manuscripta geodaetica (1990) 15:137–150

manuscripta geodaetica

Nonlinear least squares

P. J. G. Teunissen

Geodetic Computing Centre, Faculty of Geodesy, Delft University of Technology, Thijseweg 11, NL-2629 JA Delft, The Netherlands

Received October 17, 1989; Accepted March 19, 1990

Abstract

This paper discusses the numerical characteristics of a number of iterative algorithms for solving nonlinear least squares problems. The methods discussed all belong to the class of iterative descent methods. The basic principles of these methods are discussed, necessary and sufficient conditions of convergence are given, and the rates of convergence of these methods are derived. Particular attention is given to the Gauss-Newton method. This method is especially suited for solving (small residual) nonlinear least squares problems since it takes avantage of the "sum of squares" structure of the objective function to be minimized. The differential geometric approach for nonlinear adjustment of (Teunissen, 1985) is applied to facilitate a geometric interpretation of the numerical characteristics of the Gauss-Newton method.

1 Introduction

The numerical estimation of parameters is typically a problem of optimization. The estimation of parameters requires frequently the maximization or minimization of an objective function. Typical objective functions are risk functions, robust loss functions, posterior density functions, likelihood functions and (weighted or unweighted) sums of squares. Of these, the two most common methods of estimation are maximum likelihood and least squares. In maximum likelihood, the estimates of the parameters are taken as those values that maximize the likelihood function given the data. Thus if $p_{\underline{y}}(y \mid x)$ is the density of the random data vector $\underline{y}$ (the underscore indicates randomness), the optimization problem of maximum likelihood reads

$$\max_x p_{\underline{y}}(y \mid x). \tag{1}$$

In general no *direct* methods exist for solving (1) when the parameter x enters to the third or higher power in $p_{\underline{y}}(y \mid x)$. For these cases one will therefore have to take recourse to computational techniques that are *iterative* in nature. That is, one starts with an initial guess x_0 of the solution $\hat{x}$ and then proceeds to generate according to some preassigned rule a sequence $x_1, x_2, x_3, \ldots$ that hopefully converges to $\hat{x}$. Various iterative techniques exist which can be used to solve a nonlinear optimization problem like (1) (Ortega and Rheinboldt, 1970). In view of a comparison with the Gauss-Newton method we present the Steepest-Ascent (Descent) method, the Newton method and the Trust-Region method. The basic principles of these methods are discussed, sufficiency conditions of convergence are given, and the rates of convergence of these methods are derived.

In most geodetic applications it is customary to assume that the random m-vector $\underline{y}$ has a multivariate normal (or Gaussian) distribution

$$p_{\underline{y}}(y \mid x) = (2\pi)^{-m/2} \mid Q_y \mid^{-1/2} exp[-\frac{1}{2} \parallel y - A(x) \parallel^2]$$

with $\parallel . \parallel^2 = (.)^* Q_y^{-1}(.)$ and $A(.) : R^n \rightarrow R^m, m \geq n$. In this case the maximization problem (1) can be turned into the minimization problem

$$\min_x \parallel y - A(x) \parallel^2 . \tag{2}$$

This is the least squares problem.

The minimization problem (2) can be solved directly if the map $A(.)$ is *linear*, i.e. if $A(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 A(x_1) + \alpha_2 A(x_2)\ \forall \alpha_1, \alpha_2 \in R, x_1, x_2 \in R^n$. The corresponding linear least squares estimators are given by the well-known formulae

$$\begin{array}{ll} \hat{\underline{y}} = P_A \underline{y} & \hat{\underline{e}} = P_A^{\perp} \underline{y} \\ \hat{\underline{x}} = A^- P_A \underline{y} & \parallel \hat{\underline{e}} \parallel^2 = \parallel P_A^{\perp} \underline{y} \parallel^2 \end{array} \tag{3}$$

where: P_A is the orthogonal projector that projects onto the range of A and along its orthogonal complement; $P_A^{\perp} = I - P_A$; and A^- is an (arbitrary) inverse of A. The estimators $\hat{\underline{y}}$, $\hat{\underline{e}}$ and $\parallel \hat{\underline{e}} \parallel^2$ are unique, and the estimator $\hat{\underline{x}}$ is unique if and only if the map A has full rank n.

In: Manuscripta Geodaetica, Vol. 15, No. 3, pp. 137-150.

If the map $A(.)$ is *nonlinear* then generally no direct methods exist for solving (2). In this case one has to fall back on iterative techniques. One can in principle solve (2) with one of the iterative techniques mentioned earlier. These methods however do not take advantage of the special structure of the objective function of (2). The Gauss-Newton method on the other hand does take advantage of the sums of squares structure of the objective function. The method is therefore especially suited for solving (small residual) nonlinear least squares problems. The numerical characteristics of the Gauss-Newton method are discussed using concepts from differential geometry (Teunissen, 1984). The concept of normal curvature is introduced and it is shown how the optimality conditions and the rate of convergence of the Gauss-Newton method can be expressed in terms of the normal curvatures of the manifold $A(x)$. A parameter invariant convergency criterion is formulated and bounds on the error magnitudes of the computed quantities are derived in terms of curvature measures.

Differential geometric concepts have also been used in studies on nonlinear problems by [Blaha and Besette, 1989; Borre and Lauritzen, 1989; Grafarend and Schaffrin, 1989]. Nondifferential geometric numerical studies of nonlinear geodetic inverse problems can be found in [Kubik, 1967; Saito, 1973; Schek and Maier, 1976; Pope, 1982; Bähr, 1988].

2 Iterative Descent Methods

Consider the problem of finding (local or global) solutions to the problem:

$$\min_x F(x) \quad , \quad x \in R^n \quad , \quad F : R^n \rightarrow R. \tag{4}$$

The methods that will be discussed for solving the minimization problem (4) are all iterative descent algorithms. By *iterative*, we mean, that the algorithm generates a sequence of points, each point being calculated on the basis of the points preceding it. An iterative algorithm is initiated by specifying a starting point, the initial guess. By *descent*, we mean that as each new point is generated by the algorithm the corresponding value of $F(x)$ evaluated at the most recent point decreases in function value. Ideally, the sequence of points generated by the algorithm in this way converges in a finite or infinite number of steps to a solution of (4).

The methods discussed all adhere to the following scheme:

$$x_{k+1} = x_k + t_k d_k \quad , \quad k = 0, 1, 2, \ldots \tag{5}$$

i Set $k = 0$. An initial guess is provided externally.

ii Direction generation: Determine a direction vector d_k in the direction of the proposed step.

iii Line search strategy: Determine a positive scalar t_k such that $F(x_{k+1}) \leq F(x_k)$.

iv Test whether the termination criterion is met. If so, accept x_{k+1} as the solution of (4). If not, increase k by one and return to step ii.

Generally one can say that the individual methods falling under (5) differ in their choice of the direction-vector d_k and the scalar t_k. The iterative techniques fall roughly into two classes: direct search methods and gradient methods. Direct search methods are those which do not require the explicit evaluation of any partial derivatives of the function $F(x)$, but instead rely solely on values of the objective function $F(x)$, plus information gained from the earlier iterations. Gradient methods on the other hand are those which select the direction vector d_k using values of the partial derivatives of the objective function $F(x)$ with respect to the independent variables, as well as values of $F(x)$ itself, together with information gained from earlier iterations. The required derivatives, which for some methods are of order higher than the first, can be obtained either analytically or numerically using some finite difference scheme. This latter approach necessitates extra function evaluations close to the current point x_k, and effectively reduces a gradient method to one of direct search.

We will restrict ourselves to gradient methods for which the required derivatives can be obtained analytically. The descent methods that will be discussed are: the Steepest Descent method, Newton's method and the Trust Region method. But before discussing these methods we first develop the general structure of iterative descent methods.

The direction vector d_k of (5) is said to be in a *descent direction* if a positive scalar t_k exists such that

$$F(x_k + t_k d_k) < F(x_k). \tag{6}$$

If we apply Taylor's expansion to $F(x_k + t_k d_k)$ at x_k we get

$$F(x_k + t_k d_k) = F(x_k) + t_k \partial_x F(x_k)^* d_k + o(t_k).$$

The Landau order term $o(t)$ indicates the remainder in the Taylor series. It has the property $\lim_{t \rightarrow \infty} o(t)/t = 0$. This shows that if

$$\boxed{\partial_x F(x_k)^* d_k < 0} \tag{7}$$

then it is possible to choose a positive scalar t_k so that (6) holds. Direction vectors d_k that satisfy inequality (7) are thus vectors that lie in the direction of descent. The various descent directions at x_k of the function $F(x)$ are shown in figure 1. It follows from inequality (7) that the descent direction vectors can be represented as

$$d_k = -Q(x_k) \partial_x F(x_k) \tag{8}$$

Figure 1: Contours of $F(x)$ and descent directions at x_k.

where $Q(x_k)$ is an arbitrary but positive-definite matrix that may depend on x_k.

Substitution of (8) into (5) shows that the descent methods take the general form

$$\boxed{x_{k+1} = x_k - t_k Q(x_k) \partial_x F(x_k)} \tag{9}$$

The two variables in (9) are the positive scalar t_k and the positive-definite matrix $Q(x_k)$. Different choices for t_k and $Q(x_k)$ correspond with different descent algorithms.

