Spatial Analysis through the Looking Glass

Introduction

One of the delightful attractions that mathematics and geography have in common is that they offer fresh views of the world. Both fields of inquiry are circumspect almost by definition, and as soon as one looks around without prejudice, one discovers alternatives to one's provincial world view.

In mathematics, formalism and generalization are the seemingly sterile keys that open the gates to fertile fields of invention that, for its unexpected—even unexpectable—consequences, is indistinguishable from discovery. Formalism engenders the "what if" assumptions that lead to strange constructions, with their uncanny applicability to physics—non-Euclidean geometry, complex numbers, the Mandelbrot set. Generalization is the building of broader and broader analogies—the essence of cognition that Hofstadter has called 'chunking" (Hofstadter 2002). The joy of mathematics is that one gets more than what one bargained for; one discovers connections that one did not anticipate. The net effect is that the familiar is seen anew, as a "special case".

In geography, one discovers cultural symmetry, and cultural asymmetry. One realizes that relations are not absolute, but dependent on location in space and in time. Unless one is very well indoctrinated, one develops a sense of the relative, even "accidental", nature of one's point of view.

Science seeks to discover the laws which describe the rhythms and patterns of nature in time and space (Feynman 1967); in fact the essence of reality may be *relations*, or interactions, rather than any inherent objective attributes (Anandan 2003).

In this spirit, I'd like to offer here an example of a geographic-analytic point of view that I think is a refreshing departure from the customary provincial point of view in spatial analysis.

Case in point: Interpolation of spatial data

Interpolation of spatial data from irregular samples is a familiar problem in the geosciences. In the case of remotely-sensed data, the problem may be that of reconstruction of an image where data are missing; in the case of *in situ* data, the problem may be that of producing a data model on a regular spatial grid. Notwithstanding the prevailing view that "the location of events is considered an essential part of the observations" (Tobler 1966), it may be helpful to regard the data and the desired interpolation as objects in an abstract information space, not essentially dependent on the spatial coordinate system.

In this way, a data set $\mathbf{d}$ is thought of as a point in a space of as many dimensions (or coordinates) as the data have spatial locations. Let us say for specificity that there are n data points, and that the desired interpolation would have m data points, with $m > n$. The n components of the data vector $d_1 \ldots d_n$ are simply the n data values for each of the n coordinates (*i.e.* spatial locations). The desired interpolation $\hat{\mathbf{i}}$ would then be a point in a space of m dimensions (including $m - n$ additional dimensions for the missing data locations) and this interpolation would be determined by a *projection* of $\mathbf{d}$ from a space of n dimensions to one of m dimensions.

This way of looking at the problem is useful because of the importance of *context* to interpolation: any data set, like any signal, can be thought of as belonging to a class of similar data, which probably has a Gaussian distribution, in the m-dimensional information space, that is contained by a generalization of an ellipsoid (see, *e.g.*, (Goyal 2001)). The object of interpolation, then, is to project from $\mathbf{d}$ to $\hat{\mathbf{i}}$, where $\hat{\mathbf{i}}$ lies within the generalized m-dimensional ellipsoid of the expected distribution that represents *context*.

A linear *transform* of the data, such as a Fourier transform (or FFT, Fast Fourier Transform), Discrete Cosine Transform (DCT), or a Principal Component Transform (PCT), may be regarded simply as a change of the coordinate axes of the information space, not affecting the informational essence of the data, but affording a new view of the data in relation to its ellipsoid of context.

Band-limitation and its generalization

A class of data may be *band-limited*, which means that a limited number of the FFT or DCT components are significant in the representation of data sets from the class, ordinarily the longer-wavelength components. According to the Shannon-Whittaker-Kotel'nikov sampling theorem, accurate interpolation is possible if there are about two samples per shortest wavelength, not necessarily regularly-spaced samples (see, *e.g.* (Papoulis 1977, Unser 2000)). Given band-limitation, it may be much easier (and computationally more efficient) to carry out interpolation in the transform domain.

