<u>Spatial applications of least squares</u> - <u>N = 2 Case</u> (cont.)

The background field also has errors, denoted by $\varepsilon_B(\vec{\mathbf{r}}_k)$. The estimation of these errors is a major research topic in data assimilation research. In NWP, they represent the errors in the short-range forecast that provides the background field. Two possible ways to obtain these errors are:

1. Compare the forecast values with the observations at each point. The flaw here is that the obs. also have errors, which are not precisely known.
2. (NCEP method) Subtract two forecasts of different length that are valid at the same time. The difference between them represents forecast error growth.
3. Calculate them from an ensemble of many short-range forecasts made in real-time (basis of ensemble Kalman filter data assimilation)

Now, assume both $\varepsilon_o(\mathbf{r}_k)$ and $\varepsilon_B(\mathbf{r}_k)$ to be unbiased, random, and normally distributed. Also assume that ε_o and ε_B are spatially correlated (e.g., the error field has synoptic structure) but *not* with each other. We express these assumptions as:

$$< \varepsilon_o(\mathbf{r}_k)\, \varepsilon_o(\mathbf{r}_l) > \neq 0 \; ; \qquad < \varepsilon_B(\mathbf{r}_k)\, \varepsilon_B(\mathbf{r}_l) > \neq 0 \qquad (1)$$

$$< \varepsilon_o(\mathbf{r}_k)\, \varepsilon_B(\mathbf{r}_l) > = 0 \quad \text{for all k, l} \qquad (2)$$

The quantity to be minimized is given by eq. (21), Lecture #4, for N = 2:

$$I = \tfrac{1}{2} \{ [\underline{\mathbf{f}}_A - \underline{\mathbf{f}}_o]^T \underline{\underline{\mathbf{O}}}^{-1} [\underline{\mathbf{f}}_A - \underline{\mathbf{f}}_o] + [\underline{\mathbf{f}}_A - \underline{\mathbf{f}}_B]^T \underline{\underline{\mathbf{B}}}^{-1} [\underline{\mathbf{f}}_A - \underline{\mathbf{f}}_B] \} \qquad (3)$$

where $\underline{\underline{\mathbf{B}}}$ is the <u>background error covariance matrix.</u>

[Note that f_A , f_B and f_o are all defined at observation points here, so we don't have an objective analysis scheme yet!]

To provide one more example of the simplifying properties of matrix notation, here is what eq. (3) looks like in summation notation:

$$I = \frac{1}{2} \sum_{k=1}^{K} \sum_{l=1}^{K} \Big\{ \left[f_A(\vec{r}_k) - f_o(\vec{r}_k) \right] \left[f_A(\vec{r}_l) - f_o(\vec{r}_l) \right] \tilde{o}_{kl}$$

$$+ \left[f_A(\vec{r}_k) - f_B(\vec{r}_k) \right] \left[f_A(\vec{r}_l) - f_B(\vec{r}_l) \right] \tilde{b}_{kl} \Big\} \qquad (4)$$

where $\tilde{o}_{k,l}$, $\tilde{b}_{k,l}$ are the elements of $\underline{\underline{\mathbf{O}}}^{-1}$, $\underline{\underline{\mathbf{B}}}^{-1}$.

Now perform the least squares minimization by differentiating (3) or (4) w.r.t. the unknown analysis values $f_A(\mathbf{r})$. Use of eq. (4) will yield:

$$\frac{\partial I}{\partial f_A(\vec{r}_k)} = \sum_{l=1}^{K} \Big\{ \left[f_A(\vec{r}_l) - f_o(\vec{r}_l) \right] \tilde{o}_{kl} + \left[f_A(\vec{r}_l) - f_B(\vec{r}_l) \right] \tilde{b}_{kl} \Big\} = 0, \quad \forall\, k = 1(1)K.$$

whereas use of eq. (3) gives us:

$$\frac{\partial I}{\partial \underline{f}_A} = \underline{\underline{\mathbf{O}}}^{-1} [\underline{\mathbf{f}}_A - \underline{\mathbf{f}}_o] + \underline{\underline{\mathbf{B}}}^{-1} [\underline{\mathbf{f}}_A - \underline{\mathbf{f}}_B] = 0$$

or

$$\underline{\underline{\mathbf{O}}}^{-1} \underline{\mathbf{f}}_A - \underline{\underline{\mathbf{O}}}^{-1} \underline{\mathbf{f}}_o + \underline{\underline{\mathbf{B}}}^{-1} \underline{\mathbf{f}}_A - \underline{\underline{\mathbf{B}}}^{-1} \underline{\mathbf{f}}_B = 0$$

or

$$(\underline{\underline{\mathbf{B}}}^{-1} + \underline{\underline{\mathbf{O}}}^{-1}) \underline{\mathbf{f}}_A = \underline{\underline{\mathbf{B}}}^{-1} \underline{\mathbf{f}}_B + \underline{\underline{\mathbf{O}}}^{-1} \underline{\mathbf{f}}_o$$

