A continuous Newton-type method for unconstrained optimization

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

C. T. Kelley
Department of Mathematics, North Carolina State University, Box 8205, Raleigh, N. C. 27695-8205, USA
email: Tim_Kelley@ncsu.edu http://www.math.ncsu.edu/~ctk/

Li-Zhi Liao
Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Kowloon, Hong Kong, P. R. China
email: liliao@hkbu.edu.hk http://www.math.hkbu.edu.hk/~liliao/

In this paper, we propose a continuous Newton-type method in the form of an ordinary differential equation by combining the negative gradient and Newton's direction. It is shown 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 positive definite. The convergence of the new method on the set of standard test problems in the literature are also reported.

Key words: Unconstrained optimization; singularity; continuous Newton method; continuous method; stationary point; global convergence; real analytic function; pseudo-transient continuation

MSC2000 Subject Classification: Primary: 90C30; Secondary: 65K05

OR/MS subject classification: Primary: Programming/Nonlinear/Unconstrained; Secondary: Programming/Nonlinear/Algorithms

---

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

$$\min_{x\in R^n} f(x) \tag{1}$$

by using the so-called *trajectory method* or *continuous method* (see [7], [11], [12], [17], [23], [30], [36] and the references therein). Different from the conventional optimization methods, these methods adopt some kind of differential equations to define the trajectory of variable x in terms of a parameter t. By tracing this trajectory, the *stationary points* satisfying $\nabla f(x) = 0$, or hopefully, the local minima 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:

$$\begin{cases} \frac{dx}{dt} &= h(x), \quad t \ge t_0 \\ x(t_0) &= x_0, \end{cases} \tag{2}$$

where $h : U \subseteq R^n \to R^n$, and T denotes the maximal interval of existence. The curve in R^n, $\{x|x = x(t),\ t \in T\}$, is said to be the trajectory of the ordinary differential equation (ODE) (2). Without confusion, in order to simplify the following presentation, we also call $x(t)$ the trajectory of (2).

Obviously, the simplest trajectory is the one defined by the ordinary differential equation with $h(x) = -\nabla f(x)$, i.e.

$$\begin{cases} \frac{dx}{dt} &= -\nabla f(x), \\ x(t_0) &= x_0, \end{cases} \tag{3}$$

which goes back to A. Cauchy and was proposed to solve some optimization problems in [9]. This ode system has been studied extensively in [1], [2], [3], [16], and [35].

Another natural trajectory is generated by the continuous Newton's direction given by

$$\begin{cases} \frac{dx}{dt} &= -(\nabla^2 f(x))^{-1}\nabla f(x), \\ x(t_0) &= x_0. \end{cases} \tag{4}$$

However, the singularity of the Hessian matrix $\nabla^2 f(x)$ is the major obstacle for this method. In [4], Branin considered the following corresponding form

$$\nabla^2 f(x)\frac{dx}{dt} = \mp\nabla f(x), \tag{5}$$

and to change the sign of (5) whenever the trajectory $x(t)$ generated encounters a change in sign of the determinant of $\nabla^2 f(x(t))$ or arrives at a solution point of $\nabla f(x) = 0$ for finding multiple local minima. Moreover, Branin also suggested to employ the adjoint matrix, say $A(x)$, of $\nabla^2 f(x(t))$ to get around the singularity, and then replace (5) with

$$\frac{dx}{dt} = -A(x)\nabla f(x), \tag{6}$$

which is now well-defined in R^n. However, the consequence of adopting (6), the troublesome *extraneous stationary points* (or *extraneous singular points*, [4]) defined by $\{\hat{x}|A(\hat{x})\nabla f(\hat{x}) = 0,\ \nabla f(\hat{x}) \neq 0\}$ are induced (see [19] and [20] for the structure of such extraneous singular points).

An analogous modification of (6) proposed by Smale ([34], 1976) is called "global Newton equation", and has the following form in the context of the unconstrained optimization

$$\nabla^2 f(x)\frac{dx}{dt} = -\phi(x)\nabla f(x), \tag{7}$$

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 (6) immediately.

Additional research on the extended continuous Newton methods has been carried out. 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, readers can refer to [11], [12], [13].

In this paper, we propose a continuous Newton-type method (in the form of an ordinary differential equation), which combines the negative gradient and Newton's direction, and is well-defined in R^n. It is shown that our method gets around the singularities of $\nabla^2 f(x)$ and converges globally to a connected regular stationary points (points satisfying $\nabla f(x) = 0$) subset for a general function $f(x)$, and converges globally to a regular stationary point for a real analytic function[1] $f(x)$. Moreover, the trajectory defined by the proposed ordinary differential equation moves strictly downhill (meaning that the value of $f(x(t))$ is strictly decreasing as t increases); and for some convex function $f(x)$, it becomes the exact Newton trajectory of (4), and therefore, the fast convergence can be achieved.

The rest of this paper is organized as follows. In the next section, the ordinary differential equation corresponding to our continuous Newton-type method is established and the existence and the 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 examed 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 globally convergent continuous Newton-type method First, let's state some assumptions on the problem that we are interested in. Let

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

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

Assumptions:

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

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

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

[1]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.

