*Stanford Exploration Project, Report SERGEY, November 9, 2000, pages 143–**??***

Stanford Exploration Project, Report SERGEY, November 9, 2000, pages 143–??

Speculations on contouring sparse data: Gaussian curvature

Jon Claerbout and Sergey Fomel[1]

ABSTRACT

We speculate about regularizing (interpolating) sparse data. We speculate that L_1 regularization would be desirable. An example convinces us it would not. Changing direction we learn that flexed paper has zero Gaussian curvature. Unfortunately, Gaussian curvature is a nonlinear function of the altitude.

INTRODUCTION

Twenty-five years ago I[2] attended a series of lectures organized by the University of Houston called "Petroleum Geology for Geophysicists". One of the professors, Daniel Busch (if I recall correctly), proposed a data set that would be "interesting to contour". He might have said that specialized knowledge of petroleum reservoirs would be helpful. His experience was with very sparse data. I recall it being well logs from Mexico. Of special interest were (and are) sand thicknesses. He cited four wells, each with a measurement:

```
.                     5           98
.
.                           x
.
.                    100           7
```

The question for the interpreter is, what should be the value of `x`? Mathematical algorithms generally give a value of `x` near 53. As a petroleum geologist, Busch was accustomed to visualizing drainage patterns such as rivers with residual sands. For a river running northwesterly, the value of `x` would be near 6. On the other hand, he said, a paleotopography also commonly contains ridges, so maybe `x` should be roughly 100.

This example charmed me enough to remember it for 25 years and to relay it to you now, with the hope that we can do something helpful about Busch's problem (that mathematical contouring is not as good as common sense). The solutions that we are most accustomed to are the linear solutions that come from minimizing quadratic forms. Such setups generally give the average value of `x` near 54 that Busch would like to avoid.

[1]**email:** claerbout@stanford.edu, sergey@sep.stanford.edu

[2]In this paper "I" refers to the first author alone.

We might wonder if Busch's problem is too hard for any mathematical method. I think we would feel that progress had been made, however, if we uncovered a method that told us this data

```
.                              98
.                      5
.                         x
.
.                  100          7
```

suggests a river while this data

```
.                    5
.                            98
.                         x
.
.                  100          7
```

suggests a ridge. The whole problem could be more interesting in the presence of more data values. The more remote data values could actually be making the choice of a river channel or a ridge.

The goal is that the result should look more like topography than what we see arising from familiar L_2 methods. The essential aspect of real world topography is that erosion cuts drainage channels.

Will L1 solve the problem?

A motivation of this paper is to explore the idea that the L_1 norm might produce the kind of solutions that Daniel Busch would like to see. I'm far from certain of this. The behavior of the L_1 norm is well understood in *overdetermined* formulations. There L_1 is valuable because it rejects outliers (large residuals). The behavior of the L_1 norm in *underdetermined* problems is not well understood nor has it been widely observed.

One reason the behavior of the L_1 norm is not widely observed in underdetermined problems is that we do not have fast and reliable computational methods for problems of high dimensionality. For example, in Busch's example the unknowns are a cartesian mesh of all possible altitudes where just four altitudes are constrained. (Actually, we probably also need to specify the behavior at infinite distance.)

To see if L_1 might help us solve problems like the one posed by Busch, and to help us guess whether we should invest resources in L_1 solvers, I review here various examples of lesser scope.

EXAMPLES

Median of an even number of points

Consider four numerical values, say (5,7,98,100). The median value is the middle value. Since this is an even number of points, there is no middle. A way to define the median m mathematically is to choose m to minimize

$$q \quad = \quad |m-5|+|m-7|+|m-98|+|m-100| \tag{1}$$

If you plot the function $q(m)$ you find it has a flat spot between $m=7$ and $m=98$. This illustrates a principle of L_1 optimizations: The minimum residual is unique. The model, however, is not unique, but is a range. This is an interesting feature of L_1 which differs from our usual least squares L_2 experience where we never get intervals. With L_2, solutions are unique, except for a possible null space, an infinite family of added solutions. Having a 5th data point, even if very weakly weighted, would resolve the ambiguity so we might be on the track of a Busch-like solution.

Best fitting straight line

Consider the straight line that best fits a collection of data points. Suppose there are four points. Two of them are $(-10,-10)$, $(-10,10)$. I'll call these two data points the "left slot". The other two are $(10,-10)$, $(10,10)$ which I'll call the "right slot". You can easily see that straight lines that lie within both the left slot and the right slot all have the same sum of absolute distances from the line to the data. For each slot the sum is 20 so the total is 40. Thus L_1 gives us many lines inside the slots but it does not select any particular line. (This example is said to come from Albert Tarantola.)

