43rd IEEE Conference on Decision and Control
December 14-17, 2004
Atlantis, Paradise Island, Bahamas

Self-Concordant Functions for Optimization on Smooth Manifolds

Danchi Jiang, John B. Moore, and Huibo Ji

***Abstract*— This paper discusses self-concordant functions on smooth manifolds. In Euclidean space, this class of functions are utilized extensively in interior-point methods for optimization because of the associated low computational complexity. Here, the self-concordant function is carefully defined on a differential manifold. First, generalizations of the properties of self-concordant functions in Euclidean space are derived. Then, Newton decrement is defined and analyzed on the manifold that we consider. Based on this, a damped Newton algorithm is proposed for optimization of self-concordant functions, which guarantees that the solution falls in any given small neighborhood of the optimal solution, with its existence and uniqueness also proved in this paper, in a finite number of steps. It also ensures quadratic convergence within a neighborhood of the minimal point. This neighborhood can be specified by the the norm of Newton decrement. The computational complexity bound of the proposed approach is also given explicitly. This complexity bound is $O(-\ln(\epsilon))$, where ϵ is the desired precision. An interesting optimization problem is given to illustrate the proposed concept and algorithm.**

I. Introduction

Self-concordant functions play an important role in the powerful interior-point polynomial algorithms for convex programming in Euclidean space. Following the work of Nesterov and Nemirovskii [1], many articles have been published using this type of functions to construct the barrier functions for inter-point algorithms. For example, see [2], [3], [4]. 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 to 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. One of the advantages of using self-concordant 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 a polynomial time for a given precision.

This research was supported by NICTA, SEACS program

Danchi Jiang is with National ICT Australia Limited, Locked Bag 8001, Canberra ACT 2601, Australia, and jointly associated with the Dept. of Systems Engineering, RSISE, Australian National University. Email: danchi.jiang@nicta.com.au

John B. Moore is with Dept. of Systems Engineering, RSISE, Australian National University. This author is also associated with National ICT Australia Limited, Locked Bag 8001, Canberra ACT 2601, Australia. Email: John.Moore@anu.edu.au

Huibo Ji is with the Dept of Systems Engineering, RSISE, Australian National University. Email: Huibo@syseng.anu.edu.au.

The notion of self-concordant function has a deep root in geometry. In [5], 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 usually can be better understood on Riemannian manifolds. See [6] for many meaningful examples.

In fact, many optimization problem can also be better posed on manifolds rather than Euclidean space. For example, optimization problems associated with orthogonal matrices, algorithms for computation of eigenvalues and singular values, or Oja's flow in neural network pattern classification algorithms, see [6], [7], [8], [9], 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 [6], [7]. It is natural to ask, what are the self-concordant functions and associated interior-point methods on manifolds? Such a question can be easily justified by practical importance or even by theoretical completeness. The self-concordant concept has been briefly given in [10] with analysis restricted to a logarithm function. In this paper, we will give a comparatively thorough study on general self-concordant functions on manifolds.

One of the advantages of solving minimization problems on manifolds is, as pointed out in [7], to take the advantages of the intrinsic property of the constraints but the problems themself are still locally equivalent to constrained optimization problems in Euclidean spaces. Such an approach reduces the dimensions of the problems, compared to solving the original problems in their ambient Euclidean spaces.

The intrinsic approach for minimization is based on the computation on geodesic and covariant differentials. Even though the computation with respect to the intrinsic geometry sometimes might be expansive, 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 transformation can be computed by matrix exponentiation. See [7], [11]. [6] contains many particular classes of minimization problems in this category such as those on Orthogonal groups or Stiefel Manifolds. Another simple but non-trivial case is the sphere, where the geodesic and parallel transformation can be computed via trigonometric functions and vector calculation.

This paper is organized as follows: we will provide some preliminary material in Section II. Then, the self-concordant function is carefully defined to preserve as many nice properties of self-concordant functions in Euclidean space as possible in Section III. To facilitate the analysis and understanding, the Newton decrement is defined and analyzed in Section IV. Then, the existence and uniqueness of the optimal solution are proved and a damped Newton algorithm is proposed in Section IV. 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]. An interesting example is included to illustrate the proposed concept and approach in this paper in the last section.

II. Preliminaries and Notations

In this section, some concepts, techniques, and results from differential geometry will be given. However, it should be considered neither self-contained nor complete. Proofs and some concepts used will not be included. Please refer to [11] for more details.