Solving for f_A, we get

$$\underline{\mathbf{f}}_A = [\underline{\underline{\mathbf{B}}}^{-1} + \underline{\underline{\mathbf{O}}}^{-1}]^{-1} [\underline{\underline{\mathbf{B}}}^{-1} \underline{\mathbf{f}}_B + \underline{\underline{\mathbf{O}}}^{-1} \underline{\mathbf{f}}_o] \qquad (5)$$

Class exercise: Show that eq. (7) can be rewritten as

$$\underline{\mathbf{f}}_A - \underline{\mathbf{f}}_B = \underline{\underline{\mathbf{B}}} [\underline{\underline{\mathbf{B}}} + \underline{\underline{\mathbf{O}}}]^{-1} [\underline{\mathbf{f}}_O - \underline{\mathbf{f}}_B] \qquad (6)$$

[Note how (6) yields a (K,1) column vector of most likelihood estimates of the analysis values for the K obs. locations.]

This exercise is very challenging (for me!) Here are two hints - make use of the following two lemmas:

Lemma #1:

$$\mathbf{X}\,[\,\mathbf{X} + \mathbf{Y}\,]^{-1} + \mathbf{Y}\,[\,\mathbf{X} + \mathbf{Y}\,]^{-1} \;=\; [\,\mathbf{X} + \mathbf{Y}\,]\,[\,\mathbf{X} + \mathbf{Y}\,]^{-1} \;=\; \mathbf{I}$$

Similarly,

$$\mathbf{X}^{-1}[\,\mathbf{X}^{-1} + \mathbf{Y}^{-1}\,]^{-1} + \mathbf{Y}^{-1}[\,\mathbf{X}^{-1} + \mathbf{Y}^{-1}\,]^{-1} \;=\; [\,\mathbf{X}^{-1} + \mathbf{Y}^{-1}]\,[[\,\mathbf{X}^{-1} + \mathbf{Y}^{-1}]^{-1} \;=\; \mathbf{I}$$

Lemma #2: $\quad (\,\mathbf{X}\,\mathbf{Y}\,)^{-1} \;=\; \mathbf{Y}^{-1}\,\mathbf{X}^{-1}$

In eq. (6), $\underline{\mathbf{f}}_{\mathbf{o}} - \underline{\mathbf{f}}_{\mathbf{B}}$ are the <u>observation increments</u> (these might be viewed as forecast errors if the obs. were perfect, but, since they are not, we view this increment as simply a correction to the background field),

and $\underline{\mathbf{f}}_{\mathbf{A}} - \underline{\mathbf{f}}_{\mathbf{B}}$ are known as the <u>analysis increments</u>.

Again note that (6) provides analysis values at the observation locations; - it does not provide a gridded analysis.

As usual, we now determine the <u>expected analysis error variance</u> for this analysis system.

Define $\underline{\varepsilon}_A \;=\; f_A(\mathbf{r}) \;-\; f_T(\mathbf{r})$, which is a column vector of analysis errors, where f_T are the (unknown) true values. Following the text, it is a **class exercise** to obtain

$$<\underline{\varepsilon}_A \underline{\varepsilon}^T{}_A> \;=\; [\,\underline{\underline{B}}^{-1} + \underline{\underline{Q}}^{-1}\,]^{-1} \qquad (7)$$

where the LHS is an analysis error covariance matrix with elements $<\varepsilon_A(\mathbf{r}_k)\,\varepsilon_A(\mathbf{r}_l)>$

We define the analysis error $\underline{\underline{A}}$ as the RHS of (7), $\quad \underline{\underline{A}} = [\underline{\underline{B}}^{-1} + \underline{\underline{Q}}^{-1}]^{-1}$

Please show that $\underline{\underline{A}}$ can also be written as

$$\underline{\underline{A}} \;=\; \underline{\underline{B}}\,[\,\underline{\underline{B}} + \underline{\underline{Q}}\,]^{-1} \;=\; \underline{\underline{B}} - \underline{\underline{B}}\,[\,\underline{\underline{B}} + \underline{\underline{Q}}\,]^{-1}\,\underline{\underline{B}} \qquad (8)$$

Note that equations (3), (6) and (8) are the vector equivalents to the scalar equations (15), (16) and (17) in lecture 4.

Note that it is mathematically possible to minimize other than quadratic forms; i.e. – could minimize

$$I = 1/q \sum_{n=1}^{N} \frac{[s_a - s_n]^q}{\sigma_n^q} \qquad (9)$$

The choice of q depends on what you think the probability density function for the errors is.

Here [e.g., eq. (3)], we have chosen $q = 2$, or minimization in the ℓ_2 norm sense, because we believe that the normal or Gaussian distribution function is appropriate for most meteorological data errors (provided "gross errors" have been removed by a quality control procedure beforehand). Also, $q = 2$ leads to linear analysis equations.

