Analysis Local Convergence of Gauss-Newton Method

Rahmi Wahidah Siregar[1], Tulus[2] and Marwan Ramli[3]

Department of Mathematics, Universitas Sumatera Utara, Medan, Indonesia

E-mail: [1]rahmisiregar51@yahoo.com, [2]tulus_jp@yahoo.com and [3]marwan301@yahoo.com

Abstract. The Gauss-Newton method is a very efficient, simple method used to solve nonlinear least-squares problems. This can be seen as a modification of the newton method to find the minimum value of a function. In solving nonlinear problems, the Gauss Newton Algorithm is used to minimize the sum of quadratic function values, which in its completion does not require the calculation or estimate of the derivatives of the two functions f (x) hence numerically more efficient with direct or iterative processes. The Gauss Newton method studied in this study is restricted to functions of one or two variables. The results of Gauss Newton's method analysis consisted of convergence at simple roots and multiple roots. Newton's method often converges quickly, especially when the iteration begins to be close enough to the desired root. However, if iteration begins far from the searched root, this method can be missed without warning. Implementation of this method usually detects and overcomes the convergence failures.

1. Introduction

The development of an increasingly advanced world makes the problem more complex, no exception problems involving mathematical problems. One of the problems that is often encountered in mathematics and science and engineering is to find the root of the equation, ie values that satisfy f (x) = 0 (Proinov, 2009). Numerical methods are techniques in which mathematical problems are formulated in such a way that they can be solved by arithmetic operations. Numerical methods are used to solve problems where analytic calculations can not be used. This numerical method is presented in the form of algorithms that can be calculated quickly and easily. Given that the algorithm developed in the numerical method is the Gauss Newton algorithm then in the algorithm it will appear the term iteration is the repetition of the calculation process.

View the least-squares nonlinear problem:

$$min\|F(x)\|^2 \tag{1}$$
$$x\varepsilon\Omega$$

Where $\Omega \subseteq \mathbb{X}$ is the open set and $F \to Y$ is a non-linearly deferred function, with X and Y being real or complex. The problem arises when the data is good, when $\mathbb{X} = \mathbb{R}^n \ dan \ \mathbb{Y} = \mathbb{R}^m$, where m is the number of observations and n is the number of parameters that Dennis and Schnabel (1996) and Nocedal (1999) have indicated.

$F'(x)$ is denoted by a derivative of F at point $x \in \Omega$. When $F'(x)$ is an invertive derivative of the problem of stationary point found from the problem (1.1) that is the solution of the nonlinear equation $F'(x) * F(x) = 0$, Where A^* denotes the adjoint of operator A, Find the least squares solution of nonlinear equations

$$F(x) = 0 \tag{2}$$

This problem has been extensively studied by Dedieu and Kim (2002) and Dedieu and Shub (2000) for functional analysis and in the Riemannian context by Adler et al. (2002).

Where $F'(x)$ is injektif and has a closed image for all $\in \Omega$, Gauss Newton method finds the stationary point of the above problem, described as follows where point $x_0 \in \Omega$ defines

$$x_{k+1} = x_k - [F'(x_k) * F'(x_k)]^{-1} F'(x_k) * F(x_k), \quad k = 0,1, \dots$$

If the above method converges to $x_* \in \Omega$, then x_* is the stationary point of the problem (1), but we can not conclude that x_* the solution of the problem (1) or $F(x_*) = 0$. To ensure that the stationary point x_* is the solution of the problem (1), we must apply the optimal conditions. It should be pointed out that if $F'(x)$ can be reversed for all x $\in \Omega$, then Gauss Newton's method becomes Newton's method. Literatures discussing the convergence of Newton and Gauss Newton methods include Adler et al. (2002), Alvarez et al. (2008), Argyros and Hilout (2010), Chen (2008), Proinon (2010).

