TOWARDS OPTIMAL FEW-PARAMETRIC REPRESENTATION OF SPATIAL VARIATION: GEOMETRIC APPROACH AND ENVIRONMENTAL APPLICATIONS

Misha Koshelev[1,2], **Octavio Lerma**[3], **and Craig Tweedie**[3]
[1]Human Neuroimaging Lab, Division of Neuroscience
Baylor College of Medicine, Houston, TX 77030, USA
misha680hnl@gmail.com
[2]W. M. Keck Center for Interdisciplinary Bioscience Training
Houston, TX 77030, USA
[3]Cyber-ShARE Center
University of Texas, El Paso, TX 79968, USA
lolerma@episd.edu, ctweedie@utep.edu

Abstract. *In this paper, we use geometric approach to show that under reasonable assumption, the spatial variability of a field* $f(x)$*, i.e., the expected value* $F(z) \stackrel{\rm def}{=} E[(f(x+z)-f(x))^2]$*, has the form* $F(z) = \left|\sum\limits_{i=1}^n \sum\limits_{j=1}^n g_{ij}\cdot z_i\cdot z_j\right|^\alpha$*. We explain how to find* g_{ij} *and* α *from the observations, and how to optimally place sensors in view of this spatial variability.*

Need to describe spatial variability. To understand climate trends, we need to describe not only the values of temperature, humidity, wind speed and direction at a single location, we also need to know how these characteristics change from one location to the other. In other words, we need to describe spatial variability of the corresponding characteristics.

There is a similar need in other application areas. For example, to understand the brain activity within a region, in addition to describing brain activity at certain locations, we also need to describe how this brain activity changes from one location to the other, i.e., we need to describe spatial variability of the corresponding characteristics.

How to describe spatial variability: use of random variables. In general, we have a characteristic $f(x)$ that takes different values at different locations x. Since we cannot exactly predict the exact future values $f(x)$, it is reasonable to consider them random variables. Random variables $f(x)$ corresponding to different locations x form a *random field.*

How to describe spatial variability: use of normal distributions. The values $f(x)$ are determined by a large number of different factors. In statistics, the joint effect of many small independent factors is – due to the Central Limit Theorem – well described by a normal distribution; see, e.g., (Sheskin 2004). Thus, it is reasonable to assume that the variables $f(x)$ are normally distributed.

A normal distribution is uniquely determined by its first two moments, i.e., by the expected values $E[f(x)]$ and $E[f(x) \cdot f(y)]$. The values $E[f(x)]$ and $E[(f(x))^2]$ describe the behavior at a single location. Thus, to describe spatial variability, it is sufficient to describe the values $E[f(x) \cdot f(y)]$ for $x \neq y$. Since we know the values $E[(f(x))^2]$ and $E[(f(y))^2]$, describing $E[f(x) \cdot f(y)]$ is equivalent to describing the following expected value:

$$C(x, y) \stackrel{\text{def}}{=} E[(f(y) - f(x))^2] =$$

$$E[(f(y))^2] + E[(f(x))^2] - 2E[f(x) \cdot f(y)].$$

Homogeneity. Locally, the distribution is usually homogenous, i.e., does not change after a shift. Thus, if we change x to $x + z$ and y to $y + z$, we should get the same value $C(x, y)$: $C(x + z, y + z) = C(x, y)$. For $z = -z$, this leads to $C(x, y) = C(0, y - x)$. So, to describe spatial variability, it is sufficient to describe the function

$$F(z) \stackrel{\text{def}}{=} C(0, z) = E[(f(x + z) - f(x))^2].$$

Comment. For $z = 0$, the above definition leads to $F(0) = 0$.

Other natural requirements. It is reasonable to assume that $F(z)$ continuously depends on z.

It is also reasonable to assume that there is spatial variability, i.e., that $F(z) > 0$ for $z > 0$.

Another requirement is that $f(x)$ is very close to $f(y)$ only for close x and y. Formally, we will require that for some value $F_0 > 0$, the set $\{z : F(z) \leq F_0\}$ is bounded.

Comment. It should be mentioned that the spatial distribution is often *anisotropic*, i.e., depends on the direction. For example, a North-South oriented mountain range goes through the city of El Paso. The closeness to the mountain affects temperature, rainfall, wind, etc. As a result, the meteorological characteristics change much more when we move in the East-West direction than when we move in the North-South one.

