J Glob Optim (2017) 68:357–366
DOI 10.1007/s10898-016-0469-6

SHORT COMMUNICATION

Self-concordance is NP-hard

Lek-Heng Lim[1]

Received: 15 April 2016 / Accepted: 14 September 2016 / Published online: 19 September 2016

Abstract We show that deciding whether a convex function is self-concordant is in general an intractable problem.

Keywords Self-concordance · Second-order self-concordance · NP-hard · Co-NP-hard

1 Introduction

Nesterov and Nemirovskii [15] famously showed that the optimal solution of a conic programming problem can be computed to ε-accuracy in polynomial time if the cone has a self-concordant barrier function whose gradient and Hessian are both computable in polynomial time. Their work established self-concordance as a singularly important notion in modern optimization theory.

We show in this article that deciding whether a convex function is self-concordant at a point is nonetheless an NP-hard problem. In fact we will prove that deciding the self-concordance of a convex function defined locally by a cubic polynomial (which cannot be convex on all of $\mathbb{R}^n$), arguably the simplest non-trivial instance, is already an NP-hard problem. In addition to the NP-hardness of self-concordance, we will see that, unless $P = NP$, there is no fully polynomial time approximation scheme for the optimal self-concordant parameter and that deciding second-order self-concordance [10] of a quartic polynomial is also an NP-hard problem.

These hardness results are intended only to add to our understanding of self-concordance. They do not in anyway detract from the usefulness of the notion since in practice self-

This work is supported by AFOSR FA9550-13-1-0133, DARPA D15AP00109, NSF IIS 1546413, DMS 1209136, and DMS 1057064.

✉ Lek-Heng Lim
lekheng@galton.uchicago.edu

1 Computational and Applied Mathematics Initiative, Department of Statistics, University of Chicago, Chicago, IL 60637, USA

concordant barriers are *constructed* at the outset to have the requisite property (see [15, Chapter 5] and [3, Section 9.6]). We deduce the NP-hardness of self-concordance using a well-known result of Nesterov himself, namely, minimizing a cubic form over a sphere is in general NP-hard [14].

2 Self-concordance in terms of tensors

Let $\Omega \subseteq \mathbb{R}^n$ be open and $f : \Omega \to \mathbb{R}$ be in $C^d(\Omega)$, i.e., has continuous partials up to at least order d. Recall that the dth order derivative of f at $x \in \Omega$, denoted $\nabla^d f(x)$, is a *tensor* of order d [11]. To be more precise, this simply means that $\nabla^d f(x)$ is a multilinear functional on $T_x(\Omega)$, the tangent space of Ω at x, that is,

$$\nabla^d f(x) : \underbrace{T_x(\Omega) \times \cdots \times T_x(\Omega)}_{d \text{ copies}} \to \mathbb{R},$$

where

$$\begin{aligned}\nabla^d f(x)&(h_1, \dots, \alpha h_i + \beta h_i', \dots, h_n)\\ &= \alpha \nabla^d f(x)(h_1, \dots, h_i, \dots, h_n) + \beta \nabla^d f(x)(h_1, \dots, h_i', \dots, h_n) \qquad \text{for } i = 1, \dots, n.\end{aligned}$$

With respect to the standard basis $\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}$ of $T_x(\Omega)$, we may identify $T_x(\Omega) \cong \mathbb{R}^n$ and $\nabla^d f(x)$ may be regarded as a 'd-dimensional matrix' or d-*hypermatrix*,

$$\nabla^d f(x) = [a_{i_1 \cdots i_d}]_{i_1, \dots, i_d = 1}^n \in \mathbb{R}^{n \times \cdots \times n}.$$

Indeed, we must have

$$a_{i_1 \cdots i_d} = \frac{\partial^d f(x)}{\partial x_{i_1} \cdots \partial x_{i_d}},$$

and since $f \in C^d(\Omega)$, we get that $a_{i_1 \cdots i_d} = a_{i_{\pi(1)} \cdots i_{\pi(d)}}$ for all permutations π on the indices, i.e., $\nabla^d f(x)$ is a *symmetric* d-hypermatrix. Every symmetric d-hypermatrix $A = [a_{i_1 \cdots i_d}]_{i_1, \dots, i_d = 1}^n \in \mathbb{R}^{n \times \cdots \times n}$ defines a homogeneous polynomial of degree d, denoted

$$A(h, \dots, h) := \sum\nolimits_{i_1, \dots, i_d = 1}^n a_{i_1 \cdots i_d} h_{i_1} \cdots h_{i_d} \in \mathbb{R}[h_1, \dots, h_n]_d.$$

d-hypermatrices are coordinate representations of d-tensors, just as matrices are coordinate representations of linear operators and bilinear forms (both are 2-tensors). We refer the reader to [12] for more information.

