Datum Definition and Minimal Constraints

In the adjustment of geodetic or photogrammetric networks, one expresses the relationship between the observations and the point coordinate parameters by condition equations, for example the angle or the distance condition equations in geodetic applications and the collinearity equation in photogrammetric applications. These equations all contribute *relative* information rather that *absolute* information about the point positions. The usual procedure to introduce such absolute information is by constraining, i.e. fixing, certain point coordinate components. These points are referred to as *control points*. Without such control points, or other constraints, the system of normal equations would be *rank deficient* and hence not uniquely solvable. The rank deficiency is just equal to the minimum number of such constraints which would be needed to bring the system to *full rank*. In the case of a horizontal network with only angle observations, the rank deficiency is four. In the case of a horizontal network which includes at least one distance observation, the rank deficiency is three. In the case of a photogrammetric network, the rank deficiency is seven. These rank deficiencies are referred to as the *datum defect*, since the presence of the necessary control points would define the datum. Of course, there may be other causes of rank deficiency such as insufficient observations to define a point. These are another matter altogether, and are referred to as *configuration defects*. These should not occur with careful network planning and they will not be discussed further here.

If one introduces just enough constraint equations to satisfy the datum defect, then these are known as *minimal constraints*. Different sets of minimal constraints have the interesting property that, although the point coordinate estimates may vary, the observation residuals are invariant. Thus for residual analysis only, any minimal set of control points is as good as another, however in practice the choice of control points is very important so that the new network would be consistent with existing networks at shared points. More constraints than the minimal number will often be used in practice to counteract weak network geometry, or to satisfy particular requirements to match existing point coordinates. If fewer than the minimal number of constraints are used, then the system of normal equations remains rank deficient and we say that the point coordinates are not *estimable*. Even if the individual point coordinates are not estimable, functions of them may be, for example angles or distances.

For the purpose of illustration, let us take the geodetic, horizontal triangle network in Figure 1 as an example. If only the three angle observations shown are made and no control points are introduced, then the resulting normal equations, of size six by six, have rank of two and a rank deficiency of four. The fixing of two control points, or four point coordinates, would resolve this datum defect. These constraints can be implemented very simply by elimination, in effect just replacing the unknowns with numerical constants. However for generality we assume that the constraints are implemented by the general method of bordering the normal equations as in,

$$\begin{bmatrix} \mathbf{0} & C \\ C^T & N \end{bmatrix} \begin{bmatrix} \lambda \\ \Delta \end{bmatrix} = \begin{bmatrix} g \\ t \end{bmatrix} \tag{1}$$

The constraint matrix corresponding to fixing the first two points would be

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \tag{2}$$

Inner Constraints or Free Net Adjustment

If a square, symmetric matrix, such as N above, is of full rank then all of its eigenvalues are nonzero, and its eigenvectors form an orthogonal basis for the row space. If it is not of full rank, say is has order u, rank h, and therefore defect u-h. Such a matrix will have u-h zero eigenvalues. The h eigenvectors associated with the nonzero eigenvalues will form an orthogonal basis for the row space, and the u-h eigenvectors associated with the zero eigenvalues will form an orthogonal basis for the null space. In Figure 2, we see, schematically, the locus of solutions to the rank deficient equations, the null space, and the intersection point. Because of the favorable geometry, this intersection point will have both minimum magnitude and minimum variance compared to other restrictions or constraints. To achieve this solution, we may use the eigenvectors associated with the zero eigenvalues as coefficients in the constraint equations. This strategy will have the following characteristics:

(1) it will resolve the rank deficiency from the datum defect and it will therefore permit a unique solution to the system of equations, and

(2) of all the possible solutions to the rank deficient system, it will select the one with minimum magnitude and minimum variance.

This solution is known in the geodetic and photogrammetric literature as the *inner constraint* solution, or sometimes as the *free net solution.*

Now, having said all of the above, it can be stated that no one uses this strategy exactly as presented. This presentation is useful to understand the geometry of the problem but there are easier ways to construct the needed constraint matrix. The above mentioned eigenvectors provide a basis of the null space of the rank deficient matrix. But, as with any vector space, there are many (an infinite number of) such bases. One particular basis can be written directly and is therefore the one that is most often used. We will describe this one, show that it has the required properties, and then give a geometric derivation which provides much insight into the meaning of inner constraints. For the horizontal 2D network in Figure 1, with no distance observations, the following constraint matrix will have the same effect as the (harder to compute) eigenvectors.

