11-1-2021

Why Geological Regions?

Daniela Flores
The University of Texas at El Paso, floredan005@gmail.com

Olga Kosheleva
The University of Texas at El Paso, olgak@utep.edu

Vladik Kreinovich
The University of Texas at El Paso, vladik@utep.edu

Comments:

Technical Report: UTEP-CS-21-93

Why Geological Regions?

Daniela Flores, Olga Kosheleva, and Vladik Kreinovich

Abstract In most practical applications, we approximate the spatial dependence by smooth functions. The main exception is geosciences, where, to describe, e.g., how the density depends on depth and/or on spatial location, geophysicists divide the area into regions on each of which the corresponding quantity is approximately constant. In this paper, we provide a possible explanation for this difference.

1 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. This can be the description:

- of an electromagnetic field or
- of the state of the atmosphere

In most such situations, 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*. These are zones in each of which the value q is assumed to be constant.

So why, in geosciences, this different approximating approach is more successful?

Daniela Flores and Vladik Kreinovich
Department of Computer Science, University of Texas at El Paso
El Paso, Texas 79968, USA
e-mail: floredan005@gmail.com, vladik@utep.edu

Olga Kosheleva
Department of Teacher Education, University of Texas at El Paso
El Paso, Texas 79968, USA
e-mail: olgak@utep.edu

2 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

$$q(x) = \sum_{i=1}^{\infty} c_i \cdot e_i(x),$$

where $c_i = \int q(x) \cdot e_i(x)\,dx$. So, with any desired accuracy, we can approximate the function $q(x)$ as

$$q(x) \approx \sum_{i=1}^{n} c_i \cdot e_i(x),$$

for a sufficiently large n.

In practice, we only know approximate values $\widetilde{q}(x) \approx q(x)$. So we get

$$\widetilde{q}(x) \approx \sum_{i=1}^{n} \widetilde{c}_i \cdot e_i(x),$$

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?

3 How can we measure approximation accuracy: usual case

How can we measure approximation 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 $\int (q(x) - \widetilde{q}(x))^2\,dx$.

4 Geosciences are different

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. However, 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.

In all these cases, we want to make sure that the value $q(x)$ at each location x is accurately approximated, with some accuracy $\varepsilon > 0$.

5 The resulting explanation: formulation of the result

We want 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

$$A \stackrel{\text{def}}{=} \max_{x,q(x),\widetilde{q}(x)} |\widetilde{c}_i \cdot e_i(x) - c_i \cdot e_i(x)|.$$

Here:

- we denoted $c_i = \int q(x) \cdot e_i(x)\,dx$ and $\widetilde{c}_i = \int \widetilde{q}(x) \cdot e_i(x)\,dx$, and
- the maximum is taken over all the functions $q(x)$ and $\widetilde{q}(x)$ for which, for all x, we have
$$|\widetilde{q}(x) - q(x)| \le \varepsilon.$$

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.

6 Proof

We want to minimize the expression

$$A \stackrel{\text{def}}{=} \max_{x,q(x),\widetilde{q}(x)} |\widetilde{c}_i \cdot e_i(x) - c_i \cdot e_i(x)|.$$

Here, $\widetilde{c}_i \cdot e_i(x) - c_i \cdot e_i(x) = \Delta c_i \cdot e_i(x)$, where

$$\Delta c_i \stackrel{\text{def}}{=} \widetilde{c}_i - c_i = \int \Delta q(x) \cdot e_i(x)\,dx \text{ and } \Delta q(x) \stackrel{\text{def}}{=} \widetilde{q}(x) - q(x).$$

Thus,

$$A = \max_{x,\Delta q(x)} |\Delta c_i \cdot e_i(x)| = \max_{x,\Delta q(x)} \left(|\Delta c_i| \cdot |e_i(x)|\right).$$

The only condition of $\Delta q(x)$ is that $|\Delta q(x)| \le \varepsilon$.

