Optimal Interpolation (§5.4)

We now generalize the least squares method to obtain the OI equations for vectors of observations and background fields.

These equations were derived originally by Eliassen (1954), However, Lev Gandin (1963) derived the multivariate OI equations independently and applied them to objective analysis in the Soviet Union.

OI became the operational analysis scheme of choice during the 1980s and 1990s. Indeed, it is still widely used.

Later, we show that 3D-Var is equivalent to the OI method, although the method for solving it is quite different.

Optimal interpolation (OI)

We now consider the complete NWP operational problem of finding an optimum analysis of a field of model variables $\mathbf{x}_a$, given

- A background field $\mathbf{x}_b$ available at grid points in two or three dimensions
- A set of p observations $\mathbf{y}_o$ available at irregularly spaced points $\mathbf{r}_i$

For example, the unknown analysis and the known background might be two-dimensional fields of a single variable like the temperature.

Alternatively, they might be the three-dimensional field of the initial conditions for all the model prognostic variables:

$$\mathbf{x} = (p_s, T, q, u, v)$$

$$[\textbf{Here } \dim(\mathbf{x}) = N_x N_y + 4 * N_x N_y N_z]$$

The model variables are ordered by grid point and by variable, forming a single vector of length n, where n is the product of the number of points by the number of variables.

The truth, $\mathbf{x}_t$, discretized at the model points, is also a vector of length n.

We use a different variable $\mathbf{y}_o$ for the observations than for the field we want to analyze.

This is to emphasize that the observed variables are, in general, different from the model variables by being:

- (a) located in different points
- (b) (possibly) *indirect* measures of the model variables.

Examples of these measurements are radar reflectivities and Doppler shifts, satellite radiances, and global positioning system (GPS) atmospheric refractivities.

Just as for a scalar variable, the analysis is cast as the background *plus* weighted innovation:

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

The error in the analysis is

$$\varepsilon_a = \mathbf{x}_a - \mathbf{x}_t$$

So, the truth may be written

$$\mathbf{x}_t = \mathbf{x}_a - \varepsilon_a = \mathbf{x}_b + \mathbf{W}\mathbf{d} - \varepsilon_a$$

Now the truth, the analysis, and the background are vectors of length n (the total number of grid points times the number of model variables)

The weights are given by a matrix of dimension $(n \times p)$.

They are determined from statistical interpolation.

Helpful Hints

- For errors, always subtract the "truth" from the approximate or estimated quantity.
- For every matrix expression, check the orders of the components to ensure that the expression is meaningful.
- Be aware whether vectors are row or column vectors.

Forward Operator: General Remarks

In general, we do not directly observe the grid-point variables that we want to analyze.

For example, radiosonde observations are at locations that are different from the analysis grid points.

Thus, we have to perform horizontal and vertical interpolations.

We also have remote sensing instruments (like satellites and radars) that measure quantities like radiances, reflectivities, refractivities, and Doppler shifts, rather than the variables themselves.

We use an observation operator $H(\mathbf{x}_b)$ (or forward operator) to obtain, from the first guess grid field a first guess of the observations.

The observation operator H includes

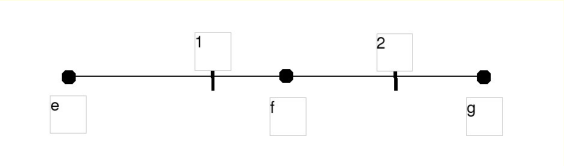
- Spatial interpolations from the first guess to the location of the observations
- Transformations that go from model variables to observed quantities (e.g., radiances)

The direct assimilation of radiances, using the forward observational model H to convert the first guess into first guess TOVS radiances has resulted in major improvements in forecast skill.

Simple Low-order Example

As a illustration, let us consider the simple case of three grid points e, f, g, and two observations, 1 and 2.

We assume that the observed and model variables are the same, so that ther is no conversion, just interpolation.



Simple example: three grid points and two observation points.

Then

$$\mathbf{x}^a = (x_e^a, x_f^a, x_g^a)^T = \begin{pmatrix} x_e^a \\ x_f^a \\ x_g^a \end{pmatrix} \quad \text{and} \quad \mathbf{x}^b = (x_e^b, x_f^b, x_g^b)^T = \begin{pmatrix} x_e^b \\ x_f^b \\ x_g^b \end{pmatrix}$$

The Observational Operator H

The forward observational operator H converts the background field into first guesses of the observations.

