J Glob Optim (2007) 38:437–457
DOI 10.1007/s10898-006-9095-z

ORIGINAL ARTICLE

Self-concordant functions for optimization on smooth manifolds

Danchi Jiang · John B. Moore · Huibo Ji

Received: 28 February 2006 / Accepted: 8 September 2006 / Published online: 11 October 2006

Abstract This paper discusses self-concordant functions on smooth manifolds. In Euclidean space, such functions are utilized extensively as barrier functions in interior-point methods for polynomial time optimization algorithms. Here, the self-concordant function is carefully defined on a differential manifold in such a way that the properties of self-concordant functions in Euclidean space are preserved. A Newton decrement is defined and analyzed for this class of functions. Based on this, a damped Newton algorithm is proposed for the optimization of self-concordant functions. Under reasonable technical assumptions such as geodesic completeness of the manifold, this algorithm is guaranteed to fall in any given small neighborhood of the optimal solution in a finite number of steps. The existence and uniqueness of the optimal solution is also proved in this paper. Hence, the optimal solution is a global one. Furthermore, it ensures a quadratic convergence within a neighborhood of the minimal point. This neighborhood can be specified in terms of the Newton decrement. The computational complexity bound of the proposed approach is also given explicitly. This complexity bound is shown to be of the order $O(-\ln(\epsilon))$, where ϵ is the desired precision. Some interesting optimization problems are given to illustrate the proposed concept and algorithm.

Keywords Self-concordant functions · Polynomial algorithms · Damped Newton method · Geodesics · Covariant differentials

---

A part of the materials has been presented at 2004 Conference on Decision and Control

D. Jiang (✉)
The School of Engineering, University of Tasmania, Private Bag 65, Hobart, TAS 7001, Australia
e-mail: Danchi.Jiang@utas.edu.au

J. B. Moore · H. Ji
Dept. of Information Engineering, RSISE, Australian National University, Canberra, ACT 0200, Australia

J. B. Moore
e-mail: John.Moore@anu.edu.au

H. Ji
e-mail: huibo@rsise.anu.edu.au

1 Introduction

Self-concordant functions play an important role in the powerful interior-point polynomial algorithms for convex programing in Euclidean space. Following the work of Nesterov and Nemirovskii [1], many articles have been published using this type of function to construct barrier functions for interior-point algorithms. For example (see [2,3]). The idea of interior-point methods is to force the constraints of the optimization problem to be satisfied using a barrier penalty function in a composite cost function. This barrier function is relatively flat in the interior of the feasible region yet approaches to infinity in approaching the boundary. As the coefficient of the barrier function in the composite cost function converges to zero, the minimal point of the composite cost function converges to that of the original minimization problem. Self-concordant functions are a class of functions such that the third-order derivative is bounded by the cube of the square root of the second-order derivative, possibly with a constant scaling factor. The main advantage of using such a class of functions as barrier functions is that the computational complexity of the minimization problem of the constructed composite function is very low so that the original minimization problem can be solved in an amount of time that is a polynomial of the required precision.

The notion of a self-concordant function has deep roots in geometry. In [4], it is shown that a Riemannian metric can be rendered by a self-concordant barrier function. Such a metric gives a good explanation of the optimal direction for optimization algorithms. As such, it can provide guidance for the construction of efficient interior-point methods. In this aspect, the optimal path is along a geodesic defined by the Riemannian metric. This path is not a straight line in Euclidean space. Indeed, optimization problems in Euclidean space can often be better understood with the structure of appropriately associated Riemannian manifolds (see [5] for many meaningful examples).

In fact, many optimization problems can also be better posed on manifolds rather than in Euclidean space. For example, optimization problems associated with orthogonal matrices, such as algorithms for the computation of eigenvalues and singular values, or Oja's flow in neural network pattern classification algorithms (see [5–8]), to list a few. On the other hand, optimization methods such as the steepest descent method, the Newton method, and other related methods can be extended to Riemannian manifolds (see [5,6,9]). It is natural to ask, what are the self-concordant functions and the associated interior-point methods on manifolds? Such a question can be justified by practical importance or by theoretical completeness. The self-concordant concept has been applied in [10] with analysis restricted to a logarithm cost function optimized on a manifold. However, polynomial complexity is not available in that case. In this paper, we will give a useful definition of a self-concordant function on a manifold and give a thorough study of such.

One of the advantages of solving minimization problems on manifolds is, as pointed out in [6], the reduction of the dimensions of the problems, compared to solving the original problems in their ambient Euclidean spaces. A typical intrinsic approach for minimization is based on the computation of geodesics and covariant differentials, which might be expansive. However, there are many meaningful cases where the computation can be very simple. One of the examples is the real compact semisimple Lie group endowed with its natural Riemannian metric. In such a case the geodesic and parallel transportation can be computed by matrix exponentiation, whose rich set of nice properties provide convenience in analysis and development of efficient algorithms (see [6,11]). Many particular classes of minimization problems in this category such as those on Orthogonal groups or Stiefel manifolds are studied in [5]. Another simple but non-trivial case is the sphere, where the geodesic and parallel transportation can be computed via trigonometric functions and vector calculation.

This paper is organized as follows: notations and assumptions are listed in Sect. 2. In Sect. 3, the self-concordant function is defined to preserve as many nice properties of the original version in Euclidean space as possible. To facilitate the analysis and understanding of the elegantly proposed damped Newton method, the Newton decrement is defined and analyzed in Sect. 4. Then, the existence and uniqueness of the optimal solution are proved and a damped Newton algorithm is proposed in Sect. 5. It is shown that this algorithm has a similar convergence property and computational complexity to the algorithm for self-concordant functions in Euclidean space proposed in [12]. Two interesting examples are included to illustrate the proposed concept and approach in this paper in the last section.

2 Assumptions and notations

In this paper, notations will be listed in the following paragraph. Definitions and other details of the concepts and related results used will not be included. Please refer to [11] for more details.

M	A smooth manifold
T_pM	Tangent space of M at point p
$\mathrm{T}M$	tangent bundle of M
$\mathrm{T}_p^\star M$	1-forms of M at point p
$\mathrm{T}^\star M$	dual bundle of TM
∇	An affine connection defined on M
∇_X^i	the ith order covariant differentials with respect to the vector field X, based on the affine connection ∇, where i is an integer
$\exp_p X$	Exponential map of vector field X based on the parallel transform defined using the affine connection ∇
X_{pq}	The tangent vector field of the shortest geodesic connecting point p and q on M, where the geodesic parameter is 0 at p and 1 at q
t_τ	Parallel transform on M, where the affine connection is determined according to the context.

In this paper, our purpose is to show how the self-concordant function can be extended to non-Euclidean space. The discussion is not intended for the most general cases to avoid unnecessary details. As such, we make the following assumptions:

Assumption 1

(1) The affine connection is symmetric, meaning that

$$\nabla_Y \nabla_X(f) = \nabla_X \nabla_Y(f),$$

for any vector field X, Y and function f. A very interesting example of the symmetric connection is the Riemannian connection.

(2) The smooth manifold M of interest is geodesic complete. Furthermore, the geodesic between two points is unique.

3 Self-concordant functions

Let M be a smooth manifold of finite dimension and ∇ a symmetric affine connection defined on it. Assume that M is geodesic complete in this paper to focus on the main ideas rather than how general the proposed concept and method can be applied. Consider a function defined on M: $f: M \to R$, which has an open domain, a closed map, meaning that $\{(f(P), P), P \in dom(f)\}$ is a closed set in the product manifold $\Re \times M$, and is at least three times differentiable.

