Why Geological Regions?

Daniela Flores[1], Olga Kosheleva[2], and Vladik Kreinovich[1]
Departments of [1]Computer Science and [2]Teacher Education
University of Texas at El Paso
500 W. University, El Paso, Texas 79968, USA
floredan005@gmail.com, olgak@utep.edu, vladik@utep.edu

Formulation of the problem. In many practical problems, we want to describe how the value of some quantity q depends on the 2D or 3D spatial location x. In most such situations – whether it is a description of an electromagnetic field or of the state of the atmosphere – we use smooth (differentiable) functions to describe the dependence $q(x)$.

However, in geological sciences, the usual description consists of dividing the spatial area into *geological regions*: zones in each of which the value q is assumed to be constant. So why, in geosciences, this different approximating approach is more successful?

Our idea. In general, a natural way to describe an unknown function is to select an orthonormal basis $e_1(x)$, $e_2(x)$, ... Then, each function $q(x)$ can be represented as $q(x) = \sum\limits_{i=1}^{\infty} c_i \cdot e_i(x)$, where $c_i = \int q(x) \cdot e_i(x)\, dx$. So, with any desired accuracy, we can approximate the function $q(x)$ as $q(x) \approx \sum\limits_{i=1}^{n} c_i \cdot e_i(x)$, for a sufficiently large n.

In practice, we only know approximate values $\widetilde{q}(x) \approx q(x)$, so we get $\widetilde{q}(x) \approx \sum\limits_{i=1}^{n} \widetilde{c}_i \cdot e_i(x)$, where $\widetilde{c}_i = \int \widetilde{q}(x) \cdot e_i(x)\, dx$. We want to select the basis $e_i(x)$ for which this approximation is as accurate as possible. How can we measure this accuracy? This depends on the application.

- In weather prediction, we are not trying to predict the temperature or the wind speed at every single location in the city: understandably, some areas will be more windy, some less windy, some slightly warmer, some slightly colder. What we want to predict is average temperature over some area, average wind speed, etc. In such situations, a reasonable measure of accuracy is the usual "average" (mean square) difference $\int (q(x) - \widetilde{q}(x))^2\, dx$.
- In contrast, in geosciences, we are usually interested in specific locations: it is useless to learn that on average, the area contains some oil, we want to know where exactly is this oil. It makes sense to predict the weather in Southern California in general, but it would be useless to just say that this is a seismic zone: we want to know which areas are more vulnerable to future earthquakes. We want to make sure that the value $q(x)$ at each location x is accurately approximated, with some accuracy $\varepsilon > 0$.

The resulting explanation. To make sure that the sum of the terms $\widetilde{c}_i \cdot e_i(x)$ approximates the sum of the terms $c_i \cdot e_i(x)$, it is reasonable to require that each term $\widetilde{c}_i \cdot e_i(x)$ is as close to the corresponding ideal term $c_i \cdot e_i(x)$ as possible. In other words, we want to minimize the worst-case approximation error $A \stackrel{\text{def}}{=} \max\limits_{x, q(x), \widetilde{q}(x)} |\widetilde{c}_i \cdot e_i(x) - c_i \cdot e_i(x)|$, where $c_i = \int q(x) \cdot e_i(x)\, dx$, $\widetilde{c}_i = \int \widetilde{q}(x) \cdot e_i(x)\, dx$, and maximum is taken overall the functions $q(x)$ and $\widetilde{q}(x)$ for which $|\widetilde{q}(x) - q(x)| \le \varepsilon$ for all x.

It turns out that the smallest value of this worst-case approximation error A is attained when the function $e_i(x)$ is piece-wise constant, This explains why such an approximation – corresponding to geological regions – is indeed very effective in geosciences.