

Three-dimensional variational assimilation (3D-Var)

In the derivation of Optimal Interpolation, we found the optimal weight matrix \mathbf{W} that minimizes the total analysis error variance.

Lorenc (1986) showed that this solution is equivalent to a specific variational assimilation problem:

Find the optimal analysis x_a field that minimizes a (scalar) cost function, where the cost function is defined as the distance between x and the background x_b , weighted by the inverse of the background error covariance, plus the distance to the observations y_o , weighted by the inverse of the observation error covariance:

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}(\mathbf{y}_o - H(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y}_o - H(\mathbf{x})) \quad (5.1)$$

Reference:

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **112**, 1177-1194.

Note: Lorenc (1986) is a classic paper on modern data assimilation – the scanned paper is provided at the course web site – read it.

In the paper of Lorenc (1986),

- Bayesian probabilistic arguments are used to derive idealized equations for finding the best analysis for NWP.
- The equations are discussed in the light of the physical characteristics of the NWP analysis problem:
 - (1) the predetermined nature of the basis for the analysis,
 - (2) the need for approximation because of large-order systems,
 - (3) the underdeterminacy of the problem when using observations alone, and
 - (4) the availability of prior relationships to resolve the underdeterminacy
- Prior relationships result from
 - (1) knowledge of the time evolution of the model (which together with the use of a time distribution of observations constitutes four-dimensional data assimilation);
 - (2) knowledge that the atmosphere varies slowly (leading to balance relationships);
 - (3) other nonlinear relationships coupling parameters and scales in the atmosphere.
- The paper also discussed other DA methods, including variational techniques, smoothing splines, Kriging, optimal interpolation, successive corrections, constrained initialization, the Kalman-Bucy filter, and adjoint model data assimilation.

- The paper showed that all these methods are related to the idealized analysis, and hence to each other. The paper also gave opinions on when particular methods might be more appropriate and discussed appropriate choices of parameters in the practical methods.

Remark: Derivative of a quadratic scalar function of a vector:

The general form of a scalar quadratic function is $F(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{d}^T \mathbf{x} + c$ where \mathbf{A} is a symmetric matrix, \mathbf{d} is a vector and c a scalar.

The gradient of this scalar function with respect to vector \mathbf{x} is defined as

$$\nabla_{\mathbf{x}} F = \frac{\partial F}{\partial \mathbf{x}} = \begin{bmatrix} \partial F / \partial x_1 \\ \partial F / \partial x_2 \\ \vdots \\ \partial F / \partial x_n \end{bmatrix},$$

which is a column vector. To find the gradient, we use the following properties of the gradient with respect to \mathbf{x} :

$$\boxed{\nabla(\mathbf{d}^T \mathbf{x}) = \nabla(\mathbf{x}^T \mathbf{d}) = \mathbf{d}} \quad (\text{since } \nabla_{\mathbf{x}} \mathbf{x}^T = \mathbf{I}, \text{ the identity matrix}), \text{ or}$$

$$\nabla(\mathbf{d}^T \mathbf{x}) = \nabla(d_1 x_1 + d_2 x_2 + \dots d_n x_n) = \begin{bmatrix} \partial(d_1 x_1 + d_2 x_2 + \dots d_n x_n) / \partial x_1 \\ \partial(d_1 x_1 + d_2 x_2 + \dots d_n x_n) / \partial x_2 \\ \cdot \\ \partial(d_1 x_1 + d_2 x_2 + \dots d_n x_n) / \partial x_n \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \cdot \\ d_n \end{bmatrix} = \mathbf{d}.$$

And

$$\nabla(\mathbf{x}^T \mathbf{A} \mathbf{x}) = 2\mathbf{A} \mathbf{x}.$$

We can see this by taking, say, time derivative of \mathbf{x} :

$$\frac{d(\mathbf{x}^T \mathbf{A} \mathbf{x})}{dt} = \dot{\mathbf{x}}^T \mathbf{A} \mathbf{x} + \mathbf{x}^T \mathbf{A} \dot{\mathbf{x}} = (\dot{\mathbf{x}}^T \mathbf{A} \mathbf{x})^T + \mathbf{x}^T \mathbf{A} \dot{\mathbf{x}} = \mathbf{x}^T \mathbf{A} \dot{\mathbf{x}} + \mathbf{x}^T \mathbf{A} \dot{\mathbf{x}} = 2\mathbf{x}^T \mathbf{A} \dot{\mathbf{x}}$$

$$d(\mathbf{x}^T \mathbf{A} \mathbf{x}) = 2\mathbf{x}^T \mathbf{A} d\mathbf{x} = 2(\mathbf{A} \mathbf{x})^T d\mathbf{x}$$

Because the differential of function $F(\mathbf{x}) = F(x_1, x_2, \dots, x_n)$ is

$$dF = \frac{\partial F}{\partial x_1} dx_1 + \frac{\partial F}{\partial x_2} dx_2 + \dots + \frac{\partial F}{\partial x_n} dx_n = \begin{bmatrix} \frac{\partial F}{\partial x_1} & \frac{\partial F}{\partial x_2} & \dots & \frac{\partial F}{\partial x_n} \end{bmatrix} \begin{bmatrix} dx_1 \\ dx_2 \\ \cdot \\ dx_n \end{bmatrix} = [\nabla_x F]^T d\mathbf{x}$$

In our current case, $F = \mathbf{x}^T \mathbf{A} \mathbf{x}$, therefore

$$\nabla(\mathbf{x}^T \mathbf{A} \mathbf{x}) = 2\mathbf{A} \mathbf{x}$$

Therefore, $F(x) = \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{d}^T \mathbf{x} + c$,

$$\nabla F(\mathbf{x}) = \mathbf{A} \mathbf{x} + \mathbf{d}, \quad \nabla^2 F(\mathbf{x}) = \mathbf{A} \quad \text{and} \quad \delta F = (\nabla F)^T \delta \mathbf{x}$$

The above will be used when deriving the gradient of quadratic-form cost function with respect to the analysis variable \mathbf{x} , which is needed when we try to find 3DVAR or 4DVAR solution.

The variational problem: Find analysis x_a that minimizes J

The minimum of $J(x)$ in (5.1) is attained for $x=x_a$, i.e., the analysis is given by the solution of

$$\nabla_{\mathbf{x}} J(\mathbf{x}_a) = 0 \quad (5.2)$$

An exact solution can be obtained in the following way.