Definition 1 f is a self-concordant function with respect to ∇ if and only if the following condition holds:

$$\left|\nabla_X^3 f(p)\right| \leq M_f \left[\nabla_X^2 f(p)\right]^{3/2}, \quad \forall X \in \mathrm{T}_p M, \quad p \in M, \tag{1}$$

where M_f is a positive constant associated with f.

As noticed in [1], if a function f is self-concordant with the constant M_f, then the function $M_{\mathrm{f}}^{-2} f$ is self-concordant with the constant 1. It can also be directly checked by simple computation. As such, we assume $M_f = 2$ for the rest of this paper.

Also notice that the second covariant differential of a self-concordant function is a positive semi-definite mapping, meaning that it is symmetric with respect to two tangent vectors and its value is always non-negative. For the simplicity of the analysis in this paper, we only consider those functions that satisfy the following assumption:

Assumption 2

$$\nabla_X^2 f(p) > 0, \quad \forall p \in \mathrm{dom}(f), \quad X \in \mathrm{T}_p M.$$

Then, the second-order covariant differentials can be used to define a Dikin-type ellipsoid $W^\circ(p; r)$ as follows:

Definition 2 For any $p \in \mathrm{dom}(f)$, and $r > 0$,

$$W^\circ(p; r) := \left\{ q \in M \mid \left[\nabla_{X_{pq}}^2 f(p)\right]^{1/2} < r \right\},$$

where X_{pq} is the vector field defined by the geodesic connecting the points p and q.

The definition of self-concordant function is based on second-order and third-order covariant differentials with respect to the same vector field X. However, it is equivalent to the case where they are calculated with respect to different vector fields. More specifically, the following property holds:

Property 1 *If f is a self-concordant function defined on M, then, the following inequality holds:*

$$\left|\nabla_{X_1} \nabla_{X_2} \nabla_{X_3} f(p)\right| \leq M_f \left[\nabla_{X_1}^2 f(p)\right]^{1/2} \left[\nabla_{X_2}^2 f(p)\right]^{1/2} \left[\nabla_{X_3}^2 f(p)\right]^{1/2}, \quad \forall X_1, X_2, X_3 \in \mathrm{T}_p M. \tag{2}$$

This property comes from the linearity of the mapping

$$\nabla_{\mathrm{d}f}^2 : \mathrm{T}_p M \times \mathrm{T}_p M \to R, \quad \text{and}$$
$$\nabla_{\mathrm{d}f}^3 : \mathrm{T}_p M \times \mathrm{T}_p M \times \mathrm{T}_p M \to R,$$

defined by

$$\nabla^2_{\mathrm{d}f}(X_1, X_2) := \nabla_{X_1}\nabla_{X_2} f(P), \quad \text{and}$$
$$\nabla^3_{\mathrm{d}f}(X_1, X_2, X_3) := \nabla_{X_1}\nabla_{X_2}\nabla_{X_3} f(p),$$

respectively, for any $X_1, X_2, X_3 \in \mathrm{T}_p M$.

Proof The following proof is inspired by that in ([1], Proposition 9.1). However, it should be pointed out that, the trilinear property of the third-order differentials is not required in our proof.

By assumption the affine connection is symmetric. Therefore, the manifold is torsion free. With a slight misuse of notions, for a given tangent vector X at p, we let X also denote the vector field generated by X using parallel transportation in a neighborhood of p where a geodesic starting from p along the direction X exists uniquely. In this case, it is easy to see that $[X_i, X_j] = 0, i \neq j$. Hence the corresponding second-order covariant differentials is symmetric, meaning

$$\nabla_{X_i}\nabla_{X_j} f(p) = \nabla_{X_j}\nabla_{X_i} f(p), \quad i, j = 1, 2, 3.$$

As such, it can be directly checked that

$$\begin{aligned} \nabla_{X_1}\nabla_{X_2}\nabla_{X_3} f(p) &= \nabla_{X_2}\nabla_{X_1}\nabla_{X_3} f(p), \\ \nabla_{X_1}\nabla_{X_2}\nabla_{X_3} f(p) &= \nabla_{X_1}\nabla_{X_3}\nabla_{X_2} f(p). \end{aligned} \tag{3}$$

Adopting the reasoning process in Proposition 9.1.1 in [1, p 361], we only need to show that the extremal value of the function:

$$\begin{aligned} W(X_1, X_2, X_3) &:= \mid \nabla_{X_1}\nabla_{X_2}\nabla_{X_3} f(p) \mid, \\ &\qquad \forall X_i \in T_M(p), \;\; \nabla^2_{X_i} f(p) = 1, \quad i = 1, 2, 3 \end{aligned} \tag{4}$$

is achieved at a triple vector in the form of (Z, Z, Z).

Assume the contrary, let $(X^\star, Y^\star, Z^\star)$ be an extremal point of W. By Lemma 9.1.2 in [1] one knows that there exists a unit vector $X_1^\star$, meaning its second covariant differential of f is 1, in the direction $X^\star + Y^\star$ such that $(X_1^\star, X_1^\star, Z^\star)$ achieves the extreme. By the same method in this lemma, one can show that there exists a unit vector $Y_1^\star$ in the one dimensional linear space spanned by $X_1^\star + Z^\star$ such that $(X_1^\star, Y_1^\star, Y_1^\star)$ achieves the extremal value. By using this method iteratively, one constructs a sequence of vectors in $\Re^n \times \Re^n \times \Re^n$:

$$\{(X^\star, Y^\star, Z^\star), (X_1^\star, X_1^\star, Z^\star), (X_1^\star, Y_1^\star, Y_1^\star), (X_2^\star, X_2^\star, Y_1^\star), (X_2^\star, Y_2^\star, Y_2^\star), \ldots\},$$

where the vector in the middle is always in the one-dimensional linear space spanned by the summation of two different vectors in the previous triple and either the first or the last vector is the same as the middle. By the assumption that $\nabla^2_X f(p) > 0$ for all tangent vector X, one can show that this sequence is convergent to a vector triple in the form of $(Z^\star, Z^\star, Z^\star)$. Hence this property is proved. □

A self-concordant function also has the following interesting property:

Property 2 $\forall p \in \mathrm{dom}(f) \subseteq M, W^\circ(p; 1) \subseteq \mathrm{dom}(f)$.

This property gives a safe bound for the line search along geodesics for optimization problems so that the search will always be in the admissible domain. We need the following lemma to prove it:

Lemma 1 *Given that f is a self-concordant function defined on a smooth manifold M and X a vector field on M, define a function $\phi(t) : \Re \to \Re$ as follows:*

$$\phi(t) := \left[\nabla^2_{\tau_{p\exp_p(tX)}X} f(\exp_p tX)\right]^{-1/2}, \tag{5}$$

where τ_{pq} is the parallel transportation from the point p to the point q and $\exp_p(X)$ is the exponential map of the vector field X at p. Then, the following results hold:

(1) $\mid \phi'(t) \mid \leq 1$.
(2) *If* $\phi(0) > 0$, *then,* $(-\phi(0), \phi(0)) \subseteq \mathrm{dom}(\phi)$.

Proof It can be calculated that

$$\phi'(t) = -\frac{\frac{\mathrm{d}}{\mathrm{d}t}\left[\nabla^2_{\tau_{p\exp_p(tX)}X} f(\exp_p tX)\right]}{2\left[\nabla^2_{\tau_{p\exp_p(tX)}X} f(\exp_p tX)\right]^{3/2}}$$
$$= -\frac{\nabla^3_{\tau_{p\exp_p(tX)}X} f(\exp_p tX)}{2\left[\nabla^2_{\tau_{p\exp_p(tX)}X} f(\exp_p tX)\right]^{3/2}}.$$

The claim (1) follows directly from the definition of self-concordant function.

