ELSEVIER Journal of Computational and Applied Mathematics 91 (1998) 249–259

JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

A model-trust region algorithm utilizing a quadratic interpolant

Brian J. McCartin *

Applied Mathematics, Kettering University, 1700 West Third Avenue, Flint, MI 48504-4898, USA

Received 12 March 1997; received in revised form 12 February 1998

Abstract

A new model-trust region algorithm for problems in unconstrained optimization and nonlinear equations utilizing a quadratic interpolant for step selection is presented and analyzed. This is offered as an alternative to the piecewise-linear interpolant employed in the widely used "double dogleg" step selection strategy. After the new step selection algorithm has been presented, we offer a summary, with proofs, of its desirable mathematical properties. Numerical results illustrating the efficacy of this new approach are presented. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Model-trust region algorithms; Nonlinear equations; Unconstrained optimization

1. Introduction

We are concerned in this paper with a new procedure for solving unconstrained optimization problems whether they arise directly or indirectly through minimization of the norm of the residual of a system of nonlinear equations. Of particular interest are problems of sufficient complexity to preclude the determination of an initial guess for the iteration that is close to the desired solution. This situation is quite common in engineering design, one such problem arising in the numerical simulation of semiconductor devices [4].

Newton's method [6] is a powerful tool for solving such problems since it exhibits local q-quadratic convergence, i.e. it is q-quadratically convergent *provided* that we start the iteration sufficiently close to the solution. In order to construct a globally convergent variant, Newton's method is hybridized with the globally, yet typically slowly, convergent Cauchy's method (steepest descent). The resulting so-called model-trust region algorithms [2] retain the best features of both methods: strong global convergence coupled with rapid local convergence (i.e. they are globally q-quadratically convergent).

* Tel.: (810) 762-7802; fax: (810) 762-9796; e-mail: bmccarti@kettering.edu

PII S0377-0427(98)00052-1

The various model-trust region algorithms differ from one another in precisely how this hybridization is achieved. This matter will be further elaborated upon in the next section.

The present paper offers an improvement to the standard "double dogleg" version of this algorithm via replacement of its piecewise-linear approximant by a quadratic interpolant. In the ensuing pages, we first introduce the model-trust region strategy. We then present our new step selection algorithm, describe its mathematical properties, and compare it to the "double dogleg" strategy. Finally, we report on the strenuous excercise of the new algorithm on a suite of standard test problems [5] as well as on the numerical simulation of a $p-n$ diode [4]. Additional details of the test problems appear in [3]. We focus herein on those aspects of our algorithm which are innovative. For details on model-trust region algorithms in general (e.g. scaling, initialization, termination, Hessian estimation, linear solver), the reader is referred to the superlative text [1].

2. Model-trust region algorithms

Consider the unconstrained minimization problem

$$\min_{x\in\mathbb{R}^n} f(x), \quad f:\mathbb{R}^n\to\mathbb{R}, \tag{1}$$

with $f(x)$ assumed twice continuously differentiable. We seek a local minimizer, x_*, for this problem. For unconstrained maximization, simply replace $f(x)$ by $-f(x)$. In what follows, all unspecified norms are Euclidean 2-norms, $\|x\| := \sqrt{x^{\mathrm{T}}x}$, and $\|x\|_A := \sqrt{x^{\mathrm{T}}Ax}$ with A symmetric and positive definite (SPD).

In Newton's method [6], the step at each stage is selected to minimize the local quadratic model:

$$m_c(x_c+s) := f_c + \nabla f_c^{\mathrm{T}} s + \tfrac{1}{2} s^{\mathrm{T}} \nabla^2 f_c s, \tag{2}$$

where a subscript denotes the point of evaluation, in this instance the current point, x_c. Applying the first derivative test for a local extremum, we arrive at the equations for the Newton step:

$$\nabla^2 f_c s_N = -\nabla f_c. \tag{3}$$

If $\nabla^2 f_c$ is not positive definite then we modify the local quadratic model as follows:

$$\hat{m}_c(x_c+s) := f_c + \nabla f_c^{\mathrm{T}} s + \tfrac{1}{2} s^{\mathrm{T}} H_c s, \tag{4}$$

where $\mu_c \geqslant 0$ is selected so that the modified model Hessian, $H_c := \nabla^2 f_c + \mu_c I$ is "safely positive definite". Assuming that $\nabla^2 f_*$ is nonsingular, $\nabla^2 f_c$ will be SPD when we are close to the local minimizer, so that $\mu_c = 0$ eventually. Further justification for this modification as well as details of its implementation are available in [1].

