Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 4, No. 2, pp. 259 – 277, May 2008

A CONTINUOUS NEWTON-TYPE METHOD FOR UNCONSTRAINED OPTIMIZATION*

Lei-Hong Zhang, C.T. Kelley† and Li-Zhi Liao‡

Abstract: In this paper, we propose a continuous Newton-type method in the form of an ordinary differential equation by combining the negative gradient and the Newton direction. We show that for a general function $f(x)$, our method converges globally to a connected subset of the stationary points of $f(x)$ under some mild conditions, and converges globally to a single stationary point for a real analytic function. The method reduces to the exact continuous Newton method if the Hessian matrix of $f(x)$ is uniformly positive definite. We report on convergence of the new method on the set of standard test problems in the literature.

Key words: *unconstrained optimization, continuous method, ODE method, global convergence, pseudo-transient continuation*

Mathematics Subject Classification: *65K10, 65L05, 90C47*

1 Introduction

In this paper we consider the solution schemes for the following unconstrained optimization problem

$$\min_{x\in R^n} f(x), \quad f \in C^2(R^n), \tag{1.1}$$

by using the so-called *continuous method* or *Ordinary Differential Equation* (**ODE**) method (see [8, 12, 13, 19, 25, 32, 38] and the references therein). Different from the conventional optimization approaches, such method adopts some kind of differential equation with the initial condition to define the trajectory of variable x in terms of a parameter t. By tracing this trajectory, the *stationary point(s)* satisfying $\nabla f(x) = 0$, or hopefully, the local minimizer of $f(x)$ can be located. To be more precise, let $x(t)$ for $t \in T \subseteq R$, be the solution of the following initial value problem (IVP):

$$\begin{cases} \frac{dx(t)}{dt} &= h(x), \quad t \geq 0 \\ x(0) &= x_0, \end{cases} \tag{1.2}$$

*This research was supported in part by FRG grants from Hong Kong Baptist University and the Research Grant Council of Hong Kong.

†The research of C. T. Kelley was supported by National Science Foundation Grants DMS-0404537 and DMS-0707220, and Army Research Office Grants W911NF-04-1-0276, W911NF-06-1-0096, W911NF-06-1-0412, and W911NF-07-1-0112.

‡Corresponding author.

where $h : R^n \to R^n$ is a continuous mapping and T denotes the maximal interval of existence of $x(t)$. The solution of (1.2), $\{x(t),\ t \in T\}$, is said to be the trajectory of (1.2), and without confusion, in order to simplify the following presentation, we also call $x(t)$ the trajectory of (1.2).

From the optimization point of view, the simplest trajectory of (1.2) can be defined with $h(x) = d_G(x) = -\nabla f(x)$, i.e.,

$$\begin{cases} \frac{dx(t)}{dt} & = d_G(x), \quad t \geq 0, \\ x(0) & = x_0, \end{cases} \tag{1.3}$$

which goes back to Cauchy and was proposed to solve some optimization problems in [10]. This method has been studied extensively later, e.g., in [1, 4, 5, 17, 37].

Another natural trajectory of (1.2) can be defined from Newton's direction, say $h(x) = d_N(x) = -(\nabla^2 f(x))^{-1}\nabla f(x)$, and the IVP becomes

$$\begin{cases} \frac{dx(t)}{dt} & = d_N(x), \quad t \geq 0, \\ x(0) & = x_0. \end{cases} \tag{1.4}$$

It is called the continuous Newton method, and if the Hessian matrix $\nabla f^2(x)$ along the trajectory $x(t)$ of (1.4) is positive definite for $t \in T$, it then follows that

$$\nabla f(x(t)) = e^{-t}\nabla f(x_0), \quad t \in T, \tag{1.5}$$

and hence

$$\frac{\nabla f(x(t))}{\|\nabla f(x(t))\|_2} = \frac{\nabla f(x_0)}{\|\nabla f(x_0)\|_2} \in S^{n-1} := \{x \in R^n | \|x\|_2 = 1\}, \quad t \in T. \tag{1.6}$$

This property is observed and utilized in [18] to show an interesting topological property of the continuous Newton method that in some neighborhood of a strict local minimizer x^* of $f(x)$, when perturbing $f(x)$ to $\tilde{f}(x)$ on an open region not containing x_0 and x^*, the trajectories $x(t)$ and $\tilde{x}(t)$ with $x(0) = \tilde{x}(0) = x_0$ defined by the continuous Newton equations for $f(x)$ and $\tilde{f}(x)$ coincide outside the perturbed region, hence coincide in their asymptotic states.

The continuous Newton method converges very fast since $\|\nabla f(x(t))\|_2$ reduces exponentially as indicated by (1.5). However, the nonpositive definiteness of the Hessian matrix $\nabla^2 f(x)$ is the major obstacle for this method. In [6], Branin then considered the following corresponding form

$$\nabla^2 f(x)\frac{dx(t)}{dt} = \mp\nabla f(x), \tag{1.7}$$

and suggested to change the sign of (1.7) whenever its trajectory $x(t)$ encounters a change in sign of the determinant of $\nabla^2 f(x(t))$ or arrives at a solution point of $\nabla f(x) = 0$ in order to find multiple local minima numerically. Moreover, Branin also suggested to employ the adjoint matrix, say $A(x)$, of $\nabla^2 f(x)$ to get around the singularity, and then to replace (1.7) with

$$\frac{dx(t)}{dt} = -A(x)\nabla f(x), \tag{1.8}$$

which is now well-defined in R^n. However, the consequence of adopting (1.8), the troublesome *extraneous singular points* (see [6]) defined by $\{\hat{x} \in R^n | A(\hat{x})\nabla f(\hat{x}) = 0,\ \nabla f(\hat{x}) \neq 0\}$ are induced (see [22, 23] for the structure of such extraneous singular points).

An analogous modification of (1.8) proposed by Smale ([36]) is called "global Newton equation", and has the following form in the context of the unconstrained optimization

$$\nabla^2 f(x)\frac{dx(t)}{dt} = -\phi(x)\nabla f(x), \tag{1.9}$$

where $\phi(x)$ is a real function suggested specifically to satisfy the following condition

$$sign(\phi(x)) = sign(\det(\nabla^2 f(x))),$$

and the simple choice of $\phi(x) = \det(\nabla^2 f(x))$ leads to the equation (1.8) immediately.

Additional research related to continuous Newton method has been carried out (see [2, 3, 12, 13, 14, 21] and the reference therein). For example, Diener et al. developed the so-called "Newton-leaves" and attempted to connect several or all of the stationary points of $f(x)$ in a single connected trajectory. For more details, reader can refer to [12, 13, 14].

In this paper, we propose a continuous Newton-type method (in the form of an ODE), which combines the negative gradient $d_G(x)$ and Newton's direction $d_N(x)$, and is well-defined in R^n. It is shown that our method gets around the singularities of $\nabla^2 f(x)$, and under certain conditions, it converges globally to a connected stationary point subset for a general function $f(x)$, and converges globally to a stationary point for a real analytic function[§] $f(x)$. Moreover, the trajectory defined by the proposed ODE moves strictly downhill with respect to $f(x)$ (meaning that the value of $f(x(t))$ is strictly decreasing as t increases); and for the uniformly convex function $f(x)$, it becomes the exact Newton trajectory of (1.4), and therefore, the fast convergence can be achieved.

The rest of this paper is organized as follows. In the next section, the ODE corresponding to our continuous Newton-type method is established and the existence and uniqueness of the trajectory are verified. The convergence analysis of this trajectory is addressed in Section 3. A powerful numerical solver for some continuous models is examined for our new continuous Newton-type method in Section 4. The encouraging numerical results on a set of standard test problems are presented in Section 5. Some concluding remarks are drawn in Section 6.

2 A continuous Newton-type Method

First, let's state some assumptions on the objective function $f(x)$ that we are going to minimize. Let

$$L = \{x \in R^n | f(x) \leq f(x_0)\}$$

be the level set of $f(x)$, and let $L_{f(x_0)}$ denote the connected component of L that contains the point x_0.

Assumptions:

(a) $\nabla^2 f(x)$ is at least locally Lipschitz continuous in R^n.

[§]A real function is said to be analytic if it possesses derivatives of all orders and agrees with its Taylor series in the neighborhood of every point.

(b) $f(x)$ is bounded from below by $f^* > -\infty$.

(c) For any $x_0 \in R^n$, $L_{f(x_0)}$ is bounded.

