THE LEVENBERG-MARQUARDT ALGORITHM: IMPLEMENTATION AND THEORY*

Jorge J. Moré

1. Introduction

Let $F: R^n \to R^m$ be continuously differentiable, and consider the nonlinear least squares problem of finding a local minimizer of

$$\Phi(x) = \frac{1}{2} \sum_{i=1}^{m} f_i^2(x) = \frac{1}{2} \|F(x)\|^2 . \tag{1.1}$$

Levenberg [1944] and Marquardt [1963] proposed a very elegant algorithm for the numerical solution of (1.1). However, most implementations are either not robust, or do not have a solid theoretical justification. In this work we discuss a robust and efficient implementation of a version of the Levenberg-Marquardt algorithm, and show that it has strong convergence properties. In addition to robustness, the main features of this implementation are the proper use of implicitly scaled variables, and the choice of the Levenberg-Marquardt parameter via a scheme due to Hebden [1973]. Numerical results illustrating the behavior of this implementation are also presented.

Notation. In all cases $\|\cdot\|$ refers to the ℓ_2 vector norm or to the induced operator norm. The Jacobian matrix of F evaluated at x is denoted by $F'(x)$, but if we have a sequence of vectors $\{x_k\}$, then J_k and f_k are used instead of $F'(x_k)$ and $F(x_k)$, respectively.

2. Derivation

The easiest way to derive the Levenberg-Marquardt algorithm is by a linearization argument. If, given $x \in R^n$, we could minimize

$$\Psi(p) = \|F(x+p)\|$$

as a function of p, then x+p would be the desired solution. Since Ψ is usually a nonlinear function of p, we linearize F(x+p) and obtain the linear least squares problem

$$\psi(p) = \|F(x) + F'(x)p\| .$$

Of course, this linearization is not valid for all values of p, and thus we consider the constrained linear least squares problem

---

*Work performed under the auspices of the U.S. Energy Research and Development Administration

(2.1) $$\min\{\psi(p): \|Dp\| \leq \Delta\} .$$

In theory D is any given nonsingular matrix, but in our implementation D is a diagonal matrix which takes into account the scaling of the problem. In either case, p lies in the hyperellipsoid

(2.2) $$E = \{p: \|Dp\| \leq \Delta\} ,$$

but if D is diagonal, then E has axes along the coordinate directions and the length of the ith semi-axis is Δ/d_i.

We now consider the solution of (2.1) in some generality, and thus the problem

(2.3) $$\min\{\|f+Jp\|: \|Dp\| \leq \Delta\}$$

where $f \in R^m$ and J is any m by n matrix. The basis for the Levenberg-Marquardt method is the result that if p^* is a solution to (2.3), then $p^* = p(\lambda)$ for some $\lambda \geq 0$ where

(2.4) $$p(\lambda) = -(J^TJ + \lambda D^TD)^{-1}J^Tf .$$

If J is rank deficient and $\lambda = 0$, then (2.4) is defined by the limiting process

$$Dp(0) \equiv \lim_{\lambda \to 0^+} Dp(\lambda) = -(JD^{-1})^+f .$$

There are two possibilities: Either $\lambda = 0$ and $\|Dp(0)\| \leq \Delta$, in which case $p(0)$ is the solution to (2.3) for which $\|Dp\|$ is least, or $\lambda > 0$ and $\|Dp(\lambda)\| = \Delta$, and then $p(\lambda)$ is the unique solution to (2.3).

The above results suggest the following iteration.

(2.5) Algorithm

(a) Given $\Delta_k > 0$, find $\lambda_k \geq 0$ such that if

$$(J_k^TJ_k + \lambda_k D_k^TD_k)p_k = -J_k^Tf_k ,$$

then either $\lambda_k = 0$ and $\|D_kp_k\| \leq \Delta_k$, or $\lambda_k > 0$ and $\|D_kp_k\| = \Delta_k$.

(b) If $\|F(x_k+p_k)\| < \|F(x_k)\|$ set $x_{k+1} = x_k+p_k$ and evaluate J_{k+1}; otherwise set $x_{k+1} = x_k$ and $J_{k+1} = J_k$.

(c) Choose Δ_{k+1} and D_{k+1}.

In the next four sections we elaborate on how (2.5) leads to a very robust and efficient implementation of the Levenberg-Marquardt algorithm.