Newton-Type Methods for Nonlinear Least Squares Using Restricted Second Order Derivative Information

H. Schwetlick[1]

Abstract: For minimizing the nonlinear sum of squares $f(x) := \frac{1}{2}\sum_{i=1}^{m} F_i(x)^2$ where $F : \mathbb{R}^n \to \mathbb{R}^m$ with $m \geq n$, the Gauss-Newton model

$$\varphi_{GN}(s) := g(x)^T s + \frac{1}{2} s^T G(x) s = \frac{1}{2}\|F(x) + F'(x)s\|^2 \approx f(x+s) - f(x)$$

where $g(x) := \nabla f(x) = F'(x)^T F(x)$, $G(x) := F'(x)^T F'(x)$ is often used as basic approximation. It requires only first order derivatives of F, and related damped or trust region Gauss-Newton methods usually work well as long as the Jacobian $F'(x)$ has full rank n and is not too badly conditioned. Otherwise, Gauss-Newton techniques usually break down, and the Newton model

$$\varphi_N(s) := g(x)^T s + \frac{1}{2} s^T H(x) s, \quad H(x) := \nabla^2 f(x) = G(x) + S(x)$$

where $S(x) := F(x) \circ F''(x) := \sum_{i=1}^{m} F_i(x) \cdot \nabla^2 F(x)$ is the alternative choice. It is, however, more expensive since it uses the Hessian H of f which contains the term S which depends on all second order derivatives of F.

In order to reduce the costs for evaluating S, a *restricted* Newton model is proposed where S is replaced by its restriction B onto the q-dimensional subspace spanned by the eigenvectors $\{u^j\}_{j=1}^{q}$ which belong to the q smallest eigenvalues of the Gauss-Newton matrix G where $q \leq n$ is a number of moderate size which may vary from step to step. Hence, B is defined by

$$B = ZZ^T S\, ZZ^T = Z\hat{S}Z^T \quad \text{with} \quad \hat{S} := (\hat{s}_{ij}) = Z^T S Z \in \mathbb{R}^{q \times q}$$

where $Z := [u^1, \ldots, u^q] \in \mathbb{R}^{n \times q}$. This means that B is chosen such that $u^T B v = u^T S v$ for all $u, v \in \operatorname{im} Z$ and $\|B\|_F$ is minimal. Then only the $q(q+1)/2$ different entries

$$\hat{s}_{ij} = F(x)^T F''(x)[u^i, u^j] = F(x)^T \lim_{\alpha,\beta \to 0} \frac{\partial^2 F(x + \alpha u^i + \beta u^j)}{\partial\alpha\, \partial\beta}, \quad i \leq j,$$

of $\hat{S}$ which contain second order derivatives have to be computed. This can preferably be done via automatic differentiation, or one can approximate them via divided differences as indicated above which requires $\sim q^2/2$ additional evaluations of F.

The restricted approximation is used within a trust region approach with adaptively chosen q, and corresponding numerical results are reported.

[1] Department of Mathematics, Technical University of Dresden
D-01062 Dresden, Germany
schwetlick@math.tu-dresden.de