It is clear that Assumption (c) is much weaker than the condition that the level set $L = \{x \in R^n | f(x) \le f(x_0)\}$ is bounded. For example, if $f(x) = \sin x,\ x \in R$, and for any given $x_0 \in R$ with $f(x_0) \ne 1$, the level set

$$L = \bigcup_{k=-\infty}^{+\infty} [(2k-1)\pi - \arcsin f(x_0), 2k\pi + \arcsin f(x_0)],\ k \in \mathbb{Z} \text{ (the integer set)},$$

is unbounded, but

$$L_{f(x_0)} = [(2\tilde{k}-1)\pi - \arcsin f(x_0), 2\tilde{k}\pi + \arcsin f(x_0)]$$

for some $\tilde{k} \in \mathbb{Z}$ with $x_0 \in L_{f(x_0)}$ is bounded. From Assumption (c), we know also that the set $L_{f(x_0)}$ is compact, and furthermore, for any $x_0 \in R^n$, the set $S_{f(x_0)}$ defined by

$$S_{f(x_0)} := S \cap L_{f(x_0)}, \tag{2.1}$$

is compact too, where S is the stationary points set of $f(x)$ given by

$$S := \{x \in R^n | \nabla f(x) = 0\}. \tag{2.2}$$

Consider the following continuous Newton-type differential equation,

$$\begin{cases} \frac{dx(t)}{dt} &= d(x), \quad t \ge 0 \\ x(0) &= x_0, \end{cases} \tag{2.3}$$

where

$$d(x) = \alpha(x) d_N(x) + \beta(x) d_G(x), \tag{2.4}$$

with

$$\alpha(x) = \begin{cases} 1, & \text{if } \lambda_{\min}(x) > \delta_2, \\ \frac{\lambda_{\min}(x)-\delta_1}{\delta_2-\delta_1}, & \text{if } \delta_1 \le \lambda_{\min}(x) \le \delta_2, \\ 0, & \text{if } \lambda_{\min}(x) < \delta_1; \end{cases} \tag{2.5}$$

and

$$\beta(x) = 1 - \alpha(x) = \begin{cases} 0, & \text{if } \lambda_{\min}(x) > \delta_2, \\ \frac{\delta_2-\lambda_{\min}(x)}{\delta_2-\delta_1}, & \text{if } \delta_1 \le \lambda_{\min}(x) \le \delta_2, \\ 1, & \text{if } \lambda_{\min}(x) < \delta_1. \end{cases} \tag{2.6}$$

Here, $\lambda_{\min}(x)$ represents the smallest eigenvalue of $\nabla^2 f(x)$, and $\delta_2 > \delta_1 > 0$ are two predefined positive constants. It should be noted that the smallest eigenvalue of $\nabla^2 f(x)$, $\lambda_{\min}(x)$, can be easily estimated from the modified Cholesky factorization in [33] numerically.

A first observation is that for a general function $f(x) \in C^2(R^n)$, (2.4) is well-defined in R^n; and when $\nabla^2 f(x)$ is uniformly positive definite, i.e., $y^T \nabla^2 f(x) y \ge \tilde{\delta} \|y\|_2^2$, for some constant $\tilde{\delta} > 0$ and $\forall y \in R^n$, the trajectory generated from (2.3) is exactly the continuous Newton trajectory for any $0 < \delta_1 < \delta_2 \le \tilde{\delta}$. Since the direction $d(x)$ in (2.4) is continuous in R^n, the Cauchy-Peano existence theorem implies that there is a solution to the IVP (2.3); for the uniqueness of the solution, furthermore, we need to prove $d(x)$ is also locally Lipschitz continuous in R^n. The first result shows that under Assumption (a), $\lambda_{\min}(x)$ is locally Lipschitz continuous in R^n, which is a direct consequence of the following Wielandt-Hoffman lemma.

Lemma 2.1 ([16], p. 396). *If A and $A+E$ are n-by-n symmetric matrices, then*

$$|\lambda_k(A+E) - \lambda_k(A)| \leq \|E\|_2, \quad k = 1, \cdots, n,$$

where $\lambda_k(A)$ designates the kth largest eigenvalue of A.

Since $\nabla^2 f(x)$ is locally Lipschitz continuous by Assumption (a), the previous lemma reveals that for any two points y, z in some neighborhood of any $x \in R^n$,

$$|\lambda_{\min}(y) - \lambda_{\min}(z)| \leq \|\nabla^2 f(y) - \nabla^2 f(z)\|_2 \leq C_1 \|y - z\|_2, \tag{2.7}$$

where $C_1 > 0$ is the Lipschitz constant of $\nabla^2 f(x)$, and hence it follows that $\lambda_{\min}(x)$ is locally Lipschitz continuous in R^n. Assumptions (a) also implies that

$$\|d_G(y) - d_G(z)\|_2 \quad \leq \quad C_2 \|y - z\|_2, \tag{2.8}$$

for a constant $C_2 > 0$ and for any two points y, z in some neighborhood of $x \in R^n$.

Moreover, from the result in [30] (p.46), we know that for any $x \in R^n$, if $\lambda_{\min}(x) > 0$, there exist a $\gamma > 0$ and a neighborhood $N_\tau(x)$ of x such that $\forall y \in N_\tau(x)$, $\nabla^2 f(y)$ is invertible and $\|(\nabla^2 f(y))^{-1}\|_2 \leq \gamma$. Hence, for any $y, z \in N_\tau(x)$, it follows

$$\begin{aligned} \|(\nabla^2 f(y))^{-1} - (\nabla^2 f(z))^{-1}\|_2 &= \|(\nabla^2 f(y))^{-1}[\nabla^2 f(z) - \nabla^2 f(y)](\nabla^2 f(z))^{-1}\|_2 \\ &\leq \gamma^2 \cdot \|\nabla^2 f(z) - \nabla^2 f(y)\|_2 \\ &\leq \gamma^2 C_1 \|z - y\|_2, \end{aligned} \tag{2.9}$$

which implies that when $\lambda_{\min}(x) > 0$, $(\nabla^2 f(x))^{-1}$ is Lipschitz continuous at x too; and additionally, it is true that

$$\|d_N(y) - d_N(z)\|_2 \quad \leq \quad C_3 \|y - z\|_2, \quad \forall y, z \in N_\tau(x), \tag{2.10}$$

for some positive constant $C_3 > 0$ and $\lambda_{\min}(x) > 0$.

Theorem 2.2. *Suppose that $f(x)$ satisfies Assumptions (a), (b) and (c), then for any $x(0) = x_0 \in R^n$, there exists a unique solution $x(t)$ to (2.3), and the maximal interval of existence of the solution can be extended to $[0, +\infty)$.*

Proof. See the Appendix. □

We next provide a general result which shows that the trajectory $x(t)$ of (2.3) will never reach the set $S_{f(x_0)}$ at finite time $t \geq 0$ provided that $\nabla f(x_0) \neq 0$. This result is the extension of Theorem 2(iii) in [26] which obtains the same conclusion for the gradient system (1.3).

Theorem 2.3. *Suppose $h : R^n \to R^n$ is locally Lipschitz continuous. Then for any $x_0 = x(0) \in R^n$ with $h(x_0) \neq 0$, the solution to the IVP (1.2), $x(t)$, satisfies $h(x(t)) \neq 0$ for any $t \in T$, where T denotes the maximal interval of existence of $x(t)$.*

Proof. The proof can be conducted along the same arguments as Theorem 2 (iii) in [26]. □

Under the Assumptions (a), (b) and (c), (6.1) together with the previous theorem reveals that $f(x(t))$ is strictly decreasing along the trajectory as t increases whenever $\nabla f(x_0) \neq 0$. This property also guarantees that there is no periodic solution for (2.3).

Theorem 2.4. *There is no periodic solution to (2.3) for any $x(0) = x_0 \in R^n$ with $\nabla f(x_0) \neq 0$.*

Proof. See the Appendix. □

3 Convergence Analysis

Since for any $x_0 \in R^n$, the solution $x(t)$ of (2.3) is unique and its maximal interval of existence can be extended to $[0, +\infty)$, we then can apply some results (refer to [31]) of the dynamical system to develop the convergence analysis.

Consider the ODE

$$\frac{dx(t)}{dt} = h(t). \tag{3.1}$$

We suppose that for any $x_0 \in R^n$, the solution $x(t)$ to (3.1) with $x(0) = x_0$ is unique and its right maximal interval of existence is $[0, +\infty)$.

Definition 3.1. A point $p \in R^n$ is an ω-limit point of the trajectory $x(t)$ of (3.1) with $x(0) = x_0$ if there is a sequence $t_i \to +\infty$ (as $i \to +\infty$) such that

$$\lim_{i \to +\infty} x(t_i) = p.$$

The set of all ω-limit points of the trajectory $x(t)$ of (3.1) with $x(0) = x_0$ is called the ω-limit set of $x(t)$ and it is denoted by Ω_{x_0}.

Some properties of the ω-limit set are summarized in the following Lemma 3.2 (see [31], p. 175).

Lemma 3.2. *The ω-limit set of a trajectory $x(t)$ of (3.1) with $x(0) = x_0$, Ω_{x_0}, is a closed subset of R^n and if $x(t)$ is contained in a compact subset of R^n, then Ω_{x_0} is non-empty, connected and compact subset of R^n.*

Denote by

$$\text{dist}(x(t), \Omega_{x_0}) = \inf_{\hat{x} \in \Omega_{x_0}} \|x(t) - \hat{x}\|_2$$

the distance of $x(t)$ to the set Ω_{x_0}. If the trajectory $x(t)$ is contained in a compact subset of R^n, it then is true that

$$\lim_{t \to +\infty} \text{dist}(x(t), \Omega_{x_0}) = 0,$$

since otherwise, there exists some $\epsilon > 0$ and a sequence $\{t_i\}$ such that $t_i \to +\infty$ and

$$\text{dist}(x(t_i), \Omega_{x_0}) > \epsilon, \quad i = 1, 2, \cdots. \tag{3.2}$$

The boundedness of $\{x(t_i)\}$ then implies that there is a convergent subsequence $\{x(t_i')\}$ with the limit point $p \in R^n$. The fact $p \in \Omega_{x_0}$ then contradicts (3.2).

Applying these results to (2.3), we can analyze the convergence of the trajectory $x(t)$ of (2.3).

