11.9 Least Squares With Equality Constraints

The idea underlying weighted least squares is that observations with large weight shall influence the solution more than observations with smaller weights. Extending this idea to the extreme implies that giving an observation infinite weight will lead to that this observation equation approximately will be fulfilled. In other words: we have a way of introducing linear constraints on the least-squares problem: ***Observation equations with large weights act as linear constraints***. This observation has been made by many geodesists.

Often constraints follow from a mathematical or physical model. In its simplest form certain prescribed coordinate values shall be left unchanged by the least-squares procedure. Much litterature has been written about all possible conditions between the number of observations, the number of constraints, the number of unknowns and the rank of the involved matrices. However we restrict ourselves to the ordinary case with more observations than unknowns, and with as many constraints as the rank of the matrix B describing the constraints.

Exact Solution

Our starting point is again the observation equations $A\boldsymbol{x} = \boldsymbol{b} - \boldsymbol{e}$ augmented with p constraints described through $B\boldsymbol{x} = \boldsymbol{d}$. We assume that the system $B\boldsymbol{x} = \boldsymbol{d}$ is consistent. The problem is

$$\min_{B\boldsymbol{x}=\boldsymbol{d}} \left\| A\boldsymbol{x} - \boldsymbol{b} \right\|^2. \tag{11.1}$$

Let B^{T} have the QR factorization:

$$Q^{\mathrm{T}} B^{\mathrm{T}} = \begin{array}{c} \begin{bmatrix} R \\ 0 \end{bmatrix} \end{array} \begin{array}{l} p \\ n-p \end{array}$$

and set

$$AQ = \underset{\quad p \quad n-p}{\begin{bmatrix} A_1 & A_2 \end{bmatrix}} \quad \text{and} \quad Q^{\mathrm{T}}\boldsymbol{x} = \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{z} \end{bmatrix} \begin{array}{l} p \\ n-p \end{array} .$$

We get $B = R^{\mathrm{T}} Q^{\mathrm{T}}$ and $B\boldsymbol{x} = R^{\mathrm{T}} Q^{\mathrm{T}} \boldsymbol{x} = R^{\mathrm{T}} \boldsymbol{y}$; and $(AQ)(Q^{\mathrm{T}}\boldsymbol{x}) = A\boldsymbol{x} = A_1\boldsymbol{y} + A_2\boldsymbol{z}$. With these transformations (11.1) becomes

$$\min_{R^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{d}} \left\| A_1\boldsymbol{y} + A_2\boldsymbol{z} - \boldsymbol{b} \right\|. \tag{11.2}$$

Hence $\boldsymbol{y}$ is determined from $R^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{d}$ and $\boldsymbol{z}$ is obtained by solving the unconstrained least-squares problem

$$\min_{\boldsymbol{z}} \left\| A_2\boldsymbol{z} - (\boldsymbol{b} - A_1\boldsymbol{y}) \right\|. \tag{11.3}$$

With known $\boldsymbol{y}$ and $\boldsymbol{z}$ the solution $\boldsymbol{x}$ becomes

$$\boldsymbol{x} = Q \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{z} \end{bmatrix}.$$

Alternatively we want to derive a solution using SVD. We are facing a new issue as we need to find a simultaneous SVD for both A and B. The problem is called a ***generalized singular value decomposition***.

We introduce the compound matrix

$$M = \begin{array}{c} \begin{bmatrix} A \\ B \end{bmatrix} \\ n \end{array} \begin{array}{c} m \\ p \\ \; \end{array} .$$

???? proved that there exist an m by m orthogonal matrix U, a p by p orthogonal matrix V, a (usually) square matrix X, and nonnegative diagonal matrices C and S so that

$$A = UCX^{\mathrm{T}} \tag{11.4}$$

$$B = VSX^{\mathrm{T}} \tag{11.5}$$

$$C^2 + S^2 = I. \tag{11.6}$$

The matrix X is n by s where $s = \min\{m + p, n\}$.