It is clear that Assumption (c) is much weaker than the condition that the level set $L = \{x|f(x) \leq f(x_0)\}$ is bounded. For example, if $f(x) = \sin x,\ x \in R$, then the level set $L = \{x|f(x) = \sin x \leq \frac{\sqrt{2}}{2}\}$ is unbounded, but $L_{f(x_0)}$ is bounded for any given x_0. From Assumption (c), we know 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

$$S_{f(x_0)} := S \cap L_{f(x_0)}, \tag{8}$$

is compact too, where S is the the stationary points set given by

$$S := \{x|\nabla f(x) = 0\}. \tag{9}$$

Consider the following continuous Newton-type differential equation,

$$\begin{cases} \frac{dx}{dt} & = d(x), \\ x(t_0) & = x_0, \end{cases} \tag{10}$$

where

$$d(x) = \begin{cases} -(\nabla^2 f(x))^{-1}\nabla f(x), & \text{if } \lambda_{\min}(x) > \delta_2, \\ -\alpha(x)(\nabla^2 f(x))^{-1}\nabla f(x) - \beta(x)\nabla f(x), & \text{if } \delta_1 \leq \lambda_{\min}(x) \leq \delta_2, \\ -\nabla f(x), & \text{if } \lambda_{\min}(x) < \delta_1, \end{cases} \tag{11}$$

where $\lambda_{\min}(x)$ represents the smallest eigenvalue of $\nabla^2 f(x)$, $\delta_2 > \delta_1 > 0$ are two predefined positive constants, and $\alpha(x)$, $\beta(x)$ are set as

$$\alpha(x) = \frac{\lambda_{\min}(x) - \delta_1}{\delta_2 - \delta_1}, \tag{12}$$

$$\beta(x) = 1 - \alpha(x) = \frac{\delta_2 - \lambda_{\min}(x)}{\delta_2 - \delta_1}. \tag{13}$$

The smallest eigenvalue of $\nabla^2 f(x)$, $\lambda_{min}(x)$, can be easily estimated from the modified Cholesky factorization in [31]. However, for convenience, we still use the MATLAB function **eig.m** to compute the $\lambda_{min}(x)$ in our numerical test in Section 4.

For simplicity of our presentation, we will use $d_N(x)$ and $d_G(x)$ to denote Newton's direction $-(\nabla^2 f(x))^{-1}\nabla f(x)$ and the negative gradient $-\nabla f(x)$ at point x, respectively. A first observation is that when $\nabla^2 f(x)$ is strictly positive and $\delta_2 > \delta_1 > 0$ are chosen properly, the trajectory generated is exactly the continuous Newton trajectory. Furthermore, (11) is well-defined in R^n, and the existence and the uniqueness of the solution to the initial value problem (10) can also be guaranteed as long as $d(x)$ is proved to be locally Lipschitz continuous in R^n. In order to get this result, we first show that $\lambda_{\min}(x)$ is locally Lipschitz continuous in R^n, which is a direct result of the Wielandt-Hoffman Theorem.

Lemma 2.1 *([15], 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 immediately leads to the fact that $\lambda_{\min}(x)$ is locally Lipschitz continuous in R^n. Moreover, from the result in [28] (Section 2.3.3, p. 46), we know that for any $\bar{x}$, if $\lambda_{\min}(\bar{x}) > 0$, there exist a $\gamma > 0$ and a neighborhood $N_\tau(\bar{x})$ of $\bar{x}$ such that $\forall x \in N_\tau(\bar{x})$, $\nabla^2 f(x)$ is invertible and $\|(\nabla^2 f(x))^{-1}\|_2 \leq \gamma$. Hence,

$$\begin{aligned} \|(\nabla^2 f(\bar{x}))^{-1} - (\nabla^2 f(x))^{-1}\|_2 &= \|(\nabla^2 f(\bar{x}))^{-1}[\nabla^2 f(x) - \nabla^2 f(\bar{x})](\nabla^2 f(x))^{-1}\|_2 \\ &\leq \gamma\|(\nabla^2 f(\bar{x}))^{-1}\|_2 \cdot \|\nabla^2 f(x) - \nabla^2 f(\bar{x})\|_2, \end{aligned}$$

which implies that $(\nabla^2 f(x))^{-1}$ is Lipschitz continuous at $\bar{x}$ too.

Theorem 2.1 *Suppose that $f(x)$ satisfies Assumptions (a), (b) and (c). Then for any $x(t_0) = x_0 \in R^n$, there exists a unique solution $x(t)$ to (10), and the maximal interval of existence of the solution can be extended to $[0, +\infty)$.*

PROOF. For any $\bar{x} \in R^n$ with $\lambda_{\min}(\bar{x}) \neq \delta_1$ or δ_2, the locally Lipschitz continuity of $d(x)$ at $\bar{x}$ can be immediately obtained from the previous discussion. If $\lambda_{\min}(\bar{x}) = \delta_1$, then $\alpha(\bar{x}) = 0$, $\beta(\bar{x}) = 1$ and $d(\bar{x}) = -\nabla f(\bar{x})$. For any x in some neighborhood of $\bar{x}$ with $\lambda_{\min}(x) < \delta_1$,

$$\|d(x) - d(\bar{x})\|_2 = \|\nabla f(\bar{x}) - \nabla f(x)\|_2 \leq L_1\|x - \bar{x}\|_2.$$

where $L_1 \geq 0$ is the corresponding Lipschitz constant. For any x in some neighborhood of $\bar{x}$ with $\lambda_{\min}(x) \geq \delta_1$, since $0 \leq \alpha(x), \beta(x) \leq 1$, and $\lambda_{\min}(x)$, $d_N(x)$ and $d_G(x)$ are locally Lipschitz continuous at $\bar{x}$, we have