You might object to having two data points at the same coordinate. By moving them apart a little, we suppose the "degeneracy" is broken, that a unique line becomes defined. Perhaps so. Never the less, it is clear that the residual is "almost minimum" for all lines inside the slots, and it is much bigger for lines outside the slots. Thus the reality of the slots remains, even where technically we might avoid them. Again, L_1 has the appealing feature that an additional data point, even if weighted weakly, could break the ambiguity.

Statics

An important example is the estimation of source and receiver time corrections. Here one has a set of observed traveltimes from the ith source to the jth receiver. After known systematic geometrical and velocity effects are removed, the time residual matrix t_{ij} remains. Then, near-source traveltimes s_i and near-receiver traveltimes r_j are estimated from the t_{ij} by minimizing the error e_{ij} in

$$e_{ij} = t_{ij} - s_i - r_j. \tag{2}$$

A trivial nonuniqueness is that an arbitrary constant added to all the s_i and subtracted from all the r_j will give the same residuals. I was surprised to discover deeper nonuniqueness lurking in a simple example. Absolute error minimization reduced a 3-by-3 matrix of t_{ij} to the e_{ij} residual matrix

$$e_{ij} = \begin{bmatrix} 0 & -12 & 4 \\ 17 & 0 & 0 \\ 0 & 10 & 0 \end{bmatrix}. \tag{3}$$

As expected theoretically (by the solution method I used), there are 5 zeros representing the 5 independent unknowns of the 6 unknowns. Note that $\sum |e_{ij}| = 43$. Now modify source and receiver times by applying $+12$ to row 1 and -12 to column 1. We have

$$\begin{bmatrix} 0 & 0 & 16 \\ 5 & 0 & 0 \\ -12 & 10 & 0 \end{bmatrix}, \tag{4}$$

still with $\sum |e_{ij}| = 43$. Now apply $+12$ to row 3 and -12 to column 3. We have

$$\begin{bmatrix} 0 & 0 & 4 \\ 5 & 0 & -12 \\ 0 & 22 & 0 \end{bmatrix}. \tag{5}$$

Furthermore, we can generate an infinite set of e_{ij} (and hence source and receiver corrections) all with the same $\sum |e_{ij}|$ by taking residuals (3)-(5) and forming any convex combination (weighted combination where each weight is positive and the weights sum to one).

The existence of a sizeable nonuniqueness with absolute error minimization leaves us the uncomfortable feeling that the mathematical uniqueness of squared error is not genuine, i.e., that the uniqueness of results with squared error is not a realistic charactorization of our certainty.

Often, however, the this unfamiliar nonuniqueness does not arise. It depends on the data, not the mathematical structure of the problem. For example, I don't know any other minimum L_1 solutions with the e_{ij} matrix:

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 7 & -11 \\ 0 & -3 & 8 \end{bmatrix}. \tag{6}$$

More details are found in (Claerbout and Muir, 1973) which is where I recovered this example.

Curve through two points

Consider values along a horizontal x-axis ranging from 1 to 100. Suppose at $x = 1$, the y value is given to be $y_1 = 1$. Likewise at $x = 100$ the y value is given to be $y_{100} = 100$. Now we are to find all the intervening points, $y_2, y_3, \ldots, y_{99}$. Let us use the L_1 criterion

$$\min_{y_2, y_3, \cdots, y_{99}} \quad |y_2 - y_1| + |y_3 - y_2| + |y_4 - y_3| + \cdots \tag{7}$$

The solution to this problem is any curve with a positive slope since all such curves result in the same value of 99 for (7) That is quite a lot of curves!

Now we begin to appreciate the strange flavor of L_1. We appreciate the idea that solutions are "intervals". But it is distressing to realize that we could often have graphical difficulty displaying the results. In practice we might need to settle for "seeing some examples." Perhaps a satisfactory way of generating those examples would be by using random starting values for the fitting.

Suppose we set up the Busch problem with L_1. Perhaps we will find the solution is not a unique surface. It might turn out to be a "mat" of variable thickness. It would be annoying to try to display the thickness, but perhaps the thickness is related to the uncertainty of the result. That should have value.

WHAT IS THE L1 NORM OF THE 2-D GRADIENT?

The idea of finding smooth solutions is to minimize a measure of the gradient. The first time I thought about doing this with L_1, I tried the wrong approach (and that put me off the track for 25 years). The wrong approach is to take the L_1 norm of the x-component of the gradient and add it to the L_1 norm of the y-component of the gradient. This is bad because it embeds the orientation of the coordinate system. Axiomatically, in science we like solutions that are independent of the human choice of a coordinate system. Thus L_1 appears to conflict with this basic requirement.

