Variational Assimilation (§5.5)

We now turn from Optimal Interpolation to another approach to objective analysis, the variational assimilation technique.

This method is of growing popularity and is now in use in several major NWP centres.

Variational assimilation has been shown to yield significant improvements in the quality of numerical forecasts.

It has also been invaluable for re-analysis:
The ERA-40 Project at ECMWF was carried out using the 3D-Var system.

The Cost Function J

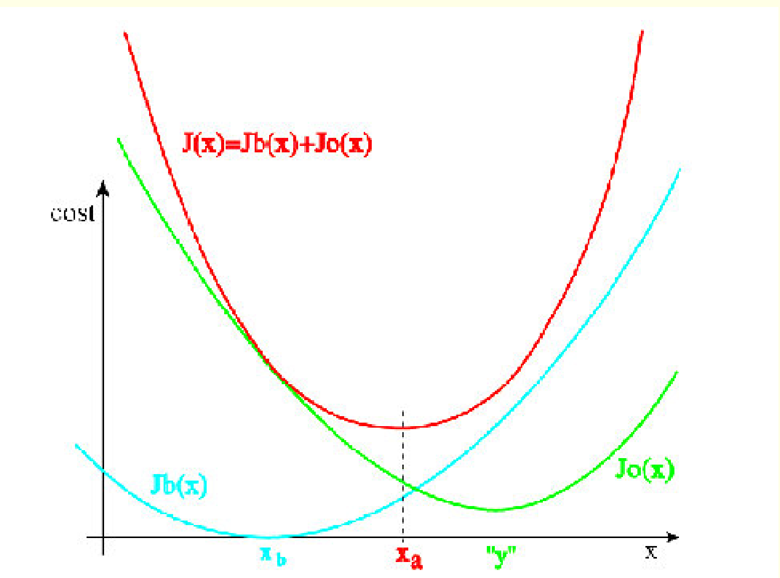
We saw, for the "two-temperature problem", an important equivalence between the least squares approach and the variational approach.

The same equivalence holds for the full 3-dimensional case.

Lorenc (1986) showed that the OI solution is equivalent to a specific variational assimilation problem: Find the optimal analysis $\mathbf{x}_a$ field that minimizes a (scalar) cost function.

The cost function is defined as the (weighted) distance between $\mathbf{x}$ and the background $\mathbf{x}_b$, *plus the (weighted) distance to the observations* $\mathbf{y}_o$,:

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



Schematic representation of the cost function in a simple one-dimensional case. J_b and J_o respectively tend to pull the analysis towards the background $\mathbf{x}_b$ and the observation $\mathbf{y}$.

Again, the cost function is

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

The minimum of $J(\mathbf{x})$ is attained for $\mathbf{x} = \mathbf{x}_a$ such that

$$\frac{\partial J}{\partial \mathbf{x}} = \nabla_{\mathbf{x}} J(\mathbf{x}_a) = 0 \qquad\qquad (n \times 1)$$

Assuming the analysis is close to the truth, we write

$$\mathbf{x} = [\mathbf{x}_b + (\mathbf{x}-\mathbf{x}_b)]$$

and assume that $\mathbf{x}-\mathbf{x}_b$ is small.

Then we can linearize the observation operator:

$$[\mathbf{y}_o - H(\mathbf{x})] = \mathbf{y}_o - H[\mathbf{x}_b + (\mathbf{x}-\mathbf{x}_b)] = \{\mathbf{y}_o - H(\mathbf{x}_b)\} - \mathbf{H}\cdot(\mathbf{x}-\mathbf{x}_b)$$

We now substitute this into the cost function:

