Lecture Notes on Statistical Analysis of Environmental Data

G. Venuti - D. Triglione

Contents

1 Gridding **5**

1.1 Position of the problem 5

1.2 Grid numbering 6

1.3 Weighted averages 8

1.4 Prediction error and choice of the weight function 12

Chapter 1

Gridding

1.1 Position of the problem

We have M data $\{Y_k;\ k = 1, 2, \ldots, M\}$ that we assume to be scalar. Each observation Y_k is related to a point P_k with coordinates $\underline{t}_k$ ($\underline{t}_k$ is thought of as a vector of cartesian coordinates in 1D, 2D or 3D) so the input data set is $\{\underline{t}_k, Y_k;\ k = 1, 2 \ldots M\}$. We assume that $\{Y_k\}$ corresponds to a sampled signal $\{s(\underline{t}_k)\}$, plus a perturbation $\{\nu_k\}$, that we shall assume as a "white noise", namely a vector of uncorrelated random variables with zero mean and the same variance $\sigma^2(\nu_k) = \sigma_\nu^2$.

The signal is assumed to be a function smooth enough to let the expression $s(\underline{t}_k)$ (pointwise value) to be meaningful. For instance we assume that $s(\underline{t})$ is at least continuous in the region A where data are sampled.

The osbervation equation then writes

$$Y_k = s(\underline{t}_k) + \nu_k \tag{1.1.1}$$

with

$$E\{\nu_k\} = 0 \quad E\{\nu_k \nu_j\} = \delta_{kj}\sigma_\nu^2 \ . \tag{1.1.2}$$

We want to estimate the values of the signal $s(\underline{t})$ at the knots of a regular grid covering the region A, so that a new vector $\underline{\widetilde{Y}}$ can be substituted to $\underline{Y} = \{Y_k\}$.

In particular, when feasible, one would also like to be able to write the covariance matrix of the vector $\underline{\widetilde{Y}}$, due to noise propagation, as well as to give information on the biases of the estimates. In order to get a result like this, one could indeed use the reduction of the signal to a linear combination of some base functions $\{\varphi_n(\underline{t}); n = 1, 2 \dots N\}$, with $N < M$, estimate the coefficients $\{a_n\}$ of this linear combination via least squares and then propagate the predicted signal $\widehat{s}(\underline{t})$ to the points of the grid. This will be studied in the next chapter. Here we consider the possibility of a direct prediction of $s(\underline{t})$ at the grid knots by means of weighted averaging. So we could say that here we study direct gridding methods.

1.2 Grid numbering

Since it will be easier in the sequel to use different orderings for grid knots, we introduce already here a simple notation and an algorithm allowing to transform a plane grid indexing into a consecutive counting and vice versa.

We note that a 1D grid is a sequence of points on the real axis with constant distance between two subsequent points

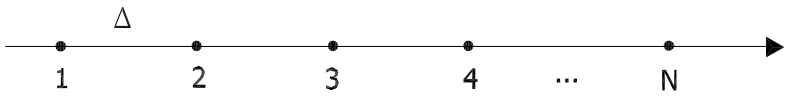

Figure 1.1: the 1D case.

So a grid is just a sequence

$$t_k = k\Delta \quad k = 1, 2 \dots N \ . \tag{1.2.1}$$

For the sake of simplicity, we shall assume that $\Delta = 1$, though this is not strictly necessary. We notice that to each point $t_k = k\Delta$ we can associate the interval I_k to the right of the point. Indeed there are $(N-1)$ such intervals. Furthermore, if we take any real t and we want to know the index of the interval in which it falls, we have

$$k = \left[\frac{t}{\Delta}\right] + 1 \tag{1.2.2}$$

where $[r]$, the integer part of the real r, is the maximum integer i such that $i \leq r$.

In 2D the situation is a little more complicated. Consider a regular (square) grid, like in fig. 1.2, of H columns and $\mathcal{K}$ rows.

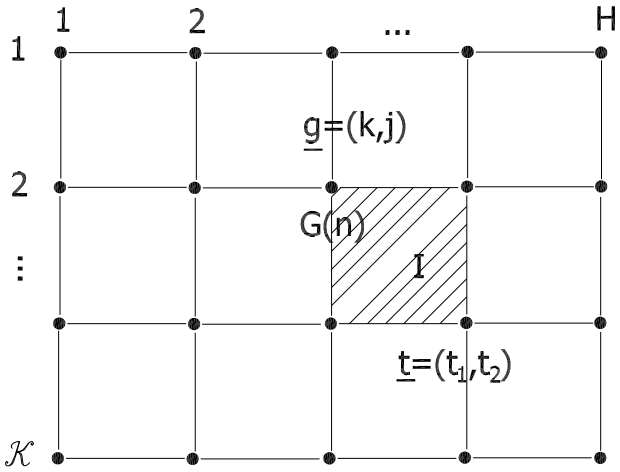

Figure 1.2: a regular grid of side Δ with H columns and $\mathcal{K}$ rows. The knot G has grid coordinates (k, j) and sequential index n.