An approach independent of coordinate rotation and translation on a grid is to minimize

$$q(u) \quad = \quad \sum \sqrt{\nabla u \cdot \nabla u} \tag{8}$$

where $u = u(x, y)$ and where the summation is over (x, y)-space. Multivariate L_1-norm problems generally reduce to a line search that is a weighted median. Hoare's algorithm makes this very fast. Unfortunately, this multidimensional generalization of L_1 does not seem to reduce to a weighted median so Hoare's algorithm is irrelevant, as might be other L_1 experiences we have seen in $1 - D$.

I discussed the $\sum \sqrt{\nabla u \cdot \nabla u}$ criterion for a while with Bill Symes. We came up with this simple problem where we would use zero side boundaries and seek the response of an impulse in the medium.

```
0  0  0  0
0  a 10  0
0  b  c  0
0  0  0  0
```

The free variables are a, b, and c. We take the x derivative diagonally to the northeast and the

y derivative diagonally to the southeast.

$$\begin{array}{rcccccccc} q(a,b,c) & = & |a| & + & \sqrt{10^2+a^2} & + & 10 & + \\ & & \sqrt{b^2+a^2} & + & \sqrt{(a-c)^2+(b-10)^2} & + & \sqrt{10^2+c^2} & + \\ & & |b| & + & \sqrt{b^2+c^2} & + & c & \end{array} \tag{9}$$

A few manual calculations quickly convinced us that the solution is $a = b = c = 0$. Thus multidimensional L_1 does not look like the answer we seek. It looks like the boundaries at infinite distance dominate the data (in this case the 10). Thus, the response to an isolated collection of spikes, might simply be the spike values where they are given and zero elsewhere.

GAUSSIAN CURVATURE

I proposed that we find out the differential equation that describes the bending of paper and use it as a regularization. The idea is to encourage a Busch-like behavior. As with $L1$, I would like to have a linear operator to preserve the uniqueness of the solution. Uniqueness gives reliability. My exerience has taught me that if a method has multiple isolated minima, I will descend into the wrong one. If the paper-bending operator is nonlinear, I could linearize it.

Bill Symes suggested the Gaussian curvature. My favorite search engine (google.com) quickly gave me several references. Indeed a sheet of paper does seem to have a Gaussian curvature of zero. The Gaussian curvature of a 2-D function vanishes wherever the the function is locally one dimensional. The Gaussian curvature is the product of the principal curvatures. The Gaussian curvature is

$$\frac{h_{xx}h_{yy} - h_{xy}^2}{1 + h_x^2 + h_y^2} \tag{10}$$

For small dips, the numerator is the important part. The numerator is the determinant of the Hessian,

$$\det \begin{vmatrix} h_{xx} & h_{xy} \\ h_{yx} & h_{yy} \end{vmatrix} \tag{11}$$

We might regularize a collection of data points by minimizing this determinant. I have begun looking for references that may have previously investigated this very basic idea. Unfortunately, the function is nonlinear. We can linearize it. Replacing h by $\bar{h} + h$ and dropping terms in h^2 we get

$$0 \quad \approx \quad (\bar{h}_{xx}\bar{h}_{yy} - \bar{h}_{xy}^2) + \bar{h}_{xx}h_{yy} + h_{xx}\bar{h}_{yy} - 2\bar{h}_{xy}h_{xy} \tag{12}$$

The most important question is: what is $\bar{h}(x,y)$? How do we initialize it, and how can we safely update it? A way to initialize $\bar{h}(x,y)$ is to approximate the initial data by a best fitting one-dimensional parabola. One way to stablize $\bar{h}(x,y)$ is to smooth it in patches.

I am reminded of "LOMOPLAN", an earlier idea I had to fit a best plane wave, then use it to define a linear operator to use as a weighting function in estimation. The idea is that a sedimentary section consists of a single local plane wave. Perhaps that two-stage least squares process is akin to linearizing the Gaussian curvature.

MORE ON GAUSSIAN CURVATURE

Given a function $u(x,y)$, its x-derivative u_x, its y-derivative u_y, and a slope parameter p, we have the planewave operator L

$$0 \quad = \quad L(u) = u_x + p u_y \tag{13}$$

which vanishes when $u(x,y)$ is not really a two-dimensional function but is a one-dimensional function $u = f(x - py)$.

Next we get two equations from the plane-wave equation, one differentiating by x, the other by y.

$$0 \quad = \quad L(u_x) \quad = \quad u_{xx} + p u_{xy} \tag{14}$$
$$0 \quad = \quad L(u_y) \quad = \quad u_{yx} + p u_{yy} \tag{15}$$

Eliminate p from these two equations by solving for it:

$$-p \quad = \quad \frac{u_{xx}}{u_{xy}} \quad = \quad \frac{u_{yx}}{u_{yy}} \tag{16}$$

or

$$G(u) \quad = \quad u_{xx}u_{yy} - u_{xy}u_{yx} \quad = \quad 0 \tag{17}$$

The plane-wave operator $L(u)$ will not vanish unless u is a plane wave going in the direction specified by p. A remarkable property of the function $G(u)$ is that it vanishes for any orientation of plane wave. If we want to test a 2-D field for one-dimensionality, the test $L(u)$ requires us to know p. The test $G(u)$ does not. Generally we do not know p and we need to estimate it by statistical means in a window of some size that we must specify. In principle, G escapes those problems (although it might be worse in practice because it is a nonlinear function of the wavefield).