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

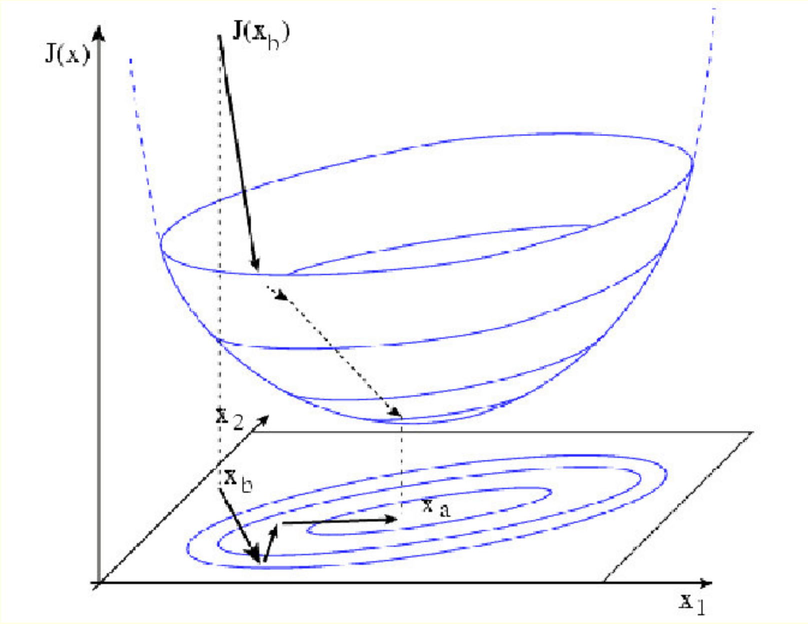
The result is:

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

Expanding the products, we get

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

The cost function is a quadratic function of the analysis increments $(\mathbf{x}-\mathbf{x}_b)$.



Schematic of the cost function in two dimensions. The minimum is found by moving down-gradient in discrete steps.

We need to compute the gradient of J.

We use the following **Lemma:**

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*** $\nabla F(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{d}$.

⋆ ⋆ ⋆

Proof:

$$F(\mathbf{x}) = \frac{1}{2}\sum_i\sum_j A_{ij}x_ix_j + \sum_i d_ix_i + c$$

So the derivative w.r.t. x_k is

$$\frac{\partial F}{\partial x_k} = \frac{1}{2}\sum_j A_{kj}x_j + \frac{1}{2}\sum_j A_{ik}x_i + d_k = \sum_j A_{kj}x_j + d_k$$

Q.E.D.

Recall the cost function was

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

Combining the first two terms, we get

$$\begin{aligned}2J(\mathbf{x}) &= (\mathbf{x}-\mathbf{x}_b)^T[\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}](\mathbf{x}-\mathbf{x}_b)\\ &- \{\mathbf{y}_o - H(\mathbf{x}_b)\}^T\mathbf{R}^{-1}\mathbf{H}(\mathbf{x}-\mathbf{x}_b)\\ &- (\mathbf{x}-\mathbf{x}_b)^T\mathbf{H}^T\mathbf{R}^{-1}\{\mathbf{y}_o - H(\mathbf{x}_b)\}\\ &+ \{\text{Term independent of } \mathbf{x}\}\end{aligned}$$

The gradient of the cost function J with respect to $\mathbf{x}$ is

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

Repeat: The gradient of J with respect to $\mathbf{x}$ is

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

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

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

We can write this as:

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

This is the solution of the 3-dimensional variational (3D-Var) analysis problem.

It looks similar to the OI result, but the weight matrix is

$$\mathbf{W} = [\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}]^{-1}\mathbf{H}^T\mathbf{R}^{-1}$$

The equivalence is ***not obvious***.

Again, the variational analysis is

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

It is a formal solution: the computation $\mathbf{x}_a$ requires the inversion of a huge matrix, which is impractical.

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

Note that the control variable for the minimization is now the analysis, not the weights as in OI.

The equivalence between the minimization of the analysis error variance and the three-dimensional variational cost function approach is an important property.

Conclusion of the foregoing

Minimization

In practical 3D-Var, we do not invert a huge matrix.

We find the minimum of $J(\mathbf{x})$ by computing the cost function for a range of values of $\mathbf{x}$ and using an optimization technique.

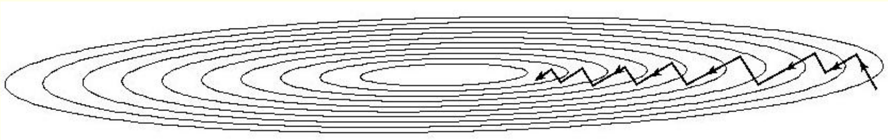
The idea is to "proceed downhill" as quickly as possible.

Examples are the Steepest Descent algorithm, Newton's method, and the Conjugate Gradient algorithm.