Remark 3.3. Let Ω_{x_0} represent the ω-limit set of trajectory $x(t)$ of (2.3) with $x(0) = x_0 \in R^n$. As indicated by (6.1), $\Omega_{x_0} \subseteq L_{f(x_0)}$; and moreover, we can say that the trajectory $x(t)$ converges to the set Ω_{x_0} as $t \to +\infty$ in the sense that

$$\lim_{t \to +\infty} \text{dist}(x(t), \Omega_{x_0}) = 0.$$

If, in addition, Ω_{x_0} contains only one point $p \in R^n$, it then implies that

$$\lim_{t \to +\infty} \mathrm{dist}(x(t), \Omega_{x_0}) = \lim_{t \to +\infty} \inf_{\hat{x} \in \{p\}} \|x(t) - \hat{x}\|_2 = \lim_{t \to +\infty} \|x(t) - p\|_2 = 0,$$

which equivalently implies that $x(t)$ converges to a single point $p \in R^n$.

The following theorem gives the convergence result for a general function $f(x)$.

Theorem 3.4. *Suppose $f(x)$ satisfies Assumptions (a), (b), and (c), and for any $x_0 \in R^n$, let $x(t)$ be the trajectory of (2.3) with $x(0) = x_0 \in R^n$, and let Ω_{x_0} denote the ω-limit set of $x(t)$. Then there exists some constant $\bar{f}$ such that*

$$\Omega_{x_0} \subseteq \{x \in R^n | f(x) = \bar{f}\} \cap S_{f(x_0)}; \tag{3.3}$$

and $x(t)$ converges to some connected subset of $S_{f(x_0)}$ as $t \to +\infty$, where $S_{f(x_0)}$ is defined by (2.1).

Proof. Since $\nabla f(x_0) = 0$ is the trivial case in which the unique trajectory becomes $x(t) \equiv x_0,\ t \geq 0$ (due to uniqueness), we just consider $\nabla f(x_0) \neq 0$.

From (6.1) and Theorem 2.3, it follows that $f(x(t))$ is strictly decreasing as t increases, but still bounded below by Assumption (b), which consequently implies that there exists a constant, say $\bar{f}$, so that

$$\lim_{t \to +\infty} f(x(t)) = \bar{f}.$$

As a result, for any $\bar{x} \in \Omega_{x_0}$, there exists a sequence $\{t_i\}_{i=1}^{+\infty}$ such that $t_i \to +\infty$, $x(t_i) \to \bar{x}$ and $f(x(t_i)) \to f(\bar{x}) = \bar{f}$ as $i \to +\infty$, which implies $\Omega_{x_0} \subseteq \{x \in R^n | f(x) = \bar{f}\}$ directly.

Furthermore, the LaSalle invariant set theorem (Theorem 3.4 in [35]) says that for any $\bar{x} \in \Omega_{x_0}$, we have $\frac{df(\bar{x})}{dt} = \nabla f(\bar{x})^T d(\bar{x}) = 0$, which is true only when $\nabla f(\bar{x}) = 0$ by (6.1). Consequently, from Lemma 3.2, Remark 3.3 and $x(t) \in L_{f(x_0)}$ for $t \geq 0$, we conclude $\Omega_{x_0} \subseteq \{x \in R^n | f(x) = \bar{f}\} \cap S_{f(x_0)}$, and complete the proof. □

Special cases of the set Ω_{x_0} below directly lead to the convergence to a stationary point, and the proof is obvious by Lemma 3.2, Remark 3.3, and Theorem 3.4.

Corollary 3.5. *Under the conditions of Theorem 3.4, suppose that $x(t)$ is the trajectory of (2.3) with $x(0) = x_0 \in R^n$. If each point in $S_{f(x_0)}$ is isolated from one another, then $x(t)$ converges to a stationary point as $t \to +\infty$; and therefore, if there is an $\bar{x} \in \Omega_{x_0}$ being a strictly local minimizer of $f(x)$, then $x(t) \to \bar{x}$ as $t \to +\infty$.*

However, in general, it should be pointed out that converging to a (single) stationary point may not be obtained, because it is known that the trajectory of (1.3) will not necessarily converge to a single point (see [19], Prop. C.12.1; and see [1] for a counterexample). By only endowing $f(x)$ to be real analytic additionally, however, converging to a single stationary point is achievable. The proof for this is based on Corollary 3.5 and similar to the proof of Theorem 2.2 in [1].

Theorem 3.6. *Suppose that $f(x)$ is a real analytic function satisfying Assumption (a), (b), and (c). Then for any $x_0 \in R^n$, the trajectory $x(t)$ of (2.3) converges to a (single) stationary point of $f(x)$ as $t \to +\infty$ for any $x(0) = x_0 \in R^n$.*

Proof. We just need to consider the case $\nabla f(x_0) \neq 0$. Let Ω_{x_0} be the ω-limit set of $x(t)$. If there exists an $\bar{x} \in \Omega_{x_0}$ such that $\lambda_{\min}(\bar{x}) > 0$, then $\bar{x}$ must be a strictly local minimizer of $f(x)$ and Corollary 3.5 completes the proof already; otherwise, $\forall \bar{x} \in \Omega_{x_0}$, $\lambda_{\min}(\bar{x}) \leq 0$. We prove next that $\bar{x}$ is the unique point in Ω_{x_0} and therefore $\lim_{t\to+\infty} x(t) = \bar{x}$.

Obviously, there exists a neighborhood $N_{\tau_1}(\bar{x})$ of $\bar{x}$ such that $\forall x \in N_{\tau_1}(\bar{x})$, $\lambda_{\min}(x) < \delta_1$ for the predefined $\delta_1 > 0$ in (2.5). Also, since $f(x)$ is real analytic, the following Łojasiewicz gradient inequality (see [27]) holds in a neighborhood $N_{\tau_2}(\bar{x})$ of $\bar{x}$,

$$\|\nabla f(x)\|_2 \geq c|f(x) - f(\bar{x})|^\sigma, \quad \forall x \in N_{\tau_2}(\bar{x}),$$

for some constants $c > 0$ and $\sigma \in [0, 1)$. We then can assume that for any sufficiently small $\epsilon > 0$, the Łojasiewicz gradient inequality and $\lambda_{\min}(x) < \delta_1$ hold in the neighborhood $N_\epsilon(\bar{x})$.

From Theorem 3.4 and $\nabla f(x_0) \neq 0$, it follows that $f(x(t)) > f(\bar{x})$ for $t \geq 0$. Then for any $x(t) \in N_\epsilon(\bar{x})$, we have

$$\frac{d[f(x(t)) - f(\bar{x})]}{dt} = -\|\nabla f(x(t))\|_2^2 \leq -c[f(x(t)) - f(\bar{x})]^\sigma \cdot \|\frac{dx(t)}{dt}\|_2,$$

or equivalently,

$$c_1 \frac{d[f(x(t)) - f(\bar{x})]^{1-\sigma}}{dt} \leq -\|\frac{dx(t)}{dt}\|_2, \tag{3.4}$$

where $c_1 = (c(1-\sigma))^{-1} > 0,\ c > 0$ and $\sigma \in [0, 1)$.

Note that $\bar{x}$ is an accumulation point and $f(x(t)) \to f(\bar{x})$ as $t \to +\infty$, there must exist some $t_1 \geq 0$ such that the following two inequalities hold simultaneously,

$$\begin{aligned} \|x(t_1) - \bar{x}\|_2 &< \frac{\epsilon}{2}, \\ c_1[f(x(t_1)) - f(\bar{x})]^{1-\sigma} &< \frac{\epsilon}{2}. \end{aligned}$$

Suppose $x(t)$ will leave $N_\epsilon(\bar{x})$ after t_1, and let t_2 be the smallest such that $\|x(t_2) - \bar{x}\|_2 = \epsilon$, then $x(t) \in N_\epsilon(\bar{x})$ for all $t \in (t_1, t_2)$. From (3.4) and the decreasing property of $f(x(t))$, we get

$$\begin{aligned} 0 < \int_{t_1}^{t_2} \|\frac{dx(t)}{dt}\|_2 dt &\leq c_1[f(x(t_1)) - f(\bar{x})]^{1-\sigma} - c_1[f(x(t_2)) - f(\bar{x})]^{1-\sigma} \\ &< c_1[f(x(t_1)) - f(\bar{x})]^{1-\sigma} < \frac{\epsilon}{2}. \end{aligned}$$

Therefore,

$$\begin{aligned} \|x(t_2) - \bar{x}\|_2 &\leq \|x(t_2) - x(t_1)\|_2 + \|x(t_1) - \bar{x}\|_2 \\ &\leq \int_{t_1}^{t_2} (\|\frac{dx(t)}{dt}\|_2) dt + \|x(t_1) - \bar{x}\|_2 < \epsilon. \end{aligned}$$

This contradiction implies that $\forall \epsilon > 0$ arbitrarily small, $\exists$ a t_1 such that $\|x(t) - \bar{x}\|_2 < \epsilon$, $\forall t \geq t_1$, this is just the definition of the convergence of $x(t)$ to $\bar{x}$ as $t \to +\infty$. □

It should be mentioned that different from Theorem 2.2 in [1], we do not rely on the angle condition

$$\frac{df(x(t))}{dt} \equiv \nabla f(x(t))^T \frac{dx(t)}{dt} \leq -\theta \|\nabla f(x(t))\|_2 \cdot \|\frac{dx(t)}{dt}\|_2, \quad \theta > 0, \tag{3.5}$$

to prove that $x(t)$ converges to a single stationary point. This is due to our special structure of (2.3). Moreover, the converging point of the trajectory $x(t)$ of (2.3) is also a stationary point, which is stronger than that of Theorem 2.2 in [1]. In general, Theorem 2.2 of [1] can still be strengthened to guarantee the convergence to a stationary point of a real analytic function $f(x)$, and an analogous version is presented as follows.

