Vol. 6 No. 4 JOURNAL OF COMPUTATIONAL MATHEMATICS October 1988

A MEMORYLESS AUGMENTED GAUSS–NEWTON METHOD FOR NONLINEAR LEAST-SQUARES PROBLEMS*

DENNIS, J. E. JR.

(*Mathematical Sciences Department, Rice University, Houston, Texas, USA* 77251)

SHENG SONG–BAI (盛松柏)

(*Nanjing University, Nanjing, China*)

PHUONG ANH VU

(*International Mathematical and Statistical Libraries*, 7500 *Bellaire Blvd, Houston, Texas, USA*)

Abstract

In this paper, we develop, analyze, and test a new algorithm for nonlinear least–squares problems. The algorithm uses a BFGS update of the Gauss–Newton Hessian when some heuristics indicate that the Gauss–Newton method may not make a good step. Some important elements are that the secant or quasi–Newton equations considered are not the obvious ones, and the method does not build up a Hessian approximation over several steps. The algorithm can be implemented easily as a modification of any Gauss–Newton code, and it seems to be useful for large residual problems.

§ 1. Introduction

Nonlinear least–squares problems are frequently encountered in practical optimization, and they are also of interest to the algorist because of their highly structured nature. In this paper, we suggest another way to use this structure in an attempt to increase the efficiency of the trust–region–Gauss–Newton or Levenberg–Marquardt algorithm ([6], [8]).

The algorithm presented here is inspired by NL2SOL ([3], [4]), in that it chooses at each iteration whether to use a Gauss–Newton quadratic model or a variable metric augmentation of the Gauss–Newton model to define the next iterate. The difference is that the variable metric augmentation used here requires less storage, less algebra, and less code than NL2SOL. However, it seems to have no better theoretical justification than the Gauss–Newton method. Still, it seems to use fewer residual and Jacobian computations than the Gauss–Newton for some large residual problems and to require little additional arithmetic at each iteration. Conversation with NL2SOL users encouraged us to undertake this research, and we publish it now in hopes that they will find it helpful and that our colleagues will find it an interesting use of secant updating ideas.

* Received April 1, 1987.

Section 2 explains the augmented local model in its various forms, and points out some overlap between our ideas and those of Al-Baali and Fletcher[1]. Section 3 contains a unified local convergence proof under standard Gauss-Newton-type assumptions for all combinations of the methods presented here. Section 4 describes a model-switching strategy and the resulting hybrid algorithm that adaptively decides whether to use the Gauss-Newton model or an augmentation at each iteration. Section 5 compares an experimental implementation of the algorithms suggested here to the LMDER implementation of the Gauss-Newton method and the NL2S1 routine from NL2SOL.

§ 2. The Augmented Model

Let F: $\Omega \subset R^p \to R^n$ be continuously differentiable, and consider the nonlinear least-squares problem of finding a local minimizer x_* for

$$\phi(x) = \frac{1}{2} F(x)^T F(x) = \frac{1}{2} \sum_{i=1}^{n} (f_i(x))^2. \tag{2.1}$$

The classical algorithm for this problem is the Gauss-Newton method which can be thought of in two ways:

First, we can linearize $F(x) - F(x_c)$ about the current parameter vector x_c to obtain the local affine model for $F(x)$,

$$F(x) \approx F(x_c) + J_c(x - x_c),$$

where $J_c = J(x_c) = F'(x_c) = \left(\frac{\partial f_i}{\partial x_j}(x_c)\right)$. Then we can seek to improve x_c by taking the next estimate x_+ to be the value of the parameter vector that solves the linear least-squares problem defined by the local affine model.

The sum-of-squares-of-residuals of this model is

$$\frac{1}{2}[F(x_c) + J_c(x - x_c)]^T [F(x_c) + J_c(x - x_c)], \tag{2.2}$$

and it can be viewed as a local quadratic model of $\phi(x)$ of the form

$$\phi(x) \approx m_c^{GN}(x) \equiv \phi(x_c) + \nabla\phi(x_c)^T(x - x_c) + \frac{1}{2}(x - x_c)^T J_c^T J_c (x - x_c). \tag{2.3}$$

A second way is to view this local quadratic model as an approximation to the Newton model

$$m_c^N(x) = \phi(x_c) + \nabla\phi(x_c)^T(x - x_c) + \frac{1}{2}(x - x_c)^T \nabla^2\phi(x_c)(x - x_c) \tag{2.4}$$

where

$$\nabla^2\phi(x_c) - J_c^T J_c = \sum_{i=1}^{n} f_i(x_c)\nabla^2 f_i(x_c) \equiv S(x_c) \tag{2.5}$$

is approximated by the zero matrix. It is easy to reason from either derivation that the difference between the two models depends on the size of the residuals $F(x_c)$ and on how nearly affine F is in a neighborhood of x_c.

Aside from the obvious advantage that the Gauss-Newton method has of not having to compute or make assumptions about the n $p \times p$ Hessians $\nabla^2 f_i(x_c)$, $i = 1, \cdots,$