Normally, H is be nonlinear (e.g., the radiative transfer equations that go from temperature and moisture vertical profiles to the satellite observed radiances).

The observation field $\mathbf{y}_o$ is a vector of length p, the number of observations.

The vector d, also of length p, is the innovation or ***observational increments*** vector:

$$\mathbf{d} = \mathbf{y}_o - H(\mathbf{x}_b)$$

Note: The operator H is a nonlinear vector function. It maps from the n-dimensional analysis space to the p-dimensional observation space.

Observation Error Variances

The observation error variances come from two different sources:

- The instrumental error variances
- Subgrid-scale variability not in the grid-average values.

The second type of error is called error of representativity.

For example, an observatory might be located in a river valley. Then local effects will be encountered.

The observational error variance R is the sum of the instrument error variance $\mathbf{R}_{\text{instr}}$ and the representativity error variance $\mathbf{R}_{\text{repr}}$, assuming that these errors are not correlated:

$$\mathbf{R} = \mathbf{R}_{\text{instr}} + \mathbf{R}_{\text{repr}}$$

Error Covariance Matrix

The error covariance matrix is obtained by multiplying the vector error

$$\varepsilon = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

by its transpose

$$\varepsilon^T = \begin{bmatrix} e_1 \; e_2 \ldots \; e_n \end{bmatrix}$$

We average over many cases, to obtain the expected value:

$$\mathbf{P} = \overline{\varepsilon\varepsilon^T} = \begin{bmatrix} \overline{e_1e_1} & \overline{e_1e_2} & \cdots & \overline{e_1e_n} \\ \overline{e_2e_1} & \overline{e_2e_2} & \cdots & \overline{e_2e_n} \\ \vdots & \vdots & & \vdots \\ \overline{e_ne_1} & \overline{e_ne_2} & \cdots & \overline{e_ne_n} \end{bmatrix}$$

The overbar represents the expected value ($E(\;)$).

There are error covariance matrices for the background field and for the observations.

Covariance matrices are symmetric and positive definite.

The diagonal elements are the variances of the vector error components $\overline{e_ie_i} = \sigma_i^2$.

If we <u>normalize</u> the covariance matrix, dividing each component by the product of the standard deviations $\overline{e_ie_j}/\sigma_i\sigma_j = \text{corr}(e_i, e_j) = \rho_{ij}$, we obtain a correlation matrix

$$\mathbf{C} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{12} & 1 & \cdots & \rho_{2n} \\ \vdots & \vdots & & \vdots \\ \rho_{1n} & \rho_{12} & \cdots & 1 \end{bmatrix}$$

⋆ ⋆ ⋆

Warning: Do not confuse $\varepsilon\varepsilon^T$ and $\varepsilon^T\varepsilon$. Write expressions for both. Experiment with the 2×2 case.

If

$$\mathbf{D} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}$$

is the diagonal matrix of the variances, then we can write

$$\mathbf{P} = \mathbf{D}^{1/2}\mathbf{C}\mathbf{D}^{1/2}$$

⋆ ⋆ ⋆

Exercise: Verify the last expression explicitly for a low-order (say, $n = 3$) matrix.

Some General Rules

The transpose of a matrix product is the product of the transposes, but in reverse order:

$$[\mathbf{AB}]^T = \mathbf{B}^T\mathbf{A}^T$$

A similar rule applies to the inverse of a product:

$$[\mathbf{AB}]^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$$

⋆ ⋆ ⋆

Exercise: Prove these statements.

Note: The transpose $\mathbf{A}^T$ exists for any matrix. However, the (two-sided) inverse only exists for non-singular square matrices.

The general form of a quadratic function is

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

where A is a symmetric matrix, d is a vector and c a scalar.

To find the gradient of this scalar function $\nabla_{\mathbf{x}}F = \partial F/\partial\mathbf{x}$ (a column vector), we use the following properties of the gradient with respect to x:

$$\nabla(\mathbf{d}^T\mathbf{x}) = \nabla(\mathbf{x}^T\mathbf{d}) = \mathbf{d} \quad \text{i.e.} \quad \frac{\partial}{\partial x_i}(d_1x_1 + \ldots d_nx_n) = d_i$$

Also,

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

Therefore,

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

Conclusion of the foregoing.