$$\begin{aligned}
\|d(x) - d(\bar{x})\|_2 &= \|\alpha(x)d_N(x) + \beta(x)d_G(x) - d_G(\bar{x})\|_2 \\
&\leq \|d_N(\bar{x})\|_2 \cdot \alpha(x) + \alpha(x)\|d_N(x) - d_N(\bar{x})\|_2 \\
&\quad + \|d_G(\bar{x})\|_2 \cdot [1 - \beta(x)] + \beta(x)\|d_G(x) - d_G(\bar{x})\|_2 \\
&\leq \frac{1}{\delta_2 - \delta_1}\|d_N(\bar{x})\|_2 \cdot [\lambda_{\min}(x) - \lambda_{\min}(\bar{x})] + \alpha(x)\|d_N(x) - d_N(\bar{x})\|_2 \\
&\quad + \frac{1}{\delta_2 - \delta_1}\|d_G(\bar{x})\|_2 \cdot [\lambda_{\min}(x) - \lambda_{\min}(\bar{x})] + \beta(x)\|d_G(x) - d_G(\bar{x})\|_2 \\
&\leq L_2\|x - \bar{x}\|_2,
\end{aligned}$$

where $L_2 \geq 0$ is the corresponding Lipschitz constant. Similarly, $d(x)$ is also locally Lipschitz continuous at $\bar{x}$ when $\lambda_{\min}(\bar{x}) = \delta_2$; and therefore, $d(x)$ is locally Lipschitz continuous in R^n, from which the existence and the uniqueness of the solution of (10) are obtained by the Picard-Lindelöf theorem.

Furthermore, note that

$$\frac{df(x(t))}{dt} = \begin{cases} -d_N(x)^T d_G(x), & \text{if } \lambda_{\min}(x) > \delta_2, \\ -\alpha(x)d_N(x)^T d_G(x) - \beta(x)\|d_G(x)\|_2^2, & \text{if } \delta_1 \leq \lambda_{\min}(x) \leq \delta_2, \\ -\|d_G(x)\|_2^2, & \text{if } \lambda_{\min}(x) < \delta_1, \end{cases} \tag{14}$$

which implies that $\frac{df(x(t))}{dt} \leq 0$ (since $d_N^T d_G \geq 0$), and $f(x(t))$ is nonincreasing along the trajectory $x(t)$ for $t \geq t_0$. Therefore, it follows that the solution $x(t)$ will always stay in the compact set $S_{f(x_0)}$, and the maximal interval of existence of the solution can be extended to $[0, +\infty)$. □

We now provide a general result which shows that the trajectory $x(t)$ of (10) will never reach set $S_{f(x_0)}$ at finite time t if $\nabla f(x(t_0)) \neq 0$.

THEOREM 2.2 *Suppose $h : U \subseteq R^n \to R^n$ is locally Lipschitz continuous on an open set U. Then for any $x_0 = x(t_0) \in U$ with $h(x_0) \neq 0$, the solution to the initial problem (2) satisfies $h(x(t)) \neq 0$ for any $t \in T$, where T denotes the maximal interval of existence of $x(t)$.*

PROOF. Suppose by contradiction that $\bar{t} \in T$ is the smallest value satisfying $h(x(\bar{t})) = h(\bar{x}) = 0$ in the right maximal interval of existence of $x(t)$. Since $h(x)$ is locally Lipschitz continuous at $\bar{x} = x(\bar{t})$, there exist a neighborhood $N_\tau(\bar{x})$ of $\bar{x}$ and $L(\bar{x}) \geq 0$ such that

$$\|h(x) - h(\bar{x})\|_2 \leq L(\bar{x})\|x - \bar{x}\|_2, \quad \forall x \in N_\tau(\bar{x}).$$

Clearly, $L(\bar{x}) > 0$; otherwise, $\bar{t}$ must not be the smallest value satisfying $h(x(t)) = 0$.

Note that $x(t)$ is continuous, there exists an $\eta > 0$ such that $0 < L(\bar{x})\eta < 1$, $\eta < \bar{t} - t_0$, and $x(t) \in N_\tau(\bar{x})$ for all $t \in [\bar{t} - \eta, \bar{t}]$. Since for any $t \in [\bar{t} - \eta, \bar{t}]$, it holds that

$$\|h(x(t))\|_2 = \|h(x(t)) - h(\bar{x})\|_2 \leq L(\bar{x})\|x(t) - \bar{x}\|_2 \leq L(\bar{x})\eta \cdot \max_{s \in [\bar{t}-\eta, \bar{t}]} \|h(x(s))\|_2.$$

This together with $0 < L(\bar{x})\eta < 1$ implies

$$\max_{s \in [\bar{t}-\eta, \bar{t}]} \|h(x(s))\|_2 = 0,$$

which contradicts the fact that $\bar{t}$ is the smallest satisfying $h(x(t)) = 0$, and thereby, $h(x(t)) \neq 0$ for any $t \in T$. □

The result of Theorem 2.2 is the extension of Theorem 2(iii) in [24] which obtained the same result for the gradient system. (14) 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 (10).

THEOREM 2.3 *There is no periodic solution to (10) for any $x(t_0) = x_0 \in R^n$ with $\nabla f(x_0) \neq 0$.*

PROOF. 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 $t \geq t_0$, which just contradicts the fact that $\frac{df(x(t))}{dt} < 0$ for any $t \geq t_0$ (by Theorem 2.2). This completes the proof. □

3. Convergence analysis Since the maximal interval of existence of the solution to (10) can be extended to $[t_0, +\infty)$, we then can apply some results of the dynamical system to develop the convergence analysis.

