International Association of Geodesy
Section IV : Theory and Evaluation
SSG 4.56: *The Differential Geometry of the Gravity Field*

Minutes
of the Chaniá Workshop
Chaniá (Crete) Greece, June 10-12, 1985

Editors
K. Katsambalos and E. Livieratos

Sponsored by

The Hellenic Ministry of Culture
The Technical University of Crete
The Department of Geodesy and Surveying,
University of Thessaloniki

Introduction by P. Teunissen

NON-LINEAR ADJUSTMENT, AN INTRODUCTORY DISCUSSION AND SOME NEW RESULTS

Let us first give a brief review of the geometry of linear least-squares adjustment. The whole process of adjustment can conveniently be divided into

a) an adjustment computational part and

b) a quality control or diagnostic part.

The adjustment computational part can be divided into

1) the actual adjustment and

2) the actual inverse mapping.

If we assume that our mathematical model consists of a linear manifold $\bar{n}$ embedded in an ambient Euclidean observation space m, with $\bar{n} = \{y_1\} + u$, $y_1 \in m$ and u a subspace of m, then the solution to the actual adjustment problem is given by

$$\hat{y} = y_1 + P_{u,\, u^\perp}(y_s - y_1) \quad \text{for some} \quad y_1 \in \bar{n}, \tag{1}$$

where $P_{u,\, u^\perp}$ stands for the projector projecting onto u and along $u^\perp$, and $y_s \in m$ stands for the given sample vector. Note that $\hat{y}$ is the unique solution to the problem of finding that point on $\bar{n}$ which has least-distance to $y_s \in m$ and that $\hat{y}$ is normally distirbuted if y_s is.

If we now represent the subspace u by a linear map A: $n \to m$, i.e. $u = An$, then the solution to the actual inverse mapping problem reads

$$\hat{x} = B(\hat{y} - y_1) \quad \text{for some } y_1 \in \bar{n}; \text{ and some } B\colon m \to n \text{ such that } ABA = A \tag{2}$$

Note that $\hat{x}$ is *not* unique if $A^{-1}(0) \neq \{0\}$ and that $\hat{x}$ is normally distributed if y_s is. As to the diagnostic part, one is usually interested in the effects of changes in either the metric of m, the distribution assumed or the linear manifold $\bar{n}$. If we focus on the latter effects, those are usually tested on their significance through the use of statistical tests. For instance, under the nulhypothesis we have:

$$\frac{\|\hat{y}-y_s\|^2}{\sigma^2\,(\dim \mathcal{M} - \dim \mathcal{U})} \sim X^2\,(\dim \mathcal{M} - \dim \mathcal{U})\,,$$

which leads to X^2 –or F– tests, and

$$\frac{<P_{\mathcal{U}^\perp,\,\mathcal{U}}\,c_i,\; y_s-y_1>}{\sigma\,\|P_{U^\perp,\,U}\,c_i\|} \sim N(0,\,1)\,,$$

which leads to data snooping through the choice of the vector c_i. Alternative approaches are possible.

Let us now consider the geometry of non-linear adjustment. Instead of the linear manifold we would now have a non-linear or curved manifold $\bar{\mathcal{n}} \subset \mathcal{M}$. The adjustment problem can then be stated as the problem of finding that point on the curved manifold $\bar{\mathcal{n}}$ which has least distance to the given sample point $y_s \in \mathcal{M}$. Questions that arise immediately are: How many of such points exist and how do we locate them? It will be clear that the non-linear problem is drastically more complex than the linear one. In general we could say that solutions to linear problems are prefabricated, where exact solutions to non-linear problems are custom made. To make a start, one might ask oneselves whether a classification of type of manifolds may help to get more insight. It turns out that for totally geodesic manifolds, $\hat{y}$ can be found without recourse to iteration methods. Examples are the 2-dim geodetic triangulation chain and the 2-dim. Helmert transformation. For the so-called ruled-type of manifolds simplifications of the computational process through dimensional reduction are possible. Examples are the 3-dim Helmert transformation, the 2 and 3 dim models for connecting geodetic networks and the 2-dim closed polygon with azimuth and distance observables. Also the properties of so-called canal manifolds could be used to explain how infinite curvature can produce non-injectivity. And finally the theory of complex manifolds (hermitian defferential geometry) may be used for 2-dim polygon adjustments is the complex plane.