If we define a vector function $\Phi : R^n \rightarrow R^n$ as

$$\Phi(x) = x - t(x) Q(x) \partial_x F(x), \tag{10}$$

with the scalar line search function t(x), equation (9) can be written in the compact form

$$\boxed{x_{k+1} = \Phi(x_k)} \tag{11}$$

Note that since $t(x)$ is positive and $Q(x)$ is positive-definite, the solutions of $x = \Phi(x)$, the socalled *fixed points* of $\Phi(x)$, are identical to the solutions of $\partial_x F(x) = 0$, i.e. the stationary points of $F(x)$. This implies that if the sequence generated by (11) converges to a fixed point of $\Phi(x)$, the sequence generated by the descent method (9) will converge to a stationary point of $F(x)$.

The iteration scheme (11) is known as the *fixed point iteration method*. It is sometimes also called the method of successive approximation and also Picard's method. The following theorem gives sufficient conditions for the fixed point method to convergence to the unique solution of $x = \Phi(x)$.

Theorem: Fixed point iteration
Let Ω be a set of $R^n, \Omega \subset R^n$.
Assume that

i $\Phi(x) \in \Omega$, $\forall x \in \Omega$

ii $\Phi(x)$ is continuous $\forall x \in \Omega$

iii $\| \Phi(x_2) - \Phi(x_1) \| \leq c \| x_2 - x_1 \|$,
$0 \leq c < 1$, $\forall x_1, x_2 \in \Omega$.

Then:

1. A solution $\hat{x}$ of $x = \Phi(x)$ exists in Ω
2. The solution $\hat{x}$ is unique
3. The fixed point algorithm converges to $\hat{x}$, that is $\lim_{k \rightarrow \infty} x_k = \hat{x}$

Proof: see e.g. (Naylor and Sell, 1982).

Although the theorem gives sufficiency conditions for the guaranteed convergence of the sequence $x_{k+1} = \Phi(x_k), k = 0, 1, 2, 3, \ldots$, to a unique fixed point, its usefulness in practical applications is unfortunately rather limited. This is due to the difficulty one has in practical applications with verifying the sufficiency conditions. Especially the verification of the inequality condition iii for all pairs of vectors in Ω is most difficult. This task becomes somewhat simpler if we may assume that $\Phi(x)$ has continuous partial derivatives and that Ω is convex. With the *mean value theorem* follows then that

$$\begin{aligned} \| \Phi(x_2) - \Phi(x_1) \| &= \| \partial_x \Phi(\bar{x})(x_2 - x_1) \| \\ &\leq \| \partial_x \Phi(\bar{x}) \| \| x_2 - x_1 \| \end{aligned}$$

with $\bar{x} = x_1 + t(x_2 - x_1), 0 \leq t \leq 1$. This result implies that we may check condition iii of the theorem by verifying whether

$$c = \max_{x \in \Omega} \| \partial_x \Phi(x) \| < 1. \tag{12}$$

With this result we are now also able to formulate more tractable convergency conditions for the class of descent methods (9). By taking the partial derivatives of (10) we get

$$\partial_x \Phi(x) = I - \sum_{\alpha=1}^{n} \partial_x q_\alpha(x) \partial_\alpha F(x) - t(x) Q(x) \partial^2_{xx} F(x) \tag{13}$$

where $q_\alpha(x), \alpha = 1, 2, \ldots, n$, are the column vectors of the positive-definite matrix $t(x)Q(x)$. Hence,

$$\begin{aligned} \| \partial_x \Phi(x) \| \leq & \| I - t(x) Q(x) \partial^2_{xx} F(x) \| + \\ & + \sum_{\alpha=1}^{n} \| \partial_x q_\alpha(x) \| | \partial_\alpha F(x) | \end{aligned} \tag{14}$$

This shows that convergence of the descent methods is guaranteed if

$$\boxed{\| I - t(x) Q(x) \partial^2_{xx} F(x) \| < 1} \tag{15}$$

and if the second term on the right hand side of (14) can be made sufficiently small. Since $\partial_x F(\hat{x}) = 0$ and $\partial_x F(x)$ is continuous, then by the very definition of continuity for each $\epsilon > 0$ there exists a $\delta > 0$ such that if $\| x - \hat{x} \| < \delta$, then $\| \partial_x F(x) - \partial_x F(\hat{x}) \| = \| \partial_x F(x) \| < \epsilon$. This implies that the second term on the right hand side

of (14) can be made sufficiently small for a sufficiently small neighborhood of $\hat{x}$. Thus convergence of the descent methods is guaranteed if (15) holds and if the initial guess is *sufficiently close* to the solution $\hat{x}$. The practical problem with the above proof of guaranteed convergence is of course still that one never knows beforehand whether the initial guess is indeed sufficiently close to $\hat{x}$. Nevertheless the above derivation shows clearly what the cause for a possible lack of convergence can be. And it also shows, see (15), how convergence can be enforced by a suitable choice for the scalar $t(x)$.

3 The Steepest Descent Method

The steepest descent method is one of the oldest iterative descent methods for solving a minimization problem. The method goes back to Cauchy (1847). The steepest descent method is characterized by the following simple choice for the positive-definite matrix $Q(x_k)$ of (9):

$$Q(x_k) = I. \tag{16}$$

The steepest descent method takes therefore the form

$$\boxed{x_{k+1} = x_k - t_k \partial_x F(x_k)} \tag{17}$$

The choice (17) is motivated by the fact that the vector $d_k = -\partial_x F(x_k)$ minimizes

$$\frac{\partial_x F(x_k)^* d_k}{(d_k^* d_k)^{\frac{1}{2}}}.$$

Thus within a *linear* approximation, the direction vector $d_k = -\partial_x F(x_k)$ points in the direction of the steepest descent of the function $F(x)$ at x_k.

One of the advantages of the steepest descent methods is its great simplicity. No partial derivatives of $F(x)$ of the order higher than the first are needed and no matrices need to be inverted. A drawback of the method is however that its performance is dependent on the more or less arbitrary choice of the variables x used to define the minimization problem. This can be seen as follows.

Suppose that R is an invertible $n \times n$ matrix. We can represent points in R^n either by the standard vector x or by $\bar{x}$ where $R\bar{x} = x$. The problem of finding x to minimize $F(x)$ is equivalent to that of finding $\bar{x}$ to minimize $G(\bar{x}) = F(R\bar{x})$. Thus using steepest descent, the direction vector in case of minimizing $G(\bar{x})$ will be $\bar{d}_k = -R^* \partial_x F(R\bar{x}_k)$ which in the original variables is $d_k = -RR^* \partial_x F(x_k)$. Thus, we see that if $RR^* \neq I$ the change of variables changes the direction of a search. Hence, a new choice of variables may substantially alter the performance characteristics of the steepest descent method.

Another drawback of the steepest descent method is that it has the tendency to *zig-zag*, when it is combined with an *exact line search strategy* and the contours of the objective function are elongated. An exact line search strategy is a strategy in which the positive scalar t_k is chosen so as to minimize $F(x_k + t_k d_k)$. If t_k is a minimizer of $F(x_k + t_k d_k)$ then

$$0 = \frac{dF}{dt}(t_k) = \partial_x F(x_k + t_k d_k)^* d_k = \partial_x F(x_{k+1})^* d_k.$$

This shows that if an exact line search is used, the successive directions of search, d_k and d_{k+1}, are orthogonal to each other. Hence the steepest descent method will obviously zig-zag when the contours of $F(x)$ are very elongated (see figure 2). The zig-zagging is absent of course when the contours of $F(x)$ are circular. In fact, the steepest descent method with an exact line search will locate the minimum of $F(x)$ in one step if the contours of $F(x)$ are circles (or hyperspheres).

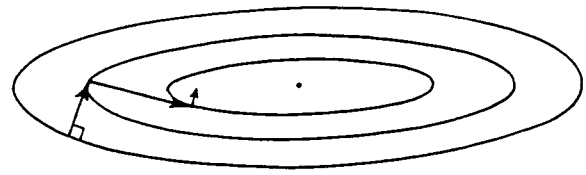

Figure 2: Zig-zagging of the steepest descent method.

An important performance measure of an iteration method is its *rate of convergence*. The rate of convergence of an iterative technique is related to the way the error magnitude at the $(k+1)$th step, $\| x_{k+1} - \hat{x} \|$, is related to the error magnitude in the previous step. The rate of convergence shows therefore whether convergence of an iteration method is rapid enough to make the whole scheme practical.

In order to derive the rate of convergence for the steepest descent method we expand (17) into a Taylor series at the solution $\hat{x}$. This gives

$$x_{k+1} - \hat{x} = [I - t_k \partial_{xx}^2 F(\hat{x})](x_k - \hat{x}) + o(\| x_k - \hat{x} \|) \tag{18}$$

If $\hat{x}$ is a local minimizer of $F(x)$ then the matrix $\partial_{xx}^2 F(\hat{x})$ is positive semi-definite and its eigenvalues may be ordered so that

$$0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n.$$

By taking the norm of (18) we therefore get

$$\begin{aligned} \| x_{k+1} - \hat{x} \| \leq max\{| 1 - t_k\lambda_1 |, | 1 - t_k\lambda_n |\}. \\ . \| x_k - \hat{x} \| + o(\| x_k - \hat{x} \|) \end{aligned} \tag{19}$$

This shows that the steepest descent method has a *linear rate of convergence* for points sufficiently close to the solution. Thus for points sufficiently close to the solution the error magnitude gets reduced by a factor $max\{| 1 - t_k\lambda_1 |, | 1 - t_k\lambda_n |\}$ at each iteration step. The closer

this factor is to 1 the slower the rate of convergence; the closer the factor is to 0 the faster the rate of convergence.