Theorem 3.7. *Let $f(x)$ be a real analytic function and let $x(t)$ be a C^1 curve in R^n with $\frac{dx(t)}{dt} = h(x)$. Assume that there exist a $\theta > 0$ and a real η such that for $t > \eta$, $x(t)$ satisfies the angle condition (3.5) and*

$$[\frac{df(x)}{dt} = 0] \Rightarrow [h(x) = 0] \Rightarrow [\nabla f(x) = 0]. \tag{3.6}$$

Then, as $t \to +\infty$, either $\|x(t)\|_2 \to \infty$ or there exists an $x^* \in R^n$ such that $x(t) \to x^*$ with $\nabla f(x^*) = 0$.

Proof. According to Theorem 2.2 in [1], we just need to verify $\nabla f(x^*) = 0$. Lemma 2 in [7] (p. 429) ensures that if there exists an $x^* \in R^n$ such that $x(t) \to x^*$ as $t \to +\infty$, then $h(x^*) = 0$, and hence by (3.6), it leads to the result. □

4 Pseudo-transient Continuation

Pseudo-transient continuation (Ψtc) is one way to solve (1.2). This Ψtc method was originally designed for finding steady-state solutions to time-dependent differential equations without computing a fully time-accurate solution. The approach can also be adapted to optimization problems. We refer to [24, 9, 15, 20, 11] for the details of the theory and some applications. In this section we will only summarize the method. We report numerical results for Ψtc in Section 5.

In the context of optimization, one would integrate (1.2) numerically, managing the "time step", say ξ_i, in a way that, while maintaining stability, would increase as rapidly as possible in order to make the transition to Newton's method near the solution. One way to implement this is the iteration

$$x_{i+1} = x_i - (\xi_i^{-1} I + H(x_i))^{-1} h(x_i), \quad i = 0, 1, \cdots, \tag{4.1}$$

where $H(x)$ is the model Hessian or $H(x) = h'(x)$. A common way to manage the time step ξ_i is "Switched Evolution Relaxation" (SER) [29]

$$\xi_{i+1} = \xi_i \|h(x_i)\| / \|h(x_{i+1})\|, \quad i = 0, 1, \cdots . \tag{4.2}$$

SER is supported by theory, and it is this approach we use in the numerical test in Section 5.

One thing we should mention is that for the sequence $\{x_i\}$ generated from Ψtc , the corresponding objective function value sequence $\{f(x_i)\}$ may not be monotonically decreasing. This is different from the continuous method (2.3) where $\frac{df(x(t))}{dt} \leq 0$, for $t \geq 0$.

5 Computational Experiments

This section deals with the numerical test of our continuous Newton-type method whose ODE is in the form of (2.3) in comparing with the continuous steepest descent method whose ODE is in the form of (1.3) by using the Matlab ODE solver. In addition, we also report the numerical results of Ψtc in solving the related ODEs. For this purpose, the set of the 17 standard test functions (except for the last Chebyquad function) for unconstrained minimization from [28] is used and tested with their dimensions ranging from 2 to 400. For each test function, the same initial value x_0 as in [28] is used. The test problems are summarized in Table 1.

Table 1. Test Problems

$No.$	$Function\ name$	n	m
$P1$	$Helical\ valley\ function$	3	3
$P2$	$Biggs\ EXP6\ function$	6	$m \geq n$
$P3$	$Gaussian\ function$	3	15
$P4$	$Powell\ badly\ scaled\ function$	2	2
$P5$	$Box\ three-dimensional\ function$	3	$m \geq n$
$P6$	$Variably\ dimensioned\ function$	n	$m = n + 2$
$P7$	$Watson\ function$	$2 \leq n \leq 31$	31
$P8$	$Penalty\ function\ I$	n	$m = n + 1$
$P9$	$Penalty\ function\ II$	n	$m = 2n$
$P10$	$Brown\ badly\ scaled\ function$	2	3
$P11$	$Brown\ and\ Dennis\ function$	4	m
$P12$	$Gulf\ research\ and\ development\ function$	3	$n \leq m \leq 100$
$P13$	$Trigonometric\ function$	n	$m = n$
$P14$	$Extended\ Rosenbrock\ function$	$n(even)$	$m = n$
$P15$	$Extended\ Powell\ singular\ function$	$n(multiple\ of\ 4)$	$m = n$
$P16$	$Beale\ function$	2	3
$P17$	$Wood\ function$	4	6

5.1 Matlab Platform

All computation in this section is performed on Matlab platform. Before presenting our numerical results, several points should be clarified. First, the minimum eigenvalue routine used in our tests is directly based on the MATLAB code **eig.m**, although the attractive modified Cholesky factorization in [33] can be used for efficiency consideration. Second, for each test function, the explicit expression of $\nabla^2 f(x)$ is provided. Third, due to the result of Theorem 3.4, we do not have to require, as Theorem 3.6 states, that the test functions are real analytic. Finally, we let $\delta_2 = 1000\delta_1$ in (2.5), and fix δ_1 to $\delta^{(0)} = 1e-9$, but if this fails for some problems, δ_1 is set as $\delta^{(1)} = 1e-4$.

All our tests are performed on a PC with Intel(R) Pentium(R)4 Processor at 3.20GHz. The nonstiff ODE solver **ODE113** is used to solve (1.3) and (2.3) with the relative tolerance **rtol** $= 1e-8$ and absolute tolerance **atol** $= 1e-9$ to control the accuracy of the integrated trajectory (see [34] for the details of these options), and $\|\frac{d}{dt}x(t)\|_\infty \leq 1e-6$ is the stopping criterion. The CPU times to obtain the acceptable solutions are summarized in Table 2 where '$*$' denotes that the algorithm cannot stop within 1000 seconds of the CPU time; and the CPU times of the continuous steepest descent ODE (1.3) and our continuous Newton-type ODE (2.3) are denoted by CPU_G and CPU_N, respectively. In addition, we also list the smallest eigenvalue (labeled as $\lambda^*_{\min}$) of the Hessian at the computed point x^* for supporting the validity of our choices of δ_1, δ_2 and for detecting whether the computed point is a local minimizer. f^*_G and f^*_N represent the final computed objective function values from (1.3) and (2.3), respectively.

Table 2. Comparison of (1.3) and (2.3) on ODE113

No.	n	m	$CPU_G(s)$	$CPU_N(s)$	$\lambda^*_{\min}$	f^*_G	f^*_N
$P1$	3	3	2.5781	0.5313	$1.4328e-000$	$6.4722e-013$	$7.9391e-013$
$P2$	6	6	128.9375	165.2656	$-4.5330e-005$	$3.5509e-005$	$3.5509e-005$
$P3$	3	15	0.0938	0.0469	$1.3966e-001$	$1.1283e-008$	$1.1282e-008$
$P4$	2	2	$*$	604.8594	$1.0059e-006$	$*$	$4.1537e-010$
$P5$	3	10	18.2656	$7.3750(\delta^{(1)})$	$9.1158e-004$	$5.6492e-010$	$5.6174e-012$
$P5$	3	20	15.3438	$7.2031(\delta^{(1)})$	$1.6145e-003$	$3.1329e-010$	$2.8701e-012$
$P6$	5	7	0.1563	0.0625	$2.0000e-000$	$1.1589e-015$	$4.0253e-011$
$P6$	10	12	0.1406	0.0781	$2.0000e-000$	$1.9155e-015$	$5.8237e-010$
$P6$	20	22	0.1719	0.1250	$2.0000e-000$	$4.6993e-016$	$1.9398e-008$
$P6$	30	32	0.1875	0.4375	$2.0000e-000$	$1.7847e-016$	$6.3986e-008$
$P7$	2	31	0.1250	0.0781	$2.3977e+001$	$5.4661e-001$	$5.4661e-001$
$P7$	6	31	$*$	1.6250	$2.8101e-003$	$*$	$2.2877e-003$
$P7$	8	31	$*$	4.8750	$7.5430e-006$	$*$	$1.8162e-005$
$P8$	4	5	20.9688	0.1250	$7.9998e-005$	$2.2514e-005$	$2.2500e-005$
$P8$	10	11	15.0313	0.1719	$1.2648e-004$	$7.0893e-005$	$7.0877e-005$
$P8$	20	21	11.3906	0.1875	$1.7887e-004$	$1.5780e-004$	$1.5778e-004$
$P8$	50	51	9.1563	0.4531	$2.8281e-004$	$4.3181e-004$	$4.3179e-004$
$P8$	100	101	8.7031	1.4531	$3.9993e-004$	$9.0253e-004$	$9.0249e-004$
$P8$	200	201	9.7031	6.5781	$5.6554e-004$	$1.8611e-003$	$1.8611e-003$
$P9$	4	8	0.1719	0.4844	$2.9693e-006$	$9.4914e-006$	$9.3763e-006$
$P9$	10	20	773.6875	0.4531	$1.8842e-005$	$2.9369e-004$	$2.9366e-004$
$P9$	20	40	$*$	0.3281	$1.3795e-004$	$*$	$6.3897e-003$
$P9$	50	100	188.5469	0.5313	$1.6645e-002$	$4.2961e-000$	$4.2961e-000$
$P9$	100	200	2.5000	1.5156	$2.2137e-001$	$9.7096e+004$	$9.7096e+004$
$P9$	200	400	14.3281	5.7188	$2.6871e+002$	$4.7116e+013$	$4.7116e+013$
$P10$	2	3	$*$	5.2188	$2.0000e-000$	$*$	$2.5763e-015$
$P11$	4	10	0.8750	0.3125	$4.7720e-000$	$1.4432e-000$	$1.4432e-000$
$P11$	4	20	4.0625	0.1563	$1.5158e+003$	$8.5822e+004$	$8.5822e+004$
$P11$	4	50	$*$	0.3594	$1.4581e+009$	$*$	$2.6684e+016$
$P11$	4	100	$*$	0.6406	$1.5186e+018$	$*$	$1.5087e+034$
$P12$	3	3	$*$	0.3438	$1.9330e-006$	$*$	$3.2312e-007$
$P13$	5	5	0.4063	0.5156	$1.5045e-001$	$4.3481e-011$	$1.5018e-011$
$P13$	10	10	0.2500	0.6875	$9.8024e-001$	$2.7951e-005$	$2.7951e-005$
$P14$	2	2	10.5625	0.1094	$3.9936e-001$	$3.9442e-012$	$2.9867e-013$
$P14$	10	10	11.2031	0.1250	$3.9936e-001$	$1.9721e-011$	$1.4933e-012$
$P14$	20	20	12.2500	0.2500	$3.9936e-001$	$3.9442e-011$	$2.9867e-012$
$P14$	50	50	15.4063	0.9844	$3.9936e-001$	$9.8606e-011$	$7.4667e-012$
$P14$	100	100	28.2813	4.5938	$3.9936e-001$	$1.9721e-010$	$1.4933e-011$
$P14$	200	200	79.7500	27.0313	$3.9936e-001$	$3.9442e-010$	$2.9867e-011$
$P14$	400	400	340.0625	212.2969	$3.9936e-001$	$7.8885e-010$	$5.9733e-011$
$P15$	4	4	234.0938	3.7656	$3.2196e-008$	$1.4476e-009$	$3.1023e-015$
$P15$	20	20	400.0781	5.3281	$3.2596e-008$	$7.2380e-009$	$1.5628e-014$
$P15$	40	40	606.6875	10.8438	$3.2228e-008$	$1.4476e-008$	$2.4472e-014$
$P15$	100	100	$*$	46.2813	$3.2281e-008$	$*$	$6.3657e-014$
$P15$	200	200	$*$	198.1563	$3.2127e-008$	$*$	$1.1339e-013$
$P16$	2	3	0.6719	0.3281	$3.0146e-001$	$2.2351e-012$	$1.0640e-013$
$P17$	4	6	23.9219	$6.7031(\delta^{(1)})$	$7.1957e-001$	$1.6888e-012$	$5.4878e-013$