We expand observation increment or innovation in the second term of Eq. (5.1) around background \mathbf{x}_b and linearize H assuming state variable \mathbf{x} and the final analysis \mathbf{x}_a remains close enough to ...:

$$\mathbf{y}_o - H(\mathbf{x}) = \mathbf{y}_o - H(\mathbf{x}_b + (\mathbf{x} - \mathbf{x}_b)) \approx \{\mathbf{y}_o - H(\mathbf{x}_b)\} - \mathbf{H}(\mathbf{x} - \mathbf{x}_b) \quad (5.3)$$

Plugging (5.3) into (5.1) we get

$$2J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + [\{\mathbf{y}_o - H(\mathbf{x}_b)\} - \mathbf{H}(\mathbf{x} - \mathbf{x}_b)]^T \mathbf{R}^{-1} [\{\mathbf{y}_o - H(\mathbf{x}_b)\} - \mathbf{H}(\mathbf{x} - \mathbf{x}_b)] \quad (5.4)$$

Expanding the products, and using the rules for transpose of matrix products:

$$\begin{aligned} 2J(\mathbf{x}) = & (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + (\mathbf{x} - \mathbf{x}_b)^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} (\mathbf{x} - \mathbf{x}_b) - \{\mathbf{y}_o - H(\mathbf{x}_b)\}^T \mathbf{R}^{-1} \mathbf{H} (\mathbf{x} - \mathbf{x}_b) \\ & - (\mathbf{x} - \mathbf{x}_b)^T \mathbf{H}^T \mathbf{R}^{-1} \{\mathbf{y}_o - H(\mathbf{x}_b)\} + \{\mathbf{y}_o - H(\mathbf{x}_b)\}^T \mathbf{R}^{-1} \{\mathbf{y}_o - H(\mathbf{x}_b)\} \end{aligned} \quad (5.5)$$

The cost function is a quadratic function of the analysis increment $\delta\mathbf{x} = (\mathbf{x} - \mathbf{x}_b)$:

$$J(\mathbf{x}) = \frac{1}{2} \delta\mathbf{x}^T (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}) \delta\mathbf{x} - \{\mathbf{y}_o - H(\mathbf{x}_b)\}^T \mathbf{R}^{-1} \mathbf{H} \delta\mathbf{x} + \frac{1}{2} \{\mathbf{y}_o - H(\mathbf{x}_b)\}^T \mathbf{R}^{-1} \{\mathbf{y}_o - H(\mathbf{x}_b)\} \quad (5.1a)$$

The above is actually called the *incremental form* of cost function because it's in terms of analysis increment $\delta\mathbf{x}$.

Remember a remark given earlier: Given a quadratic function $F(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{d}^T \mathbf{x} + c$,

where \mathbf{A} is a symmetric matrix, \mathbf{d} is a vector and c a scalar, the gradient is given by $\nabla F(\mathbf{x}) = \mathbf{A} \mathbf{x} + \mathbf{d}$. The gradient of the cost function J wrt $\delta\mathbf{x}$ is

$$\nabla J(\mathbf{x}) = \{\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}\} \delta\mathbf{x} - \mathbf{H}^T \mathbf{R}^{-1} \{\mathbf{y}_o - H(\mathbf{x}_b)\} \quad (5.6)$$

We now set $\nabla J(\mathbf{x}_a) = 0$ to ensure that J is a minimum, and obtain an equation for

$$\delta\mathbf{x}_a = (\mathbf{x}_a - \mathbf{x}_b)$$

$$(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}) \delta\mathbf{x}_a = \mathbf{H}^T \mathbf{R}^{-1} \{\mathbf{y}_o - H(\mathbf{x}_b)\} \quad (5.7)$$

or

$$\boxed{\mathbf{x}_a = \mathbf{x}_b + (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} \{\mathbf{y}_o - H(\mathbf{x}_b)\}} \quad (5.8)$$

or, in incremental form

$$\delta \mathbf{x}_a = (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} \delta \mathbf{y}_o \quad (5.8)'$$

where $\delta \mathbf{y}_o$ is the innovation vector, as defined earlier.

Eq. (5.8) is the solution of the variational (3D-Var) analysis problem, because we took a variation of the cost function, and obtained the solution by requiring that the variation to be zero for all possible variations in \mathbf{x} .

The problem is, solution (5.9) is nowhere near being easy to evaluate, because the involvement of the inverse of both background and analysis error covariance matrices. In fact, it's even harder to evaluate than the OI solution, which is

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

In practice the solution is obtained through minimization algorithms for $J(x)$ using iterative methods for minimization such as conjugate gradient or quasi-Newton.

We will discuss the typical iterative procedure for minimizing J more later.

Note that the *control variable* for the minimization (i.e., the variable with respect to which we are minimizing the cost function J) is now the *analysis*, \mathbf{x} , not the *weights*, \mathbf{W} , as in OI.

The *equivalence* between

- the minimization of the total analysis error variance (finding the optimal weights through a least squares approach – OI approach), and
- the 3D variational cost function approach (finding the optimal analysis that minimizes the distance to the observations weighted by the inverse of the error variance)

is an important property, which we will show in next section.

Equivalence between the OI and 3D-Var statistical problems

We now demonstrate the equivalence of the 3D-Var solution to the OI analysis solution obtained earlier.

We have to show that the weight matrix that multiplies the innovation $\{\mathbf{y}_o - H(\mathbf{x}_b)\} = \delta\mathbf{y}_o$ in (5.8) is the same as the weight matrix obtained with OI, i.e., that

$$\mathbf{W} = (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} \equiv (\mathbf{B} \mathbf{H}^T)(\mathbf{R} + \mathbf{H} \mathbf{B} \mathbf{H}^T)^{-1} \quad (5.9).$$

If the variables that we are observing are the same as the model variables, i.e., if $\mathbf{H} = \mathbf{H}^T = \mathbf{I}$, then it is rather straightforward to prove (5.9), using the rules for inverse and transpose of matrix products (exercise).

However, in general \mathbf{H} is rectangular, and non-invertible. The equality (5.9) can be proven by considering the following block matrix equation:

$$\begin{bmatrix} \mathbf{R} & \mathbf{H} \\ \mathbf{H}^T & -\mathbf{B}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \delta\mathbf{x}_a \end{bmatrix} = \begin{bmatrix} \delta\mathbf{y}_o \\ \mathbf{0} \end{bmatrix} \quad (5.10)$$

where \mathbf{w} is a vector to be further discussed later. We want to derive from (5.10) an equation of the form $\delta\mathbf{x}_a = \mathbf{W} \delta\mathbf{y}_o$.

Eliminating \mathbf{w} from both block rows in (5.10), we find that

$\delta\mathbf{x}_a = (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} \delta\mathbf{y}_0$, the 3D-Var version of the weight matrix.

Namely:

$$\begin{bmatrix} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{R} & \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \\ \mathbf{H}^T & -\mathbf{B}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \delta\mathbf{x}_a \end{bmatrix} = \begin{bmatrix} \mathbf{H}^T \mathbf{R}^{-1} \delta\mathbf{y}_0 \\ \mathbf{0} \end{bmatrix} \rightarrow$$

$$\begin{bmatrix} \mathbf{H}^T & \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \\ \mathbf{H}^T & -\mathbf{B}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \delta\mathbf{x}_a \end{bmatrix} = \begin{bmatrix} \mathbf{H}^T \mathbf{R}^{-1} \delta\mathbf{y}_0 \\ \mathbf{0} \end{bmatrix} \rightarrow$$

$$\begin{bmatrix} \mathbf{0} & \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} + \mathbf{B}^{-1} \\ \mathbf{H}^T & -\mathbf{B}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \delta\mathbf{x}_a \end{bmatrix} = \begin{bmatrix} \mathbf{H}^T \mathbf{R}^{-1} \delta\mathbf{y}_0 \\ \mathbf{0} \end{bmatrix} \rightarrow$$

$$(\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} + \mathbf{B}^{-1}) \delta\mathbf{x}_a = \mathbf{H}^T \mathbf{R}^{-1} \delta\mathbf{y}_0 \rightarrow$$

$$\delta\mathbf{x}_a = (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} \delta\mathbf{y}_0$$