Assume the cliam (2) is not true. Since $\phi(0) > 0$, because of the continuity of $\nabla^2_{\tau_{p\exp_p(tX)}} f(\exp_p tX)$, there is a symmetric neighborhood of 0 in the definition domain of ϕ. Let $(-\bar{t}, \bar{t})$ denote the largest of such symmetric neighborhoods. Then, at least one of the two end points is not in $\mathrm{dom}(\phi)$. Without loss of generality, assume $\bar{t}$ is this point and $\bar{t} < \phi(0)$. Because $\phi(t) \geq \phi(0) - \mid t \mid$, we have

$$\nabla^2_{\tau_{p\exp_p(tX)}} f(\exp_p tX) < \frac{1}{(\phi(0) - \mid t \mid)^2} \leq \frac{1}{(\phi(0) - \bar{t})^2} < +\infty, \quad \forall t \in (-\bar{t}, \bar{t}).$$

Hence,

$$\lim_{t \to \bar{t}-0} \nabla^2_{\tau_{p\exp_p(tX)}X} f(\exp_p tX) = \nabla^2_{\tau_{p\exp_p(\bar{t}X)}X} f(\exp_p \bar{t}X)$$
$$= \frac{1}{[\phi(0) - \|\bar{t}\|]^2} < +\infty.$$

The existence of $f(\exp_p \bar{t}X)$ comes from the assumption that f has a closed map and the fact that

$$f(\exp_p tX) = \int_0^t \left[\int_0^s \nabla^2_{\tau_{p\exp_p(\nu X)}} f(\exp_p \nu X)\mathrm{d}\nu + \nabla_X f(X)\right] \mathrm{d}s < +\infty.$$

Therefore, $\phi(\bar{t})$ is well-defined, which is contradiction to the assumption we made. As such, (2) holds.

Now we can prove Property 2.

Proof Notice, from Lemma 1, that for any vector field $X \in \mathrm{T}M$,

$$\{\exp_p(tX) \mid \quad \mid t \mid^2 < \phi^2(0)\} \subseteq \mathrm{dom}(f).$$

On the other hand,

$$\{\exp_p(tX) \mid\mid t \mid^2 < \phi^2(0)\} = \left\{\exp_p(tX) \mid\mid t \mid^2 < \frac{1}{\nabla^2_X f(p)}\right\}$$
$$= \left\{\exp_p tX \mid\mid t \mid^2 \left(\nabla^2_X f(p)\right) < 1\right\} = W^\circ(p; 1).$$

This completes the proof. □

In the following, two groups of properties will be given to reveal the relationship between two different points on a geodesic. They are delicate characteristics of self-concordant functions. In fact, they are the foundation for the polynomial complexity of self-concordant functions.

Property 3 *For any* $p, q \in \mathrm{dom}(f)$, *such that there is a geodesic contained in the definition domain of f connecting the points p and q, if f is a self-concordant function, the following results hold:*

$$\left[\nabla^2_{X_{pq}(q)} f(q)\right]^{1/2} \geq \frac{\left[\nabla^2_{X_{pq}(p)} f(p)\right]^{1/2}}{1 + \left[\nabla^2_{X_{pq}(p)} f(p)\right]^{1/2}}, \tag{6}$$

$$\nabla_{X_{pq}(q)} f(q) - \nabla_{X_{pq}(p)} f(p) \geq \frac{\nabla^2_{X_{pq}(p)} f(p)}{1 + \left[\nabla^2_{X_{pq}(p)} f(p)\right]^{1/2}}, \tag{7}$$

$$f(q) \geq f(p) + \nabla_{X_{pq}(p)} f(p) + \left[\nabla^2_{X_{pq}(p)} f(p)\right]^{1/2} - \ln\left(1 + \left[\nabla^2_{X_{pq}(p)} f(p)\right]^{1/2}\right). \tag{8}$$

Proof Let $\phi(t)$ be the same function defined in Lemma 1, where one can see that $\phi(1) \leq \phi(0) + 1$. This is equivalent to (6) taking into account that $\phi(0) = [\nabla^2_{X_{pq}(p)} f(p)]^{-1/2}$, and $\phi(1) = [\nabla^2_{X_{pq}(q)} f(q)]^{-1/2}$. Furthermore,

$$\nabla_{X_{pq}(q)} f(q) - \nabla_{X_{pq}(p)} f(p) = \int_0^1 \nabla^2_{X_{pq}(\exp_p tX_{pq})} f(\exp_p tX_{pq}) \mathrm{d}t,$$
$$= \int_0^1 \frac{1}{t^2} \nabla^2_{tX_{pq}(\exp_p tX_{pq})} f(\exp_p tX_{pq}) \mathrm{d}t, \tag{9}$$

which leads to (7) using the inequality (6).

For the inequality (8), notice that:

$$f(q) - f(p) - \nabla_{X_{pq}} f(p) = \int_0^1 (\nabla_{X_{pq}(\exp_p tX_{pq})} f\left(\exp_p tX_{pq}\right) - \nabla_{X_{pq}(p)} f(p))\mathrm{d}t$$
$$= \int_0^1 \frac{1}{t}\left[\nabla_{tX_{pq}(\exp_p tX_{pq})} f\left(\exp_p tX_{pq}\right) - \nabla_{tX_{pq}(p)} f(p)\right]\mathrm{d}t$$
$$\geq \int_0^1 \frac{\nabla^2_{tX_{pq}(p)} f(p)}{t(1+[\nabla^2_{tX_{pq}}(p)]^{1/2})}\mathrm{d}t.$$

Let $r = [\nabla^2_{X_{pq}(p)} f(p)]^{1/2}$. The last integral becomes

$$\int_0^1 \frac{tr^2}{1+tr}\,\mathrm{d}t = r - \ln(1+r),$$

which leads to the inequality (8) by replacing r with its original form.

Property 4 *For any $p \in dom(f)$, $q \in W^\circ(p;1)$, and $X \in \mathrm{T}M$, there holds:*

$$(1 - [\nabla^2_{X_{pq}(p)} f(p)]^{1/2})^2 \nabla^2_{X(p)} f(p) \leq \nabla^2_{X(q)} f(q) \leq \frac{\nabla^2_{X(p)} f(p)}{\left(1 - \left[\nabla^2_{X_{pq}(p)} f(p)\right]^{1/2}\right)^2}, \tag{10}$$

$$\nabla_{X_{pq}(q)} f(q) - \nabla_{X_{pq}(p)} f(p) \leq \frac{\nabla^2_{X_{pq}(p)} f(p)}{1 - \left[\nabla^2_{X_{pq}(p)} f(p)\right]^{1/2}}, \tag{11}$$

$$f(q) \leq f(p) + \nabla_{X_{pq}(p)} f(p) - \left[\nabla^2_{X_{pq}(p)} f(p)\right]^{1/2} - \ln(1 - \left[\nabla^2_{X_{pq}(p)} f(p)\right]^{1/2}). \tag{12}$$

Proof Let $\psi(t)$ be a function defined in the following form:

$$\psi(t) = \nabla^2_{X(\exp_p tX_{pq})} f(\exp_p tX_{pq}).$$

Then, its derivative satisfies

$$|\ \psi'(t)\ | = |\ \nabla_{X_{pq}(\exp_p tX_{pq})} \nabla^2_{X(\exp_p tX_{pq})} f(\exp_p tX_{pq})\ |$$
$$\leq 2\ |\ \nabla^2_{X_{pq}(\exp_p tX_{pq})} f(\exp_p tX_{pq})\ |^{1/2}\ \psi(t)$$
$$= \frac{2}{t}\ |\ \nabla^2_{tX_{pq}(\exp_p tX_{pq})} f(\exp_p tX_{pq})\ |^{1/2}\ \psi(t)$$
$$\leq \frac{2}{t}\frac{t\left[\nabla^2_{X_{pq}(p)} f(p)\right]^{1/2}}{1 - t\left[\nabla^2_{X_{pq}(p)} f(p)\right]^{1/2}}\psi(t). \tag{13}$$

Here the last part is obtained by applying $\phi(1) \geq \phi(0) - 1$ from Lemma 1. Integrating both sides of the inequality (13), one obtains

$$(1 - [\nabla^2_{X_{pq}(p)} f(p)]^{1/2})^2 \leq \frac{\psi(1)}{\psi(0)} \leq \frac{1}{\left(1 - \left[\nabla^2_{X_{pq}(p)} f(p)\right]^{1/2}\right)^2},$$

which is equivalent to the inequality (10).