Except for the second problem $P2$, where the computed solution x^* is a saddle point, the rest computed points are all local minima. These numerical results clearly demonstrate that our continuous Newton-type ODE (2.3) is much more efficient and reliable compared with the steepest descent ODE (1.3), and it converges globally to the regular stationary point(s).

5.2 Ψtc Approach

Though the continuous Newton-type ODE (2.3) can be successively solved by the sophisticated ODE solver **ODE113**, it seems still time-consuming since it is intended to produce a high accurate trajectory by cautiously controlling the stepsize. However, for an ODE model in optimization, either (1.3) or (2.3), the accuracy of the trajectory is of no essential consequence as long as the asymptotical point can be found, and hence the steady-state solutions of ODE are essential. According to this point, we employ Ψtc which is in the spirit of efficiently implementing an ODE model in optimization to implement (1.3) and (2.3) for (1.1).

As mentioned in Section 4, Ψtc is a very fast solver for (1.2). Even though the objective function values at the computed points generated by Ψtc would not be monotonically decreasing, yet its fast convergence would always provide an attractive and competitive approach for any ODE resulted from the optimization problem. In our Ψtc implementation for (1.3), we utilize the SER (4.2) strategy to update the time step ξ_i. Table 3 and Table 4 summarize the numerical results with the initial time steps $\xi_0 = 1e-1$ and $\xi_0 = 1e-2$ respectively, in which $Iter$ represents the number of iterations, f^* represents the final objective function value, and ξ_i^* represents the final value of the time step ξ_i.

Table 3. Numerical results of Ψtc for (1.3) with $\xi_0 = 1e-1$

$No.$	n	m	$Iter$	$CPU(s)$	f^*	ξ_i^*
$P1$	3	3	41	0.0313	$5.8305e-013$	$1.9527e+004$
$P2$	6	6	78	0.1719	$3.5505e-005$	$9.7840e+006$
$P3$	3	15	8	0.0781	$1.1279e-008$	$1.0285e+004$
$P4$	2	2	42	0.0781	$1.3039e-008$	$8.5200e+002$
$P5$	3	10	46	0.0313	$8.2370e-019$	$2.2327e+007$
$P5$	3	20	141	0.0781	$3.6143e-014$	$1.0918e+006$
$P6$	5	7	11	0.0156	$9.8752e-011$	$5.1768e+005$
$P6$	10	12	14	0.0313	$1.0315e-009$	$1.7620e+006$
$P6$	20	22	16	0.0156	$1.9000e-003$	$4.6426e+007$
$P6$	30	32	17	0.0313	$5.5257e-002$	$1.3659e+007$
$P7$	2	31	6	0.0155	$5.4661e-001$	$8.5340e+006$
$P7$	6	31	16	0.1406	$2.3000e-003$	$1.7110e+006$
$P7$	8	31	18	0.3125	$1.8162e-005$	$3.0671e+007$
$P7$	9	31	17	0.4063	$1.4375e-006$	$1.0825e+007$
$P8$	4	5	21	0.0313	$2.2501e-005$	$2.2898e+006$
$P8$	10	11	13	0.0156	$7.4403e-005$	$2.3594e+006$
$P8$	20	21	15	0.0313	$1.6347e-004$	$4.3232e+007$
$P8$	50	51	16	0.0313	$1.7000e-002$	$3.5908e+007$
$P8$	100	101	17	0.0938	$4.5525e-001$	$1.1564e+007$
$P8$	200	201	17	0.1563	$3.7352e+001$	$1.1867e+007$
$P9$	4	8	21	0.0313	$9.3763e-006$	$6.5247e+005$
$P9$	10	20	32	0.0313	$2.9367e-004$	$2.1867e+004$
$P9$	20	40	32	0.1250	$6.3897e-003$	$8.4172e+005$
$P9$	50	100	22	0.0625	$4.2961e-000$	$2.6271e+006$
$P9$	100	200	19	0.1094	$9.7096e+004$	$1.8427e+007$
$P9$	200	400	10	0.1406	$4.7116e+013$	$3.8687e+005$
$P10$	2	3	17	0.0157	$1.3580e-014$	$1.5268e+006$
$P11$	4	10	85	0.0625	$1.4433e-000$	$1.0524e+007$
$P11$	4	20	17	0.0313	$8.5822e+004$	$1.4144e+007$
$P11$	4	50	12	0.0313	$2.6684e+016$	$1.5367e+007$
$P11$	4	100	12	0.0313	$1.5087e+034$	$3.3363e+005$
$P12$	3	3	4	0.0313	$1.4000e-003$	$1.0000e+005$
$P13$	5	5	538	0.2188	$4.0773e-017$	$2.3249e+007$
$P13$	10	10	664	0.3906	$2.7951e-005$	$3.2628e+007$
$P14$	2	2	16	0.0155	$4.1877e-015$	$2.3416e+004$
$P14$	10	10	16	0.0156	$2.0939e-014$	$2.3416e+004$
$P14$	20	20	16	0.0157	$4.1877e-014$	$2.3416e+004$
$P14$	50	50	16	0.0313	$1.0469e-013$	$2.3416e+004$
$P14$	100	100	16	0.0938	$2.0939e-013$	$2.3416e+004$
$P14$	200	200	16	0.3438	$4.1877e-013$	$2.3416e+004$
$P14$	400	400	16	1.1719	$8.3754e-013$	$2.3416e+004$
$P15$	4	4	18	0.0156	$2.0684e-009$	$1.4122e+007$
$P15$	20	20	18	0.0156	$1.0342e-008$	$1.4122e+007$
$P15$	40	40	17	0.0313	$2.0684e-008$	$1.4122e+007$
$P15$	100	100	17	0.0469	$5.1711e-008$	$1.4122e+007$
$P15$	200	200	17	0.1094	$1.0342e-007$	$1.4122e+007$
$P16$	2	3	$fail$	$fail$	$fail$	$fail$
$P17$	4	6	61	0.0313	$3.5720e-019$	$2.0420e+007$