Let a smooth manifold be denoted as M. The differential structure of M is a set of local charts covering M. Each local chart is a pair of a neighborhood and a smooth mapping from this neighborhood to an open set in Euclidean space. The local charts satisfy some regular properties. The tangent space of M at a point p can be denoted as $T_M(p)$. It is the set of linear mappings from all functions passing through the point p to real numbers, satisfying the derivative condition. The set of its vector fields is denoted as T_M.

Let $T \in T_M$. Then $T(p)$ is the value (a tangent vector) of the vector field T at p. The dual of tangent vector fields is denoted as $T_M^\star$. Any element of it is called a differential 1-form. Locally, both tangent vectors $T_M(p)$ and space of differential 1-forms $T_M^\star(p)$ are vector spaces and are isospectral to $\Re^n$, where n is the dimension of the manifold.

An affine connection on a manifold M is a rule that assigns a vector field a linear mapping from T_M to itself. Let an affine connection be denoted as ∇. This linear mapping assigned to X, where $X \in T_M$, satisfies the following conditions:

$$\begin{aligned} \nabla_{fX+gY} &= f\nabla_X + g\nabla_Y; \\ \nabla_X(fY) &= f\nabla_X(Y) + (Xf)Y, \\ &\quad \text{where } f, g \in C^1(M), \text{ and } X, Y \in T_M. \end{aligned}$$

The mapping ∇_X is called a covariant differential. Adopting the convention for derivatives, we also simplify $\nabla_X\nabla_X$ as ∇_X^2. In the case where ∇ is symmetric, the Hessian matrix associated with ∇_X^2 in local coordinates is also symmetric. There always exists a symmetric connection associated with the Riemannian metric for a Riemannian manifold. This connection is also termed a Riemannian connection.

A curve $\gamma(t)$ on M defines a vector field by its derivative with respect to t. A family of vector field $Y(t)$ is said to be parallel with respect to $\gamma(t)$ if

$$\nabla_{\dot\gamma}(Y(\gamma(t)) = 0, \quad \forall t \in J \subseteq \Re.$$

A curve γ is a geodesic if the $\dot\gamma$ is parallel with respect to γ itself.

For any vector field X at a point p, there uniquely exists a geodesic $\gamma(t)$ such that $\gamma(0) = p$, and $\dot\gamma(0) = X(p)$. The mapping $X \to \gamma(1)$ is defined as an exponential map, if it exists. It is denoted as exp_pX or $exp_p(X)$. A manifold is geodesic complete if exp_pX is defined for every tangent vector at every point on the manifold. See [13].

With parallelism and geodesic, the covariant differential can also be defined as:

$$\nabla_X(W(p)) = \lim_{t\to 0} \frac{\tau_t^{-1}W(exp_p(tX)) - W(p)}{t}, \quad (1)$$

where τ_t is the parallel transformation from p to $exp(tX)$ along the $exp(tX)$, and W can be a vector field or a function. The covariant differential defined by (1) can also be extended to the cases where W is any tensor field.

A Lie group G is a manifold as well a group with the group action $\alpha\beta^{-1}$ being diffeomorphism. Given a manifold and a Lie group, if the group action on the manifold is an isomorphism, this manifold is called a homogenous space. Riemannian matric or manifold topology on a homogenous space can be induced from the Lie group.

III. Self-Concordant Functions

Let M denote a smooth manifold of finite dimension and ∇ an affine symmetric connection defined on M. Assume that M is geodesic complete in this paper. 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:

$$\mid \nabla_X^3 f(p) \mid \leq M_f[\nabla_X^2 f(p)]^{3/2}, \forall X \in T_M(p), p \in M, \quad (2)$$

where M_f is a positive constant associated with f.