Combining the inequality (10) and the formula (9), one obtains

$$\begin{aligned}\nabla_{X_{pq}(q)} f(q) - \nabla_{X_{pq}(p)} f(p) &\leq \int_0^1 \frac{1}{t^2} \frac{\nabla^2_{tX_{pq}(p)} f(p)}{\left(1 - \left[\nabla^2_{tX_{pq}(p)} f(p)\right]^{1/2}\right)^2} \mathrm{d}t \\ &= \frac{\nabla^2_{X_{pq}(p)} f(p)}{1 - \left[\nabla^2_{X_{pq}(p)} f(p)\right]^{1/2}},\end{aligned}$$

which proves the inequality (11).

Combining this result and using the same technique as that used in the proof of the last property, there holds:

$$\begin{aligned}f(q) - f(p) - \nabla_{X_{pq}} f(p) &= \int_0^1 \nabla_{X_{pq}(\exp_p tX_{pq})} f(\exp_p tX_{pq}) \mathrm{d}t - \nabla_{X_{pq}} f(p) \\ &= \int_0^1 \left\{ \frac{1}{t} [\nabla_{tX_{pq}(\exp_p tX_{pq})} f(\exp_p tX_{pq})] - \nabla_{X_{pq}} f(p) \right\} \mathrm{d}t \\ &\leq \int_0^1 \frac{\nabla^2_{tX_{pq}} f(p)}{t(1 - [\nabla^2_{tX_{pq}} f(p)]^{1/2})} \mathrm{d}t \\ &= -\left[\nabla^2_{X_{pq}(p)} f(p)\right]^{1/2} - \ln\left(1 - \left[\nabla^2_{X_{pq}(p)} f(p)\right]^{1/2}\right).\end{aligned}$$

As such, the inequality (12) is obtained by a simple transformation of this inequality. □

4 Newton decrement

Consider the following auxiliary quadratic cost defined on $\mathrm{T}_p M$:

$$N_{f,p}(X) := f(p) + \nabla_X f(p) + \frac{1}{2} \nabla^2_X f(p). \tag{14}$$

Definition 3 The Newton decrement $X_N(f,p)$ is defined as the minimal solution to the auxiliary cost function given by (14). More specifically,

$$X_N(f,p) := \arg \min_{X \in \mathrm{T}_p M} N_{f,p}(X). \tag{15}$$

Similar to the case in Euclidean space, the Newton decrement can be characterized in many ways. The following theorem summaries its properties.

Theorem 1 *Let $f\colon M \to R$ be a self-concordant function, p, a given point in dom(f) $\subseteq$ M, and X_N, its Newton decrement defined at p. The following results hold:*

$$\nabla_{X_N}\nabla_X f(p) = -\nabla_X f(p), \quad \forall X \in \mathrm{T}_p M, \tag{16}$$

$$\sqrt{\nabla^2_{X_N} f(p)} = \max\left\{\nabla_X f(p) | X \in \mathrm{T}_p M, \nabla^2_X f(p) \le 1\right\}. \tag{17}$$

Furthermore, if the quadratic form $\nabla^2_\mu\colon \mathrm{T}_p M \to T^\star_M(p)$, defined by:

$$\nabla^2_\mu(X) = \nabla^2_X \mu, \quad \forall X \in \mathrm{T}_p M$$

for a given $\mu \in T^\star_M(p)$, is of full rank, then

$$X_N = -(\nabla^2_{\mathrm{d}f})^{-1}\mathrm{d}f. \tag{18}$$

Proof Since p is a given point on the manifold M, the claimed results can be converted into their local representation in Euclidean space. More specifically, consider the following quadratic function:

$$q(x) := \frac{1}{2}x^\top Ax + b^\top x + c,$$
$$\text{where} \quad A \in R^n \times R^n, \quad A^\top = A, b \in R^n, \ c \in R. \tag{19}$$

Let $x^\star$ denote the optimal point. Then, the gradient of q at $x^\star$ must be a zero vector. i.e.,

$$y^\top(Ax^\star + b) = 0, \quad \forall y \in R^n.$$

This is the local representation of (16).

The case where ∇^2_μ is isomorphic is corresponding to the case where A is non-degenerate. Then, $x^\star = -A^{-1}b$, which is the local representation of (18).

On the other hand,

$$\begin{aligned} | \, y^\top b \, | &= | \, y^\top Ax^\star \, | = | \, y^\top A^{1/2}A^{1/2}x^\star \, | \\ &\le (y^\top Ay)^{1/2}((x^\star)^\top Ax^\star)^{1/2}, \end{aligned}$$

where the equality holds if and only if $y = x^\star$. As such,

$$\begin{aligned} \max\{y^\top b | y \in R^n, y^\top Ay \le 1\} &= \max\left\{\frac{y^\top b}{\sqrt{y^\top Ay}} | y \in R^n\right\} \\ &= \sqrt{(x^\star)^\top Ax^\star}. \end{aligned}$$

This is the local representation of (17). Therefore, the proof is complete.

5 A damped newton algorithm for self-concordant functions

Consider now the minimization problem of self-concordant functions on a smooth manifold. First, let us establish the existence of the minimal point:

Theorem 2 *Let $\lambda_f(p)$ be defined as follows:*

$$\lambda_f(p) := \max_{X \in \mathrm{T}_p M} \frac{| \nabla_X f(p) |}{\sqrt{\nabla^2_X f(p)}} \quad \textit{for } p \in \mathrm{dom}(f). \tag{20}$$

If $\lambda_f(p) < 1$ *for some* $p \in$ *dom(f), then there exists a unique point* $p_f^\star \in \mathrm{dom}(f)$ *such that*

$$f(p_f^\star) = \min\{f(p) \mid p \in \mathrm{dom}(f)\}.$$

Proof Let p_0 is a point such that $\lambda_f(p_0) < 1$. For any $q \in \mathrm{dom}(f)$ such that $f(q) < f(p_0)$, from (8) we have

$$f(q) \geq f(p_0) - \lambda_f(p_0)[\nabla^2_{X_{p_0q}} f(p_0)]^{1/2} + \left[\nabla^2_{X_{p_0q}} f(p_0)\right]^{1/2} - \ln\left(1 + \left[\nabla^2_{X_{p_0q}} f(p_0)\right]^{1/2}\right).$$

Hence,

$$\frac{\ln\left(1 + \left[\nabla^2_{X_{p_0q}} f(p_0)\right]^{1/2}\right)}{\left[\nabla^2_{X_{p_0q}} f(p_0)\right]^{1/2}} \geq 1 - \lambda.$$

Since

$$\lim_{t \to +\infty} \frac{\ln(1+t)}{t} = 0,$$

there exists a constant $c > 0$, such that

$$\left[\nabla^2_{X_{p_0q}} f(p_0)\right]^{1/2} \leq c. \tag{21}$$

Hence these X_{p_0q} contained in the compact set defined by inequality (21). Consider the map from the any tangent vector X to its geodesic $\exp(X)$. This map is continuous by the definition of geodesic. This indicates the image of compact set defined by inequality (21) is also compact. Therefore, a minimal point exists.

