IS THE MINIMUM-TRACE DATUM DEFINITION THEORETICALLY CORRECT AS APPLIED IN COMPUTING 2D AND 3D DISPLACEMENTS?

Witold Prószyński

Institute of Applied Geodesy, Warsaw University of Technology, Pl. Politechniki 1, PL - 00 661, Warszawa, Poland

Abstract

It is shown that the minimum-trace datum definition as used for computing 2D and 3D positions and displacement vectors as well as their accuracy characteristics does not yield a physically interpretable datum with respect to orientation. This is explained on the basis of the analysis of mutual relationships between datum constraints in the initial non-linear Gauss-Markov model (GMM) for a local network and the corresponding constraints in the linearized system. A property of minimum-trace datum for 2D networks is proved which, in spite of the above mentioned drawback, convinces one of the safe use of this datum in practice. The property is illustrated on a simple practical example, being a 2D linear-angular monitoring network.

To keep the contents of the paper within the limits set by the Symposium organisers, the relevant proofs, except for one, are given in a brief outline only.

1. Introduction

A physically interpretable datum is a principal element of every process of determining the positions or displacement vectors. In addition to eliminating the defect of the monitoring network, such a datum should give a real sense both to the computed quantities and their accuracy characteristics. A question arises whether the frequently used minimum-trace datum definition fulfils the second requirement. The datum originated in the sixties as a core concept of free net adjustment and has since been extensively investigated by many researchers such as Meissl (1962), Mittermeyer (1972), Pelzer (1971), Grafarend, Schaffrin (1974), Caspary (1988), Papo, Perelmuter (1989) to mention but a few. Their research concentrated on linear (linearized) models, geometrical interpretation of the datum constraints and the properties of the least squares estimators. From the analysis of those findings, which are not conclusive as regards the orientation of 2D and 3D networks, it seems that the question of physical interpretability of the minimum-trace datum needs a special treatment. Following this observation, the present paper focuses on the question how this datum defines the orientation of the network, and hence, the orientation of the displacement vectors.

The physically-oriented approach applied in this paper is reflected in the following well-known principles:
- a co-ordinate system becomes identifiable in a physical reality by attaching it uniquely to a group of physical points (i.e. network monuments). The requirement for unique attachment follows from the fact that the determined positions of these points are stochastic quantities;
- the covariance matrix for a vector of positions (or position changes) must have a physically identifiable “zero-variance base”. In a local network the zero-variance base is defined on a specified subset of the network points;
The basis of the analysis will be the relationships between the constraints in the initial non-linear GMM and those in its linearized form. It is the former model that constitutes a proper mathematical model of the positioning or monitoring task.

2. Datum-defining constraints in a non-linear GMM and in its linearized form

Let us recall both a parametric non-linear model with minimum constraints and its linearized form

\[ \begin{align*}
    F(X) + e &= y ; \\
    G(X) &= c \\
    \text{Ax} + e &= y - y_{ap} ; \\
    \text{Sx} &= 0
\end{align*} \]

(1)

where:

- \( X, x \) = the \( u \times 1 \) vectors of parameters,
- \( F(X), G(X) \) = the \( n \times 1 \) and \( d \times 1 \) functional vectors, such that
- \( A = \frac{d[F(X)]}{dx} \) \( \text{X}_{ap} \); \( S = \frac{d[G(X)]}{dx} \) \( \text{X}_{ap} \); \( A \) (\( n \times u \)) matrix; \( \text{rank}(A) = u - d \)
- \( d \) (the network defect); \( S \) (\( d \times u \)), \( \text{rank}(S) = d \); \( \text{rank} [A^T S^T]^T = u \);
- \( \text{X}_{ap} \) = the \( u \times 1 \) vector of approximate co-ordinates
- \( c \) = the \( d \times 1 \) vector of constants;
- \( e \) = the \( n \times 1 \) vector of unknown random errors;
- \( y \) = the \( n \times 1 \) vector of observations;
- \( y_{ap} \) = the \( n \times 1 \) vector of the approximate observation values;
- \( C_e \) = the \( n \times n \) covariance matrix of observations (pos. definite).

