



HAL
open science

On the Convergence Analysis of a Two-Step Modification of the Gauss-Newton Method

Roman Iakymchuk, Stepan Shakhno

► **To cite this version:**

Roman Iakymchuk, Stepan Shakhno. On the Convergence Analysis of a Two-Step Modification of the Gauss-Newton Method. PAMM. Vol. 14. Special issue: 85th Annual Meeting of the International Association of Applied Mathematics and Mechanics (GAMM), Mar 2014, Erlangen, Germany. pp.813-814. hal-00960300v2

HAL Id: hal-00960300

<https://hal.science/hal-00960300v2>

Submitted on 30 May 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On the Convergence Analysis of a Two-Step Modification of the Gauss-Newton Method

Roman Iakymchuk^{1,2*} and Stepan Shakhno^{3**}

¹ Sorbonne Universités, UPMC Univ Paris 06, ICS, F-75005 Paris, France

² Sorbonne Universités, UPMC Univ Paris 06, UMR 7606, LIP6, F-75005 Paris, France

³ Ivan Franko National University of Lviv, Universitetska Str. 1, 79000 Lviv, Ukraine

We investigate the convergence of a two-step modification of the Gauss-Newton method applying the generalized Lipschitz condition for the first and second order derivatives. The convergence order as well as the convergence radius of the method are studied and the uniqueness ball of solution of the nonlinear least squares problem is examined. Finally, we carry out numerical experiments on a set of well-known test problems.

Copyright line will be provided by the publisher

1 Introduction

Let us consider the nonlinear least squares problem [4]:

$$\min f(x) := \frac{1}{2} F(x)^T F(x), \tag{1}$$

where F is a Fréchet differentiable operator defined on R^n with its values on R^m , $m \geq n$.

For solving the problem (1), we consider a two-step modification of the Gauss-Newton method [2, 3]

$$\begin{cases} x_{k+1} = x_k - [F'(z_k)^T F'(z_k)]^{-1} F'(z_k)^T F(x_k), \\ y_{k+1} = x_{k+1} - [F'(z_k)^T F'(z_k)]^{-1} F'(z_k)^T F(x_{k+1}), \quad k = 0, 1, 2, \dots \end{cases} \tag{2}$$

where $z_k = (x_k + y_k)/2$; x_0 and y_0 ($x_0 = y_0$) are given. In case when $m = n$, this method matches the method proposed by Bartish [1] and Werner [7]. On each iteration, the method (2) computes only one value of functions F and F' . Because of that, the computation cost of each iteration of the method (2) is roughly the same as of the Gauss-Newton method [4]: for calculating y_{k+1} , it is only necessary to perform one backward substitution, which requires $O(n^2)$ floating-point operations (Flops), since the decomposition LL^T of the matrix $F'(z_k)^T F'(z_k)$, which costs $O(n^3/3)$ Flops, is computed for x_{k+1} .

The main goal of this paper is to analyze the local convergence of the method (2). Bartish et al. [2] examined the local convergence of the same using the classical Lipschitz condition for derivatives of the second order, but only for the problem (1) with zero residual. Instead, we study the convergence of the above-mentioned method using the wider generalized Lipschitz conditions [6] for derivatives of the first and second orders: such conditions use an integrable function $L(u)$ instead of the Lipschitz constant L . Besides that, we prove the convergence of the method (2) for the problem (1) with zero as well as non-zero residuals. Moreover, we find both the order and the radius of convergence of the method (2) as well as the uniqueness ball of solution of the problem (1).

2 Convergence Analysis and the Uniqueness of Solution

In this section, we investigate the convergence of the method (2) as well as examine the uniqueness of solution.

Theorem 2.1 *Let $F : R^n \rightarrow R^m$, $m \geq n$, be a twice Fréchet differentiable operator in $D \in R^n$. Assume that (1) has a solution $x_* \in D$ and a Fréchet derivative $F'(x_*)$ has full column rank. Suppose that Fréchet derivatives $F'(x)$ and $F''(x)$ on $B(x_*, r) = \{x \in D : \|x - x_*\| < r\}$ satisfy the Lipschitz conditions with L and N average:*

$$\|F'(x) - F'(x_*)\| \leq \int_0^{\rho(x)} L(u) du, \quad \|F''(x) - F''(y)\| \leq \int_0^{\|x-y\|} N(u) du, \tag{3}$$

where $x, y \in B(x_*, r)$, $\rho(x) = \|x - x_*\|$, $L(u)$ and $N(u)$ are positive nondecreasing functions. Also, the radius $r > 0$ satisfies

$$0 < \frac{(\beta/8) \int_0^r N(u)(r-u)^2 du + \beta r \int_0^{(3/2)r} L(u) du + \sqrt{2}\alpha\beta^2 \int_0^r L(u) du}{1 - \beta \int_0^r L(u) du} \leq r.$$

* Corresponding author: e-mail: roman.iakymchuk@lip6.fr, Phone: +33 (0)1 44 27 87 73

