Konrad-Zuse-Zentrum für Informationstechnik Berlin
Heilbronner Str. 10, W-1000 Berlin 31

Peter Deuflhard Florian A. Potra*

A Refined Gauss–Newton–Mysovskii Theorem

* Department of Mathematics
The University of Iowa
Iowa City, IA 52242
U.S.A.

Preprint SC 91–4 (December 1991)

Contents

Abstract

The present paper contains a generalization of a refinement of the Newton–Mysovskii theorem, recently obtained by the authors, to the case of Gauss–Newton procedures for solving nonlinear least–squares problems with full Jacobians. Invariant sufficient conditions are given that ensure the convergence of the Gauss–Newton iterates towards a solution of the problem, as well as the uniqueness of that solution in an explicitly defined neighborhood. It is shown by a counter–example that the results do not carry over to the rank-deficient case.

The authors would like to thank E. Cadano–Körnig for her quick and careful TeX–typing of the manuscript.

1. Introduction

The Gauss–Newton method is one of the most popular techniques to tackle nonlinear least squares problems. Only first–order derivative information is used to construct a useful local quadratic model. As has been shown earlier, the ordinary Gauss–Newton method converges locally for so–called *adequate* nonlinear least squares problems [2], which means a characterization of a "small residual" at the solution point in terms of second–order information — in basic agreement with results of WEDIN [8], but in contradiction with statements of GILL/MURRAY [6] who suggested such a characterization based on first–order information (in particular, in terms of the singular value distribution of the Jacobian). In contrast to WEDIN [8], however, the characterization of [2] uses norms and quantities exclusively in the parameter space — a fact, which makes an extension to the constrained case much easier and even trivial (compare [3]). In fact, a rather general class of nonlinear least squares problems is covered by the convergence theorem of [4], which is a natural generalization of the well–known Newton–Mysovskii theorem. Recently, in [5], the present authors gave a refinement of that theorem, which guarantees both quadratic convergence and uniqueness out of one frame of simple assumptions. The purpose of this paper here is to generalize that convergence theorem to the nonlinear least squares case.

2. Local Convergence of General Gauss–Newton Methods — the Full Rank Case

Consider the nonlinear least squares problem

$$g(x) := \|F(x)\|_2^2 = \min \tag{2.1}$$

where $F : D \subset \mathbb{R}^n \to \mathbb{R}^m$ is a twice continuously differentiable mapping of an open set $D \subset \mathbb{R}^n$ into $\mathbb{R}^m$ and $m \geq n$. In order not to complicate notation unnecessarily, the restriction to the finite–dimensional case is made throughout. At each step a Gauss–Newton method solves a linear least squares problem of the form

$$g_k(\Delta x) := \|F'(x^k)\Delta x + F(x^k)\|_2^2 = \min \tag{2.2}$$

to obtain a correction Δx^k, and then defines

$$x^{k+1} := x^k + \Delta x^k . \tag{2.3}$$

If $F'(x^k)$ has full rank, then (2.2) has a unique solution. Otherwise the solution set of (2.2) is a linear manifold $\mathcal{N}_k$ of dimension $p = n$–rank $(F'(x^k))$ in the parameter space $\mathbb{R}^n$.

The ordinary Gauss–Newton method takes as Δx^k the "shortest" solution of (2.2) i.e. the orthogonal projection of the origin of $\mathbb{R}^n$ onto the linear manifold $\mathcal{N}_k$

$$\Delta x^k = \Pr(\mathbb{R}^n, \mathcal{N}_k)(0) . \tag{2.4}$$

In terms of the Moore–Penrose pseudo–inverse this can be written as

$$\Delta x^k = -F'(x^k)^+ F(x^k) . \tag{2.5}$$

WEDIN [8] interprets (2.2) as the problem of finding a point y^k on the linear manifold

$$\begin{aligned} \mathcal{M}_k &:= \{y \in \mathbb{R}^m : y = F'(x^k)h + F(x^k), h \in \mathbb{R}^k\} \\ &= F(x^k) + \mathcal{R}(F'(x^k)) \end{aligned} \tag{2.6}$$

that is closest (in Euclidian norm) to the origin of $\mathbb{R}^m$. This point is unique and is given by the orthogonal projection of the origin of $\mathbb{R}^m$ onto $\mathcal{M}_k$ i.e

$$y^k = \Pr(\mathbb{R}^m, \mathcal{M}_k)(0) . \tag{2.7}$$

Let A_k denote the affine mapping

$$A_k(h) = F'(x^k)h + F(x^k) . \tag{2.8}$$

We have clearly

$$\mathcal{M}_k = A_k(\mathbb{R}^n), \ \mathcal{N}_k = A_k^{-1}(y^k) \tag{2.9}$$

so that

$$\dim(\mathcal{M}_k) + \dim(\mathcal{N}) = n . \tag{2.10}$$

Therefore $\mathcal{N}_k = \{x^k\}$ if and only if $\dim(\mathcal{M}_k) = n$ i.e. whenever $F'(x^k)$ has full rank.

