On a Trust Algorithm with Initial Radius Chosen by a Normed Metric

By

F.M. Aderibigbe
Department of Mathematical Sciences
University of Ado-Ekiti
Ado-Ekiti, Nigeria

T.A. Adewale
Department of Mathematical Sciences
Ondo State University
Akungba-Akoko, Nigeria

Abstract

A trust region algorithm is presented for the solution of nonlinear unconstrained optimization problems. The theory allows for the use of general metric in choice of the initial trust region radius (TRR). Specifically, the initial trust region radius is chosen in a non-normed but metric linear space. The algorithm is implemented in conjunction with separable programming technique using the Rosenbrock's Banana-shaped valley function for numerical experimentation.

AMS 1998: 90C

Keywords:

Non-normed but metric linear spaces, separable programming technique, trust region radius, Rosenbrock's banana-shaped valley function.

1. Introduction

Trust region methods for nonlinear optimization problems have become very popular over the last 15 years. One possible explanation of their success is their remarkable numerical reliability associated with the existence of a sound and complete convergence theory [2–4, 7–9]. The fact that they efficiently handle non-convex problems has also been considered and advantage.

In trust region methods [1–5, 7–2], the calculation of the step between iterates requires the solution of a problem of the form

$$\min\{\Psi(w) : \|w\| \le \Delta\} \tag{1.1}$$

where Δ is a positive parameter, $\|\cdot\|$ is the Euclidean norm in IR, and

$$\Psi(w) \equiv \Delta f(x)^T \underline{w} + \frac{1}{2}\underline{w}^T B \underline{w} \qquad (1.2)$$

with $\nabla f \in IR^N$ and $B \in IR^{n \times n}$ a symmetric matrix. The quadratic function Ψ generally represents a local model to the objective function defined by interpolatory data at an iterate and thus it is important to be able to solve (1.1) for any symmetry matrix *B*, in particular, for a matrix *B* with negative eigenvalues. In trust region methods it is sometimes helpful to include a scaling matrix for the variables. In this case, problem (1.1) is replaced by

$$\min\{\Psi(\underline{v})' : \|s\underline{v}\| \le \Delta\} \qquad (1.3)$$

where $s \in IR^{n \times n}$ is a non-singular matrix. The change of variables $s\underline{v} = \underline{w}$ shows that problem (1.3) is equivalent to

$$\min\{\Psi(w) : \|w\| \le \Delta\} \qquad (1.4)$$

where $\Psi(\underline{w}) \equiv \left(s^{-1}w\right)$, and that the solutions of (1.3) and (1.4) are related by $s\underline{v} \equiv \underline{w}$. Because of this equivalence, we only consider problem (1.1). Also, note that if as is usually the case, *S* is a diagonal matrix, then it is easy to explicitly carry out the change of variable and solve problem (1.4).

The use of a trust region method in a nonlinear optimization problem requires the solution of many problems of type (1.1). These problems do not usually require accurate solutions, but in all cases we must be able to find an approximate solution with a reasonable amount of computational effort, and the approximate solution found must guarantee that the trust region methods.

Goldfeld, Quandt and Trotter (1966), Hebden (1973), Fletcher (1980), Gay (1981) and Sorensen (1982), have discussed (1.1) in connection with trust region methods. Their algorithms are based on the fact that if (1.1) has a solution on the boundary of $\{\underline{w} : \|w\| \le \Delta\}$, then, in most cases, a solution of (1.1) can be found by determining $\mu \ge 0$ such that $B + \mu I$ is positive definite and

$$\left\|(B + \mu I)^{-1} \underline{\nabla f}\right\| = \Delta . \qquad (1.5)$$

In the "hard case", equation (1.5) has no solution with $B + \mu I$ positive definite, and this leads to numerical difficulties. Hebden (1973) proposed an algorithm for the solution of (1.1) which is basically sound except for its treatment of the hard case. Gay (1981) improved Hebden's scheme and showed that under certain conditions approximate solution determined by his algorithm is nearly optimal. His algorithm, however, may require a large number of iterations.

Other authors of trust region approach include R.H. Byrd, R.B. Schnabel, J.V. Burke, Reinsch, Dembo, Steihaug, L.C.W. Dixion, J.J. More, Cells, Tapia, Dennis Jr., Conn, Gould, Toint, Startenger, Stewart, M.J.D. Powell, Y. Yuan, Vardi, Shultz, M. El-Alem (who provided a convergence theory for the methods of Celis, Dennis and Tapia), Toraldo, Bertsekas.

