Statistics & Decisions, Supplement Issue No. 2, 455–466 (1985)

A NOTE ON THE USE OF GAUSS' FORMULAS IN NON-LINEAR GEODESIC ADJUSTMENT

P.J.G. Teunissen

Received: revised version: February 13, 1985

Abstract. The theory of adjustment is usually expounded by algebraic and analytical methods. It is well known, however, that the theory of linear adjustment can be represented simply and directly, using geometric reasoning, as properties of linear spaces. In Geodesy, geometric reasoning was already advocated by Tienstra (1948), who used the Ricci calculus. Generalizing, the theory of non-linear adjustment can then be represented as properties of curved manifolds.

In this note we show the principal role played by Gauss' formulas, known from differential geometry, in Gauss' method of least-squares. In particular we show the significance of

AMS 1980 subject classification: 62 P 99, 62 F 10, 86 A 30.
Key words and phrases: Gauss' method of least-squares, Gauss' iteration method, geodesic adjustment, planar Helmert transformation, local bias.

Gauss' formulas in deriving non-linearity measures, bias expressions and the local convergence factor of Gauss' iteration method.

Introduction

The study of differential geometry was begun very early in history and geodesy benefited considerably from its development. In fact, practical tasks in cartography and geodesy caused and influenced the creation of the classical theory of surfaces (Gauss (1827), Helmert (1880)). And differential geometry can now be said to constitute an essential part of the foundation of both mathematical and physical geodesy. But it was not only in the development of geodetic models that differential geometry played such a pivotal role. Also in geodetic adjustment theory, adjustment was soon considered as a geometrical problem. Very early Tienstra (1948), already advocated the use of the Ricci calculus in adjustment theory. His approach was later followed by Baarda (1967), Kooimans (1958) and many others.

Geodesists thus seem to feel very comfortable with geometric reasoning and geometric interpretations. This is partly due to their background, but also because geometric approaches often lead to simpler and more easily interpretable expressions and to simpler proofs and derivations. Recently this was illustrated again by Krarup (1982), who showed how to benefit from differential geometric methods in dealing with the problem of geodetic network adjustments on the ellipsoid. And in this note we will give yet another illustration.

We will discuss geometrically two seemingly unrelated but still geometrically related problems from the complex of problems on non-linear geodetic adjustment. Namely, the local convergence behaviour of Gauss' iteration method and the expected bias in adjusted variables. Both are of particular geodetic relevance. Gauss' method, since it is preeminently suited for geodetic non-linear adjustment problems. Not only because it is the logical generalization of the linear case, but also because geodetic adjustment problems usually have reasonable small residuals and aren't too non-linear. And bias is considered, because of its importance in especially the difficult subject-matter of geodetic error-control (e.g. (Baarda 1969), Teunissen (1984)).

As said our approach will be a geometrical one. In the next section we will therefore start reviewing some preliminaries from the theory of submanifolds, in particular Gauss' formula. In section three we then show how Gauss' formulas furnish a geometric interpretation of the local convergence behaviour of Gauss' iteration method and of the expected bias in the adjusted variables. We conclude this section with some geodetic examples.

Because of the space permitted we will refrain as much as possible form giving derivations. For derivations of differential geometric results we refer to any textbook on contemporary differential geometry (e.g. Hicks, 1974).

2. On the geometry of submanifolds

Let N be an n-dimensional differentiable manifold immersed in an m-dimensional manifold M $(m > n)$ by a mapping y . We will assume that M is a Riemannian manifold with a Riemannian connection $\bar{\nabla}$ defined by the connection coefficients Γ^k_{ij} , $i,j,k = 1,\dots,m$. If the metric of M is given by the positive definite symmetric tensorfield $\bar{g}$ of type (0,2), one can induce a metric, g say, on N by pulling the metric tensor $\bar{g}$ of M back to N . The definition of the induced metric g reads:

$$(\qquad g(X,Y) = \bar{g}(y_*(X), y_*(Y))$$

for any vector fields X, Y in N ; y_* is the tangent map of y . With the so induced metric one can then define the unique Riemannian connection, ∇ say, of N . One can, however, also follow an alternative route for introducing the connection of N . From the fact that the tangent bundle of M , restricted to $y(N)$ equals the direct sum of the tangent bundle and the normal bundle of $y(N)$, i.e.

$$T(M)_{|y(N)} = T(y(N)) \oplus T^{\perp}(y(N)) \ ,$$

follows namely that for any vector fields, X , Y of N , $\bar{\nabla}_{y_*(X)} y_*(Y)$ can be expressed in the form

$$(2) \qquad \bar{\nabla}_{y_*(X)} y_*(Y) = y_*(\nabla_X Y) + B(X,Y)$$

where $\nabla_X Y$ is a vector field tangent to N and $B(X,Y)$ is a vector field normal to $y(N)$.