Table 4. Numerical results of Ψtc *for (1.3) with* $\xi_0 = 1e - 2$

$No.$	n	m	$Iter$	$CPU(s)$	f^*	ξ_i^*
$P1$	3	3	76	0.0156	$2.4296e-012$	$2.0169e+003$
$P2$	6	6	489	0.5938	$3.5505e-005$	$1.1425e+005$
$P3$	3	15	24	0.0156	$1.1279e-008$	$5.7233e+004$
$P4$	2	2	37	0.0313	$3.3789e-007$	$1.8238e+000$
$P5$	3	10	32	0.0155	$4.0396e-013$	$1.4709e+005$
$P5$	3	20	22	0.0313	$1.3365e-017$	$1.3774e+006$
$P6$	5	7	11	0.0153	$1.1030e-010$	$5.2633e+004$
$P6$	10	12	14	0.0156	$1.0326e-009$	$1.7623e+005$
$P6$	20	22	16	0.0313	$1.8704e-003$	$4.6426e+006$
$P6$	30	32	17	0.0155	$5.5257e-002$	$1.3659e+006$
$P7$	2	31	9	0.0153	$5.4661e-001$	$4.3214e+005$
$P7$	6	31	22	0.1875	$2.2877e-003$	$3.9246e+006$
$P7$	8	31	25	0.4219	$1.8185e-005$	$2.8944e+006$
$P7$	9	31	21	0.5000	$2.7859e-006$	$1.2264e+006$
$P8$	4	5	20	0.0312	$2.2501e-005$	$1.3018e+005$
$P8$	10	11	13	0.0156	$7.4418e-005$	$2.9452e+005$
$P8$	20	21	15	0.0313	$1.6349e-004$	$4.4251e+006$
$P8$	50	51	16	0.0156	$1.7070e-002$	$3.5939e+006$
$P8$	100	101	17	0.4375	$4.5525e-001$	$1.1565e+006$
$P8$	200	201	17	0.1250	$3.7352e+001$	$1.1867e+006$
$P9$	4	8	39	0.0313	$9.3763e-006$	$3.1829e+004$
$P9$	10	20	34	0.0938	$2.9366e-004$	$1.3461e+004$
$P9$	20	40	32	0.1250	$6.3897e-003$	$4.3641e+004$
$P9$	50	100	22	0.0625	$4.2961e-000$	$2.6661e+005$
$P9$	100	200	19	0.0938	$9.7096e+004$	$1.8365e+006$
$P9$	200	400	10	0.0938	$4.7116e+013$	$3.8783e+004$
$P10$	2	3	63	0.0313	$4.7304e-013$	$1.1728e+005$
$P11$	4	10	85	0.0625	$1.4433e-000$	$1.0518e+006$
$P11$	4	20	17	0.0156	$8.5822e+004$	$1.4095e+006$
$P11$	4	50	12	0.0313	$2.6684e+016$	$1.5367e+006$
$P11$	4	100	12	0.0313	$1.5087e+034$	$3.3363e+004$
$P12$	3	3	$fail$	$fail$	$fail$	$fail$
$P13$	5	5	792	0.2656	$4.1105e-017$	$2.3344e+006$
$P13$	10	10	904	0.5938	$2.7951e-005$	$3.2390e+006$
$P14$	2	2	21	0.0156	$9.4629e-019$	$1.2025e+004$
$P14$	10	10	21	0.0156	$4.7314e-018$	$1.2025e+004$
$P14$	20	20	21	0.0153	$9.4629e-018$	$1.2025e+004$
$P14$	50	50	21	0.0313	$2.3657e-017$	$1.2025e+004$
$P14$	100	100	21	0.0625	$4.7314e-017$	$1.2025e+004$
$P14$	200	200	21	0.3125	$9.4629e-017$	$1.2025e+004$
$P14$	400	400	21	1.2031	$1.8937e-016$	$1.2025e+004$
$P15$	4	4	18	0.0313	$2.1577e-009$	$2.0004e+006$
$P15$	20	20	18	0.0154	$1.0789e-008$	$2.0004e+006$
$P15$	40	40	18	0.0156	$2.1577e-008$	$2.0004e+006$
$P15$	100	100	18	0.0625	$5.3944e-008$	$2.0004e+006$
$P15$	200	200	18	0.1406	$1.0789e-007$	$2.0004e+006$
$P16$	2	3	19	0.0156	$7.2047e-019$	$2.7779e+005$
$P17$	4	6	48	0.0313	$8.2639e-016$	$1.6817e+005$

Since the Ψtc method for solving (1.3) already adopts the Hessian of $f(x)$, there is no direct application of Ψtc to the ODE (2.3). However, we can apply Ψtc partially to solve (2.3). Our test for solving (2.3) is to adopt Newton's direction if $\lambda_{\min}(x) > \delta_2$, otherwise we adopt the Ψtc direction. The numerical results of this combined method are reported in Table 5 and Table 6, where $Iter$, f^*, ξ_i^* share the same meanings as Table 3 and Table 4; $\lambda_{\min}^*$ denotes the final computed $\lambda_{\min}(x)$. We set $\delta_1 = 1e - 7$ and $\delta_2 = 1e - 4$ in (2.4).

Table 5. Numerical results of the combined method with $\xi_0 = 1e-1$

$No.$	n	m	Iter	CPU(s)	f^*	$\lambda^*_{\min}$	ξ^*_i
$P1$	3	3	90	0.0625	$1.0225e-014$	$1.4328e-000$	$1.4013e+005$
$P2$	6	6	78	0.1250	$3.5505e-005$	$-4.4169e-005$	$9.7840e+006$
$P3$	3	15	3	0.0156	$1.1279e-008$	$1.3966e-001$	$2.6052e+003$
$P4$	2	2	36	0.0625	$5.0082e-008$	$5.7972e-005$	$7.3800e+002$
$P5$	3	10	45	0.0313	$7.5602e-002$	$-5.3429e-010$	$6.3309e+005$
$P5$	3	20	$fail$	$fail$	$fail$	$fail$	$fail$
$P6$	5	7	11	0.0156	$9.7541e-011$	$2.0000e-000$	$5.1672e+005$
$P6$	10	12	14	0.0152	$1.0314e-009$	$2.0000e-000$	$1.7620e+006$
$P6$	20	22	16	0.0313	$1.9155e-003$	$2.0000e-000$	$4.6426e+007$
$P6$	30	32	17	0.0156	$5.5257e-002$	$2.0000e-000$	$1.3659e+007$
$P7$	2	31	6	0.0156	$5.4661e-001$	$2.3977e+001$	$1.7146e+007$
$P7$	6	31	13	0.1406	$2.2877e-003$	$2.8101e-003$	$3.3248e+007$
$P7$	8	31	18	0.3281	$1.8162e-005$	$7.5430e-006$	$3.0671e+007$
$P7$	9	31	17	0.4375	$1.4375e-006$	$3.1599e-007$	$1.0825e+007$
$P8$	4	5	17	0.0156	$2.2513e-005$	$1.0022e-003$	$5.1724e+006$
$P8$	10	11	13	0.0156	$7.4402e-005$	$1.3945e-002$	$2.3004e+006$
$P8$	20	21	15	0.0154	$1.6347e-004$	$5.1832e-002$	$4.3120e+007$
$P8$	50	51	16	0.0313	$1.7043e-002$	$1.5880e-000$	$3.5905e+007$
$P8$	100	101	17	0.1250	$4.5525e-001$	$6.6031e-000$	$1.1564e+007$
$P8$	200	201	17	0.2813	$3.7352e+001$	$5.5580e+001$	$1.1867e+007$
$P9$	4	8	28	0.0155	$9.3765e-006$	$6.2659e-004$	$3.9608e+004$
$P9$	10	20	29	0.0313	$2.9366e-004$	$2.1416e-003$	$6.2639e+005$
$P9$	20	40	34	0.0625	$6.402e-003$	$2.5972e-004$	$2.0886e+006$
$P9$	50	100	22	0.0938	$4.2961e-000$	$1.7843e-002$	$2.6228e+006$
$P9$	100	200	19	0.1250	$9.7096e+004$	$2.2412e-001$	$1.8434e+007$
$P9$	200	400	10	0.2188	$4.7116e+013$	$2.6924e+002$	$3.8677e+005$
$P10$	2	3	5	0.0156	$9.8341e-010$	$2.0000e-000$	$5.6000e+000$
$P11$	4	10	85	0.0625	$1.4433e-000$	$4.7750e-000$	$1.0525e+007$
$P11$	4	20	17	0.0155	$8.5822e+004$	$1.5158e+003$	$1.4150e+007$
$P11$	4	50	12	0.0154	$2.6684e+016$	$1.4581e+009$	$1.5367e+007$
$P11$	4	100	12	0.0313	$1.5087e+034$	$1.5197e+018$	$3.3363e+005$
$P12$	3	3	2	0.0153	$1.4000e-003$	$-9.4304e-000$	$1.0000e-001$
$P13$	5	5	653	0.2656	$5.0235e-017$	$2.3764e-001$	$2.2897e+007$
$P13$	10	10	644	0.5000	$2.7951e-005$	$9.8102e-001$	$3.2449e+007$
$P14$	2	2	7	0.0155	$6.8653e-020$	$3.9944e-001$	$3.9929e+007$
$P14$	10	10	7	0.0154	$3.4326e-019$	$3.9944e-001$	$3.9929e+007$
$P14$	20	20	7	0.0156	$6.8653e-019$	$3.9944e-001$	$3.9929e+007$
$P14$	50	50	7	0.0625	$1.7163e-018$	$3.9944e-001$	$3.9929e+007$
$P14$	100	100	7	0.1250	$3.4326e-018$	$3.9944e-001$	$3.9929e+007$
$P14$	200	200	7	0.4219	$6.8653e-018$	$3.9944e-001$	$3.9929e+007$
$P14$	400	400	7	2.7031	$1.3731e-017$	$3.9944e-001$	$3.9929e+007$
$P15$	4	4	17	0.0156	$1.7193e-009$	$9.0837e-005$	$1.2294e+007$
$P15$	20	20	17	0.0313	$8.5966e-009$	$9.0837e-005$	$1.2294e+007$
$P15$	40	40	17	0.0313	$1.7193e-008$	$9.0837e-005$	$1.2294e+007$
$P15$	100	100	17	0.1094	$4.2983e-008$	$9.0837e-005$	$1.2294e+007$
$P15$	200	200	17	0.2813	$8.5966e-008$	$9.0837e-005$	$1.2294e+007$
$P16$	2	3	$fail$	$fail$	$fail$	$fail$	$fail$
$P17$	4	6	14	0.0313	$7.8770e-000$	$-1.1943e-001$	$6.7781e+007$