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

$$\begin{aligned} & \mid \nabla_{X_1}\nabla_{X_2}\nabla_{X_3} f(p) \mid \\ \leq\ & M_f[\nabla_{X_1}^2 f(p)]^{1/2}[\nabla_{X_2}^2 f(p)]^{1/2}[\nabla_{X_3}^2 f(p)]^{1/2}, \\ & \forall X_1, X_2, X_3 \in T_M(p). \end{aligned} \quad (3)$$

This property comes from the linearity of the mapping

$$\nabla_{df}^2 : T_M(p) \times T_M(p) \to R, \text{ and}$$
$$\nabla_{df}^3 : T_M(p) \times T_M(p) \times T_M(p) \to R,$$

defined by

$$\nabla_{df}^2(X_1, X_2) := \nabla_{X_1}\nabla_{X_2} f(P), \text{ and}$$
$$\nabla_{df}^3(X_1, X_2, X_3) := \nabla_{X_1}\nabla_{X_2}\nabla_{X_3} f(p),$$

respectively, for any $X_1, X_2, X_3 \in T_M(p)$. The proof in ([1], Proposition 9.1) can be adopted here to prove Property

1. Also noticed in [1], if a function f is self-concordant with the constant M_f, then, the function $M_f^2 f$ is self-concordant with the constant 1. As such, we assume $M_f = 2$ for the rest of this paper.

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 variables 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 1:

$$\nabla_X^2 f(p) > 0, \quad \forall p \in dom(f), X \in T_M(p). \tag{4}$$

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 $p \in dom(f)$, and $r > 0$,

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

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

A self-concordant function has the following interesting property:

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

We need the following lemma in order to prove this proposition:

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 $\phi(t) : \Re \to R$ as follows:

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

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

(1). $\|\phi'(t)\| \leq 1$;

(2). If $\phi(0) > 0$, then, $(-\phi(0), \phi(0)) \subseteq dom(\phi)$.

Proof: Since

$$\begin{aligned}\phi'(t) &= -\frac{\frac{d}{dt}[\nabla^2_{\tau_{pexp_p(tX)}X} f(exp_p tX)]}{2[\nabla^2_{\tau_{pexp_p(tX)}X} f(exp_p tX)]^{3/2}} \\ &= -\frac{\nabla^3_{\tau_{pexp_p(tX)}X} f(exp_p tX)}{2[\nabla^2_{\tau_{pexp_p(tX)}X} f(exp_p tX)]^{3/2}},\end{aligned}$$

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

Since $\phi(0) > 0$, because of the continuity of $\nabla^2_{\tau_{pexp_p(tX)}} f(exp_p tX)$, 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 $dom(\phi)$. 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

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

Hence,

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

The existence of $f(exp_p \bar{t} X)$ comes from the assumption that f has a closed map and the fact that $f(exp_p tX) = \int_0^t [\int_0^s \nabla^2_{\tau_{pexp_p(\nu X)}} f(exp_p \nu X) d\nu + \nabla_X f(X)] ds < +\infty$. Therefore, $\phi(\bar{t})$ is well-defined, which is contradiction to the assumption we made. As such, (2) holds. ∎

Proof of Property 2: Notice, from **Lemma 1**, that for any vector field $X \in T_M$, $\{exp_p(tX) \mid \quad \mid t \mid^2 < \phi^2(0)\} \subseteq dom(f)$. On the other hand,

$$\begin{aligned}\{exp_p(tX) \mid\mid t \mid^2 < \phi^2(0)\} &= \{exp_p(tX) \mid\mid t \mid^2 < \frac{1}{\nabla_X^2 f(p)}\} \\ &= \{exp_p tX \mid\mid t \mid^2 (\nabla_X^2 f(p)) < 1\} = W^\circ(p;1).\end{aligned}$$

This proves the **Property 2** . ∎

Property 3: For any $p, q \in dom(f)$, 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:

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

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

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

The proof to this property is omitted to save space. Interested reader may refer to [14] or ask the authors for details.

Property 4: For any $p \in dom(f)$, $q \in W^\circ(p;1)$, and $X \in T_M$, there holds:

$$\begin{aligned}&(1 - [\nabla^2_{X_{pq}(p)} f(p)]^{1/2})^2 \nabla^2_{X(p)} f(p) \leq \nabla^2_{X(q)} f(q) \\ &\leq \frac{\nabla^2_{X(p)} f(p)}{(1 - [\nabla^2_{X_{pq}(p)} f(p)]^{1/2})^2},\end{aligned} \tag{9}$$

$$\nabla_{X_{pq}(q)} f(q) - \nabla_{X_{pq}(p)} f(p) \leq \frac{\nabla^2_{X_{pq}(p)} f(p)}{1 - \nabla^2_{X_{pq}(p)} f(p)} \tag{10}$$

$$\begin{aligned}&f(q) \leq f(p) + \nabla_{X_{pq}(p)} f(p) - [\nabla^2_{X_{pq}(p)} f(p)]^{1/2} \\ &- \ln(1 - [\nabla^2_{X_{pq}(p)} f(p)]^{1/2}).\end{aligned} \tag{11}$$

The proof to this property is omitted to save space. Interested reader may refer to [14] or ask the authors for details.