Expression (2) constitutes the generalized Gauss' formulas and they can be considered as one of the fundamental equations for submanifolds. From the affine properties of the connection $\bar{\nabla}$ follows that ∇ is an affine connection for N and that $B(X,Y)$ is bilinear. And since $\bar{\nabla}$ has no torsion also ∇ is torsionfree and $B(X,Y)$ is symmetric. And finally, since $\bar{\nabla}$ is metric one will find that also ∇ is metric. Hence ∇ is the unique, so-called induced, Riemannian connection of N .

Our appreciation of Gauss' orthogonal decomposition formula (2) now stems from the fact that it decomposes the connection of the ambient space M into an intrinsic and extrinsic part. A tangential part which produces the induced connection, a rule according to which one may compare tangent vectors situated at different tangent spaces of the manifold N . And a normal part which can be shown to measure the curvature of the submanifold $y(N)$ in M .

Let us first consider the tangential component of (2). The connection coefficients $\Gamma^{\gamma}_{\alpha\beta}$, $\alpha,\beta,\gamma = 1,\ldots,n$ of the induced connection applied to the coordinate vector fields ∂_α of N

can be shown to describe the local behaviour of the coordinate curves in N. They are not the components of a tensor, i.e. they do not follow the tensorial transformation law, and hence depend on the choice of local coordinates in N. Since N is Riemannian it follows that generally one cannot find local coordinates for which these connection coefficients vanish identically. There are, however, two exceptions. The first is when N can be considered to be Euclidean. And the second when N is one dimensional. In this case namely N and its tangent space TN are identifiable. Coordinatization with the arc length or a linear function thereof will then make the connection coefficients vanish identically.

Although generally on cannot make the connection coefficients vanish for every point of the manifold N, there exist local coordinates for which they will vanish in just one point, x_o say. And as one can expect these coordinates, $\bar{x}^{\alpha}$, $\alpha = 1,\ldots,n$, say, ought to be generalizations of the arc length. They are known as Riemannian or normal coordinates and are defined as

$$(3) \qquad \bar{x}^{\alpha} = (x^{\alpha} - x_o^{\alpha}) + \frac{1}{2}\Gamma^{\alpha}_{\beta\gamma}(x^{\beta} - x_o^{\beta})(x^{\gamma} - x_o^{\gamma}), \quad \alpha,\beta,\gamma = 1,\ldots,n$$

As to the normal component of (2), assume N_p, $p = 1,\ldots,(m-n)$, to be an orthonormal basis of $T(y(N))$. Then $B(\partial_{\alpha},\partial_{\beta}) = B^p_{\alpha\beta}N_p$ and $g(B(\partial_{\alpha},\partial_{\beta}),N_q) = B^q_{\alpha\beta}$. For a particular normal direction, N_q say, the eigenvalues, with respect to the induced metric, of $B^q_{\alpha\beta}$ then determine the so-called principal curvatures of the submanifold $y(N)$ for N_q. Thus the principal curvatures are the eigenvalues of the generalized eigenvalue problem:

$$\det(B^q_{\alpha\beta} - \lambda^q g_{\alpha\beta}) = 0 ,$$

or the eigenvalues of

$$(4) \qquad \det(\bar{g}(g^{\gamma\alpha}B(\partial_{\alpha},\partial_{\beta}),N_q) - \lambda^q \delta^{\gamma}_{\beta}) = 0 .$$

The corresponding eigenvectors then determine the mutually orthogonal principal direction for N_q. The trace of B

gives us the mean curvature H^q of $y(N)$ for N_q :

(5) $$H^q = \frac{1}{n} \sum_{\alpha=1}^{n} \lambda_\alpha^q = \frac{1}{n} g^{\alpha\beta} B_{\alpha\beta}^q .$$

And the unique mean curvature normal η of the submanifold $y(N)$ is defined as

(6) $$\eta = H^q N_q \ , \ q = 1,\ldots,(m-n) .$$

Submanifolds for which the mean curvature normal η vanishes identically are called minimal submanifolds. And those for which the normalfield B vanishes identically are called totally geodesic.