In summary, it follows that the ordinary Gauss–Newton method

$$x^{k+1} = x^k - F'(x^k)^+ F(x^k), \ k = 0, 1, \ldots \tag{2.11}$$

is well defined, and at each step the correction $\Delta x^k = x^{k+1} - x^k$ is the projection of the origin of $\mathbb{R}^n$ onto the linear manifold $\mathcal{N}_k$ formed by the solutions of the linear least squares problem (2.2).

It is possible to consider more general Gauss–Newton algorithms of the form

$$x^{k+1} = x^k - \Gamma_F(x^k),\ F(x^k),\ \ k = 0, 1, \ldots \tag{2.12}$$

where $\Gamma_F(x^k)$ is a linear mapping that generalizes the Moore–Penrose pseudo–inverse. The following theorem due to DEUFLHARD/HEINDL [4] gives sufficient conditions for the convergence of the sequence of iterates $\{x^k\}$ generated by (2.12).

Theorem 2.1 *Let $F : D \subset \mathbb{R}^n \to \mathbb{R}^m$ be a continuously differentiable mapping defined on an open convex set $D \subset \mathbb{R}^n$, with values in $\mathbb{R}^m$, and let $\Gamma_F : D \to \mathbb{R}^{n\times m}$ be a given mapping. Assume that one can find a starting point $x^0 \in D$, a mapping $\kappa : D \to \mathbb{R}^+$ and constants $\alpha,\ \omega,\ \overline{\kappa} \geq 0$ such that*

a) $\|\Gamma_F(x^0)F(x^0)\| \leq \alpha$

b) $\|\Gamma_F(y)(F'(u + s\cdot v) - F'(u))v\| \leq s\cdot\omega\|v\|^2$
$0 \leq s \leq 1$,

c) $\|\Gamma_F(y)(I - F'(x)\Gamma_F(x))F(x)\| \leq \kappa(x)\|y - x\|\ \ \forall x, y \in D$ (2.13)

d) $\kappa(x) \leq \overline{\kappa} < \forall x \in D$
$h : \frac{1}{2}\alpha\omega < 1 - \overline{\kappa}$

e $\overline{S}(x^0, \rho) \subset D$ *with* $\rho_0 := \alpha/(1 - \overline{\kappa} - h)$.

Then the iterates $\{x^k\}$ from (2.12) *are well-defined, remain in $\overline{S}(x^0, \rho_0)$ and converge to some $x^* \in \overline{S}(x^0, \rho_0)$ with*

$$\Gamma_F(x^*)F(x^*) = 0\ . \tag{2.14}$$

The convergence rate can be estimated according to

$$\|x^{k+1} - x^k\| \leq \left(\kappa(x^{k-1}) + \frac{\omega}{2}\|x^k - x^{k-1}\|\right)\|x^k - x^{k+1}\|\ . \tag{2.15}$$

Note that for the ordinary Gauss–Newton method, where $\Gamma_F(y) = F'(y)^+$, condition (2.13.c) reduces to

$$\|(F'(y) - F'(x)^+)\ \Pr(\mathbb{R}^m, \ker(F'(x)^T))F(x)\| \leq \kappa(x)\|y - x\|,\quad \forall x, y \in D\ , \tag{2.16}$$

while equation (2.14) will become

$$F'(x^*)^T\ F(x^*) = 0\ . \tag{2.17}$$

In the case $n = m$, and $F'(x)$ has full rank (i.e. nonsingular) then $F'(x)^+ = F'(x)^{-1}$ and (2.16) is trivially satisfied with $\kappa(x) = 0$ because $\ker(F'(x)^T) = \{0\}$. In this case Theorem 2.1 reduces to the affine invariant version of the Newton–Mysovskii theorem of DEUFLHARD/HEINDL [4]. As we mentioned in the introduction, that theorem has been recently refined by DEUFLHARD/POTRA [5] so that it guarantees not only the convergence of the iterates towards a solution but also the uniqueness of that solution. We will now generalize that result to the case of the ordinary Gauss–Newton when $m \geq n$ and $F'(x)$ has full rank. As will turn out, condition (2.13.c) can be replaced by a condition of the form

$$\|(F'(y)^+ - F'(x)^+)F(x)\| \leq \lambda(x)\|x - y\| \ \ \forall x,\, y \in D\ . \tag{2.18}$$

We note that (2.18) does not necessarily imply (2.16).