IV. Newton Decrement

Consider the following auxiliary quadratic cost defined on $T_M(p)$:

$$N_{f,p}(X) := f(p) + \nabla_X f(p) + \frac{1}{2} \nabla_X^2 f(p). \tag{12}$$

Definition 3: The Newton decrement $X_N(f,p)$ is defined as the minimal solution to the auxiliary cost function given by (12). More specifically,

$$X_N(f,p) := \arg\min_{X\in T_M(p)} N_{f,p}(X). \tag{13}$$

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 $f : M \to R$ be a self-concordant function, p, a given point on the manifold M, and X_N, its Newton decrement defined at p. The following results hold:

$$\nabla_{X_N}\nabla_X f(p) = -\nabla_X f(p), \quad \forall X\in T_M(p), \tag{14}$$

$$\sqrt{\nabla^2_{X_N} f(p)} = \max\{\nabla_X f(p) | X\in T_M(p), \nabla^2_X f(p)\le 1\}. \tag{15}$$

Furthermore, if the bilinear map $\nabla^2_\mu : T_M(p)\to T^\star_M(p)$, defined by:

$$\nabla^2_\mu(X) = \nabla^2_X\mu, \quad \forall X\in T_M(p),$$

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

$$X_N = -(\nabla^2_{df})^{-1}df. \tag{16}$$

Proof: Since p is a given point on the manifold M, the claimed results can be converted into their local representation in Euclidean space. By such way this result can be proved. Detsils are omitted because of length limitation. Interested reader may refer to [14] or ask the authors for details. ■

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

$$\lambda_f(p) := \max_{X\in T_M(p)} \frac{|\nabla_X f(p)|}{\sqrt{\nabla^2_X f(p)}}, \quad \text{for } p\in dom(f). \tag{17}$$

If $\lambda_f(p) < 1$ for some $p\in dom(f)$, then there exists a unique point $p^\star_f \in dom(f)$ such that

$$f(p^\star_f) = \min\{f(p) \mid p\in dom(f)\}.$$

Proof is omitted to save space. Interested readers may refer to [14] or ask the authors for details.

Consider the following damped Newton method:

Algorithm 1: **(Damped Newton Algorithm)**

Step 0: find a feasible point $p_0\in dom(f)$.

Step k: $p_k = exp_{p_{k-1}} \frac{1}{1+\lambda_f(p_{k-1})} X_N$,

where $exp_{X_N(f,p_{k-1})} f(p_{k-1})$ 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 $p^\star_f$, and p is any admissible point.

(1). The following inequality holds:

$$[\nabla^2_{X_{pp^\star_f}} f(p)]^{1/2} \le \frac{\lambda_f(p)}{1-\lambda_f(p)}. \tag{18}$$

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

$$0 \le f(p) - f(p^\star_f) \le -\lambda_f(p) - \ln(1-\lambda_f(p)). \tag{19}$$

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

$$\begin{aligned} f(p^\star_f) \quad &\le \quad f(p_k) \le f(p_{k-1}) - (\lambda_f(p_{k-1}) \\ &\quad - \ln(1+\lambda_f(p_{k-1}))). \end{aligned} \tag{20}$$

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

$$\nabla_{X_{pp^\star_f}} f(p) \ge \frac{r^2(p)}{1+r(p)}. \tag{21}$$

On the other hand, there holds

$$\nabla_{X_{pp^\star_f}} f(p) \le \lambda_f(p) r(p),$$

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 (18).

(2). Based on (8) and the inequality (18) obtained above, one has:

$$\begin{aligned} f(p^\star_f) - f(p) &\ge \nabla_{X_{pp^\star_f}} f(p) + r(p) - \ln(1+r(p)) \\ &\ge r(p) - \ln(1+r_f(p)) - \lambda_f(p) r(p). \end{aligned} \tag{22}$$

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

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

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

The solution to this group of equations satisfies

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

As such, at the minimal point there holds:

$$g(x_0,y_0) = x_0 - x_0y_0 + y_0 = x_0(1-y_0) + y_0 > 0,$$

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

$$g(x,y)\ge 0, \quad \forall x\ge 0, 1 > y\ge 0.$$

Applying this inequality to (22), we obtain the right side of the inequality (19).