On the other hand, let $p^\star$ denote a minimal point. Then,

$$\begin{aligned} f(q) &\geq f(p^\star) + [\nabla^2_{X_{p^\star q}} f(p^\star)]^{1/2} - \ln(1 + [\nabla^2_{X_{p^\star q}} f(p^\star)]^{1/2}) \\ &> f(p^\star), \quad \forall q \in \mathrm{dom}(f), \ q \neq p^\star. \end{aligned} \tag{22}$$

The uniqueness is proved.

Consider the following damped Newton method:

Algorithm 1 *(Damped Newton algorithm)*
Step 0: Find a feasible point $p_0 \in \mathrm{dom}(f)$.
Step k: $p_k = \exp_{p_{k-1}} \frac{1}{1+\lambda_f(p_{k-1})} X_N$,
where $\exp_{p_{k-1}} tX_N$ is the exponential map of the Newton decrement at p_{k-1}.

The following theorem establishes the convergence properties of the proposed damped Newton algorithm:

Theorem 3 *Let the minimal point of* $f(p)$ *be denoted as* $p^\star$, *and* p *is any admissible point in* $W^\circ(p^\star; 1)$.

(1) *The following inequality holds:*

$$\left[\nabla^2_{X_{pp^\star}} f(p)\right]^{1/2} \leq \frac{\lambda_f(p)}{1 - \lambda_f(p)}. \tag{23}$$

(2) *If* $\lambda_f(p) < 1$, *then*

$$0 \le f(p) - f(p^\star) \le -\lambda_f(p) - \ln(1 - \lambda_f(p)). \tag{24}$$

(3) *For the proposed Damped Newton Method algorithm, there holds:*

$$f(p^\star) \le f(p_k) \le f(p_{k-1}) - (\lambda_f(p_{k-1}) - \ln(1 + \lambda_f(p_{k-1}))). \tag{25}$$

Proof

(1) Let $[\nabla^2_{X_{pp^\star}} f(p)]^{1/2}$ be denoted as $r(p)$. In view of (7) and (11) we have:

$$\frac{r^2}{1-r} \ge -\nabla_{X_{pp^\star}} f(p) \ge \frac{r^2(p)}{1+r(p)} \ge 0. \tag{26}$$

On the other hand, there holds

$$| \nabla_{X_{pp^\star}} f(p) | \le \lambda_f(p) r(p)$$

by the definition of $\lambda_f(p)$. Therefore,

$$\lambda_f(p) \ge \frac{r(p)}{1+r(p)},$$

where r can be solved as follows:

$$r(p) \le \frac{\lambda_f(p)}{1-\lambda_f(p)},$$

which is (23).

(2) Based on (8) and the inequality (23) obtained above, one has:

$$\begin{aligned} f(p^\star) - f(p) &\ge \nabla_{X_{pp^\star}} f(p) + r(p) - \ln(1 + r(p)) \\ &\ge r(p) - \ln(1 + r(p)) - \lambda_f(p) r(p). \end{aligned} \tag{27}$$

Let an auxiliary function $g(x, y)$ be defined as:

$$g(x, y) = x - \ln(1 + x) - xy + y - \ln(1 - y), \quad \forall x \ge 0, \ \ 1 > y \ge 0.$$

It can be easily checked that

$$g(x, 0) = x - \ln(1 + x) \ge 0$$

and

$$g(0, y) = y - \ln(1 - y) \ge 0.$$

If there is a point (x_0, y_0) such that $g(x_0, y_0) < 0$, this function must have a minimal interior point. The gradient will be zero at such a point. However, it can be calculated that

$$\frac{\partial g}{\partial x} |_{(x_0,y_0)} = 1 - \frac{1}{1+x_0} - y_0 = 0,$$
$$\frac{\partial g}{\partial y} |_{(x_0,y_0)} = -x_0 + 1 + \frac{1}{1-y_0} = 0.$$

The solution to this system of equations satisfies

$$(1 - y_0)(1 + x_0) = 1.$$

As such, at the minimal point there holds:

$$g(x_0, y_0) = x_0 - x_0 y_0 + y_0 = x_0(1 - y_0) + y_0 > 0,$$

which is a contradiction. Therefore, the minimum, if it exists, is achieved at the boundary. Hence,

$$g(x, y) \geq 0, \qquad \forall x \geq 0, 1 > y \geq 0.$$

Applying this inequality to (27), we obtain the right side of the inequality (24).

(3) It is clear that $p_{k+1} \in W^{\circ}(p_k, 1)$ since

$$\nabla^2_{\frac{1}{1+\lambda_f(p_k)} X_N} f(p_k) \overset{(4.4)}{=} \left[\frac{1}{1 + \lambda_f(p_k)}\right]^2 \lambda_f(p_k)^2 < 1.$$

Applying (12), there holds

$$\begin{aligned} f(p_{k+1}) &\leq f(p_k) + \frac{1}{1 + \lambda_f(p_k)} \nabla_{X_N} f(p_k) - \frac{1}{1 + \lambda_f(p_k)} [\nabla^2_{X_N} f(p_k)]^{1/2} \\ &\quad - \ln\left(1 - \frac{1}{1 + \lambda_f(p_k)} [\nabla^2_{X_N} f(p_k)]^{1/2}\right) \\ &= f(p_k) - \lambda_f(p_k) + \ln(1 + \lambda_f(p_k)) \end{aligned}$$

by the definition of $\lambda_f(p_k)$ and the results in Theorem 1. Therefore, the inequality (25) is proved.

Notice that the two functions

$$\lambda - \ln(1 + \lambda), \qquad \forall \lambda \in (0, +\infty)$$

and

$$-\lambda - \ln(1 - \lambda), \qquad \forall \lambda \in (0, 1)$$

are positive and monotonically increasing. The results proved in Theorem 3 have already given a set of error bounds for the function $f(p)$ and estimation of the variable point p based on $\lambda_f(p)$. More specifically, the inequality (25) implies the following results:

Corollary 1 *For the Damped Newton algorithm, the $\lambda_f(p_k)$ is bounded as follows:*

$$\lambda_f(p_k) - \ln(1 + \lambda_f(p_k)) \leq f(p_k) - f(p^\star). \tag{28}$$

Furthermore, for a given precision $\epsilon > 0$, the number of iterations, denoted as N, required such that $\lambda_f(p_N) < \epsilon$ is less than or equal to $\frac{f(p_0) - f(p^\star)}{\epsilon - \ln(1+\epsilon)}$.

For the convergence rate, the following theorem reveals the quadratic convergence and the computational complexity of the damped Newton algorithm proposed in this paper.

Theorem 4 *For the damped Newton algorithm proposed in this paper, the following result holds:*

$$\lambda_f(p_{k+1}) \leq 2\lambda_f^2(p_k). \tag{29}$$

Proof For a vector field $X \in \mathrm{T}M$, construct an auxiliary function $\psi(t)$ as follows:

$$\psi(t) := \nabla_X f(\exp_{p_k} tX_N(p_k)), \quad \forall t \in \left(0, \frac{1}{1+\lambda_f(p_k)}\right). \tag{30}$$

Now, with a slight misuse of the notation that ignores the vector field adaptation:

$$\begin{aligned}
\psi'(t) &= \nabla_{X_N(p_k)} \nabla_X f(\exp_{p_k} tX_N(p_k)), \\
\psi''(t) &= \nabla^2_{X_N} \nabla_X f(\exp_{p_k} tX_N(p_k)) \\
&\leq 2\nabla^2_{X_N} f(\exp_{p_k} tX_N(p_k))[\nabla^2_X f(\exp_{p_k} tX_N(p_k))]^{1/2}.
\end{aligned}$$

The last inequality is obtained in view of Property 1.