More specifically, the constraints which constitute the definition of the datum for a local network should be written as \( G(X_b) = c \) and \( S_b x_b = 0 \) to indicate that they are defined on a chosen subset \( P_b \) of the network points.

The initial non-linear GMM, when used for positioning, can be interpreted as a description of the following task (see Fig.1a – for 2D case):

- construct figure \( F_g(y) \) and locate it uniquely in the co-ordinate system, i.e. according to the specified values of the location parameters \( G(X) \), here \( X_0 = X_0(X), Y_0 = Y_0(X), \alpha_0 = \alpha_0(X) \).

![Fig. 1](https://example.com/fig1.png)

Fig. 1 Specification of the positioning task (a) and the monitoring task (b)
In the case of monitoring the displacements the description of the task will be as follows (see Fig. 1b):
- construct figure $F_g(y_1)$ and figure $F_g(y_2)$, corresponding to two different time moments, and locate each uniquely in the co-ordinate system, i.e. according to the specified values of location parameters $G(X)$, here $X_0 = X_o(X)$, $Y_o = Y_o(X)$, $\alpha_o = \alpha_o(X)$.

Note that we obtain the superposition of the figures $F_g(y_1), F_g(y_2)$ in the indirect way, i.e. through locating each in the same datum.

Since $\sigma_{X_0} = 0; \sigma_{Y_0} = 0; \sigma_{\alpha_o} = 0$, the point $P_0$ and the line $l_0$ defined on the positions of the reference points constitute also the zero-variance base for the accuracy characteristics of computed positions or displacement vectors.

From the analysis presented so far it follows that the minimum constraints $G(X) = c$ should satisfy the following requirements:

i. **physical interpretability** – each constraint should correspond to a specified degree of freedom of $F_g(y)$ in a co-ordinate system;

ii. **differentiability** – the possibility of operating with a linear model $Sx = 0$.

On this basis we may formulate the corresponding requirements for linear constraints $Sx = 0$ when they are created for a linearized model

$$Ax + e = y - y_{ap} \quad; \quad e - (0, C_e) \quad (\text{see (1)})$$

without prior specifying $G(X) = c$:

i. **integrability** – the possibility of obtaining the relationships of the type $G(X) = c$.

ii. **physical interpretability** – each relationship obtained from integration should be identifiable as a constraint corresponding to a specified degree of freedom of $F_g(y)$ in a co-ordinate system.

3. **Shortcomings of the minimum-trace datum definition**

It is a specific case of datum for a local network, belonging to the class of the minimum-constraints datums. It uses $S = S_M$ (or more specifically $S_M = [S_{M,b} \ 0]$) such that $A S_M^T = 0$, which along with $v^T C_{e^{-1}} v = \min$ yields $\|\hat{S}_b\| = \min$ (and hence $\text{Tr}(C_{x,b}) = \min$).

Also each of the matrices $S = BS_M$, where $B$ is non-singular, can be used equivalently.

The constraints are formulated here at a level of the linearized model so the relationships corresponding to (1) can be written as follows

$$F(X) + e = y \quad; \quad e - (0, C_e) \quad \Downarrow \quad Ax + e = y - y_{ap} \quad; \quad e - (0, C_e)$$

$$G(X) = c \quad \Downarrow \quad S_M x = 0 \quad (2)$$

According to Section 2, the linear constraints $S_M x = 0$ should satisfy the requirements of integrability and physical interpretability. It is easy to see that such is the case with positional constraints, where we obtain after integration $X_C = \text{const.}; Y_C = \text{const.}; Z_C = \text{const.}, C$ being
the gravity centre of the group of reference points. Direct integrating of the orientation and scale constraints yields the corresponding constraints for the initial GMM in the form \( G(X, X_{ap}) = c \). They do not fulfil the requirement of being \( X_{ap} \) independent, i.e. \( G(X) = c \).

