J Optim Theory Appl
DOI 10.1007/s10957-016-1038-3

Local Stability and Local Convergence of the Basic Trust-Region Method

Bui Ngoc Muoi[1] · **Nguyen Dong Yen**[2]

Received: 24 February 2016 / Accepted: 14 November 2016

Abstract It is proved that the iterative sequence constructed by the Basic Trust-Region Algorithm (see Conn et al. in Trust-region methods, MPS-SIAM series on optimization, Philadelphia, 2000), which uses the Cauchy point method, is locally stable and linearly convergent in a neighborhood of a nonsingular local minimizer.

Keywords Unconstrained smooth optimization problem · Basic trust-region method · Iterative sequence · Stability · Linear convergence

Mathematics Subject Classification 90C30 · 90C26 · 49M37

1 Introduction

The Trust-Region Method (see, e.g., [1]) has been applied for solving unconstrained optimization problems, linearly constrained optimization problems, and optimization problems with general constraints.

Together with such traditional solution schemes as the gradient, steepest descent gradient, gradient projection, conjugate gradient, feasible direction, Newton, Lagrange

Communicated by Suliman Saleh Al-Homidan.

Dedicated to Professor B.T. Polyak on the occasion of his 80th birthday.

✉ Nguyen Dong Yen
ndyen@math.ac.vn

Bui Ngoc Muoi
muoisp2@gmail.com

1 Hanoi Pedagogical University No. 2, Xuan Hoa, Phuc Yen, Vinh Phuc, Vietnam

2 Institute of Mathematics, Vietnam Academy of Science and Technology, Hanoi, Vietnam

Published online: 28 November 2016

multiplier, penalty function, barrier function, subgradient, bundle, and interior point methods (see, e.g., [2–5]), the Trust-Region Method is considered one of the most effective solution methods for optimization problems.

In this paper, we will make a focus on the local stability and local convergence of iterative sequences generated by the Trust-Region Method. Specifically, we will look for a result similar to that of Le Thi et al. [6, Theorem 3]. Namely, based on some arguments from Ruszczynski's book [4, pp. 226–227], which were used for establishing the convergence of the steepest descent method, we will be able to prove the *local stability* and *local convergence* of the iterative sequences created by the Basic Trust-Region Algorithm (BTR Algorithm) for the unconstrained optimization problem with a twice-continuously differentiable objective function. Moreover, the *linear convergence rate* of the iterative sequences will be obtained.

Our research is close to the spirit not only of [6, Theorem 3], but also of [7, Theorem 4], where convergence and stability of line search methods for unconstrained optimization are investigated.

The organization of the present paper is as follows. Section 2 recalls the BTR Algorithm and gives two illustrative numerical examples. Section 3 establishes the local stability, the local convergence, and the linear convergence rate of the BTR Algorithm. Some concluding remarks are given in Sect. 4.

2 The Basic Trust-Region Algorithm

Given a twice-continuously differentiable real-valued function $f : \mathrm{IR}^n \to \mathrm{IR}$, we consider the unconstrained optimization problem

$$\min\{f(x) \ : \ x \in \mathrm{IR}^n\}. \tag{1}$$

With an arbitrary initial point $x_0 \in \mathrm{IR}^n$, the Basic Trust-Region Algorithm [1, p. 116] (BTR Algorithm for brevity) creates an iterative sequence $\{x_k\}$ such that the movement from x_k to x_{k+1} reduces the *approximate objective function* at step k, which is denoted by $m_k(x)$, of $f(x)$. One of the most popular approximations consists in replacing $f(x)$ by the linear-quadratic part of Taylor's expansion of $f(x)$ at x_k. At the k-th iteration, instead of IR^n, one works with a ball of a suitable radius Δ_k centered at x_k. The rule for choosing Δ_k, in order to ensure the highest possible calculation rate, is an important issue of this method. To be more precise, the ratio between the reduction amount of the objective function and the reduction amount of its approximate objective function at step k, that is the function $m_k(x)$, is the basis for defining the radius Δ_{k+1}. The auxiliary problem at step k is the one of minimizing $m_k(x)$ on the closed ball $\bar{B}(x_k, \Delta_k) := \{x \in \mathrm{IR}^n \ : \ \|x - x_k\| \le \Delta_k\}$. Under some conditions, the iterative sequence $\{x_k\}$ converges to a stationary point of problem (1). The trust-region algorithm is a reducing objective function algorithm, that is $f(x_{k+1}) \le f(x_k)$ for all k.

We will make a focus on the local stability and local convergence of iterative sequences generated by the Trust-Region Method. Specifically, we will look for a result similar to that of Le Thi et al. [6, Theorem 3], where the authors have proved that: *If $\bar{x}$ is an isolated local solution of a quadratic programming problem with a polyhedral convex constraint set C, then there exist positive constants δ and μ such* *that, for any* $x_0 \in C \cap B(\bar{x}, \delta)$ *with* $B(\bar{x}, \delta) := \{x \in \mathrm{IR}^n \ : \ \|x - \bar{x}\| < \delta\}$, *the iterative sequence* $\{x_k\}$ *generated by the DCA Projection Algorithm with the initial point* x_0 *has the following properties:*

(i) $x_k \in C \cap B(\bar{x}, \mu)$ *for all* $k \geq 0$;
(ii) $x_k \to \bar{x}$ *as* $k \to \infty$.

Note that [7, Definition 2], the above property (i) is called *stability* in the sense of Lyapunov, while the property (ii) is called *attraction*. If both (i) and (ii) are valid, then, in the terminology of [7, Definition 2], $\bar{x}$ is *asymptotically stable* w.r.t. the iteration method under consideration.

Applied to the unconstrained problem (1), the Basic Trust-Region Algorithm from [1, p. 116] can be described as follows.