The object of *principal component analysis* is to determine the major, semi-major, etc., axes of the m-dimensional dataset distribution ellipsoid in information space, and to define a new coordinate system aligned with those axes. There are typically few axes of significant extent, relative to the total number of dimensions of the information space, so members of the class can be distinguished by a minimum number of components. Likewise, by extension of the sampling theorem, reconstruction of data sets which are known to be members of the class can be based on very few point samples (as shown, for example, by (Everson & Siroich 1995), for the case of images of faces).

Many data classes, such as grayscale images, can be modeled as *autoregressive* sources, for which adjacent data values are related by a correlation coefficient ρ close to 1 plus a small noise term $\mathbf{z}$:

$$\mathbf{d}_k = \rho \mathbf{d}_{k-1} + \mathbf{z}_k \ .$$

For such data classes, the PCT can be approximated bt a DCT (Goyal 2001), which is the reason that the first JPEG image-compression standard was based on the DCT.

In consideration of various linear transforms by which the members of a data class might be represented by a limited number of components, a generalization of the concept of band-limitation suggests itself: if data to be interpolated belong to a class whose distribution ellipsoid can be aligned with the coordinate axes under some transform, then the data can be called "band-limited" in that transform space. More to the point, interpolation can then be carried out in a computationally efficient manner.

Interpolation in practice

Notwithstanding the ideal that interpolation should not introduce information not present in the given data (Briggs 1974), interpolation is necessarily a hypothesis about the character of the interpolated surface (Lam 1983). That is so because any interpolation is the addition of hypothetical information components to given information components, based on known constraints, as the data vector of given components is projected to an information space of higher dimension. The essence of information is context (Martin 2004). Therefore the problem of interpolation is to venture a least-biased projection of a limited vector of given data onto a more complete vector of interpolated data. With the best choice of transform, the additional components will have value zero, while the original data components (*i.e.* in the standard coordinates) will remain unchanged.

Common interpolation methods

Splines

For many purposes, a least-biased estimate is assumed to be that the interpolated surface should be smooth (or band-limited after FFT), so *minimum curvature* algorithms (employing splines) have been favored, and found to be the best for most scattered data sets (Rauth & Strohmer 1998). Briggs (Briggs 1974) advocated such a method for interpolation of geophysical fields, which was refined by Smith & Wessel's (Smith & Wessel 1990) method of "splines in tension" for gridding such data as that of seafloor topography, which may exhibit more abrupt changes of direction than do gravity or magnetic fields.

Equivalent source

The *equivalent source* field-interpolation technique of Dampney (Dampney 1969), extended by Cordell (Cordell 1992) is an algorithm employing geophysical inversion, solving for field sources consistent with given data, then calculating intervening field values from the posited sources. Its limitations are with respect to propriety of a field data model and with respect to computational efficiency.

Kriging

This is a statistically-based interpolation technique based on estimation of how data values vary as a function of distance (see, *e.g.*, (Journel 1989)). In this technique, hypotheses about the character of the surface are based strictly on the given data. It is explicitly a least-biased estimation technique, but it tends to be computationally intensive (Doucette & Beard 2000).

Iterative Fourier-based method

Computationally attractive iterative algorithms for reconstruction (or interpolation) of band-limited signals from irregular samples were devised by Marvasti (Marvasti 1991) and refined by Feichtinger,

Gröchenig, Strohmer, and others (Feichtinger & Strohmer 1993, Strohmer 1997). In these algorithms, the error of a projection (referred back to the original basis of coordinates) is fed back into the algorithm, its projection added to the first approximation, and so on iteratively until a good "fit" to the given sample values is achieved.

Bandwidth in geoscientific data

The iterative Fourier-based methods are most effective for interpolation when the bandwidth of the data to be interpolated is known *a priori*. If the bandwidth is incorrectly assumed, the interpolation result will be unsatisfactory (Martin 2004): if the upper band limit is underestimated, the interpolation will be too "smooth", missing some of the sample points; if the upper band limit is overestimated, the interpolation will exhibit wild excursions even as it passes through all of the sample points (like curve-fitting with a polynomial of high degree).

Many data in geosciences belong to a class with $1/f$ spectral characteristic, where the magnitude of FFT components vary inversely with the "frequency", or wavenumber. For example, seafloor topography exhibits $1/f$ spectral characteristic over seven orders of magnitude (Bell 1975). In general, it is difficult to devise an appropriate spectral filter in order to implement the iterative Fourier-based method (Martin 2004).