$$C = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ Y_i^o & -X_i^o & Y_j^o & -X_j^o & Y_k^o & -X_k^o \\ X_i^o & Y_i^o & X_j^o & Y_j^o & X_k^o & Y_k^o \end{bmatrix} \tag{3}$$

If distance observations were present then the datum defect would be one less, and the fourth row of the matrix in Equation 3 would not be needed. For 3D networks, as usually found in photogrammetry, the datum defect resulting from no control point information is seven. Therefore one needs seven constraint equations to resolve the defect. Using this approach the required inner constraint matrix can be written directly,

$$C = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 1 & 0 & 0 & 1 & \cdots \\ 0 & Z_i^o & -Y_i^0 & 0 & Z_j^o & -Y_j^o & \cdots \\ -Z_i^o & 0 & X_i^o & -Z_j^o & 0 & X_j^o & \cdots \\ Y_i^o & -X_i^o & 0 & Y_j^o & -X_j^o & 0 & \cdots \\ X_i^o & Y_i^o & Z_i^o & X_j^o & Y_j^o & Z_j^o & \cdots \end{bmatrix} \tag{4}$$

If distance observations were present then the datum defect would be only six, and the last row in the matrix of Equation 4 could be eliminated. The X,Y,Z terms in Equations 3 and 4 represent the current values of the approximations of the point coordinate parameters, as the solution proceeds to iterative convergence.

In order to demonstrate the plausibility of the above statements, it will be shown that the rows of the matrix in Equation 3 are orthogonal to the coefficients of the angle condition equation. This condition equation represents the one most widely used in 2D, horizontal triangulation networks. For the clockwise angle at point i, in Figure 1, from point j to point k, the following row vector represents the coefficients of the linearized angle condition equation. For reference see Mikhail, 1980,

$$b = \begin{bmatrix} \frac{\partial F_\theta}{\partial X_i} & \frac{\partial F_\theta}{\partial Y_i} & \frac{\partial F_\theta}{\partial X_j} & \frac{\partial F_\theta}{\partial Y_j} & \frac{\partial F_\theta}{\partial X_k} & \frac{\partial F_\theta}{\partial Y_k} \end{bmatrix} \tag{5}$$

$$b=\left[\frac{Y_k^o-Y_i^o}{(S_{ik}^o)^2}-\frac{Y_j^o-Y_i^o}{(S_{ij}^o)^2},-\frac{X_k^o-X_i^o}{(S_{ik}^o)^2}+\frac{X_j^o-X_i^o}{(S_{ij}^o)^2},\frac{Y_j^o-Y_i^o}{(S_{ij}^o)^2},-\frac{X_j^o-X_i^o}{(S_{ij}^o)^2},-\frac{Y_k^o-Y_i^o}{(S_{ik}^o)^2},\frac{X_k^o-X_i^o}{(S_{ik}^o)^2}\right] \tag{6}$$

If one takes the inner product of **b** with the rows of **C** the result is a vector of zeros. In other words the rows of **C** are orthogonal to **b**.

$$bC^T = [0 \;\; 0 \;\; 0 \;\; 0] \tag{7}$$

and,

$$NC^T = B^T W B C^T = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{8}$$

where **B** in the last equation represents the full matrix of condition equation coefficients. Thus we see that the rows of **C** are orthogonal to the rows of **N**, and therefore **C** can serve as a set of inner constraints for the matrix **N**. A similar demonstration can be made with the **C** of Equation 4 and the collinearity equations as commonly used in 3D networks in photogrammetry.

The following development is based upon Leick, 1982. It shows the geometrical meaning of using the inner constraints just described. The development will be summarized for the 2D case and a simple extension can be made for the 3D case. Consider a similarity transformation (four parameter) between the adjusted coordinates, $\mathbf{X}_a$, and the approximate coordinates, $\mathbf{X}_o$,

$$X_a = T + (1+k)R_\alpha X_o \tag{9}$$

in which **T** is the translation vector, (1+k) is the scale factor, and **R** is the rotation matrix of a small angle. Written out,

$$\begin{bmatrix} x_a \\ y_a \end{bmatrix} = \begin{bmatrix} t_x \\ t_y \end{bmatrix} + (1+k)\begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix}\begin{bmatrix} x_o \\ y_o \end{bmatrix} \tag{10}$$

