

## 11. The Numerical Alternative

The functional analysis approach to the norm minimization problem in Hilbert space shows how an exact theory can be worked out retaining the essential asymmetry between data and model: data are finite in number, but the model is a function that lives on an infinite-dimensional space. But real-world situations in which the Gram matrix can be calculated exactly, while not nonexistent, are in the minority. I mention for your edification two of my papers where exact Gram matrices are in fact found: Shure, L., Parker, R. L., and G. E. Backus: Harmonic splines for geomagnetic modeling, in *Phys. Earth Planet. Inter.*, 28, 215-29, 1982 and Parker, R. L.: Calibration of the pass-through magnetometer-I. Theory, in *Geophys. J. Internat.*, 142, 371-83, 2000. Numerical methods will in any case appear at some point in every calculation and so we will now develop a treatment based on linear algebra that is founded on a finite-dimensional approximation for the forward problem.

The key is that the linear functional that captures the solution to the forward problem, written in Chapter 2 of GIT as

$$(g_j, m) = d_j, \quad j = 1, 2, \dots, M \quad (1)$$

can be written in an approximation as a finite sum

$$g_j^T m = d_j, \quad j = 1, 2, \dots, M \quad (2)$$

where  $g_j, m \in \mathbb{R}^N$ ; so now the model and representers are vectors of dimension  $N$ . Of course (2) is more compactly written

$$Gm = d \quad (3)$$

where  $G \in \mathbb{R}^{M \times N}$  and  $d \in \mathbb{R}^M$  and the rows of the matrix  $G$  are the (row) vectors  $g_j^T$ . At this point in the class when the models are simple, we tend to think of  $M$ , the number of data, being a lot smaller than the  $N$  the dimension of the model space, but when big data sets are involved that isn't always the case.

In the vast majority of practical cases the inner product in (1) is representing an integral, as in the inner product of  $L_2$ . To make the transition to (2) one must **discretize** the underlying continuous problem. The most primitive way of doing this which, quite honestly I prefer, is to break the model space up into boxes and replace the integral by a simple sum. On the real line, for example, in the simple one-dimensional magnetization problem, we could use uniform sampling,  $\Delta x$ , and then

$$\int_a^b f(x) dx = \sum_{k=1}^N w_k f(a + (k-1)\Delta x) + \varepsilon \quad (4)$$

where  $\Delta x = (b-a)/(N-1)$ ,  $w_k$  are a set of weights, and  $\varepsilon$  is the error approximation. The easiest, and in most circumstances least accurate formula of this kind is the **trapezoidal rule** where one takes:

$$w_1 = w_N = \frac{1}{2}\Delta x, \quad w_2 = w_3 = \dots w_{N-1} = \Delta x \tag{5}$$

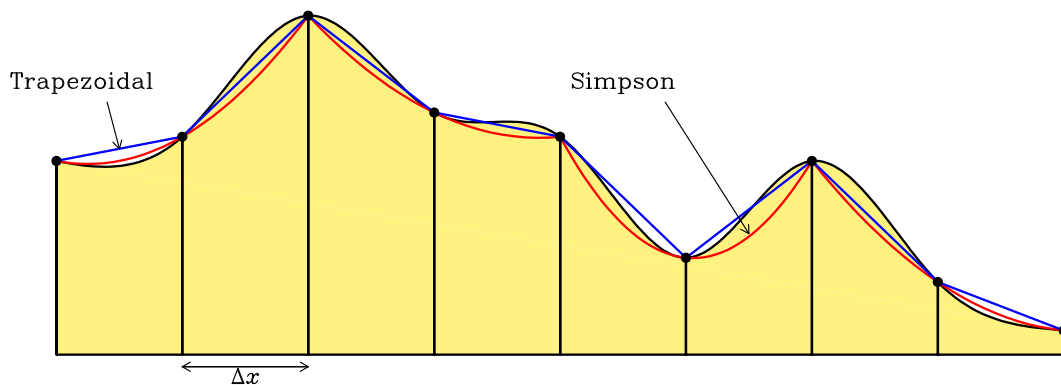
which has the effect of replacing the original function  $f(x)$  by the straight-line approximation, show in Figure 11.1. The error term then tends to zero as  $\Delta x^2$ , provided  $f$  is smooth enough to be twice differentiable. Another favorite is **Simpson's rule**; here

$$w_1 = w_N = \frac{1}{3} \Delta x, \quad w_2 = w_4 = \dots w_{N-1} = \frac{4}{3} \Delta x \tag{6}$$