Similarly, we may also address the root-finding problem:

$$F(x_*) = 0, \; x_* \in \mathbb{R}^n, \quad F:\mathbb{R}^n\to\mathbb{R}^n, \tag{5}$$

via its reformulation as the unconstrained minimization problem:

$$\min_{x\in\mathbb{R}^n} f(x), \quad f(x) := \tfrac{1}{2}F(x)^{\mathrm{T}}F(x) = \tfrac{1}{2}\|F(x)\|^2. \tag{6}$$

Taylor series expansion about the current iterate, x_c, yields the following local approximation to $f(x)$:

$$m_c(x_c+s) := \tfrac{1}{2}F(x_c)^{\mathrm{T}}F(x_c) + [J(x_c)^{\mathrm{T}}F(x_c)]^{\mathrm{T}}s + \tfrac{1}{2}s^{\mathrm{T}}[J(x_c)^{\mathrm{T}}J(x_c) + \sum_{i=1}^{n} F_i(x_c)\nabla^2 F_i(x_c)]s. \quad (7)$$

Rather than having to compute the Hessian matrices, $\nabla^2 F_i$ $(i = 1,\dots,n)$, we observe that their coefficients vanish at x_*. Thus, we omit the above summation and introduce the modified local quadratic model:

$$\begin{aligned}\hat{m}_c(x_c+s) &:= \tfrac{1}{2}F(x_c)^{\mathrm{T}}F(x_c) + [J(x_c)^{\mathrm{T}}F(x_c)]^{\mathrm{T}}s + \tfrac{1}{2}s^{\mathrm{T}}[J(x_c)^{\mathrm{T}}J(x_c)]s \\ &= \tfrac{1}{2}M_c(x_c+s)^{\mathrm{T}}M_c(x_c+s);\ M_c(x_c+s) := F(x_c) + J(x_c)s. \end{aligned} \quad (8)$$

Since the gradients of $\hat{m}_c$, m_c, and f are identical at x_c, they share descent directions from that point. In particular, $p_{\mathrm{SD}} := -J^{\mathrm{T}}F$ (steepest descent direction) and $p_{\mathrm{N}} := -J^{-1}F$ (Newton direction) are shared by all three functions. Eq. (8) may be recast in the form of Eq. (4) by the indentifications $\nabla f_c = J_c^{\mathrm{T}}F_c$ and $H_c = J_c^{\mathrm{T}}J_c$, which is positive definite provided that J_c is nonsingular.

A variant of Newton's method for either unconstrained minimization or nonlinear equations with attractive global convergence properties may be derived as follows. A pure Newton iteration would correspond to selecting s to minimize $\hat{m}_c$ regardless of the quality of this local quadratic model. The model-trust region algorithms improve upon this by utilizing an estimate, δ_c, of the radius of the region about x_c in which $\hat{m}_c$ adequately represents f. Thus, the next iterate is determined by the step, s, which solves the locally constrained minimization problem:

$$\min_s[\hat{m}_c(x_c+s) = f_c + \nabla f_c^{\mathrm{T}}s + \tfrac{1}{2}s^{\mathrm{T}}H_c s] \quad \text{s.t. } \|s\| \leqslant \delta_c. \quad (9)$$

This stands in stark contrast to restricted-step Newton methods which retain the Newton step direction but simply reduce the step length when far from x_*.