** e-mail: s.shakhno@franko.lviv.ua, Phone: +38 03 22 39 43 04

Then, for all $x_0 = y_0 \in B(x_*, r)$ the sequences $\{x_k\}$ and $\{y_k\}$ generated by the method (2) are well defined, remain in $B(x_*, r)$ for all $k \geq 0$, and converge to x_* such that

$$\begin{aligned} \rho(x_{k+1}) &= \|x_{k+1} - x_*\| \leq \gamma \rho(x_k)^3 + \eta \rho(x_k) \rho(y_k) + \theta [\rho(x_k) + \rho(y_k)]/2, \\ \rho(y_{k+1}) &= \|y_{k+1} - x_*\| \leq \gamma \rho(x_{k+1})^3 + (\eta/3)(\rho(x_k) + \rho(y_k) + \rho(x_{k+1}))\rho(x_{k+1}) + \theta [\rho(x_k) + \rho(y_k)]/2, \\ r_{k+1} &= \max\{\|x_{k+1} - x_*\|, \|y_{k+1} - x_*\|\} \leq q r_k \leq \dots \leq q^{k+1} r_0, \end{aligned}$$

where

$$\begin{aligned} 0 < q &= \frac{(\beta/8) \int_0^{\rho(x_0)} N(u)(\rho(x_0) - u)^2 du + \beta \rho(x_0) \int_0^{(3/2)\rho(x_0)} L(u) du + \sqrt{2} \alpha \beta^2 \int_0^{\rho(x_0)} L(u) du}{\rho(x_0) \left(1 - \beta \int_0^{\rho(x_0)} L(u) du\right)} < 1, \\ \gamma &= \frac{\beta \int_0^{\rho(x_0)} N(u)(\rho(x_0) - u)^2 du}{8 \rho(x_0)^3 \left(1 - \beta \int_0^{\rho(x_0)} L(u) du\right)}, \quad \eta = \frac{\beta \int_0^{(3/2)\rho(x_0)} L(u) du}{\rho(x_0) \left(1 - \beta \int_0^{\rho(x_0)} L(u) du\right)}, \quad \theta = \frac{\sqrt{2} \alpha \beta^2 \int_0^{\rho(x_0)} L(u) du}{\rho(x_0) \left(1 - \beta \int_0^{\rho(x_0)} L(u) du\right)}, \\ \alpha &= \|F(x_*)\|, \quad \beta = \|(F'(x_*)^T F'(x_*))^{-1} F'(x_*)^T\|. \end{aligned} \tag{4}$$

Corollary 2.2 Order of convergence of the iterative process (2) in case of zero residual is equal $1 + \sqrt{2}$.

Theorem 2.3 Suppose x_* satisfies (1) and $F'(x_*)$ has full rank. Besides that, $F(x)$ has a continuous derivative $F'(x)$ in $B(x_*, r)$ that satisfies the Lipschitz condition with L average (3). Let $r > 0$ satisfies

$$\beta \int_0^r L(u)(r - u) du + \alpha \beta_0 \int_0^r L(u) du \leq r,$$

where α and β are defined in (4) and $\beta_0 = \|[F'(x_*)^T F'(x_*)]^{-1}\|$. Then, x_* is a unique solution of the problem (1) in $B(x_*, r)$.

3 Numerical Results

We carried out a set experiments on widely used test problems and compared the number of iterations under which the Gauss-Newton method, the Secant method [5], and the method (2) converge to solution. We used the same initial points for all methods, but different stopping criteria such as $\|F'^T(x_k)F(x_k)\| \leq \varepsilon$ or $\|A_k^T F(x_k)\| \leq \varepsilon$; the later is applied when the method utilizes the divided differences. In Table 1 we present the amount of iterations spent by each methods to compute an approximation to the solution of the examples from [5] with the accuracy $\varepsilon = 10^{-8}$.

Table 1: The number of iterations to solution.

Example	Gauss-Newton method	Secant method [5]	Method (2)
Rosenbrock function	3	3	2
Kowalik-Osborne function	10	17	10
Box-3D function	6	8	5
Gnedenko-Veibull distribution	6	8	4
Freidenstein-Ross function	44	19	8
Wood function	51	74	49

To conclude, the method (2) is not only more efficient than the Gauss-Newton and Secant methods in terms of the convergence order, but also in terms of the amount of iterations to solution on a variety of tests. Furthermore, the method (2) has promising characteristics for parallelization, which we plan to utilize for constructing and developing new parallel methods for solving the problem (1).

Acknowledgements This work undertaken (partially) in the framework of CALSIMLAB is supported by the public grant ANR-11-LABX-0037-01 overseen by the French National Research Agency (ANR) as part of the ‘‘Investissements d’Avenir’’ program (reference: ANR-11-IDEX-0004-02).

References

- [1] M. Ia. Bartish. *Dopov. AN URSR. Ser. A.*, 30:387–391, 1968.
- [2] M. Ia. Bartish and A. I. Chypurko. *Visnyk of Lviv Univ. Ser. Appl. Math. and Infor.*, 1:3–7, 1999.
- [3] M. Ia. Bartish, A. I. Chypurko, and S. M. Shakhno. *Visnyk of Lviv Univ. Ser. Mech. Math.*, 42:35–38, 1995.
- [4] J. M. Dennis and R. B. Schnabel. Prentice-Hall, New York, 1983.
- [5] S. M. Shakhno and O. P. Gnatyshyn. *Applied Mathematics and Computation*, 161:253–264, 2005.
- [6] X. Wang. *IMA Journal of Numerical Analysis*, 20:123–134, 2000.
- [7] W. Werner. *Numer. Math.*, 32:333–342, 1979.