The BTR Algorithm:

Step 0: Initialization. An initial point x_0 and an initial trust-region radius Δ_0 are given. The constants η_1, η_2, γ_1, and γ_2 are also given and satisfy

$$0 < \eta_1 \leq \eta_2 < 1, \quad 0 < \gamma_1 \leq \gamma_2 < 1. \tag{2}$$

Compute $f(x_0)$ and set $k = 0$.

Step 1: Model definition. Equip IR^n with the inner product $\langle x, y\rangle$, the Euclidean norm $\|x\| = \langle x, x\rangle^{1/2}$, and define a model $m_k(x)$ (see, e.g., the model m_k in formula (6) below) in the closed ball $\bar{B}_k = \left\{x \in \mathrm{IR}^n \ : \ \|x - x_k\| \leq \Delta_k\right\}$.

Step 2: Step calculation. Find $s_k \in \mathrm{IR}^n$ satisfying $\|s_k\| \leq \Delta_k$, i.e., $x_k + s_k \in \bar{B}_k$, such that the function m_k "sufficiently decreases" when the argument is shifted from x_k to the point $x_k + s_k$.

Step 3: Acceptance of the trial point. Compute $f(x_k + s_k)$ and the ratio

$$\rho_k = \frac{f(x_k) - f(x_k + s_k)}{m_k(x_k) - m_k(x_k + s_k)}. \tag{3}$$

If $\rho_k \geq \eta_1$, then define $x_{k+1} = x_k + s_k$. Otherwise, let $x_{k+1} = x_k$.

Step 4: Trust-region radius update. Set

$$\Delta_{k+1} \in \begin{cases} [\Delta_k, +\infty[& \text{as } \ \rho_k \geq \eta_2, \\ [\gamma_1 \Delta_k, \Delta_k] & \text{as } \ \rho_k \in [\eta_1, \eta_2[, \\ [\gamma_1 \Delta_k, \gamma_2 \Delta_k] & \text{as } \ \rho_k < \eta_1. \end{cases} \tag{4}$$

Increase k by 1 and go to **Step 1**.

Finding an optimal choice of the constants $\eta_1, \eta_2, \gamma_1, \gamma_2$ (in order to attain the highest possible computation speed) is beyond our study in this paper. As an example, one can choose

$$\eta_1 = 0.01, \quad \eta_2 = 0.9, \quad \gamma_1 = \gamma_2 = 0.5.$$

However, other values of $\eta_1, \eta_2, \gamma_1, and \gamma_2$ can be used, provided that condition (2) is satisfied.

As $\Delta_k > 0$ and $\Delta_{k+1} > 0$, there exists a unique real number $\mu_k > 0$ such that $\Delta_{k+1} = \mu_k \Delta_k$. By introducing this coefficient μ_k, we can represent formula (4) for the trust-region radius update in the following simpler form:

$$\mu_k \in \begin{cases} [1, +\infty[& \text{as } \rho_k \in [\eta_2, +\infty[, \\ [\gamma_1, 1] & \text{as } \rho_k \in [\eta_1, \eta_2[, \\ [\gamma_1, \gamma_2] & \text{as } \rho_k \in]-\infty, \eta_1[. \end{cases} \tag{5}$$

At each step k, the choice of μ_k satisfying (5) is indeterministic. For instance, if $\rho_k \in [\eta_2, +\infty[$ then, as μ_k, one can select any number from $[1, +\infty[$. To create softwares, one usually keeps a fixed "strategy"; for example $\mu_k = 1.2$ for any k with $\rho_k \in [\eta_2, +\infty[$. (Of course, one can choose $\mu_k = 2$, or $\mu_k = 1$.) For the cases $\rho_k \in [\eta_1, \eta_2[$ and $\rho_k \in]-\infty, \eta_1[$, one can proceed similarly. It is worthy to stress that the convergence theorems in [1, Chapter 6] have been proved for any choice of μ_k, provided that (5) is satisfied for all k.

Iterations for which $\rho_k \geq \eta_1$ are called *successful*, and iterations for which $\rho_k \geq \eta_2$ are called *very successful*.

In practice, one often chooses

$$m_k(x) = f(x_k) + \langle g_k, x - x_k \rangle + \frac{1}{2}\langle x - x_k, H_k(x - x_k)\rangle, \tag{6}$$

where $g_k = \nabla f(x_k)$ is the gradient of f at x_k, and $H_k = \nabla^2 f(x_k)$ is the Hessian of f at x_k. Thus $m_k(x)$ is a linear-quadratic function.

If we assume that our model is quadratic, then it is easy to find a global minimum of $m_k(x)$ on the *Cauchy arc*, that is the segment

$$\left\{x_k - tg_k \ : \ t \geq 0, \ x_k - tg_k \in \bar{B}_k\right\}.$$

If $g_k \neq 0$, then a *Cauchy point* is given by the formula

$$x_k^C = x_k - t_k^C g_k, \tag{7}$$

where the value t_k^C is a solution of the minimization problem

$$\min\left\{m_k(x_k - tg_k) : 0 \leq t \leq \frac{\Delta_k}{\|g_k\|}\right\}. \tag{8}$$

If $g_k = 0$, then one puts $x_k^C = x_k$. Hence, in the notation of Step 2, $s_k = 0$ if $g_k = 0$, and $s_k = -t_k^C g_k$ if $g_k \neq 0$.

Let us clarify the performance of the BTR Algorithm by two numerical examples.

