Appendix E

Smoothness over Graphs

This appendix defines the mathematical notion of smoothness over geographic areas using concepts from graph theory. The ideas here have relatively few practical uses in the book, but they do convey the essential unity of the priors we offer, defined across any type of underlying variable. For priors defined over discretized continuous variables, we use analogous ideas to reduce the task of specifying the (non)spatial contiguity matrix to the choice of a single number (the order of the derivative in the smoothing functional). Unfortunately, a parallel reduction of effort is not possible for geographic space, although similar ideas apply.

We begin by denoting by G the set of cross-sectional indices. When cross sections vary by age, or similar variables, the set G is a discrete or continuous set endowed with a natural metric, and it is easy to formalize the notion of smoothness using derivatives or their discretized versions. When the cross-sectional indices are labels like country, familiar calculus does not help to define a notion of smoothness, but graph theory does.

In fact, when we have a number of cross sections that we want to pool (partially), it is natural to represent each of them as a point on the two-dimensional plane and join by a line segment points corresponding to cross sections that we consider "neighbors" of each other. This construction is called a *graph*, where we call the points *vertices* and the line segments *edges* of the graph. We denote the vertices and edges by V and E, respectively. Both vertices and edges are numbered using increasing positive integers: any numbering scheme is allowed, as long as it is used consistently. If i and j are two vertices connected by the edge e, we assign to e a *weight* $w(e) \equiv s_{ij}$, which represents our notion of how close the two cross sections are. The quantity $\rho(i, j) \equiv \frac{1}{\sqrt{s_{ij}}}$ is called the *length* of the edge e, and it is thought of as the distance between cross sections i and j. In the simplest case s, which is called the *adjacency matrix* of the graph, is a matrix of zeros and ones, where the ones denote country pairs that are considered as neighbors. If no edge exists between i and j, we set $s_{ij} = 0$. Vertex i could be connected to itself (in which case we would have a *loop*), so that we could have $s_{ii} \neq 0$; we will see later that for our purposes the value of s_{ii} is irrelevant, so we arbitrarily set $s_{ii} = 0$. The number of edges connected to the vertex i is called the *degree* of i, and we denote it by $s_i^+ \equiv \sum_j s_{ij}$ (in other words, s_i^+ is the sum of the elements of the i-th row of s, which is the number of neighbors).

If we have a function f defined over the graph, that is, $f : V \to \mathbb{R}$, it is possible to introduce the notion of a gradient. This is done by introducing the *oriented incidence matrix* Q, which is a $\#V \times \#E$ matrix whose rows and columns are indexed by the indices for V and E. To define Q, we first need to *orient* the graph, that is we assign a direction to each edge, so that edge e will point from some vertex i (the initial vertex) to some other vertex j (the terminal vertex). The orientation is arbitrary for our purposes, as long as it is fixed once for all. The matrix Q is built by setting entry Q_{ie} to $\sqrt{w(e)}$ if i is the initial vertex of e, to $-\sqrt{w(e)}$ if i is the terminal vertex of e, and 0 otherwise.

For example, if the first of row of Q is $(0, 0, 1, 0, -2)$, then vertex 1 is the terminal vertex of edge 5, which has a weight $w(5) = 4$, and the initial vertex of edge 3, which has a weight $w(3) = 1$. Notice that because each edge must have one initial and one terminal point, the columns of Q have even numbers of nonzero elements, and must sum up to 0.

Now that we have the matrix Q, we define a meaningful differential operator. At any given vertex, we think of the edges connected to that point as abstract "directions" from which one can leave that vertex. Therefore, given a function defined over V, it makes sense to characterize its local variation in terms of how much the function changes along each direction, that is, to assign to the edge e running from vertex i to vertex j the quantity:

$$\frac{f(j) - f(i)}{\rho(i, j)},$$

which obviously resembles a derivative. The matrix Q allows us to group all the quantities of this type in one single vector, which we think of as the gradient of the function f. In fact, denote by f the vector $(f)_i \equiv f(i) \;\; \forall i \in V$, and suppose e is the edge that runs from i to j; then, by construction:

$$(Q'f)_e = \frac{f(j) - f(i)}{\rho(i, j)}.$$

Therefore, we think of $Q'f$ as the gradient of f, and as Q' as the gradient operator. A measure of the smoothness of a function f defined over V is therefore obviously the quantity $\|Q'f\|^2$. The definition of smoothness we give in chapter 4 is operationally equivalent to this. A simple result of graph theory shows, however, that we need not compute the matrix Q' of the gradient in order to compute $\|Q'f\|^2$. In fact, the gradient operator is strictly connected to another important operator defined over the graph, that is, the *Laplacian*, defined more simply in terms of the adjacency matrix s:

$$W \equiv s^+ - s,$$

where the following relationship is well known:

$$W = QQ'.$$

A general smoothness functional for a function f defined over the graph therefore has the following form:

$$H[f] = \|Q'f\|^2 = f \cdot Wf = \sum_{ij} W_{ij} f_i f_j,$$

which can alternatively be written as

$$H[f] = \frac{1}{2} \sum_{ij} s_{ij}(f_i - f_j)^2.$$