If the positive scalar t_k is taken to be equal to one in each iteration step (this is the simplest line search strategy), the rate of convergence of the steepest descent method becomes approximately

$$\boxed{\| x_{k+1} - \hat{x} \| \leq max\{| 1 - \lambda_1 |, | 1 - \lambda_n |\} \| x_k - \hat{x} \|} \tag{20}$$

This shows that the error magnitude gets reduced if the extreme eigenvalues of $\partial^2_{xx} F(\hat{x})$ satisfy

$$0 < \lambda_1, \lambda_n < 2. \tag{21}$$

Hence, see also (15), local convergence cannot be guaranteed if one or more eigenvalues of the positive semi-definite matrix $\partial^2_{xx} F(\hat{x})$ lie outside the open interval (0,2).

The rate of convergence of (20) can be improved and local convergence can be guaranteed, however, if the positive scalar t_k is chosen so as to minimize $max\{| 1 - t_k\lambda_1 |, | 1 - t_k\lambda_n |\}$. It follows from figure 3 that the corresponding optimal choice for t_k is

$$t_k = 2/(\lambda_1 + \lambda_n). \tag{22}$$

With this choice for t_k it follows from (19) that instead of (20) we have

$$\boxed{\| x_{k+1} - \hat{x} \| \leq \left[\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right] \| x_k - \hat{x} \|} \tag{23}$$

Note that if the matrix $\partial^2_{xx} F(\hat{x})$ is positive definite then the factor $(\lambda_n - \lambda_1)/(\lambda_n + \lambda_1)$ is always less than one and local convergence is guaranteed. This factor is close to one if the condition number, λ_n/λ_1, of the matrix $\partial^2_{xx} F(\hat{x})$ is large, i.e. if the contours of $F(x)$ are very elongated near the solution $\hat{x}$.

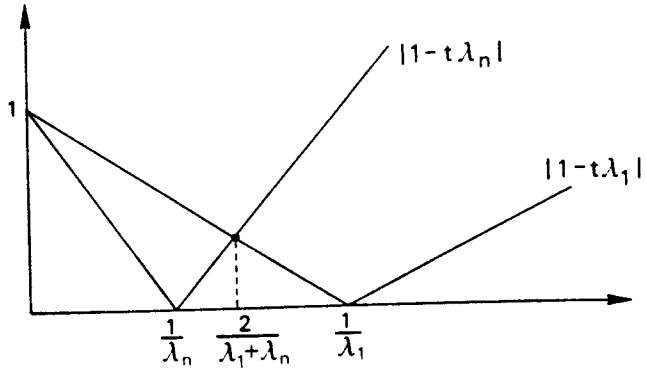

Figure 3: Optimal choice for t is $t = 2/(\lambda_1 + \lambda_n)$.

4 Newton's Method

Newton's method is characterized by the following choice for the positive-definite matrix $t_k Q(x_k)$ of (9):

$$t_k Q(x_k) = [\partial^2_{xx} F(x)]^{-1}. \tag{24}$$

Newton's method takes therefore the form

$$\boxed{x_{k+1} = x_k - [\partial^2_{xx} F(x_k)]^{-1} \partial_x F(x_k)} \tag{25}$$

We will give two motivations for the choice (24). The first one goes back to the basic idea on the basis of which Newton's method was originally introduced. Newton's method was originally conceived as an iterative technique for solving a system of nonlinear equations. The basic idea of the method is best explained for a function $G(x)$ with one variable x. Let the nonlinear equation which needs to be solved be

$$G(x) = 0. \tag{26}$$

The graph of this function is plotted in figure 4.

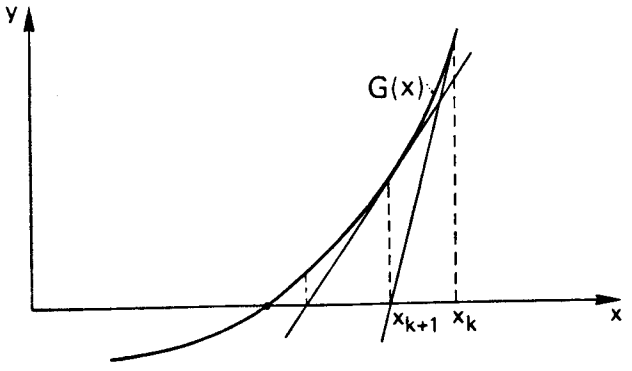

Figure 4: Newton's method: $\frac{dG}{dx}(x_k) = G(x_k)/(x_k - x_{k+1})$.

At a given point x_k the graph of the function $G(x)$ is approximated by its tangent, and an approximate solution to equation (26) is taken to be the point x_{k+1} where the tangent crosses the x-axis. The process is then repeated from this new point. This procedure defines a sequence of points according to the recurrence relation

$$x_{k+1} = x_k - [d_x G(x_k)]^{-1} G(x_k). \tag{27}$$

If we replace $G(x)$ in (26) by $d_x F(x)$, the recurrence relation for determining a stationary point of $F(x)$ becomes

$$x_{k+1} = x_k - [d^2_{xx} F(x)]^{-1} d_x F(x). \tag{28}$$

In order to generalize this result to the multivariate case, note that in the above procedure the original nonlinear equation, $G(x) = 0$ or $d_x F(x) = 0$, is linearized about the point x_k and then solved for x_{k+1}. If we apply this procedure to the system of nonlinear equations $\partial_x F(x) = 0$, linearization gives

$$0 = \partial_x F(x_k) + \partial^2_{xx} F(x_k)(x_{k+1} - x_k) \tag{29}$$

from which x_{k+1} follows as (25).

Since Newton's method is based on a linearization of $\partial_x F(x)$, one can interpret the method as one that computes the minimum of a *quadratic approximation* of $F(x)$ at each iteration step. This shows the distinct difference with the steepest descent method. The steepest descent method is namely based on a *linear approximation* of $F(x)$ at each iteration step. As a consequence, if the function $F(x)$ is quadratic, Newton's method will locate the minimum in one iteration step, whereas the steepest descent method needs in general an infinite number of iteration steps.

The second motivation for the choice (24) is based on inequality (15). Note that with (24) inequality (15) is trivially fulfilled, which implies that Newton's method has a guaranteed convergence for points sufficiently close to the solution. Thus, contrary to the steepest descent method no line search is needed to enforce local convergence.

In order to derive the rate of convergence for Newton's method we expand (25) into a Taylor series at the solution $\hat{x}$. This gives

$$x_{k+1} - \hat{x} = -\frac{1}{2}(x_k - \hat{x})^*[\partial_x[\partial^2_{xx}F(\hat{x})]^{-1}\partial^2_{xx}F(\hat{x})]. \\ .(x_k - \hat{x}) + o(\| x_k - \hat{x} \|^2)$$

or,
with $[\partial^2_{xx}F(\hat{x})]^{-1}\partial^3_{xxx}F(\hat{x}) = -\partial_x[\partial^2_{xx}F(\hat{x})]^{-1}\partial^2_{xx}F(\hat{x})$,

$$\boxed{\begin{array}{l} x_{k+1} - \hat{x} = \\ \frac{1}{2}(x_k - \hat{x})^* \left[[\partial^2_{xx}F(\hat{x})]^{-1}\partial^3_{xxx}F(\hat{x})\right](x_k - \hat{x}) \\ \qquad + o(\| x_k - \hat{x} \|^2) \end{array}} \quad (30)$$

This shows that Newton's method has a *quadratic* rate of convergence.

Although the information requirements associated with the evaluation, storage and inversion of the matrix $\partial^2_{xx}F(x)$ as required by Newton's method are rather heavy, the method has proved, due to its guaranteed local convergence and quadratic rate of convergence, to be extremely effective in dealing with general minimization problems. Difficulties with Newton's method occur however when the matrix $\partial^2_{xx}F(x)$ is non-invertible or when it fails to be positive definite. These difficulties can be overcome by using a so-called trust region method. This method, which can be considered as a regularized version of Newton's method, will be discussed in the next section.

5 The Trust Region Method

The trust region method was introduced by Levenberg (1944) for nonlinear least squares, modified by Marquardt (1963) and further developed and generalized by Goldfeld, Quandt and Trotter (1966). The method is characterized by the following choice for the positive definite matrix $t_k Q(x_k)$ of (9):

$$t_k Q(x_k) = [\partial^2_{xx}F(x_k) + \alpha_k R]^{-1} \quad (31)$$

where α_k is a non-negative scalar and R is a positive definite matrix. The trust region method takes therefore the form

$$\boxed{x_{k+1}(\alpha_k) = x_k - [\partial^2_{xx}F(x_k) + \alpha_k R]^{-1}\partial_x F(x_k)} \quad (32)$$

This formula already shows some of the basic ideas underlying the trust region method. Since matrix R is positive definite by assumption, a sufficiently large α_k ensures the positiveness of (31). Thus by adjusting α_k, a possible lack of positive definiteness of $\partial^2_{xx}F(x_k)$ can be circumvented and a descent direction can be generated. Furthermore note that for $R = I$, the trust region method can be interpreted as a *compromise* between Newton's method and the method of steepest descent. For $\alpha_k = 0$, we get

$$x^N_{k+1} = x_k - [\partial^2_{xx}F(x_k)]^{-1}\partial_x F(x_k) \quad (33)$$

which is Newton's method, and for large α_k we have approximately

$$x^{sd}_{k+1} = x_k - \alpha_k^{-1}\partial_x F(x_k) \quad (34)$$

which is the steepest descent method with α_k^{-1} playing the role of the line search scalar t_k. Thus the direction of search of the trust region method interpolates between the Newton direction and the steepest descent direction (see figure 5). Since Newton's method is based on a quadratic approximation of $F(x)$ and the method of steepest descent is based on a linear approximation, it seems that with the trust region method one can, by adjusting α_k, control the approximation used for $F(x)$.