Assuming a small angle, and near unity scale factor, and assuming that products of small quantities may be disregarded, we obtain,

$$\begin{bmatrix} x_a \\ y_a \end{bmatrix} = \begin{bmatrix} t_x \\ t_y \end{bmatrix} + \begin{bmatrix} x_o \\ y_o \end{bmatrix} + \alpha \begin{bmatrix} y_o \\ -x_o \end{bmatrix} + k \begin{bmatrix} x_o \\ y_o \end{bmatrix} \tag{11}$$

Since this represents a step in the iterative solution,

$$\begin{bmatrix} x_a \\ y_a \end{bmatrix} = \begin{bmatrix} x_o + dx \\ y_o + dy \end{bmatrix} \tag{12}$$

we can combine the last two equations to obtain,

$$\begin{bmatrix} dx \\ dy \end{bmatrix} = \begin{bmatrix} t_x \\ t_y \end{bmatrix} + \alpha \begin{bmatrix} y_o \\ -x_o \end{bmatrix} + k \begin{bmatrix} x_o \\ y_o \end{bmatrix} \tag{13}$$

rearranging,

$$\begin{bmatrix} dx \\ dy \end{bmatrix} = \begin{bmatrix} 1 & 0 & y_o & x_o \\ 0 & 1 & -x_o & y_o \end{bmatrix} \begin{bmatrix} t_x \\ t_y \\ \alpha \\ k \end{bmatrix} \tag{14}$$

We write these equations for every point in the network,

$$\begin{bmatrix} 1 & 0 & 1 & 0 & \cdots & 1 & 0 \\ 0 & 1 & 0 & 1 & \cdots & 0 & 1 \\ y_{10} & -x_{10} & y_{20} & -x_{20} & \cdots & y_{n0} & -x_{n0} \\ x_{10} & y_{10} & x_{20} & y_{20} & \cdots & x_{n0} & y_{n0} \end{bmatrix} \begin{bmatrix} dx_1 \\ dy_1 \\ dx_2 \\ dy_2 \\ \vdots \\ dx_n \\ dy_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{20}$$

The coefficient matrix here, it will be noticed, is identical in form to that of Equation 3. It was shown previously that this set of equations, considered as constraint equations, will form a basis for the null space of the datum deficient normal equations, and hence can be used as the inner constraint matrix. Thus the geometric interpretation of the inner constraint solution is that when advancing from one iteration to the next, there will be no net shift, rotation, or scale change between the approximate and the refined coordinate positions. Thus rather than arbitrarily fixing two points (four coordinate components) out of many, one fixes four geometric relationships. All points then play equal roles in "connecting" ~~to~~ the network to the coordinate system.

This can have dramatic effects upon the *a posteriori* confidence ellipses of the network points. If a point is fixed then its confidence ellipse vanishes. For a three point network, after fixing two points, all of the error is cast into the uncertainty of the third point. With the inner constraint or free net solution, however, each point has a finite confidence ellipse which reflects its strength of determination in the network. Furthermore, as was noted earlier, the inner constraint solution yields a minimum variance solution, and the trace of the variance-covariance matrix (the sum of the variances) will be a minimum among all constrained solutions. It should be emphasized that this property is most useful in network analysis and pre-analysis. In actual practice, the network must eventually be constrained at fixed points, in order to have the best consistency with the existing control points. The simple three point network shown in Figure 1 is again depicted in Figures 3 and 4 with confidence or error ellipses derived from two point constraints and inner constraints respectively. Of course everything which has been described here relating to two-dimensional networks, has an equivalent and obvious expression in three-dimensional networks, as found in photogrammetry. The usual three-dimensional inner constraint matrix has been shown in Equation 4.