DEFINITION 3.1 *A point $p \in U$ is an ω-limit point of the trajectory $x(t)$ of dynamical system $\frac{dx}{dt} = h(x)$ with $x(t_0) = x_0$ if there is a sequence $t_n \to +\infty$ such that*

$$\lim_{n \to +\infty} x(t_n) = p.$$

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

LEMMA 3.1 *(see [29]) The ω-limit set of a trajectory $x(t)$ of (2) is closed in U and if $x(t)$ is contained in a compact subset of R^n, then Ω_{x_0} is non-empty, connected and compact.*

REMARK 3.1 *Let Ω_{x_0} represent the ω-limit set of trajectory of (10). As indicated by (14), $\Omega_{x_0} \subseteq L_{f(x_0)}$; and moreover, from Definition 3.1, we can say that the trajectory $x(t)$ converges to the set Ω_{x_0} as $t \to +\infty$ in the sense that for any $\epsilon > 0$, there exists a $t_\epsilon \geq t_0$ such that $\forall t > t_\epsilon$, it follows that*

$$d(x(t), \Omega_{x_0}) = \inf_{\hat{x} \in \Omega_{x_0}} \|x(t) - \hat{x}\|_2 < \epsilon.$$

Ω_{x_0} is also said to be attracting for the trajectory $x(t)$. If, in addition, Ω_{x_0} contains only one point, "attracting" equivalently means convergence to a single point.

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

THEOREM 3.1 *Suppose $f(x)$ satisfy Assumptions (a), (b), and (c), and let $x(t)$ be the trajectory of (10) with $x(t_0) = x_0 \in R^n$. Then there exists some constant $\bar{f}$ such that*

$$\Omega_{x_0} \subseteq \{x | f(x) = \bar{f}\} \cap S_{f(x_0)}; \qquad (15)$$

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 (8).

PROOF. Since $\nabla f(x_0) = 0$ is the trivial case in which the unique trajectory becomes $x(t) \equiv x_0$, $t \geq t_0$ (due to uniqueness), we just consider $\nabla f(x_0) \neq 0$.

From (14) and Theorem 2.2, 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 $\Omega_{x_0} \subseteq \{x | f(x) = \bar{f}\}$ directly.

Furthermore, the LaSalle invariant set theorem (Theorem 3.4 in [33]) 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 (14). Therefore, consequently, from Lemma 3.1, Remark 3.1 and $x(t) \in L_{f(x_0)}$ for $t \geq t_0$, we conclude $\Omega_{x_0} \subseteq \{x | f(x) = \bar{f}\} \cap S_{f(x_0)}$, 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.

Corollary 3.1 *Under the conditions of Theorem 3.1, suppose that $x(t)$ is the trajectory of (10) with $x(t_0) = x_0 \in R^n$. 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 (3) will not necessarily converge to a single point (see [17], Prop. C.12.1; and see [1] for the counterexample and more general version). By only endowing $f(x)$ to be real analytic additionally, however, the single limit-point convergence is achievable. The proof for this is based on Corollary 3.1 and similar to the proof of Theorem 2.2 in [1].

Theorem 3.2 *Suppose that $f(x)$ is a real analytic function satisfying Assumption (a), (b), and (c). Then the trajectory $x(t)$ of (10) converges to a (single) stationary point of $f(x)$ as $t \to +\infty$ for any $x(t_0) = x_0 \in R^n$.*

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 $\exists$ 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 3.1 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 any predefined $\delta_1 > 0$ in (10). Also, since $f(x)$ is real analytic, the following Łojasiewicz gradient inequality (see [25]) 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 3.1 and $\nabla f(x_0) \neq 0$, it follows that $f(x(t)) > f(\bar{x})$ for $t \geq t_0$. 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,

$$c_1 \frac{d[f(x(t)) - f(\bar{x})]^{1-\sigma}}{dt} \leq -\|\frac{dx(t)}{dt}\|_2, \tag{16}$$

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 $t_1 \geq t_0$ 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 (16) 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, to obtain the single-limit convergence, we do not impose the angle condition

$$\frac{df(x(t))}{dt} \equiv \nabla f(x(t))^T \frac{dx(t)}{dt} \leq -\theta \|\nabla f(x(t))\|_2 \cdot \|\frac{dx(t)}{dt}\|_2, \quad \theta > 0, \tag{17}$$

on Theorem 3.2 as Theorem 2.2 in [1] does, this is due to the special structure of (10). Moreover, the converging point of $x(t)$ is also a regular 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 regular stationary point of a real analytic function $\psi(x)$, and an analogous version is presented as follows.

THEOREM 3.3 *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 (17) and*

$$[\frac{df(x)}{dt} = 0] \Rightarrow [h(x) = 0] \Rightarrow [\nabla f(x) = 0]. \tag{18}$$

Then, as $t \to +\infty$, either $\|x(t)\|_2 \to \infty$ or there exists $x^* \in R^n$ such that $\|x(t)\|_2 \to x^*$ 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 [6] (p. 429) ensures $h(x^*) = 0$ and the condition (18) leads to the result. □

4. Pseudo-transient continuation

Pseudo-transient continuation (Ψtc) is one way to implement the method (10). This method was originally designed as a method 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 [22, 8, 14, 18, 10] for the details of the theory and some applications. In this section we will only summarize the method. Some numerical results of Ψtc will be reported in Section 5.

In the context of optimization, one would integrate (2) numerically, managing the time step 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 do this is the iteration

$$x_{n+1} = x_n - (\delta_n^{-1} I + H(x_n))^{-1} h(x_n), \tag{19}$$

where $H(x)$ is the model Hessian or $H(x) = h'(x)$. One common way to manage the "time step" δ_n is "Switched Evolution Relaxation" (SER) [27]

$$\delta_{n+1} = \delta_n \|h(x_n)\| / \|h(x_{n+1})\|. \tag{20}$$

SER is supported by theory, and it is this approach we use in this section.

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

5. Computational experiments

This section deals with the numerical test of our continuous Newton-type method (10) in comparing with the continuous steepest descent method (3) by using the Matlab ODE solvers. 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 [26] is used and tested with their dimensions ranging from 2 to 400. For each test function, we use the same initial value x_0 as in [26]. The test problems are summarized in the following table.