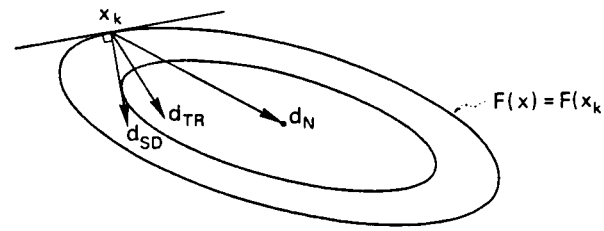

Figure 5: d_N = Newton direction, d_{SD} = Steepest Descent direction, d_{TR} = Trust Region direction.

In order to get a better understanding of this phenomenon, let us study the different approximations involved. We start from the following Taylor series expansion of $F(x)$:

$$F(x) = a(x) + o(\| x - x_k \|^2) \quad (35)$$

with

$$a(x) = F(x_k) + \partial_x F(x_k)^*(x - x_k) + \\ + \frac{1}{2}(x - x_k)^* \partial^2_{xx} F(x_k)(x - x_k) \qquad (36)$$

As we know Newton's method is based on the approximation $a(x)$ of $F(x)$. The Newton solution x^N_{k+1} follows then from minimizing (provided this is possible) the quadratic function $a(x)$. Thus $a(x^N_{k+1}) \leq a(x) \forall x \in R^n$ and $a(x^N_{k+1}) < a(x_k) = F(x_k)$ if $x^N_{k+1} \neq x_k$. From this and $F(x^N_{k+1}) = a(x^N_{k+1}) + o(\| x^N_{k+1} - x_k \|^2)$ follows that the objective function gets reduced, i.e. $F(x^N_{k+1}) < F(x_k)$, if $o(\| x^N_{k+1} - x_k \|^2)$ is small enough. Thus descent of the objective function occurs when $a(x^N_{k+1})$ can still be considered an *adequate approximation* of $F(x^N_{k+1})$.

Problems may occur however when this approximation is not adequate, that is, when $o(\| x^N_{k+1} - x_k \|^2)$ is too large. This may happen if matrix $[\partial^2_{xx} F(x_k)]^{-1}$ of (33) is "large", i.e. when the matrix $\partial^2_{xx} F(x_k)$ is poorly conditioned and thus the contours of $a(x)$ are very elongated. If the approximation of $F(x)$ by $a(x)$ is inadequate one can improve the approximation by restricting the region for x. In the steepest descent method this is achieved by replacing $x - x_k$ in (35) and (36) by $t_k d_k$, by taking d_k in the direction of steepest descent, $-\partial_x F$, and then by adjusting t_k so that $\| x^{sd}_{k+1} - x_k \| = \| t_k d_k \|$ is sufficiently small and $F(x^{sd}_{k+1}) = F(x_k + t_k d_k) < F(x_k)$ holds. This idea of a line search along the direction vector d_k to restrict the region of x by validity of the Taylor approximation can in principle also be applied to Newton's method. Instead of (33) one gets then

$$x^N_{k+1} = x_k - t_k [\partial^2_{xx} F(x_k)]^{-1} \partial_x F(x_k). \qquad (37)$$

By taking t_k sufficiently small one can then again ensure that $F(x^N_{k+1}) < F(x_k)$, provided that $\partial^2_{xx} F(x_k)$ is positive-definite. The problem with this modification of Newton's method is however that it cannot deal with those cases where $\partial^2_{xx} F(x_k)$ lacks positive definiteness or is singular. The basic idea of the trust region method is now to replace the one dimensional steepest descent-like restriction along a fixed direction d_k by an n-dimensional restricted *region* for x. That is, in the trust region method again the quadratic approximation (36) is used, but now with the additional restriction that x should lie within an ellipsoidal region for which $a(x)$ is *trusted* to be an adequate approximation of $F(x)$ (see figure 6).

If we define the quadratic form

$$A(x) = a(x) + \frac{1}{2}\alpha_k (x - x_k)^* R (x - x_k)$$

it follows that if $\partial^2_{xx} F(x_k) + \alpha_k R$ is positive definite then

$$A(x_{k+1}(\alpha_k)) \leq A(x) \ \ \forall x \in R^n.$$

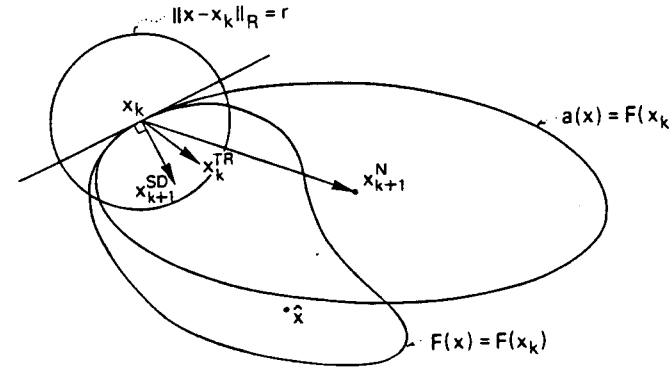

Figure 6: The trust region:
$\| x - x_k \|_R = [(x - x_k)^* R (x - x_k)]^{\frac{1}{2}} \leq r.$

This implies that

$$a(x_{k+1}(\alpha_k)) \leq a(x)$$

$$\forall x \in \{x \mid \| x - x_k \|_R \leq \| x_{k+1}(\alpha_k) - x_k \|_R\}$$

Thus $x_{k+1}(\alpha_k))$ of (32) minimizes $a(x)$ over the ellipsoid $\| x - x_k \|_R \leq \| x_{k+1}(\alpha_k) - x_k \|_R$.

The "radius" $r(\alpha_k) = \| x_{k+1}(\alpha_k) - x_k \|_R$ is a decreasing function of α_k. In order to show this, we consider the problem

$$\partial^2_{xx} F(x_k) e_i = \lambda_i R e_i \ , \ i = 1, \ldots, n. \qquad (38)$$

The eigenvectors e_i , $i = 1, \ldots, n$, form a basis of R^n and they can be chosen so that

$$e_i^* R e_j = \delta_{ij} \ , \ i, j = 1, \ldots, n. \qquad (39)$$

Suppose that

$$\partial_x F(x_k) = \sum_{i=1}^{n} c_i R e_i \neq 0.$$

Then with (38),

$$x_{k+1}(\alpha_k) - x_k = -[\partial^2_{xx} F(x_k) + \alpha_k R]^{-1} \partial_x F(x_k) \\ = -\sum_{i=1}^{n} \frac{c_i}{\lambda_i + \alpha_k} e_i$$

and thus with (39)

$$r(\alpha_k) = \| x_{k+1}(\alpha_k) - x_k \|_R = \left[\sum_{i=1}^{n} \left(\frac{c_i}{\lambda_i + \alpha_k} \right)^2 \right]^{1/2} \qquad (40)$$

This shows the monotone decreasing property of $r(\alpha_k)$ if $\lambda_i + \alpha_k > 0$, i.e. if $\partial^2_{xx} F(x_k) + \alpha_k R$ is positive definite. The trust region is therefore an expanding ellipsoid if α_k gets smaller, and a contracting ellipsoid if α_k gets larger.

The trust region method operates now as follows. At the kth-iteration the point $x_{k+1}(\alpha_k)$ of (32) is computed for a certain $\alpha_k > 0$. Then the actual reduction $F(x_{k+1}(\alpha_k)) - F(x_k)$ is compared with the predicted reduction $a(x_{k+1}(\alpha_k)) - a(x_k) = a(x_{k+1}(\alpha_k)) - F(x_k)$. If the prediction is poor, the parameter α_k is increased in order to contract the trust region, and the computations are repeated; otherwise $x_{k+1}(\alpha_k)$ is accepted as the new iteration point.

6 The Gauss-Newton Method

We will now consider the nonlinear least squares problem, i.e. the minimization problem in which the objective function is a weighted sum of squared terms

$$F(x) = \frac{1}{2} \parallel y - A(x) \parallel^2 \tag{41}$$

where $\| \cdot \|^2 = (.)^* Q_y^{-1}(.)$; Q_y is positive definite; y is an m-dimensional data vector, and $A(.)$ is a nonlinear vector function or map from R^n into R^m. The factor $\frac{1}{2}$ in (41) is merely introduced for convenience.

For varying values of x, $A(x)$ traces locally an n-dimensional surface or manifold embedded in R^m. If the metric of R^m is described by the positive definite matrix Q_y^{-1}, the scalar $\| y - A(x) \|$ equals the distance from point y to the point $A(x)$ on the manifold. Hence, the problem of minimizing $F(x)$ corresponds to the problem of finding that point on the manifold, say $\hat{y} = A(\hat{x})$, which has least distance to y. This geometry of the nonlinear least squares problem is sketched in figure 7.