The maximized expression $|\Delta c_i| \cdot |e_i(x)|$ is the product of two terms:

- the term $|\Delta c_i|$ only depends on $\Delta q(x)$, and
- the term $|e_i(x)|$ only depends on x.

Thus,

$$A = \left(\max_{\Delta q(x)} |\Delta c_i|\right) \cdot \left(\max_y |e_i(y)|\right).$$

The largest value of the sum $\Delta c_i = \int \Delta q(x) \cdot e_i(x)\,dx$ is attained when each term $\Delta q(x) \cdot e_i(x)$ is the largest.

- When $e_i(x) \ge 0$, maximum is attained when $\Delta q(x)$ is the largest $\Delta q(x) = \varepsilon$, then $\Delta q(x) \cdot e_i(x) = \varepsilon \cdot e_i(x)$.
- When $e_i(x) \le 0$, maximum is attained when $\Delta q(x)$ is the smallest $\Delta q(x) = -\varepsilon$, then $\Delta q(x) \cdot e_i(x) = -\varepsilon \cdot e_i(x)$.

In both cases, the largest value is equal to $\varepsilon \cdot |e_i(x)|$. Thus:

$$\max_{\Delta q(x)} |\Delta c_i| = \max_{\Delta q(x)} \left| \int \Delta q(x) \cdot e_i(x)\,dx \right| = \int \varepsilon \cdot |e_i(x)|\,dx = \varepsilon \cdot \int |e_i(x)|\,dx.$$

So,

$$A = \varepsilon \cdot \left(\int |e_i(x)|\,dx \right) \cdot \max_y |e_i(y)|.$$

Minimizing A is equivalent to minimizing

$$J \stackrel{\text{def}}{=} \frac{A}{\varepsilon} = \left(\int |e_i(x)|\,dx \right) \cdot \max_y |e_i(y)|.$$

The functions $e_i(x)$ are orthonormal, so

$$\int e_i^2(x)\,dx = \int |e_i(x)| \cdot |e_i(x)|\,dx = 1.$$

For each x, we have $|e_i(x)| \le \max_y |e_i(y)|$. So:

$$1 = \int |e_i(x)| \cdot |e_i(x)|\,dx \le \int \left(\max_y |e_i(y)|\right) \cdot |e_i(x)|\,dx =$$

$$\max_y |e_i(y)| \cdot \int |e_i(x)|\,dx = J.$$

If at least for one x, we have $|e_i(x)| \cdot |e_i(x)| < \left(\max_y |e_i(y)|\right) \cdot |e_i(x)|$, then $1 < J$.

The smallest possible value $J = 1$ is therefore attained if for all x, we have:

$$|e_i(x)| \cdot |e_i(x)| = \left(\max_y |e_i(y)|\right) \cdot |e_i(x)|. \qquad (1)$$

- If $|e_i(x)| = 0$, the equality (1) is satisfied.
- If $|e_i(x)| \neq 0$, then we can divide both side of the equality (1) by $|e_i(x)|$ and get

$$|e_i(x)| = \max_y |e_i(y)|.$$

So, for each x, the value of $e_i(x)$ is:

- either equal to 0,
- or equal to $\pm \max_y |e_i(y)|$.

Thus, the optimal function $e_i(x)$ is indeed piecewise-constant. The statement is proven.

Comment. Ideas of this proof are similar to the ideas from [1].

Acknowledgments

This work was supported in part by the National Science Foundation grants:

- 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science), and
- HRD-1834620 and HRD-2034030 (CAHSI Includes).

It was also supported:

- by the AT&T Fellowship in Information Technology, and
- by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478.

The authors are thankful to all the participants of the 26th Annual UTEP/NMSU Workshop on Mathematics, Computer Science, and Computational Science (El Paso, Texas, November 5, 2021) for valuable discussions.

References

1. A. E. Brito and O. Kosheleva, “Interval + Image = Wavelet: For Image Processing under Interval Uncertainty, Wavelets are Optimal”, *Reliable Computing*, 1998, Vol. 4, No. 3, pp. 291–301.