The usual definition of self-concordance requires that $f \in C^3(\Omega)$ and in which case it is given by a condition involving the matrix $\nabla^2 f(x) \in \mathbb{R}^{n \times n}$ and the 3-hypermatrix $\nabla^3 f(x) \in \mathbb{R}^{n \times n \times n}$.

Definition 1 (*Nesterov–Nemirovskii*) Let $\Omega \subseteq \mathbb{R}^n$ be a convex open set. Then $f : \Omega \to \mathbb{R}$ is said to be *self-concordant* with parameter $\sigma > 0$ at $x \in \Omega$ if

$$\nabla^2 f(x)(h, h) \geq 0 \tag{1}$$

and

$$[\nabla^3 f(x)(h, h, h)]^2 \leq 4\sigma [\nabla^2 f(x)(h, h)]^3 \tag{2}$$

for all $h \in \mathbb{R}^n$; f is self-concordant on Ω if (1) and (2) hold for all $x \in \Omega$. The set of self-concordant functions on Ω with parameter σ is denoted by $S_\sigma(\Omega)$.

By (1), a function self-concordant on Ω is necessarily convex on Ω. A minor deviation from [15] is that σ above is really the reciprocal of the self-concordance parameter as defined in [15, Definition 2.1.1]. Our hardness results would be independent of the choice of σ. Note that

$$\nabla^2 f(x)(h,h) = \sum_{i,j=1}^{n} \frac{\partial^2 f(x)}{\partial x_i \partial x_j} h_i h_j, \quad \nabla^3 f(x)(h,h,h) = \sum_{i,j,k=1}^{n} \frac{\partial^3 f(x)}{\partial x_i \partial x_j \partial x_k} h_i h_j h_k.$$

So for a fixed $x \in \Omega$, $\nabla^2 f(x)(h,h)$ is a quadratic form in h and $\nabla^3 f(x)(h,h,h)$ is a cubic form in h. It is well-known that deciding (1) at any fixed x is a polynomial-time problem (but not so for deciding it over all $x \in \Omega$, see [1]). Hence given a $\sigma > 0$, deciding self-concordance at x essentially boils down to (2): Is the square of a given cubic form globally bounded above by the cube of a given quadratic form? We shall see in the next sections that this decision problem is NP-hard.

While we will think of $\nabla^2 f(x)$ as a matrix and $\nabla^3 f(x)$ as a hypermatrix, we wish to highlight that (2) is really a condition on $\nabla^2 f(x)$ regarded as a 2-tensor and $\nabla^3 f(x)$ regarded as a 3-tensor, i.e., (2) is independent of the choice of coordinates, a property that follows from the *affine invariance* of self-concordance [15, Proposition 2.1.1]. Self-concordance on Ω is therefore a global condition about the tensor fields $\nabla^3 f$ and $\nabla^2 f$.

3 Maximizing a cubic form over a sphere

We include a proof that the clique and stability numbers of a graph with n vertices and m edges may be expressed as the maximum values of cubic forms (in $n+m$ variables) over the unit sphere $\mathbb{S}^{n+m-1}$. This, or at least the stability number version, is known but the reference usually cited [14, Theorem 4] contains some typos that have been reproduced elsewhere[1]. We take the opportunity to provide a corrected version below. Our proof follows Motzkin–Straus Theorem [13] and the similar result of Nesterov [14, Theorem 4] for stability number.

Let $G = (V, E)$ be an undirected graph with n vertices and m edges. We shall require that $E \neq \varnothing$ throughout, so $n \geq 2$ and $m \geq 1$. Recall that $S \subseteq V$ is a *clique* in G if $\{i, j\} \in E$ for all $i, j \in S$ and $S \subseteq V$ is *stable* in G if $\{i, j\} \notin E$ for all $i, j \in S$. The *clique number* and *stability number* of G are respectively:

$$\omega(G) = \max\{|S| : S \subseteq V \text{ is clique}\}, \quad \alpha(G) = \max\{|S| : S \subseteq V \text{ is stable}\}.$$

Motzkin and Straus [13] showed that $\omega(G)$ may be expressed in terms of the maximum value of a simple quadratic polynomial over the unit simplex. Although not in [13], it is straightforward to see that essentially the same proof also yields a similar expression for $\alpha(G)$.

Theorem 1 (Motzkin–Straus) *Let $\Delta^n = \{x \in \mathbb{R}^n : x_1 + \cdots + x_n = 1,\ x_i \geq 0\}$ denote the unit simplex in $\mathbb{R}^n$. Then the clique number $\omega(G)$ and stability number $\alpha(G)$ may be determined via quadratic optimization over simplices:*

$$1 - \frac{1}{\omega(G)} = 2 \max_{x \in \Delta^n} \sum\nolimits_{\{i,j\} \in E} x_i x_j, \quad 1 - \frac{1}{\alpha(G)} = 2 \max_{x \in \Delta^n} \sum\nolimits_{\{i,j\} \notin E} x_i x_j. \tag{3}$$

[1] For example, [4, Theorem 3.4]. To see that the expression is incorrect, take a graph with three vertices and one edge, the left-hand side gives $1/\sqrt{2}$ and the right-hand side gives 1.

