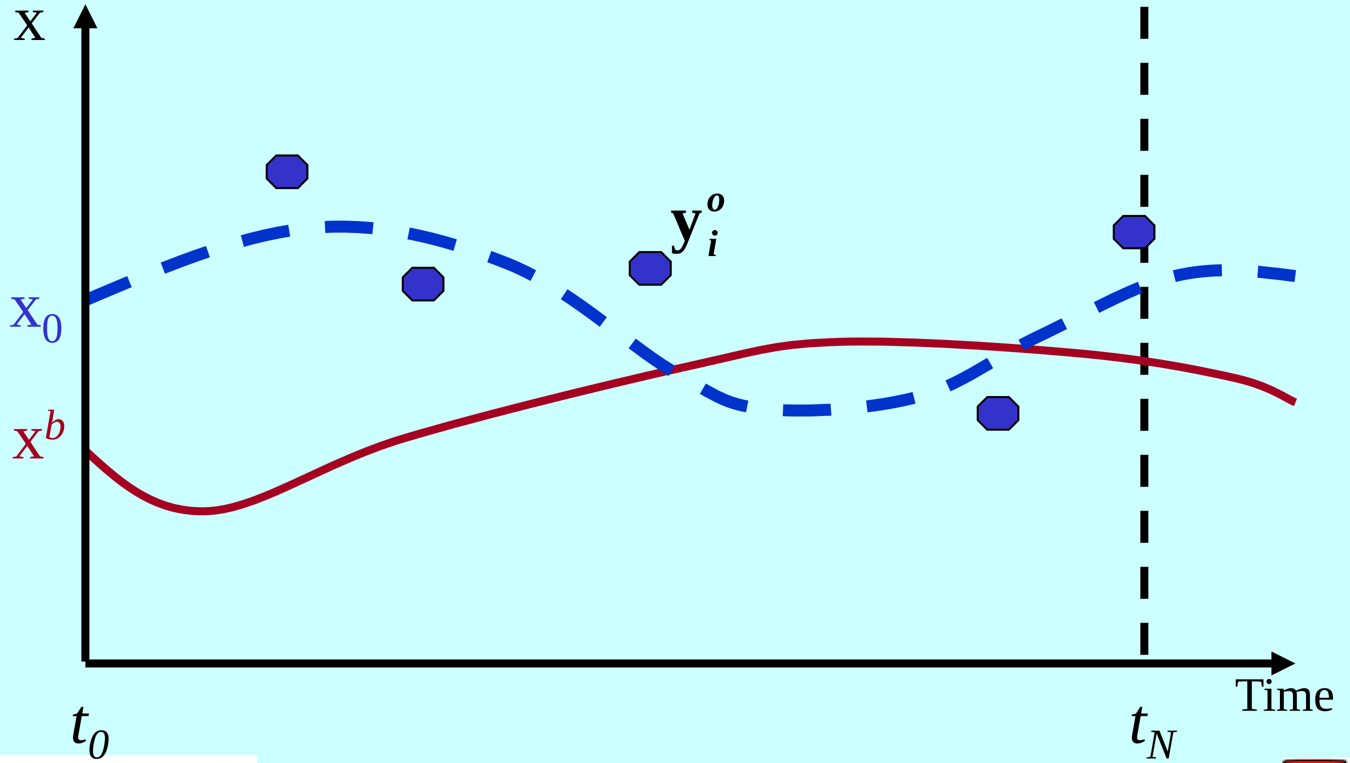
We will present 4D-Var first and 3D-Var as a variant of this.

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

Four-dimensional variational assimilation (4D-Var)

Aim: Find the best estimate of the true state of the system (*analysis*), consistent with both observations distributed in time and the system dynamics.