So the number of knots in the grid is $\mathcal{K} \times H$. Each knot, for instance G in Fig. 1.2, has then two "natural" grid coordinates, which for this specific example are

$$k = 2 \ , \ j = 3 \ .$$

At the same time, we could decide to order the knots through a unique progressive index n, running from 1 to $\mathcal{K} \times H$, starting from the left upper corner and moving along the lines from left to right and from top to bottom. We want to see what is the rule that allows assigning to the knot with grid coordinates $\underline{g} = (k, j)$ the order index n. By looking at the figure, one discovers that

$$n = (k-1)H + j \ . \tag{1.2.3}$$

Vice versa, given $n, (k,j) = \underline{g}$ are derived from formulas

$$\begin{cases} k = \left[\dfrac{n-1}{H}\right] + 1 \\ j = n - (k-1)H \end{cases} \tag{1.2.4}$$

In addition, to each grid point G we can associate a square which has G as upper left corner. This square can then be indexed by $\underline{g}(G)$ or $n(G)$, and we will denote it as $I_{\underline{g}}$ or I_n. We note that indeed there are $(H-1)(\mathcal{K}-1)$ of such squares.

1.3 Weighted averages

We will use in what follows a weight function $\mathcal{W}(\vartheta)$ with the following characteristics: $\mathcal{W}(\vartheta)$ is continuously differentiable and

$$\begin{cases} \mathcal{W}(0) = 1 \\ \mathcal{W}'(\vartheta) < 0\ , \quad \vartheta > 0 \\ \mathcal{W}'(\vartheta) \to 0\ , \quad \vartheta \to \infty\ . \end{cases} \tag{1.3.1}$$

Note that it results automatically that

$$0 < \mathcal{W}(\vartheta) \leq 1\ . \tag{1.3.2}$$

Esempio 1.3.1 Useful examples of weight functions are

$$\mathcal{W}(\vartheta) = \frac{\varepsilon^\alpha}{\varepsilon^\alpha + \vartheta^\alpha}\ , \quad \varepsilon, \alpha > 0 \tag{1.3.3}$$

$$\mathcal{W}(\vartheta) = e^{-\alpha\vartheta}\ , \quad \alpha > 0\ . \tag{1.3.4}$$

We note that by decreasing ε in (1.3.3) we make $\mathcal{W}(\vartheta)$ steeper at the origin and by increasing α in both (1.3.3), (1.3.4), we accelerate the decay of $\mathcal{W}(\vartheta)$ to 0.

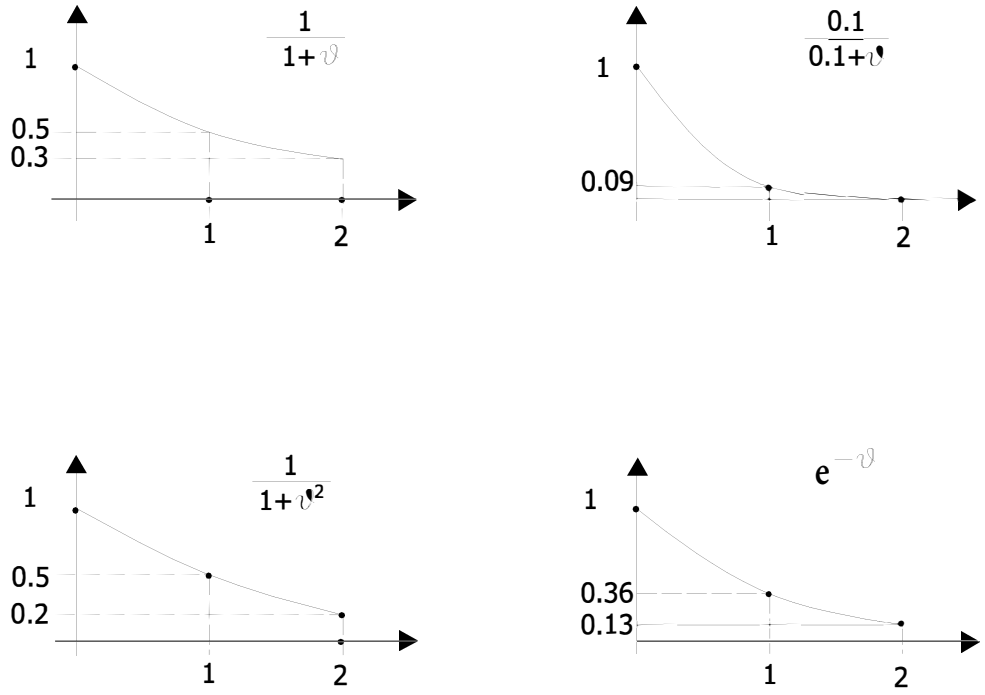

Figure 1.1: examples of weight functions.