The ideal case for interpolation is to have a large number of data sets, to which principal-component analysis might be applied, in order to arrive at a figure of generalized "bandwidth", in terms of number of significant principal components.

Computing principal components

Given a number s of data sets (data vectors $\mathbf{d}_k$) of the same presumed class, and of the same length n, principal component analysis may be carried out as follows using Matlab: first one constructs the s x n matrix $\mathbf{P}$ whose s rows are the s representative data sets, each of length n, from the data class[1]. The covariance matrix $\mathbf{C}$ is computed:

$$\mathbf{C} = \text{cov}(\mathbf{P}) \ ,$$

and after this, a matrix $\mathbf{V}$ of eigenvectors and a diagonal matrix $\mathbf{D}$ of eigenvalues is computed:

$$[\mathbf{V}, \mathbf{D}] = \text{eig}(\mathbf{C}) \ .$$

The columns of $\mathbf{V}$ are the principal component vectors, in reverse order of importance, which importance is indicated by the magnitude of the corresponding eigenvalues, on the diagonal of $\mathbf{D}$.

The columns of $\mathbf{V}$ thus derived can be reassembled to construct an inverse transform matrix for interpolation based on the hypothesis of generalized "band limitation", as described below.

Matrix computation for interpolation in a transform base

A linear transform $\mathbf{T}$, which is an operator used to go from the representation of a data set $\mathbf{i}$ in the standard base to the representation $\acute{\mathbf{i}}$ in the transform base, is a system of equations that can be expressed as the matrix multiplication

$$\mathbf{Ti} = \acute{\mathbf{i}} \ .$$

Likewise the inverse transform can be written

$$\mathbf{T}^{-1}\acute{\mathbf{i}} = \mathbf{i} \ .$$

Here we assume $\mathbf{T}$ to be an m x m matrix, where m is the the number of points in the desired interpolation.

If a set of samples $\mathbf{d}$ are taken at points $\mathbf{p}$ of $\mathbf{i}$, so that we can write

$$\mathbf{d} = \mathbf{i}(\mathbf{p}) \ ,$$

then to represent the data in the transform base, the general system of equations to solve can be written

$$\mathbf{T}_{\mathbf{p}}^{-1}\acute{\mathbf{d}} = \mathbf{d} \ ,$$

---

[1]For spatial data of two or three dimensions, the data can be arranged into a one-dimensional vector, for example concatenating the rows of an image in alternating order.

where $\mathbf{T_p^{-1}}$ is $\mathbf{T^{-1}}$ having only the rows corresponding to $\mathbf{p}$. In Matlab notation,

$$\mathbf{T_p^{-1}} = \mathbf{T^{-1}}(\mathbf{p},:) \ .$$

After a least-squares estimate of transform coefficients $\hat{\hat{\mathbf{d}}}$ is found (again in Matlab notation) by

$$\hat{\hat{\mathbf{d}}} = \mathbf{T_p^{-1}} \backslash \mathbf{d} \ , \tag{1}$$

the interpolation $\hat{\mathbf{i}}$ is computed as

$$\hat{\mathbf{i}} = \mathbf{T^{-1}} \hat{\hat{\mathbf{d}}} \ .$$

Equation 1 will not in general be well-determined since $m > n$. However, the key point is that a "band-limited" matrix computation can be formulated when solution is sought for a limited set of significant components. This is done by removing the columns of the inverse transform matrices $\mathbf{T_p^{-1}}$ and $\mathbf{T^{-1}}$ corresponding to components that are considered to be insignificant in the context of the class of data.

Non-iterative method proven with FFT

The above-described method of matrix computation was used in conjunction with FFT in a Matlab routine "FFTINTERP"[2], and found to run faster than an "Adaptive Weight, Conjugate Gradient" iterative routine (Martin 2004). It was also found to be more readily applicable to data characterized by $1/f$ spectra since *the source bandwidth need not be specified*. Instead, the bandwidth of interpolation is determined (by default) by the number of samples.