On the other hand, eliminating $\delta\mathbf{x}_a$ from both block rows,

$$\begin{bmatrix} \mathbf{R} & \mathbf{H} \\ \mathbf{HBH}^T & -\mathbf{HBB}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \delta \mathbf{x}_a \end{bmatrix} = \begin{bmatrix} \delta \mathbf{y}_0 \\ \mathbf{0} \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{R} & \mathbf{H} \\ \mathbf{HBH}^T & -\mathbf{H} \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \delta \mathbf{x}_a \end{bmatrix} = \begin{bmatrix} \delta \mathbf{y}_0 \\ \mathbf{0} \end{bmatrix} \rightarrow$$

$$\begin{bmatrix} \mathbf{R} & \mathbf{H} \\ \mathbf{HBH}^T + \mathbf{R} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \delta \mathbf{x}_a \end{bmatrix} = \begin{bmatrix} \delta \mathbf{y}_0 \\ \delta \mathbf{y}_0 \end{bmatrix} \rightarrow$$

$$(\mathbf{HBH}^T + \mathbf{R})\mathbf{w} = \delta \mathbf{y}_0$$

we obtain an equation for the vector \mathbf{w} ,

$$\mathbf{w} = (\mathbf{R} + \mathbf{HBH}^T)^{-1} \delta \mathbf{y}_0.$$

From this, replacing \mathbf{w} in the second block row of (5.10),

$$\mathbf{H}^T \mathbf{w} - \mathbf{B}^{-1} \delta \mathbf{x}_a = \mathbf{0} \rightarrow \mathbf{H}^T (\mathbf{R} + \mathbf{HBH}^T)^{-1} \delta \mathbf{y}_0 - \mathbf{B}^{-1} \delta \mathbf{x}_a = \mathbf{0} \rightarrow$$

we obtain the OI version of the weight matrix: $\delta \mathbf{x}_a = \mathbf{BH}^T (\mathbf{HBH}^T + \mathbf{R})^{-1} \delta \mathbf{y}_0$.

This demonstrates the formal equivalence of the problems solved by 3D-Var and OI. Because the methods of solution are different, however, their results are different, and most centers have adopted the 3D-Var approach.

Another way of showing the equivalence of two formulae is (note that the **O** is our **R** and the x' is our \mathbf{x}_a):

$$\begin{aligned}
x' &= x_b + [B^{-1} + H^T O^{-1} H]^{-1} H^T O^{-1} (y - H x_b) \\
&= x_b + [B^{-1} (I + B H^T O^{-1} H)]^{-1} H^T O^{-1} (y - H x_b) \\
&= x_b + [I + B H^T O^{-1} H]^{-1} B H^T O^{-1} (y - H x_b)
\end{aligned} \tag{A1}$$

Let note $Z = B H^T O^{-1}$, then (A1) becomes:

$$\begin{aligned}
x' &= x_b + [I + Z H]^{-1} Z [y - H x_b] \\
&= x_b + [Z^{-1} [I + Z H]]^{-1} [y - H x_b] \\
&= x_b + [Z^{-1} + H]^{-1} [y - H x_b] \\
&= x_b + [[I + H Z] Z^{-1}]^{-1} [y - H x_b] \\
&= x_b + Z [I + H Z]^{-1} [y - H x_b]
\end{aligned} \tag{A5}$$

And thus:

$$\begin{aligned}
x' &= x_b + B H^T O^{-1} [I + H B H^T O^{-1}]^{-1} [y - H x_b] \\
&= x_b + B H^T [[I + H B H^T O^{-1}] O]^{-1} [y - H x_b]
\end{aligned} \tag{A6}$$

And finally:

$$x' = x_b + B H^T [O + H B H^T]^{-1} [y - H x_b] \tag{A7}$$

Why do we choose the above cost function?

The cost function (5.1) can also be derived based on a Bayesian approach.

Baye's theorem of probability states that "the posterior probability of an event A occurring, given that event B is known to have occurred, is proportional to the prior probability of A, multiplied by the probability of B occurring given that A is known to have occurred".

$$p(A/B) = \frac{p(B/A)p(A)}{p(B)}.$$

If A is event $\mathbf{x} = \mathbf{x}_t$, B is event $\mathbf{y} = \mathbf{y}_o$, then the probability the *a posteriori* probability distribution of the true field, given the new observations y_o , is

$$p(\mathbf{x} = \mathbf{x}_t | \mathbf{y} = \mathbf{y}_o) = \frac{p(\mathbf{y} = \mathbf{y}_o | \mathbf{x}) p_B(\mathbf{x} = \mathbf{x}_t)}{p(\mathbf{y} = \mathbf{y}_o)} \quad (5.11)$$

The Bayesian estimate of the true state is the one that maximizes the *a posteriori* probability (5.11), i.e., we want to find an analysis that has a maximum probability to occur and this probability can be obtained from the above equation if the probability distribution of the background (p_B) and the probability distribution of the observation $p(\mathbf{y} = \mathbf{y}_o | \mathbf{x})$ are known.

We assume that the true field is a realization of a random process defined by the prior (multi-dimensional Gaussian) probability distribution function (given the background field),

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

(Remember of the Gaussian probability distribution for scalar x :

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2} \frac{(x - x_0)^2}{\sigma^2}\right\}$$

The denominator in (5.11) is the “climatological” distribution of observations, and does not depend on the current true state x .

$$\text{Since } p(\mathbf{y}_o | \mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\mathbf{R}|^{1/2}} \exp\left\{-\frac{1}{2}[(\mathbf{y}_o - H(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y}_o - H(\mathbf{x}))]\right\}$$

therefore

$$p(\mathbf{x} | \mathbf{y}_o) \propto \frac{1}{(2\pi)^{p/2} |\mathbf{R}|^{1/2}} \frac{1}{(2\pi)^{n/2} |\mathbf{B}|^{1/2}} \exp \left\{ -\frac{1}{2} [(\mathbf{y}_o - H(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y}_o - H(\mathbf{x})) + (\mathbf{x}_b - \mathbf{x})^T \mathbf{B}^{-1} (\mathbf{x}_b - \mathbf{x})] \right\}$$

The \mathbf{x} that minimizes the minus of the exponent is the \mathbf{x} that maximizes the probability of the analysis!

Therefore the maximum of the a posteriori probability is attained when the numerator is maximum, and is given by the minimum of the cost function (5.1).

This is, of course, only valid when probability distributions are Gaussian!