Notice that:

$$\begin{aligned}
\nabla^2_{X_N} f(\exp_{p_k} tX_N(p_k)) &= \frac{1}{t^2} \nabla^2_{tX_N} f(\exp_{p_k} tX_N(p_k)) \\
&\leq \frac{1}{t^2} \frac{\nabla^2_{tX_N} f(p_k)}{\left(1 - \left[\nabla^2_{tX_N} f(p_k)\right]^{1/2}\right)^2} \\
&= \frac{\lambda_f^2(p_k)}{(1 - t\lambda_f(p_k))^2}
\end{aligned} \tag{31}$$

by applying the inequality (10) and the fact that $p_{k+1} \in W^\circ(p_k, 1)$. Then, by applying the left part of the inequality (10) we have

$$\begin{aligned}
\psi''(t) &\leq \frac{2\lambda_f^2(p_k)}{\left(1 - t\lambda_f(p_k)\right)^2} \left[\nabla^2_X f(\exp_{p_k} tX_N(p_k))\right]^{1/2} \\
&\leq \frac{2\lambda_f^2(p_k)}{\left(1 - t\lambda_f(p_k)\right)^2} \frac{1}{1 - t\lambda_f(p_k)} [\nabla^2_X f(p_k)]^{1/2} \\
&= \frac{2\lambda_f^2(p_k)}{\left(1 - t\lambda_f(p_k)\right)^3} [\nabla^2_X f(p_k)]^{1/2}.
\end{aligned}$$

Let $\frac{1}{1+\lambda_f(p_k)}$ be denoted as r. Then,

$$\begin{aligned}
\psi(r) &= \psi(0) + \int_0^r \psi'(\tau)\mathrm{d}\tau \\
&= \psi(0) + \int_0^r \{\psi'(0) + \int_0^\tau \psi''(t)\mathrm{d}t\}\mathrm{d}\tau \\
&\leq \nabla_X f(p_k) + r\nabla_{X_N}\nabla_X f(p_k) \\
&\quad + [\nabla_X^2 f(p_k)]^{1/2} \int_0^r \int_0^\tau \frac{2\lambda_f^2(p_k)}{(1 - t\lambda_f(p_k))^3}\mathrm{d}t\,\mathrm{d}\tau \\
&= \nabla_X f(p_k) + r\nabla_{X_N}\nabla_X f(p_k) + [\nabla_X^2 f(p_k)]^{1/2}\frac{\lambda_f^2(p_k)r^2}{1-\lambda_f(p_k)r}.
\end{aligned}$$

By using the (16) of the Newton decrement in Theorem 1, the last inequality results in:

$$\begin{aligned}
\psi(r) &\leq (1-r)\nabla_X f(p_k) + \left[\nabla_X^2 f(p_k)\right]^{1/2}\frac{\lambda_f^2(p_k)r^2}{1-\lambda_f(p_k)r} \\
&\leq (1-r)\lambda_f(p_k)\left[\nabla_X^2 f(p_k)\right]^{1/2} + \left[\nabla_X^2 f(p_k)\right]^{1/2}\frac{\lambda_f^2(p_k)r^2}{1-\lambda_f(p_k)r} \\
&= \frac{2\lambda_f^2(p_k)}{1+\lambda_f(p_k)}\left[\nabla_X^2 f(p_k)\right]^{1/2},
\end{aligned}$$

where $[\nabla_X^2 f(p_k)]^{1/2}$ can be estimated by the left part of the inequality (10). That is:

$$\begin{aligned}
\psi(r) &\leq \frac{2\lambda_f^2(p_k)}{1+\lambda_f(p_k)}\frac{1}{1-r\lambda_f(p_k)}[\nabla_X^2(\exp_{p_k} rX_N(p_k))]^{1/2} \\
&= 2\lambda_f^2(p_k)[\nabla_X^2(\exp_{p_k} rX_N(p_k))]^{1/2}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\lambda_f(p_{k+1}) &= \max_{X\in \mathrm{T}M(p_{k+1})} \frac{\nabla_X f(p_{k+1})}{\left[\nabla_X^2 f(p_{k+1})\right]^{1/2}} \\
&= \max_{X\in \mathrm{T}M(p_{k+1})} \frac{\psi(r)}{\left[\nabla_X^2 f(\exp_{p_k} rX_N(p_k))\right]^{1/2}} \\
&\leq 2\lambda_f^2(p_k).
\end{aligned}$$

The proof is complete.

It is clear that $\lambda_f(p_{k+1}) < \lambda_f(p_k)$ if $\lambda_f(p_k) < \frac{1}{2}$. It can also be proved by simple analysis that

$$-\lambda - \ln(1-\lambda) < \lambda, \qquad \forall \lambda \in (0, 1/2).$$

Remark 1 Under the condition that $\lambda(p) < 1, p \in M$ and starting from an admissible initial point, the convergence results of the proposed damped Newton algorithm can be summarized as:

(1) For not more than $\frac{f(p_0)-f(p^\star)}{1/2-\ln(3/2)}$ steps $\lambda_f(p_k)$ will fall into the interval $(0, 1/2)$.
(2) $\lambda_f(p_k)$ will monotonically and quadratically converge to zero starting from any point p_{k_0} such that $\lambda_f(p_{k_0}) \in (0, 1/2)$.
(3) $[\nabla^2_{X_{pp^\star}} f(p)]^{1/2} \leq 2\lambda_f(p), \quad \forall \lambda_f(p) \in (0, 1/2)$.
(4) $0 \leq f(p) - f(p^\star) \leq \lambda_f(p), \quad \forall \lambda_f(p) \in (0, 1/2)$.
(5) For a given precision $\epsilon \in (0, 1/2)$, the maximal number of iterations such that $\lambda_f(p_k) \leq \epsilon$, is not more than

$$\min\left\{\frac{f(p_0)-f(p^\star)}{\epsilon - \ln(1+\epsilon)}, \min_{\alpha\in(0,1/2)}\left(\frac{f(p_0)-f(p^\star)}{\alpha-\ln(1+\alpha)} + \frac{\ln\epsilon}{\ln\alpha}\right)\right\}.$$

This amount is then bounded by

$$\frac{f(p_0)-f(p^\star)}{1/2-\ln(3)+\ln(2)} + \frac{-\ln\epsilon}{\ln(2)}.$$

6 Illustrative examples

This section includes two simple examples to illustrate the proposed damped Newton method:

Example 1

Consider the following optimization problem:

$$\begin{aligned} \text{minimize} \quad & f(x) := x_1 + x_2, \\ \text{subject to:} \quad & x_1, x_2 > 0, x = (x_1, x_2)^\top \in H, \end{aligned}$$

where H is a hyperbola satisfying $x_1 x_2 = 1$. The Riemannian metric is defined as the induced metric from the ambient Eucleadean space. Let $\mathrm{T}_x H$ be the tangent space of H at x. i. e., $\mathrm{T}_x H = \{\alpha h | h = (-x_1, x_2), \alpha \in R\}$. Then, the geodesic on the hyperbola can be calculated as:

$$\exp_x tX = Ax, \tag{32}$$

where $\mathrm{A} = \begin{pmatrix} e^{-\alpha t} & 0 \\ 0 & e^{\alpha t}\end{pmatrix}$ and $X \in T_x H$. Hence, covariant differentials can be calculated as follows:

$$\begin{aligned} \nabla_X f(x) &= (-x_1 + x_2)\alpha, \\ \nabla_h^2 f(X) &= (x_1 + x_2)\alpha^2, \\ \nabla_h^3 f(X) &= (-x_1 + x_2)\alpha^3. \end{aligned}$$

It can be seen that $\nabla_X^2 f(x)$ is positive definite. Then

$$\frac{\left(\nabla_X^3 f(x)\right)^2}{\left(\nabla_X^2 f(x)\right)^3} = \frac{(x_2 - x_1)^2}{(x_1 + x_2)^3} \leq 1. \tag{33}$$

As such, $f(x)$ is a self-concordant function.

