Applied Numerical Mathematics 111 (2017) 92–110

Contents lists available at ScienceDirect

Applied Numerical Mathematics

www.elsevier.com/locate/apnum

Approximate Gauss–Newton methods for solving underdetermined nonlinear least squares problems

Ji-Feng Bao [a,b,1], Chong Li [c,2], Wei-Ping Shen [d], Jen-Chih Yao [e,3], Sy-Ming Guu [f,*,4]

[a] School of Mathematics, Physics and Information Science, Zhejiang Ocean University, Zhoushan, Zhejiang 316022, PR China
[b] Key Laboratory of Oceanographic Big Data Mining & Application of Zhejiang Province, Zhoushan, Zhejiang 316022, PR China
[c] Department of Mathematics, Zhejiang University, Hangzhou 310027, PR China
[d] Department of Mathematics, Zhejiang Normal University, Jinhua 321004, PR China
[e] Center for General Education, China Medical University, Taichung 40402, Taiwan
[f] Graduate Institute of Business and Management, Department of Neurology, College of Management, Chang Gung University, Chang Gung Memorial Hospital, Taoyuan City, Taiwan

ARTICLE INFO

Article history:
Received 24 October 2015
Received in revised form 15 August 2016
Accepted 21 August 2016
Available online 2 September 2016

Keywords:
Nonlinear least squares problems
Approximate Gauss–Newton methods
Lipschitz condition

ABSTRACT

We propose several approximate Gauss–Newton methods, i.e., the truncated, perturbed, and truncated-perturbed GN methods, for solving underdetermined nonlinear least squares problems. Under the assumption that the Fréchet derivatives are Lipschitz continuous and of full row rank, Kantorovich-type convergence criteria of the truncated GN method are established and local convergence theorems are presented with the radii of convergence balls obtained. As consequences of the convergence results for the truncated GN method, convergence theorems of the perturbed and truncated-perturbed GN methods are also presented. Finally, numerical experiments are presented where the comparisons with the standard inexact Gauss–Newton method and the inexact trust-region method for bound-constrained least squares problems [23] are made.

1. Introduction

Let D be an open set and $f : D \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be a nonlinear operator with the continuous Fréchet derivative denoted by f'. Let $\phi : D \subseteq \mathbb{R}^n \to \mathbb{R}$ be defined by

$$\phi(x) := \frac{1}{2} f(x)^T f(x), \quad \text{for each } x \in D.$$

* Corresponding author.
E-mail addresses: feelwinter@163.com (J.-F. Bao), cli@zju.edu.cn (C. Li), shenweiping@zjnu.cn (W.-P. Shen), yaojc@mail.cmu.edu.tw (J.-C. Yao), iesmguu@mail.cgu.edu.tw, iesmguu@gmail.com (S.-M. Guu).
[1] This author was supported by the Open Foundation from Marine Sciences in the Most Important Subjects of Zhejiang (20160101) and Scientific Research Foundation of Zhejiang Ocean University (21065013613).
[2] This author was supported in part by the National Natural Science Foundation of China (grants 11571308, 11371325).
[3] This author is supported in part by the Grant MOST 105-2221-E-039-MY3.
[4] Research was partially supported by MOST 105-2221-E-182-052 and BMRPD17.

http://dx.doi.org/10.1016/j.apnum.2016.08.007
0168-9274/

The nonlinear least squares problem (NLSP) of f is defined by

$$\min_{x\in D} \phi(x). \tag{1.1}$$

The NLSP (1.1) arises most commonly from data-fitting applications [3,8,14]. This model is particularly useful in the formulation of a parameterized system in which a chemical, physical, financial, or economic application could use the function ϕ to measure the discrepancy between the model and the output of the system at various observation points. By minimizing the function ϕ, they select the parameter values which best match the model to the data [2,21].