Defining

$$s(\lambda) := -\left(H_c + \frac{\lambda}{1-\lambda}I\right)^{-1}\nabla f_c, \quad (10)$$

we may now state the fundamental theorem of model-trust region algorithms (see Fig. 1):

Theorem 1 (Fundamental Theorem of model-trust region algorithms). *The solution to the locally constrained minimization problem, Eq.* (9), *is given by*

$$s_+ = \begin{cases} s(0) = s_{\mathrm{N}}, & \text{if } \|s(0)\| \leqslant \delta_c \\ s(\lambda_+) \ni \|s(\lambda_+)\| = \delta_c, & \text{otherwise.} \end{cases} \quad (11)$$

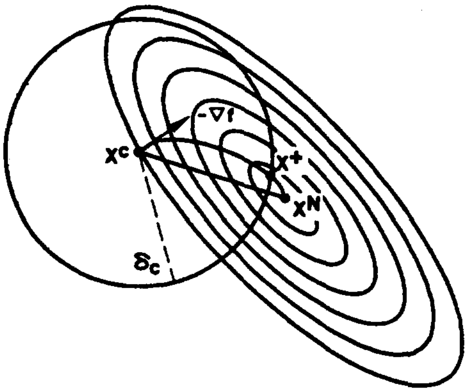

Fig. 1. Model-trust region.

Proof. See [1, pp. 131–132]. □

Note that the curve, $s(\lambda)$, has the following properties [1]: $s(0) = s_N$, $s(1) = 0$, $s'(1) = \nabla f_c$. Thus, the associated step-directions range between the Newton (δ_c large) and the Cauchy (δ_c small) directions. Also, $s(\lambda)$ is a descent direction from x_c for all $0 \leqslant \lambda < 1$ and distance from x_c increases as λ decreases. In practice, the difficulty in applying Theorem 1 lies in the complexity of finding λ_+. If we attempt to find it exactly, as is done in the optimal "hook step" strategy [1], then systems of equations of the form

$$\left(H_c + \frac{\lambda_k}{1-\lambda_k} I\right) s_k = -\nabla f_c \tag{12}$$

need to be solved at each step.

A popular alternative to incurring this added computational burden is the use of the "double dogleg" strategy [1] whereby the curve $s(\lambda)$, $0 \leqslant \lambda \leqslant 1$, is approximated by a piecewise-linear curve $x_c \to x_{SD} \to \bar{x} \to x_N$ as shown in Fig. 2. Here, x_{SD} is the Cauchy point, x_N is the Newton point, $\bar{x} := x_c + \alpha s_N$ ($\gamma \leqslant \alpha \leqslant 1$), and γ is the ratio of the Cauchy step length to the length of the projection of the Newton step onto the Cauchy direction which we will show in the next section to be bounded by 1. The new point, x_+, is chosen so that $\|x_+ - x_c\| = \delta_c$, unless $\|H_c^{-1}\nabla f_c\| \leqslant \delta_c$ in which case $x_+ = x_N$.

It is shown in [1] that with such a choice for α (they suggest setting $\alpha = 0.8\gamma + 0.2$), the "double dogleg" curve possesses two very important properties. Firstly, as a point proceeds along the "double dogleg" curve from x_c to x_N, distance from x_c increases monotonically, so that this process is well defined. Secondly, the value of the modified local quadratic model, $\hat{m}_c(x_c + s)$ decreases monotonically as a point traverses this curve in the same direction, so that this process is reasonable. In the next section, we present a new quadratic interpolant to $s(\lambda)$ with these same desirable properties.

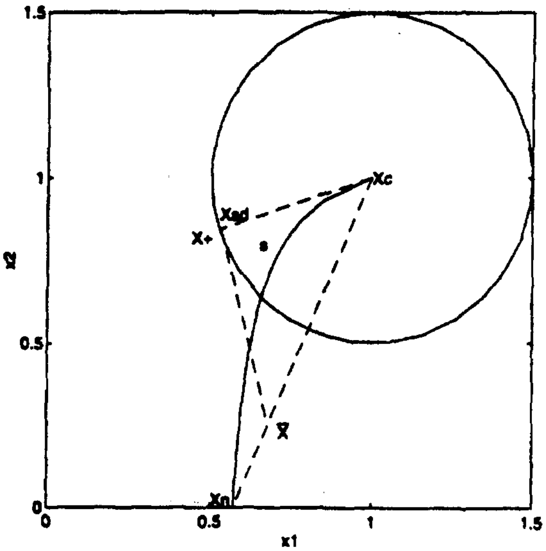

Fig. 2. Double dogleg strategy.