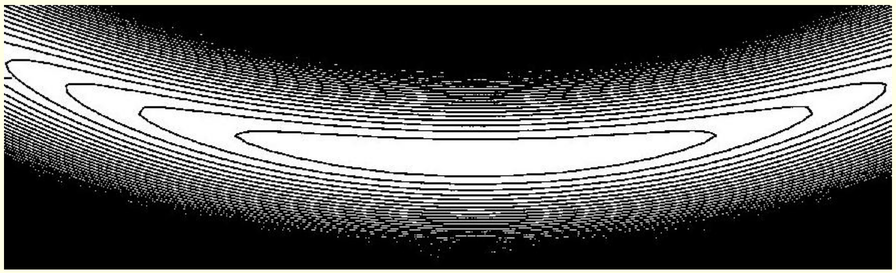
For a selection of techniques, see Numerical Recipes, which may be inspected online before purchase.

The location of the minimum depends greatly on the nature of the function J.

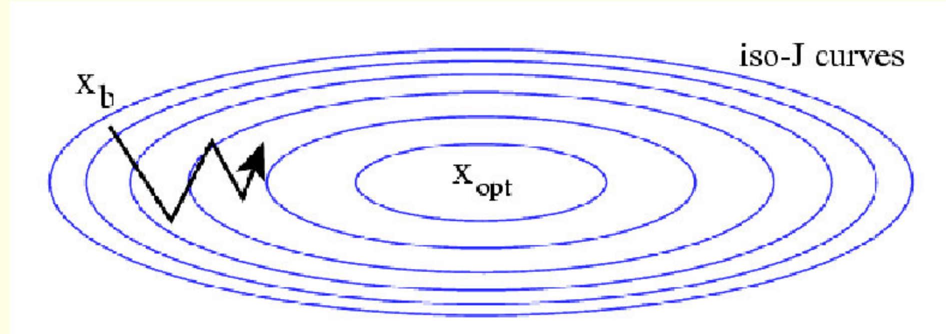
As an example, for two dimensions, we consider the shape of the "surface" $J = J(x, y)$.



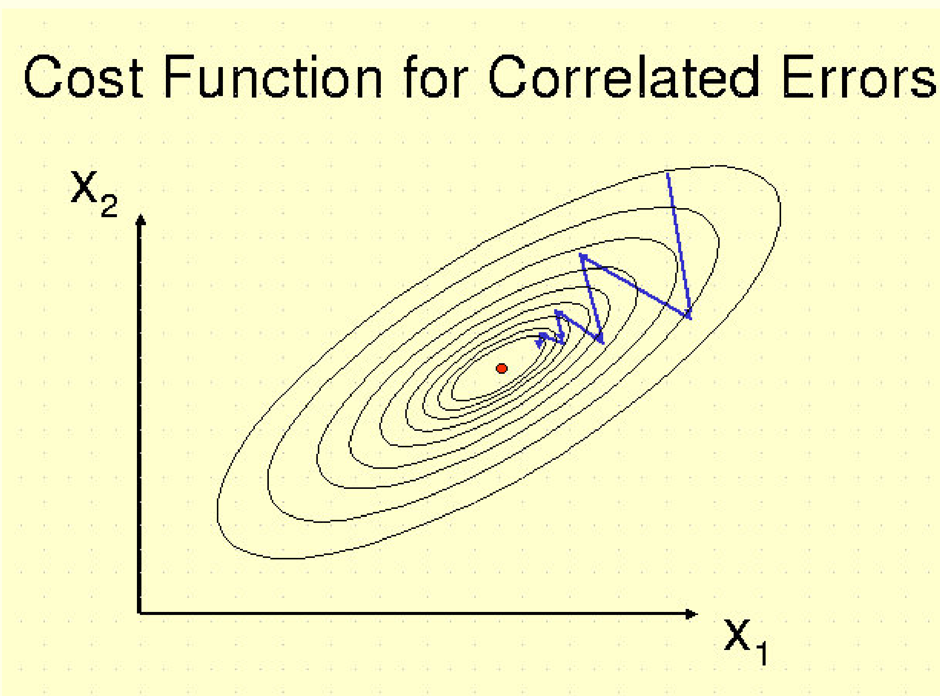
For a purely elliptic surface, the minimum is easily located.