Since deciding if a clique of a given size exists is an NP-complete problem [9], an immediate consequence is that computing the clique number of a given graph is NP-hard, and by the Motzkin–Straus theorem so is the maximization of a quadratic form on a simplex.

In an unpublished manuscript [14, Theorem 4], Nesterov used Motzkin–Straus Theorem to obtain an alternate expression (5) for stability number involving the maximum value of a cubic form over a sphere. In the following we derive a similar expression (4) for the clique number, which yields slightly simpler expressions for our discussions in Sects. 4 and 6, and may perhaps be of independent interest.

Theorem 2 (Nesterov) *Let $G = (V, E)$ be an undirected graph with n vertices and m edges. Let $\mathbb{S}^{d-1} = \{x \in \mathbb{R}^d : \|x\| = 1\}$ denote the unit ℓ^2-sphere in $\mathbb{R}^d$. The clique number $\omega(G)$ and stability number $\alpha(G)$ may be determined via cubic optimization over spheres:*

$$1 - \frac{1}{\omega(G)} = \frac{27}{2} \max_{(v,w)\in\mathbb{S}^{n+m-1}} \Big[\sum\nolimits_{\{i,j\}\in E} v_i v_j w_{ij}\Big]^2, \tag{4}$$

$$1 - \frac{1}{\alpha(G)} = \frac{27}{2} \max_{(v,w)\in\mathbb{S}^{n+m-1}} \Big[\sum\nolimits_{\{i,j\}\notin E} v_i v_j w_{ij}\Big]^2. \tag{5}$$

Proof This follows from Motzkin–Straus Theorem and the equalities

$$\max_{x\in\Delta^n} \sum\nolimits_{\{i,j\}\in E} x_i x_j = \max_{v\in\mathbb{S}^{n-1}} \sum\nolimits_{\{i,j\}\in E} v_i^2 v_j^2 \tag{6}$$

$$= \max_{v\in\mathbb{S}^{n-1}, w\in\mathbb{S}^{m-1}} \Big[\sum\nolimits_{\{i,j\}\in E} v_i v_j w_{ij}\Big]^2 \tag{7}$$

$$= \frac{27}{4} \max_{(v,w)\in\mathbb{S}^{n+m-1}} \Big[\sum\nolimits_{\{i,j\}\in E} v_i v_j w_{ij}\Big]^2. \tag{8}$$

(6) comes from substituting $x_i = v_i^2$, $i = 1, \ldots, n$. As for (7), Cauchy–Schwarz yields

$$\sum\nolimits_{\{i,j\}\in E} v_i v_j w_{ij} \le \Big[\sum\nolimits_{\{i,j\}\in E} v_i^2 v_j^2\Big]^{\frac{1}{2}} \Big[\sum\nolimits_{\{i,j\}\in E} w_{ij}^2\Big]^{\frac{1}{2}}$$

and so

$$\max_{\|v\|=\|w\|=1} \sum\nolimits_{\{i,j\}\in E} v_i v_j w_{ij} \le \max_{\|v\|=1}\Big[\sum\nolimits_{\{i,j\}\in E} v_i^2 v_j^2\Big]^{\frac{1}{2}} \max_{\|w\|=1}\Big[\sum\nolimits_{\{i,j\}\in E} w_{ij}^2\Big]^{\frac{1}{2}}$$

$$= \max_{\|v\|=1}\Big[\sum\nolimits_{\{i,j\}\in E} v_i^2 v_j^2\Big]^{\frac{1}{2}} =: \alpha. \tag{9}$$

Let the maximum value α be attained at $\overline{v} \in \mathbb{S}^{n-1}$. We set $\overline{w}_{ij} = \overline{v}_i \overline{v}_j / \alpha$ for all $\{i, j\} \in E$ (note that $\alpha > 0$ if $E \neq \varnothing$). Observe that

$$\sum\nolimits_{\{i,j\}\in E} \overline{v}_i \overline{v}_j \overline{w}_{ij} = \frac{1}{\alpha} \sum\nolimits_{\{i,j\}\in E} \overline{v}_i^2 \overline{v}_j^2 = \alpha,$$

and $\overline{w} \in \mathbb{S}^{m-1}$ since

$$\sum\nolimits_{\{i,j\}\in E} \overline{w}_{ij}^2 = \frac{1}{\alpha^2} \sum\nolimits_{\{i,j\}\in E} \overline{v}_i^2 \overline{v}_j^2 = 1.$$