We need to select a few-parametric family of functions $F(z)$. In different practical situations, we have different functions $F(z) \geq 0$. To describe all such situations, it is desirable to have a parametric family $\mathcal{F}$ of possible functions $F(z)$.

Often, we only have a limited amount of data, so we can only statistically significantly determine a small number of parameters of the function $F(z)$. For example, in environmental sciences, we have a limited number of observations in remote areas such as most areas of Arctic and Antarctica. In brain research, we also often only have limited data. To cover such situations, it is desirable to have simple, few-parametric families $\mathcal{F}$.

Desired properties of few-parametric families. The numerical value of a physical characteristic depends on the choice of a measuring unit. For example, for length, if we change from inches to cm, the numerical values increase by 2.54. In general, if we use a new unit which is λ times smaller than the previous one, then numerical values $f(x)$ increase by λ, and the resulting values of $F(z)$ increase by λ^2. In principle, we can have an arbitrary positive value $C = \lambda^2$, so it is reasonable to require that the family $\mathcal{F}$ contains, with every function $F(z)$, also all functions $C \cdot F(z)$ for every $C > 0$.

Another possible change is a change in spatial coordinates. In some applications, the usual coordinates work best, in other applications, polar, cylindrical, or other coordinates are more appropriate. Locally, each smooth coordinate transformation $x_i \to f_i(x_1, \ldots, x_n)$ can be well approximated by a linear function $x_i \to \sum\limits_{j=1}^{n} a_{ij} \cdot x_j + a_i$, i.e., in matrix terms, $x \to Ax + a$. Under this transformation, the difference $z = y - x$ is replaced with Az. It is therefore reasonable to require that the the family $\mathcal{F}$ contains, with every function $F(z)$, also all functions $F(Az)$ for all non-degenerate matrices A.

It turns out that these two requirements are sufficient to determine few-parametric families $\mathcal{F}$ with the smallest possible number of parameters.

Main result. *Let $\mathcal{F}$ be a $\dfrac{n\cdot(n+1)}{2}$-parametric family of continuous functions $F(z)$ from $\mathrm{I\!R}^n$ to $\mathrm{I\!R}$ for which $F(z)=0$, $F(z)>0$ for $z\neq 0$, and for some $F_0>0$, the set $\{z : F(z)\le F_0\}$ is bounded. Let us also assume that the family $\mathcal{F}$ contains, with every function $F(z)$, also all functions $C\cdot F(z)$ for all $C>0$ and all functions $F(Az)$ for all non-degenerate matrices A. Then, every function $F\in\mathcal{F}$ has the form*

$$F(z)=\left|\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}\cdot z_i\cdot z_j\right|^{\alpha}$$

for some real values α and g_{ij}.

Proof. In this proof, similarly to (Li et al. 2002), we will use ellipsoids centered at 0, i.e., ellipsoids $E=\{z : \sum g_{ij}\cdot z_i\cdot z_j\le 1\}$. We will call them *c-ellipsoids* (c for *centered*). To describe all such c-ellipsoids, we need to describe all symmetric matrices g_{ij}, so the family of c-ellipsoids is $\dfrac{n\cdot(n+1)}{2}$-dimensional. The border $\{z : \sum g_{ij}\cdot z_i\cdot z_j=1\}$ of an ellipsoid E will be denoted by ∂E.

1°. Let $F\in\mathcal{F}$. Let us first prove that there is a c-ellipsoid E_0 on whose border ∂E_0 we have $F(z)=F_0$ for all $z\in\partial E_0$.

1.1°. By definition of the class $\mathcal{F}$, the set $S\stackrel{\text{def}}{=}\{z : F(z)\le F_0\}$ is bounded, and each function $F\in\mathcal{F}$ is continuous. Since $F(z)$ is continuous, the set S is closed.

Every bounded set can be enclosed into a c-ellipsoid. It is known (see, e.g., (Busemann 2005)) that, among all ellipsoids containing a given closed bounded set, there is exactly one ellipsoid with the smallest volume.

Let E_0 denote the c-ellipsoid with the smallest volume that contains the set S. We will say that this ellipsoid *corresponds* to the function $F(z)$.