BLUE

We consider multiple regression or Best Linear Unbiased Estimation (BLUE).

We start with two time series of vectors

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix} \qquad \mathbf{y}(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_p(t) \end{bmatrix}$$

We assume (*nlog*) that they are centered about their mean value, $E(\mathbf{x}) = 0$, $E(\mathbf{y}) = 0$, i.e., vectors of anomalies.

We derive the best linear unbiased estimation of x in terms of y, i.e., the optimal value of the weight matrix W in the multiple linear regression

$$\mathbf{x}_a(t) = \mathbf{W}\mathbf{y}(t)$$

This approximates the true relationship

$$\mathbf{x}(t) = \mathbf{W}\mathbf{y}(t) - \varepsilon(t)$$

where $\varepsilon(t) = \mathbf{x}_a(t) - \mathbf{x}(t)$ is the linear regression ("analysis") error, and W is an $n \times p$ matrix that minimizes the mean squared error $E(\varepsilon^T\varepsilon)$.

To derive W we write the regression equation matrix components explicitly:

$$x_i(t) = \sum_{k=1}^{p} w_{ik} y_k(t) - \varepsilon_i(t)$$

Then

$$\sum_{i=1}^{n} \varepsilon_i^2(t) = \sum_{i=1}^{n} \left[\sum_{k=1}^{p} w_{ik} y_k(t) - x_i(t) \right]^2 \qquad = \varepsilon^T\varepsilon$$

The derivative of this with respect to the weight matrix components is

$$\begin{aligned} \frac{\partial}{\partial w_{ij}} \sum_{i=1}^{n} \varepsilon_i^2(t) &= 2 \left[\sum_{k=1}^{p} w_{ik} y_k(t) - x_i(t) \right] [y_j(t)] \\ &= 2 \left[\sum_{k=1}^{p} w_{ik} y_k(t) y_j(t) - x_i(t) y_j(t) \right] \end{aligned}$$

Setting this to zero, and taking the long-time average, we get a system of equations for w_{ik}:

$$\sum_{k=1}^{p} w_{ik} \overline{y_k(t) y_j(t)} = \overline{x_i(t) y_j(t)}$$

This is a linear system of equations for the weights w_{ik}.

We will re-cast the system in matrix form:

$$\mathbf{W}\overline{\mathbf{y}\mathbf{y}^T} = \overline{\mathbf{x}\mathbf{y}^T}$$

In matrix form, the derivative of the error variance is

$$\frac{\partial}{\partial \mathbf{W}}(\varepsilon^T\varepsilon) = \frac{\partial}{\partial \mathbf{W}} \left[(\mathbf{y}^T\mathbf{W}^T - \mathbf{x}^T)(\mathbf{W}\mathbf{y} - \mathbf{x}) \right]$$

Expanding, and taking the derivative, we get

$$\frac{\partial}{\partial \mathbf{W}} \varepsilon^T\varepsilon = 2 \left\{ \left[\mathbf{W}\mathbf{y}(t)\mathbf{y}^T(t) \right] - \left[\mathbf{x}(t)\mathbf{y}^T(t) \right] \right\}$$

If we take a long time mean, and choose W to minimize the mean squared error, we get the normal equation

$$\mathbf{W} E\left(\mathbf{y}\mathbf{y}^T\right) - E\left(\mathbf{x}\mathbf{y}^T\right) = 0$$

or

$$\mathbf{W} = E\left(\mathbf{x}\mathbf{y}^T\right) \left[E\left(\mathbf{y}\mathbf{y}^T\right) \right]^{-1}$$

This gives the best linear unbiased estimation

$$\mathbf{x}_a(t) = \mathbf{W}\mathbf{y}(t) \, .$$

Formal Derivation of BLUE

The analysis error covariance can be written

$$\varepsilon^T\varepsilon = (\mathbf{y}^T\mathbf{W}^T - \mathbf{x}^T)(\mathbf{W}\mathbf{y} - \mathbf{x})$$

We proceed heuristically, formally differentiating and gathering terms taking account of the matrix orders.