Hence equality is attained in (9) and we have (7). We deduce (8) from

$$
\begin{aligned}
\max_{\|(v,w)\|=1} \sum\nolimits_{\{i,j\}\in E} v_i v_j w_{ij} &= \max_{\|v\|^2+\|w\|^2=1} \sum\nolimits_{\{i,j\}\in E} v_i v_j w_{ij} \\
&= \sup_{\beta\in(0,1)} \left[\max_{\|v\|^2=\beta,\ \|w\|^2=1-\beta} \sum\nolimits_{\{i,j\}\in E} v_i v_j w_{ij} \right] \\
&= \sup_{\beta\in(0,1)} \left[\max_{\|v/\sqrt{\beta}\|^2=1, \|w/\sqrt{1-\beta}\|^2=1} \sum\nolimits_{\{i,j\}\in E} \right. \\
&\qquad \left. \frac{v_i}{\sqrt{\beta}} \frac{v_j}{\sqrt{\beta}} \frac{w_{ij}}{\sqrt{1-\beta}} \times \beta\sqrt{1-\beta} \right] \\
&= \left[\max_{\|v\|^2=1,\ \|w\|^2=1} \sum\nolimits_{\{i,j\}\in E} v_i v_j w_{ij} \right] \times \sup_{\beta\in(0,1)} \beta\sqrt{1-\beta} \\
&= \frac{2}{3\sqrt{3}} \max_{\|v\|=\|w\|=1} \sum\nolimits_{\{i,j\}\in E} v_i v_j w_{ij}.
\end{aligned}
$$

The maximum value of $\sum_{\{i,j\}\in E} v_i v_j w_{ij}$, whether over $\mathbb{S}^{n-1} \times \mathbb{S}^{m-1}$ or over $\mathbb{S}^{n+m-1}$, can always be attained with $v \geq 0$ and $w \geq 0$, thereby allowing one to take square in (7) and (8). The same proof works for stability number with the replacement of index of summation '$\{i, j\} \in E$' by '$\{i, j\} \notin E$'. □

4 Complexity of deciding self-concordance

The recent resolution of Shor's conjecture by Ahmadi, Olshevsky, Parrilo, and Tsitsiklis [1] shows that deciding the convexity of a quartic polynomial globally over $\mathbb{R}^n$ is NP-hard. So the self-concordance of a function that is not a priori known to be convex is NP-hard since deciding whether (1) holds for all $x \in \Omega$ in Definition 1 is already an NP-hard problem. Our complexity result assumes more stringent conditions: (i) Our functions are smooth and convex in Ω and so (1) is always satisfied and self-concordance reduces to checking (2). (ii) We show that (2) is already NP-hard to check at a single point $x \in \Omega$.

Throughout the following we will require the *inputs* to our problems to take values in an algebraic number field[2], e.g., $A \in \mathbb{Q}^{n\times n\times n}$, $q \in \mathbb{Q}$, to ensure a finite bit-length input. Since an NP-hard problem need not be in the class NP, an NP-hard decision problem can be posed over the reals (e.g. is there an $h \in \mathbb{R}^n$ such that $[A(h, h, h)]^2 \leq q[h^\mathsf{T} h]^3$ holds?) without any issue as it is not required to have a polynomial-time checkable certificate.

We will now formulate a decision problem that will lead us to the NP-hardness of self-concordance. Let $G = (V, E)$ be an undirected graph with n vertices and m edges where $n \geq 2$ and $m \geq 1$. We order the n vertices arbitrarily so that $V = \{1, \ldots, n\}$. We also order the m edges arbitrarily so that

$$E = \{\{i_k, j_k\} : k = 1, \ldots, m\}.$$

[2] We will only encounter simple quadratic and cubic extensions of $\mathbb{Q}$, see (18) and (11). For $q \in \mathbb{Q}$, elements of $\mathbb{Q}(\sqrt{q})$ and $\mathbb{Q}(\sqrt[3]{q})$ may be written as $a + b\sqrt{q}$ and $a + b\sqrt[3]{q} + c(\sqrt[3]{q})^2$ respectively with $a, b, c \in \mathbb{Q}$, thus representable by pairs and triples of rational numbers.