3. The quadratic interpolant

In the new step selection algorithm (see Fig. 3), we approximate s_+ by replacing $s(\lambda)$ by the quadratic interpolant:

$$\sigma(\eta) := (\eta - 1)[(\eta - 1)s_N + \eta\beta\nabla f_c], \quad 0 \leqslant \eta \leqslant 1, \tag{13}$$

when s_N lies outside the trust region. For the moment, we leave $\beta > 0$ unspecified. We next summarize some general properties of $\sigma(\eta)$.

Theorem 2. *For any $\beta > 0$, $x_c + \sigma(\eta)$ passes through x_c and x_N and is tangent to $x_c + s(\lambda)$ at x_c. Moreover, $\sigma(\eta)$ is a descent direction from x_c for any $0 \leqslant \eta < 1$.*

Proof. $\sigma(0) = s_N \Rightarrow$ interpolation at x_N. $\sigma(1) = 0 \Rightarrow$ interpolation at x_c. $\sigma'(\eta) = 2(\eta - 1)s_N + (2\eta - 1)\beta\nabla f_c \Rightarrow \sigma'(1) = \beta\nabla f_c \Rightarrow$ tangency at x_c. $\nabla f_c^T\sigma(\eta) = -(\eta-1)^2\|\nabla f_c\|^2_{H^{-1}} + \eta(\eta-1)\|\nabla f_c\|^2 < 0 \Rightarrow$ $\sigma(\eta)$ is a descent direction from x_c for $0 \leqslant \eta < 1$. □

Ideally, the free parameter, β, would be chosen to achieve tangency at x_N also. However, this is equivalent to selecting β so that $-2s_N - \beta\nabla f_c \parallel -H_c^{-1}s_N$ and would thus require that an additional system of equations be solved. We wish to avoid this added cost. In the next section, we show that the choice

$$\beta := \sqrt{\frac{-2s_N^T\nabla f_c}{\nabla f_c^T H_c \nabla f_c}} \tag{14}$$

endows the curve $x_c + \sigma(\eta)$ with certain highly desirable properties, thus obviating the need for multiple linear solves at each step of the model-trust region iteration.

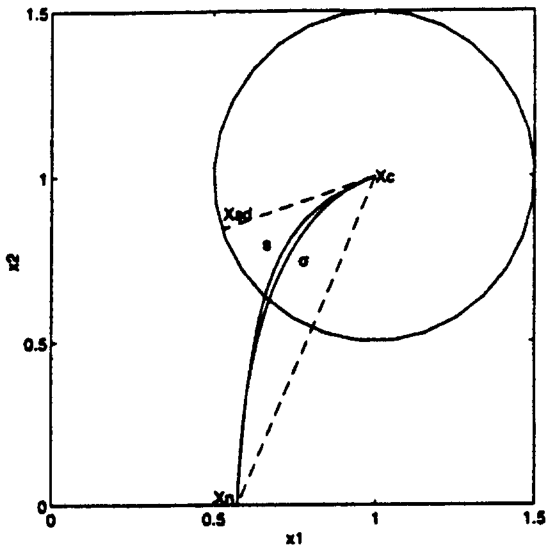

Fig. 3. New step selection algorithm.

4. Mathematical properties

In order to fully explore the consequences of the above choice for β, Eq. (14), we need to lay some preliminary groundwork. We begin with:

Lemma 1. *Let $A, B, C \in \mathbb{R}^{n\times n}$ with A and C positive definite and $x, y \in \mathbb{R}^n$; then the following are equivalent:*

$$|x^{\mathrm{T}}By|^2 \leqslant (x^{\mathrm{T}}Ax)(y^{\mathrm{T}}Cy), \tag{15}$$

$$2|x^{\mathrm{T}}By| \leqslant x^{\mathrm{T}}Ax + y^{\mathrm{T}}Cy, \tag{16}$$

$$\rho(B^{\mathrm{T}}A^{-1}BC^{-1}) \leqslant 1. \tag{17}$$

Proof. This is Theorem 7.7.7 [2, p. 473] specialized to real, square matrices. □

This lemma has the following important consequence.

