A New Class of Self-Concordant Barriers from Separable Spectral Functions

Javier Peña* and **Hristo S. Sendov†**

September 7, 2007

Abstract

Given a separable strongly self-concordant function $f : \mathbb{R}^n \to \mathbb{R}$, we show the associated spectral function $F(X) = (f \circ \lambda)(X)$ is also strongly self-concordant function. In addition, there is a universal constant $\mathscr{O}$ such that, if $f(x)$ is separable self-concordant barrier then $\mathscr{O}^2 F(X)$ is a self-concordant barrier. We estimate that for the universal constant we have $\mathscr{O} \leq 22$. This generalizes the relationship between the standard logarithmic barriers $-\sum_{i=1}^n \log x_i$ and $-\log\det X$ and gives a partial solution to a conjecture of L. Tunçel.
Keywords: Self-concordant barrier, strongly self-concordant, self-concordant function, spectral function, eigenvalue, symmetric matrix
Mathematics Subject Classification (2000) 65K10, 73K40, 46N10, 15A18

1 Introduction

1.1 Self-concordant barriers

A real-valued function F defined on an open convex subset Q of a Euclidean space $\mathscr{H}$ is called *$\mathscr{O}$-self-concordant* if it is three times continuously differentiable and satisfies the condition

$$(1) \qquad |\nabla^3 F(X)[H,H,H]| \quad \leq \quad 2\mathscr{O}(\nabla^2 F(X)[H,H])^{3/2},$$

for all $H \in \mathscr{H}$, $X \in Q$. Here $\nabla^k F(X)[H,\ldots,H] = \frac{d^k}{dt^k} F(X+tH)\big|_{t=0}$ is the k-th directional derivative at X along the direction H. The function F is called *strongly $\mathscr{O}$-self-concordant* if in addition to (1) it satisfies

$$(2) \qquad F(X^r) \to +\infty \text{ for any sequence } X^r \to X \in \partial Q.$$

*Carnegie Mellon University, Tepper School of Business, Pittsburgh, PA 15213. Email: jfp@andrew.cmu.edu.

†Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Ontario, Canada N6A 5B7. Email: hssendov@stats.uwo.ca. Research supported by NSERC.

We say that F is (strongly) self-concordant if it is (strongly) $\mathscr{O}$-self-concordant for some constant $\mathscr{O}$. Finally, the function F is called *ϑ-self-concordant barrier* if it is strongly 1-self-concordant and

$$|\nabla F(X)[H]| \quad \leq \quad \sqrt{\vartheta}\,(\nabla^2 F(X)[H,H])^{1/2}, \tag{3}$$

for all $H \in \mathscr{H}$, $X \in Q$. We simply say that F is a self-concordant barrier, if it is ϑ-self-concordant barrier for some constant ϑ. The constants $\mathscr{O}$ and ϑ are fixed and depend on the function F only. Self-concordant barrier functions are the heart of interior-point algorithms. The parameter ϑ determines the speed of the underlying interior-point method: smaller parameters ensure that the interior-point algorithm using F runs faster. (It can be shown [3, Corollary 4.3.1] that for any ϑ-self-concordant barrier $\vartheta \geq 1$.) For detailed treatment of of these notions and their applications to interior-point methods see the monographs [5], [6], or most recently [3].

Theorem 2.5.1 in [5] shows that if $\dim(\mathscr{H}) = n$ then every convex domain in $\mathscr{H}$ admits a Cn-self-concordant barrier, called *the universal barrier*, where C is an absolute constant. The abstract nature of this important theoretical result does not reveal how quickly and efficiently to compute the barrier in practice. In addition it says nothing about the smallest possible parameter ϑ for which there is a ϑ-self-concordant barrier for a given convex set Q. These two issues are among the most important ones in the theory of self-concordant barriers.

