<u>Biased observations</u>

In the least squares derivation in the previous lecture, we assumed unbiased observations. Here we will show how to account for bias - that is, when $<\varepsilon_n> \neq 0$, the obs. are said to be biased, and thus there will be a bias in the analysis s_a.

Fortunately, if the bias is known, it is easy to produce unbiased estimates.

We start with eq. (5) from the previous derivation and replace s_n with $s_n - <\varepsilon_n>$

[Note that this redefines the error variance σ_n^2 to be $\sigma_n^2 = <\varepsilon_n^2> - <\varepsilon_n>^2$

So, eq. (5) is

$$I_b = \tfrac{1}{2} \Sigma \, \sigma_n^{-2} [\, s_a - (s_n - <\varepsilon_n>)\,]^2 \qquad (11)$$

Minimization of I_b w.r.t. s_a yields

$$s_a = \frac{\sum_{n=1}^{N} \sigma_n^{-2} (s_n - \langle \varepsilon_n \rangle)}{\sum_{n=1}^{N} \sigma_n^{-2}} \qquad (12)$$

To check this result, we should examine the <u>bias of the analysis estimate</u> to see if it is zero.

Since the bias of the analysis estimate ε_a is defined by $<\varepsilon_a> = <s_a - s>$, then we can use eq. (12) to write

$$<\varepsilon_a> = \left\langle \frac{\sum_{n=1}^{N} \sigma_n^{-2} (s_n - \langle \varepsilon_n \rangle)}{\sum_{n=1}^{N} \sigma_n^{-2}} \right\rangle - \langle s \rangle \qquad (13)$$

It is a **class exercise** to show that RHS of (13) is equal to 0.

<u>Therefore, if we know the bias, we can always create an unbiased analysis estimate</u>.

Because, of this, we shall usually assume unbiased errors in the future to keep the analysis simpler.

Finally, we note that we can write our expression for I (eq. 5) in a simpler form, that is,

$$I = \sum_{n=1}^{N} w_n d_n^2 \qquad (14)$$

where $\quad d_n = s_a - s_n \quad$ is the <u>residual</u> of the nth observation

and $\quad w_n = \dfrac{\sigma_n^{-2}}{2} \quad$ is the *a priori* (specified) weight.

These w_n are often called Gauss precision moduli and they are inversely proportional to the observational error variance - thus small errors yield larger weights.

Class exercise

Consider the N = 2 case for the equations developed for I, the quantity to be minimized (eq. 5); s_a, the analysis value (eq. 6); and $< \varepsilon_a^2 >$, the expected error variance of the analysis (eq. 9). Denote the two observations of s as s_o and s_b, with corresponding error variances of σ_o^2 and σ_b^2 respectively. Show that these three equations become:

$$I = \frac{(s_a - s_o)^2}{2\sigma_o^2} + \frac{(s_a - s_b)^2}{2\sigma_b^2} \tag{15}$$

$$s_a = \frac{\sigma_o^{-2} s_o + \sigma_b^{2} s_b}{\sigma_o^{-2} + \sigma_b^{-2}} = s_b + \frac{\sigma_b^{2}}{\sigma_o^2 + \sigma_b^2}(s_o - s_b) \tag{16}$$

$$\langle \varepsilon_a^2 \rangle = \sigma_b^2 - \frac{\sigma_b^4}{\sigma_o^2 + \sigma_b^2} = \frac{\sigma_b^2 \sigma_o^2}{\sigma_o^2 + \sigma_b^2} = \left[\sigma_o^{-2} + \sigma_b^{-2}\right]^{-1} \tag{17}$$

These equations represent the scalar or zero-dimensional application of least squares estimation. Although very simple, they foreshadow very accurately the results we will see from much more sophisticated analysis schemes. On the next page, we will start considering vector versions of least squares – although only in the spatial dimensions.

Spatial applications of least squares

Now we define a dependent state variable (e.g., temperature) $f(\vec{r})$, where $\vec{r} = (x, y, z)$ is a 3-D spatial coordinate.

We first do the N = 1 case - one observation at each point.

Denote $f_o(\vec{r}_k)$ as an observation of f at station $\vec{r}_k$ with expected obs. error variance of

$< \varepsilon_o^2 (\vec{\mathbf{r}}_k) >$. Also assume K observations whose errors are normally distributed, unbiased, and spatially uncorrelated. This last assumption means that the covariance

$$< \varepsilon_o (\vec{\mathbf{r}}_k)\, \varepsilon_o (\vec{\mathbf{r}}_l) > \;=\; 0, \quad \text{for } k \neq l \qquad [\; k = l \text{ gives us } < \varepsilon_o^2 (\vec{\mathbf{r}}_k) >]$$

Finally, define $f_A (\vec{\mathbf{r}})$ as the analyzed field of $f(\vec{\mathbf{r}})$.