$$w_3 = w_5 = \dots w_{N-2} = \frac{2}{3} \Delta x$$

and the number of samples  $N$  must be odd. Now instead of straight lines between sample points, the effective approximation is that of parabolic arcs (quadratic interpolation). The error decreases as  $\Delta x^4$  but the function must be four-times differentiable. If one relaxes the constraint the sample in  $x$  be even, one can increase the accuracy of the approximation considerably, with **Gaussian quadrature**. Here one looks for the highest degree polynomial approximation to the integral; this idea is dealt with in every text on numerical analysis. Less well known are the rules for integrating over surfaces and volume elements, often by building upon the Gaussian method. For these see Stroud, A. H., *Approximate Calculation of Multiple Integrals* Prentice-Hall Book Co. 1971.

Another idea is to say the model itself consists of a piece-wise constant or a piece-wise linear function, and to perform the integral analytically over the chosen region. This can be done for many 3-dimensional polyhedral shapes for potential fields (gravity and magnetism), and has the advantage of avoiding large errors that can arise when the observer is very near, or actually in contact with, the source material. See Blakely, R. J., *Potential Theory in Gravity and Magnetic Applications*, Cambridge University Press, New York, 1995, for lots of messy formulas, and



**Figure 11.1:** Trapezoidal and Simpson quadrature.

FORTRAN code too! This approach may be necessary when the representer is unbounded (as it is with gravity or magnetic measurements made on the surface of the source region) because then trapezoidal or Simpson quadrature can give infinite answers.

Let us try some of these ideas out on the 1-dimensional crustal magnetization problem. First consider the norm minimizing model. We are going to numerical approximations of the norm:

$$\|f\| = \left( \int_{-\infty}^{\infty} f(x)^2 dx \right)^{1/2} \quad (7)$$

and the corresponding inner product. The first thing that must go is the infinite interval of integration, which was a mathematical fiction anyhow. To represent the integral by a finite sum we will have at least two ways: (a) we can just truncate the interval to a finite one; (b) we could use a change of variable, that maps the real line onto some interval  $(a, b)$  and then we use (4)—a possible candidate might be  $x = h \tan \theta$  which sends the real line into  $(-\frac{1}{2}\pi, \frac{1}{2}\pi)$ . Here I will pursue (a). We will simply assume that the magnetization is confined between  $\pm 25$  km, adding 2.5 times the water depth  $h$  to each end of the interval containing the measurements. So now the model norm is found from

$$\|m\|^2 = m^T W m \quad (8)$$

where  $m \in \mathbb{R}^N$  a vector of evenly spaced samples of the model, and  $W \in \mathbb{R}^{N \times N}$  is the diagonal matrix:

$$W = \text{diag}(w_1, w_2, \dots, w_N) \quad (9)$$

and we choose the weights  $w_k$  according some integration rule, say Simpson. Then the statement that the model fits the data looks like this

$$G W m = d. \quad (10)$$

Here each row of  $G \in \mathbb{R}^{M \times N}$  is a vector of samples of the representer  $g_j$  and the  $N$  points  $\xi_k$ :

$$G_{jk} = g_j(\xi_k) = \frac{\mu_0 \Delta z}{2\pi} \frac{h^2 - (\xi_k - x_j)^2}{(h^2 + (\xi_k - x_j)^2)^2}, \quad j = 1, 2, \dots, M, \quad k = 1, 2, \dots, N. \quad (11)$$

The matrix  $W$  is doing the integration here. It is natural to choose  $N > M$ , more model points than data, and so we have an underdetermined LS problem to solve. It is not quite in the form we considered earlier in sections 5 and 6, but that can easily be fixed by a simple substitution. Suppose I introduce  $n \in \mathbb{R}^N$  as

$$m = W^{-1/2} n \quad (12)$$

where by  $W^{-1/2}$  I mean the diagonal matrix  $\text{diag}(w_1^{-1/2}, w_2^{-1/2}, \dots, w_N^{-1/2})$ . Then in terms of  $n$  (8) becomes

$$\|m\|^2 = (W^{-1/2}n)^T W (W^{-1/2}n) = n^T (W^{-1/2})^T W W^{-1/2}n \quad (13)$$

$$= n^T n. \quad (14)$$

So the norm of  $m$  in  $L_2$  becomes the Euclidean length of the vector  $n$ . And putting  $n$  into (10) gives

$$(GW^{1/2})n = d \quad (15)$$

This is an ordinary underdetermined LS problem for  $n$  with the matrix  $A = GW^{1/2} \in \mathbb{R}^{M \times N}$ . We can use QR on this for example if we need extra stability. If we use the normal equations, recall the solution 6(20):