The routine "FFTINTERP" is not to be confused with the Matlab routine "INTERPFT", which requires regularly-spaced samples. It does, however, converge to the same result for regular samples, and runs faster.

The "FFTINTERP" routine demonstrates the efficiency of interpolation in a transform space, as well as the idea of adaptive band limiting, based on number of samples. By substituting a different inverse transform matrix in the algorithm (such as a PCT), the idea of generalized "bandwidth" can be demonstrated and applied analogously.

Conclusion

A practical computation algorithm for interpolation of irregularly-spaced data on a transform basis has been presented, motivated by a conceptual understanding of data sets residing in information space. This has been offered as proof of the idea that even the customary spatial point of view of the geographer is not absolute.

The relation between generalized "bandwidth" and "context" of the data to be interpolated was discussed, leading to explicit computation formulas involving partial inverse transform matrices. This was offered as a substantial example of how a formal change in point of view affords one the tools to understand and solve problems from a fresh point of view.

References

Anandan, J. (2003). "Laws, symmetries, and reality" *International Journal of Theoretical Physics*. **42**(9): 1943–1955.

Bell, T. (1975). "Statistical features of sea floor topography" *Deep Sea Research*. **22**: 883–892.

Briggs, I. (1974). "Machine contouring using minimum curvature" *Geophysics*. **39**: 39–48.

Cordell, L. (1992). "A scattered equivalent-source method for interpolation and gridding of potential field data in three dimensions" *Geophysics*. **57**(4): 629–636.

Dampney, C. (1969). "The equivalent source technique" *Geophysics*. **34**(1): 39–53.

Doucette, P. & Beard, K. (2000). "Exploring the capability of some GIS surface interpolators for DEM gap fill" *Photogrammetric Engineering and Remote Sensing*. **66**: 881–888.

---

[2]Available: http://martinandriener.com/Papers.

Everson, R. & Siroich, L. (1995). "Karhunen-Loève procedure for gappy data" *J.Opt. Soc. Am. A.* **12**(8): 1657–1664.

Feichtinger, H. & Strohmer, T. (1993). "Fast iterative reconstruction of band-limited images from irregular sampling values" In D. Chetverikov & W. Kropatsch (eds), *Proceedings, 5th International Conference, CAIP '93.* Budapest pp. 82–91.

Feynman, R. (1967). *The Character of Physical Law.* M.I.T. Press. Cambridge, MA.

Goyal, V. (2001). "Theoretical foundations of transform coding" *IEEE Signal Processing Magazine.* **18**(5): 9–21.

Hofstadter, D. (2002). "Analogy as the core of cognition" In D. Gentner, K. Holyoak & B. Kokinov (eds), *The Analogical Mind.* MIT Press. Cambridge, MA In "15",pp. 499–538.

Journel, A. (1989). *Fundamentals of Geostatistics in Five Lessons.* American Geophysical Union. Washington, DC.

Lam, N. (1983). "Spatial interpolation methods: A review" *American Cartographer.* **10**(2): 129–149.

Martin, P. (2004). *"Spatial Interpolation in Other Dimensions"* Master's thesis Oregon State University Corvallis.
URL: *http://martinandriener.com/Thesis*

Marvasti, F. (1991). "Recovery of signals from nonuniform samples using iterative methods" *IEEE Trans. Sig. Proc..* **39**: 872–877.

Papoulis, A. (1977). *Signal Analysis.* McGraw-Hill. New York.

Rauth, M. & Strohmer, T. (1998). "Smooth approximation of potential fields from noisy scattered data" *Geophysics.* **63**: 85–94.

Smith, W. & Wessel, P. (1990). "Gridding with continuous curvature splines in tension" *Geophysics.* **55**: 293–305.

Strohmer, T. (1997). "Computationally attractive reconstruction of bandlimited images from irregular samples" *IEEE Trans. Image Proc..* **6**(4): 540–548.

Tobler, W. (1966). "Spectral analysis of spatial series" *Proceedings, 1966 URISA Conference..* pp. 179–186.

Unser, M. (2000). "Sampling—50 years after Shannon" *Proc. IEEE.* **88**: 569–587.