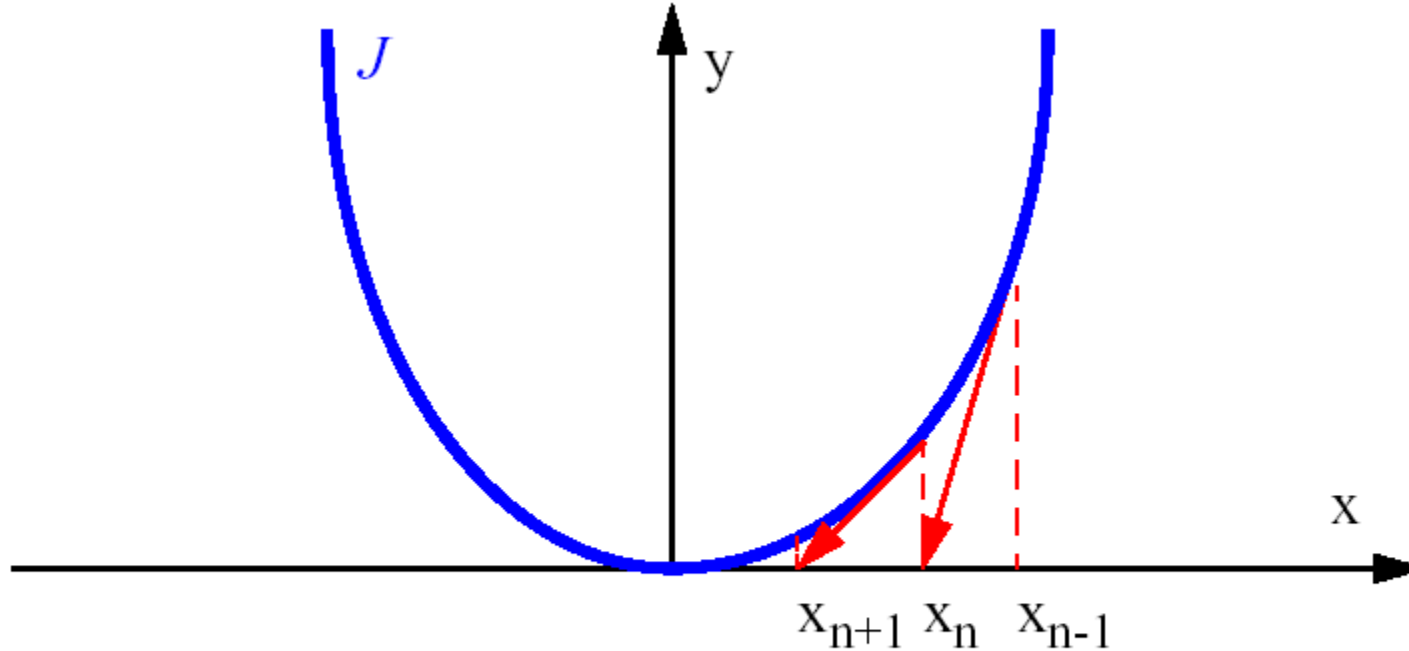
Reference: Lorenc (1986). Read also sections 5.3.2 and 5.5 of Kalnay's book.

Numerical minimization of the cost function

In practice, 3DVAR does not find its solution by evaluating Eq. (5.8)

$\mathbf{x}_a = \mathbf{x}_b + (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} \{\mathbf{y}_o - H(\mathbf{x}_b)\}$, or Eq. (5.8)'. The minimum is found by numerical minimization, which is usually an iterative procedure.

3DVAR minimization: An illustration of gradient descent



Linear model: J quadratic

Schematics of variational cost-function minimization

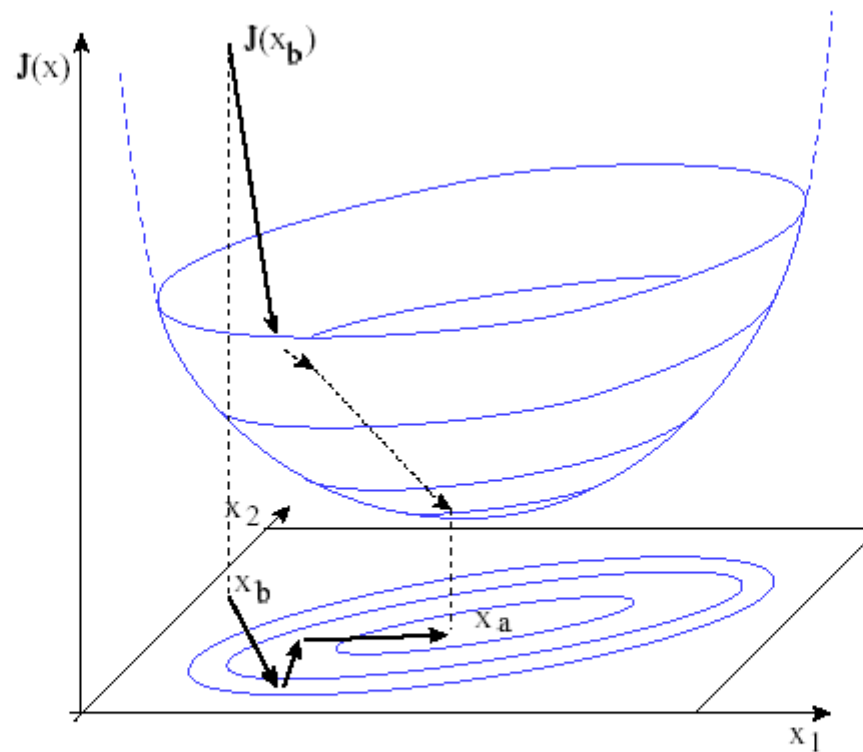


Figure 11. Schematic representation of the variational cost-function minimization (here in a two-variable model space): the quadratic cost-function has the shape of a paraboloid, or bowl, with the minimum at the optimal analysis x_a . The minimization works by performing several line-searches to move the control variable x to areas where the cost-function is smaller, usually by looking at the local slope (the gradient) of the cost-function.

Typical iterative procedure for numerically minimizing J :

1. Starting from background \mathbf{x}_b as the first guess, compute the cost function J according to (5.1) or (5.1a)
2. Compute the gradient of cost function J with respect to the analysis variable \mathbf{x} or $\delta\mathbf{x}$
3. Call certain optimization subroutine (using e.g., the conjugate gradient method), passing into the subroutine the cost function and gradient vector, and determine the amount of correction to \mathbf{x} or $\delta\mathbf{x}$, so that

$$\mathbf{x}^{n+1} = \mathbf{x}^n + \alpha f(\nabla J)$$

where n is the number of iteration, and α is the optimal step size factor. $f(\nabla J)$ is the optimal descent direction obtained by combining the gradients from several former iterations when using the conjugate gradient method. For the simple steepest decent algorithm, $f(\nabla J) = -\nabla J$. We should always have