$$n = A^T (A A^T)^{-1} d = (GW^{1/2})^T [GW^{1/2}(GW^{1/2})^T]^{-1} d \quad (16)$$

$$= W^{1/2} G^T [G W G^T]^{-1} d. \quad (17)$$

The matrix  $GWG^T \in \mathbb{R}^{N \times N}$  is the *numerical approximation of the Gram matrix*. So  $a = [G W G^T]^{-1} d$  is the approximation for the vector of coefficients  $\alpha$  in 10(4), for example, and you will easily verify that after we multiply through by  $W^{-1/2}$  as indicated in (12), we get

$$m = G^T [G W G^T]^{-1} d = G^T a \quad (18)$$

the equation which is the numerical equivalent to

$$m = \sum_{j=1}^M \alpha_j g_j \quad (19)$$

the standard solution from the recipe from Hilbert space.

The numerical results for the idealized magnetic profile are shown in Figure 11.2; here I have chosen the number of sample points  $N$  to be 101. Three solutions are plotted superimposed: the two numerical models using Simpson's rule and trapezoidal rule and the analytic result found by solving the system in  $L_2(-\infty, \infty)$ . A surprising result is that the analytic model and trapezoidal rule track almost exactly, and the Simpson rule solution is slightly different, visible as the slightly less wiggly curve in the magnified picture. The 2-norm of all of these solutions is 15.4; the distance between the Simpson solution and the exact one is 0.51 in the norm, while it is only 0.0079 between the trapezoidal solution and the exact result.

When we come to a seminorm minimization, one that could penalize the gradient for example, we can replace (8) by

$$\|Pm\|^2 = (Dm)^T W_1 Dm = m^T (D^T W_1 D) m \quad (20)$$

where  $D \in \mathbb{R}^{N-1 \times N}$  is the upper triangular matrix approximating the first difference:

$$D = \frac{1}{\Delta x} \begin{bmatrix} -1 & 1 & & & & \\ & -1 & 1 & & & \\ & & -1 & 1 & & \\ & & & \dots & & \\ & & & & -1 & 1 \end{bmatrix} \quad (21)$$

where entries that are not  $\pm 1$  are zeros. And, since  $D$  is not square,  $W_1$  is a weight matrix in  $\mathbb{R}^{N-1 \times N-1}$ . We cannot perform the same trick as in the norm minimization, because the matrix  $(D^T W_1 D)$  is not of full rank, (a constant vector is mapped to zero) and cannot therefore be inverted. Inversion is implied when one forms  $W^{-1/2}$  from  $W$ . So we have a form of LS problem that we did not meet earlier:

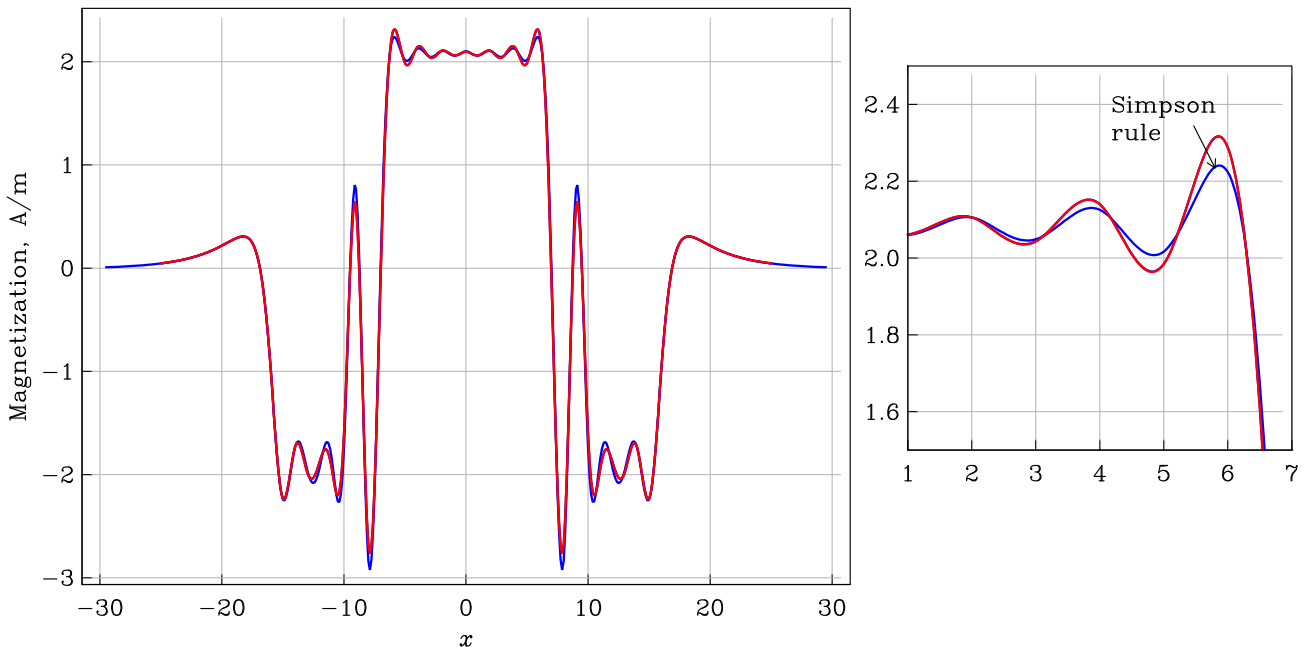
$$x_0 = \arg \min_{Ax=y} x^T Bx \quad (22)$$