Theorem 2.2 *Let F be a continuously differentiable mapping defined on an open convex set $D \subset \mathbb{R}^n$ with values in $\mathbb{R}^m$ and suppose there are a starting point $x^0 \in D$, a mapping $\lambda : D \to \mathbb{R}^+$ and constants α, ω, $\overline{\lambda}$ such that*

a) $\|F'(x^0)^+F(x^0)\| \leq \alpha$

b) $\|F'(u)^+(F'(u + sv) - F'(u))v\| \leq s\omega\|v\|^2$
$\forall s \in (0,1),\ y \in D,\ u \in D,\ v \in D \cap \mathcal{R}(F'(u)^+)$

c) $\|(F'(y)^+ - F'(x)^+)F(x)\| \leq \lambda(x)\|x - y\|,\ \forall x, y \in D$ (2.19)

d) $\lambda(x) \leq \overline{\lambda} < 1\ \forall x \in D$
$h := \frac{1}{2}\alpha\omega < 1 - \overline{\lambda}$

e) $\overline{S}(x^0, \rho) \subset D$ *with* $\rho := \alpha/(1 - \overline{\lambda} - h)$

Then the iterates $\{x^k\}$ from (2.11) are well defined and converge to some $x^* \in \overline{S}(x^0, \rho)$ with

$$F'(x^*)^T F(x^*) = 0\ . \tag{2.20}$$

The convergence rate is characterized by

$$\|x^{k+1} - x^k\| \leq \left(\lambda(x^k) + \frac{\omega}{2}\|x^k - x^{k-1}\|\right)\|x^k - x^{k-1}\|\ . \tag{2.21}$$

Proof. By using the fact that in the full rank case $F'(x)^+F'(x) = I_n$, we can write

$$\begin{aligned}
x^{k+1} - x^k &= -F'(x^k)^+F(x^k) \\
&= -F'(x^k)^+F(x^k) + x^k - x^{k-1} + F'(x^{k-1})^+F(x^{k-1}) \\
&= -F'(x^k)^+F(x^k) + F'(x^{k-1})^+F'(x^{k-1})(x^k - x^{k-1}) \\
&\quad +F'(x^{k-1})^+F(x^{k-1}) \\
&= (F'(x^{k-1})^+ - F'(x^k)^+)F(x^k) \\
&\quad +F'(x^{k-1})^+\big(F(x^k) - F(x^{k+1}) - F'(x^{k-1})(x^k - x^{k-1})\big)
\end{aligned}$$

The last two terms may be majorated by virtue of (2.19.b, c) to obtain (2.26). Then the theorem follows by standard techniques (see [5]) ■

We note that condition (2.18) appears in BEN–ISRAEL [1] and the above Theorem 2.2 may be regarded as a parameter space based version of Theorem 2.1 of [1]. Furthermore, we have

Proposition 2.3 [1]. *Under the hypothesis of Theorem 2.2, x^* is an isolated zero of the equation*

$$F'(x^*)^T F(x) = 0 \tag{2.22}$$

in the set $D \cap \mathcal{T}_*$ where $\mathcal{T}_*$ is the linear manifold

$$\mathcal{T}_* = x^* + \mathcal{R}(F'(x^*)^T). \tag{2.23}$$

■

In case $F'(x^*)$ has full rank then $\mathcal{T}_* = \mathbb{R}^n$ so that we obtain:

Corollary 2.4 *Under the hypothesis of Theorem 2.2 suppose that $F'(x^*)$ is full rank. Then x^* is an isolated zero of the equation (2.22) in D.*

■

The above corollary says that there is a neighborhood of x^* which contains no other solution of the equation (2.22) besides x^*. We remark that this does not imply the fact that x is an isolated zero of the equation

$$F'(x^T)F(x) = 0 \,. \tag{2.24}$$

In the following theorem we give sufficient conditions under which the iterates produced by the ordinary Gauss–Newton method converge to an isolated zero of the equation (2.24).