Seeking such a form we shall consider the constraints \( S_M x = 0 \) with the approximate co-ordinates being replaced by the corresponding variables. Then, on integrating the scale constraint we obtain immediately \( M_{p,c} = \text{const.} \), which by analogy to concepts used in mechanics (Leyko 1997), may be called a polar moment of inertia for the group of reference points with respect to its gravity centre \( C \). As regards the orientation constraints in \( S \), the proof given in Appendix B allows one to draw the following conclusions:

3D networks - for \( n \geq 3 \) the orientation constraints are not integrable;

2D networks - for \( n \geq 3 \) the orientation constraint is not integrable.

Summing up the above, we may define free net adjustment as solving the initial functional model as in (1) with the constraints as follows:

\[
G_p(X) = c_p; \quad G_\alpha(X, X_{ap}) = c_\alpha; \quad G_s(X) = c_s,
\]

where \( p, \alpha, s \) denote positional, orientation and scale constraints respectively.

For a 2D network without scale defect the model can be interpreted as the description of the following task (see Fig.2):

- construct figure \( F_g(y) \) and locate it uniquely in the co-ordinate system according to the specified values of the location parameters \( G_p(X) \), i.e. \( X_C, Y_C \), and the zero value of the relative orientation parameter \( G_\alpha(X, X_{ap}) \), i.e. \( A(X, X_{ap}) \) - being the algebraic sum of the hatched areas.

For \( n \geq 3 \) the orientation of the resulting figure \( F_g(X) \) cannot be defined directly relative to the axes of the co-ordinate system, i.e. through \( G_\alpha(X) \). Conversely, since the line \( l_0 \) cannot be defined we cannot reproduce the orientation of the X0Y system and we do not have the zero-variance base for network orientation.

The possibility of improving the accuracy of \( X_{ap} \) in the successive iterations, even till \( |\tilde{X} - X_{ap}| \) becomes negligibly small, does not lead to finding the required orientation parameter \( G_\alpha(X) \).
4. Advantageous properties of the minimum-trace datum definition

It can be proved (see Appendix C) that for each point of 2D network the following property holds (see Fig.3):

- the tangential component $t_M$ of the solution vector obtained with minimum trace datum is a weighted mean $t_w$ of the tangential components $t_{(j)}$ ($j = 1, 2, \ldots, n$) of solution vectors obtained with the datums, each being „point C fixed, the line CP$_j$ fixed”; the weights being $p_{(i)} = L_j^2$ (for the formula and the proof see Appendix C).

![Diagram](image)

Fig.3 Specific property of the minimum-trace datum for 2D networks

This implies the following properties:

* being a weighted mean, $t_M$ is always inside the interval $< t_{\min}, t_{\max} >$ and thus inside the interval obtained with physically interpretable datums;
* for $t_M$ as a weighted mean, we have $\sigma_{t,M} < \sigma_{t,max}$.

The above properties are consistent with those formulated for minimum-constraints datums with the gravity centre fixed (see relationships (A1), Appendix A). Note also the properties of minimum-trace datum related to the reference points (inequalities (A5) and (A6), Appendix A).

**Practical example.** Figure 1 shows the tangential components $t_M$ and $t_{(j)}$ $j = 1, 2, \ldots, 6$ of displacement vectors for some points of the angular-linear monitoring network. We can see that the arrangement of the t – values is similar for all the three network points.

5. Concluding remarks

On the basis of the analysis carried out in this paper one may formulate the following conclusions:

- for $n \geq 3$ the minimum-trace datum definition does not specify a physically interpretable reference base for the orientation of 2D and 3D networks, and hence for the displacement vectors and their accuracy characteristics;
- it is an advantageous property of the minimum-trace datum definition for 2D networks that the resulting solution vectors for each network point fall in between those obtained with physically interpretable datums. One may expect that the analogous properties can be found for 3D networks;
Appendix A: Properties of the minimum-constraints datums with fixed gravity centre (2D and 3D networks)

Let us consider the following two options of minimum-constraints datum for a network without scale defect:

- \[ S_M x = 0 \] minimum-trace datum
- \[ S_N x = 0 \] datum with fixed gravity centre and orientation constraints other than those above,

- in view of practical applications the advantageous property as above compensates for the lack of physically interpretable reference base for orientation;
- the accuracy of the approximate co-ordinates, which can be improved in successive iteration steps, is neither the cause nor adds to the shortcomings of the minimum-trace datum definition.