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
$P11$	*Brwon 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 In this subsection, all computation 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 [31] can be used. Second, for each test function, the explicit expression of $\nabla^2 f(x)$ is employed. Third, because of Theorem 3.1, we do not have to require, as Theorem 3.2 states, that the test functions are real analytic. Finally, we let $\delta_2 = 1000\delta_1$ in (10) and fix $\delta_1 = \delta^{(0)} = 10^{-9}$, but if this fails for some problems, $\delta_1 = \delta^{(1)} = 10^{-4}$ would be used.

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 with the settings **RelTol** $= 10^{-8}$, **AbsTol** $= 10^{-9}$ and $\|\frac{d(x(t))}{dt}\|_\infty \leq \epsilon = 10^{-6}$ being the stopping criterion. The CPU times to obtain the acceptable solutions are summarized in Table 2 where '$*$' denotes that the method cannot stop within 1000 seconds of the CPU time; and the CPU times of the continuous steepest descent method (3) and our continuous Newton-type method (10) 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 (3) and (10), respectively.

Table 2. Comparison of (3) and (10) on ODE113

$No.$	n	m	$CPU_G(s)$	$CPU_N(s)$	$\lambda^*_{\min}$	f^*_G	f^*_N
$P1$	3	3	2.5781	0.5313	$1.4328D-00$	$6.4722D-13$	$7.9391D-13$
$P2$	6	6	128.9375	165.2656	$-4.5330D-05$	$3.5509D-05$	$3.5509D-05$
$P3$	3	15	0.0938	0.0469	$1.3966D-01$	$1.1283D-08$	$1.1282D-08$
$P4$	2	2	$*$	604.8594	$1.0059D-06$	$*$	$4.1537D-10$
$P5$	3	10	18.2656	$7.3750(\delta^{(1)})$	$9.1158D-04$	$5.6492D-10$	$5.6174D-12$
$P5$	3	20	15.3438	$7.2031(\delta^{(1)})$	$1.6145D-03$	$3.1329D-10$	$2.8701D-12$
$P6$	5	7	0.1563	0.0625	$2.0000D-00$	$1.1589D-15$	$4.0253D-11$
$P6$	10	12	0.1406	0.0781	$2.0000D-00$	$1.9155D-15$	$5.8237D-10$
$P6$	20	22	0.1719	0.1250	$2.0000D-00$	$4.6993D-16$	$1.9398D-08$
$P6$	30	32	0.1875	0.4375	$2.0000D-00$	$1.7847D-16$	$6.3986D-08$
$P7$	2	31	0.1250	0.0781	$2.3977D+01$	$5.4661D-01$	$5.4661D-01$
$P7$	6	31	$*$	1.6250	$2.8101D-03$	$*$	$2.2877D-03$
$P7$	8	31	$*$	4.8750	$7.5430D-06$	$*$	$1.8162D-05$
$P8$	4	5	20.9688	0.1250	$7.9998D-05$	$2.2514D-05$	$2.2500D-05$
$P8$	10	11	15.0313	0.1719	$1.2648D-04$	$7.0893D-05$	$7.0877D-05$
$P8$	20	21	11.3906	0.1875	$1.7887D-04$	$1.5780D-04$	$1.5778D-04$
$P8$	50	51	9.1563	0.4531	$2.8281D-04$	$4.3181D-04$	$4.3179D-04$
$P8$	100	101	8.7031	1.4531	$3.9993D-04$	$9.0253D-04$	$9.0249D-04$
$P8$	200	201	9.7031	6.5781	$5.6554D-04$	$1.8611D-03$	$1.8611D-03$
$P9$	4	8	0.1719	0.4844	$2.9693D-06$	$9.4914D-06$	$9.3763D-06$
$P9$	10	20	773.6875	0.4531	$1.8842D-05$	$2.9369D-04$	$2.9366D-04$
$P9$	20	40	$*$	0.3281	$1.3795D-04$	$*$	$6.3897D-03$
$P9$	50	100	188.5469	0.5313	$1.6645D-02$	$4.2961D-00$	$4.2961D-00$
$P9$	100	200	2.5000	1.5156	$2.2137D-01$	$9.7096D+04$	$9.7096D+04$
$P9$	200	400	14.3281	5.7188	$2.6871D+02$	$4.7116D+13$	$4.7116D+13$
$P10$	2	3	$*$	5.2188	$2.0000D-00$	$*$	$2.5763D-15$
$P11$	4	10	0.8750	0.3125	$4.7720D-00$	$1.4432D-00$	$1.4432D-00$
$P11$	4	20	4.0625	0.1563	$1.5158D+03$	$8.5822D+04$	$8.5822D+04$
$P11$	4	50	$*$	0.3594	$1.4581D+09$	$*$	$2.6684D+16$
$P11$	4	100	$*$	0.6406	$1.5186D+18$	$*$	$1.5087D+34$
$P12$	3	3	$*$	0.3438	$1.9330D-06$	$*$	$3.2312D-07$
$P13$	5	5	0.4063	0.5156	$1.5045D-01$	$4.3481D-11$	$1.5018D-11$
$P13$	10	10	0.2500	0.6875	$9.8024D-01$	$2.7951D-05$	$2.7951D-05$
$P14$	2	2	10.5625	0.1094	$3.9936D-01$	$3.9442D-12$	$2.9867D-13$
$P14$	10	10	11.2031	0.1250	$3.9936D-01$	$1.9721D-11$	$1.4933D-12$
$P14$	20	20	12.2500	0.2500	$3.9936D-01$	$3.9442D-11$	$2.9867D-12$
$P14$	50	50	15.4063	0.9844	$3.9936D-01$	$9.8606D-11$	$7.4667D-12$
$P14$	100	100	28.2813	4.5938	$3.9936D-01$	$1.9721D-10$	$1.4933D-11$
$P14$	200	200	79.7500	27.0313	$3.9936D-01$	$3.9442D-10$	$2.9867D-11$
$P14$	400	400	340.0625	212.2969	$3.9936D-01$	$7.8885D-10$	$5.9733D-11$
$P15$	4	4	234.0938	3.7656	$3.2196D-08$	$1.4476D-09$	$3.1023D-15$
$P15$	20	20	400.0781	5.3281	$3.2596D-08$	$7.2380D-09$	$1.5628D-14$
$P15$	40	40	606.6875	10.8438	$3.2228D-08$	$1.4476D-08$	$2.4472D-14$
$P15$	100	100	$*$	46.2813	$3.2281D-08$	$*$	$6.3657D-14$
$P15$	200	200	$*$	198.1563	$3.2127D-08$	$*$	$1.1339D-13$
$P16$	2	3	0.6719	0.3281	$3.0146D-01$	$2.2351D-12$	$1.0640D-13$
$P17$	4	6	23.9219	$6.7031(\delta^{(1)})$	$7.1957D-01$	$1.6888D-12$	$5.4878D-13$