We propose an algorithm for the solution of (1.1) which is guaranteed to produce a nearly optimal solution in a finite number of steps.

2. Problem Formulation: Trust Region Methods in Unconstrained Minimization

Let $\Re^n \rightarrow \Re$ be a twice continuously differentiable function with gradiant $\underline{\Delta} f$ and Hessian $\underline{\Delta}^2 f$. In Newton's method with a trust region strategy, each iterate x^2 has a bound Δ^2 such that

$$f\left(\underline{x}^2 + \underline{w}\right) = f\left(x^k\right) + \psi^k(\underline{w}), \ \left\|\underline{w}\right\| \leq \Delta^k. \tag{2.1}$$

where

$$\psi^k(\underline{w}) = \underline{\nabla} f\left(x^T\right)\underline{w} + \frac{1}{2}\underline{w}^T \nabla^2 f\left(x^k\right)\underline{w}. \tag{2.2}$$

This implies that ψ^k

Is a model of the reduction of *f* within a neighbourhood of the iterate x^k. This suggests that it may be desirable to compute a step $\underline{p}^k$ which approximately solves the problem

$$\min\left\{\psi^k(\underline{w}) : \left\|\underline{w}\right\| \leq \Delta^k\right\}. \tag{2.3}$$

If the step is satisfactory in the sense that $\underline{x}^k + \underline{p}^k$ produces a sufficient reduction in *f*, then Δ^k can be increased, if the step is unsatisfactory then Δ^k should be decreased. In this paper, we have chosen the initial radius Δ° by a no-normed metric. The motivation for this choice is the inequality.

$$\frac{|A+B|}{1+|A+B|} \leq \frac{|A|}{1+|A|} + \frac{|B|}{1+|B|} \tag{2.4}$$

which is valid for *A* and *B* real or complex [12]. A proof of (2.4) follows readily from the observation that $t(1+t)^{-1}$ increases as the real variable *t* increases if $t > -1$ (checking by computing the derivative). It can also be shown that if *d* is a metric on a non-empty set *x*, then the function

$$f\left(x_1, x_2\right) = \frac{d\left(x_1, x_2\right)}{1 + d\left(x_1, x_2\right)}, x_1, x_2 \in X \tag{2.5}$$

is also a metric on *x*.

Proof:

Since d is a metric on X, f clearly satisfies

1. $f(x_1, x_2) \geq 0$ and $f(x_1, x_1) = 0 = \dfrac{d(x_1, x_2)}{1 + d(x_1, x_2)} = \dfrac{0}{1+0}$

$$f(x_1, x_2) = \frac{d(x_1, x_2)}{1 + d(x_1, x_2)} > 0 \text{ since } d(x_1, x_2) > 0$$

2. $f(x_1, x_2) = \dfrac{d(x_1, x_2)}{1 + d(x_1, x_2)} = \dfrac{d(x_1, x_2)}{1 + d(x_1, x_2)} f(x_1, x_2)$

3. If $x_1 \neq x_2$, then $f(x_1, x_2) > 0$

4. To show triangle inequality, i.e., $f(x_1, x_3) \leq f(x_1, x_2) + f(x_1, x_3)$
   Let $x_1, x_2, x_3 \in X$, then

$$f(x_1, x_2) = \frac{d(x_1, x_3)}{1 + d(x_1, x_2) + d(x_2, x_3)} + \frac{d(x_2, x_3)}{1 + d(x_1, x_2) + d(x_2, x_3)}$$

that is,

$$f(x_1, x_3) \leq f(x_1, x_2) + f(x_2, x_3)$$

Thus, f is a metric on X.
Let us define the metric

$$d_2(x_1, x_2) = f|x_1, x_2| \text{ then } d_2 \text{ is a matrix on } \Re^p.$$

Summary of the Algorithm

The algorithm solves the problem

$$\min f(\underline{x})\underline{x} \in \Re^n.$$

Step 0: Compute $\underline{\nabla} f(\underline{x}^\circ), H_c = \nabla^2 f(x^\circ), \underline{S}_0 = -\nabla^2 f(\underline{x}^\circ)^{-1} \underline{\nabla} f(\underline{x}^\circ) = \underline{S}_c^N$

Choose $L_c = \dfrac{\|S_c\|}{1 + \|S_c\|}$

Step 1: Calculate $\left\|\underline{S}_c^N\right\|, \underline{S}^{c.p.} + E' * \underline{\nabla} f(x_c), E' = \dfrac{\|\nabla f(x_0)\|^2}{\underline{\nabla} f(\underline{x}_c)^T H_c \underline{\nabla} f(\underline{x}_c)}$,