Throughout the whole work we assume that $g : \mathbb{R} \to \mathbb{R}$ is defined on $\operatorname{dom} g := (a,b)$, where the interval (a,b) may be finite of infinite. In this case, the ϑ-self-concordant barrier conditions for g can be stated more succinctly as:

$$|g'''(x)| \quad \leq \quad 2\,(g''(x))^{3/2} \text{ for all } x \in (a,b); \tag{4}$$

$$g(x^r) \quad \to \quad +\infty \text{ for any sequence } x^r \to x \text{ on the boundary of } (a,b);\text{ and} \tag{5}$$

$$|g'(x)| \quad \leq \quad \sqrt{\vartheta}\,(g''(x))^{1/2} \text{ for all } x \in (a,b). \tag{6}$$

1.2 Separable spectral functions

By M^n we denote the Euclidean space of $n \times n$ real matrices with the standard inner product $\langle X,Y\rangle = \operatorname{tr}(X^TY)$ between any $X,Y \in M^n$. By S^n we denote the subspace of $n \times n$ symmetric matrices with inner product inherited from M^n. For any $X \in S^n$, by $\lambda(X)$ we denote the real vector of n eigenvalues of X, ordered non-increasingly, that is

$$\lambda_1(X) \geq \lambda_2(X) \geq \cdots \geq \lambda_n(X).$$

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a symmetric function, that is, $f(Px) = f(x)$ for every permutation matrix P and vector x in the domain of f. The composition $F := f \circ \lambda$ defines a *spectral function* on S^n with the characterizing property that $F(UXU^T) = F(X)$ for every orthogonal matrix U and X in the domain of F. The spectral functions are in one-to-one correspondence with the symmetric functions. Knowing the spectral function we can recover the underlying symmetric function by restricting F to the subspace of diagonal matrices. The following questions (see [2, p.168]) provide the main motivation for our work.

Conjecture 1.1 (L.Tunçel, 1995) *The function f is a self-concordant (barrier) if and only if $F = f \circ \lambda$ is.*

Our main result, Theorem 1.2, gives a partial solution to this conjecture. In one direction the conjecture is easy because functions that are self-concordant (barriers) remain such when restricted to affine subspaces, see [5, Proposition 5.1.1].

Example 1.1 An example supporting the conjecture is the standard logarithmic n-self-concordant barrier on the positive orthant $\mathbb{R}^n_+$:

$$f(x_1,...,x_n) = -\sum_{i=1}^{n} \log(x_i).$$

Here, $-\log(x)$ is a 1-self-concordant barrier on $(0,+\infty)$ and the separable, symmetric function $f(x_1,...,x_n)$ is an n-self-concordant barrier on $\mathbb{R}^n_{++}$. The corresponding spectral function

$$F(X) = -\log\det(X),$$

is an n-self-concordant barrier on S^n_+, the positive semi-definite cone. (See for instance [5, Theorem 5.4.5], [6, page 25], or [3, Theorem 4.3.3].)

We generalize Example 1.1 by showing that the conjecture is true in the more general setting when f is an arbitrary separable symmetric function. Thus, for the rest of the paper we assume that

$$f(x_1,...,x_n) := g(x_1) + \cdots + g(x_n), \tag{7}$$

where $g(x)$ is a self-concordant (barrier) on the interval (a,b) of the real line. The interval (a,b) may be finite or infinite. Clearly, f is k times (continuously) differentiable at the vector $\lambda(X)$ if and only if g is k times (continuously) differentiable at every eigenvalue $\lambda_i(X)$, $i = 1,...,n$. It can be seen that if g is a ϑ-self-concordant barrier on (a,b) then f is a $n\vartheta$-self-concordant barrier on $(a,b)^n$. Define the *separable spectral function* $F : S^n \to \mathbb{R}$ by

$$F(X) := (f \circ \lambda)(X) = g(\lambda_1(X)) + \cdots + g(\lambda_n(X)). \tag{8}$$

The set

$$\operatorname{dom} F := \{X \in S^n \mid \lambda_i(X) \in (a,b) \text{ for all } i = 1,...,n\}$$

is the domain of $F(X)$. The main result in this work is the following theorem.