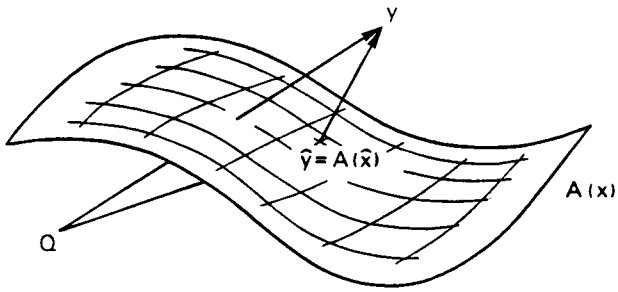

Figure 7: Geometry of nonlinear least squares.

The minimizer of (41) can in principle be located by one of the iterative descent methods of the previous sections. Since

$$\partial_x F(x) = -\partial_x A(x)^* Q_y^{-1} e(x), \tag{42}$$

with $e(x) = y - A(x)$, the steepest-descent method takes the form

$$\boxed{x_{k+1} = x_k + t_k \partial_x A(x_k)^* Q_y^{-1} e(x)} \tag{43}$$

And since

$$\partial_{xx}^2 F(x) = \partial_x A(x)^* Q_y^{-1} \partial_x A(x) - e(x)^* Q_y^{-1} \partial_{xx}^2 A(x)$$

Newton's method takes the form

$$\boxed{\begin{aligned} x_{k+1} = x_k + [&\partial_x A(x_k)^* Q_y^{-1} \partial_x A(x_k) + \\ &- e(x)^* Q_y^{-1} \partial_{xx}^2 A(x_k)]^{-1} \partial_x A(x_k)^* Q_y^{-1} e(x) \end{aligned}} \tag{44}$$

Although the steepest-descent method and Newton's method are iterative methods that can locate the minimizer of (41), they do not take advantage of the special structure of the objective function (41). A method which does take advantage of the "sum of squares" structure of the objective function is the Gauss-Newton method. This method is therefore especially suited for solving nonlinear least squares problems.

The Gauss-Newton method belongs to the same class of iterative descent methods as the steepest-descent method, Newton's method and the trust-region method. The method is characterized by the following choice for the positive-definite matrix $Q(x_k)$ of (9)

$$Q(x_k) = [\partial_x A(x_k)^* Q_y^{-1} \partial_x A(x_k)]^{-1}. \tag{45}$$

With (42) follows therefore that the Gauss-Newton method takes the form

$$\boxed{\begin{aligned} x_{k+1} = x_k + t_k [&\partial_x A(x_k)^* Q_y^{-1} \cdot \\ &\cdot \partial_x A(x_k)]^{-1} \partial_x A(x_k)^* Q_y^{-1} e(x) \end{aligned}} \tag{46}$$

Note that the Newton direction reduces to the Gauss-Newton direction if $\partial_{xx}^2 F(x)$ is replaced by $[\partial_{xx}^2 F(x) + \alpha(x) n^* Q_y^{-1} \partial_{xx}^2 A(x)]$, with $n = e(x)/ \| e(x) \|$ and $\alpha = \| e(x) \|$. But the particular choice (45) is perhaps best motivated if we draw a parallel with *linear* least squares problems. A least squares problem is said to be linear if the map $A(x)$ is linear. If $A(x)$ is linear, the minimizer of $F(x)$ follows from solving a system of linear equations. Thus the nonlinearity of $\partial_x F(x)$ is due to the nonlinearity of $A(x)$. The idea is therefore to approximate $F(x)$ by a function which is obtained by replacing $A(x)$ in $\| y - A(x) \|$ by its linearized version $A(x_k) + \partial_x A(x_k)(x - x_k)$. Hence, instead of using a Taylor-expansion of $F(x) = \| y - A(x) \|$ with first or second order terms as is done in case of the steepest descent method or Newton's method, one approximates $F(x)$ through a linearization *within* the norm. The resulting approximation

$$\| y - A(x_k) - \partial_x A(x_k) d_k \|$$

is then minimized as function of d_k. This gives the solution $d_k = -Q(x_k) \partial_x F(x_k)$. Thus the Gauss-Newton direction d_k can be seen as the solution of a linear(ized) least squares problem. The geometry of the Gauss-Newton method is therefore also one of *orthogonal projection*. That is, the vector $\partial_x A(x_k) d_k$, which lies in the tangent space of the manifold $A(x)$ at $A(x_k)$, is the orthogonal projection of the residual vector $e(x) = y - A(x)$ onto this tangent space (see figure 8).

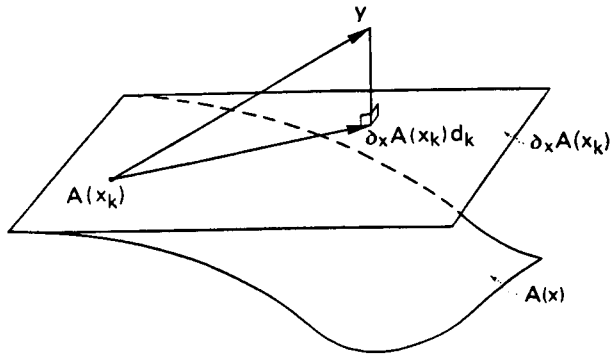

Figure 8: Orthogonal projection onto tangent space of $A(x)$ at $A(x_k)$.

This geometric interpretation of the Gauss-Newton method already makes intuitively clear that the geometry of the manifold $A(x)$ must play an important role in the local behaviour of the method. The role of the geometry of the manifold will be made precise in the following sections.

7 Geometry of Optimality Conditions

It is well known that a point $\hat{x}$ is a (local or global) minimum of the objective function $F(x)$ if

$$\begin{array}{ll} a) & \partial_x F(\hat{x}) = 0 \\ b) & \partial^2_{xx} F(\hat{x}) > 0 \end{array}$$

When applied to the objective function $F(x) = \frac{1}{2} \parallel y - A(x) \parallel^2 = \frac{1}{2} \parallel e(x) \parallel^2$ these necessary and sufficient conditions become

$$\begin{array}{ll} a) & \partial_x F(\hat{x}) = -\partial_x A(\hat{x})^* Q_y^{-1} e(\hat{x}) = 0 \\ b) & \partial^2_{xx} F(\hat{x}) = Q(\hat{x})^{-1} - e(\hat{x})^* Q_y^{-1} \partial^2_{xx} A(\hat{x}) > 0 \end{array} \tag{47}$$

Both these conditions can be given an interesting geometric interpretation (Teunissen 1984, 1985). The geometric interpretation of (47a) is rather simple. Equation (47a) states namely that the residual vector $e(x)$ should be *orthogonal* to the tangent space of manifold $A(x)$ at the solution $\hat{x}$. The interpretation of (47b) is somewhat more complicated. In order to interpret (47b) geometrically we first introduce the concept of *normal curvature* [Krarup, 1982; Teunissen, 1984].

The nonlinearity of the manifold can be defined as the deviation of the manifold from its tangent space in the neighborhood of the point of tangency. It is convenient to introduce the function

$$d(x) = n^* Q_y^{-1} [A(x) - A(x_0)], \tag{48}$$

where n is a unitvector normal to the tangent space of the manifold at the point $A(x_0)$. Thus

$$n^* Q_y^{-1} \partial_x A(x_0) = 0 \quad and \quad n^* Q_y^{-1} n = 1. \tag{49}$$

Note that we assume the metric of R^m to be described by the positive-definite matrix Q_y^{-1}.

The function $d(x)$ describes the perpendicular distance along n from the tangent space to the point $A(x)$ on the manifold. If we assume as before that the map $A(x)$ is sufficiently smooth, substitution of its Taylor expansion at x_0 in (48) gives with (49),

$$d(x) = \frac{1}{2} \Delta x^* [n^* Q_y^{-1} \partial^2_{xx} A(x_0)] \Delta x + \cdots \tag{50}$$

The *second fundamental form* of the manifold is defined as the quadratic form

$$\begin{array}{l} II = v^* [n^* Q_y^{-1} \partial^2_{xx} A(x)] v \ , \\ with \quad v \in R^n \ , \quad n \in R(\partial_x A(x))^{\perp} \subset R^m \end{array} \tag{51}$$

Thus for small values of v the function $2d(x)$ can be approximated by the second fundamental form II with errors of third or higher orders in v. A study of the form II will therefore give information about the shape of the manifold $A(x)$ near the point of tangency.

With the second fundamental form it is now a small step to introduce the concept of *normal curvature*. In Gaussian surface theory the normal curvature is defined as the ratio of the second fundamental form and the first fundamental form [Spivak, 1979]. The *first fundamental form* is defined as

$$\begin{array}{l} I = v^* Q(x)^{-1} v \ , \\ with \quad Q(x)^{-1} = \partial_x A(x)^* Q_y^{-1} \partial_x A(x). \end{array} \tag{52}$$

With the first fundamental form one can compute the arclength of a curve in the manifold. The matrix $Q(x)^{-1}$ is known as the induced metric of the manifold. Note that $Q(x)^{-1}$ corresponds to the *normal matrix* of the problem of linearized least-squares inversion.