National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

4D-Var cost function

Minimize

$$\mathcal{J}(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) + \frac{1}{2}\sum_{i=0}^{N}(\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i)^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i)$$

with respect to x_0, subject to

$$\mathbf{x}_{i+1} = \mathcal{M}_i(\mathbf{x}_i),$$

x^b - *a priori* (background) state – Size of order 10^8 - 10^9

y_i - Observations – Size of order 10^6 - 10^7

H_i - Observation operator

B - Background error covariance matrix

R_i - Observation error covariance matrix

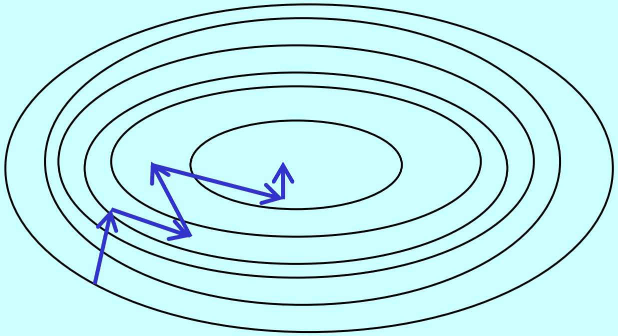
National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

For N=0 we have 3D-Var

University of Reading

Numerical minimization - Gradient descent methods

Iterative methods, where each successive iteration is based on the value of the function and its gradient at the current iteration.



$$\mathbf{x}_0^{(k+1)} = \mathbf{x}_0^{(k)} - \alpha\, \varphi(\mathbf{x}_0^{(k)})$$

where α is a step length and φ is a direction that depends on $J(\mathbf{x}_0^{(k)})$ and its gradient.

Problem: How do we calculate the gradient of $J(\mathbf{x}_0^{(k)})$ with respect to $\mathbf{x}_0^{(k)}$?

Method of Lagrange multipliers

We introduce Lagrange multipliers $\boldsymbol{\lambda}_i$ at time t_i and define the Lagrangian

$$\mathcal{L}(\mathbf{x}_i, \boldsymbol{\lambda}_i) = \mathcal{J}(\mathbf{x}_0) + \sum_{i=0}^{N-1} \boldsymbol{\lambda}_{i+1}^{\mathrm{T}} (\mathbf{x}_{i+1} - \mathcal{M}_i(\mathbf{x}_i))$$

Then necessary conditions for a minimum of the cost function subject to the constraint are found by taking variations with respect to $\boldsymbol{\lambda}_i$ and $\mathbf{x}_i$.

Variations with respect to $\boldsymbol{\lambda}_i$ simply give the original constraint.

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

$$\mathcal{L}(\mathbf{x}_i, \boldsymbol{\lambda}_i) = \mathcal{J}(\mathbf{x}_0) + \sum_{i=0}^{N-1} \boldsymbol{\lambda}_{i+1}^{\mathrm{T}}(\mathbf{x}_{i+1} - \mathcal{M}_i(\mathbf{x}_i))$$

Variations with respect to $\mathbf{x}_i$ give the *adjoint* equations

$$\boldsymbol{\lambda}_i = \mathbf{M}_i^T \boldsymbol{\lambda}_{i+1} - \mathbf{H}_i^T \mathbf{R}_i^{-1}(\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i)$$

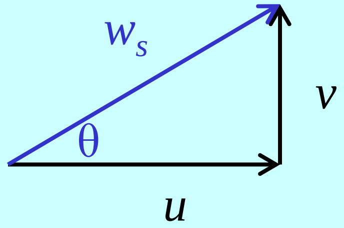
with boundary condition $\boldsymbol{\lambda}_{N+1} = 0$.

Then at initial time we have

$$\nabla \mathcal{J}(\mathbf{x}_0) = -\boldsymbol{\lambda}_0 + \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b).$$

An aside – What are the linear operators **H** & **M**?

Suppose we observe the wind speed w_s.