Comment. The existence of the smallest-volume ellipsoid follows from the fact that every continuous function on a compact set attains its minimum. Uniqueness follows from the fact that if we have two c-ellipsoids E and E' of the same volume containing the same set, then we can select coordinates in which both matrices are diagonal, i.e., have the form $\sum g_i \cdot z_i^2 \le 1$ and $\sum g'_i \cdot z_i^2 \le 1$; then, for $g''_i = \dfrac{g_i + g'_i}{2}$, the ellipsoid $\sum g''_i \cdot z_i^2 \le 1$ also contains the bounded set and, as can be easily shown, has a strictly smaller volume than E and E'.

$1.2°$. It is known that every c-ellipsoid E in appropriate affine coordinates becomes a unit ball $\{z : \sum z_i^2 \le 1\}$. In other words, every ellipsoid can be obtained from a unit ball by an appropriate affine transformation. By combining the affine transformations corresponding to E and to E_0, we conclude that E can be obtained from the ellipsoid E_0 by an affine transformation $z \to Az$.

Under an affine transformation, the ratio of volumes is preserved. So, since E_0 the c-ellipsoid with the smallest volume contains the set $S = \{z : F(z) \le F_0\}$, E is the c-ellipsoid with the smallest volume containing the set $S' = \{z : F'(z) \le F_0\}$, where $F'(z) \stackrel{\text{def}}{=} F(Az) \in \mathcal{F}$.

Different ellipsoids correspond to different functions $F'(z)$, so we have as many such functions $F'(z)$ as there are ellipsoids – i.e., a $\dfrac{n \cdot (n+1)}{2}$-dimensional family.

$1.3°$. There are many affine transformations (rotations) that preserve the ball; in particular, for every two points on a unit sphere, there is a rotation that transforms one into another.

Thus, there are many affine transformations that preserve every ellipsoid E. In particular, for every two points $z, z' \in \partial E$ on this ellipsoid's border, there is an affine transformation that preserves ∂E and transforms z into z'.

For the ellipsoid E_0, let us denote, by G_0, the group of all affine transformations that preserve ∂E_0.

$1.4°$. Let us show that the border ∂E_0 of the ellipsoid E_0 contains some points from the set $S \stackrel{\text{def}}{=} \{z : F(z) \le z_0\}$.

We will prove this by contradiction. Let us assume that the border ∂E_0 of the ellipsoid E_0 does not contain any points from the set S. Then, we can proportionally shrink E_0 and get a new c-ellipsoid with the smaller volume that still contains S. This contradicts to the fact that E_0 has the smallest volume. The statement is proven.

1.5°. Let us prove that for all $z \in \partial E_0 \cap S$, we have $F(z) = F_0$.

Indeed, since $z \in S$, by definition of the set S, we have $F(z) \le F_0$. On the other hand, since z belongs to the border ∂E_0, the point z is a limit of points z_n from outside E_0: $z_n \to z$. Outside E_0, there are no points from S, so for all $z_n \notin E_0$, we have $F(z_n) > F_0$. Since the function $F(z)$ is continuous, in the limit $z_n \to z$, we get $F(z) \ge F_0$. From $F(z) \le F_0$ and $F(z) \ge F_0$, we conclude that $F(z) = F_0$.

1.6°. Finally, let us prove that every point $z \in \partial E_0$ belongs to the set S; due to Part 1.5 of this proof, this will imply that $F(z) = F_0$ for all $z \in \partial E_0$.

We will prove this statement by contradiction. Let us assume that not every point $z \in \partial E_0$ belongs to the set S. Since transformations from G_0 transform every point $z \in \partial E_0$ into every other point $z' \in \partial E_0$, this means that not all transformations from G_0 preserve the intersection $\partial E_0 \cap S$. Thus, transformations that preserve the intersection form a subgroup $G_0' \subset G_0$. Subgroups of the group of rotations are well known, they have smaller dimension than G_0. Thus, we have a finite-parametric family of transformations (of dimension ≥ 1) that preserve ∂E_0 and turn the set $S = \{z : F(z) \le F_0\}$ into a different set S' – i.e., which turn $F(z)$ into a different function $F'(z)$ for which the ellipsoid E_0 is the same. Thus, we have an at least 1-dimensional family of functions $F'(z)$ corresponding to E_0.