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

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

Applying (11), there holds

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

by the definition of $\lambda_f(p_k)$ and the results in **Theorem 1**. Hence, the inequality (20) is proved. ■

Notice that the two functions

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

and

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

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 (20) implies the following results:

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

$$\lambda_f(p_k) - \ln(1 + \lambda_f(p_k)) \leq f(p_k) - f(p_f^\star). \tag{23}$$

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 $\frac{f(p_0) - f(p_f^\star)}{\epsilon - \ln(1+\epsilon)}$.

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:

$$\lambda_f(p_{k+1}) \leq 2\lambda_f^2(p_k). \tag{24}$$

The proof to this result is omitted because of length limitation. Interested reader may refer to [14] or ask the authors for details.

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

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

Remark 1: The convergence results of the proposed damped Newton algorithm can be summarized as:

(1). For not more than $\frac{f(p_0) - f(p_f^\star)}{1/2 - \ln(3/2)}$ 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). $[\nabla^2_{X_{pp_f^\star}} f(p)]^{1/2} \leq 2\lambda_f(p), \quad \forall \lambda_f(p) \in (0, 1/2)$.

(4). $0 \leq f(p) - f(p_f^\star) \leq \lambda_f(p), \quad \forall \lambda_f(p) \in (0, 1/2)$.

(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

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

This amount is then bounded by

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

V. An Illustrative Example

Consider the following simple optimization problem:

$$\begin{aligned} \min \quad & f(x) := -ln(x_1 x_2) \\ \text{subject to:} \quad & x_1, x_2 > 0, x = (x_1, x_2) \in S^1, \end{aligned}$$

where S^1 is unit circle. We define a Riemannian metric as the induced metric from the ambient Euclidean space. Let $x \in S^1$ and $h \in T_x S^1$ have unit length. Then the geodesic on the unit circle is $\exp_x th = xcos(t) + hsin(t)$. Hence, the following covariant differentials can be calculated:

$$\begin{aligned} \nabla_h f(x) &= -\frac{h_1}{x_1} - \frac{h_2}{x_2} \\ \nabla^2_h f(x) &= \frac{h_1^2}{x_1^2} + \frac{h_2^2}{x_2^2} + 2 \\ \nabla^3_h f(x) &= -(\frac{h_1^3}{x_1^3} + \frac{h_1}{x_1} + \frac{h_2^3}{x_2^3} + \frac{h_2}{x_2}) \end{aligned}$$

It is obvious that $\nabla^2_h f(x)$ is positive definite.
Let $x \in S^1$ and $h \in T_x S^1$ and $\|h\| = 1$. Notice that $h = (-x_2, x_1)$ or $h = (x_2, -x_1)$. Therefore,

$$\frac{(\nabla^3_h f(x))^2}{(\nabla^2_h f(x))^3} = \frac{4(\frac{h_1^3}{x_1^3} + \frac{h_1}{x_1} + \frac{h_2^3}{x_2^3} + \frac{h_2}{x_2})^2}{(\frac{h_1^2}{x_1^2} + \frac{h_2^2}{x_2^2} + 2)^3} \leq 4. \tag{25}$$

As such, f(x) is self-concordant function. Now the damped Newton algorithm proposed in this paper becomes:

Algorithm 2: (Damped Newton Algorithm)

step 0: randomly generate a feasible initial point x_0.

step k: calculate the $k-$th step according to:

$$x^k = x^{k-1} cos(\frac{\|X_N\|}{1+\lambda(x^{k-1})}) + \frac{X_N}{\|X_N\|} sin(\frac{\|X_N\|}{1+\lambda(x^{k-1})}),$$