Table 6. Numerical results of the combined method with $\xi_0 = 1e - 2$

$No.$	n	m	Iter	CPU(s)	f^*	$\lambda^*_{\min}$	ξ^*_i
$P1$	3	3	90	0.0625	$1.0225e-014$	$1.4328e-000$	$1.4013e+005$
$P2$	6	6	78	0.1250	$3.5505e-005$	$-4.4169e-005$	$9.7840e+006$
$P3$	3	15	3	0.0155	$1.1279e-008$	$1.3966e-001$	$2.6052e+003$
$P4$	2	2	36	0.0938	$8.6100e-009$	$4.9937e-005$	$7.6421e+002$
$P5$	3	10	46	0.0313	$7.5602e-002$	$-8.3937e-012$	$2.6963e+005$
$P5$	3	20	145	0.1094	$9.5334e-002$	$-8.8936e-008$	$7.9937e+006$
$P6$	5	7	11	0.0153	$9.7541e-011$	$2.0000e-000$	$5.1672e+005$
$P6$	10	12	14	0.0156	$1.0314e-009$	$2.0000e-000$	$1.7620e+006$
$P6$	20	22	16	0.0313	$1.9155e-003$	$2.0000e-000$	$4.6426e+007$
$P6$	30	32	17	0.0156	$5.5257e-002$	$2.0000e-000$	$1.3659e+007$
$P7$	2	31	6	0.0156	$5.4661e-001$	$2.3977e+001$	$1.7146e+007$
$P7$	6	31	16	0.1406	$2.2877e-003$	$2.8101e-003$	$1.7110e+006$
$P7$	8	31	18	0.3281	$1.8162e-005$	$7.5430e-006$	$3.0671e+007$
$P7$	9	31	17	0.4375	$1.4375e-006$	$3.1599e-007$	$1.0825e+007$
$P8$	4	5	17	0.0156	$2.2513e-005$	$1.0022e-003$	$5.1724e+006$
$P8$	10	11	13	0.0156	$7.4402e-005$	$1.3945e-002$	$2.3004e+006$
$P8$	20	21	15	0.0153	$1.6347e-004$	$5.1832e-002$	$4.3120e+007$
$P8$	50	51	16	0.0313	$1.7043e-002$	$1.5880e-000$	$3.5905e+007$
$P8$	100	101	17	0.1250	$4.5525e-001$	$6.6031e-000$	$1.1564e+007$
$P8$	200	201	17	0.2813	$3.7352e+001$	$5.5580e+001$	$1.1867e+007$
$P9$	4	8	29	0.0156	$9.3763e-006$	$3.9279e-005$	$2.3035e+005$
$P9$	10	20	29	0.0313	$2.9366e-004$	$2.1416e-003$	$6.2639e+005$
$P9$	20	40	34	0.0625	$6.4022e-003$	$2.0922e-004$	$1.2226e+006$
$P9$	50	100	22	0.0938	$4.2961e-000$	$1.7843e-002$	$2.6228e+006$
$P9$	100	200	19	0.1250	$9.7096e+004$	$2.2412e-001$	$1.8434e+007$
$P9$	200	400	10	0.2188	$4.7116e+013$	$2.6924e+002$	$3.8677e+005$
$P10$	2	3	5	0.0150	$9.8341e-010$	$2.0000e-000$	$5.6000e+000$
$P11$	4	10	85	0.0625	$1.4433e-000$	$4.7750e-000$	$1.0525e+007$
$P11$	4	20	17	0.0156	$8.5822e+004$	$1.5158e+003$	$1.4150e+007$
$P11$	4	50	12	0.0156	$2.6684e+016$	$1.4581e+009$	$1.5367e+007$
$P11$	4	100	12	0.0313	$1.5087e+034$	$1.5197e+018$	$3.3363e+005$
$P12$	3	3	2	0.0156	$1.4000e-003$	$-9.4304e-000$	$1.0000e-001$
$P13$	5	5	684	0.2813	$5.1161e-017$	$2.3764e-001$	$2.3107e+007$
$P13$	10	10	644	0.5000	$2.7951e-005$	$9.8102e-001$	$3.2449e+007$
$P14$	2	2	7	0.0156	$6.8653e-020$	$3.9944e-001$	$3.9929e+007$
$P14$	10	10	7	0.0151	$3.4326e-019$	$3.9944e-001$	$3.9929e+007$
$P14$	20	20	7	0.0151	$6.8653e-019$	$3.9944e-001$	$3.9929e+007$
$P14$	50	50	7	0.0625	$1.7163e-018$	$3.9944e-001$	$3.9929e+007$
$P14$	100	100	7	0.1250	$3.4326e-018$	$3.9944e-001$	$3.9929e+007$
$P14$	200	200	7	0.4219	$6.8653e-018$	$3.9944e-001$	$3.9929e+007$
$P14$	400	400	7	2.7031	$1.3731e-017$	$3.9944e-001$	$3.9929e+007$
$P15$	4	4	17	0.0156	$1.7231e-009$	$9.6064e-005$	$1.2314e+007$
$P15$	20	20	17	0.0153	$8.6154e-009$	$9.6064e-005$	$1.2314e+007$
$P15$	40	40	17	0.0313	$1.7231e-008$	$9.6064e-005$	$1.2314e+007$
$P15$	100	100	17	0.1094	$4.3077e-008$	$9.6064e-005$	$1.2314e+007$
$P15$	200	200	17	0.2500	$8.6154e-008$	$9.6064e-005$	$1.2314e+007$
$P16$	2	3	$fail$	$fail$	$fail$	$fail$	$fail$
$P17$	4	6	14	0.0313	$7.8770e-000$	$-1.1943e-001$	$6.7781e+007$

Compared with the results of Table 3 and Table 4, we can see that the combined method works well. However a specially designated Ψtc method for (2.3) can be expected to have better performance, but this is beyond the scope of this paper.

6 Concluding Remarks

By combining Newton's direction and the steepest descent direction, a new continuous Newton-type method, whose ODE is given by (2.3), is proposed in this paper. The direction

$$d(x) = \alpha(x)d_N(x) + (1 - \alpha(x))d_G(x)$$

of the new ODE (2.3) is actually the convex combination of Newton's direction $d_N(x)$ and the negative gradient direction $d_G(x)$. The convergence of this ODE is fully addressed in Section 3. Our numerical results reported in Section 5 clearly illustrate that our new method works well numerically. However, we should point out that the optimal choice of the parameters δ_1 and δ_2 in (2.5) is somehow problem dependent, which can be seen from the numerical results of problems $P5$ and $P17$ in Table 2. Theoretically, (2.3) prefers small δ_1 and δ_2, and the smaller values of δ_1 and δ_2 are, the closer trajectory of the proposed method to the continuous Newton trajectory, and hence the faster convergence. However, if δ_1 and δ_2 are too small, it could cause numerical difficulties and instability. From our preliminary numerical results, the values in the examples appear to be proper and they seem to work well in practice. Even though the Ψtc method cannot be applied directly to solve (2.3), yet a partial implementation of the Ψtc on the new ODE (2.3) also works well as shown in Table 5 and Table 6.

References

[1] P.-A. Absil, R. Mahony, B. Andrews, Convergence of the iterates of descent methods for analytic cost functions, *SIAM J. Optim.* 16 (2005) 531–547.

[2] R.G. Airapetyan and A.G. Ramm, *Dynamical systems and discrete methods for solving nonlinear ill-posed problems*, Appl. Math. Reviews, 1 (2000), Ed. G. Anastassiou, World Sci. Publishers, pp. 491–536.

[3] F. Alvarez D and J. M. Pérez C, A dynamical system associated with Newton's method for parametric approximations of convex minimization problems, *Appl. Math. Optim.* 38 (1998) 193–217.

[4] N. Andrei, Gradient flow algorithm for unconstrained optimization, ICI Technical Report, April, 2004.

[5] C.A. Botsaris, Differential gradient methods, *J. Math. Anal. Appl.* 63 (1978) 177–198.

[6] F.H. Branin, Jr., A widely convergent method for finding multiple solutions of simultaneous nonlinear equations, *IBM J. Res. Develop.* 16 (1972) 504–522.

[7] M. Braun, *Differential Equations and their Applications: an Introduction to Applied Mathematics*, Springer-Verlag, New York, 1993.

[8] A.A. Brown and M.C. Bartholomew-Biggs, Some effective methods for unconstrained optimization based on the solution of systems of ordinary differential equations, *J. Optim Theory and Appl.* 62 (1988) 211–224.

[9] T. Coffey, C.T. Kelley, and D.E. Keyes, Pseudo-transient continuation and differential-algebraic equations, *SIAM J. Sci. Comp.* 25 (2003) 553–569.

[10] R. Courant, Variational methods for the solution of problems of equilibrium and vibration, *Bull. Amer. Math. Soc.* 49 (1943) 1–43.