Note that finding the stationary points of ϕ is equivalent to solving the nonlinear gradient equation

$$\nabla\phi(x) := f'(x)^T f(x) = 0. \tag{1.2}$$

Based on this equivalence, Newton's method for solving nonlinear equation (1.2) can be applied to solving NLSP (1.1) (cf. [13]). However, Newton's method for solving (1.2) requires the computation of the Hessian matrix of ϕ at each iteration x_k:

$$\nabla^2\phi(x_k) := f'(x_k)^T f'(x_k) + R(x_k), \tag{1.3}$$

where $R(x_k)$ is the second order term which may be difficult to obtain especially for large scale problems [9]. In order to make the procedure more efficient, Newton's method could be approximated by ignoring the second-order term in the Hessian matrix (1.3) and this yields the Gauss–Newton (GN) method. More precisely, the GN step d_k is defined by the minimum norm solution of the following equation:

$$f'(x_k)^T f'(x_k)d_k = -f'(x_k)^T f(x_k). \tag{1.4}$$

That is, in terms of $f'(x_k)^\dagger$, the GN step d_k is given by

$$d_k := -f'(x_k)^\dagger f(x_k), \tag{1.5}$$

where $f'(x_k)^\dagger$ represents the Moore–Penrose inverse of $f'(x_k)$ (see section 2 for the definition). Many authors have studied the local as well as semi-local convergence of the GN method; see for example [6,11,12,16,21,22].

However, as expressed in (1.4), the GN method has the following two disadvantages from the point of view of practical calculation: (a) it requires the computation of the Fréchet derivative $f'(x_k)$ at each outer iteration; (b) it requires solving the equation (1.4) exactly at each outer iteration. This sometimes makes the GN method inefficient especially when the problem size is large. Noting these two disadvantages, Gratton et al. designed in [9] some approximate GN methods for solving the NLSP (1.1) with $m \geq n$. To overcome the disadvantage in (a), the perturbed GN method was proposed in [9] where the Fréchet derivative $f'(x_k)$ was replaced by a perturbed Fréchet derivative J_k which is much easier or computationally less expensive to calculate. More precisely, the perturbed GN method finds the step d_k such that

$$J_k^T J_k d_k = -J_k^T f(x_k). \tag{1.6}$$

As for the drawback in (b), a natural approach is to solve the equation (1.4) inexactly instead of exactly and this yields the truncated GN method proposed in [9]. For the residual r_k and the iteration x_k, the truncated GN method finds the step d_k such that

$$f'(x_k)^T f'(x_k)d_k = -f'(x_k)^T f(x_k) + r_k. \tag{1.7}$$

The third approximate GN method designed in [9] is the truncated-perturbed GN method which avoids both disadvantages in (a) and (b) and, for the residual r_k and the iteration x_k, finds the step d_k such that

$$J_k^T J_k d_k = -J_k^T f(x_k) + r_k.$$

Under the assumption that the (perturbed) Fréchet derivatives are of full column rank, the truncated, perturbed, and truncated-perturbed GN methods were proved to be convergent in [9]. However, in the case when $m < n$ or without the full column rank assumption of the (perturbed) Fréchet derivatives, the sequences $\{x_k\}$ generated by these approximate GN methods are not convergent in general. For example, we consider the function $f : \mathbb{R}^2 \to \mathbb{R}^1$ defined by

$$f(x) := x^{(1)}x^{(2)}, \quad \text{for each } x = (x^{(1)}, x^{(2)})^T \in \mathbb{R}^2.$$

Let $x_0 = (x_0^{(1)}, x_0^{(2)})^T$ be any point in $\mathbb{R}^2$. For each $k \geq 1$, we define $x_k := (0, k!x_0^{(2)})^T$. Then $d_0 = (-x_0^{(1)}, 0)^T$ and $d_k = (0, kk!x_0^{(2)})^T$ for each $k \geq 1$. Thus, one can check that the sequences $\{x_k\}$ and $\{d_k\}$ satisfy (1.7) with $r_k \equiv 0$. This implies that $\{x_k\}$ is generated by the truncated GN method proposed in [9] (with $r_k \equiv 0$). Obviously, $\{x_k\}$ doesn't converge. (Note that in the case when each $r_k \equiv 0$ and each search direction d_k is chosen to be the minimum norm solution of (1.7), the approximate GN method is reduced to the GN method; hence, $\{x_k\}$ is convergent.)

The purpose of this paper is, in the case when $m \leq n$ (i.e., the NLSP (1.1) is underdetermined), trying to propose some approximate GN methods for solving the NLSP (1.1) and study the convergence of the proposed method under the full row rank assumption. For this purpose, note that, in the case when $f'(x_k)$ is of full row rank, $f'(x_k)^\dagger = f'(x_k)^T(f'(x_k)f'(x_k)^T)^{-1}$ (see section 2 for details) and thus, solving (1.5) is equivalent to solving the equation