Now we can apply the proposed damped Newton algorithm.

Algorithm 2 (*Damped Newton algorithm*)
step 0: Randomly generate a feasible initial point x^0.
step k: Calculate the kth step according to:

$$x^k = \exp_{x^{k-1}}\left(\frac{1}{1+\lambda(x^{k-1})}X_N\right) = A^{k-1}x^{k-1},$$

where

$$A^{k-1} = \begin{pmatrix} e^{-\frac{1}{1+\lambda(x^{k-1})}\alpha^{k-1}} & 0 \\ 0 & e^{\frac{1}{1+\lambda(x^{k-1})}\alpha^{k-1}} \end{pmatrix},$$

$$\lambda(x^{k-1}) = \sqrt{(x_1^{k-1} + x_2^{k-1})\alpha^2}$$

$$X_N = \alpha^{k-1}\left(-x_1^{k-1}, x_2^{k-1}\right)^T, \quad \alpha^{k-1} = -\frac{x_2^{k-1} - x_1^{k-1}}{x_1^{k-1} + x_2^{k-1}}.$$

The simultaion result is shown in Table 1.

Example 2

Consider the following optimization problem:

$$\min f(x) := -\ln x_1 - \cdots - \ln x_n, \qquad \text{subject to} : \begin{cases} x = (x_1, x_2, \ldots, x_n) \in S^{n-1}, \\ 0 < x_1, \ldots, x_n < 1, \end{cases} \tag{34}$$

where S^{n-1} is the unit sphere with $x^T x = 1$. Here, we define a Riemannian metric as the induced metric from the ambient Euclidean space, i.e., $< y, z >= y^T z$ where

Table 1 The simulation result for Example 1

Step k	x	$f(x)$	$\lambda(x)$
0	(6.0000, 0.1667)	6.1667	2
1	(0.2525, 3.9601)	4.2126	2.3490
2	(2.9852, 0.3350)	3.3202	1.8064
3	(0.4208, 2.3762)	2.7971	1.4545
4	(1.9173, 0.5216)	2.4389	1.1692
5	(0.6487, 1.5417)	2.1903	0.8938
6	(1.2471, 0.8018)	2.0490	0.6034
7	(0.9379, 1.0662)	2.0041	0.3111
8	(1.0057, 0.9943)	2.0000	0.0906
9	(1.0000, 1.0000)	2.0000	0.0081
10	(1.0000, 1.0000)	2.0000	0.0001
11	(1.0000, 1.0000)	2.0000	0.0000

$y, z \in \mathrm{T}_x S^{n-1}$. Let $x \in S^{n-1}$ and $h = (h_1, h_2, \ldots, h_n) \in \mathrm{T}_x S^{n-1}$ have unit length (if not, we can normalize it). Then, the geodesic on the sphere is $\exp_x th = x \cos(t) + h \sin(t)$, and the parallel transportation along the geodesic $\tau h = h \cos(t) - x \sin(t)$. Therefore, the following covariant differentials can be calculated:

$$\nabla_h f(x) = -\frac{h_1}{x_1} - \frac{h_2}{x_2} -, \ldots, -\frac{h_n}{x_n},$$
$$\nabla_h^2 f(x) = \frac{h_1^2}{x_1^2} + \frac{h_2^2}{x_2^2} +, \ldots, + \frac{h_n^2}{x_n^2} + n,$$
$$\nabla_h^3 f(x) = -2\left(\frac{h_1^3}{x_1^3} + \frac{h_1}{x_1} + \frac{h_2^3}{x_2^3} + \frac{h_2}{x_2} +, \ldots, + \frac{h_n^3}{x_n^3} + \frac{h_n}{x_n}\right)$$

The following procedure is to prove that the function f is self-concordant defined on S^{n-1}.

It can be easily checked that for any $h \in \mathrm{T}_x S^{n-1}$, the second covariant differentials $\nabla_h^2 f(x)$ are always positive.

Let $y_1 = \frac{h_1}{x_1}, \ldots, y_n = \frac{h_n}{x_n}, y = (y_1, y_2, \ldots, y_n)^T, b = ((y_1^2+1), (y_2^2+1), \ldots, (y_n^2+1))^T$. Then, we have

$$\begin{aligned}
M_f &= \frac{\left(\nabla_h^3 f(x)\right)^2}{(\nabla_h^2 f(x))^3} \\
&= \frac{4\left(y_1^3 +, \ldots, + y_n^3 + y_1 + \cdots + y_n\right)}{\left(y_1^2 + \cdots + y_n^2 + n\right)^3} \\
&= 4\frac{(y_1(y_1^2+1) + \cdots + y_n(y_n^2+1))^2}{((y_1^2+1) + \cdots + (y_n^2+1))^3} \\
&\le \frac{4\|y\|^2\|b\|^2}{\left((y_1^2+1) + \cdots + (y_n^2+1)\right)^2\left((y_1^2+1) + \cdots + (y_n^2+1)\right)} \\
&\le \frac{4(y_1^2 + \cdots + y_n^2)}{(y_1^2+1) + \cdots + (y_n^2+1)} \\
&\le 4.
\end{aligned} \tag{35}$$

Therefore, the function f is self-concordant function defined on S^{n-1}.

We apply the damped Newton algorithm on this problem. In particular, $n = 10$. Figure 1 shows the results of the damped Newton method on function f, where $x^\star$ is the optimal solution and the logarithm function is 10-based.

The result clearly demonstrates the quadratic convergence of the proposed algorithm.

To illustrate the advantage of the proposed algorithms on manifold, this optimization problem is also solved in Euclidean space using the Lagrangian multiplier method [13], as a comparison example. Given the Lagrangian multiplier $\alpha \in R$, the Lagrangian form of the original problem is

$$L(x, \alpha) = -\ln x_1 - \cdots - \ln x_n + \alpha(1 - x^T x). \tag{36}$$

Then, Newton method can be applied in Euclidean space to find the critical point (x^*, α^*) of L. However, because the parametrization of problem in Euclidean space