Then we have $\mathbf{x} = \begin{pmatrix} u \\ v \end{pmatrix}$, $\mathbf{y} = w_s$ and $\mathbf{y} = H(\mathbf{x})$

with

$$H(\mathbf{x}) = \sqrt{u^2 + v^2}$$

Then

$$\mathbf{H} = \begin{pmatrix} \frac{\partial H}{\partial u} & \frac{\partial H}{\partial v} \end{pmatrix} = \begin{pmatrix} \frac{u}{\sqrt{u^2+v^2}} & \frac{v}{\sqrt{u^2+v^2}} \end{pmatrix}$$

So where have we got to?

We wish to minimize

$$\mathcal{J}(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) + \frac{1}{2}\sum_{i=0}^{N}(\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i)^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i)$$

with respect to x_0, subject to

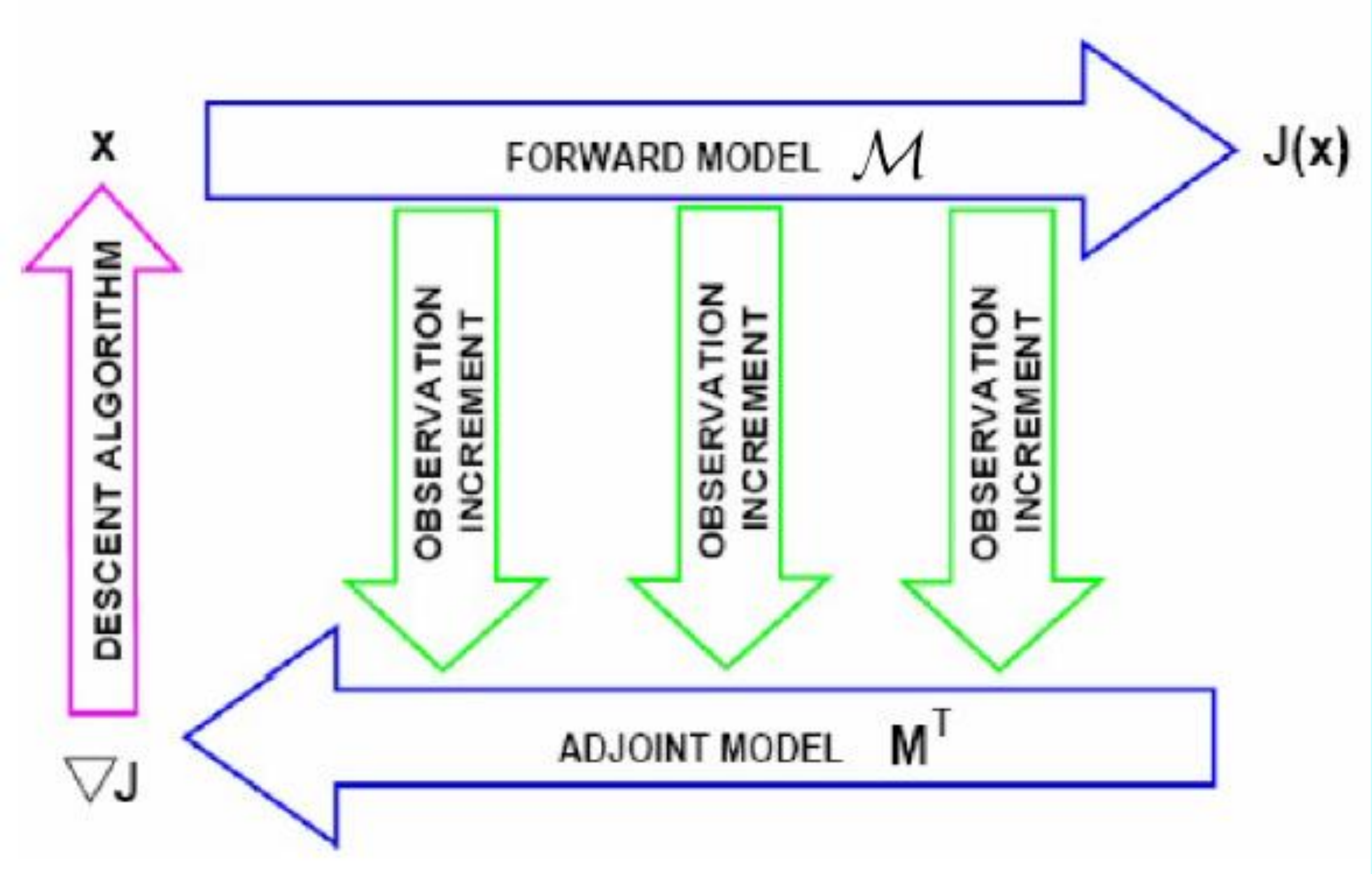
$$\mathbf{x}_{i+1} = \mathcal{M}_i(\mathbf{x}_i),$$