3. The local convergence factor of Gauss' iteration method and local bias expressions

Now that we have dealt with the necessary preliminaries we can start considering the adjustment problem of orthogonally projecting a given sample point $y_s \in M$ onto the submanifold $y(N)$ and show how Gauss' decomposition formula (2) comes to our use. We assume M to be Euclidean and $y_s - \varepsilon \in y(N)$, with ε randomly distributed and $E\{\varepsilon^i \varepsilon^j\} = \sigma^2 g^{ij}$, $E\{\varepsilon^i\} = 0$. For $\hat{y}$ to be the adjusted point we have as necessary condition that the residual vector $y_s - \hat{y}$ should be orthogonal to the tangent space $T_{\hat{y}}(y(N)))$ of $y(N)$ at $\hat{y}$, i.e. we have that

(7) $$\bar{g}(y_*(\partial_\alpha), y_s - \hat{y}) = 0 ,$$

must hold at $\hat{y} \in y(N)$. Due, however, to the assumed non-linearity of the mapping y , the tangent space $T_{\hat{y}}(y(N))$ is generally unknown a priori. Hence our adjustment problem cannot be solved directly. Expression (7) does, however, suggest a way of solving our adjustment problem. Instead of orthogonally projecting y_s onto the tangent space $T_{\hat{y}}(y(N))$

one can take as a first approximation the orthogonal projection of y_s onto a chosen nearby tangent space $T_{y_k}(y(N))$. Of course then

$$(8) \qquad \bar{g}(y_*(\partial_\alpha), y_s - y_k) \neq 0 ,$$

at $y_k \in y(N)$. But by pulling the non-orthogonality as measured by (8) back to the Riemannian manifold N , we get

$$(9) \qquad g(\partial_\alpha, \Delta x_{k+1}) = \bar{g}(y_*(\partial_\alpha), y_s - y_k) \text{ , with } \Delta x_{k+1} \in T_{x_k}(N) ,$$

which suggests in local coordinates the following iteration procedure

$$(10) \qquad x^\beta_{k+1} - x^\beta_k = g^{\beta\alpha}\partial_\alpha y^i g_{ij}(y^j_s - y^j_k) \qquad \alpha, \beta = 1,\ldots,n \text{ ; } i,j = 1,\ldots,m \text{ .}$$

This is Gauss' iteration method and it consists of successively solving a linear least-distance adjustment problem until condition (7) is met. Because of the reasonable small residuals and the moderate non-linearity of geodetic adjustment problems, the method is usually used in geodesy. Now in order to understand the local behaviour of Gauss' iteration method, consider what happens geometrically when (9) is applied. At each iteration step $k+1$ the sample point y_s is orthogonally projected onto a new tangent space $T_{y_{k+1}}(y(N))$, which will be close to the previous one, $T_{y_k}(y(N))$. Hence the rate in which the tangential part of $y_s - y_k$ decreases will depend on the rate of change of the tangent spaces. And since curvature is classically defined to measure the rate of change of tangents, one can expect the local behaviour of Gauss' iteration method to depend on the principal curvatures of the submanifold $y(N)$.

Now assume the submanifold $y(N)$ to be totally geodesic. Then all tangent spaces of $y(N)$ become identifiable with the submanifold itself. Hence the approximation of projecting onto the wrong tangent space will be absent then. The only approximation left, consists of interpreting (10) as the coordinatized version of (9). Namely in expression (9) Δx_{k+1} is a tangent vector of the tangent space $T_{x_k}N$ of N at x_k, whereas in (10) the components of the tangent vector Δx_{k+1} are interpreted as the components of the separation vector between the two points x_k and x_{k+1} of N. But no such separation vector exists in Riemannian manifolds. In Riemannian geometry each local geometric object has namely its own official place of residence; it can interact with other objects residing there; but it cannot interact with any object at another point, until it has been carefully transported to this point. And this transport is failing in the above interpretation. Hence the approximation involved will depend on the induced connection of N. And one can therefore expect that in this case the local behaviour of Gauss' iteration method is determined by the connection coefficients of N.