Theorem 3. *Assume that H_c is SPD; then*

$$|\nabla f_c||^2 \leqslant ||\nabla f_c||_{H_c} \cdot ||\nabla f_c||_{H_c^{-1}}, \tag{18}$$

$$||\nabla f_c||^2 \leqslant \tfrac{1}{2} \cdot [||\nabla f_c||^2_{H_c} + ||\nabla f_c||^2_{H_c^{-1}}], \tag{19}$$

$$||\nabla f_c|| \leqslant \tfrac{1}{2}[||\nabla f_c||_{H_c} + ||\nabla f_c||_{H_c^{-1}}]. \tag{20}$$

Proof. Let $A = H_c$, $B = I$, $C = H_c^{-1}$, $x = y = \nabla f_c$ in Lemma 1; then $\rho(B^{\mathrm{T}}A^{-1}BC^{-1}) = \rho(I) = 1$. Thus,

(i) $\nabla f_c^{\mathrm{T}} \nabla f_c \leqslant (\nabla f_c^{\mathrm{T}} H_c \nabla f_c)^{1/2} \cdot (\nabla f_c^{\mathrm{T}} H_c^{-1} \nabla f_c)^{1/2}$.

(ii) $2\nabla f_c^T \nabla f_c \leqslant \nabla f_c^T H_c \nabla f_c + \nabla f_c^T H_c^{-1} \nabla f_c$.

(iii) Taking the arithmetic mean of (i) and (ii) yields $||\nabla f_c||^2 \leqslant \frac{1}{4}[||\nabla f_c||_{H_c}^2 + 2||\nabla f_c||_{H_c}||\nabla f_c||_{H_c^{-1}} + ||\nabla f_c||_{H_c^{-1}}^2] = \frac{1}{4}[||\nabla f_c||_{H_c} + ||\nabla f_c||_{H_c^{-1}}]^2$. Taking the square root of both sides of this results in (iii). □

This brings us to the fundamental result:

Theorem 4 (Cauchy-Newton inequality). *The size of the Cauchy step is always bounded above by the length of the projection of the Newton step onto the Cauchy step:*

$$||s_{SD}|| \leqslant \left|\left| s_N^T \frac{s_{SD}}{||s_{SD}||} \right|\right|. \tag{21}$$

Proof.

$$s_{SD} := -\frac{||\nabla f_c||^2}{\nabla f_c^T H_c \nabla f_c} \cdot \nabla f_c \Rightarrow ||s_{SD}|| = \frac{||\nabla f_c||^3}{\nabla f_c^T H_c \nabla f_c};$$

$$s_N := -H_c^{-1} \nabla f_c \Rightarrow ||projection\ of\ s_N\ onto\ \nabla f_c|| = \frac{\nabla f_c^T H_c^{-1} \nabla f_c}{||\nabla f_c||}.$$

By Eq. (18), $||\nabla f_c||^4 \leqslant ||\nabla f_c||_H^2 \cdot ||\nabla f_c||_{H^{-1}}^2 \Rightarrow \frac{||\nabla f_c||^3}{\nabla f_c^T H_c \nabla f_c} \leqslant \frac{\nabla f_c^T H_c^{-1} \nabla f_c}{||\nabla f_c||}$, which we now recognize as the desired inequality, Eq. (21). □

In what follows, we require an additional result:

Lemma 2.

$$\beta \leqslant -\sqrt{2}\frac{s_N^T \nabla f_c}{||\nabla f_c||^2} \leqslant -\sqrt{2} \cdot \frac{||s_N||^2}{s_N^T \nabla f_c}. \tag{22}$$

Proof. By Eq. (18), $||\nabla f_c||^4 \leqslant (\nabla f_c^T H_c^{-1} \nabla f_c) \cdot (\nabla f_c^T H_c \nabla f_c) = (-s_N^T \nabla f_c) \cdot (\nabla f_c^T H_c \nabla f_c)$. Thus,

$$\beta^2 = \frac{-2 s_N^T \nabla f_c}{\nabla f_c^T H_c \nabla f_c} \leqslant 2 \cdot \frac{(s_N^T \nabla f_c)^2}{||\nabla f_c||^4}.$$