Example 2.1 Consider the unconstrained problem

$$\min\left\{f(x) = x_1^2 + 2x_2^2 \ : \ x = (x_1, x_2) \in \mathrm{IR}^2\right\}. \tag{9}$$

Table 1 Computation results for Example 2.1

k	$x_{k,1}$	$x_{k,2}$	$f(x_k)$
1	−1.6838	2.0513	11.2509
2	−0.9244	0.2011	0.9354
3	−0.1269	−0.1459	0.0587
4	−0.0580	0.0126	0.0037
5	−0.0080	−0.0091	0.0002
6	−0.0036	0.0008	0.0000
7	−0.0005	−0.0006	0.0000
8	−0.0002	0.0000	0.0000

To find an approximate solution of problem (9), we apply the BTR Algorithm with the Cauchy point method, the initial point $x_0 = (-2, 3)$, the initial trust-region radius $\varDelta_0 = 1$, and the tolerance $\varepsilon = 10^{-3}$. The parameters $\eta_1, \eta_2, \gamma_1, \gamma_2$ are chosen as follows: $\eta_1 = 0.25,\ \eta_2 = 0.75,\ \gamma_1 = \gamma_2 = 0.5$. The gradient vector and the Hessian matrix at the k-th iteration are, respectively,

$$g_k = \begin{bmatrix} 2x_{k,1} \\ 4x_{k,2} \end{bmatrix} \quad \text{and} \quad H_k = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}.$$

The approximate objective function is selected by the rule (6) with

$$H_k = \nabla^2 f(x_k).$$

Here, the Cauchy point is $x_k^C = x_k - t_k^C g_k$ with $t_k^C = \tau_k \frac{\varDelta_k}{\|g_k\|}$, where

$$\tau_k = \begin{cases} 1, & \text{if} \quad \langle g_k, H_k g_k \rangle \leq 0, \\ \min\left\{\|g_k\|^3 \big/ (\varDelta_k \langle g_k, H_k g_k \rangle), 1\right\}, & \text{if} \quad \langle g_k, H_k g_k \rangle > 0. \end{cases}$$

To satisfy (4), we can update the trust-region radius by the formula

$$\varDelta_{k+1} = \mu_k \varDelta_k,$$

where

$$\mu_k = \begin{cases} 2, & \text{if } \rho_k \in [\eta_2, +\infty[, \\ 0.8, & \text{if } \rho_k \in [\eta_1, \eta_2[, \\ 0.5, & \text{if } \rho_k \in]-\infty, \eta_1[. \end{cases}$$

The computation results in MATLAB are given in Table 1.

In Example 2.1, the objective function is strongly convex, quadratic.

Table 2 Computation results for Example 2.2

k	x_k	$f(x_k)$
1	2.5000	6.6406
2	1.7606	0.8521
3	1.3151	−0.1169
4	1.0861	−0.2419
5	1.0093	−0.2499
6	1.0001	−0.2500

Example 2.2 Consider the following nonconvex, nonquadratic problem

$$\min \left\{ f(x) = \frac{1}{4}x^4 - \frac{1}{2}x^2 \,:\, x \in \mathrm{IR} \right\}. \tag{10}$$

It has two local minimum points $x = -1$ and $x = 1$, and one local maximum point $x = 0$. Chose $\Delta_0 = 0.5$, and $\varepsilon = 10^{-3}$. The parameters $\eta_1, \eta_2, \gamma_1, \gamma_2$, the formula for determining the Cauchy point, and the trust-region radius update rule are the same as in the preceding example. For the case $x_0 = 3$, the computation results in MATLAB are given in Table 2. Now, we let x_0 to be one of the 21 equidistant points from the segment $[-2, 2]$. The computation results obtained by the BTR Algorithm in MATLAB with these initial points are shown in Table 3. For any $x_0 \in]-2, 0[$, the iterative sequence converges to $x_* = -1$. For any $x_0 \in]0, 2[$, the iterative sequence converges to $x_* = 1$. For $x_0 = 0$, the algorithm stops right at the first step, because this x_0 is a stationary point of (10).

3 Local Convergence and Stability

We are going to establish an analogue of [6, Theorem 3], where the authors have proved the local stability and local convergence of iterative sequences generated by the DCA Projection Algorithm (see Algorithm A in [6, p. 484]). Here, a linear convergence rate will be also obtained.

Definition 3.1 One calls x_* a *nonsingular local minimizer* of (1) if

$$\nabla f(x_*) = 0 \quad \text{and} \quad \nabla^2 f(x_*) \succ 0 \tag{11}$$

where, for a square matrix A, the writing $A \succ 0$ means that A is positive definite.

Definition 3.2 (See, e.g., [8]) Let $\{x_k\}$ be a sequence in IR^n that converges to x_*. We say that the convergence is *R–linear* if

$$\limsup_{k\to\infty} \|x_k - x_*\|^{1/k} < 1.$$

Table 3 Further computation results for Example 2.2

j	x_0	x_*	$f(x_k)$
0	−2.0000	−1.0000	−0.2500
1	−1.8000	−1.0002	−0.2500
2	−1.6000	−1.0000	−0.2500
3	−1.4000	−1.0005	−0.2500
4	−1.2000	−1.0000	−0.2500
5	−1.0000	−1.0000	−0.2500
6	−0.8000	−1.0003	−0.2500
7	−0.6000	−1.0000	−0.2500
8	−0.4000	−1.0000	−0.2500
9	−0.2000	−1.0000	−0.2500
10	0.0000	0.0000	0.0000
11	0.2000	1.0000	−0.2500
12	0.4000	1.0000	−0.2500
13	0.6000	1.0005	−0.2500
14	0.8000	1.0003	−0.2500
15	1.0000	1.0000	−0.2500
16	1.2000	1.0000	−0.2500
17	1.4000	1.0005	−0.2500
18	1.6000	1.0000	−0.2500
19	1.8000	1.0002	−0.2500
20	2.0000	1.0000	−0.2500

For the sake of completeness of our presentation, we provide an elementary, direct proof of the following well-known fact.