Theorem 2.5 *Let $D \subset \mathbb{R}^n$ be convex and open and let $F : D \to \mathbb{R}^m$, $m \geq n$ be a continuously differentiable mapping whose Jacobian $F'(x)$ has full rank for all $x \in D$. Suppose that the nonlinear equation (2.24) has a solution $x^* \in D$ and that there are constants $\omega > 0$ and $0 \leq \lambda_* < 1$ such that*

$$\begin{aligned} &a)\quad \|F'(x)^+(F'(x+sv) - F'(x)v\| \leq s\omega\|v\|^2 \\ &\quad\ \forall s \in (0,1),\ x \in D,\ v \in D \\ &b)\quad \|(F'(x)^+ - F'(x^*)^+)F(x^*)\| \leq \lambda_*\|x - x^*\|,\ \forall x \in D \\ &c)\quad B_\sigma(x^*) := \{x \in \mathbb{R}^n : \|x - x^*\| < \sigma = 2(1-\lambda_*)/\omega\} \subset D \end{aligned} \tag{2.25}$$

Then for any starting point $x^0 \in D$ such that

$$\rho := \|x - x^0\| < \frac{1}{\omega}(1 - \lambda_*) =: \sigma \tag{2.26}$$

the Gauss–Newton iterates (2.11) are well defined, converge to x^* and satisfy the following inequalities

$$\|x^{k+1} - x^*\| \leq \frac{\omega}{2}\|x^k - x^*\|^2 + \lambda_*\|x^k - x^*\| \leq c\|x^k - x^*\| \leq \rho \tag{2.27}$$

where

$$c = \frac{\omega}{2}\rho + \lambda_* < 1\,. \tag{2.28}$$

Moreover x^* is the unique solution of the nonlinear equation (2.24) in the open ball $B_\sigma(x^*)$.

Proof. If $F'(x)$ is of full rank then $F'(x)^+F(x) = I$ so that we can write

$$\begin{aligned} x^{k+1} - x^* &= x^k - x^* - F'(x^k)^+F(x^k) \\ &= F'(x^k)^+(F(x^*) - F(x^k) - F'(x^k)(x^* - x^k)) \\ &\quad - (F'(x^k)^+ - F'(x^*)^+)F(x^*) \end{aligned}$$

and by using (2.25) we obtain

$$\begin{aligned} \|x^{k+1} - x^*\| &\leq \frac{\omega}{2}\|x^k - x^*\|^2 + \lambda_*\|x^k - x^*\| \\ &= \left(\frac{\omega}{2}\|x^k - x^*\| + \lambda_*\right)\|x^k - x^*\| = c_k\|x^k - x^*\| \end{aligned}$$

For $k = 0$ we have

$$c_0 = c = \frac{\omega}{2}\rho + \lambda_* < \frac{\omega}{2}\sigma + \lambda_* = 1$$

and therefore we have

$$\|x^1 - x^0\| \le c_0\|x^0 - x^*\| < \|x^0 - x^*\| \le \rho .$$

By induction we can prove that

$$\|x^{k+1} - x^*\| \le c_k\|x^k - x^*\| \le c_0\|x^k - x^*\|$$

wherefrom we deduce the convergence of the sequence $\{x^k\}$ because $c_0 < 1$. Suppose now that there is another solution x^{**} of equation (2.24) such that $\rho = \|x^* - x^{**}\| < \sigma$. If we take $x^0 = x^{**}$ then the ordinary Gauss–Newton method (2.11) gives $x^1 = x^{**}$ and by applying (2.27) we have

$$\|x^{**} - x^*\| \le c_0\|x^{**} - x^*\|,\ 0 < c < 1$$

wherefrom it follows that $x^* = x^{**}$. ■

In case our problem is compatible, i.e. if $F(x^*) = 0$, then $\lambda_* = 0$ and (2.27) shows that the sequence $\{x^k\}$ converges *even quadratically.* In the particular case $n = m$ Theorem 2.5 reduces to the refined Newton–Mysovskii theorem obtained in [5].

3. A Counter–Example for the Rank–Deficient Case

In order to throw some light into the rank–deficient case, we present a counter-example which shows that Theorem 2.5 cannot be extended in a straightforward way to the case where the Jacobian $F'(x)$ is rank deficient. Let $n = 2$, $m = 3$ and

$$x = \begin{pmatrix} \xi \\ \eta \end{pmatrix},\ F(x) = \tfrac{1}{2}\begin{pmatrix} \xi^2 - \eta^2 \\ \xi^2 - \eta^2 \\ 2 \end{pmatrix}. \tag{3.1}$$