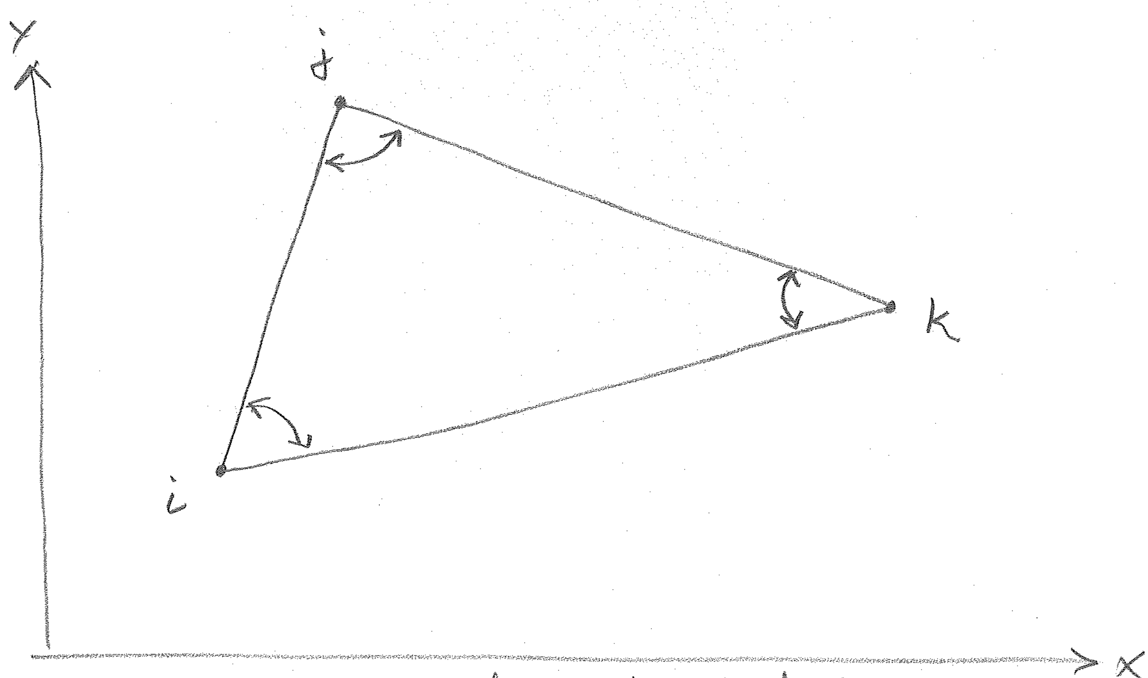

Figure 1. Triangle Network with only Angle Observations

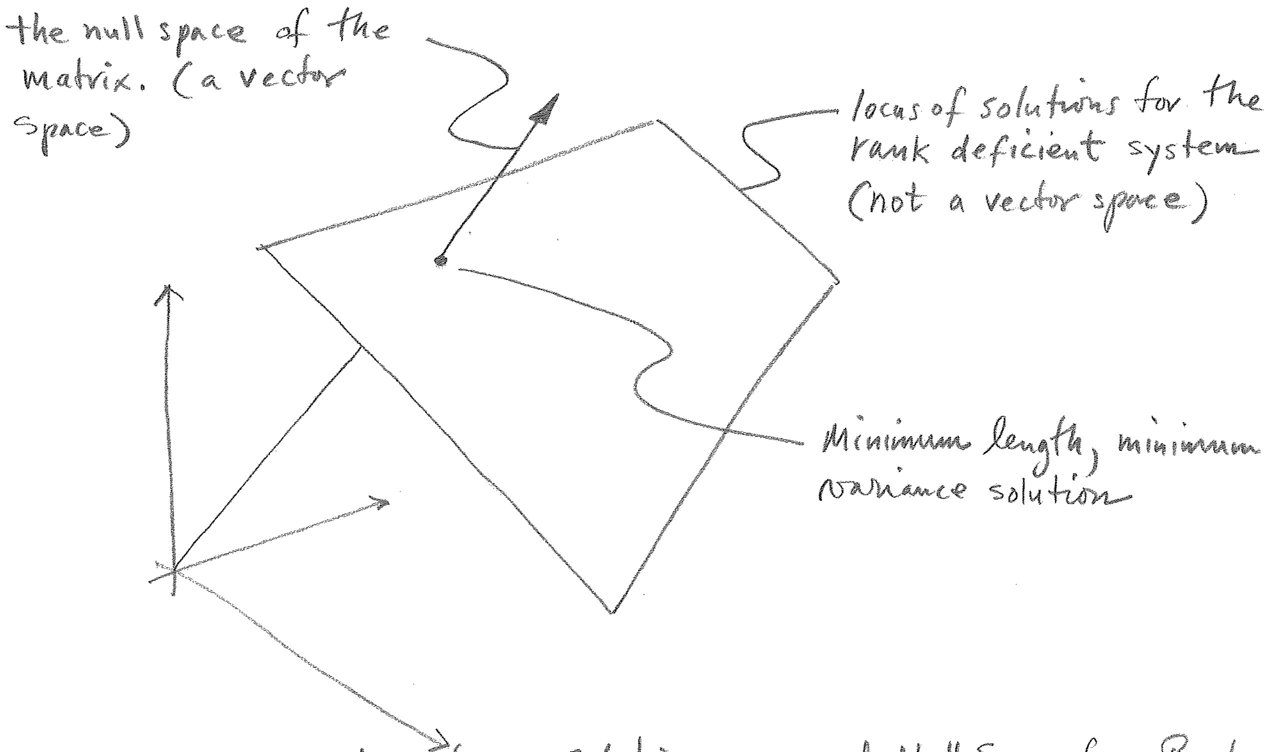

Figure 2. Intersection of Solution space and Null Space for a Rank Deficient System