On each iteration we have to calculate *J* and its gradient

- To calculate *J* we need to run the nonlinear model
- To calculate the gradient of *J* we need one run of the adjoint model (backward in time)

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

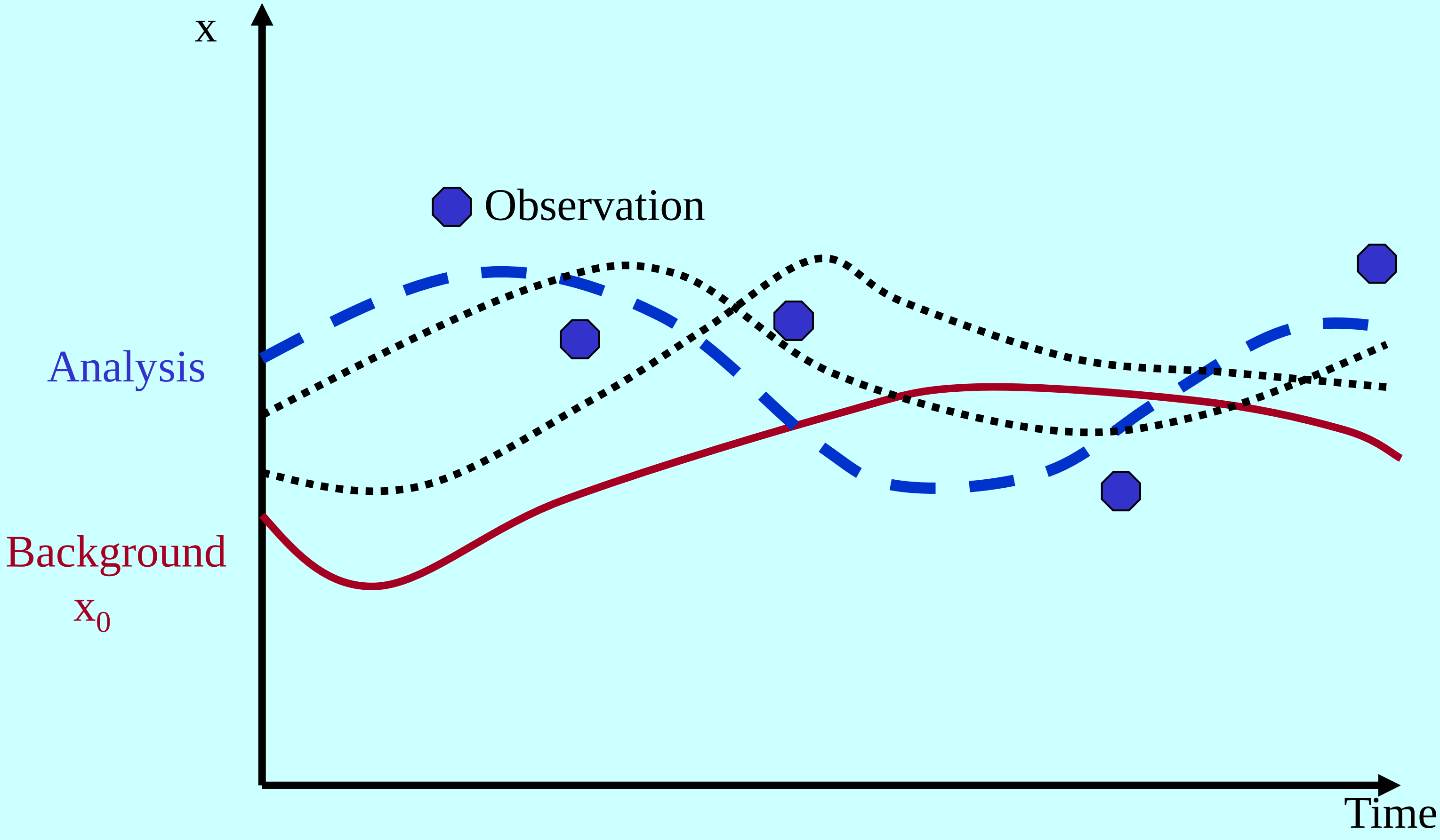


BUT this can be computationally expensive!

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

Incremental 4D-Var



Incremental 4D-Var

We solve a series of linearized minimization problems

$$\begin{aligned}\tilde{\mathcal{J}}^{(k)}[\delta\mathbf{x}_0^{(k)}] &= \frac{1}{2}(\delta\mathbf{x}_0^{(k)} - [\mathbf{x}^b - \mathbf{x}_0{}^{(k)}])^{\mathrm{T}}\mathbf{B}^{-1}(\delta\mathbf{x}_0^{(k)} - [\mathbf{x}^b - \mathbf{x}_0{}^{(k)}]) \\ &+ \frac{1}{2}\sum_{i=0}^{N}(\mathbf{H}_i\delta\mathbf{x}_i^{(k)} - \mathbf{d}_i^{(k)})^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathbf{H}_i\delta\mathbf{x}_i^{(k)} - \mathbf{d}_i^{(k)})\end{aligned}$$

with

$$\begin{aligned}\mathbf{d}_i &= \mathbf{y}_i - \mathcal{H}_i[\mathbf{x}_i^{(k)}] \\ \delta\mathbf{x}_{i+1} &= \mathbf{M}_i\delta\mathbf{x}_i\end{aligned}$$

and update using

$$\mathbf{x}_0^{(k+1)} = \mathbf{x}_0^{(k)} + \delta\mathbf{x}_0^{(k)}$$

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

Comments on incremental formulation

- Inner loop cost function is linear quadratic, so has a unique minimum.
- Can simplify the linear model (low resolution, simplified physics) in order to save computational time.