Define $B_G = [b_{ijk}]_{i,j,k=1}^{n+m} \in \mathbb{Q}^{(n+m)\times(n+m)\times(n+m)}$ by

$$b_{ijk} = \begin{cases} 1 & \text{if } i = i_k \in V,\ j = j_k \in V,\ \{i_k, j_k\} \in E, \\ 0 & \text{otherwise.} \end{cases}$$

B_G is not a symmetric hypermatrix. Let $A_G = [a_{ijk}]_{i,j,k=1}^{n+m} \in \mathbb{Q}^{(n+m)\times(n+m)\times(n+m)}$ be the symmetrization of B_G, i.e.,

$$a_{ijk} = \frac{1}{3!}(b_{ijk} + b_{ikj} + b_{jik} + b_{jki} + b_{kij} + b_{kji})$$

for all $i, j, k \in \{1, \ldots, n+m\}$. So A_G is symmetric, i.e., $a_{ijk} = a_{ikj} = a_{jik} = a_{jki} = a_{kij} = a_{kji}$, and furthermore $A_G(h,h,h) = B_G(h,h,h)$. Let us denote the coordinates of $h \in \mathbb{R}^{n+m}$ by

$$h = (v_1, \ldots, v_n, w_{i_1 j_1}, \ldots, w_{i_m j_m}).$$

In which case,

$$\begin{aligned} A_G(h,h,h) = B_G(h,h,h) &= \sum\nolimits_{i,j,k=1}^{m+n} b_{ijk} h_i h_j h_k \\ &= \sum\nolimits_{k=1}^{m} v_{i_k} v_{j_k} w_{i_k j_k} = \sum\nolimits_{\{i,j\}\in E} v_i v_j w_{ij}. \end{aligned}$$

By Theorem 2,

$$\max_{h\neq 0}\left[\frac{A_G(h,h,h)}{\|h\|^3}\right]^2 = \max_{\|h\|=1}[A_G(h,h,h)]^2 = \frac{2}{27}\left(1 - \frac{1}{\omega(G)}\right). \tag{10}$$

The CLIQUE problem asks if for a given graph G and a given $k \in \mathbb{N}$, whether G has a clique of size k. CLIQUE is well-known to be NP-complete [9]. So deciding if $\omega(G) \geq k$, or equivalently, $\omega(G) > k-1$, is an NP-complete problem; by (10), so is deciding if

$$\max_{h\neq 0}\left[\frac{A_G(h,h,h)}{\|h\|^3}\right]^2 > \frac{2}{27}\left(1 - \frac{1}{k-1}\right).$$

Hence deciding if there exists an $h \in \mathbb{R}^{n+m}$ for which

$$[A_G(h,h,h)]^2 > \frac{2}{27}\left(1 - \frac{1}{k-1}\right)[h^\top h]^3$$

is an NP-hard problem. As $q = \frac{2}{27}[1 - (k-1)^{-1}] \in \mathbb{Q}$, this problem is of the form:

Problem 1 Given a symmetric $A \in \mathbb{Q}^{(n+m)\times(n+m)\times(n+m)}$ and a positive $q \in \mathbb{Q}$, does there exists an $h \in \mathbb{R}^{n+m}$ for which $[A(h,h,h)]^2 > q[h^\top h]^3$?

Let $\sigma \in \mathbb{Q}, \sigma > 0$, be a self-concordance parameter and let

$$\gamma := \frac{1}{3}\left[\frac{1}{2\sigma}\left(1 - \frac{1}{k-1}\right)\right]^{1/3}. \tag{11}$$

We follow the notation in Sect. 2. Let Ω be the ε-ball $B_\varepsilon(0) = \{x \in \mathbb{R}^{n+m} : \|x\| < \varepsilon\}$ where $\varepsilon > 0$ is to be chosen later. We are interested in deciding self-concordance at $x = 0$ of the cubic polynomial $f : \Omega \to \mathbb{R}$ defined by

$$f(x) = \frac{\gamma}{2}x^\top x + A_G(x,x,x) = \frac{\gamma}{2}\sum\nolimits_{i=1}^{n+m} x_i^2 + \sum\nolimits_{i,j,k=1}^{n+m} a_{ijk} x_i x_j x_k.$$

We have $\nabla^2 f(0) = \gamma I$ where I is the $(n+m)\times(n+m)$ identity matrix. Since $\gamma > 0$, $\nabla^2 f(x)$ is strictly positive definite in a neighborhood of $x = 0$ and so there is some $B_\varepsilon(0)$ on which f is convex, giving us our ε. Also, $\nabla^3 f(0) = A_G$.

Hence $\nabla^2 f(0)(h,h) = \gamma h^\mathsf{T} h = \gamma \|h\|_2^2$, $\nabla^3 f(0)(h,h,h) = A_G(h,h,h)$, and f is self-concordant at the origin with parameter $\sigma \in \mathbb{Q}$ if and only if

$$[A_G(h,h,h)]^2 \le 4\sigma\gamma^3[h^\mathsf{T} h]^3 = \frac{2}{27}\left(1 - \frac{1}{k-1}\right)[h^\mathsf{T} h]^3$$

for all $h \in \mathbb{R}^{n+m}$. This problem is of the form:

Problem 2 Given a symmetric $A \in \mathbb{Q}^{(n+m)\times(n+m)\times(n+m)}$ and a positive $q \in \mathbb{Q}$, is it true that for every $h \in \mathbb{R}^{n+m}$, we have $[A(h,h,h)]^2 \le q[h^\mathsf{T} h]^3$?