References


Appendix A: Properties of the minimum-constraints datums with fixed gravity centre (2D and 3D networks)

Let us consider the following two options of minimum-constraints datum for a network without scale defect:

- \[ S_M x = 0 \] minimum-trace datum
- \[ S_N x = 0 \] datum with fixed gravity centre and orientation constraints other than those above,

Fig.4 The tangential components of displacement vectors in a monitoring network
where: \( S_M = [S_{M,b} \ 0] = \begin{bmatrix} S_{p,b} & 0 \\ S_{\beta,b} & 0 \end{bmatrix}; \ S_N = [S_{N,b} \ 0] = \begin{bmatrix} S_{p,b} & 0 \\ S_{\alpha,b} & 0 \end{bmatrix}; \ x = \begin{bmatrix} x_b \\ x_c \end{bmatrix} \)

the indices \( b \) and \( c \) denote those network points which form the reference base (b) and those which do not belong to this base (c); \( p \) – stands for positions, \( \alpha \) and \( \beta \) - stand for orientation.

Let \( \hat{L}_i \) denote the length of the line joining the gravity centre \( C \) and a point \( P_i \) and \( \sigma_{\hat{L}_i} \) - the standard deviation of obtained from the LS estimation \( i, L \hat{=} i, P \).

PROPERTIES. For any network point both \( \hat{L} \) and \( \sigma_{\hat{L}} \) are invariant to the choice of orientation constraints, i.e.

\[
\hat{L}_i(S_N) = \hat{L}_i(S_M), \quad \sigma_{\hat{L}_i}(S_N) = \sigma_{\hat{L}_i}(S_M) \quad i = 1,2,...,n
\]  

(A1)

PROOF. Applying a well-known similarity transformation (here – isometry transformation) – see Baarda (1973), we get

\[
\hat{x}_N = G \hat{x}_M
\]  

(A2)

where

\[
G = I - S_{M*}(S_N S_{M*})^{-1} S_N; \quad S_{M*} = \begin{bmatrix} S_{p,b} & S_{p,c} \\ S_{\beta,b} & S_{\beta,c} \end{bmatrix};
\]

\[
\hat{x}_M = \begin{bmatrix} \hat{x}_{M,b} \\ \hat{x}_{M,c} \end{bmatrix}; \quad \hat{x}_N = \begin{bmatrix} \hat{x}_{N,b} \\ \hat{x}_{N,c} \end{bmatrix}
\]

and carrying out a sequence of simple operations (including the reductions of mutually orthogonal factors), we obtain

\[
\hat{x}_N = \begin{bmatrix} Q & 0 \\ -U & I_c \end{bmatrix} \begin{bmatrix} \hat{x}_{M,b} \\ \hat{x}_{M,c} \end{bmatrix}, \quad \text{(A3)}
\]

where

\[
Q = I_b - S_{\beta,b}^T(S_{\alpha,b} S_{\beta,b})^{-1} S_{\alpha,b}; \quad U = S_{\beta,c}^T(S_{\alpha,b} S_{\beta,b})^{-1} S_{\alpha,b}.
\]

In a linear model (2) we can represent \( \hat{L}_i \) as \( \hat{L}_i = L_{ap} + g_i \hat{x}^T \),

where \( g_i = [0,0,0,...,\Delta X_i,\Delta Y_i,\Delta Z_i,...,0,0,0] \cdot L_{ap}^{-1} \)

It is easy to check that \( g_{i,b} S_{\beta,b}^T = 0 \) and \( g_{i,c} S_{\beta,c}^T = 0 \), and hence

\[
g_{i,b} \hat{x}_{N,b}^T = g_{i,b} \hat{x}_{M,b}^T \quad \text{and} \quad g_{i,c} \hat{x}_{N,c}^T = g_{i,c} \hat{x}_{M,c}^T, \quad \text{(A4)}
\]

which yields immediately \( \hat{L}_i(S_N) = \hat{L}_i(S_M) \) for \( i = 1,2,...,n \).