- Equivalent to an approximate Gauss-Newton procedure – Convergence results proved by *Lawless, Gratton & Nichols, QJRMS, 2005; Gratton, Lawless & Nichols, SIAM J. on Optimization, 2007.*
- Used in several operational centres, including ECMWF and Met Office.

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

3D-FGAT (First guess at appropriate time)

We solve a series of linearized minimization problems

$$
\begin{aligned}
\tilde{\mathcal{J}}^{(k)}[\delta \mathbf{x}_0^{(k)}] &= \frac{1}{2}(\delta \mathbf{x}_0^{(k)} - [\mathbf{x}^b - \mathbf{x}_0{}^{(k)}])^{\mathrm{T}} \mathbf{B}^{-1} (\delta \mathbf{x}_0^{(k)} - [\mathbf{x}^b - \mathbf{x}_0{}^{(k)}]) \\
&+ \frac{1}{2} \sum_{i=0}^{N} (\mathbf{H}_i \delta \mathbf{x}_i^{(k)} - \mathbf{d}_i^{(k)})^{\mathrm{T}} \mathbf{R}_i^{-1} (\mathbf{H}_i \delta \mathbf{x}_i^{(k)} - \mathbf{d}_i^{(k)})
\end{aligned}
$$

with

$$
\begin{aligned}
\mathbf{d}_i &= \mathbf{y}_i - \mathcal{H}_i[\mathbf{x}_i^{(k)}] \\
\delta \mathbf{x}_{i+1} &= \mathbf{M}_i \delta \mathbf{x}_i
\end{aligned}
$$

Replace this equation with $\delta \mathbf{x}_{i+1} = \delta \mathbf{x}_i$

and update using

$$
\mathbf{x}_0^{(k+1)} = \mathbf{x}_0^{(k)} + \delta \mathbf{x}_0^{(k)}
$$

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

Properties of 4D-Var

- Observations are treated at correct time.
- Use of dynamics means that more information can be obtained from observations.
- Covariances are implicitly evolved.
- In practice development of linear and adjoint models may be complex, but can be done at level of code.
- Standard formulation assumes model is perfect. Weak-constraint 4D-Var being developed to relax this assumption.