Problems 1 and 2 are mathematically equivalent, being logical complements of each other. However they may or may not have the same computational complexity depending on our choice of *reduction* [16]. Using *Cook reduction*, also know as polynomial-time Turing reduction [16], Problems 1 and 2 have equivalent computational complexity, i.e., deciding self-concordance is NP-hard. However, using *Karp reduction*, also know as polynomial-time many-one reduction [16], the NP-hardness of Problem 1 implies the co-NP-hardness of Problem 2. In either case, our conclusion is that self-concordance is intractable.

Theorem 3 *Deciding whether a cubic polynomial is self-concordant at the origin is NP-hard under Cook reduction and co-NP-hard under Karp reduction.*

Deciding self-concordance on the whole of Ω is of course at least as hard as deciding self-concordance at a point in Ω and we obtain the following.

Corollary 1 *For any Ω and any $\sigma > 0$, membership in $S_\sigma(\Omega)$ is NP-hard.*

The argument in this section clearly works not just for cubic polynomials but for any $f \in C^3(\Omega)$ as long as $0 \in \Omega$, $\nabla^2 f(0) = \gamma I$, and $\nabla^3 f(0) = A_G$ — other derivatives and the remainder term in the Taylor expansion of f at $x = 0$ may be chosen arbitrarily as long as f stays convex in Ω. This flexibility allows one to extend the construction above to functions with other desired properties. For instance, we may want an example where $\Omega = \mathbb{R}^n$ and since cubic polynomials cannot be convex on the whole of $\mathbb{R}^n$, we will need a quartic f and therefore need to choose $\nabla^4 f(0)$ accordingly; or we may want an example where f is a barrier function, which is equivalent to f having an epigraph $\{(x,t) \in \mathbb{R}^{n+1} : x \in \Omega,\ f(x) \le t\}$ that is closed. One may trivially replace 0 by any point $a \in \mathbb{R}^n$ by considering the function $f_a(x) = f(x-a)$ on $\Omega = B_\varepsilon(a)$.

While we have proved our hardness result for functions on $\Omega \subseteq \mathbb{R}^n$, it is easy to extend this to any $\mathbb{R}$-vector space, for example, symmetric matrices $\mathbb{S}^{n\times n}$ or polynomials $\mathbb{R}[x_1,\ldots,x_n]$, or even Riemannian manifolds with a non-trivial class of geodesically convex functions (i.e., not just the constant functions). Since self-concordance at a point is a local property, a choice of coordinate patch would transform the problem to one in $\mathbb{R}^n$; and by our remark at the end of Sect. 2, it will in fact be independent of our choice of coordinates.

The reason our conclusion in Theorem 3 is given as NP- and co-NP-hardness as opposed to NP- and co-NP-completeness is that checking a condition like $[A(h,h,h)]^2 \le q[h^\mathsf{T} h]^3$, even if the certificate h is in $\mathbb{Q}^n$, could well require exponential time when time complexity is measured in units of *bit* operations.

5 Inapproximability of optimal self-concordance parameter

Let $A \in \mathbb{Q}^{n\times n\times n}$ be symmetric and $f : \Omega \to \mathbb{R}$ be defined by the cubic polynomial $f(x) = \frac{1}{2}x^\mathsf{T} x + A(x, x, x)$. As in Sect. 4, Ω is chosen to be a neighborhood of the origin so that f is convex on Ω. The condition (2) for self-concordance of f at $x = 0$ with parameter $\sigma > 0$ may be written as

$$|A(h, h, h)| \le 2\sqrt{\sigma}\|h\|_2^3 \tag{12}$$

for all $h \in \mathbb{R}^n$. This is equivalent to requiring

$$\max_{h\neq 0} \frac{A(h, h, h)}{\|h\|_2^3} \le 2\sqrt{\sigma}, \tag{13}$$

as $A(-h, -h, -h) = -A(h, h, h)$ and we may drop the absolute value in (12).

Since $A \in \mathbb{R}^{n\times n\times n}$ is a symmetric 3-hypermatrix, the *spectral norm* [6] of A may be expressed as follows:

$$\|A\|_{2,2,2} := \max_{h_1,h_2,h_3\neq 0} \frac{A(h_1, h_2, h_3)}{\|h_1\|_2\|h_2\|_2\|h_3\|_2} = \max_{h\neq 0} \frac{A(h, h, h)}{\|h\|_2^3},$$

where the second equality follows from Banach's result on the polarization constant of Hilbert spaces (see [2] and [17]). Hence the optimal self-concordance parameter of f at $x = 0$, i.e., the smallest value of σ so that (13) holds, is given by

$$\sigma_{\text{opt}} = \frac{1}{4}\|A\|_{2,2,2}^2. \tag{14}$$

The spectral norm of a 3-hypermatrix is NP-hard to approximate to within a certain constant factor by [6, Theorem 1.11], which we state here for easy reference.