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 method (10) is much more efficient and reliable compared with the gradient method (3), and converges globally to the regular stationary point(s).

5.2 Ψtc approach As we mentioned in Section 4, Ψtc is a very fast solver for (2). Even though the points generated by Ψtc would not have a monotonically decreasing objective function value in general, yet its fast convergence would always provide an attractive and competitive approach for any dynamical system resulted from the optimization problem. In our Ψtc implementation for (3), we set tol=[1e-9,1e-8], maxit=5000, mode=1, and qflag=0 (for more details, please see [22]). The following two tables summarize the numerical results, where $Iter$ represents the number of iterations, f^* represents the final objective function value, and δ_n^* represents the final value of δ_n. In addition, in the following four tables, '0*' denotes the 0 second return of function 'cputime' in MATLAB.

Table 3. Numerical results of Ψtc for (3) with dt=1e-1

$No.$	n	m	$Iter$	$CPU(s)$	f^*	$\frac{1}{\delta_n^*}$
$P1$	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^*	$9.8752e-011$	$5.1768e-005$
$P6$	10	12	14	0.0313	$1.0315e-009$	$1.7620e-006$
$P6$	20	22	16	0^*	$1.9000e-003$	$4.6426e-007$
$P6$	30	32	17	0.0313	$5.5257e-002$	$1.3659e-007$
$P7$	2	31	6	0^*	$5.4661e-001$	$8.5340e-006$
$P7$	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^*	$7.4403e-005$	$2.3594e-006$
$P8$	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^*	$1.3580e-014$	$1.5268e-006$
$P11$	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^*	$4.1877e-015$	$2.3416e-004$
$P14$	10	10	16	0^*	$2.0939e-014$	$2.3416e-004$
$P14$	20	20	16	0^*	$4.1877e-014$	$2.3416e-004$
$P14$	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^*	$2.0684e-009$	$1.4122e-007$
$P15$	20	20	18	0^*	$1.0342e-008$	$1.4122e-007$
$P15$	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 (3) with dt=1e-2

$No.$	n	m	$Iter$	$CPU(s)$	f^*	$\frac{1}{\delta_n^*}$
$P1$	3	3	76	0.0156	$2.4296D-12$	$2.0169e-003$
$P2$	6	6	489	0.5938	$3.5505D-05$	$1.1425e-005$
$P3$	3	15	24	0^*	$1.1279D-08$	$5.7233e-004$
$P4$	2	2	37	0.0313	$3.3789D-07$	$1.8238e-000$
$P5$	3	10	32	0^*	$4.0396D-13$	$1.4709e-005$
$P5$	3	20	22	0.0313	$1.3365D-17$	$1.3774e-006$
$P6$	5	7	11	0^*	$1.1030D-10$	$5.2633e-004$
$P6$	10	12	14	0^*	$1.0326D-09$	$1.7623e-005$
$P6$	20	22	16	0.0313	$1.8704D-03$	$4.6426e-006$
$P6$	30	32	17	0^*	$5.5257D-02$	$1.3659e-006$
$P7$	2	31	9	0^*	$5.4661D-01$	$4.3214e-005$
$P7$	6	31	22	0.1875	$2.2877D-03$	$3.9246e-006$
$P7$	8	31	25	0.4219	$1.8185D-05$	$2.8944e-006$
$P7$	9	31	21	0.5000	$2.7859D-06$	$1.2264e-006$
$P8$	4	5	20	0.0312	$2.2501D-05$	$1.3018e-005$
$P8$	10	11	13	0^*	$7.4418D-05$	$2.9452e-005$
$P8$	20	21	15	0.0313	$1.6349D-04$	$4.4251e-006$
$P8$	50	51	16	0^*	$1.7070D-02$	$3.5939e-006$
$P8$	100	101	17	0.4375	$4.5525D-01$	$1.1565e-006$
$P8$	200	201	17	0.1250	$3.7352D+01$	$1.1867e-006$
$P9$	4	8	39	0.0313	$9.3763D-06$	$3.1829e-004$
$P9$	10	20	34	0.0938	$2.9366D-04$	$1.3461e-004$
$P9$	20	40	32	0.1250	$6.3897D-03$	$4.3641e-004$
$P9$	50	100	22	0.0625	$4.2961D-00$	$2.6661e-005$
$P9$	100	200	19	0.0938	$9.7096D+04$	$1.8365e-006$
$P9$	200	400	10	0.0938	$4.7116D+13$	$3.8783e-004$
$P10$	2	3	63	0.0313	$4.7304D-13$	$1.1728e-005$
$P11$	4	10	85	0.0625	$1.4433D-00$	$1.0518e-006$
$P11$	4	20	17	0^*	$8.5822D+04$	$1.4095e-006$
$P11$	4	50	12	0.0313	$2.6684D+16$	$1.5367e-006$
$P11$	4	100	12	0.0313	$1.5087D+34$	$3.3363e-004$
$P12$	3	3	$fail$	$fail$	$fail$	$fail$
$P13$	5	5	792	0.2656	$4.1105D-17$	$2.3344e-006$
$P13$	10	10	904	0.5938	$2.7951D-05$	$3.2390e-006$
$P14$	2	2	21	0^*	$9.4629D-19$	$1.2025e-004$
$P14$	10	10	21	0^*	$4.7314D-18$	$1.2025e-004$
$P14$	20	20	21	0^*	$9.4629D-18$	$1.2025e-004$
$P14$	50	50	21	0.0313	$2.3657D-17$	$1.2025e-004$
$P14$	100	100	21	0.0625	$4.7314D-17$	$1.2025e-004$
$P14$	200	200	21	0.3125	$9.4629D-17$	$1.2025e-004$
$P14$	400	400	21	1.2031	$1.8937D-16$	$1.2025e-004$
$P15$	4	4	18	0.0313	$2.1577D-09$	$2.0004e-006$
$P15$	20	20	18	0^*	$1.0789D-08$	$2.0004e-006$
$P15$	40	40	18	0^*	$2.1577D-08$	$2.0004e-006$
$P15$	100	100	18	0.0625	$5.3944D-08$	$2.0004e-006$
$P15$	200	200	18	0.1406	$1.0789D-07$	$2.0004e-006$
$P16$	2	3	19	0^*	$7.2047D-19$	$2.7779e-005$
$P17$	4	6	48	0.0313	$8.2639D-16$	$1.6817e-005$