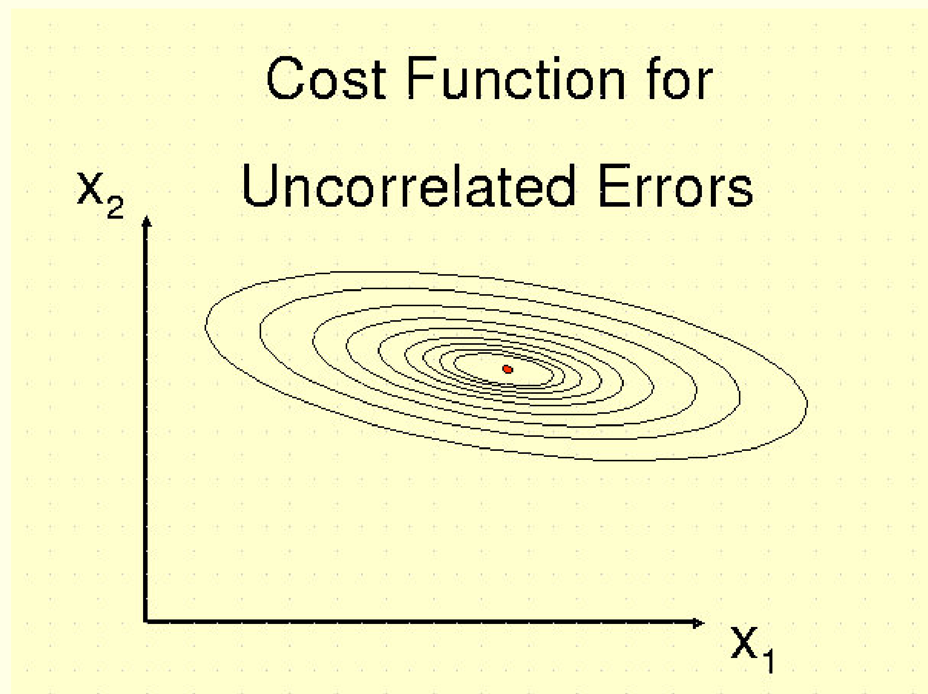


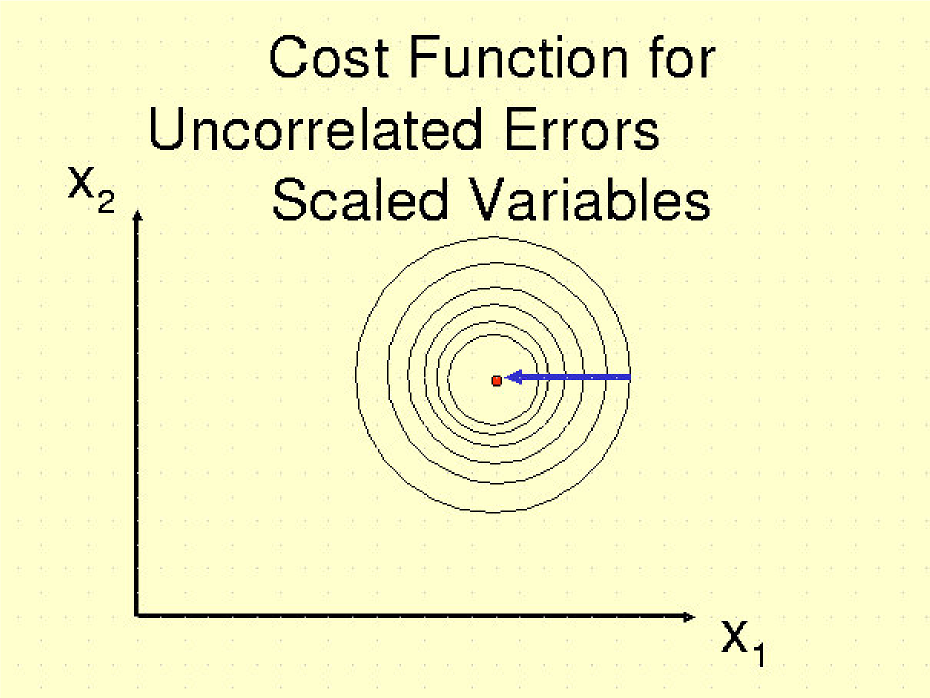
For a banana shaped surface, the minimum is much harder to find.



The "narrow-valley" effect. The minimization can spend many iterations zigzagging towards the minimum.







Equivalence between OI and 3D-Var

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

$$\underbrace{(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}}_{\text{(3D-Var)}} = \underbrace{(\mathbf{B}\mathbf{H}^T)(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}}_{\text{(OI)}}$$

This identity is a variant of the Sherman–Morrison–Woodbury formula (Golub and Van Loan, 1996).