The first half of the inequality now follows by taking square roots. The second half of the inequality is a direct consequence of the Cauchy–Schwarz inequality. □

We are finally in a position to fully describe the mathematical properties of the curve $x_c + \sigma(\eta)$ with our choice of β. We summarize these as:

Theorem 5. *With $\sigma(\eta)$ given by Eq.* (13) *and β defined by Eq.* (14):

(i) *Distance from x_c increases monotonically as we proceed along the curve, $x_c + \sigma(\eta)$, from x_c $(\eta = 1)$ to x_N $(\eta = 0)$.*

(ii) *The value of the local quadratic model, $\hat{m}_c(x_c + \sigma(\eta))$, decreases monotonically as we traverse the curve, $x_c + \sigma(\eta)$, from x_c $(\eta = 1)$ to x_N $(\eta = 0)$.*

Proof.

(i) A direction, p, from $x_c + \sigma(\eta)$ increases distance from x_c iff $p^T\sigma(\eta) > 0$. The tangent vector to the path $x_c + \sigma(\eta)$ from x_c to x_N is parallel to $p := -\sigma'(\eta)$, so that we need to establish that $\sigma'(\eta)^T\sigma(\eta) < 0$ for $0 < \eta < 1$. But, $\sigma'(\eta) = 2(\eta-1)s_N + (2\eta-1)\beta\nabla f_c$, so that $\sigma'(\eta)^T\sigma(\eta) = (\eta-1)\cdot u(\eta)$ where $u(\eta) := \eta^2\cdot[2||\nabla f_c||^2\beta^2 + 4s_N^T\nabla f_c\beta + 2||s_N||^2] - \eta\cdot[||\nabla f_c||^2\beta^2 + 5s_N^T\nabla f_c\beta + 4||s_N||^2] + [s_N^T\nabla f_c\beta + 2||s_N||^2]$. We need to establish that $u(\eta) > 0$ for $0 < \eta < 1$. Note that $u(0) = 2||s_N||^2 + s_N^T\nabla f_c\beta \geqslant 0$ by Lemma 2, and that $u(1) = \beta^2||\nabla f_c||^2 > 0$. Thus, either both roots of this quadratic lie in $(0,1)$ or neither one does. Furthermore, the product of the roots [7] of this quadratic exceeds 1, again by Lemma 2, so that both roots cannot lie in $(0,1)$ and hence neither root can either. Hence, $u(\eta) > 0$ for $0 < \eta < 1$.

(ii) The directional derivative of $\hat{m}_c(x_c + \sigma(\eta))$ is $[\nabla f_c + H_c\sigma(\eta)]^T \cdot \sigma'(\eta)/||\sigma'(\eta)||$ and must be shown to be positive for $0 < \eta < 1$. We will accomplish this by establishing that $d(\eta) := [\nabla f_c^T + (\eta-1)^2 s_N^T H_c + \eta(\eta-1)\beta\nabla f_c^T H_c]\cdot[2(\eta-1)s_N + (2\eta-1)\beta\nabla f_c] > 0$ for $0 < \eta < 1$, since it has the same sign as the directional derivative. Note that $d(0) = 0$, $d(1) = \beta||\nabla f_c||^2 > 0$, $d'(0) = -4\beta||\nabla f_c||^2 - 6s_N^T\nabla f_c > 0$ — once again, by Lemma 2, and $d'(1) = 2\beta||\nabla f_c||^2 > 0$. Thus, it will suffice to exclude either of the other two roots of the cubic, $d(\eta)$, from the interval $(0,1)$. This is achieved by rewriting $d(\eta) = \eta\cdot Q(\eta)$ with the quadratic $Q(\eta) := A\eta^2 + B\eta + C$ and $A := 2\beta^2\nabla f_c^T H_c\nabla f_c - 4\beta||\nabla f_c||^2 - 2s_N^T\nabla f_c$, $B := -3\beta^2\nabla f_c^T H_c\nabla f_c + 9\beta||\nabla f_c||^2 + 6s_N^T\nabla f_c$, $C := \beta^2\nabla f_c^T H_c\nabla f_c - 4\beta||\nabla f_c||^2 - 4s_N^T\nabla f_c$. The product of the roots [7] of $Q(\eta)$ is $C/A \geqslant 1$ provided that $\beta^2 \leqslant -2\cdot s_N^T\nabla f_c/\nabla f_c^T H_c\nabla f_c$ which is assured by Eq. (14). Thus, for our choice of β, the two roots of $Q(\eta)$ cannot both reside on $(0,1)$, so that neither can separately. Hence, we have that $d(\eta)$ is nonnegative on $[0,1]$ and is positive on $(0,1)$. □