As to the distributional properties of the non-linear least-squares estimators it is hard to give general results. However, for their first moments one can show that to an approximation of the order σ^4, the following holds:

$$E\,\{\hat{y}-\tilde{y}\} = \frac{1}{2}\,\sigma^2 n \bar{N} \quad \text{and} \quad E\,\{\hat{x}^\gamma - x^\gamma\} = -\frac{1}{2}\,\sigma^2\, g^{\alpha\beta}\,\Gamma^\gamma_{\alpha\beta}$$

where: $n = \dim \bar{\mathcal{n}}$; $\bar{N}$ is the unique mean curvature normal of $\bar{\mathcal{n}}$ at $\tilde{y}$, and $\Gamma^\gamma_{\alpha\beta}$ are the Christoffel symbols of the second kind.

Note that it follows from the above, that the bias in $\hat{y}$ depends on the extrinsic curvature behaviour of $\bar{\mathcal{n}}$ and is therefore invariant to reparameteriza-

tions, whereas the bias in the parameters $\hat{x}^\gamma$ is given by the average trace of the induced connection.

Despite that classifications of manifolds are useful, usually the methods of location are still iterative in nature. One can discriminate between local and global methods of location. If we restrict ourselves to immersions (the iterative methods for submersions are rather complex), the iterative scheme

$$x_{q+1}^\beta = x_q{}^\beta + \Delta x_q{}^\beta, \quad \beta = 1, \ldots, \dim \mathcal{n}, \tag{3}$$

for local methods typically consists of the following steps:

i) Set $q=0$. An inital guess $x_0{}^\beta$ in provided either externally or through the use of observational data.

ii) Determine an incremental vector $\Delta x_q{}^\beta$ in the direction of the proposed step

iii) Test whether termination criterion is met. If so, accept x_{q+1}^β as $\hat{x}$. If not, increase q by one and return to (ii).

Questions that arise are: when does the iteration method converge; what it its rate of convergence and when do we have a local minimum of the least-distance problem. Most iteration methods take the following form for determining the incremental vector:

$$\Delta x_{q+1}^\beta = h^{\beta\alpha}(x_q)\, \partial_\alpha y^i(x_q)\, g_{ij}\, (y_s^j - y^j(x_q)),$$

where $y^i(x)$, $i = 1, \ldots \dim \mathcal{m}$, are the coordinate functions of the non-linear map defining the immersion and g_{ij} is the coordinate epression for the Euclidean metric of $\mathcal{m}$.

The so-called Newton method, Levenberg-Marquardt compromise and Gauss method differ in their choise for $h_{\alpha\beta}$:

Newton:

$$h_{\alpha\beta}(x_q) = g_{\alpha\beta}(x_q) + \partial^2_{\alpha\beta} y^i(x_q)\, g_{ij}\, (y_s^j - y^j(x_q)),$$

where $g_{\alpha\beta}$ is the induced metric tensor of the parameter space $\mathcal{n}$.

L-M compromise:

$$h_{\alpha\beta}(x_q) = g_{\alpha\beta}(x_q) + \mu\delta_{\alpha\beta},$$

where μ is a scalar still to be determined.

Gauss:

$$h_{\alpha\beta}(x_q) = g_{\alpha\beta}(x_q).$$

We will restrict ourselves to the last method, not only since it can be considered

as the natural generalization of the linear case, but also since it takes full advantage of the distance structure of the adjustment problem. Moreover, it can be shown that this method is preeminently suited for small residual adjustment problems and moderately curved manifolds. Some results are:

1. A necessary condition for convergence is that the length of the least-squares residual vector should be less than the reciprocal of the in absolute value maximum principal extrinsic curvature of the manifold $\bar{n}$ at $\hat{y}$.
2. The local convergence behaviour is linear and the rate of convergence is given by

$$\lim_{q\to\infty} \frac{(x_{q+1}^{\alpha}-\hat{x}^{\alpha})\, g_{\alpha\beta}(\hat{x})\,(x_{q+1}^{\beta}-\hat{x}^{\beta})}{(x_q{}^{\alpha}-\hat{x}^{\alpha})\, g_{\alpha\beta}(\hat{x})\,(x_q{}^{\beta}-\hat{x}^{\beta})} \leq k^2 \,\|\hat{y}-y_s\|^2 ,$$

   where k is the in absolute value maximum principal extrinsic curvature of $\bar{n}$ at $\hat{y}$.
3. The convergence behaviour is invariant to admissable parameter transformations.
4. $\hat{y}$ is a strict local minimum of the least-diastance problem if $\|\hat{y}-y_s\|$ is less than the reciprocal of the largest principal extrinsic curvature of $\bar{n}$ at $\hat{y}$.
5. If either the extrinsic curvatures are zero at $\hat{y}$ or the residual vector at $\hat{y}$ is zero, then local convergence is quadratic with a rate determined by the Christoffel symplols of the second kind.

For the global variant of Gauss method we have instead of (3) that

$$x_{q+1}^{\beta} = x_q{}^{\beta} + t_q\, \Delta x_q{}^{\beta} ,$$

where the scalar t_q is chosen, such that the length of the residual vector decreases in each iteration step. One can show that the optimal choice for t_q would be

$$t_q = \frac{2}{2 - (k^1+k^n)\,\|\hat{y}-y_s\|} , \tag{4}$$

where k^1 and k^n are respectively the largest and smallest principal extrinsic curvatures of $\bar{n}$ at $\hat{y}$. Note that equation (4) shows that for moderately curved manifolds, $t_q = 1$ is a close to optimal choice. Also note that if in case of Gauss method the scalar t_q is taken infinitesimally small in each iteration step we would get the autonomous dynamical system of first order differential equations

$$\frac{dx^{\beta}}{dt} = -\left(\mathrm{grad}\cdot \frac{1}{2}\,\|y_s-y(x)\|^2\right)^{\beta} .$$

This shows that in principle numerical integration techniques can be applied to solve our adjustment problem. Open questions are then how to apply:

1. the qualitative theory of global behaviour of dynamical systems, which is concerned with the existence of equilibrium behaviour of dyn. systems together with questions of local and global stability;
2. Morse theory (Morse inequalities), which studies the equilibrium configuration of a gradient system; and
3. Singularity theory, which studies the critical points of smooth functions.

Discussion

SCHWARZ: Could you classify some cases in our geodetic problems, where we could express non-linear or large curvature in a wrong way? I am talking about cases where we need very good approximate values for the parameters in order to ensure convergence.

TEUNISSEN: One such case is the ellipsoidal one (the reduction of the observations on the ellipsoidal surface), but here the curvature is very small. Some problems can be solved with the conventional reduction, some others analytically. I can not yet identify geodetic problems with severe curvature conditions. In general, it would be difficult to give such a classification.

HEIN: Peter Meissl used to say that we never have to go to a non-linear adjustment, since we always have some good apriori information to be used for a Gaussian iteration.

TEUNISSEN: In 90 percent of the cases, when we have to perform an iteration, everything works all right. In the rest of the cases, we want to know what went wrong, and we try to understand a little better the non-linear physical problem itself.

DERMANIS: I would like to get into the distribution problem, which is critical for hypothesis testing. It is difficult to go from something which is originally normally distributed to something else which is not normally distributed, or which is approximately normally distributed but assumed to be so. In reallity, the original observations are never normally distributed.

TEUNISSEN: We should try to find-out what is the effect of such an assumption on normally distributed observations on the final results.

DERMANIS: If we forget the usual observation equations and we go into the condition equations, or to the combined model, there are always alternative ways to write the equations. For non linear example, a square root can be

eliminated by raising to the second power, a denominator term can be transfered to the nominator at the other side of an equation, etc. There are several alternatives by which the linearization leads to different results. The problem is to choose the best among equivallent forms. A very good example comes from an experiment when we tried to fit a surface to a dome-like structure. The square of the distance did not work in our observation equations, but its square root did. The problem was coming from the non-linearity of the equations. I think that what we need are some criteria for choosing the best form among possible alternatives.

TEUNISSEN: There are indeed several alternatives by which linearization leads to a different situation numerically. However, all these forms are equivalent in the sense that they have the same curvature. Hence, the performance of Gauss' method is in theory not affected by these changes.