Applied Mathematics and Computation 241 (2014) 401–408

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

Local convergence analysis of proximal Gauss–Newton method for penalized nonlinear least squares problems

Ioannis K. Argyros [a], Á. Alberto Magreñán [b,*]

[a] Cameron University, Department of Mathematics Sciences, Lawton, OK 73505, USA
[b] Universidad Internacional de La Rioja, Departamento de TFG/TFM, 26005 Logroño, La Rioja, Spain

ARTICLE INFO

Keywords:
Least squares problems
Proximal Newton–Gauss methods
Hilbert space
Majorant function
Center majorant function

ABSTRACT

We present a local convergence analysis of the proximal Gauss–Newton method for solving penalized nonlinear least squares problems in a Hilbert space setting. Using more precise majorant conditions than in earlier studies such as (Allende and Gonçalves) [1], (Ferreira et al., 2011) [9] and a combination of a majorant and a center majorant function, we provide: a larger radius of convergence; tighter error estimates on the distances involved and a clearer relationship between the majorant function and the associated least squares problem. Moreover, these advantages are obtained under the same computational cost as in earlier studies using only the majorant function.

1. Introduction

Let $\mathcal{X}$ and Y be Hilbert spaces. Let $\mathcal{D} \subseteq \mathcal{X}$ be open set and $F : \mathcal{D} \longrightarrow \mathcal{Y}$ be continuously Fréchet-differentiable. Moreover, let $J : \mathcal{D} \to \mathbb{R} \cup \{\infty\}$ be proper, convex and lower semicontinuous. In this study we are concerned with the problem of approximating a locally unique solution $x^\star$ of the penalized nonlinear least squares problem

$$\min x \in \mathcal{D} \|F(x)\|^2 + J(x). \tag{1.1}$$

A solution $x^\star \in \mathcal{D}$ of (1.1) is also called a least squares solution of the equation $F(x) = 0$.

Many problems from computational sciences and other disciplines can be brought in a form similar to Eq. (1.1) using Mathematical Modelling [3,6,14,16]. For example in data fitting, we have $\mathcal{X} = \mathbb{R}^i, \mathcal{Y} = \mathbb{R}^j, i$ is the number of parameters and j is the number of observations.

The solution of (1.1) can rarely be found in closed form. That is why the solution methods for these equations are usually iterative. In particular, the practice of numerical analysis for finding such solutions is essentially connected to Newton-type methods [1–3,5,4,6,7,14,17]. The study about convergence matter of iterative procedures is usually centered on two types: semilocal and local convergence analysis. The semilocal convergence matter is, based on the information around an initial point, to give criteria ensuring the convergence of iterative procedures; while the local one is, based on the information around a solution, to find estimates of the radii of convergence balls. A plethora of sufficient conditions for the local as well as the semilocal convergence of Newton-type methods as well as an error analysis for such methods can be found in [1–7,9,11–20].

* Corresponding author.
E-mail addresses: iargyros@cameron.edu (I.K. Argyros), angel.magrenan@unir.net (Á.A. Magreñán).

http://dx.doi.org/10.1016/j.amc.2014.04.087
0096-3003/

If $J = 0$, we obtain the well known Gauss–Newton method defined by

$$x_{n+1} = x_n - F'(x_n)^+ F(x_n), \quad \text{for each } n = 0, 1, 2, \ldots, \tag{1.2}$$

where $x_0 \in \mathcal{D}$ is an initial point [12] and $F'(x_n)^+$ is the Moore–Penrose inverse of the linear operator $F'(x_n)$. In the present paper we use the proximal Gauss–Newton method (to be precised in Section 2, see (2.6)) for solving penalized nonlinear least squares problem (1.1). Notice that if $J = 0, x^\star$ is a solution of (1.1), $F(x^\star) = 0$ and $F'(x^\star)$ is invertible, then the theories of Gauss–Newton methods merge into those of Newton method. A survey of convergence results under various Lipschitz-type conditions for Gauss–Newton-type methods can be found in [2,6] (see also [5,9,10,12,15,18]). The convergence of these methods requires among other hypotheses that F' satisfies a Lipschitz condition or F'' is bounded in $\mathcal{D}$. Several authors have relaxed these hypotheses [4,8–10,15]. In particular, Allende and Gonçalves [1] and Ferreira et al. [9,10] have used the majorant condition in the local as well as semilocal convergence of Newton-type method. Argyros and Hilout [3–7] have also used the majorant condition to provide a tighter convergence analysis and weaker convergence criteria for Newton-type method. The local convergence of inexact Gauss–Newton method was examined by Ferreira et al. [9] using the majorant condition. It was shown that this condition is better that Wang's condition [15,20] in some sence. A certain relationship between the majorant function and operator F was established that unifies two previously unrelated results pertaining to inexact Gauss–Newton methods, which are the result for analytical functions and the one for operators with Lipschitz derivative.