Note that (i) precludes the quadratic trajectory, $x_c + \sigma(\eta)$; $\eta \in [0,1]$, from "overshooting" x_N and "turning back on itself", so that there will be a unique point of intersection of the path with the trust region boundary. Also, (ii) asserts that it is pointless to stop along the curve, $x_c + \sigma(\eta)$, before reaching the trust region boundary since the local quadratic model, $\hat{m}_c$, which we "trust" is still decreasing.

Thus, it is well defined and reasonble to specify η_+ as the smallest positive root of the quartic

$$q(\eta) := ||\sigma(\eta)||^2 - \delta_c^2, \tag{23}$$

and to select as our next iterate (see Fig. 3)

$$x_+ := x_c + \sigma(\eta_+). \tag{24}$$

Since the quartic, Eq. (23), has leading coefficient and constant term of opposite sign, it has at least two real roots, one positive and one negative [7, p. 105]. In fact, by Theorem 5 (i), $q(\eta)$ has a unique root in $[0,1]$ provided that $\delta_c < ||s_N||$.

Letting $\Delta f := f(x_+) - f(x_c)$ and $\Delta f_p := \hat{m}_c(x_+) - f(x_c)$, if $\Delta f > 10^{-4}\cdot\nabla f_c^T(x_+ - x_c)$ then we reject x_+ and reduce δ_c by a quadratic backtrack strategy [1]. Otherwise, we check to see if the decrease in f was dramatic enough to justify an enlarged trust region with the present model. If not, then the model is updated with a δ_c determined by a comparison of the actual decrease in f, Δf, with the predicted decrease, Δf_p. This portion of the algorithm is not new and further details on the model-trust region update are available [1]. The complete model-trust region algorithm is depicted in Fig. 4.

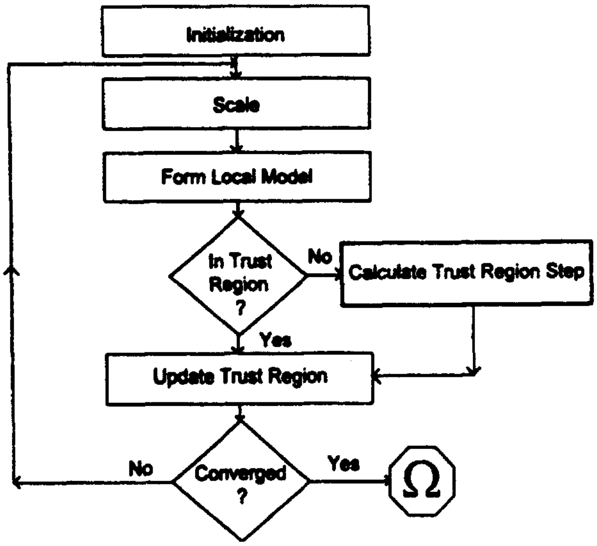

Fig. 4. Model-trust region algorithm.