It is also known that the Gauss Newton method may fail or even fail to be properly defined by Dennis (1996) and Dedieu (2000). To ensure that the method is well defined and convergence to the stationary point of the problem some conditions to be worn such as the classical convergence analysis shown by Dennis (1996) and Nocedal (1999) require that Lipschitz comply that the initial iteration is close enough for the solution but can not make us see clearly how big the convergence of spherical radius for analytic functions, Dedineu and Shub (2000) have given an approximation of the convergence radius and criteria for the convergence of the Newtonian Gauss method. One of the obstacles in using Gauss Newton's method is to compute the value of the function derivative. This is not always easy if done manually, especially for certain functions, even if calculations are done by calculator or computer. Therefore it is necessary to find the appropriate software to implement the newton method which does not require manual derivative calculation, matlab can be used for this purpose. The Gauss-Newton algorithm is a very efficient, simple method used to solve nonlinear least-squares problems (Cox et al., 2004). This can be seen as a modification of the newton method to find the minimum value of a function. In solving non-linear problems, the Gauss Newton Algorithm is used to minimize the sum of quadratic function values, which in its completion does not require the calculation or estimate of the derivatives of the two functions f (x) hence numerically more efficient with direct or iterative processes. The Gauss Newton method studied in this study is restricted to functions of one or two variables. The analysis of Gauss Newton's method involves convergence of simple roots and multiple roots. Newton's method often converges quickly, especially when the iteration begins to be close enough to the desired root. However, if iteration begins far from the searched root, this method can be missed without warning. Implementation of this method usually detects and overcomes the convergence failures.

The speed of convergence is basically a value that measures how quickly the solution of iterated results leads to the actual solution. Let $f(x) = 0$ be the equation to be solved with the exact solution x^* dan $x_k (k = 1,2, \dots)$ is the numerical solution of the ke-k iteration result. Difference $|x_k - x^*|$ Measure how far the numerical solution of the actual solution is to the k-iteration or usually referred to as the error (in the k-iteration) and of course the desired property is $|x_{k+1} - x^*| < |x_k - x^*|$. This study will analyze the local convergence of the Gauss Newton method and solve some examples of the application of the Gauss Newton method.

2. Literature Review

2.1 Non-linear least-squares problems

In mathematical terms, optimization usually involves maximizing or minimizing. In the least nonlinear least squares problem is a nonlinear quadratic minimization problem. This is called the least squares because it minimizes the sum of squares of the function. This type of problem occurs just as the model functions into data: if $\phi(x;t)$ represents the function model with t as the independent variable, then every $r_j(x) = \phi(x;t_j) - y_j$, where $d(t_j, y_j)$ is the set of data points. Two common

algorithms to solve the quadratic problem are the Gauss-Newton method and the Levenberg Marquardt Algorithm.

The least squares problem is to minimize the difference between a set of data and model functions that approach the data. Considering a set of data $d(t_j, y_j)$ and the function model $\phi(x; t_j)$, we obtain the function difference with the equation $r_j(x) = \phi(x; t_j) - y_j$ where y_j is the y component at the data point on t_j. Therefore the purpose function f (x) is the sum of the squares of the nonlinear functions:

$$f(x) = \frac{1}{2} \sum_{j=1}^{m} r_j^2(x) \qquad (3)$$

This equation includes the sum of all r_j components of the vector r the remaining components of m given by the vector

$$r(x) = (r_1(x), r_2(x), \ldots, r_m(x))^T$$

When calculating the gradient of f (x), it is necessary to find the residual gradient vector. Jacobian J (x) is the matrix of all $\nabla r_j(x)$:

$$J(x) = \left[\frac{\partial r_j}{\partial x_i}\right] j = 1, \ldots, m; i = 1, \ldots n = \begin{bmatrix} \nabla r_1(x)^T \\ \nabla r_2(x)^T \\ \vdots \\ \nabla r_m(x)^T \end{bmatrix}$$

Gradient and Hessian can be expressed in terms of Jacobian:

$$\nabla f(x) = \sum_{j=1}^{m} r_j(x) \nabla r_j(x) = J(x)^T r(x)$$

$$\nabla^2 f(x) = \sum_{j=1}^{m} r_j(x) \nabla r_j(x)^T + \sum_{j=1}^{m} r_j(x) \nabla^2 r_j(x)$$

$$= J(x)^T J(x) + \sum_{j=1}^{m} r_j(x) \nabla^2 r_j(x)$$

The Gauss-Newton method is a very efficient, simple method used to solve nonlinear least-squares problems (Cox et al., 2004). This can be seen as a modification of the newton method to find the minimum value of a function. In solving non-linear problems, the Gauss Newton Algorithm is used to minimize the sum of quadratic function values, which in its completion does not require the calculation or estimate of the derivatives of the two functions f (x) hence numerically more efficient with direct or iterative processes. The Gauss Newton method studied in this study is restricted to functions of one or two variables. The analysis of Gauss Newton's method involves convergence of simple roots and multiple roots. Newton's method often converges quickly, especially when the iteration begins to be close enough to the desired root. However, if iteration begins far from the searched root, this method can be missed without warning. Implementation of this method usually detects and overcomes the convergence failures.

