16. NNLS and BVLS

Linear programming is able to solve linear inverse problems, with inequality constraints imposed, provided we can tolerate a different norm for measuring the misfit between model predictions and observations. While the sup norm $\|\cdot\|_\infty$ and $\|\cdot\|_1$ can both be treated in the LP setting, only the latter is really useful, because relying on the sup norm gives too much leverage to the noisiest data, clearly an inadvisable state of affairs. To exploit a quadratic measure of misfit, like the familiar

$$X[m]^2 = \sum_{j=1}^{M} \frac{(d_j - L_j[m])^2}{\sigma_j^2} \tag{1}$$

requires us to consider **Quadratic Programming**. Those QP problems which are convex are much easier to solve reliably than the non-convex ones, and I shall concentrate on two special convex cases that I have found invaluable in my scientific career. First, is the special QP problem called **Non-Negative Least Squares** (NNLS). This is simply the regular least-squares problem with the positivity constraint attached:

$$x_* = \arg \min_{x \geq 0} \| Ax - y \|_2 \tag{2}$$

where this notation, used extensively in the optimization literature, means that x_* is the vector x that achieves the minimum value (assuming it is unique); here $x, x_* \in \mathbb{R}^N$, $y \in \mathbb{R}^M$, and $A \in \mathbb{R}^{M \times N}$. In (2) there is no restriction on the relative sizes of M and N; the problem is interesting and nontrivial whether $N > M$, which is the natural geophysical situation, or not. There is a general property of optimization problems, somewhat similar to the Fundamental Theorem of LP called the Kuhn-Tucker conditions, which in NNLS leads to an interesting and valuable result: a solution to (2) exists in which no more than $M-1$ components of x_* are positive (and so by implication, at least $N-M+1$ of them must be zero, when $M < N$). When the dimension of the model space is large, as it will be in a discretized version of a continuous problem, this means the same thing as it did in the LP examples, that the solution vector x is mostly zeroes, with a few positive spikes, delta functions in the limit of a continuum. In many practical problems we find that the number of positive elements in the solution vector, while it can be as large s $M-1$, often much smaller than M.

Let me give a little graphical demonstration of the ideas as illustrated in Figure 16.1, which I will explain. First, consider the domain of the solution set in R^N, the set $x \geq 0$, called the **positive orthant**. It is a convex region in the space, whose edges comprise the positive extensions of all the positive unit vectors. In three dimensions, the edges of the positive orthant (called the positive octant in $\mathbb{R}^3$) are the positive x, y, and z axes. Now consider mapping the positive orthant into R^M with the linear map A, where we will assume that $M < N$. The image of the orthant will will be a fan-shaped region as shown in the Figure, where $M = 2$ and

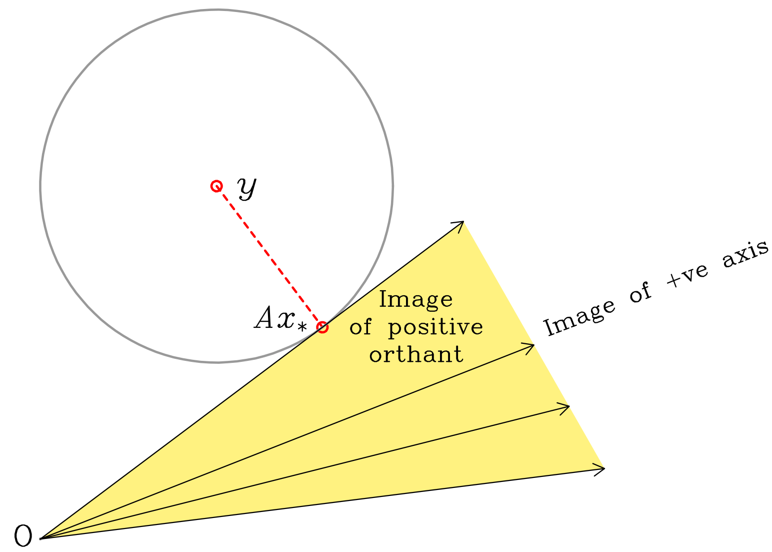


Figure 16.1: Image of the positive orthant and its edges under a linear map A taking $\mathbb{R}^4$ into $\mathbb{R}^2$. The point Ax_* is the closest point in the image set to y.

$N = 4$. The boundary of the image of the orthant comprises surfaces in R^M, and a surface in R^M locally is of dimension one less: $M - 1$; most of the positive axes are mapped into the interior of the image, the shaded region. Now consider the minimization problem (2). If an exact fit cannot be found, the point y lies outside the image. Then the smallest distance is achieved at a point in the boundary of the image, as shown, which is Ax_*. In the Figure, where $M = 2$ it is clear the point x_* falls *on the image of one of the positive axes* in $\mathbb{R}^N$; in higher dimensions x_* lies in the subspace spanned by no more than $M - 1$ images of the positive axes, because the others are not in the boundary in R_M. If several of the positive axes are mapped into a single line in the range space, we can always choose just one of them. This illustrates the assertion that the norm minimizer in (2) need have no more than $M - 1$ positive components.