With (51) and (52) the normal curvature becomes

$$\boxed{k_n(v) = \frac{II}{I} = \frac{v^* [n^* Q_y^{-1} \partial^2_{xx} A(x)] v}{v^* Q(x)^{-1} v}} \tag{53}$$

The extreme values of this ratio are the *principal normal curvatures*. They follow as the eigenvalues of the generalized eigenvalue problem

$$\mid n^* Q_y^{-1} \partial^2_{xx} A(x) - \lambda Q(x)^{-1} \mid = 0.$$

Since in the classical Gaussian surface theory $A(x)$ is a map from R^2 into R^3, the dimension of the range space of $\partial_x A(x)$, $R(\partial_x A(x))$, is two and the dimension of

its orthogonal complement, $R(\partial_x A(x))^\perp$, is one. Thus in the classical case one has just one second fundamental form and two principal normal curvatures. In our case however, $A(x)$ is a map from R^n into R^m. Therefore $dim\, R(\partial_x A(x)) = n$ and $dim\, R(\partial_x A(x))^\perp = m - n$. This implies that in our case the number of principal normal curvatures equals $n.(m-n)$. We will denote the n-number of principal normal curvatures for the normal direction n by

$$k_n^1 \leq k_n^2 \leq \cdots \leq k_n^n. \tag{54}$$

It should be noted that the normal curvature is *invariant* under a change of variables in $A(x)$. This can be seen as follows. Let $x(\bar{x})$ be a one-to-one map from R^n to R^n. Then

$$\begin{aligned} \partial^2_{\bar{x}\bar{x}} A(\bar{x}) &= \partial_{\bar{x}} x^* \partial^2_{xx} A(x(\bar{x})) \partial_{\bar{x}} x + \partial_x A(x(\bar{x})) \partial^2_{\bar{x}\bar{x}} x \\ Q(\bar{x})^{-1} &= \partial_{\bar{x}} x^* Q(x(\bar{x}))^{-1} \partial_{\bar{x}} x \\ v &= \partial_{\bar{x}} x \bar{v}. \end{aligned}$$

Substitution into

$$k_n(\bar{v}) = \frac{\bar{v}^*(n^* Q_y^{-1} \partial^2_{\bar{x}\bar{x}} A(\bar{x}))\bar{v}}{\bar{v}^* Q(\bar{x})^{-1} \bar{v}}$$

shows then, since $n^* Q_y^{-1} \partial_x A(x(\bar{x})) = 0$, that $k_n(\bar{v}) = k_n(v)$. This invariance of the normal curvature under a change of variables implies that the curvature $k_n(v)$ is an *intrinsic* property of the manifold $A(x)$ embedded in R^m.

In order to relate the normal curvature to condition (47b), note that (47b) is equivalent to

$$\frac{v^*(e(\hat{x})^* Q_y^{-1} \partial^2_{xx} A(\hat{x}))v}{v^* Q(\hat{x})^{-1} v} < 1 \quad \forall\, v \in R^n. \tag{55}$$

Hence, if we introduce the unit normal vector

$$\hat{n} = \frac{e(\hat{x})}{\| e(\hat{x}) \|}$$

we may write (55) with the help of (53) also as

$$k_{\hat{n}}(v) \| e(\hat{x}) \| < 1 \quad \forall\, v \in R^n. \tag{56}$$

This important result shows that condition (47b) is governed by two distinct quantities, namely the curvature of the manifold and the amount of inconsistency of the observation vector.

We can now rephrase the necessary and sufficient conditions of (47) in geometric terms as

$$\boxed{\begin{array}{ll} a) & e(\hat{x}) \perp R(\partial_x A(\hat{x})) \\ b) & k_{\hat{n}}^n \| e(\hat{x}) \| < 1 \end{array}} \tag{57}$$

Note that both these conditions are invariant under a change of variables.

As an exemplification of (57), assume that $A(x)$ is a circle embedded in R^2 with curvature k. Figure 9 shows for this case four possible situations that may occur. In all four cases the point $\hat{x}$ is a stationary point and satisfies condition (57a). In figure 9a, $A(\hat{x})$ is a minimizer because of negative curvature. In figure 9b, $A(\hat{x})$ is a minimizer since the curvature, although positive, is still small enough relative to $\| \hat{e} \|$. In figure 9c, $A(\hat{x})$ is a nonunique minimizer. And in figure 9d, $A(\hat{x})$ is a maximizer instead of a minimizer.

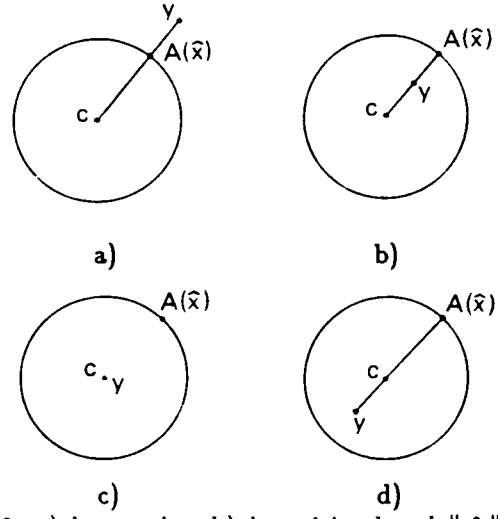

Figure 9: a) k negative, b) k positive but $k \| \hat{e} \| < 1$, c) k positive and $k \| \hat{e} \| = 1$, d) k positive and $k \| \hat{e} \| > 1$.

8 Local Convergence of the Gauss-Newton Method

In order to derive the rate of convergence of the Gauss-Newton method we expand (46) into a Taylor series at the solution $\hat{x}$. With $\partial_x A(\hat{x})^* Q_y^{-1} e(\hat{x}) = 0$, this gives

$$\begin{aligned} x_{k+1} - \hat{x} = [(1-t_k)I + t_k Q(\hat{x})[e(\hat{x})^* Q_y^{-1} \partial^2_{xx} A(\hat{x})]]. \\ .(x_k - \hat{x}) + o(\| x_k - \hat{x} \|) \end{aligned} \tag{58}$$

This shows that the Gauss-Newton method has a *linear rate of convergence* for points sufficiently close to the solution $\hat{x}$. If we take the eigenvectors v_i , $i = 1, \ldots, n$, of the generalized eigenvalue problem

$$\hat{n}^* Q_y^{-1} \partial^2_{xx} A(\hat{x}) v_i = k_{\hat{n}}^i Q(\hat{x})^{-1} v_i$$

as base vectors of the tangentspace of the manifold $A(x)$ at $\hat{x}$, and reparametrize $x_{k+1} - \hat{x}$ and $x_k - \hat{x}$ as

$$x_{k+1} - \hat{x} = \sum_{i=1}^n u_{k+1}^i v_i \quad and \quad x_k - \hat{x} = \sum_{i=1}^n u_k^i v_i$$

we can write (58) in terms of the principal normal curvatures as

$$u^i_{k+1} = [(1-t_k) + t_k k^i_{\hat{n}} \parallel e(\hat{x}) \parallel]u^i_k + o(\parallel x_k - \hat{x} \parallel) \quad (59)$$

This expression shows that the local convergence of the Gauss-Newton method is *invariant* against a change of variables. Hence, the rate of convergence of the Gauss-Newton method cannot be speed up or slowed down by a particular choice of parametrization. If the positive scalar t_k in (59) is taken to be equal to one, the rate of convergence becomes approximately

$$\boxed{\parallel x_{k+1} - \hat{x} \parallel \leq [max\{\mid k^1_{\hat{n}} \mid, \mid k^n_{\hat{n}} \mid\} \parallel e(\hat{x}) \parallel] \parallel x_k - \hat{x} \parallel} \quad (60)$$

The parameter norm in this expression is with respect to the *induced* metric $Q(\hat{x})^{-1}$. Expression (60) shows that the error magnitude gets reduced if

$$max\{\mid k^1_{\hat{n}} \mid, \mid k^n_{\hat{n}} \mid\} \parallel e(\hat{x}) \parallel < 1 \quad (61)$$

i.e. if the observation point y lies within a *hypersphere* with centre $A(\hat{x})$ and a radius equal to the inverse of the in absolute value largest curvature. If this is the case then by virtue of (15) local convergence of the Gauss-Newton method is *guaranteed*. Note however that local convergence is *not necessarily* ensured by the fact that $A(\hat{x})$ is a local minimum of $\parallel y - A(x) \parallel$. This follows if we compare inequality (61) with (57b). In figure 9a for instance, $A(\hat{x})$ is a minimizer, but $\parallel e(\hat{x}) \parallel$ may still be too large for convergence to occur.

As an exemplification of (60), assume that $A(x)$ represents a unit circle, that $Q_y = I$ and $y = (1.5, 0.0)^*$. The least squares solution is then given by $\hat{x} = 0$ and the local convergence factor by

$$k_{\hat{n}} \parallel e(\hat{x}) \parallel = -0.5. \quad (62)$$

The results of the Gauss-Newton iteration are given in table 1. They clearly show that the error magnitude gets reduced by the factor (62) in each iteration step. Also note the oscillatory character of the iteration. Oscillation or *overshoot* generally occurs if the curvatures are negative (confer (59) for $t_k = 1$). *Undershoot* on the other hand occurs if the curvatures are positive.

k	$A^1(x) = cosx$	$A^2(x) = sinx$	x_k
1	0.96235	-0.27180	-0.27526
2	0.99124	0.13205	0.13244
3	0.99785	-0.06560	-0.06564
4	0.99946	0.03274	0.03275
5	0.99987	-0.01637	-0.01637
6	0.99997	0.00818	0.00818

Table 1: Gauss-Newton iteration for orthogonal projection onto a unit circle.