In [7] motivated by the elegant work in [10] and optimization considerations we presented a new local convergence analysis for inexact Gauss–Newton-like methods by using a majorant and center majorant function (which is a special case of the majorant function) instead of just a majorant function with the following advantages: larger radius of convergence; tighter error estimates on the distances $\|x_n - x^\star\|$ for each $n = 0, 1, \cdots$ and a clearer relationship between the majorant function and the associated least squares problems (1.1). Moreover, these advantages are obtained under the same computational cost, since as we will see in Section 3 and Section 4, the computation of the majorant function requires the computation of the center-majorant function. Furthermore, these advantages are very important in computational mathematics, since we have a wider choice of initial guesses x_0 and fewer computations to obtain a desired error tolerance on the distances $\|x_n - x^\star\|$ for each $n = 0, 1, \cdots$. In the present paper, we obtain the same advantages over the work by Allende and Gonçalves [1] but using the proximal Gauss–Newton method [6,18].

The paper is organized as follows. In order to make the paper as self contained as possible, we provide the necessary background in Section 2. Section 3 contains the local convergence analysis of inexact Gauss–Newton-like methods. Some proofs are abbreviated to avoid repetitions with the corresponding ones in [18]. Special cases and applications are given in the concluding Section 4.

2. Background

Let $U(x, r)$ and $\overline{U}(x, r)$ stand, respectively, for the open and closed ball in $\mathcal{X}$ with center $x \in \mathcal{D}$ and radius $r > 0$. Let $A : \mathcal{X} \longrightarrow \mathcal{Y}$ be continuous linear and injective with closed image, the Moore–Penrose inverse [3] $A^+ : \mathcal{Y} \longrightarrow \mathcal{X}$ is defined by $A^+ = (A^\star A)^{-1} A^\star$. $\mathcal{I}$ denotes the identity operator on $\mathcal{X}$ (or $\mathcal{Y}$). Let $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ be the space of bounded linear operators from $\mathcal{X}$ into $\mathcal{Y}$. Let $M \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$, the $Ker(M)$ and $Im(M)$ denote the Kernel and image of M, respectively and M^* its adjoint operator. Let $M \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ with a closed image. Recall that the Moore–Pentose inverse of M is the linear operator $M^+ \in \mathcal{L}(\mathcal{Y}, \mathcal{X})$ which satisfies

$$MM^+M = M, \quad M^+MM^+ = M, \quad (MM^+)^* = MM^+, \quad (M^+M)^* = M^+M. \tag{2.1}$$

It follows from (2.1) that if $\prod_S$ denotes the projection of X onto subspace S, then

$$M^+M = I_X - \prod_{Ker(M)}, \quad MM^+ = \prod_{Im(M)}. \tag{2.2}$$

Moreover, if M is injective, then

$$M^+ = (M^*M)^{-1}M^*, \quad M^+M = I_X, \quad \|M^+\|^2 = \|(M^*M)^{-1}\|. \tag{2.3}$$

Lemma 2.1 (*[3,6,14] Banach's Lemma*). *Let $A : \mathcal{X} \longrightarrow \mathcal{X}$ be a continuous linear operator. If $\|A - \mathcal{I}\| < 1$ then $A^{-1} \in \mathcal{L}(\mathcal{X}, \mathcal{X})$ and $\|A^{-1}\| \leqslant 1/(1 - \|A - \mathcal{I}\|)$.*

Lemma 2.2 (*[1,3,6,10]*). *Let $A, E : \mathcal{X} \longrightarrow \mathcal{Y}$ be two continuous linear operators with closed images. Suppose $B = A + E$, A is injective and $\|EA^+\| < 1$. Then, B is injective.*

Lemma 2.3 (*[1,3,6,10]*). *Let $A, E : \mathcal{X} \longrightarrow \mathcal{Y}$ be two continuous linear operators with closed images. Suppose $B = A + E$ and $\|A^+\| \|E\| < 1$. Then, the following estimates hold*