Theorem 4 (Hillar–Lim) *Let $A \in \mathbb{Q}^{n\times n\times n}$ and N be the input size of A in bits. Then it is NP-hard to approximate $\|A\|_{2,2,2}$ to within a factor of $1 - \varepsilon$ where*

$$\varepsilon = 1 - \left(1 + \frac{1}{N(N-1)}\right)^{-1/2} = \frac{1}{2N(N-1)} + O\left(\frac{1}{N^4}\right).$$

By (14) and Theorem 4, σ_{opt} is NP-hard to approximate to within a factor of $\frac{1}{4}(1-\varepsilon)^2$ and consequently we have the following inapproximability result.

Corollary 2 *There is no polynomial time approximation scheme for determining the optimal self-concordance parameter σ_{opt} unless P = NP.*

See [5] and [7] for more extensive approximability results and approximation algorithms (that are not PTAS). The results in [7] for quartic polynomials would apply to the optimal second-order self-concordance parameter in the next section.

6 Complexity of deciding second-order self-concordance

There is also an interesting notion of second-order self-concordance due to Jarre [10]. This requires that $f \in C^4(\Omega)$ and is given by a condition involving the matrix $\nabla^2 f(x) \in \mathbb{R}^{n\times n}$ and the 4-hypermatrix $\nabla^4 f(x) \in \mathbb{R}^{n\times n\times n\times n}$.

Definition 2 (*Jarre*) If $\Omega \subseteq \mathbb{R}^n$ is a convex open set, then $f : \Omega \to \mathbb{R}$ is said to be *second-order self-concordant* with parameter $\tau > 0$ at $x \in \Omega$ if

$$\nabla^2 f(x)(h, h) \geq 0 \tag{15}$$

and

$$\nabla^4 f(x)(h, h, h, h) \leq 6\tau\big[\nabla^2 f(x)(h, h)\big]^2 \tag{16}$$

for all $h \in \mathbb{R}^n$; f is second-order self-concordant on Ω if it is so for all $x \in \Omega$.

Note that

$$\nabla^4 f(x)(h, h, h, h) = \sum\nolimits_{i,j,k,l=1}^{n} \frac{\partial^4 f(x)}{\partial x_i \partial x_j \partial x_k \partial x_l} h_i h_j h_k h_l,$$

is a quartic polynomial in h for any fixed $x \in \Omega$.

We follow the same argument in Sect. 4 to show that deciding (16) is NP-hard. This time the result would be deduced from Motzkin–Straus Theorem except that for better parallelism with Sect. 4, we will use the quartic-maximization-over-sphere form (6) instead of the quadratic-maximization-over-simplex form (3).

Given a graph $G = (V, E)$ with n vertices and m edges where $n \geq 2$ and $m \geq 1$, we define $B_G \in \mathbb{Q}^{n\times n\times n\times n}$ by

$$b_{ijkl} = \begin{cases} 1 & i = k,\ j = l, \text{ and } \{i, j\} \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Let $A_G = [a_{ijkl}] \in \mathbb{Q}^{n\times n\times n\times n}$ be the symmetrization of B_G, i.e.,

$$a_{ijkl} = \frac{1}{4!}(b_{ijkl} + b_{ijlk} + \cdots + b_{lkji}),$$

where the indices run over all 24 possible permutations. So $A = [a_{ijkl}]_{i,j,k,l=1}^n \in \mathbb{Q}^{n\times n\times n\times n}$ is symmetric and $A_G(h, h, h, h) = B_G(h, h, h, h)$. As in (6),

$$\max_{\|h\|=1} A_G(h, h, h, h) = \max_{h\in\mathbb{S}^{n-1}} \sum\nolimits_{\{i,j\}\in E} h_i^2 h_j^2 = \frac{1}{2}\left(1 - \frac{1}{\omega(G)}\right)$$

by Motzkin–Straus Theorem. As in Sect. 4, for $k \geq 2$, deciding if a k-clique exists in G is equivalent to deciding if $\omega(G) > k - 1$. So deciding if

$$A_G(h, h, h, h) > \frac{1}{2}\left(1 - \frac{1}{k-1}\right)[h^{\mathsf{T}} h]^2 \tag{17}$$

for some $h \in \mathbb{R}^n$ is NP-hard. Given self-concordance parameter $\tau \in \mathbb{Q}$, $\tau > 0$, let

$$\gamma := \left[\frac{1}{12\tau}\left(1 - \frac{1}{k-1}\right)\right]^{1/2}. \tag{18}$$