The derivative with respect to the weights is

$$\frac{\partial\varepsilon^T\varepsilon}{\partial\mathbf{W}} = -2\left(\mathbf{W}\mathbf{y} - \mathbf{x}\right)\mathbf{y}^T$$

Setting this to zero and taking time means give the normal equations:

$$\mathbf{W} = E\left(\mathbf{x}\mathbf{y}^T\right)\left[E\left(\mathbf{y}\mathbf{y}^T\right)\right]^{-1}$$

Statistical Assumptions

We define the background error and the analysis error as vectors of length n:

$$\varepsilon_b = \mathbf{x}_b - \mathbf{x}_t$$
$$\varepsilon_a = \mathbf{x}_a - \mathbf{x}_t$$

The p observations available at irregularly spaced points $\mathbf{y}_o(\mathbf{r}_k)$ have observational errors

$$\varepsilon_{ok} = \mathbf{y}_o(\mathbf{r}_k) - \mathbf{y}_t(\mathbf{r}_k) = \mathbf{y}_o - H(\mathbf{x}_t)$$

We don't know the truth, $\mathbf{x}_t$, thus we don't know the errors of the available background and observations ...

... but we can make a number of assumptions about their statistical properties.

The background and observations are assumed to be unbiased:

$$\left.\begin{array}{l} E\{\varepsilon_b\} = E\{\mathbf{x}_b\} - E\{\mathbf{x}_t\} = 0 \\ E\{\varepsilon_o\} = E\{\mathbf{y}_o\} - E\{\mathbf{y}_t\} = 0 \end{array}\right\}$$

We define the error covariance matrices for the analysis, background and observations respectively:

$$\left.\begin{array}{llll} \mathbf{P}_a & = \mathbf{A} = E\{\varepsilon_a\varepsilon_a^T\} & & (n\times n) \\ \mathbf{P}_b & = \mathbf{B} = E\{\varepsilon_b\varepsilon_b^T\} & & (n\times n) \\ \mathbf{P}_o & = \mathbf{R} = E\{\varepsilon_o\varepsilon_o^T\} & & (p\times p) \end{array}\right\}$$

The nonlinear observation operator, H, that transforms analysis variables into observed variables can be linearized as

$$H(\mathbf{x} + \delta\mathbf{x}) = H(\mathbf{x}) + \mathbf{H}\delta\mathbf{x}$$

Here $\mathbf{H}$ is a $p\times n$ matrix, called the linear observation operator with elements

$$\mathbf{H}_{ij} = \frac{\partial H_i}{\partial x_j} \qquad (p\times n)$$

Note that H is a nonlinear <u>vector function</u> while $\mathbf{H}$ is a <u>matrix</u>.

We assume that the background field is a good approximation of the truth.

Then the analysis and the observations are equal to the background values plus small increments $\varepsilon_b = \mathbf{x}_b - \mathbf{x}_t$.

So, the innovation vector $\mathbf{d} = \mathbf{y}_0 - H(\mathbf{x}_b)$ can be written

$$\begin{aligned} \mathbf{d} &= \mathbf{y}_o - H(\mathbf{x}_b) = \mathbf{y}_o - H(\mathbf{x}_t + (\mathbf{x}_b - \mathbf{x}_t)) \\ &= \mathbf{y}_o - H(\mathbf{x}_t) - \mathbf{H}(\mathbf{x}_b - \mathbf{x}_t) = \varepsilon_o - \mathbf{H}\varepsilon_b \end{aligned}$$

Here we use

$$H(\mathbf{x} + \varepsilon) = H(\mathbf{x}) + \left(\frac{\partial H}{\partial\mathbf{x}}\right)_{\mathbf{x}}\varepsilon = H(\mathbf{x}) + \mathbf{H}\varepsilon$$

The $\mathbf{H}$ matrix transforms vectors in analysis space into their corresponding values in observation space.

Its transpose or adjoint $\mathbf{H}^T$ transforms vectors in observation space to vectors in analysis space.

The background error covariance, B, and the observation error covariance, R, are assumed to be known.

We assume that the observation and background errors are uncorrelated:

$$E\{\varepsilon_o \varepsilon_b^T\} = 0$$

We will now use the best linear unbiased estimation formula

$$\mathbf{W} = E\left(\mathbf{x}\mathbf{y}^T\right)\left[E\left(\mathbf{y}\mathbf{y}^T\right)\right]^{-1}$$

to derive the optimal weight matrix W.