Weak-constraint 4D-Var

- The difference between the observation and the model trajectory may be due to model error rather than observation error.
- In weak constraint 4D-Var we do not assume that the model is correct, but assume

$$\mathbf{x}_{i+1} = \mathcal{M}_i(\mathbf{x}_i) + \eta_{i+1}, \qquad \eta \sim \mathcal{N}(\mathbf{Q}_i)$$

- Then we can either solve for $\mathbf{x}$ at each time, or for the initial $\mathbf{x}_0$ and all the η_i.

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

Error formulation

$$J(\mathbf{x}_0, \eta_1, \ldots, \eta_N) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) + \frac{1}{2}\sum_{i=0}^{N}(\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i))^T \mathbf{R}_i^{-1}(\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i)) + \frac{1}{2}\sum_{i=1}^{N}\eta_i^T \mathbf{Q}_i^{-1}\eta_i.$$

State formulation

$$J(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_N) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) + \frac{1}{2}\sum_{i=0}^{N}(\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i))^T \mathbf{R}_i^{-1}(\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i)) + \frac{1}{2}\sum_{i=1}^{N}(\mathbf{x}_i - \mathbf{x}_{i-1})^T \mathbf{Q}_i^{-1}(\mathbf{x}_i - \mathbf{x}_{i-1}).$$

Comments on weak-constraint formulation

- No longer assumes that the model is perfect, so in theory is more realistic.
- The problem becomes much bigger and more complicated to solve.
- We don't really know how to specify the model error covariance matrix **Q**.
- Even though the two formulations are equivalent, they have different mathematical properties.

References – Variational methods

Courtier, P., Thepaut, J-N. and Hollingsworth A. (1994), A strategy for operational implementation of 4D-Var, using an incremental approach, *Quart. J. Roy. Meteor. Soc.*, 120, 1367–1387.

Daužickaitė, I., Lawless, A.S., Scott, J.A. and van Leeuwen. P.J. (2021), Randomised preconditioning for the forcing formulation of weak constraint 4D-Var. *Quart. J. Royal Met. Soc.*, 147, 3719-3734.

Daužickaitė, I., Lawless, A. S., Scott, J. A., & van Leeuwen, P. J. (2021). On time-parallel preconditioning for the state formulation of incremental weak constraint 4D-Var. *Quart. J. Royal Met. Soc.*, 147, 3521-3529.

Gratton, S., Lawless, A.S. and Nichols, N.K. (2007), Approximate Gauss-Newton methods for nonlinear least squares problems, SIAM J. on Optimization, 18, 106-132.

Lawless, A.S., Gratton, S. and Nichols, N.K. (2005), An investigation of incremental 4D-Var using non-tangent linear models, Quart. J. Royal Met. Soc., 131, 459-476.

Lawless, A.S. (2013), Variational data assimilation for very large environmental problems. In *Large Scale Inverse Problems: Computational Methods and Applications in the Earth Sciences* (2013), Eds. Cullen, M.J.P., Freitag, M. A., Kindermann, S., Scheichl, R., Radon Series on Computational and Applied Mathematics 13. De Gruyter, pp. 55-90.

Talagrand, O and Courtier, P. (1987), Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory, *Quart. J. Roy. Meteor. Soc.*, 113, 1311-1328.

Trémolet, Y. (2006) Accounting for an imperfect model in 4D-Var. *Quart. J. Roy. Meteor. Soc*, 132, 2483–2504.