We have

$$F'(x) = \begin{pmatrix} \xi & -\eta \\ \xi & -\eta \\ 0 & 0 \end{pmatrix} \tag{3.2}$$

$$F'(x)^+ = \frac{1}{2(\xi^2+\eta^2)} \begin{pmatrix} \xi & \xi & 0 \\ -\eta & -\eta & 0 \end{pmatrix} . \tag{3.3}$$

For convenience let us take

$$D = \left\{ \begin{pmatrix} \xi \\ \eta \end{pmatrix} \in \mathbb{R}^2 : \xi > 0 \,,\ \eta > 0 \,,\ \xi + \eta > 1 \right\} . \tag{3.4}$$

Then it is easily seen that the hypotheses of Theorem 2.1 are satisfied for $\|\cdot\| = \|\cdot\|_2$ and

$$x^0 = \begin{pmatrix} 4 \\ 3 \end{pmatrix} ,\ \alpha = \tfrac{7}{20} ,\ \overline{\kappa} = 0,\ \omega = 1 , \qquad h = \tfrac{7}{40} ,\ \rho = \tfrac{14}{33} . \tag{3.5}$$

Therefore the Gauss–Newton iterates will converge quadratically to a solution of the equation $F'(x)^+F(x)$, which in our case reduces to

$$\xi^2 = \eta^2 . \tag{3.6}$$

We note that in our case the solution sets of (2.1), (2.24), and (3.6) coincide and clearly this solution set contains no isolated points. We also note that the Jacobian of the iteration mapping

$$H(x) = x - F'(x)^+F(x) \tag{3.7}$$

at a solution point with $\xi = \eta$ is given by

$$H' \begin{pmatrix} \xi \\ \xi \end{pmatrix} = \tfrac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} . \tag{3.8}$$

The above matrix has the eigenvalues equal to 0 and 1 and therefore

$$\|H' \begin{pmatrix} \xi \\ \xi \end{pmatrix} \| = 1 . \tag{3.9}$$

Therefore the quadratic convergence of the Gauss–Newton iterates cannot be proved in a standard way. However, from Theorem 2.1 it follows that

$$\|x^{k+1} - x^k\| \le \tfrac{1}{2}\|x^k - x^{k-1}\|^2 \quad k = 1, 2, \ldots \tag{3.10}$$

and by using Proposition 6.4 of [7] we deduce that there is a constant $\delta > 0$ such that

$$\|x^{k+1} - x^*\| \le \delta \|x^k - x^*\|^2 \quad k = 0, 1, \ldots \tag{3.11}$$

Generally speaking, the Gauss–Newton method will converge within some subspace, if the nullspace projection remains constant throughout D. Even more, convergence can be guaranteed, if this projection itself satisfies a Lipschitz condition. These cases, however, are left to some future investigation.

References

[1] A. Ben–Israel: *A Newton–Raphson method for the solution of systems of equations.* J. Math. Anal. Appl. **15**, pp. 243–252 (1966).

[2] P. Deuflhard, V. Apostolescu: *An underrelaxed Gauss–Newton method for equality constrained nonlinear least squares problems.* In: J. Stoer (ed.), "Optimization Techniques". Proc. 8th IFIP Conf., Würzburg (1977), part 2, Springer Lecture Notes Control Inf. Sci. **7**, pp. 22–32 (1978).

[3] P. Deuflhard, V. Apostolescu: *A study of the Gauss–Newton method for the solution of nonlinear least squares problem.* In: Frehse/Pallaschke/Trottenberg (eds.), "Special Topics of Applied Mathematics". North–Holland, Amsterdam, pp. 129–150 (1980).

[4] P. Deuflhard, G. Heindl: *Affine invariant convergence theorems for Newton's method and extensions to related methods.* SIAM J. Numer. Anal. **16**, pp. 1–10 (1979).

[5] P. Deuflhard, F. A. Potra: *Asymptotic mesh independence for Newton–Galerkin methods via a refined Mysovskii theorem.* Preprint SC 90–9, Konrad–Zuse–Zentrum Berlin (1990). To appear in SIAM J. Numer. Anal.

[6] P. E. Gill, W. Murray: *Nonlinear least squares and nonlinearly constrained optimization.* Proc. Dundee Conf. Numer. Anal. (1975), Springer Lecture Notes **506**, Berlin, pp. 134–147 (1976).

[7] F. A. Potra, V. Ptak: *Nondiscrete Induction and Iterative Processes.* Pittman Research Notes in Mathematics **103**, Boston (1984).

[8] Wedin, Per–Ake: *The non–linear least squares problem from a numerical point of view, I Geometrical Properties.* Lund University, Department of Computer Sciences (1972).