We may now define $f : \Omega \to \mathbb{R}$ accordingly as the quartic polynomial

$$f(x) = \frac{\gamma}{2} x^{\mathsf{T}} x + A_G(x, x, x, x) = \frac{\gamma}{2}\sum\nolimits_{i=1}^n x_i^2 + \sum\nolimits_{i,j,k,l=1}^n a_{ijkl} x_i x_j x_k x_l.$$

Hence $\nabla^2 f(0)(h, h) = \gamma h^{\mathsf{T}} h = \gamma\|h\|_2^2$ and $\nabla^4 f(0)(h, h, h, h) = A_G(h, h, h, h)$. Again we choose Ω to be a neighborhood of the origin so that f is convex on Ω. So f is second-order self-concordant at $x = 0$ with parameter τ if and only if

$$A_G(h, h, h, h) \leq 6\tau\gamma^2[h^{\mathsf{T}} h]^2 = \frac{1}{2}\left(1 - \frac{1}{k-1}\right)[h^{\mathsf{T}} h]^2 \tag{19}$$

is satisfied for all $h \in \mathbb{R}^n$. Since the problem of deciding if there exists an $h \in \mathbb{R}^n$ satisfying (17) and the problem of deciding if (19) is satisfied for all $h \in \mathbb{R}^n$ are logical complements and the former is NP-hard, we have the following.

Theorem 5 *Deciding if a quartic polynomial is second-order self-concordant at the origin is NP-hard under Cook reduction and co-NP-hard under Karp reduction.*

Self-concordance and second-order self-concordance are conditions involving high-order tensors (orders 3 and 4 respectively), their NP-hardness serves as yet another reminder of the complexity of tensor problems [6].

Acknowledgements We would like to thank the anonymous referees for their careful reading and helpful suggestions.

References

1. Ahmadi, A.A., Olshevsky, A., Parrilo, P.A., Tsitsiklis, J.N.: NP-hardness of deciding convexity of quartic polynomials and related problems. Math. Progr. Ser. A **137**(1–2), 453–476 (2013)
2. Banach, S.: Über homogene polynome in (L^2). Stud. Math. **7**(1), 36–44 (1938)
3. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004)
4. De Klerk, E.: The complexity of optimizing over a simplex, hypercube or sphere: a short survey. Cent. Eur. J. Oper. Res. **16**(2), 111–125 (2008)
5. He, S., Li, Z., Zhang, S.: Approximation algorithms for homogeneous polynomial optimization with quadratic constraints. Math. Progr. Ser. B **125**(2), 353–383 (2010)
6. Hillar, C. J., Lim, L.-H.: Most tensor problems are NP-hard. J. ACM **60**(6), 39 (**Art. 45**) (2013)
7. Hou, K., So, A.M.-C.: Hardness and approximation results for L_p-ball constrained homogeneous polynomial optimization problems. Math. Oper. Res. **39**(4), 1084–1108 (2014)
8. Jiang, B., Li, Z., Zhang, S.: On cones of nonnegative quartic forms. Found. Comput. Math. (2016). doi:10.1007/s10208-015-9286-4
9. Karp, R.M.: Reducibility among combinatorial problems. In: Miller, R.E., Thatcher, J.W. (eds.) Complexity of Computer Computations, pp. 85–103. Plenum, New York (1972)
10. Jarre, F.: A new line-search step based on the Weierstrass $\wp$-function for minimizing a class of logarithmic barrier functions. Numer. Math. **68**(1), 81–94 (1994)
11. Lang, S.: Differential and Riemannian Manifolds, 3rd edn., Graduate Texts in Mathematics, vol. 160. Springer, New York (1995)
12. Lim, L.-H.: Tensors and hypermatrices. In: Hogben, L. (ed.) Handbook of Linear Algebra, 2nd edn. CRC Press, Boca Raton (2013)
13. Motzkin, T., Straus, E.G.: Maxima for graphs and a new proof of a theorem of Turán. Can. J. Math. **17**, 533–540 (1965)
14. Nesterov, Yu.: Random walk in a simplex and quadratic optimization over convex polytopes. Preprint (2003). http://edoc.bib.ucl.ac.be:83/archive/00000238/01/dp2003-71
15. Nesterov, Yu., Nemirovskii, A.: Interior-Point Polynomial Algorithms in Convex Programming. SIAM Studies in Applied Mathematics, vol. 13. SIAM, Philadelphia (1994)
16. Papadimitriou, C.H.: Computational Complexity. Addison-Wesley, Reading (1994)
17. Pappas, A., Sarantopoulos, Y., Tonge, A.: Norm attaining polynomials. Bull. Lond. Math. Soc. **39**(2), 255–264 (2007)