From (A4) it follows that

\[
g_i C_{\hat{x},N} g_i^T = g_i C_{\hat{x},M} g_i^T
\]

and thus \( \sigma_{\hat{L}_i}(S_N) = \sigma_{\hat{L}_i}(S_M) \) for \( i = 1,2,...,n \) (Q.E.D.)
Figure 5 illustrates these properties for a 2D network. The solution vector $x_{N,o}$ and the error bar $B_{N,o}$ correspond to a datum having an orientation constraint $d\alpha = 0$, where $\alpha$ is a bearing of the line $CP_i$. The analogous bounds but in the form of planes exist for 3D solution vectors and the standard ellipsoids.

The properties (A1) can be extended upon the networks with scale deficiency.

Let $r_i, t_i^r, t_i^t$ denote respectively the radial and the tangential components of the solution vector $x_i$ for the $i$–th network point.

Since $\|x_{N,b}\| < \|x_{N,b}\|$ and $\text{Tr}(C_{x_{M,b}}) < \text{Tr}(C_{x_{N,b}})$, and according to (A1)

\[
\begin{align*}
 r_i(S_N) &= r_i(S_M) \\
 \sigma_{t_i}(S_N) &= \sigma_{t_i}(S_M)
\end{align*}
\]

we obtain immediately

\[
\sum_{(b)} (t_{M}^r)^2 + \sum_{(b)} (t_{M}^t)^2 < \sum_{(b)} (t_{N}^r)^2 + \sum_{(b)} (t_{N}^t)^2
\]

and

\[
\sum_{(b)} \sigma_{t_i,M}^2 + \sum_{(b)} \sigma_{t_i,N}^2 < \sum_{(b)} \sigma_{t_i,N}^2 + \sum_{(b)} \sigma_{t_i,N}^2
\]

and for 2D networks

\[
\sum_{(b)} t_{M}^2 < \sum_{(b)} t_{N}^2 \\
\sum_{(b)} \sigma_{t_i,M}^2 < \sum_{(b)} \sigma_{t_i,N}^2
\]

(A5)

(A6)

**Appendix B: Integrating the constraints equations in minimum-trace datum definition**

Let $(x,y,z) = (x_1, \ldots, x_n, y_1, \ldots, y_n, z_1, \ldots, z_n)$ denote the standard co-ordinates system on $R^n$ where $n \geq 3$. Let $D$ denote the following Pfaffian system on $R^n$:

\[
\begin{align*}
\sum_{i=1}^{n} dx_i &= 0; \\
\sum_{i=1}^{n} x_i dy_i - y_i dx_i &= 0 \\
\sum_{i=1}^{n} dy_i &= 0; \\
\sum_{i=1}^{n} y_i dz_i - z_i dy_i &= 0 \\
\sum_{i=1}^{n} dz_i &= 0; \\
\sum_{i=1}^{n} z_i dx_i - x_i dz_i &= 0
\end{align*}
\]
PROPOSITION 1 \( D \) is not integrable.

Proof. Let us denote

\[
\alpha_1 = \sum_{i=1}^{n} dx_i \quad \beta_1 = \sum_{i=1}^{n} x_i dy_i - y_i dx_i
\]

\[
\alpha_2 = \sum_{i=1}^{n} dy_i \quad \beta_2 = \sum_{i=1}^{n} y_i dz_i - z_i dy_i
\]

\[
\alpha_3 = \sum_{i=1}^{n} dz_i \quad \beta_3 = \sum_{i=1}^{n} z_i dx_i - x_i dz_i
\]

It is easy to see that

\[
\alpha_1 = d(\sum_{i=1}^{n} x_i) \quad \alpha_2 = d(\sum_{i=1}^{n} y_i) \quad \alpha_3 = d(\sum_{i=1}^{n} z_i)
\]

Therefore \( d\alpha_i = 0, \ i = 1,2,3 \) and \( d\beta_1 = 2 \sum_{i=1}^{n} dx_i \wedge dy_i \).