In differential geometry, a quantity appears that is known as the "Gaussian curvature". For small vertical motions u, this Gaussian curvature reduces to our expression G.

The gentle flexure of a sheet of paper follows the principle that the Gaussian curvature vanishes. The deformation must be one dimensional. We were first attracted to Gaussian curvature as a way of interpolating sparse data where the data represents a wavefield or a sedimentary earth model. We would seek the interpolation that was most "paper like", which minimized the integral of the square of the Gaussian curvature. Unfortunately, G is already a quadratic function of u even before we square G to minimize a positive value.

Thin-plate versus biharmonic equation

The biharmonic equation uses the Laplacian operator twice: The biharmonic equation results from minimizing the quadratic form

$$B(u) \quad = \quad u'(\partial_{xx} + \partial_{yy})'(\partial_{xx} + \partial_{yy})u \tag{18}$$
$$B(u) \quad = \quad u'(\partial'_{xx}\partial_{xx} + \partial'_{xx}\partial_{yy} + \partial'_{yy}\partial_{xx} + \partial'_{yy}\partial_{yy})u \tag{19}$$

To minimize, simply cancel off u' and set to zero. The thin plate equation resembles the biharmonic equation but differs in a subtle but important way. The quadratic form minimized for a thin plate is:

$$T(u) \quad = \quad v'v \qquad \text{where } v = \begin{bmatrix} \partial_{xx} \\ \partial_{xy} \\ \partial_{yx} \\ \partial_{yy} \end{bmatrix} u \tag{20}$$

or

$$T(u) \quad = \quad u'(\partial'_{xx}\partial_{xx} + \partial'_{yy}\partial_{yy} + \partial'_{xy}\partial_{xy} + \partial'_{yx}\partial_{yx})u \tag{21}$$

Again, we find the associated differential equation by canceling off the u'.

What is bothering me is that the dispersion relations look the same but the quadratic forms look different. The difference between the biharmonic quadratic form and the thin plate quadratic form lies in the cross term. Let us form half this difference $G = (B-T)/2$.

$$G(u) \quad = \quad u_{xx}u_{yy} - u_{xy}u_{xy} \tag{22}$$

We see the difference has turned out to be the Gaussian curvature. Although the difference is a quadratic form, it is not uniformly positive or negative, as it can have both signs.

By means of rotation of coordinates, we can diagonalize the matrix

$$\begin{bmatrix} u_{xx} & u_{xy} \\ u_{yx} & u_{yy} \end{bmatrix} \xrightarrow{\text{rotation}} \begin{bmatrix} K_1 & 0 \\ 0 & K_2 \end{bmatrix} \tag{23}$$

Thus we can think of $G = K_1K_2$ as the product of the curvatures while the biharmonic quadratic form is the square of the sum $B = (K_1 + K_2)^2$. In terms of curvatures, in the rotated coordinates the thin plate operator is

$$T \quad = \quad B - 2G \tag{24}$$

$$T \quad = \quad (K_1 + K_2)^2 - 2K_1K_2 \tag{25}$$

$$T \quad = \quad K_1^2 + K_2^2 \tag{26}$$

which is the sum of the squares of the curvatures.

The meaning is this: The biharmonic equation zeroes $B = (K_1 + K_2)^2$ so its solution could be expected to have many places of $K_1 = -K_2$ where the curvature on one axis is the negative of that on the other axis. In other words, solving the biharmonic equation might give us a function containing many saddles. On the other hand, the thin-plate equation $T = K_1^2 + K_2^2$ tries to eliminate both curvatures (not allowing credit for when one cancels the other). However, with respect to optimization, both quadratic forms are equivalent.

CONCLUSION

In conclusion, we do not see any immediate action items. The notion of minimizing Gaussian curvature is appealing, but it is nonlinear, which means that solutions depend on the starting location. Physically, when flexing paper, the final deformation probably depends on the deformation history. For practical purposes the thin plate operator is the Laplacian squared. If we are going to try to minimize the Gaussian curvature, a nonlinear criteria, we should probably begin from the thin plate which is unique.

REFERENCES

Claerbout, J. F., and Muir, F., 1973, Robust modeling with erratic data: Geophysics, **38**, no. 5, 826–844.