It is also known that the Gauss Newton method may fail or even fail to be properly defined by Dennis (1996) and Dedieu (2000). To ensure that this method is well defined and convergence to the stationary point of the problem of some conditions to be worn eg a classical convergence analysis is

indicated by see Dennis (1996) and Nocedal (1999) necessitates meeting F 'Lipschiatz that the initial iteration is close enough for the solution But it can not make us see clearly how big the convergence of the spherical radius for analytic functions, Dedineu and Shub (2000) have provided an approximation of the convergence radius and criteria for the convergence of the Newtonian Gauss method.

Several literatures discussing convergence are presented by Jinhai Chen (2006) discussing the convergence of the newtonian gauss method by solving the nonlinear least squares problem, Li (2011) discusses the theory of convex processes with the convergence of repetitive sequences generated by the Gauss-Newton method for problems Convex inclusions, Ferreira and Oliveira (2011) discuss about the local convergence of the proximal Gauss-Newton method to solve nonlinear least squares problems in Hilbert space settings.

2.2 Newton Method

The newton method is one of the open methods for determining root solutions of nonlinear equations or determining one of the roots of a nonlinear equation with f (x) = 0, with the following key principles :

1. This method approaches the curve f (x) with the tangent line (gradient) at a starting point.
2. The next assessment value is the point of intersection between the gradient of the curve   with the x axis.

Let's say : $f(x) = e^x - 5x^2 = 0$, with $x_0 = 1$

Table 1. Iterations of the Newton Method for $f(x) = e^x - 5x^2 = 0$

iterasi	x(k)	f(x)	f'(x)
0	1	-2,2822	-7,2812
1	0,6866	-0,3701	-4,8791
2	0,6107	-0,0228	-4,2653
3	0,6053	0,0003	-4,2212
4	0,6051	0	-4,2210

2.3 Gauss Newton Method

The Gauss-Newton method is a very efficient, simple method used to solve nonlinear least-squares problems (Cox et al., 2004). This can be seen as a modification of the newton method to find the minimum value of a function. In solving non-linear problems, the Gauss Newton Algorithm is used to minimize the sum of quadratic function values, which in its completion does not require the calculation or estimate of the derivatives of the two functions f (x) hence numerically more efficient with direct or iterative processes.

Gauss Newton's Algorithm

Step 0. Take $x_0 \in \mathbb{R}^n, k := 0$.

Step 1: If $\|g_k\| \le \epsilon$, stop.

Step 2: Perform until convergent

   Step 2.1: Complete $(J(x_k)^T J(x_k) s_k = -J^T(x_k) f(x_k)$, for s_k.

   Step 2.2: Set $x_{k+1} = x_k + s_k, k := k + 1$. kembali ke langkah 1. Go back to step 1.

Description: that at each iteration, step 2.1 of this method is equivalent to solving the linearisation of least squares problem.

$$\min_{s} \frac{1}{2} \|J(x_k)s + f(x_k)\|_2^2$$

The Gauss-Newton method can also be written as a fixed point iteration as follows

$$x_{k+1} = G(x_k)$$

Where $G(x) \equiv x - J^+(x)f(x)$ and $J(x)^+ \equiv (J^T(x)J(x))^{-1}J^T(x)$ is defined in Moore -Penrose Pseudo-Invers on J (x). In solving non-linear problems, the Gauss-Newton Algorithm is a modification of the newton method to find the minimum value of a function. Gauss-Newton is important for rapid convergence close to the solution, but like Newton's method, efficiency depends on having an accurate early guess.

2.4 Convergence of the Gauss-Newton Method

An efficient condition for the convergence of the Gauss-Newton method is known in this case where the normal equations for linearisation of least squares problems in (3.16) are completed in step 2.1 at each iteration. Given some of the results. The following assumptions are made in order to build the theory:

A.1 There is $x^* \in \mathbb{R}^n$ such that $J^T(x^*)f(x^*) = 0$

A.2 Jacobian $J(x^*)$ matrix at x^* full rank n.

With the notation ρ (A) to show the spectral radius of the matrix A at n x n, and defined:

$$\varrho = \rho((J(x^*)^T J(x^*))^{-1} Q(x^*))$$

Here's the theorem on the local convergence of the Gauss-Newton method.