The idea now is to make a prediction-filtering of the data, to derive an estimate of $s(\underline{t})$ at grid points, because this:

- facilitates the storing of data because by using the order number n we don't need to store the plane coordinates of $\underline{t}_k$ together with Y_k,
- allows for further numerical elaboration, for instance the Discrete Fourier Transform.

Naturally, there are as well disadvantages in gridding data: for instance, we loose in part the information of the original data and, in particular, we introduce a model error (also called *omission error*) due to the fact that $s(\underline{t})$ has not a precise theoretical model.

The prediction formula is a weighted average of the data

$$\widetilde{Y}_n = \widetilde{Y}_{\underline{g}} = \frac{\sum_{k=1}^{M} \mathcal{W}(|\underline{g} - \underline{t}_k|) Y_k}{\sum_{k=1}^{M} \mathcal{W}(|\underline{g} - \underline{t}_k|)} \ . \tag{1.3.5}$$

Note that by means of the denominator, in formula (1.3.5) the weights are "normalized", so that the new weights

$$\overline{\mathcal{W}}(\underline{g}, \underline{t}_k) = \frac{\mathcal{W}(|\underline{g} - \underline{t}_k|)}{\sum_{k=1}^{M} \mathcal{W}(|\underline{g} - \underline{t}_k|)} \tag{1.3.6}$$

satisfy the condition

$$\sum_{k=1}^{M} \overline{\mathcal{W}}(\underline{g}, \underline{t}_k) = 1 \tag{1.3.7}$$

This is essential if we want that every time we have Y_k = constant = C the same C is predicted at any other point; a request that is indeed quite reasonable.

Naturally the application of (1.3.5) could become very heavy if the number of observations and the number of grid knots is very high (e.g. more than 10^6), so that often the weighted average approach is applied by restricting the averaging area.

For instance for each knot $\underline{g}$ we could define an "influence" area $I_{\underline{g}}$ as the one composed by the 4 squares that have $\underline{g}$ in common, namely if $\underline{g} \leftrightarrow n$,

$$I_{\underline{g}} = I_n \cup I_{n-1} \cup I_{n-H} \cup I_{n-H-1} \ . \tag{1.3.8}$$

Equation (1.3.8) has to be suitably modified at border or corner points (see Fig. 1.2). Once this is done we define weighted averages on the influence area

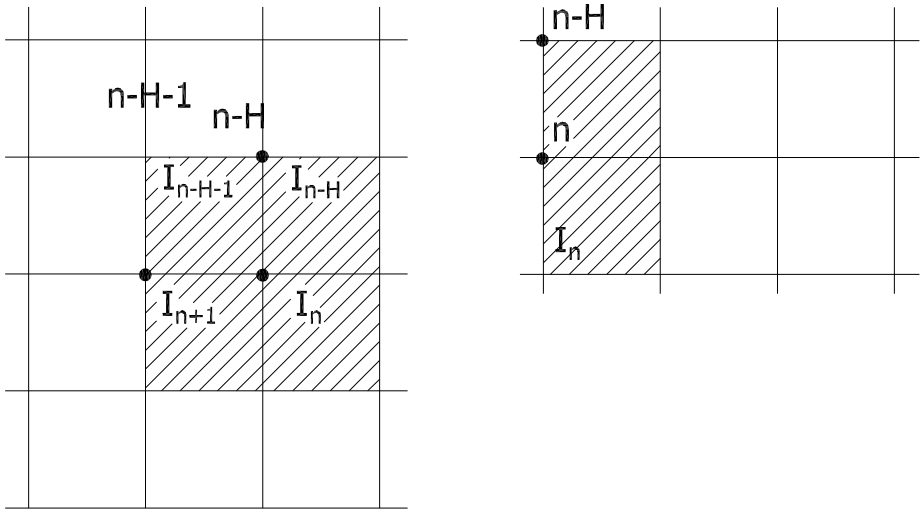

Figure 1.2: the influence area at an internal and at a border point.