We define

\[
\gamma = d\beta_1 \wedge \alpha_1 \wedge \alpha_2 \wedge \alpha_3 \wedge \beta_1 \wedge \beta_2 \wedge \beta_3
\]

\( \gamma \) is an 8-form on \( \mathbb{R}^{3n} \).

By direct calculation we obtain

\[
\gamma = \left( \frac{\partial}{\partial x_1}, \frac{\partial}{\partial y_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial y_2}, \frac{\partial}{\partial z_1}, \frac{\partial}{\partial x_3}, \frac{\partial}{\partial y_3}, \frac{\partial}{\partial z_3} \right)
\]

\[
= 2(2x_3 y_2 y_3 - 2x_2 y_3^2 - 2x_3 y_1 y_3 + 2x_1 y_3^2 + 2x_2 y_1 y_3 - 2x_1 y_2 y_3
\]

\[
- x_3 y_2^2 + x_2 y_2 y_3 - x_1 y_2 y_3 - x_2 y_1 y_2 + x_1 y_2^2
\]

\[
+ x_2 y_1 y_3 + x_3 y_1^2 - x_1 y_1 y_3 - x_2 y_1^2 + x_1 y_1 y_2)
\]

which is a polynomial of the 3rd degree.

As \( \gamma \) does not vanish on any open subset of \( \mathbb{R}^{3n} \), by Frobenius theorem (see Chern, Chen and Lam 1999) \( D \) is not integrable (Q.E.D.).

Appendix C: A specific property of the minimum-trace datum definition for 2D networks

Let \( x_{N(j)} \) \( j = 1,2,...,n \) denote the solution vectors obtained with minimum-constraints datums each with the gravity centre fixed and the orientation constraint \( d\alpha_j = 0 \) and \( x_M \) - the solution vector obtained with minimum-trace datum.

Let \( t_{N(j)} = Tx_{N(j)} \) \( j = 1,2,...,n \) and \( t_M = Tx_M \), where \( T \) is the \((n\times2n)\) rotation matrix, be the respective tangential components of the solution vectors.
PROPERTY

\[ \sum_{j=1}^{n} p(j) t_{N(j)} = t_{M} \quad j = 1, 2, \ldots, n, \quad \text{where} \quad p(j) = \frac{L_j^2}{\sum_{j=1}^{n} L_j^2} \]  

(C1)

PROOF Since \( x_{N(j)} = G_{N(j)} x_{M} \) (see A2) the relationship (C1) takes the form

\[ \sum_{j=1}^{n} p(j) T G_{N(j)} \cdot x_{M} = T x_{M} \]  

(C2)

It can be shown that (C2) holds true, if

\[ (\sum_{j=1}^{n} p(j) T G_{N(j)} - T) G_{o} = 0 \]  

(C3)

where \( G_{o} = I - S_{M}(S_{M} S_{M}^{T})^{-1} S_{M} \).

Making use of (A3) we can write the condition (C3) in the form

\[ T \begin{pmatrix} p(1) Q_{1} & 0 \\ -U_{1} & I_{c} \end{pmatrix} + \ldots + p(n) \begin{pmatrix} Q_{n} & 0 \\ -U_{n} & I_{c} \end{pmatrix} - \begin{pmatrix} I_{b} & 0 \\ 0 & I_{c} \end{pmatrix} \right] G_{o} = 0 \]

and further on

\[ T \begin{pmatrix} \sum p(j) Q_{j} - I_{b} & 0 \\ -\sum p(j) U_{j} & 0 \end{pmatrix} \right] G_{o} = 0 \]

Verifying that this condition is satisfied completes the proof.