Now to describe the local behaviour of Gauss' iteration method quantitatively we Taylorize the right-hand side of (10) at the point $\hat{x} = y^{-1}(\hat{y})$ and find

$$(11)\qquad x^{\gamma}_{k+1} - \hat{x}^{\gamma} = \bar{g}(g^{\gamma\alpha}B(\partial_{\alpha},\partial_{\beta}), y_s - \hat{y})(x^{\beta}_{k} - \hat{x}^{\beta}) + \text{second and higher order terms.}$$

And indeed we find that the local linear convergence behaviour of Gauss' iteration method depends on the normal component of Gauss' formula (2). The local convergence factor equals namely the product of the in absolute value largest principal curvature of the submanifold $y(N)$ for the normal direction of the residual vector $y_s - \hat{y}$ with the length of the residual vector.

Expression (11) also makes clear that if $y_s \in y(N)$ or if the submanifold $y(N)$ is totally geodesic then Gauss' iteration method is locally quadratic convergent. In this case we find that

(12) $$x^{\gamma}_{k+1} - \hat{x}^{\gamma} = \frac{1}{2}\Gamma^{\gamma}_{\alpha\beta}(x^{\alpha}_{k} - \hat{x}^{\alpha})(x^{\beta}_{k} - \hat{x}^{\beta}) + \text{third and higher order terms,}$$

which shows that, as expected, the locally quadratic convergence behaviour is determined by the connection coefficient of N.

The above expressions (11) and (12) thus make clear what role is played by the tangential and normal components of Gauss' formula (2) in the local convergence behaviour of Gauss' iteration method. In a similar way we can now also use Gauss decomposition formula to obtain a geometric interpretation of the bias in the estimators $\hat{x}$ and $\hat{y} = y(\hat{x})$. The reasoning is quite analogous to the above given one. Using the results of Box (1971) one will find that

(13) $$E\{\hat{x}^{\gamma} - x^{\gamma}\} = -\frac{1}{2}\sigma^{2} g^{\alpha\beta}\Gamma^{\gamma}_{\alpha\beta},$$

and

(14) $$E\{y(\hat{x}) - y(x)\} = \frac{1}{2}\sigma^{2} g^{\alpha\beta}B(\partial_{\alpha},\partial_{\beta}) = \frac{1}{2}n\sigma^{2}\eta.$$

Thus the bias in the x-parameters depends on the connection coefficients of N, whereas the bias in the adjusted sample point is given by the mean curvature of the submanifold $y(N)$. Hence the bias in the x-parameters can be manipulated by a change of parameter-choice, whereas the bias in $\hat{y}$ is invariant to such a change of parameters. In particular one can reduce the bias in $\hat{x}$ to zero by choosing Riemannian coordinates. To know the bias in $\hat{y}$, i.e. to know the curvature of submanifold $y(N)$, is of particular importance for geodetic error-control.

Note by the way that one can write (13) and (14) also as

$$E\{\hat{x}^{\gamma} - x^{\gamma}\} = \frac{1}{2}\sigma^{2}\mathrm{Div.}(dx^{\gamma}) \quad \text{and}$$

$$E\{y(\hat{x}) - y(x)\} = \frac{1}{2}\sigma^{2}\Delta y,$$

where $\mathrm{Div.}(dx^{\gamma})$ is the generalized divergence operator as applied to the 1-form dx^{γ} of the dual- or co-tangent space $T_x^{*}N$ and Δ is Beltrami's second differential operator.

We conclude this note with two geodetic examples.

Example 1: The planar Helmert transformation:

In geodesy the Helmert transformation is often used to connect or combine geodetic networks. The model reads

$$x_i = \lambda\cos\phi\bar{x}_i + \lambda\sin\phi\bar{y}_i + t_x$$

$$y_i = -\lambda\sin\phi x_i + \lambda\cos\phi\bar{y}_i + t_y \ ,$$

for $i = 1,\ldots,\frac{1}{2}n$, where n is the number of network points involved; x_i, y_i are the observed cartesian coordinates of the network points; λ, ϕ, t_x, t_y are the unknown scale-, orientation-, and translation parameters; and $\bar{x}_i$, $\bar{y}_i$ are the given fixed coordinates of the network points in a reference system.