Lemma 3.1 *If x_* is a nonsingular local minimizer point of (1), then there exist positive constants $\delta, \alpha_1, \alpha_2$ with $\alpha_1 < \alpha_2$, such that*

$$\alpha_1 I \preceq \nabla^2 f(x) \preceq \alpha_2 I \quad \forall x \in \bar{B}(x_*, \delta), \tag{12}$$

where I stands for the $n \times n$ unit matrix; relation $A \preceq B$ between two square matrices means that $B - A$ is positive semidefinite.

Proof Since f is twice-continuously differentiable, the function

$$g(x, v) := \left\langle \nabla^2 f(x) v, v \right\rangle$$

is continuous on $\mathrm{IR}^n \times \mathrm{IR}^n$. The formula $\varphi(x) = \min\{g(x, v) : v \in S^{n-1}\}$, with $S^{n-1} = \{v \in \mathrm{IR}^n : \|v\| = 1\}$ being the unit sphere, defines a continuous function on IR^n. Indeed, for every $\bar{x} \in \mathrm{IR}^n$, by the Weierstrass theorem, we can find $\bar{v} \in S^{n-1}$ such that $\varphi(\bar{x}) = g(\bar{x}, \bar{v})$. Let $\varepsilon > 0$ be given arbitrarily.

Due to the continuity of $\nabla^2 f(x)$, there exists an open neighborhood U_1 of x_* such that

$$\left\langle \nabla^2 f(x)\bar{v}, \bar{v} \right\rangle \le \left\langle \nabla^2 f(\bar{x})\bar{v}, \bar{v} \right\rangle + \varepsilon \quad \forall x \in U_1.$$

Then, for any $x \in U_1$, it holds that

$$\begin{aligned} \varphi(x) &\le g(x, \bar{v}) = \left\langle \nabla^2 f(x)\bar{v}, \bar{v} \right\rangle \le \left\langle \nabla^2 f(\bar{x})\bar{v}, \bar{v} \right\rangle + \varepsilon \\ &= \varphi(\bar{x}) + \varepsilon. \end{aligned}$$

Hence $\varphi(x) - \varphi(\bar{x}) \le \varepsilon$ for every $x \in U_1$.

For any $v \in S^{n-1}$, since the function $h(x, v') := g(\bar{x}, v') - g(x, v')$ is continuous at $(\bar{x}, v)$ and $h(\bar{x}, v) = 0$, there exist open neighborhoods U_v and W_v of $\bar{x}$ and v, respectively, such that

$$g(\bar{x}, v') - g(x, v') \le \varepsilon \quad \forall x \in U_v, \ \forall v' \in W_v. \tag{13}$$

By the compactness of S^{n-1}, from the open cover $\{W_v\}_{v \in S^{n-1}}$ one can extract a finite subcover $\{W_{v_i}\}_{i=\overline{1,k}}$. Put $U_2 = \bigcap_{i=1}^{k} U_{v_i}$. Let $v' \in S^{n-1}$ and $x \in U_2$ be given arbitrarily. Select an index $i \in \{1, \ldots, k\}$ such that $v \in W_{v_i}$. Applying (13) for $v = v_i$ yields

$$g(\bar{x}, v') - \varepsilon \le g(x, v').$$

Therefore,

$$\varphi(\bar{x}) - \varepsilon = g(\bar{x}, \bar{v}) - \varepsilon \le g(\bar{x}, v') - \varepsilon \le g(x, v').$$

It follows that $\varphi(\bar{x}) - \varepsilon \le \min\{g(x, v') \,:\, v' \in S^{n-1}\} = \varphi(x)$.

So, we have $|\varphi(x) - \varphi(\bar{x})| \le \varepsilon$ for every $x \in U_1 \cap U_2$. The continuity of φ has been proved.

Since $\varphi(x_*)$ is the smallest eigenvalue of the matrix $\nabla^2 f(x_*)$ and since x_* is a nonsingular local minimizer of (1), we have $\varphi(x_*) > 0$. By the continuity of φ at x_*, for any $\alpha_1 \in \left]0, \varphi(x_*)\right[$, there exists $\delta_1 > 0$ such that

$$\varphi(x) \ge \alpha_1 \quad \forall x \in \bar{B}(x_*, \delta_1).$$

As $\varphi(x)$ is the smallest eigenvalue of $\nabla^2 f(x)$, this implies that $\alpha_1 I \preceq \nabla^2 f(x)$ for all $x \in \bar{B}(x_*, \delta_1)$.

Similarly, by considering the function $\psi(x) := \max\{g(x, v) \,:\, v \in S^{n-1}\}$ and arguing as above, for any $\alpha_2 \in \left]\varphi(x_*), +\infty\right[$, we can find $\delta_2 > 0$ such that

$$\psi(x) \le \alpha_2 \quad \forall x \in \bar{B}(x_*, \delta_2).$$

As $\psi(x)$ is the largest eigenvalue of $\nabla^2 f(x)$, the latter implies that $\nabla^2 f(x) \preceq \alpha_2 I$ for all $x \in \bar{B}(x_*, \delta_2)$.

Setting $\delta = \min\{\delta_1, \delta_2\}$, we get the property in (12). □

The forthcoming remark is due to the referee of this paper.

Remark 3.1 The proof of Lemma 3.1 may be shortened considerably if one uses a classical result of matrix theory saying that the eigenvalues of a matrix are continuous functions of its entries. This is of course also true for the greatest and smallest eigenvalues of a symmetric matrix. Thus, if $\varphi(x)$ and $\psi(x)$ are, respectively, the greatest eigenvalue and the smallest eigenvalue of $\nabla^2 f(x)$, then φ and ψ are continuous. Then, one can go directly to the last part of the above proof.