The diagonal entries are ordered as follows: $0 = C_{11} = \cdots = C_{qq} < C_{q+1,q+1} \leq \cdots \leq C_{pp} < C_{p+1,p+1} = \cdots = C_{nn} = 1$, and $S_{11} \geq \cdots \geq S_{pp} > 0$. For $i = p+1, \ldots, n$ $S_{ii} = 0$ and from (11.6) follows that $C_{ii} = 1$ for the same indices. The ***generalized singular values*** are C_{ii}/S_{ii}. In MATLAB the general call is [U, V, X, C, S] = gsvd(A,B). Here the nonzero entries of S are always on its main diagonal. If $m \geq n$ the nonzero entries of C are also on its main diagonal. But if $m < n$, the nonzero diagonal of C is diag(C, n−m). This allows the diagonal entries to be ordered so that the generalized singular values are nondecreasing. The number of singular values of X equals the nonzero singular values of M. We apply this result to problem (11.1)

$$\min_{x} \|A\boldsymbol{x} - \boldsymbol{b}\| \quad \text{subject to} \quad B\boldsymbol{x} = \boldsymbol{d}.$$

Multiplying by an orthogonal matrix U inside the norm sign does not change the length; the constraints are multiplied by V^{T}:

$$\min_{x} \|U^{\mathrm{T}}A\boldsymbol{x} - U^{\mathrm{T}}\boldsymbol{b}\| \quad \text{subject to} \quad V^{\mathrm{T}}B\boldsymbol{x} = V^{\mathrm{T}}\boldsymbol{d}.$$

From (11.4) and (11.5) follows $U^{\mathrm{T}}A = CX^{\mathrm{T}}$ and $V^{\mathrm{T}}B = SX^{\mathrm{T}}$ and we get

$$\min_{x} \|CX^{\mathrm{T}}\boldsymbol{x} - U^{\mathrm{T}}\boldsymbol{b}\| \quad \text{subject to} \quad SX^{\mathrm{T}}\boldsymbol{x} = V^{\mathrm{T}}\boldsymbol{d}.$$

Introducing a column oriented notation of matrices $U = [\,\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m\,]$, $V = [\,\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\,]$, and $X = [\,\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\,]$ we obtain

$$\left(X^{\mathrm{T}}\boldsymbol{x}\right)_i = \begin{cases} \dfrac{\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{d}}{S_{ii}} & \text{for } i = 1, \ldots, p \\[2ex] \dfrac{\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{b}}{C_{ii}} & \text{for } i = p+1, \ldots, n. \end{cases}$$

Now let $\left(X^{\mathrm{T}}\right)^{-1} = W = [\, \boldsymbol{w}_1, \ldots, \boldsymbol{w}_n \,]$ then the solution $\boldsymbol{x}_c$ to (11.1) is

$$\boldsymbol{x}_c = \sum_{i=1}^{p} \frac{\boldsymbol{v}_i^{\mathrm{T}} \boldsymbol{d}}{S_{ii}} \boldsymbol{w}_i + \sum_{i=p+1}^{n} \frac{\boldsymbol{u}_i^{\mathrm{T}} \boldsymbol{b}}{C_{ii}} \boldsymbol{w}_i. \tag{11.7}$$

For reasons of comparison we quote the ordinary solution $\boldsymbol{x}$ without constraints

$$\boldsymbol{x} = \sum_{i=q+1}^{p} \frac{\boldsymbol{u}_i^{\mathrm{T}} \boldsymbol{b}}{C_{ii}} \boldsymbol{w}_i + \sum_{i=p+1}^{n} \boldsymbol{u}_i^{\mathrm{T}} \boldsymbol{b} \boldsymbol{w}_i. \tag{11.8}$$