where  $A \in \mathbb{R}^{M \times N}$ , with  $N > M$ ,  $B \in \mathbb{R}^{N \times N}$  and  $B$  is not necessarily of full rank. We can solve this with the introduction of  $M$  Lagrange multipliers, or equivalently a vector  $\mu \in \mathbb{R}^M$ . Then the unconstrained function is

$$u(x, \mu) = x^T Bx - \mu^T (Ax - y). \quad (23)$$

The usual differentiation leads to a linear system to be solved:

$$\begin{bmatrix} 2B & -A^T \\ A & O_M \end{bmatrix} \begin{pmatrix} x \\ \mu \end{pmatrix} = \begin{pmatrix} 0 \\ y \end{pmatrix}. \quad (24)$$



**Figure 11.2:** Numerical and analytic solutions.

This is a rather large matrix, but in our example  $B = D^T W_1 D$  is tridiagonal, and therefore mostly zeros, and the lower right matrix is all zeros, so sparse techniques are applicable to speed up the solution if the  $M$  and  $N$  get large. The relationship between this solution and the one found in Hilbert space in GIT pp 74-78 is not at all obvious. I will not get into the correspondence, because we will hardly ever use (24) in future.

Now that we have machinery for solving more realistic problems, let us return to the real magnetic anomaly profile taken near the seafloor at the Juan de Fuca Rise, plotted in Figure 8.1 of the notes. The forward problem is solved with 8(8-9), which I will repeat here for convenience, but you will need to look at Figure 8.2 also.

$$d_j = \int g(\mathbf{r}_j, x) m(x) dx = \sum_k w_k g(\mathbf{r}_j, x_k) m(x_k) \quad (25)$$

where  $w_k$  are the quadrature weights, and the representer is

$$g(\mathbf{r}_j, x) = \frac{\mu_0}{2\pi} \left[ \frac{\hat{\mathbf{z}} \cdot (\mathbf{r}_j - \mathbf{s}_+(x))}{|\mathbf{r}_j - \mathbf{s}_+(x)|^2} - \frac{\hat{\mathbf{z}} \cdot (\mathbf{r}_j - \mathbf{s}_-(x))}{|\mathbf{r}_j - \mathbf{s}_-(x)|^2} \right]. \quad (26)$$

Notice I have inserted the approximation that  $\hat{\mathbf{M}}_0 = \hat{\mathbf{B}}_0 = \hat{\mathbf{z}}$ , so that  $\hat{\mathbf{I}} = \hat{\mathbf{z}}$ . We form the matrix  $G \in \mathbb{R}^{M \times N}$ , whose rows are the representers:

$$G_{jk} = g(\mathbf{r}_j, x_k), \quad j = 1, 2, \dots, M, \quad k = 1, 2, \dots, N. \quad (27)$$