$$J(\mathbf{x}^{n+1}) \leq J(\mathbf{x}^n).$$

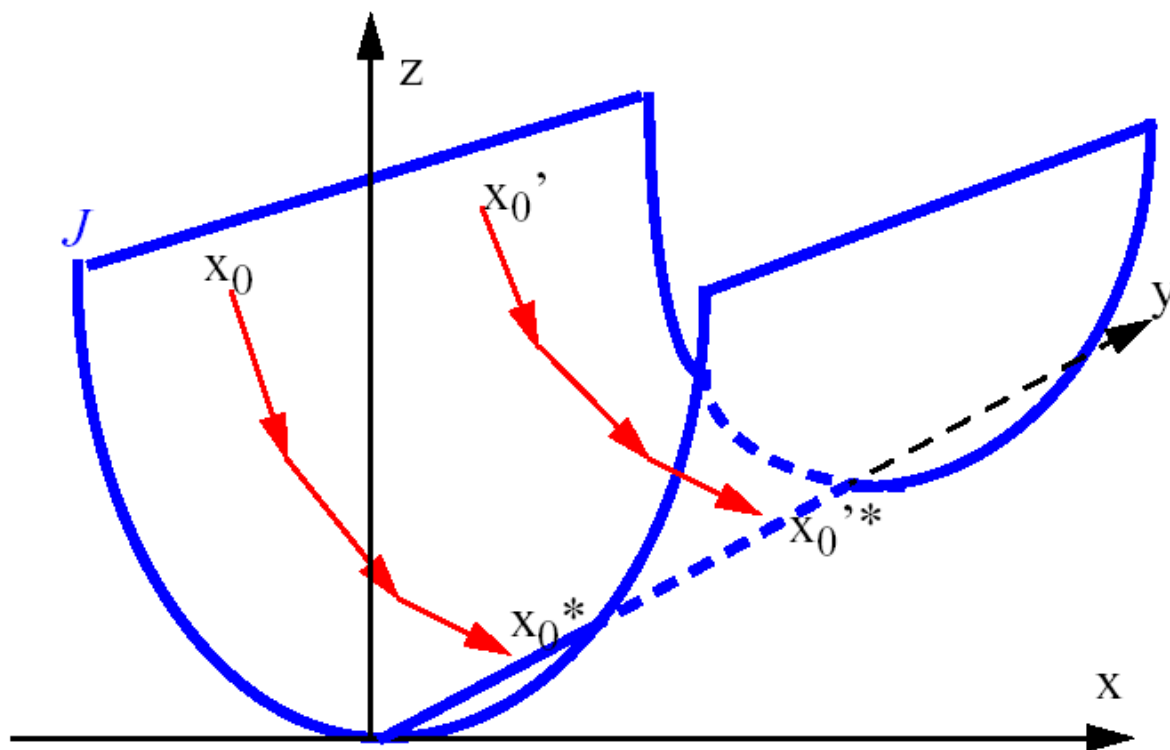
4. Check if the optimal solution has been found by computing the norm of the gradients or the value of J itself to see if either is less than a prescribed tolerance. If not, go back to step 1, with the updated value \mathbf{x}^{n+1} , and repeat the

steps until convergence is reached. The solution obtained at convergence is the optimal solution that minimizes J .

Several issues related to the minimization

1. **Undetermined problem** - In the absence of background, the problem is generally underdetermined.
2. **Non-linearity** - When the observations models are not linear, the uniqueness of the minimum is not guaranteed.
3. **Computational cost** - Because the minimization is an iterative process, it can be very costly

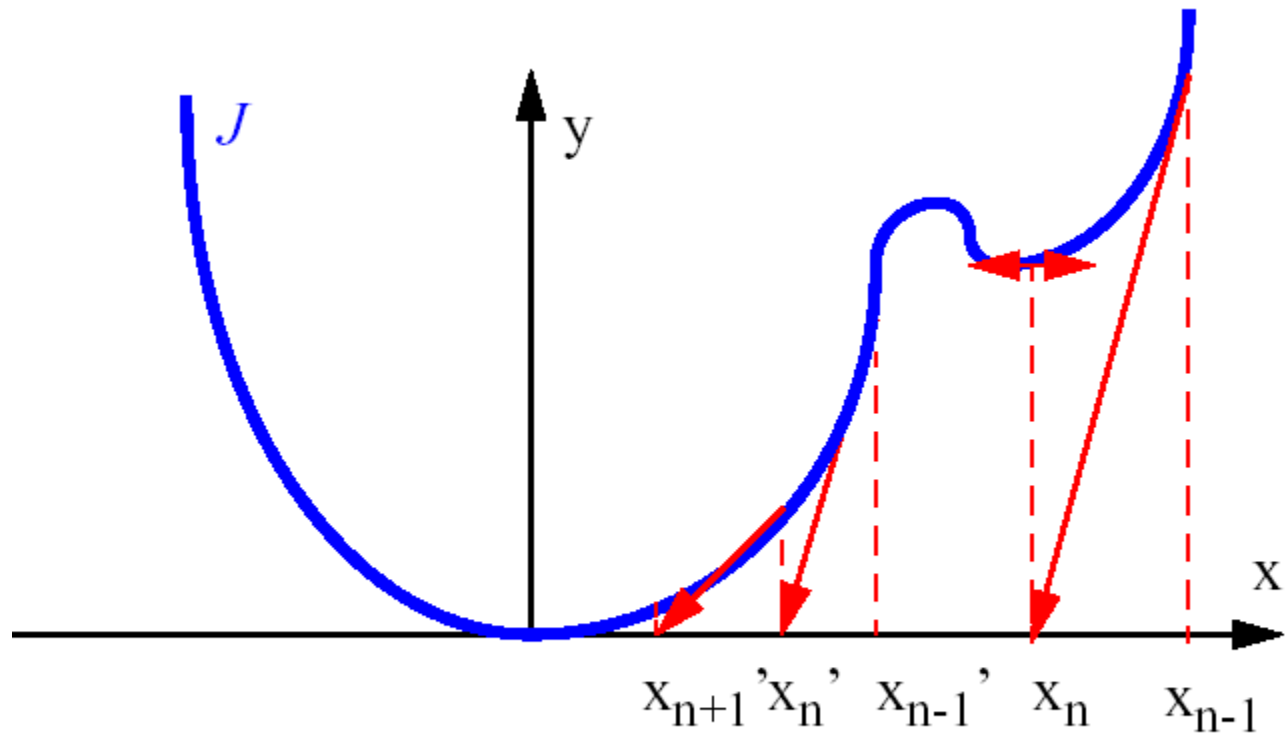
3D-VAR: Main issues: Missing background - undetermined problem



Without background term

The state vector is a 3 dimensional array $X = [x, y, z]$ but only $Y = [x, z]$ is observed. Different starting point x_0 leads to different solutions x_0^* .

3D-VAR: Main issues: Nonlinearities



non-linear effects

When observation operators are non-linear, the minimum is not unique.

3D-VAR: Main issues: Computational cost

Because the minimization is iterative, 3D-VAR can be very expensive

How to reduce the computational cost ?

- 1) Code optimization - through vectorization, multitasking and parallelization -
- 2) Incremental approach - Reduce the computations per iteration, and linearize around increasingly accurate 'first guess' state
- 3) Preconditioning - Reduce the number of iterations.

Incremental 3DVAR (and 4DVAR)

Write cost function J in (5.1) in terms of increment $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}_b$ and innovation $\delta\mathbf{y}_o = \mathbf{y}_o - H(\mathbf{x}_b)$:

$$2J(\mathbf{x}) = \delta\mathbf{x}^T \mathbf{B}^{-1} \delta\mathbf{x} + [\delta\mathbf{y}_o - \mathbf{H}\delta\mathbf{x}]^T \mathbf{R}^{-1} [\delta\mathbf{y}_o - \mathbf{H}\delta\mathbf{x}].$$

which is the same as Eq. (5.4) and involves linearization operation

$H(\mathbf{x}) \approx H(\mathbf{x}_b) + \mathbf{H}(\mathbf{x} - \mathbf{x}_b)$. The closer is \mathbf{x}_b to the final optimal analysis, the better is linearization approximation.