By applying an affine transformation, we get a similar family of functions for every ellipsoid. The family of ellipsoids is already $\dfrac{n \cdot (n+1)}{2}$-dimensional, and for each of them, there is an $\ge$ 1-dimensional family of functions – thus, we get a $\ge \left(\dfrac{n \cdot (n+1)}{2} + 1\right)$-dimensional family of functions $F'(z)$ – which contradicts to our assumption that the whole family $\mathcal{F}$ is no more than $\frac{n \cdot (n+1)}{2}$-dimensional. This contradiction shows that indeed $\partial E \subseteq S$.

2°. The ellipsoid E_0 corresponding to the function $F(z)$ has the form $\{z : \|z\|^2 \le 1\}$, where $\|z\|^2 \stackrel{\text{def}}{=} \sum\limits_{i,j} g_{ij} \cdot z_i \cdot z_j$. Let us prove that the function $F(z)$ has the form $F(z) = h(\|z\|)$ for some function $h(t)$ from real numbers to real numbers.

In other words, we need to prove that for every value v, the function $F(z)$ has a constant value on the border $\partial E_v \stackrel{\text{def}}{=} \{z : \|z\|^2 = v\}$ of the ellipsoid $E_v \stackrel{\text{def}}{=} \{z : \|z\|^2 \le v\}$ which is obtained from E_0 by an appropriate dilation (homothety).

Indeed, if the function $F(z)$ had two different values on different points $z, z' \in \partial E_v$, then, similarly to Part 1.6 of this proof, we would be able to apply appropriate affine transformations and get a $\ge$ 1-parametric family of functions $F'(z)$ corresponding to the same ellipsoid E_0 and thus, a $\ge \left(\frac{n \cdot (n+1)}{2} + 1\right)$-dimensional family of functions $F'(z)$ – which contradicts to our assumption that $\dim(\mathcal{F}) \le \frac{n \cdot (n+1)}{2}$.

3°. To complete the proof, let us show that $h(t) = \text{const} \cdot t^{\alpha}$.

Let us consider the functions $F(z)$ corresponding to all c-ellipsoids E which have the same volume $V(E)$ as E_0: $V(E) = V(E_0)$. The dimension of the family of all such ellipsoids is $\frac{n \cdot (n+1)}{2} - 1$.

For every function $F(z) = h(\|z\|) \in \mathcal{F}$, and for every two real numbers $C > 0$ and $k > 0$, the family $\mathcal{F}$ contains the function $C \cdot F(k \cdot z) = C \cdot h(k \cdot \|z\|)$. The corresponding transformations form a 2-dimensional multiplicative group.

The resulting family of functions cannot be fully 2-dimensional, since then, by considering such a family for every ellipsoid E with $V(E) = V(E_0)$, we would have a family of dimension

$$\ge \left(\frac{n \cdot (n+1)}{2} - 1\right) + 2 = \frac{n \cdot (n+1)}{2} + 1 > \frac{n \cdot (n+1)}{2}$$

inside the family $\mathcal{F}$. Thus, in the 2-dimensional transformation group, there is a ≥ 1-dimensional subgroup that keeps the function $h(t)$ invariant.

All subgroups of the 2-dimensional transformation group are well known, so we have $C(k) \cdot h(k \cdot t) = h(t)$ for some $C(k)$, and hence, $h(k \cdot t) = C^{-1}(k) \cdot h(t)$. It is known (see, e.g., (Aczel 2006)), that every continuous function that satisfies this functional equation has the form $h(t) = A \cdot t^{\alpha}$ for some A and α. The statement is proven, and so is our main result.

Discussion: relation to Riemannian geometry. In general, the values g_{ij} describing spatial variability differ from one location to another. Thus, to describe spatial variability, we need to describe the values $g_{ij}(x)$ corresponding to different locations x. Mathematically, this is equivalent to describing a Riemannian metric.

How to determine g_{ij} and α from the empirical data? Based on the recorded values $f(x,t)$ at different locations x at different times $t = 1, \ldots, T$, we can estimate $C(z) = E[(f(x+z) - f(x))^2]$ as

$$C(z) = \frac{1}{T} \cdot \sum_{t=1}^{T} (f(x+z,t) - f(x,t))^2.$$

We can then use the following iterative procedure to find g_{ij} and α. Initially, we take $g_{ij}^{(0)} = \delta_{ij}$, i.e., $g_{ii}^{(0)} = 1$ and $g_{ij}^{(0)} = 0$ when $i \neq j$. At each iteration k, we start with the values $g_{ij}^{(k-1)}$, and do the following.