The innovation is

$$\mathbf{d} = \mathbf{y}_o - H(\mathbf{x}_b) = \varepsilon_o - \mathbf{H}\varepsilon_b$$

So, the optimal weight matrix W that minimizes $\varepsilon_a^T \varepsilon_a$ is

$$\mathbf{W} = E\{(\mathbf{x} - \mathbf{x}_b)[\mathbf{y}_o - H(\mathbf{x}_b)]^T\}\left[E\{[\mathbf{y}_o - H(\mathbf{x}_b)][\mathbf{y}_o - H(\mathbf{x}_b)]^T\}\right]^{-1}$$

This can be written as

$$\mathbf{W} = E[(-\varepsilon_b)(\varepsilon_o - \mathbf{H}\varepsilon_b)^T]\ \{E[(\varepsilon_o - \mathbf{H}\varepsilon_b)(\varepsilon_o - \mathbf{H}\varepsilon_b)^T]\}^{-1}$$

We may expand it as

$$\mathbf{W} = \left[E(\varepsilon_b \varepsilon_b^T)\mathbf{H}\right]\left[E(\varepsilon_o \varepsilon_o^T) + \mathbf{H}E(\varepsilon_b \varepsilon_b^T)\mathbf{H}^T\right]^{-1}$$

Substituting the definitions of background error covariance B and observational error covariance R into this, we obtain the optimal weight matrix:

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

Repeat:

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

Using the relationship

$$\varepsilon_a = \varepsilon_b + \mathbf{W}[\varepsilon_0 - \mathbf{H}(\varepsilon_b)]$$

we can derive the analysis error covariance $E\{\varepsilon_a \varepsilon_a^T\}$.

It is

$$\begin{aligned}\mathbf{P}_a &= E\{\varepsilon_a \varepsilon_a^T\} = E\{\varepsilon_b \varepsilon_b^T + \varepsilon_b(\varepsilon_o - \mathbf{H}\varepsilon_b)^T\mathbf{W}^T \\ &\quad + \mathbf{W}(\varepsilon_o - \mathbf{H}\varepsilon_b)\varepsilon_b^T + \mathbf{W}(\varepsilon_o - \mathbf{H}\varepsilon_b)(\varepsilon_o - \mathbf{H}\varepsilon_b)^T\mathbf{W}^T\} \\ &= \mathbf{B} - \mathbf{B}\mathbf{H}^T\mathbf{W}^T - \mathbf{W}\mathbf{H}\mathbf{B} + \mathbf{W}\mathbf{R}\mathbf{W}^T + \mathbf{W}\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{W}^T\end{aligned}$$

Substituting

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

we obtain

$$\mathbf{P}_a = (\mathbf{I} - \mathbf{W}\mathbf{H})\mathbf{B}$$

The Full Set of OI Equations

For convenience, we collect the full set of basic equations of OI, and then examine their meaning in detail.

They are formally similar to the equations for the scalar least squares *'two temperatures problem'*.

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

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

$$\mathbf{P}_a = (\mathbf{I} - \mathbf{W}\mathbf{H})\mathbf{B}$$

The interpretation of these equations is very similar to the scalar case discussed earlier.

The Analysis Equation

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

This equation says:

The analysis is obtained by adding to the background field the product of the optimal weight matrix and the innovation.

The first guess of the observations is obtained by applying the observation operator H to the background vector.

Note that from $H(\mathbf{x} + \delta\mathbf{x}) = H(\mathbf{x}) + \mathbf{H}\delta\mathbf{x}$, we get

$$H(\mathbf{x}_b) = H(\mathbf{x}_t) + \mathbf{H}(\mathbf{x}_b - \mathbf{x}_t) = H(\mathbf{x}_t) + \mathbf{H}\varepsilon_b\,,$$

where the matrix H is the linear tangent perturbation of H.

The Optimal Weight Matrix

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

This equation says:

The optimal weight matrix is given by the background error covariance in the observation space $(\mathbf{B}\mathbf{H}^T)$ multiplied by the inverse of the total error covariance.

Note that the larger the background error covariance compared to the observation error covariance, the larger the correction to the first guess.

Check the result if $\mathbf{R} = 0$.

Analysis Error Covariance Matrix

$$\mathbf{P}_a = (\mathbf{I} - \mathbf{W}\mathbf{H})\mathbf{B}$$

This equation says:

The error covariance of the analysis is given by the error covariance of the background, reduced by a matrix equal to the identity matrix minus the optimal weight matrix.

Note that I is the $n \times n$ identity matrix.

End of §5.4.1