We will also need next auxiliary result, which is a local version of Lemma 5.6 from [4]. Note that the arguments given by Ruszczynski [4, p. 226] are fully applied to this setting.

Lemma 3.2 *If $\delta, \alpha_1, \alpha_2, \alpha_1 < \alpha_2$, are some positive constants satisfying* (12)*, then*

$$\|\nabla f(x)\|^2 \geq \alpha_1 \left(1 + \frac{\alpha_1}{\alpha_2}\right) [f(x) - f(x_*)] \quad \forall x \in \bar{B}(x_*, \delta). \tag{14}$$

We are now in the position to prove the main result of this paper. In accordance with the discussion given in Sect. 2, the assertions (i) and (ii) below show that the nonsingular local minimizer x_* in question is asymptotically stable w.r.t. the BTR Algorithm. The assertion (iii) tells us that the latter is linearly convergent. Our proof is based on Lemma 3.1, Lemma 3.2, and some arguments from [4, Theorem 5.7, p. 227]. Note that Theorem 5.7 from [4] is about the R-linear convergence rate of the *steepest descent method*, which is quite different from the BTR Algorithm.

Theorem 3.1 *Suppose that x_* is a nonsingular local minimizer of* (1)*. Then, there exist $\delta > 0$ and $\delta_1 > 0$ such that, for every $x_0 \in \bar{B}(x_*, \delta)$, the iterative sequence $\{x_k\}$ generated by the BTR Algorithm with the initial point x_0 and the Cauchy point method, where the $m_k(x)$ is the sum of the first three terms of the Taylor expansion of f at x_k, has the following properties:*

(i) $x_k \in \bar{B}(x_*, \delta_1)$ *for all k;*
(ii) $\{x_k\}$ *converges to x_*;*
(iii) *The convergence rate of $\{x_k\}$ to x_* is R-linear.*

Proof By Lemma 3.1, we can find positive constants $\bar{\delta}$, α_1, and α_2 such that (12) holds for $\delta = \bar{\delta}$ and for all $x \in \bar{B}(x_*, \bar{\delta})$. Because x_* is a nonsingular local minimizer of (1), there exists $\bar{\delta} > 0$ such that f is strongly convex on $\bar{B}(x_*, \bar{\delta})$, i.e., one can find $\alpha > 0$ satisfying

$$f\left((1-t)x_1 + tx_2\right) \leq (1-t)f(x_1) + tf(x_2) - \alpha t(1-t)\|x_1 - x_2\|^2$$

for all $t \in]0, 1[$, $x_1, x_2 \in \bar{B}\left(x_*, \bar{\delta}\right)$ (see [5]). Then, x_* is the unique local solution of (1) belonging to $\bar{B}(x_*, \bar{\delta})$, and (14) is fulfilled with $\delta = \bar{\delta}$.

As $\nabla^2 f(x)$ and $\nabla f(x)$ are continuous in $\bar{B}(x_*, \bar{\delta})$, there exists $\delta > 0$ such that $\delta_1 := \delta \left(\frac{\alpha_2}{\alpha_1}\right)^{1/2}$ satisfies the condition $\delta_1 < \bar{\delta}$ and, at the same time, if $x \in \bar{B}(x_*, \delta_1) \setminus \{x_*\}$, $x_1, x_2 \in \bar{B}(x_*, \delta_1)$ then

$$\frac{\Delta_0}{\|\nabla f(x)\|} \geq \frac{1}{\alpha_1}, \quad \|\nabla^2 f(x_1) - \nabla^2 f(x_2)\| \leq \frac{\alpha_1^2}{3\alpha_2}, \tag{15}$$

and

$$2-\frac{\left\langle\nabla f(x),\nabla^2 f(x_1)\nabla f(x)\right\rangle}{\left\langle\nabla f(x),\nabla^2 f(x_2)\nabla f(x)\right\rangle}\geq\eta_2. \tag{16}$$

Condition (16) can be satisfied because, due to the first inequality in (12), we have

$$\begin{aligned}
&\left|\frac{\left\langle\nabla f(x),\nabla^2 f(x_1)\nabla f(x)\right\rangle}{\left\langle\nabla f(x),\nabla^2 f(x_2)\nabla f(x)\right\rangle}-1\right|\\
&=\left|\frac{\left\langle\nabla f(x),\left(\nabla^2 f(x_1)-\nabla^2 f(x_2)\right)\nabla f(x)\right\rangle}{\left\langle\nabla f(x),\nabla^2 f(x_2)\nabla f(x)\right\rangle}\right|\\
&\leq\frac{\|\nabla^2 f(x_1)-\nabla^2 f(x_2)\|\,\|\nabla f(x)\|^2}{\alpha_1\|\nabla f(x)\|^2}\\
&=\frac{1}{\alpha_1}\|\nabla^2 f(x_1)-\nabla^2 f(x_2)\|\\
&\leq\frac{1}{\alpha_1}\left(\|\nabla^2 f(x_1)-\nabla^2 f(x_*)\|+\|\nabla^2 f(x_*)-\nabla^2 f(x_2)\|\right).
\end{aligned}$$

By the continuity of $\nabla^2 f(x)$, the last expression goes to 0 when $\delta_1\to 0$. Hence, recalling that $\eta_2\in]0,1[$, we can find $\delta>0$ such that (16) is fulfilled for all $x\in\bar{B}(x_*,\delta_1)\setminus\{x_*\}$ and $x_1,x_2\in\bar{B}(x_*,\delta_1)$.

We now show that the assertions (i)–(iii) of the theorem are valid for the chosen positive numbers δ and δ_1.

Since $\alpha_1\leq\alpha_2$ and $\delta_1=\delta\left(\frac{\alpha_2}{\alpha_1}\right)^{1/2}$, we have $\delta\leq\delta_1$; so $0<\delta\leq\delta_1<\bar{\delta}$.