J is minimized to find $\delta\mathbf{x}^*$, a new ‘background’ is found $\mathbf{x}_b^* = \mathbf{x}_b + \delta\mathbf{x}^*$, the linearization is re-performed around the new ‘background’ which should be closer to the optimal solution therefore $\delta\mathbf{x}$ is smaller and the linearization more valid.

Another good thing about the incremental approach is that the nonlinear observation operator H in the innovation $\delta\mathbf{y}_o = \mathbf{y}_o - H(\mathbf{x}_b)$ can be calculated at different (usually higher) resolution from that of analysis. In the case of 4DVAR, the nonlinear forecast model is part of H , which can be run at higher resolution to achieve better accuracy while the analysis, which is the expensive part of 4DVAR, can be performed at lower resolution. This is the essence of the incremental approach initially proposed by Courtier et al (1994) and the computational saving made operational implementation at ECMRF possible.

Reference:

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

Important Note: In the above, we said a ‘new background’ is found by the variational minimization (the inner iteration loops) around which new linearization of H is performed. With this ‘new background’, ‘outer loop’ iterations can be performed where the ‘background’ can be updated several times. However, keep in mind that this ‘new background’ should only be used in the linearization of H . In the background term, \mathbf{B} still

represents the error covariance of the ‘original’ forecast background. Therefore, in the background term, the original background \mathbf{x}_b should be used.

In terms of equations, if we let $\mathbf{x}_f = \mathbf{x}_b + \delta\mathbf{x}_f$ be the ‘new background’ or new first guess, a result of the previous variational minimization,

$$\mathbf{x}_f = \mathbf{x}_b + \delta\mathbf{x}_f$$

$$\mathbf{x} - \mathbf{x}_f = (\mathbf{x}_b + \delta\mathbf{x}) - (\mathbf{x}_b + \delta\mathbf{x}_f) = \delta\mathbf{x} - \delta\mathbf{x}_f$$

$$\mathbf{x} - \mathbf{x}_b = \mathbf{x} - \mathbf{x}_f + \delta\mathbf{x}_f = \delta\mathbf{x} - \delta\mathbf{x}_f + \delta\mathbf{x}_f$$

$$2J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + [\{\mathbf{y}_o - H(\mathbf{x}_b)\} - \mathbf{H}(\mathbf{x} - \mathbf{x}_b)]^T \mathbf{R}^{-1} [\{\mathbf{y}_o - H(\mathbf{x}_b)\} - \mathbf{H}(\mathbf{x} - \mathbf{x}_b)]$$

$$2J(\mathbf{x}) = \delta\mathbf{x}^T \mathbf{B}^{-1} \delta\mathbf{x} + [\{\mathbf{y}_o - H(\mathbf{x}_f)\} - \mathbf{H}(\mathbf{x} - \mathbf{x}_f)]^T \mathbf{R}^{-1} [\{\mathbf{y}_o - H(\mathbf{x}_f)\} - \mathbf{H}(\mathbf{x} - \mathbf{x}_f)]$$

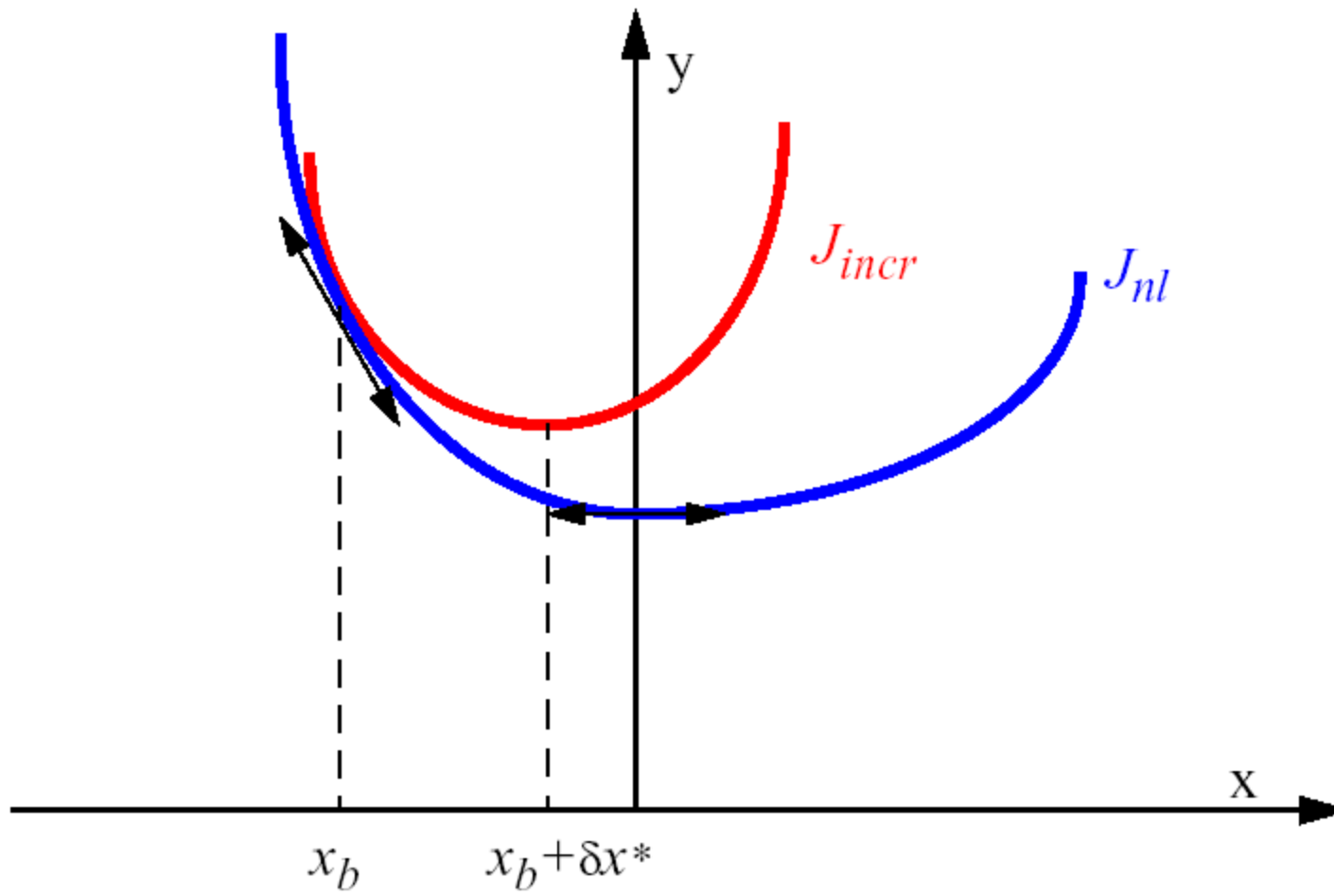
$$2J(\mathbf{x}) = (\delta\mathbf{x} - \delta\mathbf{x}_f + \delta\mathbf{x}_f)^T \mathbf{B}^{-1} (\delta\mathbf{x} - \delta\mathbf{x}_f + \delta\mathbf{x}_f) + [\{\mathbf{y}_o - H(\mathbf{x}_f)\} - \mathbf{H}(\delta\mathbf{x} - \delta\mathbf{x}_f)]^T \mathbf{R}^{-1} [\{\mathbf{y}_o - H(\mathbf{x}_f)\} - \mathbf{H}(\delta\mathbf{x} - \delta\mathbf{x}_f)]$$

$$- \mathbf{H}(\delta\mathbf{x} - \delta\mathbf{x}_f)]^T \mathbf{R}^{-1} [\{\mathbf{y}_o - H(\mathbf{x}_f)\} - \mathbf{H}(\delta\mathbf{x} - \delta\mathbf{x}_f)]$$