First, we estimate $\alpha^{(k)}$ from the condition $C(z) \approx \left|\sum g_{ij}^{(k-1)} \cdot z_i \cdot z_j\right|^{\alpha}$ We can find this α by taking the logarithms of both sides and applying the Least Squares Method to the resulting system of linear equations with unknown α:

$$\ln C(z) \approx \alpha \cdot \ln \left(\sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij}^{(k-1)} \cdot z_i \cdot z_j \right).$$

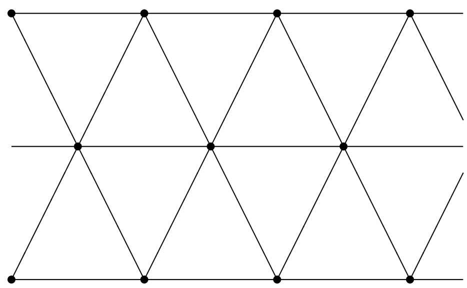
Once $\alpha^{(k)}$ is computed, we estimate $g_{ij}^{(k)}$ by applying the Least Squares Method to the following system of linear equations with unknown g_{ij}: $(C(z))^{1/\alpha^{(k)}} \approx \sum\limits_{i=1}^{n} \sum\limits_{j=1}^{n} g_{ij} \cdot z_i \cdot z_j$.

Towards optimal sensor location. We want to place the sensors so as to reconstruct the value of $f(x)$ at all locations x with the desired accuracy ε. (Thus, in the spatial direction along which $f(x)$ changes faster we should place sensors more frequently.)

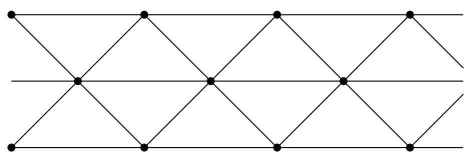
In precise terms, we want to place sensors in such a way that for each spatial location x, there is a sensor location s for which

$$E[(f(x) - f(s)]^2 = \left| \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij} \cdot (x_i - s_i) \cdot (x_j - s_j) \right|^{\alpha} \leq \varepsilon^2.$$

For every symmetric matrix g_{ij}, there are affine coordinates – formed by its eigenvectors – in which this matrix become a unit matrix. In this case, the above condition simply means that every location must be ε-close to a sensor location. It is known that under such condition, the asymptotically smallest number of sensors is provided by an equilateral triangle grid, i.e., a grid formed by equilateral triangles; see, e.g., (Kershner 1939).



Hence, in general, the sensor grid can be obtained from the equilateral triangle one by an appropriate affine transformation.



In other words, we should place sensors along the grid parallel to eigenvectors of the matrix g_{ij}.

Future work. We plan to apply the above model to describe the spatial dependence of the meteorological data and to make the corresponding recommendations on the optimal sensor placement.

A similar problem of spatial distribution. Instead of spatial *variation*, we can consider a similar problem of spatial *distributions*, i.e., the problem of describing low-dimensional affine-invariant families of probability density functions – families that contain, with every function $\rho(x)$, the function $(\det(A))^{-1} \cdot \rho(Ax + a)$. Similar ellipsoid arguments – but with general ellipsoids instead of c-ellipsoids – show that in this case, every distribution from the corresponding family has the form $\rho(x) = h(\|x - a\|)$ for some function $h(t)$ and some vector a, where $\|z\|^2 = \sum\limits_{i=1}^{n} \sum\limits_{j=1}^{n} g_{ij} \cdot z_i \cdot z_j$ for some values g_{ij}.

Acknowledgment. This work was supported by a training fellowship from the National Library of Medicine (NLM) to the Keck Center for Interdisciplinary Bioscience Training of the Gulf Coast Consortia (NLM Grant No. T15 LM007093), and by the National Science Foundation grant HRD-0734825.

References

J. Aczel, *Lectures on Functional Equations and Their Applications*, Dover, New York, 2006.

H. Busemann, *Geometry of Geodesics*, Dover, New York, 2005.

R. Kershner, “The number of circles covering a set”, *American Journal of Mathematics*, 1939, Vol. 61, No. 3, pp. 665–671.

S. Li, Y. Ogura, and V. Kreinovich, *Limit Theorems and Applications of Set Valued and Fuzzy Valued Random Variables*, Kluwer Academic Publishers, Dordrecht, 2002.

D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman & Hall/CRC, Boca Raton, Florida, 2004.