Take an arbitrary point $x_0\in\bar{B}(x_*,\delta)$. If $x_0=x_*$, then $x_k=x_*$ for $k=0,1,2,\ldots$. Hence, in this case, the assertions (i)–(iii) are true. Thus, it remains to consider the case $x_0\neq x_*$. By the choice of $\delta,\delta_1,\bar{\delta}$ and x_0, x_0 cannot be a local solution of (1). Applying the necessary and sufficient optimality condition in the Fermat form for the convex minimization problem with an open constraint set

$$\min\{f(x)\ :\ x\in B(x_*,\bar{\delta})\},$$

we can assert that $g_0=\nabla f(x_0)$ is a nonzero vector.

As f is strongly convex in $\bar{B}(x_*,\bar{\delta})$, $\nabla^2 f(x)\succ 0$ for every $x\in B(x_*,\bar{\delta})$ (see, e.g., [5]).

Since t_0^C is a solution of the problem (8) with m_k as in the formulation of the theorem and $H_k=\nabla^2 f(x_k)\succ 0$, we see that t_0^C is the unique solution of the problem

$$\min\left\{m_0\left(x_0-tg_0\right):\ 0\leq t\leq\frac{\Delta_0}{\|g_0\|}\right\}$$

with

$$m_0\left(x_0-tg_0\right)=m_0(x_0)-t\|g_0\|^2+\frac{1}{2}t^2\left\langle g_0,\nabla^2 f(x_0)g_0\right\rangle. \tag{17}$$

Put

$$t_0^* = \frac{\|g_0\|^2}{\langle g_0, \nabla^2 f(x_0) g_0 \rangle} \tag{18}$$

and note that the derivative of the strongly convex quadratic function on the right-hand side of (17) vanishes at t_0^*.

From (18), (12), and (15), we see that

$$t_0^* \leq \frac{1}{\alpha_1} \leq \frac{\Delta_0}{\|g_0\|};$$

hence

$$t_0^C = t_0^* = \frac{\|g_0\|^2}{\langle g_0, \nabla^2 f(x_0) g_0 \rangle}, \quad t_0^C \leq \frac{1}{\alpha_1}. \tag{19}$$

The second-order Taylor expansion for $f(x)$ at x_0 leads to the formula

$$f(x_0 - t_0^C g_0) = f(x_0) - t_0^C \|g_0\|^2 + \frac{1}{2}(t_0^C)^2 \left\langle g_0, \nabla^2 f(\xi) g_0 \right\rangle \tag{20}$$

with $\xi = x_0 - \theta t_0^C g_0$, for some $\theta \in [0, 1]$.

By (3), (17), (19), and (20), we have

$$\begin{aligned}
\rho_0 &= \frac{f(x_0) - f(x_0 - t_0^C g_0)}{m_0(x_0) - m_0(x_0 - t_0^C g_0)} \\
&= \frac{t_0^C \|g_0\|^2 - \frac{1}{2}(t_0^C)^2 \left\langle g_0, \nabla^2 f(\xi) g_0 \right\rangle}{t_0^C \|g_0\|^2 - \frac{1}{2}(t_0^C)^2 \left\langle g_0, \nabla^2 f(x_0) g_0 \right\rangle} \\
&= \frac{\frac{\|g_0\|^4}{\langle g_0, \nabla^2 f(x_0) g_0 \rangle} - \frac{1}{2} \frac{\|g_0\|^4}{\langle g_0, \nabla^2 f(x_0) g_0 \rangle^2} \left\langle g_0, \nabla^2 f(\xi) g_0 \right\rangle}{\frac{\|g_0\|^4}{\langle g_0, \nabla^2 f(x_0) g_0 \rangle} - \frac{1}{2} \frac{\|g_0\|^4}{\langle g_0, \nabla^2 f(x_0) g_0 \rangle}} \\
&= 2 - \frac{\langle g_0, \nabla^2 f(\xi) g_0 \rangle}{\langle g_0, \nabla^2 f(x_0) g_0 \rangle}.
\end{aligned}$$

Therefore, from (16) we can deduce that $\rho_0 \geq \eta_2$. Hence the passage from x_0 to $x_0 - t_0^C g_0$ is a very successful iteration. In addition, one has $x_1 = x_0 - t_0^C g_0$ and $\Delta_1 \geq \Delta_0$.

On the other hand, because $\langle g_0, \nabla^2 f(x_0) g_0 \rangle \leq \alpha_2 \|g_0\|^2$ by the second inequality in (12) and the function $t \mapsto m_0(x_0 - t g_0)$ attains its global minimum on the half-line

$\{x_0 - tg_0 \, : \, t \geq 0\}$ at t_0^C, one has

$$m_0(x_1) = m_0(x_0 - t_0^C g_0) \leq m_0(x_0 - \frac{1}{\alpha_2} g_0)$$
$$= f(x_0) - \frac{1}{\alpha_2}\|g_0\|^2 + \frac{1}{2\alpha_2^2}\left\langle g_0, \nabla^2 f(x_0) g_0 \right\rangle$$
$$\leq f(x_0) - \frac{1}{2\alpha_2}\|g_0\|^2. \tag{21}$$

By (17),

$$m_0(x_0 - t_0^C g_0) = f(x_0) - t_0^C \|g_0\|^2 + \frac{1}{2}(t_0^C)^2 \left\langle g_0, \nabla^2 f(x_0) g_0 \right\rangle.$$

Since $x_1 = x_0 - t_0^C g_0$, the last equality and (20) imply that

$$m_0(x_1) - f(x_1) = \frac{1}{2}(t_0^C)^2 \left\langle g_0, (\nabla^2 f(x_0) - \nabla^2 f(\xi)) g_0 \right\rangle.$$