[11] P. Deuflhard, Adaptive pseudo-transient continuation for nonlinear steady state problems, Tech. Rep. 02-14, Konrad-Zuse-Zentrum für Informationstechnik, Berlin, March 2002.

[12] I. Diener, On the global convergence of path-following methods to determine all solutions to a system of nonlinear equations, *Math. Program.* 39 (1987) 181–188.

[13] I. Diener, Trajectory nets connecting all critical points of a smooth function, *Math. Program.* 36 (1986) 340–352.

[14] I. Diener and R. Schaback, An extended continuous Newton method, *J. Optim. Theory and Appl.* 67 (1990) 57–77.

[15] K.R. Fowler and C.T. Kelley, Pseudo-transient continuation for nonsmooth nonlinear equations, *SIAM J. Numer. Anal.* 43 (2005) 1385–1406.

[16] G.H. Golub and C.F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, third edition, 1996.

[17] Q.-M. Han, L.-Z. Liao, H.D. Qi, and L.Q. Qi, Stability analysis of gradient-based neural networks for optimization problems, *J. Global Optim.* 19 (2001) 363–381.

[18] R. Hauser and J. Nedić, The continuous Newton-Raphson method can look ahead, *SIAM J. Optim.* 15 (2005) 915–925.

[19] U. Helmke and J. B. Moore, *Optimization and Dynamical Systems*, Springer-Verlag, London, UK, 1994.

[20] D.J. Higham, Trust region algorithms and time step selection, *SIAM J. Numer. Anal.* 37 (1999) 194–210.

[21] V. Janovsky and V. Seige, Qualitative analysis of Newton's flow, *SIAM J. Numer. Anal.* 33 (1996) 2068–2097.

[22] H.TH. Jongen, P. Jonker and F. Twilt, A note on Branin's method for finding the critical points of smooth functions, in *Parametric Optimization and Related Topics, 1987*, Int. Conf., Plaue/GDR, 1985, pp. 209–228.

[23] H.TH. Jongen, P. Jonker and F. Twilt, *Nonlinear Optimization in* R^n*, volume II of Methoden und Verfahren der mathematischen Physik*, Bd 32. Peter Lang Verlag, Frankfurt a.M., 1986.

[24] C.T. Kelley and D.E. Keyes, Convergence analysis of pseudo-transient continuation, *SIAM J. Numer. Anal.* 35 (1998) 508–523.

[25] L.-Z. Liao, H.D. Qi, and L.Q. Qi, Neurodynamical optimization, *J. Global Optim.* 28 (2004) 175–195.

[26] L.-Z. Liao, L.Q. Qi, and H.W. Tam, A gradient-based continuous method for large-scale optimization problems, *J. Global Optim.* 31 (2005) 271–286.

[27] S. Łojasiewicz, *Ensembles semi-analytiques*, Inst. Hautes Études Sci., Bures-sur-Yvette, France, 1965.

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

[29] W. Mulder and B.V. Leer, Experiments with implicit upwind methods for the Euler equations, *J. Comp. Phys.* 59 (1985), 232–246.

[30] J.M. Ortega and W.C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

[31] L. Perko, *Differential Equations and Dynamical Systems*, Springer-Verlag, New York, 1991.

[32] A.G. Ramm, Linear ill-posed problems and dynamical systems, *J. Math. Anal. Appl.* 258 (2001) 448–456.

[33] R.B. Schnabel and E. Eskow, A new modified Cholesky factorization, *SIAM J. Sci. Stat. Comput.* 11 (1990) 1136–1158.

[34] L.F. Shampine and M.W. Reichelt, The MATLAB ODE Suite, *SIAM J. Scientific Comput.* 18 (1997) 1–22.

[35] J.J.E. Slotine and W. Li, *Applied Nonlinear Control*, Prentice Hall, New Jersey, 1991.

[36] S. Smale, A convergent process of price adjustment and global Newton methods, *J. Math. Economics* 3 (1976) 107–120.

[37] J.A. Sturua and S.K. Zavriev, A trajectory algorithm based on the gradient method I. The search on the quasioptimal trajectories, *J. Global Optim.* 4 (1991) 375–388.

[38] K. Tanabe, Differential geometric methods in nonlinear programming, Brookhaven Laboratory, New York, Technical Report 26730-AMD831, *Math. Software* 10 (1979) 200–316.

Manuscript received 13 July 2007
revised 5 September 2007
accepted for publication 28 September 2007

Lei-Hong Zhang
Department of Mathematics, Hong Kong Baptist University
Kowloon Tong, Kowloon, Hong Kong, P. R. China
E-mail address: lhzhang@math.hkbu.edu.hk

C.T. Kelley
Department of Mathematics, North Carolina State University
Box 8205, Raleigh, N. C. 27695-8205, USA
E-mail address: Tim_Kelley@ncsu.edu

Li-Zhi Liao
Department of Mathematics, Hong Kong Baptist University
Kowloon Tong, Kowloon, Hong Kong, P. R. China
E-mail address: liliao@hkbu.edu.hk

Appendix:

Proof of Theorem 2.2. We first prove $d(x)$ defined by (2.4) is locally Lipschitz continuous in R^n. For any $\bar{x} \in R^n$ with $\lambda_{\min}(\bar{x}) \neq \delta_1$ and $\lambda_{\min}(\bar{x}) \neq \delta_2$, the local Lipschitz continuity of $d(x)$ at $\bar{x}$ can be immediately obtained from (2.7), (2.8), and (2.10). Suppose now $\lambda_{\min}(\bar{x}) = \delta_1$. Let $\rho > 0$ be sufficiently small so that (2.7), (2.8), (2.9), and (2.10) hold in the neighborhood $N_\rho(\bar{x})$ of $\bar{x}$, and moreover, for any $y \in N_\rho(\bar{x})$, $\delta_2 > \lambda_{\min}(y)$. Consider any two points $y, z \in N_\rho(\bar{x})$, if

$$(\lambda_{\min}(y) - \delta_1)(\lambda_{\min}(z) - \delta_1) \geq 0,$$

it then follows

$$\|d(y) - d(z)\|_2 \leq C\|y - z\|_2$$

for some constant $C > 0$. Otherwise, we assume $\delta_2 > \lambda_{\min}(y) > \delta_1 > \lambda_{\min}(z)$, and hence

$$\begin{aligned}
\|d(y) - d(z)\|_2 &= \|\alpha(y)d_N(y) + \beta(y)d_G(y) - \alpha(z)d_N(z) - \beta(z)d_G(z)\|_2 \\
&\leq |\alpha(y) - \alpha(z)| \cdot \|d_N(y)\|_2 + \alpha(z)\|d_N(y) - d_N(z)\|_2 \\
&+ |\beta(y) - \beta(z)| \cdot \|d_G(y)\|_2 + \beta(z)\|d_G(y) - d_G(z)\|_2 \\
&\leq \frac{\lambda_{\min}(y) - \lambda_{\min}(z)}{\delta_2 - \delta_1} \cdot \|d_N(y)\|_2 + \|d_N(y) - d_N(z)\|_2 \\
&+ \frac{\lambda_{\min}(y) - \lambda_{\min}(z)}{\delta_2 - \delta_1} \cdot \|d_G(y)\|_2 + \|d_G(y) - d_G(z)\|_2 \\
&\leq (\frac{C_1\|d_N(y)\|_2}{\delta_2 - \delta_1} + \frac{C_1\|d_G(y)\|_2}{\delta_2 - \delta_1} + C_2 + C_3) \cdot \|y - z\|_2 \\
&\leq C_4\|y - z\|_2
\end{aligned}$$

for some $C_4 > 0$. This leads to the local Lipschitz condition of $d(x)$ at $\bar{x}$.

With the same arguments as the case $\lambda_{\min}(\bar{x}) = \delta_2$, we finally show that $d(x)$ is locally Lipschitz continuous in R^n, from which the existence and uniqueness of the solution of (2.3) are obtained by the Picard-Lindelöf theorem.

Furthermore,

$$\begin{aligned}
\frac{df(x(t))}{dt} &= \nabla f(x)^T d(x) \\
&= \begin{cases} -d_N(x)^T d_G(x), & \text{if } \lambda_{\min}(x) > \delta_2, \\ -\frac{\lambda_{\min}(x) - \delta_1}{\delta_2 - \delta_1} d_N(x)^T d_G(x) - \frac{\delta_2 - \lambda_{\min}(x)}{\delta_2 - \delta_1}\|d_G(x)\|_2^2, & \text{if } \delta_1 \leq \lambda_{\min}(x) \leq \delta_2, \\ -\|d_G(x)\|_2^2, & \text{if } \lambda_{\min}(x) < \delta_1, \end{cases}
\end{aligned} \tag{6.1}$$

which implies that $\frac{df(x(t))}{dt} \leq 0$ and $f(x(t))$ is nonincreasing along the trajectory $x(t)$ for $t \geq 0$. Therefore, the solution $x(t)$ will always stay in the compact set $L_{f(x_0)}$, and the maximal interval of existence can be extended to $[0, +\infty)$. □

Proof of Theorem 2.4. Suppose there is a periodic solution $x(t)$ with its minimal period $\hat{T} > 0$, then $f(x(t+\hat{T})) = f(x(t))$, for $t \geq 0$, which just contradicts the fact that $\frac{df(x(t))}{dt} < 0$ for any $t \geq 0$ (by Theorem 2.3). This completes the proof. □