Let $\delta\mathbf{x} - \delta\mathbf{x}_f = \delta\mathbf{x}'$

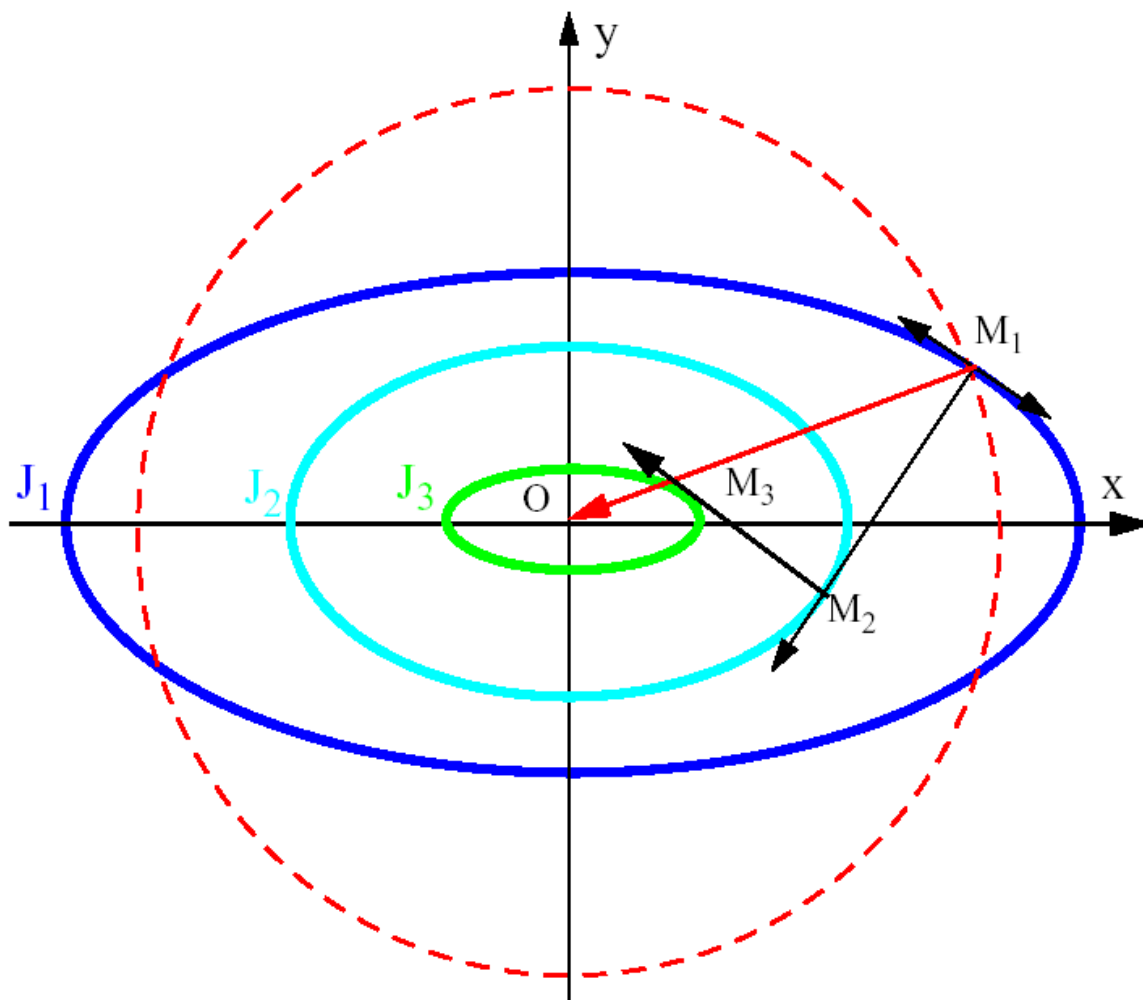
$$2J(\mathbf{x}) = (\delta\mathbf{x}' + \delta\mathbf{x}_f)^T \mathbf{B}^{-1} (\delta\mathbf{x}' + \delta\mathbf{x}_f) + [\{\mathbf{y}_o - H(\mathbf{x}_f)\} - \mathbf{H}\delta\mathbf{x}']^T \mathbf{R}^{-1} [\{\mathbf{y}_o - H(\mathbf{x}_f)\} - \mathbf{H}\delta\mathbf{x}']$$

which is a quadratic function of $\delta \mathbf{x}'$. A solution for $\delta \mathbf{x}'$ can be found using the variational minimization algorithm and new updated solution is then $\mathbf{x} = \mathbf{x}_b + \delta \mathbf{x}' + \delta \mathbf{x}_f$, which can serve as the new \mathbf{x}_f in the next outer iteration.



Incremental approach: linearization

Preconditioning: 2-D example



$$J = \frac{1}{2} \left(\frac{x^2}{a^2} + \frac{y^2}{b^2} \right), \quad M_2 = M_1 - \alpha \nabla J, \quad O = M_1 - \overline{OM_1}$$

Goal of Preconditioning: To find coordinate rotation so that $\overline{OM}_1 = R\nabla J$. Then we can find the minimum in one step!

It turns out that this is true when $a = b$ or a variable transform is performed of x ($x' = xb/a$) so that a becomes equal to $b \rightarrow$

$$J = \frac{1}{2b^2}(x'^2 + y^2).$$

Lorenc (1997, see also Huang 2000) proposed the following variable transform:

$$\delta\mathbf{x} = \sqrt{\mathbf{B}}\mathbf{v}$$

so that the cost function $J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}(\mathbf{y}_o - H(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y}_o - H(\mathbf{x}))$

becomes

$$J = \frac{1}{2}\mathbf{v}^T \mathbf{v} + \frac{1}{2}[H(\sqrt{\mathbf{B}}\mathbf{v} + \mathbf{x}_b) - \mathbf{y}_o]^T \mathbf{R}^{-1}[H(\sqrt{\mathbf{B}}\mathbf{v} + \mathbf{x}_b) - \mathbf{y}_o]$$

Transform 1) eliminated \mathbf{B}^{-1} and 2) improved the conditioning of the minimization (when \mathbf{B} dominates over \mathbf{R}).

Reference:

Lorenc, A. C., 1997: Development of an operational variational assimilation scheme. *J. Met Soc Japan*, **75**, 339-346.