$\left\|\underline{S}^{c.p.}\right\|, \underline{S}_c = \underline{S}^{c.p.} + E' * \left(\underline{S}_c^N - \underline{S}^{c.p.}\right)$, set $M = |K|, K$ is the least eigenvalue of $\nabla^2 f\left(x^k\right)$

If $\left\|S_c^N \leq L_c\right\|$ and $\left\|\underline{S}^{c.p.}\right\| \leq L_c$, then go to step 2, else,

Compute $\underline{S}_c = -(H + MI)^{-1} \underline{\nabla} f\left(\underline{x}_c\right)$ and go to step 2.

Step 2: Set

$$\underline{x}_{k1} = \underline{x}_k + \frac{L_c^c \underline{S}_c}{\left\|\underline{\nabla} f\left(\underline{x}_k\right)\right\|}$$

Step 3: Terminate if convergence is achieved.

Minimize of Rosenbrocks Banana Shaped Valley Function

ITRN	GRADIENT NORM	FUNCTION VALUE
0	232.867688	24.2000
20	227.345905	23.352732
40	221.548129	22.479063
60	216.302612	21.702862
80	210.781113	20.900625
100	205.259720	20.113773
120	199.738441	19.342500
140	194.217282	18.587004
160	188.696252	17.847491
180	174.894294	16.070021
200	169.373787	15.388135
220	163.853454	14.723219

3. Computational Results

The algorithm is implemented in Turbo Pascal on Rasenbrock's banana-shaped valley function run on IBM compatible machine in which B is positive definite. In this situation a trust region method terminates. The test problem is

$$f\left(x_1, x_2\right) = \left(1 - x_1\right)^2 + 100\left(x_1^2 - x_2\right)^2, \underline{x}_0 = (-1.2, 1)^T$$

The results of the computation are displayed on the table provided.

4. Conclusions and Perspectives

In this paper, we have proposed a new model of trust region algorithm which is characterized by monotonic convergence. The number of iterations to obtain convergence for the test function is rather high but this can be improved by modifying the steps taken to reach the minimum from the Cauchy point.

References

1. R.G. Bartle (1964), *The Elements of Real Analysis,* J. Wiley & Sons Inc., NY. Pp. 67–69.

2. J.V. Burke (1990), *A Robust Trust Region Method for Constrained Nonlinear Programming Problems,*ANLMCS, p. 131–0190, Preprint.

3. M.R. Celis, J.E. Dennis and R.A. Tapia (1983), *A Trust Region Strategy for Nonlinear Equality Constrained Optimization,* SIAM Proceedings, Book series, PR. 20.

4. M.R. Conn, N. Gould, A. Stratenaer and Phil Toint (1993), *Global Convergence of a Class of Trust Region Algorithms for Optimization using Inexact Projection on Convex Constraints,* SIAM J. Optimization, Vol. 1., pp. 164–221.

5. J.E. Dennis, Jnr., R.B. Schnabel (1983), *Numerical Methods for Unconstrained Optimization and Nonlinear Equations,* Prentice-Hall Inc., Englewood Cliffs, New Jersey, pp. 129–147.

6. L.C.W. Dixon (1987), *Use of Automatic Differentiation for Calculating Hessians and Newton Steps,* Preprint, ANL.

7. A. Griewank (1990), *Direct Calculation of Newton Steps without Accumulating Jacobians,* Preprint MCS: p. 132–0290, ANL.

8. J.J. More (1988), *Trust Regions and Projected Gradient,* ANL/MCS – TM – 107.

9. J.J. More and D.C. Sorensen (1983), *Computing A Trust Region Step,* SIAM J. Sci. Stat. Computing, 4(1), pp. 553-572.

10. J.J. More and D.C. Sorensen (1983), *Recent Developments in Algorithms and Software for Trust Region Methods in Mathematical Programnming: The State of the Art,* A. Bachen, M. Grotschel, and Korte, Eds, Springer-Verlag, pp. 258–287.

11. M.D.J. Powell and Y. Yuan (1991), *A Trust Region Algorithm for Equality Constrained Optimization, Mathematical Programming,* North Holland, 49, pp. 189–211.

12. D.C. Sorensen (1982), *Newton's Method with a Model Trust Region Modification,* SIAM J. Numer. Anal. Pp. 409–426.