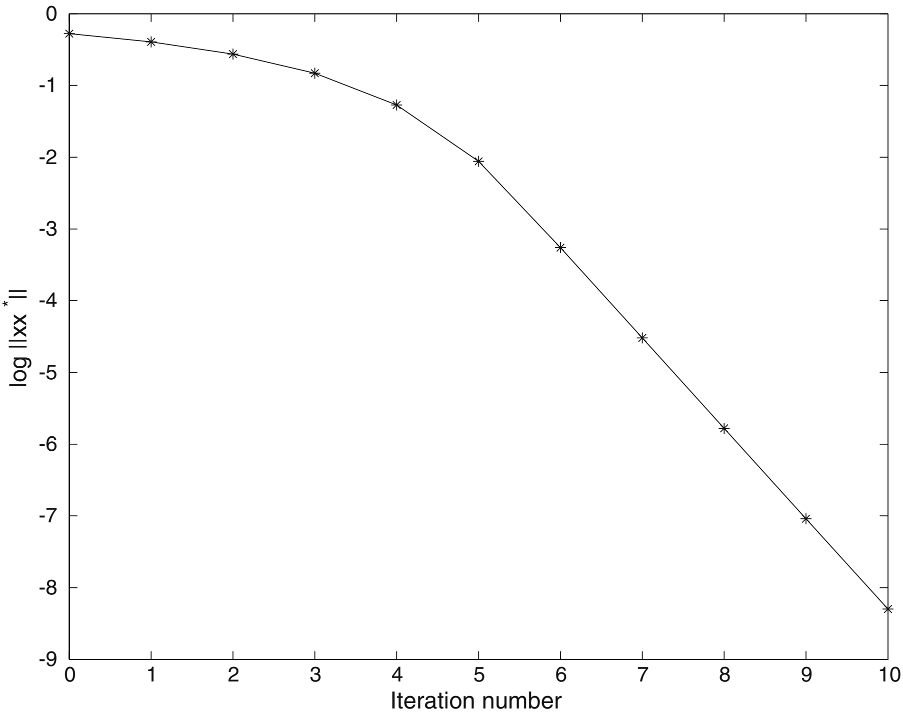

Fig. 1 The results of damped Newton method of Example 2

may not necessarily reflect its intrinsic global property, the efficiency and convergence of the Newton method will depend on the choice of initial points, which cannot be guaranteed. Figure 2 shows the simulation result obtained starting from a randomly generated initial point. The performance is inferior to that shown in Fig. 1.

7 Conclusions

This paper reports our effort to generalize the definition and known results for self-concordant functions in Euclidean space to manifolds. It lays a comparative solid foundation to facilitate the construction of barrier functions for interior-point algorithms on manifolds.

For the proposed self-concordant function defined on a general class of smooth manifolds, a number of desirable properties are obtained. These include the feasibility of a Dikin-type ellipsoid and several inequalities that characterizes the similarity between self-concordant functions and quadratic functions along the geodesics of the manifold in various inequalities. Under the convexity condition on manifold defined by second-order covariant differential, it is also shown that the optimal solution is global.

A Newton decrement is defined for this specific class of functions. This concept is analyzed in regard to the relationship between first-order covariant derivatives along Newton direction and along general direction, and to the maximal ratio of the norm of first-order covariant derivative and that of second-order derivative. The later facilitate

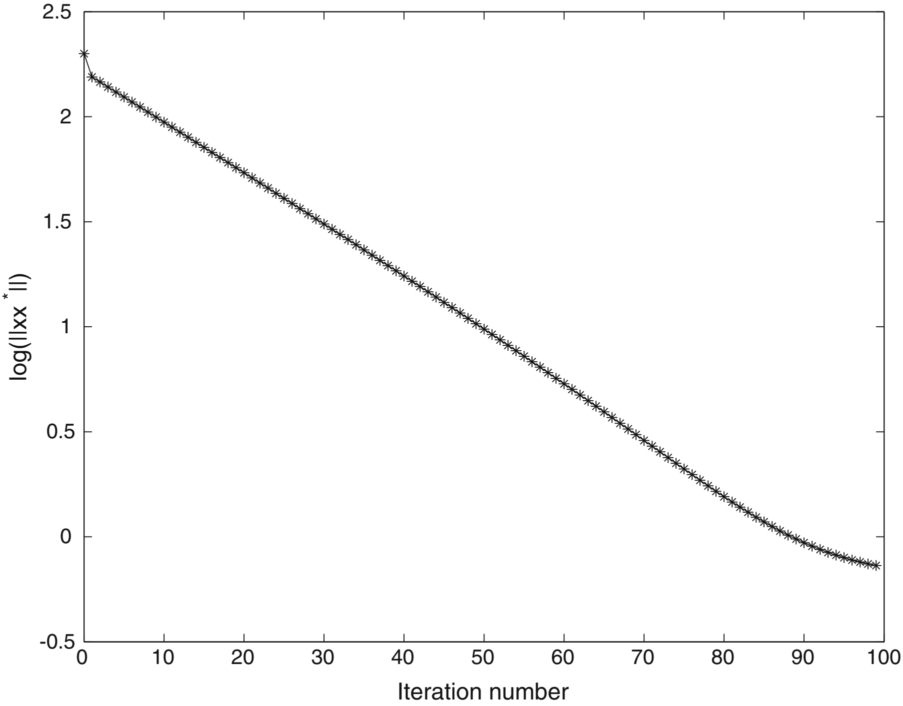

Fig. 2 The results of Newton method in Euclidean space for Example 2

the definition of the index $\lambda_f(p)$. With those theoretical preparation, the existence of global optimal solution is shown when $\lambda_f(p) < 1$ holds for a point p.

A damped Newton algorithm is proposed. Its computational complexity is carefully analyzed and precise bound is shown to be $O(-\ln(\epsilon))$.

Two simple but meaningful optimization problems are given to illustrate the efficiency of the proposed concept and algorithm.

Acknowledgements The authors are very grateful for the valuable comments given by anonymous reviewers. The first author would like to thank Dr. Jochen Trumph for valuable discussions and suggestions. His research was Partially supported by the IRGS Grant J0015016 from UTAS. John B. Moore was Partially supported by an ARC Large Research Grant A00105829. Huibo Ji's work was in part supported by National ICT Australia.

References

1. Nesterov, Y., Nemirovskii, A.: Interior-Point polynomial algorithms in convex programming. Studies in Applied Mathematics, vol. 13. SIAM, Philadelphia, PA (1994)
2. Glineur, F.: Improving complexity of structured convex optimization problems usiing self-concordant barriers. Eur. J. Oper. Res. **143**, 291–310 (2002)
3. Faybusovich, L.: Infinite-dimensional semidefinite programming: regularized determinants and self concordant barriers. Fields Inst. Commun. **8**, 39–49 (1998)

4. Nesterov, Y., Todd, M.: On the Riemannian geometry defined by self-concordant barriers and interior-point methods. Found. Comput. Math. **2**, 333–361 (2002)
5. Helmke, U., Moore, J.B.: Optimization and Dynamical Systems. Springer-Verlag, London (1994)
6. Smith, S.: Optimization techniques on Riemannian manifolds. Fields Inst. Commun. **3**, 113–146 (1994)
7. Edelman, A., Arias, T., Smith, S.: The geometry of algorithms with orthogonality constraints. SIAM J. Matrix Anal. Appl. **20**, 303–353 (1998)
8. Yan, W., Helmke, U., Moore, J.B.: Global analysis of Oja's flow for neural networks. IEEE Trans. Neural Netw. **5**, 674–683 (1993)
9. Addler, R.L., Dedieu J-P, M., Joesph, Y., Martens, M., Shub, M.: Newton's method on riemannian manifolds and a geometric model for the human spine. IMA J. Numer. Anal. **22**, 359–390 (2002)
10. Udriste, C.: Optimization methods on Riemannian manifolds. Algebera, Groups Geom. **14**, 339–359 (1997)
11. Helgason, S.: Differential Geometry, Lie Groups, and Symmetric Spaces. Academic Press, New York (1978)
12. Nesterov, Y.: Introductory Lectures on Convex Optimization: A Basic Course. Kluwer Academic Publishers Dordrecht, 2004
13. Bertsekas, D.: Nonlinear Programming, 2nd edn. Athena Scientific, Belmonk, MA (1999)