Physical-space Statistical Analysis Scheme (PSAS)

Da Silva et al (1995) introduced another scheme related to 3D-Var and OI, where the minimization is performed in the (physical) space of the observations, rather than in the model space as in the 3D-Var scheme.

They solve the same OI/3D-Var equation (5.8) written as in the OI approach,

$$\delta \mathbf{x}_a = (\mathbf{B}\mathbf{H}^T)(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} \delta \mathbf{y}_o \quad (5.12)$$

but separate it into two steps:

$$\mathbf{w} = (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} \delta \mathbf{y}_o \quad (5.13)$$

followed by

$$\delta \mathbf{x}_a = (\mathbf{B}\mathbf{H}^T)\mathbf{w}. \quad (5.14)$$

The first step is the most computing intensive, and is solved by minimization of a cost function:

$$J(\mathbf{w}) = \frac{1}{2} \mathbf{w}^T (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T) \mathbf{w} - \mathbf{w}^T \delta \mathbf{y}_o \quad (5.15)$$

Note that \mathbf{w} is a vector of length p , the number of observations. If p is much smaller than the state vector length n , this is a more efficient method to achieve results similar to those of 3D-Var.

The exercise in the previous section indicates that the intermediate solution vector \mathbf{w} is also given by

$$\mathbf{w} = (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} \delta\mathbf{y}_o$$

therefore the \mathbf{w} here is the same \mathbf{w} in Eq. (5.10). From Eq. (5.10), we have

$$\mathbf{R}\mathbf{w} + \mathbf{H}\delta\mathbf{x}_a = \delta\mathbf{y}_o \rightarrow$$

$$\mathbf{w} = \mathbf{R}^{-1}(\delta\mathbf{y}_o - \mathbf{H}\delta\mathbf{x}_a) = \mathbf{R}^{-1}[\mathbf{y}_o - \{H(\mathbf{x}_b) + \mathbf{H}(\mathbf{x}_a - \mathbf{x}_b)\}] = \mathbf{R}^{-1}(\mathbf{y}_o - H(\mathbf{x}_a)), \quad (5.16)$$

i.e, it is the misfit of the observations to the analysis weighted by the inverse of the observation covariance matrix.

Other 3DVAR topics

3DVAR without inversion of \mathbf{B} (see Huang 2000 MWR paper, also Lorenc's approach that also does preconditioning)

Huang (2000) used variable transform

$$\mathbf{v} = \mathbf{B}^{-1} \delta \mathbf{x}$$

So that

$$J = \frac{1}{2} \mathbf{v}^T \mathbf{B} \mathbf{v} + \frac{1}{2} [H(\mathbf{B} \mathbf{v} + \mathbf{x}_b) - \mathbf{y}_o]^T \mathbf{R}^{-1} [H(\mathbf{B} \mathbf{v} + \mathbf{x}_b) - \mathbf{y}_o].$$

This avoids inversion of \mathbf{B} in the background term. Such a formulation is used in NCEP operational DA system GSI although it does not improve the conditioning of the background term.

Use of filter to model background error covariances

Use explicit Gaussian filter (Huang 2000). Using implicit recursive filter (Lorenc 1992; Hayden and Purser 1995)

Huang, X.-Y., 2000: Variational analysis using spatial filters. *Mon. Wea. Rev.*, **128**, 2588–2600.

Lorenc, A. C., 1992: Iterative analysis using covariance functions and filters. *Quart. J. Roy. Met. Soc.*, **118**, 569-591.

Hayden, C. M. and R. J. Purser, 1995: Recursive filter objective analysis of meteorological fields: applications to NESDIS operational processing. *J. Appl. Meteor.*, **34**, 3-15.

Anisotropic and/or spatially inhomogeneous background error covariances

Inclusion of equation constraints in the cost function (EC notes)

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

- the J_c term.

For example, if we believe the horizontal flow is nearly non-divergence, we can define

$$J_c = w \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right)^2 \text{ where } w \text{ a weight the determines the how strictly this (weak)}$$

constraint is enforced. It's a weak constraint because this condition is not strictly enforced.

Final comments on the relative advantages between 3D-Var, PSAS and OI

Although 3D-Var, Optimal Interpolation, and PSAS have been shown to formally solve the same problem, there are important differences in the method of solution.

As indicated before, in practice OI requires introducing a number of approximations, and solving the analysis locally, grid point by grid point, or small volume by small volume (Lorenc, 1981). This in turn requires using a "radius of influence" and selecting only the stations closest to the grid point or volume being analyzed. The background error covariance matrix has to be also locally approximated.

Despite their formal equivalence, *3D-Var* (and the closely related *PSAS*) *has several important advantages with respect to OI*, because the cost function is minimized using global minimization algorithms, and as a result it makes many of the simplifying approximations required by OI unnecessary.

No need for data selection and background error localization

In 3D-Var (and PSAS) there is no data selection, all available data are used simultaneously. This avoids jumpiness in the boundaries between regions that have selected different observations.

In OI, the background error covariance has been crudely obtained assuming, for example, separability of the correlation into products of horizontal and vertical Gaussian correlations, and that the background errors are in geostrophic balance. The background

error covariance matrix for 3D-Var, although it may still require simplifying assumptions, can be defined with a more general, global approach, rather than the local approximations used in OI.

Easy incorporation of balance constraints

It is possible to add constraints to the cost function without increasing the cost of the minimization much. For example, Parrish and Derber (1992) included a "penalty" term in the cost function forcing *simultaneously* the analysis increments to approximately satisfy the linear global balance equation. With the global balance equation, the NCEP global model spin up was reduced by more than an order of magnitude compared to the results obtained with OI. With the 3D-Var it became unnecessary for the first time to perform a separate initialization step in the analysis cycle.

The use of nonlinear observation operators

It is possible to incorporate important nonlinear relationships between observed variables and model variables in the H operator in the minimization of the cost function (1) by performing "inner" iterations with the linearized \mathbf{H} observation operator kept constant and "outer" iterations in which it is updated. This is harder to do in the OI approach.

The introduction of 3D-Var has allowed 3D variational assimilation of radiances (Derber and Wu, 1998). In this approach, there is no attempt to perform retrievals, and instead, each satellite sensor is taken as an independent observation with uncorrelated errors. As a result, for each satellite observation spot (pixel), even if some channel measurements are

rejected because of cloud contamination, others may still be used. In addition, because all data are assimilated simultaneously, information from one channel at a certain location can influence the use of satellite data at a different geographical location. The quality control of the observations becomes easier and more reliable when it is made in the space of the observations than in the space of the retrievals.

It is possible to include quality control of the observations within the 3D-VAR analysis (see section 5.8 of Kalnay book)

Cohn et al (1998) have shown that the observation-space form of 3D-Var (PSAS) offers opportunities, through the grouping of data, to improve the preconditioning of the problem in an iterative solution.

Because PSAS cost function is formulated in an ‘observation space’, it is difficult to incorporate equation constraint in PSAS, which is an important disadvantage of PSAS. For this reason, PSAS or observation space 3DVAR is used much less often (US Navy’s and NASA’s VAR DA systems are formulated in the observation space).