Since $y(N)$ is totally geodesic in this case, Gauss' iteration method will show a locally quadratic convergence behaviour and the adjusted coordinates will be unbiased. Assuming $g_{ij} = \delta_{ij}$, the bias in the adjusted parameters follows from

$$g^{\alpha\beta}\Gamma^{\gamma}_{\alpha\beta} = g^{22}\Gamma^{\gamma}_{22} \ ,$$

with $g^{22}\Gamma^{1}_{22} = -(\lambda \sum_{i=1}^{n} (r_i^2 - R^2))^{-1}$, $r_i^2 = \bar{x}_i^2 + \bar{y}_i^2$,

$$R^2 = \left(\frac{\sum_{i=1}^{n} \bar{x}_i}{n}\right)^2 + \left(\frac{\sum_{i=1}^{n} \bar{y}_i}{n}\right)^2 , \text{ and } g^{22}\Gamma^{\gamma}_{22} = 0 \text{ for } \gamma = 2,3,4 \ .$$

Example 2: A planar geodetic triangulation chain:

Assume that a planar geodetic triangulation chain as depicted in figure 1 has been measured

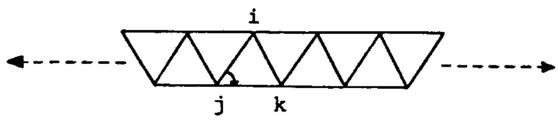

Figure 1

The observations equations then take the form of

$$\alpha_{ijk} = \tan^{-1}\frac{x_{jk}}{y_{jk}} - \tan^{-1}\frac{x_{ji}}{y_{ji}}, \tag{15}$$

where α_{ijk} is the observed angle in point P_j and x_{ji}, y_{ji} are the unknown cartesian coordinate differences of the network points. Despite the non-linearity of (15) one will then find again that $y(N)$ is totally geodesic.

References

[1] Baarda, W. (1967): Statistical Concepts in Geodesy, Netherlands Geodetic Commission, Publications on Geodesy, New Series, Vol. 2, No. 4.

[2] Baarda, W. (1969): A Testing Procedure for Use in Geodetic Networks, Netherlands Geodetic Commission, Publication on Geodesy, New Series, Vol. 2, No. 5.

[3] Box, M.J. (1971): Bias in Non-Linear Estimation (with Discussion). J.R. Statist. Soc. B32. 171-201.

[4] Gauss, C.F. (1827): Allgemeine Flächentheorie (Disquisitiones Generales Circa Superficies Curvas), Deutsch herausgegeben von A. Wangerin, Leipzig 1889.

[5] Gauss, C.F. (1887): Abhandlungen zur Methode der Kleinsten Quadrate, Deutsch herausgegeben vom Königl. Preussischen Geodätischen Institut, Berlin 1887.

Grafarend, E.W. (1981): Kommentar eines Geodäten zu einer Arbeit E.B. Christoffels, in E.B. Christoffel, The Influence of his Work on Mathematics and the Physical Sciences. Edited by P.L. Butzer and F. Fehér, Birkhäuser Verlag, 1981.

Helmert, F.R. (1880): Die mathemat. und physikal. Theorieen der höheren Geodäsie, Leipzig.

[8] Hicks, N.J. (1974): Notes on Differential Geometry. Van Nostrand Reinhold Mathematical Studies.

[9] Kooimans, A.H. (1958): Principles of the Calculus of Observations, Rapport Spécial, Neuvieème Congres International des géomètres, Pays-Bas, 301-310.

Krarup, T. (1982): Non-Linear Adjustment and Curvature, in Forty Years of Thought, Delft, 145-159.

[11 Tienstra, J.M. (1948): The Foundation of the Calculus of Observations and the Method of Least-Squares, Bulletin Geódêsique, No. 10, 1948.

Tienstra, J.M. (1956): Theory of the Adjustment of Normally Distributed Oservations, N.V. Uitgeverij Argus, Amsterdam, 1956.

[13] Teunissen, P.J.G. (1984): Quality Control in Geodetic Networks. Lecture notes International School of Geodesy. 3rd Course: Optimization and Design of Geodetic Networks, Erice-Trapani-Sicily, 25 April - 10 May 1984.

Peter J.G. Teunissen
Delft University of Technology
Department of Geodesy
Thijsseweg 11, Delft
The Netherlands