Since the Ψtc method for solving (3) already adopts the Hessian of $f(x)$, therefore, there is no direct application of Ψtc to the dynamical system (10). However, we can apply Ψtc partially to solve (10). Our test for solving (10) 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 the following 2 tables, where $Iter$, f^*, δ_n^* share the same meanings as Table 3 and Table 4; λ^* denotes the final computed $\lambda_{min}(x)$. We set $\delta_2 = 1.e-4$ in (11).

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

$No.$	n	m	Iter	CPU(s)	f^*	λ^*	$\frac{1}{\delta_n^*}$
$P1$	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^*	$1.1279e-008$	$1.3966e-001$	$2.6052e-003$
$P4$	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^*	$9.7541e-011$	$2.0000e-000$	$5.1672e-005$
$P6$	10	12	14	0^*	$1.0314e-009$	$2.0000e-000$	$1.7620e-006$
$P6$	20	22	16	0.0313	$1.9155e-003$	$2.0000e-000$	$4.6426e-007$
$P6$	30	32	17	0^*	$5.5257e-002$	$2.0000e-000$	$1.3659e-007$
$P7$	2	31	6	0^*	$5.4661e-001$	$2.3977e+001$	$1.7146e-007$
$P7$	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^*	$2.2513e-005$	$1.0022e-003$	$5.1724e-006$
$P8$	10	11	13	0^*	$7.4402e-005$	$1.3945e-002$	$2.3004e-006$
$P8$	20	21	15	0^*	$1.6347e-004$	$5.1832e-002$	$4.3120e-007$
$P8$	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^*	$9.3765e-006$	$6.2659e-004$	$3.9608e-004$
$P9$	10	20	29	0.0313	$2.9366e-004$	$2.1416e-003$	$6.2639e-005$
$P9$	20	40	34	0.0625	$6.4023e-003$	$2.5972e-004$	$2.0886e-006$
$P9$	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^*	$9.8341e-010$	$2.0000e-000$	$5.6000e-000$
$P11$	4	10	85	0.0625	$1.4433e-000$	$4.7750e-000$	$1.0525e-007$
$P11$	4	20	17	0^*	$8.5822e+004$	$1.5158e+003$	$1.4150e-007$
$P11$	4	50	12	0^*	$2.6684e+016$	$1.4581e+009$	$1.5367e-007$
$P11$	4	100	12	0.0313	$1.5087e+034$	$1.5197e+018$	$3.3363e-005$
$P12$	3	3	2	0^*	$1.4000e-003$	$-9.4304e-000$	$1.0000e+001$
$P13$	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^*	$6.8653e-020$	$3.9944e-001$	$3.9929e-007$
$P14$	10	10	7	0^*	$3.4326e-019$	$3.9944e-001$	$3.9929e-007$
$P14$	20	20	7	0^*	$6.8653e-019$	$3.9944e-001$	$3.9929e-007$
$P14$	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^*	$1.7193e-009$	$9.0837e-005$	$1.2294e-007$
$P15$	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 $dt = 1e-2$