Theorem 1.2 *The function $F(X)$ is strongly self-concordant on* $\operatorname{dom} F$ *if and only if $g(x)$ is strongly self-concordant on (a,b). More precisely, if $g(x)$ is strongly* 1*-self-concordant then $F(X)$ is strongly* 22*-self-concordant.*

In addition, if $g(x)$ is a ϑ-self-concordant barrier then $484F(X)$ is a $484n\vartheta$-self-concordant barrier.

The constant 484 in the statement of the theorem can be improved slightly but the reduction comes at the expense of a more complicated and lengthy proof. Since the dimension of the space S^n is $n(n+1)/2$, the universal self-concordant barrier constructed in [5, Theorem 2.5.1] has parameter $\vartheta = O(n^2)$, while the parameter in our result is $O(n)$. This is largely due to the regular nature of the convex set $\operatorname{dom} F$, that is, $X \in \operatorname{dom} F$ if and only if $U^T X U \in \operatorname{dom} F$ for any orthogonal U. A method for constructing self-concordant barriers for the conic hull of the epigraph of certain univariate functions is given in [4].

The following result is crucial, see [8]. (In a different context, for $k \le 3$, it may also be found in [1, Theorem V.3.3 and Exercise V.3.9(ii)].)

Theorem 1.3 *The separable spectral function F is k times (continuously) differentiable at the matrix X if and only if g is k times (continuously) differentiable at each $\lambda_i(X)$, $i = 1,...,n$.*

Since g is convex $g''(x) \ge 0$ for all $x \in \operatorname{dom} g$. Easily, we may assume that

$$g''(x) > 0 \quad \text{for all } x \in \operatorname{dom} g. \tag{9}$$

Indeed, if there is one point x_0 where $g''(x_0) = 0$ then by Corollary 2.1.1 in [5], $g'' \equiv 0$ and Theorem 1.2 becomes trivial.

We show that $F(X)$ satisfies conditions (1), (2) and (3) for X in $\operatorname{dom} F$. Condition (2) is easily implied by condition (5). In Section 4 we show that condition (6) implies that (3) holds with parameter $n\vartheta$. In the remaining sections we show that conditions (5) and (4) imply that (1) holds with $\mathscr{O} = 22$. The reason why we need (5) in order to conclude (1) lies with Lemma 6.3, holding under the assumption that g is a *nondegenerate* strongly 1-self-concordant function. The term “nondegenerate” refers to the condition $g''(x) > 0$ for all $x \in (a,b)$. (For more details refer to the definition of the class $\vec{SC}$ on page 23 in [6], see also Theorem 2.2.5, Theorem 2.5.2 and Theorem 2.5.3 there as well.)

Finally, since F is three times continuously differentiable it is enough to show that conditions (1) and (3) hold for all X in a dense subset of $\operatorname{dom} F$: the set of matrices with distinct eigenvalues.

2 Tensor notation

Let $\mathscr{H}$ be a finite dimensional Euclidean space. A k-tensor on $\mathscr{H}$ is a function $T : \mathscr{H} \times \cdots \times \mathscr{H} \to \mathbb{R}$ linear in each of its k arguments separately. If e_1,...,e_n is a fixed basis of $\mathscr{H}$ then we denote $T^{i_1 \dots i_k} = T[e_{i_1}, ..., e_{i_k}]$. By linearity, for any $h_1, ..., h_k \in \mathscr{H}$, we have

$$T[h_1, ..., h_k] = \sum_{i_1 \dots i_k = 1}^{n,\dots,n} T^{i_1,\dots,i_k} h_1^{i_1} \cdots h_k^{i_k}.$$

When $\mathscr{H} = \mathbb{R}^n$ we fix the standard basis $\{e_i\}_{i=1}^n$, where e_i is the all zero vector with one 1 in position i. When $\mathscr{H} = M^n$ we fix the basis $\{H_{pq}\}_{p,q=1}^{n,n}$ where $H_{pq}^{ij} = \delta_{ip}\delta_{jq}$.