Note that (60) may also be used to obtain a numerical estimate of the curvature (see also table 1)! This estimate can then be used to estimate the bias in the least squares estimators (Teunissen, 1988; Teunissen and Knickmeyer, 1988) and to diagnose the statistical significance of nonlinearity (Teunissen, 1989a,b).

If inequality (61) is not satisfied, one can enforce local convergence by a suitable choice for the positive scalar t_k of (59). It follows (compare with our discussion of the steepest descent method) that the optimal choice for t_k is

$$t_k = 1/[1 - \parallel e(\hat{x}) \parallel \frac{1}{2}(k^1_{\hat{n}} + k^n_{\hat{n}})]. \quad (63)$$

With this choice for t_k it follows from (59) that instead of (60) we have

$$\boxed{\parallel x_{k+1} - \hat{x} \parallel \leq \left[\frac{\parallel e(\hat{x}) \parallel (k^n_{\hat{n}} - k^1_{\hat{n}})}{2 - \parallel e(\hat{x}) \parallel (k^n_{\hat{n}} + k^1_{\hat{n}})}\right] \parallel x_k - \hat{x} \parallel} \quad (64)$$

In this case local convergence is guaranteed if (57b) holds. Equation (63) shows that the simplest choice $t_k = 1$ is close to optimal if either $\parallel e(\hat{x}) \parallel$ is small enough or the average of the extreme curvatures is small enough. Thus for points sufficiently close to the solution, the simplest line search strategy can be considered adequate if the manifold is moderately curved at $\hat{x}$ and/or the observation point is close enough to the manifold.

So far it was assumed that we were dealing with a curved manifold with inconsistent data. But what happens with the local convergence behaviour of the Gauss-Newton method if either the manifold is *flat* (zero-curvature) or the data is *consistent* (zero-residual vector)? In order to answer this question we first note that for a flat manifold, the orthogonal projector

$$P_{\partial_x A} = \partial_x A(x)[\partial_x A(x)^* Q_y^{-1} \partial_x A(x)]^{-1} \partial_x A(x)^* Q_y^{-1}$$

is constant and independent of x, and

$$A(\hat{x}) = \bar{y} + P_{\partial_x A}(y - \bar{y}) \quad \forall\, \bar{y} \in A(x)$$

With this follows that

$$\begin{aligned} &Q(x_k)\partial_x A(x_k)^* Q_y^{-1}(y - A(x_k)) = \\ &Q(x_k)\partial_x A(x_k)^* Q_y^{-1} P_{\partial_x A}(y - A(x_k)) = \\ &Q(x_k)\partial_x A(x_k)^* Q_y^{-1}(A(\hat{x}) - A(x_k)) \end{aligned}$$

This result shows that for both the cases of a flat manifold and consistent data, the observation vector y in (46) may be replaced by $A(\hat{x})$. From a Taylor series expansion at $\hat{x}$ of (46) with y replaced by $A(\hat{x})$ follows then with $t_k = 1$ and the identity

$$\begin{aligned} \partial_x Q(x) Q(x)^{-1} + Q(x)\partial^2_{xx} A(x) Q_y^{-1} \partial_x A(x) + \\ + Q(x)\partial_x A(x)^* Q_y^{-1} \partial^2_{xx} A(x) \quad = \quad 0 \end{aligned}$$

that

$$x_{k+1} - \hat{x} = \tfrac{1}{2} Q(\hat{x}) \partial_x A(\hat{x})^* Q_y^{-1} [(x_k - \hat{x})^* . \\ .\partial^2_{xx} A(\hat{x})(x_k - \hat{x})] + o(\| x_k - \hat{x} \|^2) \quad (65)$$

This shows that the rate of convergence of the Gauss-Newton method is *quadratic* in case of flat manifolds and/or consistent data. If we take the norm of (65) we get

$$\| x_{k+1} - \hat{x} \| = \frac{1}{2} \| P_{\partial_x A} (x_k - \hat{x})^* \partial^2_{xx} A(\hat{x})(x_k - \hat{x}) \| \quad (66)$$

This shows that the convergence factor is determined by the *tangential* components of $\partial^2_{xx} A(\hat{x})$. Compare this with for instance (58), where the convergence factor depends on the *normal* component of $\partial^2_{xx} A(\hat{x})$.

If we let $t_\alpha, \alpha = 1, \ldots, n$, be an orthonormal basis of $R(\partial_x A(\hat{x}))$ we can write the orthogonal projector $P_{\partial_x A(\hat{x})}$ as

$$P_{\partial_x A(\hat{x})} = \sum_{\alpha=1}^{n} t_\alpha t_\alpha^* Q_y^{-1}.$$

Substitution into (66) gives

$$\| x_{k+1} - \hat{x} \| = \\ \tfrac{1}{2} \| \textstyle\sum_{\alpha=1}^{n} t_\alpha (x_k - \hat{x})^* [t_\alpha^* Q_y^{-1} \partial^2_{xx} A(\hat{x})](x_k - \hat{x}) \| \quad (67)$$

In analogy to (53) we define the *tangential curvature* by

$$\boxed{\gamma_t(v) = \frac{v^* [t^* Q_y^{-1} \partial^2_{xx} A(x)] v}{v^* Q(x)^{-1} v}} \quad (68)$$

with $v \in R^n$ and $t \in R(\partial_x A(x))$. It can be shown [Teunissen, 1985] that (68) is closely related to the concept of *covariant differentiation* of differential geometry. The $\gamma_t(v)$ describe therefore the nonlinearity of the parameter curves in the manifold. Note that (68) in contrast to (53), is *not* invariant under a change of variables in $A(x)$.

With (68), equation (67) can be written as

$$\boxed{\| x_{k+1} - \hat{x} \| = \frac{1}{2} [\sum_{\alpha=1}^{n} \gamma^2_{t_\alpha}]^{1/2} \| x_k - \hat{x} \|^2} \quad (69)$$

Compare this result with (60).

As an exemplification of (65), assume that $A(x)$ represents a straight line, that $A^1(x) = exp(10x)$, $A^2(x) = exp(10x)$, $Q_y = I$ and $y = (0, 2e)^*$. The least squares solution is then given by $\hat{x} = 0.1$ and the convergence factor by

$$\frac{1}{2} Q(\hat{x}) \partial_x A(\hat{x})^* Q_y^{-1} \partial^2_{xx} A(\hat{x}) = 5.$$

The results of the Gauss-Newton iteration are given in table 2. They clearly show the quadratic rate of convergence. Also note that e.g. $(x_4 - \hat{x}) = 5(x_3 - \hat{x})^2$.

k	$A^1(x) = exp(10x)$	$A^2(x) = exp(10x)$	x_k
1	5.57494	5.57494	0.17183
2	3.33967	3.33967	0.12059
3	2.77267	2.77267	0.10198
4	2.71881	2.71881	0.10002
5	2.71828	2.71828	0.10000

Table 2: Gauss-Newton iteration for orthogonal projection onto a straight line with a nonlinear parametrization.

9 A Convergence Criterion

Every iteration method needs one or more termination criteria in order to be able to test whether the iteration should be continued or not. Apart from the computer-time to termination and/or the number of iterations, the most important termination criterion is the one which measures the success in obtaining an optimal solution. Since iterative descent methods try to locate a stationary point of the objective function $F(x)$, convergence can be declared if the gradient of $F(x)$, evaluated at the current iteration point x_k, falls below a preset tolerance level. A test for convergence is therefore

$$\| \partial_x F(x_k) \| < \epsilon.$$

For our least squares problem this becomes

$$\| \partial_x A(x_k)^* Q_y^{-1} e(x_k) \| < \epsilon. \quad (70)$$

In order to make the norm of the gradient *invariant* to a change of variables and thus insensitive to scale changes we choose to take the norm in (70) with respect to the induced metric $Q(x_k)^{-1}$. This gives for the Gauss-Newton method the convergency test

$$\boxed{\| x_{k+1} - x_k \| < \epsilon} \quad (71)$$

Note that since $\partial_x A(x_k)(x_{k+1} - x_k) = P_{\partial_x A(x_k)}$ the convergency test can also be written as

$$\boxed{\| P_{\partial_x A(x_k)} e(x_k) \| < \epsilon} \quad (72)$$

where the norm is with respect to the metric Q_y^{-1}.

In order to apply the convergency test we need to choose a value for the tolerance level ϵ. On what should we base our choice for ϵ? It seems natural to base the choice for ϵ on the quality of the observation vector y. Since x_k is the *exact* least squares solution of the *perturbed* minimization problem

$$\min_x \| [y - P_{\partial_x A(x_k)} e(x_k)] - A(x) \|$$

(see figure 10), it follows that the tolerance level ϵ of (72) should be chosen such that a perturbation of y with $P_{\partial_x A(x_k)} e(x_k)$ is considered insignificant.