The quantity to be minimized in this case, written in the notation of eq. (14), is

$$I = \sum_{k=1}^{K} w_k\, d_k^2 = \tfrac{1}{2} \sum_{k=1}^{K} < \varepsilon^2_o (\vec{\mathbf{r}}_k) >^{-1} [f_o (\vec{\mathbf{r}}_k) - f_A (\vec{\mathbf{r}}_k)]^2 \tag{18}$$

Minimization of (18) w.r.t. each unknown analysis value $f_A (\mathbf{r}_k)$ yields

$$f_A (\vec{\mathbf{r}}_k) = f_o (\vec{\mathbf{r}}_k) \tag{19}$$

That is, since there is only a single observation per station, the most probable value is that observation. Note that (19) is the true solution only if have error-free observations.

Also note that (18) is usually minimized w.r.t. some constraint; e.g. – a polynomial fit.

We now introduce matrix notation. Equation (18) is now

$$I = \tfrac{1}{2} [\underline{\mathbf{f}}_A - \underline{\mathbf{f}}_o]^T \,\underline{\underline{\mathbf{w}}}\, [\underline{\mathbf{f}}_A - \underline{\mathbf{f}}_o] \tag{20}$$

where $\underline{\mathbf{f}}_A$, $\underline{\mathbf{f}}_o$ are column vectors of length K:

and $\underline{\underline{\mathbf{w}}}$ is a K x K (square) matrix with diagonal elements only:

$$w_k \sim \; < \varepsilon^2_o (\vec{\mathbf{r}}_k) >$$

Note that eq. (20) is written in this form to ensure that the matrices are conformal for multiplication. Thus, in (20), the $[\quad]^T$ term creates a (1, K) row vector that is conformable with $\underline{\underline{\mathbf{w}}}$ whereas $[\quad]$ is not.

To illustrate, $\underset{(1,K)}{[\mathbf{f}_A - \mathbf{f}_o]^T}\; \underset{(K\times K)}{\mathbf{w}}$ is conformal and yields a [1, K] row vector that is conformal with the [K, 1] column vector $[\, \mathbf{f}_A - \mathbf{f}_o \,]$. The result of this second multiplication (1, K)(K, 1) yields a (1, 1) matrix - i.e. – a <u>scalar</u>.

Eq. (1) is often written is the form

$$I = \tfrac{1}{2}\,[\underline{\mathbf{f}}_A - \underline{\mathbf{f}}_o]^T\, \underline{\underline{\mathbf{O}}}^{-1}\, [\underline{\mathbf{f}}_A - \underline{\mathbf{f}}_o] \tag{21}$$

where $\underline{\underline{\mathbf{O}}}$ is a K x K <u>observational error variance matrix</u> with elements $< \varepsilon_o^2(\vec{\mathbf{r}}_k) >$

Here, $\underline{\underline{\mathbf{O}}}$ is a diagonal matrix, but if the errors are <u>spatially correlated</u>, that is,

$< \varepsilon_o(\vec{\mathbf{r}}_k)\, \varepsilon_o(\vec{\mathbf{r}}_l) > \neq 0$, then $\underline{\underline{\mathbf{O}}}$ is a "full" <u>observational error covariance matrix</u>.

Note that $\underline{\underline{\mathbf{O}}}$ needs to be nonsingular so that its inverse exists.

<u>**Spatial applications of least squares**</u> - <u>**N = 2 Case**</u>

Now, as in the zero-dimensional application, we have 2 pieces of information: an <u>observation</u> $f_o(\vec{\mathbf{r}})$ and a <u>**background estimate**</u> of f, $f_B(\vec{\mathbf{r}})$. Note that in the older literature, $f_B(\vec{\mathbf{r}})$ is often called the "first guess field". Please read p. 25-27 in Daley's book, especially Fig. 1.12, to see the role of the background field.

The background field is vital in atmospheric analysis problems since the observations are not uniformly distributed. Sources for $f_B(\vec{\mathbf{r}})$ could be climatology, a previous analysis or a previous forecast valid at the time of the new analysis - the latter is the most common in NWP applications.