The mathematical proof (an elegant demonstration) is given in Kalnay, §5.5.1.

It demonstrates the formal equivalence of the problems solved by 3D-Var and OI.

However, because the methods of solution are different, their results are different, and many centers have now adopted the 3D-Var approach.

Straightforward Demonstration of Equivalence

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

Inverse of left-hand side:

$$\mathbf{R}\mathbf{H}^{-T}(\mathbf{B}^{-1} + \mathbf{H}^{-T}\mathbf{R}^{-1}\mathbf{H})$$

$$(\mathbf{R}\mathbf{H}^{-T}\mathbf{B}^{-1} + \mathbf{H})$$

$$(\mathbf{R}\mathbf{H}^{-T}\mathbf{B}^{-1} + \mathbf{H})\,\mathbf{B}\mathbf{H}^T\mathbf{H}^{-T}\mathbf{B}^{-1}$$

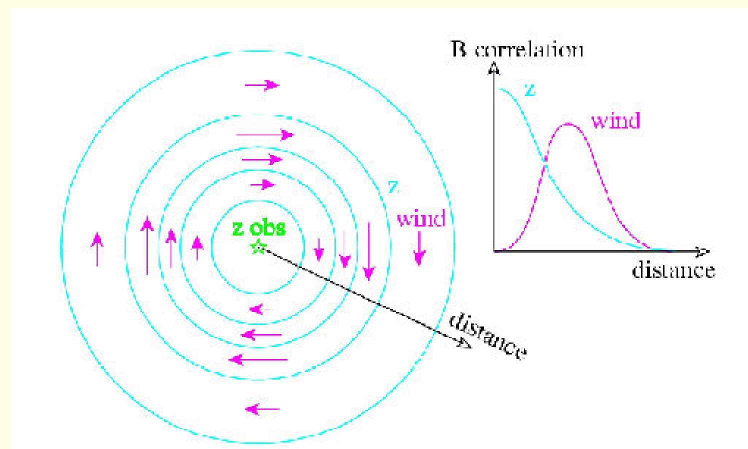
$$(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)\mathbf{H}^{-T}\mathbf{B}^{-1}$$

The inverse of this is equal to the right-hand side

$$\mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}$$

Thus 3D-Var and OI are *mathematically* equivalent.

J_B: The Conventional Method



Typical "structure function" used in OI. The autocorelation of height is an isotropic Gaussian function. By geostrophy, the cross correlation with the tangential wind is maximum where the radial gradient of the height correlation is maximum.

J_B: The "NMC method"

Most NWP centres have now adopted the "NMC method" for estimating the forecast error covariance.

The structure of the background error covariance is estimated as the average difference between two short-range model forecasts verifying at the same time.

$$\mathbf{B} \approx \alpha E\{[\mathbf{x}_f(48\,\mathbf{h}) - \mathbf{x}_f(24\,\mathbf{h})][\mathbf{x}_f(48\,\mathbf{h}) - \mathbf{x}_f(24\,\mathbf{h})]^T\}$$

The magnitude of the covariance is then appropriately scaled.

The model–forecast differences themselves provide a multivariate global forecast difference covariance.

This method has been shown to produce better results than previous estimates computed from *forecast minus observation* estimates.

Comparison of 3D-Var and OI

3D-Var has several important advantages with respect to OI, because the cost function J is minimized using global minimization algorithms.

As a result, many of the simplifying approximations required by OI are unnecessary in 3D-Var.

- In 3D-Var there is no data selection; all available data are used simultaneously. This avoids jumpiness in the boundaries between regions that have selected different observations.
- The background error covariance matrix for 3D-Var can be defined with a more general, global approach, rather than the local approximations used in OI.
- It is possible to add constraints to the cost function without increasing the cost of the minimization. These can be used to control spurious noise.

- For example, we may require the analysis increments to approximately satisfy the linear global balance equation.
- With the implementation of 3D-Var at NCEP, it became unnecessary to perform a separate initialization step in the analysis cycle.
- It is possible to incorporate nonlinear relationships between observed variables and model variables in the H operator. This is harder to do in the OI approach.
- 3D-Var has allowed three-dimensional variational assimilation of radiances.
- 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.

End of §5.5