only

$$\widetilde{Y}_n = \frac{\sum_{\underline{t}_k \in I_{\underline{g}(n)}} \mathcal{W}(|\underline{g}(n) - \underline{t}_k|) Y_k}{\sum_{\underline{t}_k \in I_{\underline{g}(n)}} \mathcal{W}(|\underline{g}(n) - \underline{t}_k|)} \tag{1.3.9}$$

1.4 Prediction error and choice of the weight function

We want, first of all, to have an idea of the variance or the r.m.s. of the prediction error, defined as

$$\begin{aligned} e(\underline{g}) &= \widetilde{Y}_{\underline{g}} - s(\underline{g}) = \\ &= \frac{\sum_{I_{\underline{g}}} \mathcal{W}(|\underline{t}_k - \underline{g}|)(s(\underline{t}_k) + \nu_k)}{\sum \mathcal{W}(|\underline{t}_k - \underline{g}|)} - s(\underline{g}) = \\ &= \left[\frac{\sum_{I_{\underline{g}}} \mathcal{W}(|\underline{t}_k - \underline{g}|)s(\underline{t}_k)}{\sum_{I_{\underline{g}}\mathcal{W}}(|\underline{t}_k - \underline{g}|)} - s(\underline{g})\right] + \left[\frac{\sum_{I_{\underline{g}}} \mathcal{W}(|\underline{t}_k - \underline{g}|)\nu_k}{\sum_{I_{\underline{g}}} \mathcal{W}(|\underline{t}_k - \underline{g}|)}\right] \\ &= e_m(\underline{g}) + e_\nu(\underline{g}) \end{aligned} \tag{1.4.1}$$

As we see we can split the error into a model prediction error plus a random error related to noise propagation. Naturally the model error $e_m(\underline{g})$ in (1.4.1) is deterministic and can assume any value depending on the true value of $s(\underline{g})$. So we cannot define any precise rule to estimate $e(\underline{t})$ because it could have a behaviour strongly dependent on the point $\underline{g}$. Yet, if we assume a kind of statistical homogeneity, at least at level of orders of magnitude, of $e_m(\underline{g})$, it would be reasonable to compute an index like

$$\mathcal{E}^2 = \frac{1}{N}\sum_{n=1}^{N} e^2(\underline{g}(n)) \ . \tag{1.4.2}$$

Nevertheless (1.4.2) is typically of no use because we usually don't know $s(\underline{g})$ and we cannot compute $e(\underline{g})$.

Nevertheless we can use the same criterion, namely the same prediction operator (1.3.9), centered to each of the points where we have observations.

In this way we can produce the predicted value

$$\widetilde{Y}_k = \frac{\sum_{I_{\underline{t}_k}} \mathcal{W}(|\underline{t}_k - \underline{t}_j|)Y_j}{\sum_{I_{\underline{t}_k}} \mathcal{W}(|\underline{t}_k - \underline{t}_j|)} \tag{1.4.3}$$

and compute the error

$$e(\underline{t}_k) = Y_k - \widetilde{Y}_k \ . \tag{1.4.4}$$

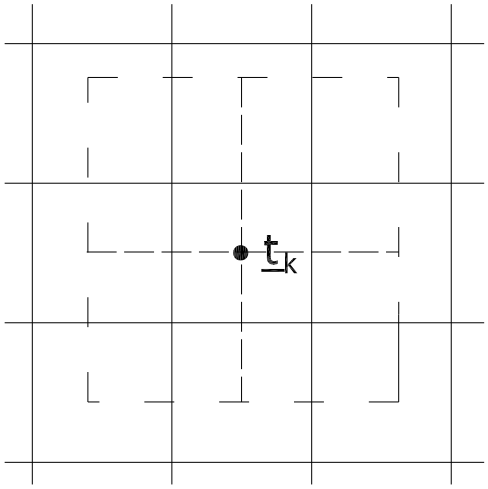

Figure 1.1: the influence function $I_{\underline{t}_k}$.

Once $e(\underline{t}_k)$ are known we can estimate

$$\mathcal{E}^2 = \frac{1}{M}\sum_{k=1}^{M} e^2(\underline{t}_k) \ . \tag{1.4.5}$$

This is known as the *leave-one-out* technique.

It has to be noticed that when (1.4.5) is too heavy from the computational point of view, one can always use a subsample of observation points. In this case though it is important that the selected leave-one-out points be well enough distributed in the data area, to be representative of the general mean square prediction error.

Another remark is that once $\mathcal{E}^2$ is suitably defined, we can use this as a criterion to select the appropriate weight function; in fact it is clear that the weight function that will produce the smaller $\mathcal{E}^2$ will be preferable with respect to the others.