A choice of $q = 1$ corresponds to a "long-tailed" error distribution – meaning large errors have a higher probability than with a Gaussian one.

See Daley for more discussion of the differences between ℓ_1 and ℓ_2 minimization.

Global function fitting

We shall briefly discuss global function fitting, and then examine some of the pitfalls of function fitting in general. Recall that Lecture 3 gave an example of local (polynomial) fitting, where the fit of a cluster of data was accomplished solely to find the value at a grid point in the center of the cluster.

Global function fitting is the process of fitting specified functions to *all* observations in an analysis domain.

Again, let $f(\mathbf{r})$ be the dependent variable, where $\vec{r}$ is a 1-D, 2-D or 3-D spatial coordinate.

Assume the analyzed field $f_A(\mathbf{r})$ can be represented by a finite series of ordered basis functions $h_1(\mathbf{r}), h_2(\mathbf{r}), \ldots\ldots h_m(\mathbf{r}) \ldots\ldots h_M(\mathbf{r})$, $m = 0(1)M$. Then our function model can be written as

$$f_A(\mathbf{r}) = \sum_{m=0}^{M} c_m \, h_m(\mathbf{r}) \qquad (1)$$

Note that m could be summing over 1, 2 or 3D space.

Assume K observations $f_o(\mathbf{r}_k)$, $k = 1(1)K$ over the entire domain.

To obtain a global fit in the least squares sense, we need to minimize

$$I = \sum_{k=1}^{K} w_k\, d_k^2 = \tfrac{1}{2} \sum_{k=1}^{K} w_k \left[\sum_{m=0}^{M} c_m\, h_m(\mathbf{r}_k) - f_O(\mathbf{r}_k) \right]^2 \qquad (2)$$

where $w_k = < \varepsilon_o^2(\mathbf{r}_k) >^{-1}$ are the observational error variances.

To minimize (2), differentiate w.r.t. each coefficient c_m and set to zero:

$$\sum_{k=1}^{K} w_k\, h_m(\mathbf{r}_k) \left[\sum_{\mu=0}^{M} c_\mu\, h_\mu(\mathbf{r}_k) - f_O(\mathbf{r}_k) \right] = 0$$

or

$$\sum_{\mu=0}^{M} c_\mu \sum_{k=1}^{K} w_k\, h_m(\mathbf{r}_k)\, h_\mu(\mathbf{r}_k) = \sum_{k=1}^{K} w_k\, h_m(\mathbf{r}_k)\, f_O(\mathbf{r}_k) \qquad (3)$$

Eq. (3) are the <u>normal equations</u> for this problem, from which we need to solve for the c_μ

If we repeat this process in matrix form, eq. (1) becomes

$$\underline{\mathbf{f}}_A = \underline{\underline{\mathbf{H}}}\, \underline{\mathbf{c}} \qquad (4)$$

where $\underline{\mathbf{f}}$ is a column vector of analysis values $f_A(\mathbf{r}_k)$ of length K and $\underline{\underline{\mathbf{H}}}$ is a K x (M+1) rectangular matrix with elements $h_{km} \sim h_m(\mathbf{r}_k)$.

The derivation of the normal equations starts with our usual quadratic form for I:

$$I = \tfrac{1}{2} \{ [\underline{\mathbf{f}}_A - \underline{\mathbf{f}}_o]^T \underline{\underline{\mathbf{O}}}^{-1} [\underline{\mathbf{f}}_A - \underline{\mathbf{f}}_o] \}$$

Where $\underline{\underline{\mathbf{O}}}$ is a KxK diagonal matrix whose elements are $< \varepsilon_o^2(\mathbf{r}_k) >$.

Using (4), we have $I = \tfrac{1}{2} \{ [\underline{\underline{\mathbf{H}}}\, \underline{\mathbf{c}} - \underline{\mathbf{f}}_o]^T \mathbf{O}^{-1} [\underline{\underline{\mathbf{H}}}\, \underline{\mathbf{c}} - \underline{\mathbf{f}}_o] \}$

Which is equivalent to eq. (2). Taking the derivative w.r.t. $\underline{\mathbf{c}}$ and setting to zero yields

$$\underline{\underline{\mathbf{H}}}^T \underline{\underline{\mathbf{O}}}^{-1} [\underline{\underline{\mathbf{H}}}\, \underline{\mathbf{c}} - \underline{\mathbf{f}}_o] = 0$$

or

$$\underline{\underline{\mathbf{H}}}^T \underline{\underline{\mathbf{O}}}^{-1} \underline{\underline{\mathbf{H}}}\, \underline{\mathbf{c}} = \underline{\underline{\mathbf{H}}}^T \underline{\underline{\mathbf{O}}}^{-1} \underline{\mathbf{f}}_o$$

We examine the solution to this equation in Lecture #6.