Combining this with (21) gives

$$f(x_1) = m_0(x_1) + \frac{1}{2}(t_0^C)^2 \left\langle g_0, (\nabla^2 f(\xi) - \nabla^2 f(x_0)) g_0 \right\rangle$$
$$\leq f(x_0) - \frac{1}{2\alpha_2}\|g_0\|^2 + \frac{1}{2}(t_0^C)^2 \left\langle g_0, (\nabla^2 f(\xi) - \nabla^2 f(x_0)) g_0 \right\rangle. \tag{22}$$

From (22), (19), (15), and Lemma 3.2, it follows that

$$\begin{aligned} & f(x_1) - f(x_*) \\ & \leq f(x_0) - f(x_*) - \frac{1}{2\alpha_2}\|g_0\|^2 + \frac{1}{2}(t_0^C)^2 \left\langle g_0, (\nabla^2 f(\xi) - \nabla^2 f(x_0)) g_0 \right\rangle \\ & \leq f(x_0) - f(x_*) - \frac{1}{2\alpha_2}\|g_0\|^2 + \frac{1}{2\alpha_1^2}\|g_0\|^2 \|\nabla^2 f(\xi) - \nabla^2 f(x_0)\| \\ & \leq f(x_0) - f(x_*) - \frac{1}{2\alpha_2}\|g_0\|^2 + \frac{1}{6\alpha_2}\|g_0\|^2 \\ & = f(x_0) - f(x_*) - \frac{\|g_0\|^2}{3\alpha_2} \\ & \leq \left(1 - \frac{\alpha_1}{3\alpha_2} - \frac{\alpha_1^2}{3\alpha_2^2}\right)[f(x_0) - f(x_*)]. \end{aligned} \tag{23}$$

By the Taylor expansion of f at x_*, for every $x \in \bar{B}(x_*, \bar{\delta})$, we have

$$f(x) - f(x_*) = \frac{1}{2}\left\langle x - x_*, \nabla^2 f(\bar{x})(x - x_*) \right\rangle,$$

where $\bar{x} = (1-\theta)x_* + \theta x$ with $\theta \in [0, 1]$ depending on x. Applying the Cauchy–Schwartz inequality and (12) yields

$$\frac{\alpha_1}{2}\|x - x_*\|^2 \leq f(x) - f(x_*) \leq \frac{\alpha_2}{2}\|x - x_*\|^2. \tag{24}$$

Due to (24), from (23) we have

$$\begin{aligned}
\frac{\alpha_1}{2}\|x_1 - x_*\|^2 &\leq f(x_1) - f(x_*) \\
&\leq \left(1 - \frac{\alpha_1}{3\alpha_2} - \frac{\alpha_1^2}{3\alpha_2^2}\right)[f(x_0) - f(x_*)] \\
&\leq \left(1 - \frac{\alpha_1}{3\alpha_2} - \frac{\alpha_1^2}{3\alpha_2^2}\right)\frac{\alpha_2}{2}\|x_0 - x_*\|^2.
\end{aligned} \tag{25}$$

Hence,

$$\begin{aligned}
\|x_1 - x_*\| &\leq \left(1 - \frac{\alpha_1}{3\alpha_2} - \frac{\alpha_1^2}{3\alpha_2^2}\right)^{1/2}\left(\frac{\alpha_2}{\alpha_1}\right)^{1/2}\|x_0 - x_*\| \\
&\leq \left(\frac{\alpha_2}{\alpha_1}\right)^{1/2}\|x_0 - x_*\| \\
&\leq \delta\left(\frac{\alpha_2}{\alpha_1}\right)^{1/2} = \delta_1.
\end{aligned}$$

It follows that $x_1 \in \bar{B}(x_*, \delta_1)$. Arguing similarly by induction, we can conclude that each iteration step k is very successful, $x_{k+1} = x_k - t_k^C g_k$, and $\Delta_k \geq \Delta_{k-1}$, where

$$t_k^C = \frac{\|g_k\|^2}{\langle g_k, \nabla^2 f(x_k) g_k\rangle} \leq \frac{1}{\alpha_1} \leq \frac{\Delta_0}{\|g_k\|} \leq \frac{\Delta_k}{\|g_k\|}.$$

Similarly as in (23), we have

$$\begin{aligned}
f(x_k) - f(x_*) &\leq \left(1 - \frac{\alpha_1}{3\alpha_2} - \frac{\alpha_1^2}{3\alpha_2^2}\right)\left[f(x_{k-1}) - f(x_*)\right] \\
&\leq \left(1 - \frac{\alpha_1}{3\alpha_2} - \frac{\alpha_1^2}{3\alpha_2^2}\right)^k [f(x_0) - f(x_*)].
\end{aligned}$$

Combining this with the inequalities similar to those which have been given in (25), we see that

$$\begin{aligned}\frac{\alpha_1}{2}\|x_k - x_*\|^2 &\le f(x_k) - f(x_*)\\ &\le \left(1 - \frac{\alpha_1}{3\alpha_2} - \frac{\alpha_1^2}{3\alpha_2^2}\right)\left[f(x_{k-1}) - f(x_*)\right]\\ &\le \left(1 - \frac{\alpha_1}{3\alpha_2} - \frac{\alpha_1^2}{3\alpha_2^2}\right)^k [f(x_0) - f(x_*)]\\ &\le \left(1 - \frac{\alpha_1}{3\alpha_2} - \frac{\alpha_1^2}{3\alpha_2^2}\right)^k \frac{\alpha_2}{2}\|x_0 - x_*\|^2.\end{aligned}$$