The solution to the NNLS problem can be extended like the LP problem to include nonpositive unknowns. One can add linear equality constraints on the unknowns by including them in the A matrix with large positive weights, so that the minimization process essentially satisfies those rows exactly and, as before, if the weight is large enough, the contribution from those rows to the misfit budget is negligible.

In MATLAB an algorithm for NNLS is provided called `lsqnonneg`. It is very slow since it is written as a MATLAB M-file and not optimized as the linear system solution is or the QR and eigenvalue codes are. There is a bug in the MATLAB code concerning the "warm start" feature—it doesn't work and generates wrong answers! For homework problems the

MATLAB code is fine, but any realistic problem you will need a Fortran program called `nnls.f` by Lawson and Hanson which you can get from me.

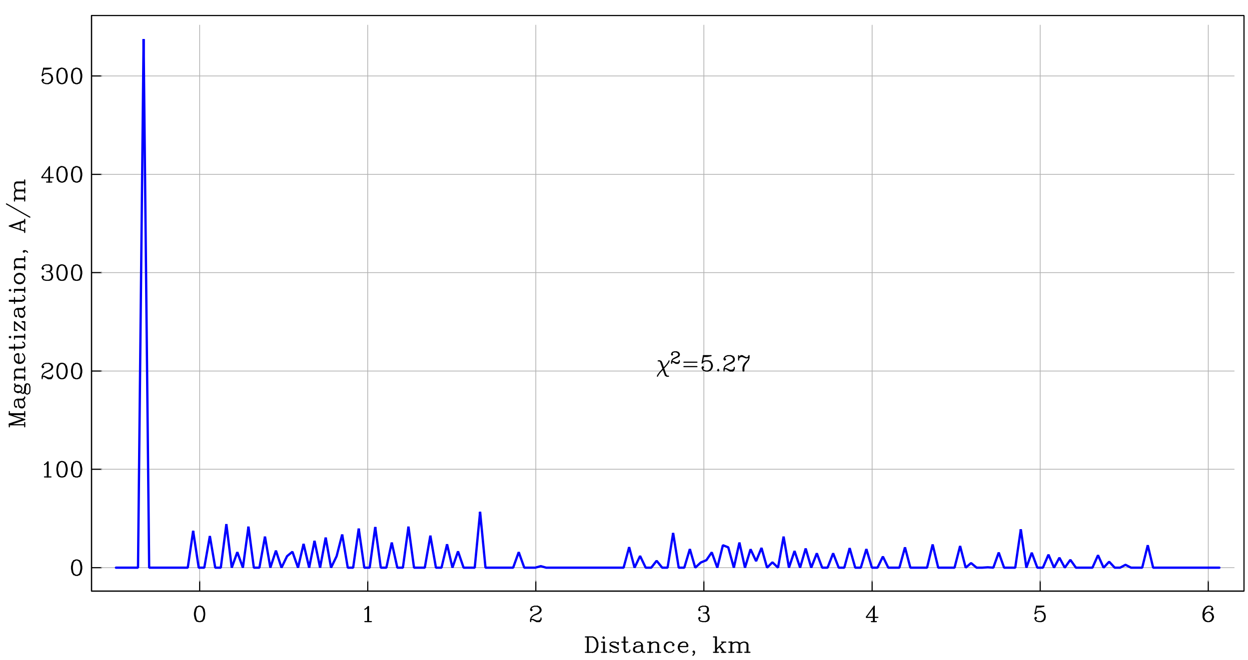
Suppose instead of merely demanding positivity of the unknown in (2) we wanted some (or all) of the components to lie between specified bounds:

$$l_j \le x_j \le u_j \tag{3}$$

This can be accomplished with NNLS; I leave this as an exercise for the student. However, the size of the problem to be solved is greatly increased and causes unnecessary waste of computer time, because there exists a specific Fortran program for this task called `bvls.f` which stands for **Bounded Variable Least-Squares**. Again I can provide you with the source should you ever need it.