5. The Dennis–Schnabel example

Now, consider the example [1]

$$f(x_1,x_2) := x_1^4 + x_1^2 + x_2^2, \tag{25}$$

with

$$x_c = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \delta_c = \tfrac{1}{2}. \tag{26}$$

This produces

$$\nabla f_c = \begin{bmatrix} 6 \\ 2 \end{bmatrix}, \qquad H_c = \begin{bmatrix} 14 & 0 \\ 0 & 2 \end{bmatrix} \tag{27}$$

and

$$s_N = \begin{bmatrix} -3/7 \\ -1 \end{bmatrix}, \qquad s_{SD} = \begin{bmatrix} -0.469 \\ -0.156 \end{bmatrix}. \tag{28}$$

The new algorithm yields (see Fig. 3) $\beta = 0.1336$, $\eta_+ = 0.444$, $\sigma_+ = [-0.330, -0.375]^T$, while $s_+ = [-0.343, -0.365]^T$. Comparison to the standard "double dogleg" step of $[-0.340, -0.669]^T$ (see Fig. 2) indicates that the new algorithm provides a dramatic increase in accuracy at essentially no additional cost.

Table 1
Summary of numerical results

Problem	Starting point	Scaled $f(x)$	No. of steps
Rosenbrock Banana function ($n = 2$)	$[-1.2, 1]^T$	$9.86 \cdot 10^{-32}$	2
Freudenstein–Roth function ($n = 2$)	$[6, 5]^T$	$7.32 \cdot 10^{-29}$	5
———	$[0.5, -2]^T$	$6.91 \cdot 10^{-29}$	19 (11+8)
Powell badly scaled function ($n = 2$)	$[0, 1]^T$	$3.83 \cdot 10^{-27}$	12
Box 3D function ($n = 3$)	$[0, 10, 20]^T$	$4.48 \cdot 10^{-32}$	5
Helical valley function ($n = 3$)	$[-1, 0, 0]^T$	$2.89 \cdot 10^{-28}$	13
Powell singular function ($n = 4$)	$[3, -1, 0, 1]^T$	$2.50 \cdot 10^{-13}$	20
P-N diode ($n = 195$)	Space-charge neutral	$3.52 \cdot 10^{-25}$	7

6. Numerical results

With the dual aims of establishing the validity and the utility of our algorithm, Table 1 presents the results of a sequence of numerical examples. They run the gamut from low-dimensional systems with known solutions [5] to a cutting-edge problem in semiconductor device simulation (*p–n* diode) [4]. All simulations were performed in double precision using the Fortran77-3.0.1 compiler under the SunOS-5.5.1 operating system on a Sun Microsystems E3000 UltraEnterprise workstation. Complete details are available elsewhere [3].

7. Conclusion

In the preceding sections, we have presented a flexible and powerful new model-trust region algorithm for unconstrained optimization and systems of nonlinear equations based upon quadratic interpolation. The mathematical properties of this algorithm have been described and full proofs have been provided. Comparison to the popular "double dogleg" strategy has been made. The new algorithm has been validated on a suite of widely used test problems from the literature [5] and its practical utility has been demonstrated on a challenging problem from semiconductor device simulation [4]. The full details pertaining to these test cases have appeared elsewhere [3]. In closing, we make the observation that methods based upon higher order interpolation might be contemplated. However, the concomitant increase in algorithmic complexity would demand numerical and/or analytical justification.

8. Acknowledgements

The author would like to thank Ms. Barbara Rowe for her assistance in the production of this paper.

References

[1] J.E. Dennis, R.B. Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Prentice-Hall, Englewood Cliffs, NJ, 1983.

[2] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge, 1990.

[3] B.J. McCartin, A new model-trust region algorithm for systems of nonlinear equations, Proc. 13th Ann. Conf. on Applied Mathematics (CAM97), University of Central Oklahoma, 1997, pp. 172–186.

[4] B.J. McCartin, R.H. Hobbs, R.E. LaBarre, P.E. Kirschner, Solution of the discrete semiconductor device equations, in: J.J.H. Miller (Ed.), NASECODE IV: Proc. 4th Internat. Conf. on the Numerical Analysis of Semiconductor Devices and Integrated Circuits, Boole Press, 1985, pp. 411–416.

[5] J.J. Moré, B.S. Garbow, K.E. Hillstrom, Testing unconstrained optimization software, ACM Trans. Math. Software 7 (1) (1981) 17–41.

[6] J.J. Moré, D.C. Sorenson, Newton's Method, in: G.H. Golub (Ed.), Studies in Numerical Analysis, MAA, 1984, pp. 29–82.

[7] J.V. Uspensky, Theory of Equations, McGraw-Hill, New York, 1948.