As before I am going to sample evenly, and it is important to have more model points than data. Also we cannot go to infinity, in  $x$ , and as before I take a modest extension of 0.5 km, a small multiple of the observation height, at each end. Below I give you a code fragment of MATLAB that generates the matrix  $G$ . On entry to this part of the code, there are vectors  $\mathbf{x}_0, \mathbf{z}_0 \in \mathbb{R}^M$  containing the observer coordinates, and  $\mathbf{x}_b, \mathbf{z}_b \in \mathbb{R}^N$  holding coordinates of the basement topography. What I want to illustrate is how the MATLAB function `meshgrid` makes it possible to prepare a matrix like  $G$  in a way that almost exactly mirrors the algebra: compare the code with (26). The only somewhat odd feature is that the arrays like  $\mathbf{z}_0$  etc have a *row* dimension inherited from the *second* vector in the argument, where I would expect the row dimension to be defined by the first vector.

```
[Zo Zb] = meshgrid(z0, zb);
[Xo Xb] = meshgrid(x0, xb);

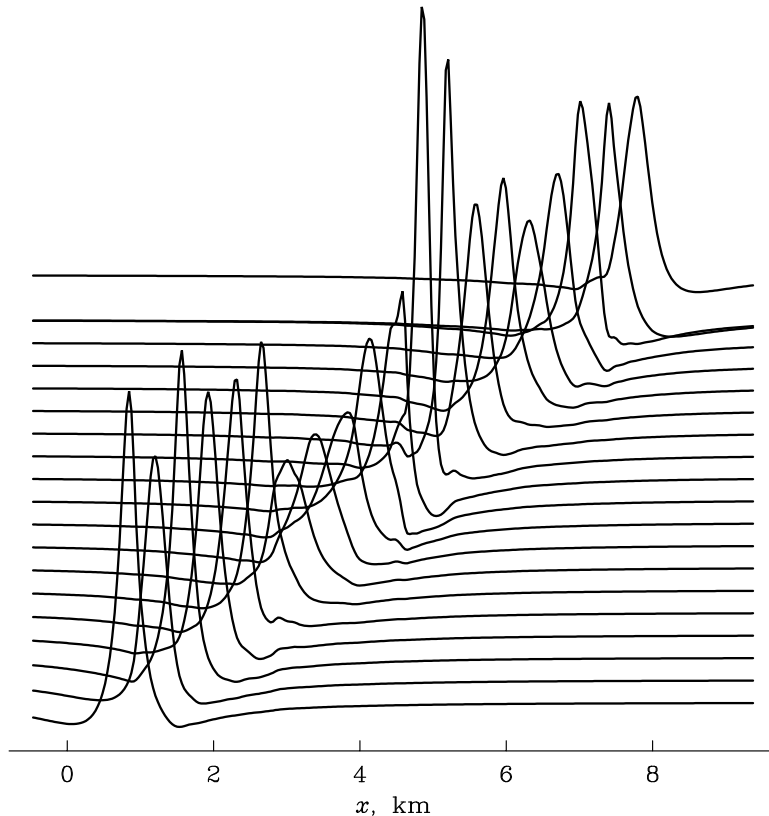
G1 = (Zo - Zb) ./ ((Zo-Zb).^2 + (Xo-Xb).^2);
Zb = Zb - dz;
G2 = (Zo - Zb) ./ ((Zo-Zb).^2 + (Xo-Xb).^2);
G = (mu0/(2*pi)) * (G1 - G2);
```

So now  $G$  has columns of representers, which is convenient if one wishes to plot them. What do the representers for the realistic problem look like?

In Figure 11.3 I plot every fifth one. Notice they resemble the simple  $g_j$  representers of the original  $L_2$  problem, but with variable amplitudes and irregular shapes. To speed execution it would be reasonable to set the small-amplitude portions to zero thus making  $G$  sparse.

Next we calculate the smallest  $m$  in  $L_2$ . I use the trapezoidal rule for simplicity, and for stability apply QR to equation (15). The horrible result appears in Figure 11.4. This is the smallest norm model, yet the magnetic intensities are *two orders of magnitude larger* than those typically found in marine basalts! What has gone wrong? This inverse problem was based on real field data, not artificially generated exact numbers. We see here the effect of instability in the presence of noise in the measurements—quite small errors have been amplified grotesquely, even though are we attempting to find small solutions. It is time to lift the artificial demand that the data must be fitted exactly.

**Figure 11.3:** Representers from equation (26) plotted with vertical offsets.



### References

Blakely, R. J., *Potential Theory in Gravity and Magnetic Applications*, Cambridge Univ. Press, New York, 1995.

Useful reference for classical geophysical treatment of potential fields, although the inverse theory presented is somewhat weak.

Parker, R. L., Calibration of the pass-through magnetometer-I. Theory, *Geophys. J. Internat.*, 142, 371-83, 2000.

Shure, L., Parker, R. L., and G. E. Backus: Harmonic splines for geomagnetic modeling, in *Phys. Earth Planet. Inter.*, 28, 215-29, 1982.

Stroud, A. H., *Approximate Calculation of Multiple Integrals* Prentice-Hall Book Co. 1971.

Still the classic book on integration over plane and solid regions.

**Figure 11.4:** Minimum 2-norm model fitting near-bottom magnetic profile exactly.