It is possible to compute how much the constraints increase the residuals. The constrained minus unconstrained residuals are

$$\boldsymbol{e}_c - \boldsymbol{e} = A(\boldsymbol{x} - \boldsymbol{x}_c).$$

Remember that $A\boldsymbol{x}_i = C_{ii} \boldsymbol{u}_i$ for $i = 1, \ldots, n$, then by using (11.7) and (11.8) we get

$$\|\boldsymbol{e}_c\|^2 = \|\boldsymbol{e}\|^2 + \sum_{i=q+1}^{p} \left(\boldsymbol{u}_i^{\mathrm{T}} \boldsymbol{b} - \frac{C_{ii}}{S_{ii}} \boldsymbol{v}_i^{\mathrm{T}} \boldsymbol{d} \right)^2.$$

Both the QR decomposition and the generalized singular value decomposition solution are coded in the M-file clsq.

Approximate Solution

Next we solve the unconstrained least-squares problem

$$\min_x \left\| \begin{bmatrix} A \\ \lambda B \end{bmatrix} \boldsymbol{x} - \begin{bmatrix} \boldsymbol{b} \\ \lambda \boldsymbol{d} \end{bmatrix} \right\|$$

for large λ. We find the solution in terms of the GSVD and reuse some of the above notation. We start by forming the normal equations

$$(A^{\mathrm{T}} A + \lambda^2 B^{\mathrm{T}} B)\boldsymbol{x} = A^{\mathrm{T}} \boldsymbol{b} + \lambda^2 B^{\mathrm{T}} \boldsymbol{d}$$

and substitute (11.4) and (11.5):

$$(X C^{\mathrm{T}} U^{\mathrm{T}} U C X^{\mathrm{T}} + \lambda^2 X S^{\mathrm{T}} V^{\mathrm{T}} V S X^{\mathrm{T}})\boldsymbol{x} = X C^{\mathrm{T}} U^{\mathrm{T}} \boldsymbol{b} + \lambda^2 X S^{\mathrm{T}} V^{\mathrm{T}} \boldsymbol{d}$$

or

$$\begin{aligned} (X C^{\mathrm{T}} C X^{\mathrm{T}} + \lambda^2 X S^{\mathrm{T}} S X^{\mathrm{T}})\boldsymbol{x} &= X C^{\mathrm{T}} U^{\mathrm{T}} \boldsymbol{b} + \lambda^2 X S^{\mathrm{T}} V^{\mathrm{T}} \boldsymbol{d} \\ (C^{\mathrm{T}} C + \lambda^2 S^{\mathrm{T}} S) X^{\mathrm{T}} \boldsymbol{x} &= C^{\mathrm{T}} U^{\mathrm{T}} \boldsymbol{b} + \lambda^2 S^{\mathrm{T}} V^{\mathrm{T}} \boldsymbol{d}. \end{aligned}$$

The solution is

$$\boldsymbol{x}(\lambda) = \sum_{i=1}^{p} \frac{C_{ii} \boldsymbol{u}_i^{\mathrm{T}} \boldsymbol{b} + \lambda^2 S_{ii} \boldsymbol{v}_i^{\mathrm{T}} \boldsymbol{d}}{C_{ii}^2 + \lambda^2 S_{ii}^2} \boldsymbol{w}_i + \sum_{i=p+1}^{n} \frac{\boldsymbol{u}_i^{\mathrm{T}} \boldsymbol{b}}{C_{ii}} \boldsymbol{w}_i. \tag{11.9}$$

Subtracting (11.7) from (11.9) yields

$$\boldsymbol{x}(\lambda) - \boldsymbol{x}_c = \sum_{i=1}^{p} \frac{C_{ii}(S_{ii}\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{b} - C_{ii}\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{d})}{S_{ii}(C_{ii}^2 + \lambda^2 S_{ii}^2)}\boldsymbol{w}_i. \tag{11.10}$$

For $\lambda \to \infty$ we have $\boldsymbol{x}(\lambda) \to \boldsymbol{x}_c$. The appeal of the method of weighting is that it does not call for special subroutines: an ordinary least-squares solver will do.

We have implemented the procedure as M-file wlsq.

The present presentation relies heavily on Golub & van Loan and Åke Björck.