It follows that

$$\|x_k - x_*\| \le \left(1 - \frac{\alpha_1}{3\alpha_2} - \frac{\alpha_1^2}{3\alpha_2^2}\right)^{k/2} \left(\frac{\alpha_2}{\alpha_1}\right)^{1/2} \|x_0 - x_*\| \le \delta_1. \tag{26}$$

From this we obtain assertion (i). Since

$$0 < 1 - \frac{\alpha_1}{3\alpha_2} - \frac{\alpha_1^2}{3\alpha_2^2} < 1,$$

from the first inequality in (26), we have $x_k \to x_*$ when $k \to \infty$. Assertion (ii) has been proved. By (26),

$$\begin{aligned}&\limsup_{k\to\infty} \|x_k - x_*\|^{1/k}\\ &\quad\le \limsup_{k\to\infty}\left[\left(1 - \frac{\alpha_1}{3\alpha_2} - \frac{\alpha_1^2}{3\alpha_2^2}\right)^{1/2} \left(\frac{\alpha_2}{\alpha_1}\right)^{\frac{1}{2k}} \|x_0 - x_*\|^{\frac{1}{k}}\right] < 1.\end{aligned}$$

This justifies the R-linear convergence rate of $\{x_k\}$ stated in (iii) and completes the proof. □

We conclude this section by a simple example showing that the nonsingularity assumption on the local solution x_* is vital for Theorem 3.1. Namely, we will see that, if the equality $\nabla f(x_*) = 0$ in (11) is satisfied, but the condition $\nabla^2 f(x_*) \succ 0$ there is replaced by the weaker condition $\nabla^2 f(x_*) \succeq 0$ then, despite to our efforts in choosing the constants $\Delta_0, \eta_1, \eta_2, \gamma_1$ and γ_2, the BTR Algorithm with the Cauchy arc point method is neither locally stable, nor locally convergent around x_*.

Example 3.1 Consider the problem

$$\min\left\{f(x) = x^3 \ :\ x \in \mathrm{IR}\right\}.$$

To analyze the stability and convergence of iterative sequences generated by the BTR Algorithm around the stationary point $x_* := 0$, we let the constants $\Delta_0, \eta_1, \eta_2, \gamma_1$ and

γ_2 be chosen arbitrarily. Here we have $\nabla f(x_*) = 0$ and $\nabla^2 f(x_*) = 0$. Suppose that $x_0 < 0$. The approximate objective function $m_0(x)$ given by (6) is the following one

$$m_0(x_0 - tg_0) = 27x_0^5 t^2 - 9x_0^4 t + x_0^3.$$

Since $\nabla_t m_0(x_0 - tg_0) = 54x_0^5 t - 9x_0^4 < 0$ for all $t \in]0, \frac{\Delta_0}{\|g_0\|}]$, (8) has the unique solution $t_0^C = \frac{\Delta_0}{\|g_0\|} = \frac{\Delta_0}{3x_0^2}$. By (7), $x_0^C = x_0 - t_0^C g_0 = x_0 - \Delta_0$. Since $s_0 = -\Delta_0$, using (3), we find that

$$\rho_0 = 1 + \frac{\Delta_0^2}{3x_0(x_0 - \Delta_0)} > 1 \geq \eta_2.$$

So, the current iteration step $k = 0$ is very successful. Therefore, by (4), the new trust-region radius Δ_1 needs just to satisfy the inequality $\Delta_1 \geq \Delta_0$. Continuing the computation in the same way, we get $t_k^C = \frac{\Delta_k}{3x_k^2}$ and, by induction,

$$x_k^C = x_k - \Delta_k = x_0 - \Delta_0 - \Delta_1 - \cdots - \Delta_k.$$

As $\rho_k > 1 \geq \eta_2$, the $k-$th iteration is very successful. This implies that the new trust-region radius Δ_{k+1} needs just to satisfy the condition $\Delta_{k+1} \geq \Delta_k$. Hence,

$$x_k = x_0 - \Delta_0 - \Delta_1 - \cdots - \Delta_k \leq x_0 - (k+1)\Delta_0$$

for all k. This yields $x_k \to -\infty$ as $k \to \infty$. So the BTR Algorithm is neither locally stable nor locally convergent in a neighborhood of $x_* = 0$.

4 Conclusions

We have established the local stability, local convergence, and R-linear convergence rate of the Basic Trust-Region Algorithm for the unconstrained twice-continuously differentiable optimization problem in a neighborhood of a nonsingular local minimizer. It is worthy to find out whether similar results are also valid for constrained optimization problems, or not.

Acknowledgements Research of B.N. Muoi was supported by Hanoi Pedagogical University No. 2 (Vietnam) under Grant Number C.2016.12. Research of N.D. Yen was supported by the project VAST.HTQT.Pháp.03/16-17 between VAST and CNRS. We thank the anonymous referee for helpful comments and suggestions.

References

1. Conn, A.R., Gould, N.I.M., Toint, P.L.: Trust-Region Methods. MPS-SIAM Series on Optimization, Mathematical Programming Society and Society for Industrial and Applied Mathematics, Philadelphia (2000)
2. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer, New York (1999)

3. Polyak, B.T.: Introduction to Optimization. Optimization Software, Inc., New York (1987)
4. Ruszczynski, A.: Nonlinear Optimization. Princeton University Press, Princeton (2005)
5. Vasilev, F.P.: Numerical Methods for Solving Extremal Problems (in Russian), 2nd edn. Nauka, Moscow (1988)
6. Le Thi, H.A., Pham Dinh, T., Yen, N.D.: Properties of two DC algorithms in quadratic programming. J. Global Optim. **49**, 481–495 (2011)
7. Leong, W.J., Goh, B.S.: Convergence and stability of line search methods for unconstrained optimization. Acta Appl. Math. **127**, 155–167 (2013)
8. Ortega, J.M., Rheinboldt, W.C.: Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, New York (1970)