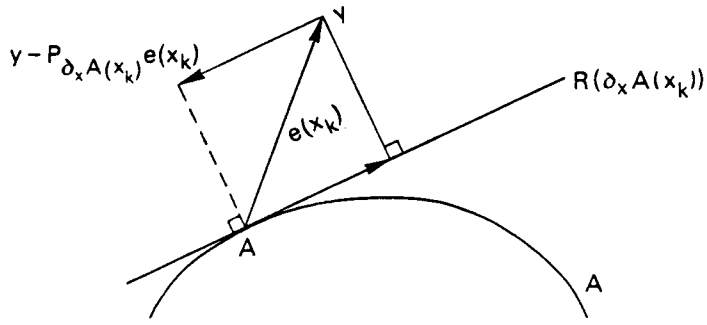

Figure 10: Perturbation of y with $P_{\partial_x A(x)} e(x_k)$.

Once convergence is declared it is of interest to know the error magnitude of the computed quantities x_k, $A(x_k)$ and $\| e(x_k) \|$.

In order to determine the relation between the above convergency indicator and the error magnitude of x_k, we expand $P_{\partial_x A(x_k)} e(x_k)$ in a Taylor series at $\hat{x}$. This gives

$$P_{\partial_x A(x_k)} e(x_k) = \partial_x A(\hat{x})[Q(\hat{x})\partial^2_{xx} A(\hat{x}) Q_y^{-1} e(\hat{x}) + -I](x_k - \hat{x}) + o(\| x_k - \hat{x} \|)$$

From this follows, if $\hat{x}$ is a local minimum of $\| y - A(x) \|$ and thus $k^n_{\hat{n}} \| e(\hat{x}) \| < 1$ (see (57b)), that

$$\left.\begin{array}{c} [1 - k^n_{\hat{n}} \| \hat{e} \|] \| x_k - \hat{x} \| \\ +o(\| x_k - \hat{x} \|) \end{array}\right\} \leq$$

$$\| P_{\partial_x A(x_k)} e(x_k) \|$$

$$\leq \left\{ \begin{array}{c} [1 - k^1_{\hat{n}} \| \hat{e} \|] \| x_k - \hat{x} \| \\ +o(\| x_k - \hat{x} \|) \end{array}\right.$$

Hence, we have the approximate interval

$$\boxed{\frac{\| x_{k+1} - x_k \|}{1 - k^1_{\hat{n}} \| \hat{e} \|} \leq \| x_k - \hat{x} \| \leq \frac{\| x_{k+1} - x_k \|}{1 - k^n_{\hat{n}} \| \hat{e} \|}} \quad (73)$$

In an analogous way we can derive the approximate intervals

$$\boxed{\frac{\| x_{k+1} - x_k \|}{1 - k^1_{\hat{n}} \| \hat{e} \|} \leq \| A(x_k) - A(\hat{x}) \| \leq \frac{\| x_{k+1} - x_k \|}{1 - k^n_{\hat{n}} \| \hat{e} \|}} \quad (74)$$

and

$$\boxed{\begin{array}{c} \frac{\| x_{k+1} - x_k \|^2}{1 - k^1_{\hat{n}} \| \hat{e} \|} \leq \\ \| e(x_k) \|^2 - \| e(\hat{x}) \|^2 \\ \leq \frac{\| x_{k+1} - x_k \|^2}{1 - k^n_{\hat{n}} \| \hat{e} \|} \end{array}} \quad (75)$$

And similarly we find for the error magnitude of the individual parameters, the upper bound

$$\boxed{| x^\alpha_k - \hat{x}^\alpha | \leq \sigma_{\hat{x}^\alpha} \frac{\| x_{k+1} - x_k \|}{1 - k^n_{\hat{n}} \| \hat{e} \|}} \quad (76)$$

with $\sigma_{\hat{x}^\alpha}$ the square root of the αth-diagonal element of $Q(\hat{x})$.

All the above inequalities show how the error magnitudes of the computed quantities are related to the convergency indicator $\| x_{k+1} - x_k \|$. In particular note that $\| x_k - \hat{x} \|$ can be large if $k^n_{\hat{n}} \| \hat{e} \|$ is close to one. This happens if y is close to the centre of curvature of $k^n_{\hat{n}}$, in which case the objective function $\| e(x) \|$ is flat near $\hat{x}$.

10 Conclusions

In this paper the numerical characteristics of a class of iterative descent methods for solving nonlinear least squares problems were discussed. Particular attention was given to the Gauss-Newton method. The advantage of this method when compared with the Steepest-Descent method is that it is invariant against a change of variables. It was shown that the Gauss-Newton method takes advantage of the "sum of squares" structure of the objective function and that it can be interpreted as a "regularized" version of the Newton method. The numerical characteristics of the Gauss-Newton method were presented in terms of differential geometric concepts. It was shown that the Gauss-Newton method exhibits a linear rate of convergence determined by the length of the residual vector and the maximum curvature of the manifold. Thus local convergence is guaranteed if the data point lies within a hypersphere with centre $A(\hat{x})$ and a radius equal to the inverse of the in absolute value largest curvature. The simplest line search strategy was shown to be adequate if the manifold is moderately curved and/or the datapoint is close enough to the manifold. In the special case of zero curvature and/or consistent data the Gauss-Newton method exhibits a quadratic rate of convergence which is determined by the nonlinearity of the parameter curves. Finally, a parameter invariant convergency criterion was formulated and bounds on the error magnitudes of the computed quantities were derived.

Acknowledgement

The expert typesetting of this work by Ms. ir. H.M.E. Verhoef using the program LaTeX is greatly appreciated.

References

Bähr, H.G. (1988): *A quadratic approach to the non-linear treatment of non-redundant observations*, Manuscripta Geodaetica, Vol. 13, No. 3, pp. 191-197.

Blaha, G. and R.P. Besette (1989): *Nonlinear least-squares method via an isomorphic geometrical setup*, Bulletin Geodesique , Vol. 63, No. 2, pp. 115-138.

Borre, K. and S.L. Lauritzen (1983): *Some Geometric Aspects of Adjustment*, In: Festschrift to Torben Krarup, Ed. E. Kejlso et al., Geod. Inst., No. 58, pp. 70-90.

Cauchy, A. (1847): *Méthode générale pour la résolution des systèmes d'équations simultanées*, C.R. Acad. Sci. Paris, Vol. 25, pp. 536-538.

Goldfeld, S.M., R.E. Quandt and H.F. Trotter (1966): *Maximization by quadratic hill climbing*, Econometrica, Vol. 34, pp. 541-551.

Grafarend, E.W. and B. Schaffrin (1989): *The Geometry of Nonlinear Adjustment - the Planar Trisection Problem*, In: Festschrift to Torben Krarup, Ed. E. Kejlso et al., Geod. Inst., No. 58, pp. 149-172.

Krarup, T. (1982): *Non-linear adjustment and curvature*, In: Forty Years of Thought, Delft, pp. 145-159.

Kubik, K.K. (1967): *Iterative Methoden zur Lösung des nichtlinearen Ausgleichungsproblemes*, Zeitschrift für Vermessungswesen, Vol. 91, No. 6, pp. 214-225.

Levenberg, K. (1944): *A method for the solution of certain nonlinear problems in least-squares*, Quart. Appl. Math. Vol. 2, pp. 164-168.

Marquardt, D.W. (1963): *An algorithm for least-squares estimation of nonlinear parameters*, J. SIAM II, pp. 431-441.

Naylor, A.W. and G.R. Sell (1982): *Linear Operator Theory in Engineering and Science*, Springer Verlag.

Ortega, J.M. and W.C. Rheinboldt (1970): *Iterative solution of nonlinear equations in several variables*, Academic Press, New York.

Pope, A. (1982): *Two approaches to non-linear least-squares adjustments*, The Canadian Surveyor, Vol. 28, No. 5, pp. 663-669.

Saito, T. (1973): *The non-linear least-squares of condition equations*, Bulletin Geodesique, Vol. 110, pp. 367-395.

Schek, H.J. and Ph. Maier (1976): *Nichtlineare Normalgleichungen zur Bestimmung der Unbekannten und deren Kovarianzmatrix*, Zeitschrift für Vermessungswesen, Vol. 101, No. 4, pp. 140-159.

Spivak, M. (1979): *A Comprehensive Introduction to Differential Geometry*, Vol. II, Publish or Perish, Inc.

Teunissen, P.J.G. (1984): *A note on the use of Gauss' formula in nonlinear geodetic adjustments*, Statistics and Descisions, No. 2, pp. 455-466.

Teunissen, P.J.G. (1985): *The geometry of geodetic inverse linear mapping and nonlinear adjustment*, Netherlands Geodetic Commission, Publications on Geodesy, New Series, Vol. 8, No. 1, Delft.

Teunissen, P.J.G. (1988): *First and Second Order Moments of Nonlinear Least-Squares Estimators*, will be published in Bulletin Geodesique.

Teunissen, P.J.G. and E.H. Knickmeyer (1988): *Nonlinearity and Least Squares*, CISM Journal ACSGC, Vol. 42, No. 4, pp. 321-330.

Teunissen, P.J.G. (1989a): *A Note on the Bias in the Symmetric Helmert Transformation*, In: Festschrift to Torben Krarup, Ed. E. Kejlso et al., Geod. Inst., No. 58, pp. 335-342.

Teunissen, P.J.G. (1989b): *Nonlinear Inversion of Geodetic and Geophysical Data: Diagnosing Nonlinearity*, Invited paper, The Ron Mather Symposium on Four-Dimensional Geodesy, Sydney, Australia.