Lecture 24: Interpolation and Smoothing

©Christopher S. Bretherton

Winter 2015

Ref: Daley, R., 1991: *Atmospheric Data Analysis.* Cambridge University Press. 457 pp.

24.1 Motivation

We often need to interpolate and smooth data, which may be taken at irregular locations and times. We also may need to synthesize data of different types (e. g. measured surface winds and sea-level pressure). Lastly, we may need to assimilate data into a diagnostic or predictive model, either to help make forecasts or to enforce physical consistency constraints between multiple data types. These are the origins of the term 'objective analysis'.

This is a huge, complex and sophisticated area of research and knowledge and in the remaining lectures we will only scratch at a few parts of its surface, namely smoothing, optimal interpolation and Kalman filtering for data assimilation.

24.2 Smoothing

We often have data with measurement errors or small-scale fluctuations from which we wish to generate a smooth field, often on a regular grid of points.

Software packages such as Matlab have various functions for interpolating irregularly-spaced data to a regular grid without smoothing, depending on whether the data is given on some kind of mesh of points (e. g. **interp1, interp2**) or at an unstructured set of points (**griddata**).

Low-pass filters can be useful for smoothing regularly-spaced time series. For smoothing irregularly spaced data, **kernel smoothing** can be a good option (http://en.wikipedia.org/wiki/Kernel_smoother). Given noisy measurements $x(t_i)$ of some process at irregularly spaced times t_i, the smoother is a weighted average

$$\tilde{x}(t) = \frac{\sum_{i=1}^{N} K(t-t_i)x(t_i)}{\sum_{i=1}^{N} K(t-t_i)} \tag{24.2.1}$$

where K is the smoothing kernel. A common choice is the Gaussian kernel

$$K(T) = \exp\left(-T^2/2\tau^2\right) \tag{24.2.2}$$

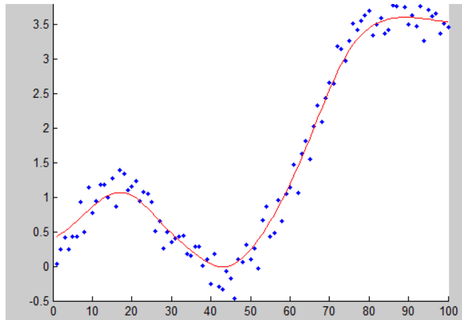

Figure 1: Gaussian kernel smoother (http://en.wikipedia.org/wiki/Kernel_smoother)

where the smoothing width τ is a user-chosen adjustable parameter chosen based on the typical spacing between sample times and the desired degree of small-scale noise filtering.

24.3 Optimal interpolation

Optimal interpolation is an approach to synthesizing multiple types of data with different known uncertainties. For instance, we might have two measurements of wind speed at the same location using different instruments with different error characteristics, or we might have a prior estimate (with some uncertainty) of what the wind speed should be (e. g. from a large-scale forecast model). How can we combine these measurements into a best guess at the wind speed with minimum uncertainty? To be consistent with the following discussion of Kalman filtering, we frame the problem as finding the optimal estimate of of some set of variables $\mathbf{y}$ by combining a prior estimate $\mathbf{y}^p$ with known uncertainties with new observations $\mathbf{y}^o$, also with known (but different) uncertainties.

24.4 Observations of a single quantity

We start with the simplest 1-variable case. We assume the prior estimate y^p is has a normal distribution with standard deviation σ_p and mean equal to the true value y. We assume the observation y^o is has a normal distribution with standard deviation σ_o and mean equal to the true value y, and that the errors in the prior and the observation are uncorrelated.

We regard the new observation as giving a correction that updates the prior estimate:

$$\hat{y} = y^p + k(y^o - y^p) = (1-k)y^p + ky^o, \qquad (0 < k < 1) \tag{24.4.1}$$

Taking the nudging factor $k \approx 0$ would weight the prior much more than the new observation; taking $k \approx 1$ would weight the new observations much more than the prior. In fact, k is chosen to minimize a *penalty function*

$$R(\hat{y}) = \left(\frac{y^o - \hat{y}}{\sigma_o}\right)^2 + \left(\frac{y^p - \hat{y}}{\sigma_p}\right)^2.$$

Given normally distributed variables, R measures the likelihood of obtaining both our prior and our new observation given an actual value $\hat{y}$, given that the observation and prior are mutually uncorrelated. A larger value of R corresponds to a smaller likelihood, so we choose $\hat{x}$ to minimize R:

$$0 = \frac{\partial R}{\partial \hat{y}} = -2\frac{y^o - \hat{y}}{\sigma_o^2} - 2\frac{y^p - \hat{y}}{\sigma_p^2}$$

$$\Rightarrow \quad \hat{y} = \hat{\sigma}^2 \left(\frac{y^o}{\sigma_o^2} + \frac{y^p}{\sigma_p^2}\right) \tag{24.4.2}$$

$$\text{where} \quad \frac{1}{\hat{\sigma}^2} = \frac{1}{\sigma_o^2} + \frac{1}{\sigma_p^2} \tag{24.4.3}$$

is the variance of the updated estimate $\hat{y}$ (Reader: prove this for yourself). Equations (24.4.1), (24.4.2) and (24.4.3) imply

$$k = \frac{\hat{\sigma}^2}{\sigma_o^2} = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_o^2}. \tag{24.4.4}$$

As we'd expect, k is near to 1 if the prior has large variance (uncertainty) compared to the observation, and k is near to zero if the opposite is true. The updated variance can also be written in terms of k, nicely showing how it is reduced by the new observation:

$$\hat{\sigma}^2 = \frac{\sigma_p^2 \sigma_o^2}{\sigma_p^2 + \sigma_o^2} = (1-k)\sigma_p^2 \tag{24.4.5}$$

24.5 Observations of multiple correlated quantities

This process can be carried through similarly (with a bit more pain) for an m-vector of data $\mathbf{y}$ with $m \times m$ error covariance matrix $\mathbf{C}^p$ for the prior and $\mathbf{C}^o$ for the observations.

$$\hat{\mathbf{y}} = \mathbf{y}_p + \underbrace{\mathbf{C}^p(\mathbf{C}_y^p + \mathbf{C}^o)^{-1}}_{K}(\mathbf{y}^o - \mathbf{y}^p) \tag{24.5.1}$$

Note that this involves inversion of a full $m \times m$ matrix, which may be impractical for large m. It also requires specification of the error covariance matrices, which may not be known or easy to find. Nevertheless, this is the basis for the 3DVAR algorithm that was long used for assimilating data into numerical weather prediction models (and still is, in some cases.)