$No.$	n	m	Iter	CPU(s)	f^*	λ^*	$\frac{1}{\delta_n^*}$
$P1$	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^*	$1.1279e-008$	$1.3966e-001$	$2.6052e-003$
$P4$	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^*	$9.7541e-011$	$2.0000e-000$	$5.1672e-005$
$P6$	10	12	14	0^*	$1.0314e-009$	$2.0000e-000$	$1.7620e-006$
$P6$	20	22	16	0.0313	$1.9155e-003$	$2.0000e-000$	$4.6426e-007$
$P6$	30	32	17	0^*	$5.5257e-002$	$2.0000e-000$	$1.3659e-007$
$P7$	2	31	6	0^*	$5.4661e-001$	$2.3977e+001$	$1.7146e-007$
$P7$	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^*	$2.2513e-005$	$1.0022e-003$	$5.1724e-006$
$P8$	10	11	13	0^*	$7.4402e-005$	$1.3945e-002$	$2.3004e-006$
$P8$	20	21	15	0^*	$1.6347e-004$	$5.1832e-002$	$4.3120e-007$
$P8$	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^*	$9.3763e-006$	$3.9279e-005$	$2.3035e-005$
$P9$	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^*	$9.8341e-010$	$2.0000e-000$	$5.6000e-000$
$P11$	4	10	85	0.0625	$1.4433e-000$	$4.7750e-000$	$1.0525e-007$
$P11$	4	20	17	0^*	$8.5822e+004$	$1.5158e+003$	$1.4150e-007$
$P11$	4	50	12	0^*	$2.6684e+016$	$1.4581e+009$	$1.5367e-007$
$P11$	4	100	12	0.0313	$1.5087e+034$	$1.5197e+018$	$3.3363e-005$
$P12$	3	3	2	0^*	$1.4000e-003$	$-9.4304e-000$	$1.0000e+001$
$P13$	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^*	$6.8653e-020$	$3.9944e-001$	$3.9929e-007$
$P14$	10	10	7	0^*	$3.4326e-019$	$3.9944e-001$	$3.9929e-007$
$P14$	20	20	7	0^*	$6.8653e-019$	$3.9944e-001$	$3.9929e-007$
$P14$	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^*	$1.7231e-009$	$9.6064e-005$	$1.2314e-007$
$P15$	20	20	17	0^*	$8.6154e-009$	$9.6064e-005$	$1.2314e-007$
$P15$	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	$7.8770e-000$	$-1.1943e-001$	$6.7781e-007$

From the previous 2 tables, we can see that the combined method works well for (10). However a more efficient Ψtc method designated for (10) should work even better. But this is beyond the scope of this paper.

6. Concluding remarks By combining the Newton's direction and the steepest descent direction, a new dynamical system (10) is proposed in this paper. The convergence and stability of this dynamical system are 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 (10) is somehow problem dependent, this can be seen from the numerical results of problems $P5$ and $P17$ in Table 2. Even though the Ψtc method can not be applied directly to solve (10), yet a combination of Newton's direction and the Ψtc direction also works well as shown in Table 5 and Table 6.

Finally, in [32], a globally convergent iterative algorithm for unconstrained optimization was proposed, which actually combines Newton's direction and the steepest descent direction within each iteration. The method involves some complicated controls, line search strategies, and direction search which are intended to satisfy the angle condition for global convergence. However, in our continuous Newton-type method (10), we set a natural way to define the trajectory without line search and angle condition, but the global convergence is guaranteed and our preliminary computational experiment shows its efficiency and reliability.

Acknowledgments. The research of the first and third authors was supported in part by FRG grants from Hong Kong Baptist University and the Research Grant Council of Hong Kong. The research of the second author was supported by National Science Foundation Grants DMS-0404537, and Army Research Office Grants W911NF-04-1-0276, W911NF-06-1-0096, and W911NF-06-1-0412.

References

[1] P.-A. Absil, R. Mahony, B. Andrews, *Convergence of the iterates of descent methods for analytic cost functions*, SIAM J. Optim., Vol. 16, No. 2, pp. 531–547, 2005.

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

[3] C. A. Botsaris, *Differential gradient methods*, J. Math. Anal. Appl., vol.63, pp.177-198, 1978.

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

[5] F. H. Branin, Jr. and S. K. Hoo, *A method for finding multiple extrema of a function of N variables, Proceedings of the Conference on Numerical methods for nonlinear Optimization, University of Dundee, Scotland, June 28-July 1, 1971*, Numerical Methods of Nonlinear Optimization, Academic Press, London, 1972.

[6] M. Braun, *Differential equations and their applications: an introduction to applied mathematics*, Springer-Verlag, New York, 1993.

[7] 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., Vol 62, **2**, pp.211–224, 1988.

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

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

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

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

[12] I. Diener, *Trajectory nets connecting all critical points of a smooth function*, Math. Prog., **36**, pp.340–352, 1986.

[13] I. Diener and R. Schaback, *An extended continuous Newton method*, J. Optim. Theory and Appl., **67(1)**, pp.57–77, 1990.

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

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

[16] 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**, pp. 363–381, 2001.

[17] U. Helmke and J. B. Moore, *Optimization and dynamical systems*, Springer-Verlag, London, UK, 1994.

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

[19] 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, pp. 209–228, 1987. Int. Conf., Plaue/GDR, 1985.

[20] 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.

[21] H. TH. Jongen, P. Jonker, and F. Twilt, *The continuous, Desingularized Newton method for meromorphic functions*, Acta Applicandae Mathematicae **13**, pp.81–121, 1988.

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

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

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

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

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

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

[28] J. M. Ortega and W. C. Rheinboldt, *Iterative solution of nonlinear equations in several variables*, Academic Press, New York, NY, 1970.

[29] L. Perko, *Differential equations and dynamical systems*, Springer-Verlag, New York, 1991.

[30] A. G. Ramm, *Linear ill-posed problems and dynamical systems.* J. Math. Anal. Appl., 258, No. 1 (2001), pp. 448–456.

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

[32] Y. Shi, *Globally convergent Algorithms for Unconstrained Optimization*, Comput. Optim. Appli., **16**, pp.295–308, 2000.

[33] J. J. E. Slotine and W. Li, *Applied nonlinear control*, Prentice Hall, New Jersey,1991.

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

[35] 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(1)**, pp.375–388, 1991.

[36] K. Tanabe, *Differential geometric methods in nonlinear programming*, Brookhaven Laboratory, New York, Technical Report 26730-AMD831, Math. Software, **10:3**, pp.200–316, 1979.