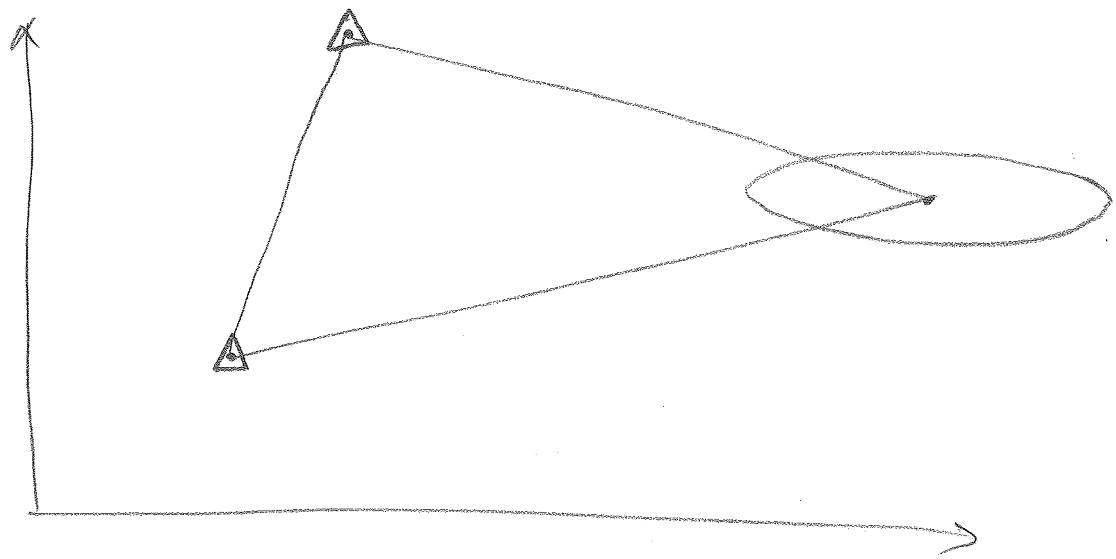

Figure 3. Error Ellipse with Fixed Point Constraints

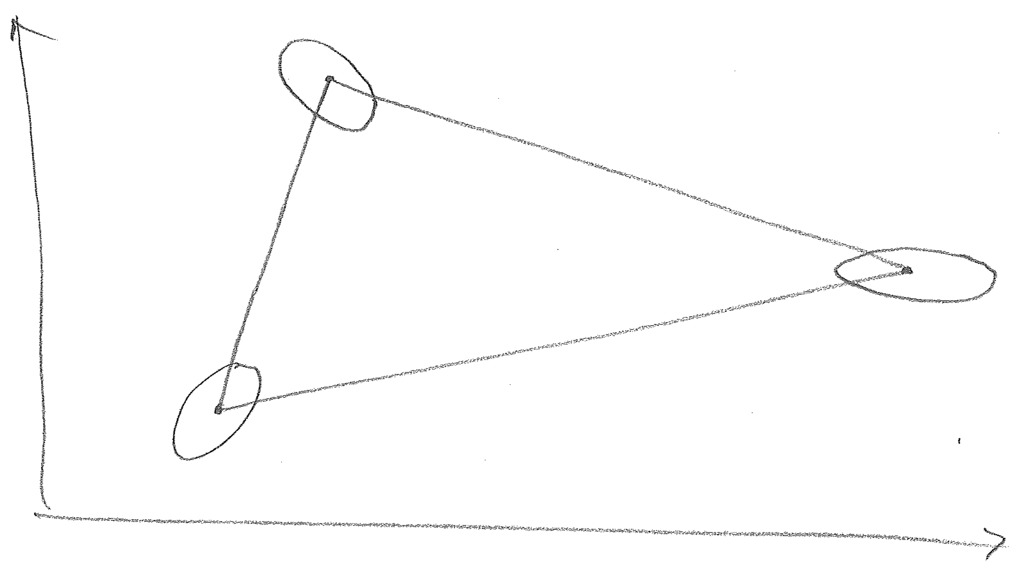

Figure 4. Error Ellipses with Inner or Free Net Constraints.