As I mentioned in my discussion of ideal bodies, once we introduce inequality constraints even with a linear forward problem we face the possibility that solutions to the constrained optimization system may not exist at all; the data and the conditions may be inconsistent. In LP language this is a statement that there is no feasible solution. The NNLS and BVLS problems always have solutions, but are set up in a way to test whether the imposed conditions are consistent with observation. Returning to the near-bottom magnetic anomaly problem, we can now ask the question, "Are there any magnetization models that fit the data without going negative somewhere?" This we do by simply minimizing the data misfit with NNLS over the set of non-negative m:

Figure 16.2: Minimum misfit, all positive magnetization solution.



$$\min_{m \geq 0} \| \hat{d} - Bm \|_2 \tag{4}$$

If this value is larger than a reasonable tolerance T according to the χ^2 distribution, then we have demonstrated the need for a magnetic reversal in the section. Notice that we now don't specify a particular interval—we'll allow the model complete freedom. The Figure shows the results of this calculation. An amazingly small misfit can be achieved, but at the expense of some rather large amplitudes. We should not be too worried about the peak on the left, since it is in a zone not controlled by data since the magnetic field measurements begin at $x = 0$. We could use BVLS to check how large the positive amplitudes need to be. Instead, however, I will demonstrate an alternative, perhaps more familiar looking approach: regularization. We simply add to (4) a term penalizing the size of the model in the 2-norm:

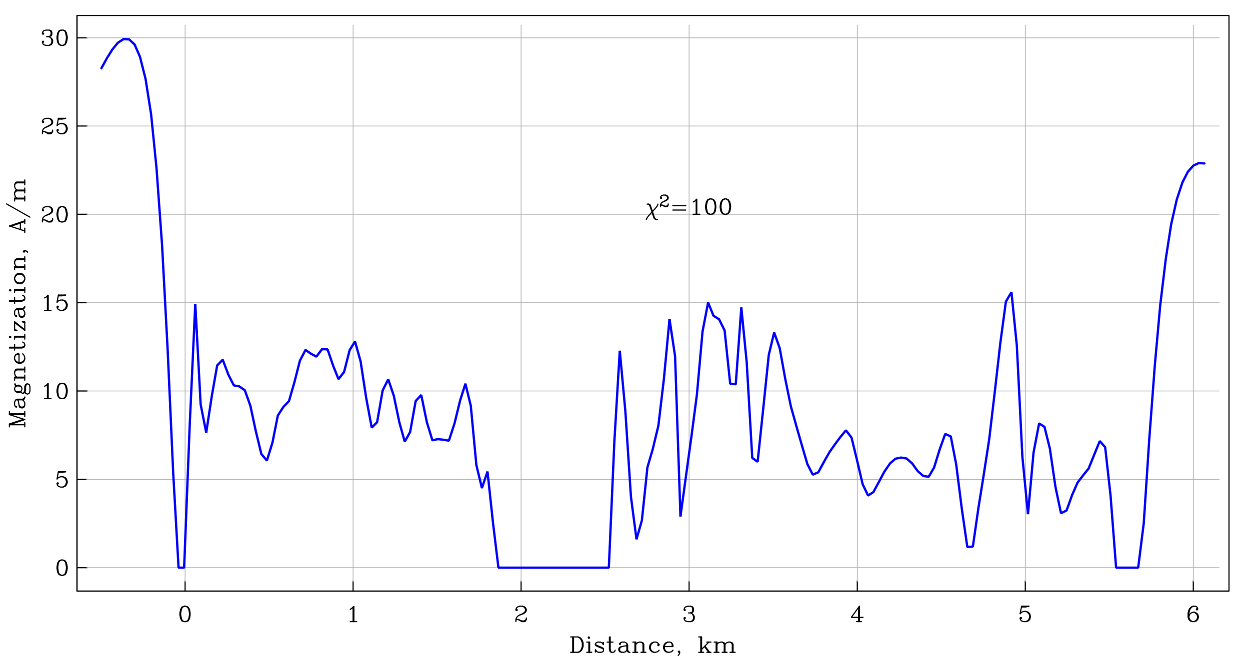
$$\min_{m \geq 0} \| \hat{d} - Bm \|^2 + w \| m \|^2 \tag{5}$$

where the 2-norm is implied, and $w > 0$ is a weight, which we can treat as a Lagrange multiplier. We rewrite (5) as

$$\min_{m \geq 0} \left\| \begin{bmatrix} B \\ w^{1/2} I \end{bmatrix} m - \begin{bmatrix} \hat{d} \\ 0 \end{bmatrix} \right\|^2 \tag{6}$$

which is exactly in the from of the NNLS problem. Now sweep though positive values of w: for small values we get misfits close to the one shown in Figure 16.2, but as w increases the misfit term increases as (5) pays more attention to the size of m. When the misfit reaches the expected tolerance, we inspect the solution, which is of course still everywhere non-

Figure 16.3: Regularized, all positive magnetization solution.



negative. The result is plotted in Figure 16.3, where the misfit has been allowed to rise to $\chi^2 = 100$ the expected value for 100 data. The model magnetizations are not extraordinary, and we must conclude that there are reasonable-looking solutions without negative segments: *the reversed magnetization is not demanded by these data.*