Theorem 1: Suppose assuming A.1 and A.2 if $\varrho < 1$, so that iteration of gauss newton converges locally to x^* for that $\varepsilon > 0$ such that $\{x_k\}$ the generation of gauss newton algorithms converges at x^* for all $x_0 \in D \equiv \{x | ||x - x^*||_2 < \varepsilon\}$.

Proof: Theorem 1 has a geometric interpretation as described in previous research by Wedin (1974) and Ake Bjorck (1996) in his research showing that the S surface in $\mathbb{R}^m$ is given by the parametric representation $y = f(x), x \in \mathbb{R}^n$, and let M be the point at S with the coordinates $f(x^*)$, by taking O as the origin of the coornidic system. The OM vector is orthogonal to the plane of contact with the surface S to M.

Theorem 2: In the study of Wedin (1974) suppose assume theorem 1 and that $f(x^*)$ is not zero, so

$$\varrho = ||f(x^*)||_2 \chi$$

Where χ is the maximum maximum curvature of surface S at point M corresponding to the normal direction $w^* = f(x^*)/||f(x^*)||_2$.

In residuals with a zero value, where $f(x^*) = 0$, the relation in equation continues. In this case the origin of O lies on the surface of S and χ denotes the maximum principle curvature S toward normal to the tangent surface at O. When we have $Q(x^*) = 0$ dan karena $\varrho = 0$, the result is still valid.

For Gauss Newton's method to converge, therefore, it is sufficient to maximize the principle of curvature χ of the surface S at the point $f(x^*)$ to satisfy $\frac{1}{\chi} > ||f(x^*)||_2$. This condition applies if and only if $\nabla^2 \phi(x^*)$ is positive definite at x^* and ensures that x^* is a local minimum of the objective function φ indicated by Ake Bjorck (1996). The relation in equation 3.19 shows that the convergence condition in theorem 3.3 is invariant in the transformation of the problem of nonlinear least squares by local diffeomorphism because of the sum $||f(x^*)||_2 \chi$ Properties in the study by Wedin (1974). The evidence of this result depends on the iterative theory for the iterative process with continuous stationary point iterations shown by Ortega and Rheinboldt (1970) studies. This theory ensures local convergence at a linear level. However, in Dennis and Schanbel's (1983) study, the condition is more rigorous for local convergence. Conditions that provide a higher degree of order of convergence can be inferred from this theory.

3. Conclusion

The Gauss-Newton method is a highly efficient iterative method used to solve nonlinear least-squares problems. This can be seen as a modification of the newton method to find the minimum value of a function. The Gauss-Newton method studied in this study is restricted to functions of one or two variables. The analysis of Gauss Newton's method involves convergence of simple roots and multiple roots. Gauss Newton's method often converges quickly, especially when the iteration begins to be close enough to the desired root. But if iteration begins far from the searched root, this method can slip off without warning, its convergence may be slow, it probably will not converge at all. Implementation of this method usually detects and overcomes the convergence failures.

References:

[1]. O.P. Ferreira, Local convergence of Newton's method in Banach space from the viewpoint of the majorant principle, IMA J. Numer. Anal. 29 (3) (2009) 746–759.

[2]. R.L. Adler, J.P. Dedieu, J.Y. Margulies, M. Martens, M. Shub, Newton's method on Riemannian manifolds and a geometric model for the human spine, IMA J. Numer. Anal. 22 (3) (2002) 359–390.

[3]. F. Alvarez, J. Bolte, J. Munier, A unifying local convergence result for Newton's method in Riemannian manifolds, Found. Comput. Math. 8 (2) (2008) 197–226.

[4]. I.K. Argyros, S. Hilout, Improved generalized differentiability conditions for Newton-like methods, J. Complexity 22 (3) (2010) 316–333.

[5]. L. Blum, F. Cucker, M. Shub, S. Smale, Complexity and Real Computation, Springer-Verlag, New York, 1998. With a foreword by Richard M. Karp.

[6]. Chapra, Canale (1996). Metode Numerik Jilid 1. . PT Gelora Aksara Pratama. Erlangga.

[7]. J. Chen, The convergence analysis of inexact Gauss–Newton methods for nonlinear problems, Comput. Optim. Appl. 40 (1) (2008) 97–118.

[8]. J. Chen, W. Li, Convergence of Gauss–Newton's method and uniqueness of the solution, Appl. Math. Comput. 170 (1) (2005) 686–705.

[9]. J. Chen, W. Li, Convergence behaviour of inexact Newton methods under weak Lipschitz condition, J. Comput. Appl. Math. 191 (1) (2006) 143–164.