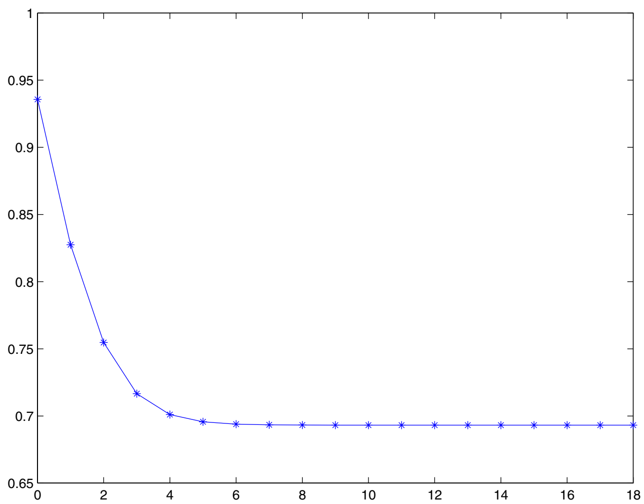
where

$$X_N = (\frac{x_1 x_2^2 (x_1^2 - x_2^2)}{x_2^4 + x_1^2 x_2^4 + x_1^4(1+x_2^2)}, \frac{-x_1^4 x_2 + x_1^2 x_2^3}{x_2^4 + x_1^2 x_2^4 + x_1^4(1+x_2^2)})^T$$

and

$$\lambda(x) = \sqrt{2 + \frac{X_{N1}^2}{\|X_N\| x_1^2} + \frac{X_{N2}^2}{\|X_N\| x_2^2}}.$$

The following figure is a simulation result with the initial point $(0.4359, 0.9000)^T$. It demonstrates the quadratic convergence of the proposed algorithm.



VI. CONCLUSIONS

This paper reports our effort to generalize the self-concordant functions to manifold. It lays a comparative solid foundation to facilitate the construction of barrier functions for interior-point algorithms on manifolds.

First, the self-concordant function is carefully defined on a general class of smooth manifold. Many desirable properties have been revealed. Those include the feasibility of Dikin-type ellipsoid and those that reflect the similarity of self-concordant functions and quadratic functions along the geodesics of the manifold in various inequalities. Then, the Newton decrement is defined for this specific class of functions. This concept is analyzed in regards 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 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 $O(-\ln(\epsilon))$.

A simple but meaningful optimization problem is given as an example to illustrate the importance of the proposed concept and algorithm. The optimization problem cannot be directly solved using the standard interior point method for matrix inequalities because the feasible set does not contain an interior point. However, it can be nicely handled by the proposed self-concordant concept and damped Newton algorithm.

VII. ACKNOWLEDGMENT

The first author would like to thank Dr. Jochen Trumph for valuable discussions and suggestions.

REFERENCES

[1] Y. Nesterov and A. Nemirovskii. *Interior-Point Polynomial Algorithms in Convex Programming*. studies in applied mathematics 13. SIAM, Philadelphia, 1994.

[2] F. Glineur. Improving complexity of structured convex optimization problems usiing self-concordant barriers. *European Journal of Operational Research*, 143:291–310, 2002.

[3] L. Faybusovich. Infinitedimensional semidefinite programming:regularized determinants and selfconcordant barriers. *Fields Institute Communications*, 8:39–49, 1998.

[4] S. Zhang. *A new self-dual embedding method for convex programming*. Department of Systems Engineering & Engineering Management The Chinese University of Hong Kong, Technical Report SEEM2001-09, 2001.

[5] Y. Nesterov and M. Todd. On the Riemannian geometry defined by self-concordant barriers and interior-point methods. *Foundations of Computational Mathematics*, pages 333–361, 2002.

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

[7] S. Smith. Optimization techniques on Riemannian manifolds. *Fields Institute Communications*, 3:113–146, 1994.

[8] A Edelman, T. Arias, and S.T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications*.

[9] W. Yan, U. Helmke, and J. B. Moore. Global analysis of Oja's flow for neural networks. *IEEE Transactions on Neural Networks*, 5:674–683, 1993.

[10] C. Udriste. Optimization methods on Riemannian manifolds. *Algebera, Groups and Geometries*, 14:339–359, 1997.

[11] S. Halgason. *Differential Geometry, Lie Groups, and Symmetric Spaces*. Academic Press, New York, 1978.

[12] Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Kluwer Academic Publishers, 2004.

[13] C. Udriste. *Convex Functions and Optimization Methods on Riemannian Manifolds*. Mathematics and Its Applications 297. Kluwer Academic Publishers, London, 1994.

[14] Danchi Jiang, John B. Moore, and Huibo Ji. Self-concordant functions for optimization on smooth manifolds. *Submitted to SIAM Journal on Optimization*, 2004.