(Here δ_{ip} is the Kronecker symbol equal to 1 if $i = p$ and 0 otherwise.) Every $2k$-tensor on $\mathbb{R}^n$ can be viewed as a k-tensor on M^n:

$$T[H_1,...,H_k] = \sum_{\substack{p_s,q_s=1\\ s=1,...,k}}^{n,...,n} T^{i_1...i_k}_{j_1\cdots j_k} H_1^{i_1 j_1}\cdots H_k^{i_k j_k}.$$

By P^k we denote the set of all $k\times k$ permutation matrices and the set of all permutations on k elements. We often represent a permutation with its cycle decomposition. The set of all k-tensors on $\mathbb{R}^n$ is denoted by $\mathscr{T}^{k,n}$. Dot product between two tensors in $\mathscr{T}^{k,n}$ is defined by:

$$\langle T_1, T_2\rangle = \sum_{p_1,...,p_k=1}^{n} T_1^{p_1\cdots p_k} T_2^{p_1\cdots p_k},$$

inducing the norm $\|\cdot\|$ on $\mathscr{T}^{k,n}$. Given a tensor $T\in\mathscr{T}^{k,n}$ and a permutation $\sigma\in P^k$ we define the $2k$-tensor $\operatorname{Diag}^{\sigma} T$ on $\mathbb{R}^n$ by

$$(\operatorname{Diag}^{\sigma} T)^{i_1\dots i_k}_{j_1\cdots j_k} = \begin{cases} T^{i_1\dots i_k} & \text{if } i_s = j_{\sigma(s)}\ \forall s=1,...,k \\ 0 & \text{otherwise.}\end{cases}$$

Given a permutation $\sigma\in P^k$ we define the σ-Hadamard product between k matrices H_1,...,H_k in M^n ($k\geq 1$) to be the k-tensor

$$(H_1\circ_\sigma H_2 \circ_\sigma \cdots \circ_\sigma H_k)^{i_1\dots i_k} = H_1^{i_1 i_{\sigma^{-1}(1)}}\cdots H_k^{i_k i_{\sigma^{-1}(k)}}.$$

The σ-Hadamard product is linear in each of its arguments separately. A few simple particular cases follow. When $k=2$ and $\sigma=(12)$ (the permutation on $\{1,2\}$ transposing 1 and 2), then $H_1\circ_\sigma H_2 = H_1\circ H_2^T$, where '$\circ$' denotes the ordinary Hadamard product. When $k=2$ and $\sigma=(1)(2)$ (the identity permutation on $\{1,2\}$), then $H_1\circ_\sigma H_2 = (\operatorname{diag} H_1)(\operatorname{diag} H_2)^T$, where $\operatorname{diag} H$ is the vector of diagonal entries of H. Finally, when $k=1$ we have $\circ_\sigma H_1 = \operatorname{diag} H_1$.

Denote by O^n the group of all $n\times n$ orthogonal matrices. Conjugation by an orthogonal matrix $U\in O^n$ is the isometry $T\in\mathscr{T}^{k,n}\mapsto UTU^T\in\mathscr{T}^{k,n}$ defined by

$$(UTU^T)^{i_1\dots i_k} = \sum_{p_1,...,p_k=1}^{n,...,n} T^{p_1\cdots p_k} U^{i_1 p_1}\cdots U^{i_k p_k}.$$

The action is also associative: $V(UTU^T)V^T = (VU)T(VU)^T$ for all $U,V\in O^n$, and the operations are connected by the following multi-linear duality relationship (see [7]):

Theorem 2.1 *For any k-tensor T on $\mathbb{R}^n$, any matrices H_1,...,H_k in M^n, any orthogonal matrix U in O^n, and any permutation σ in P^k we have*

$$\langle T, \tilde H_1\circ_\sigma\cdots\circ_\sigma \tilde H_k\rangle = \big(U(\operatorname{Diag}^{\sigma} T)U^T\big)[H_1,...,H_k], \tag{10}$$

where $\tilde H_i = U^T H_i U$, $i=1,...,k$.

3 Description of the derivatives

Suppose that function $g : (a,b) \to \mathbb{R}$ is three times differentiable and let functions f and F be defined as in (7) and (8). In this section we describe a formula for the first three derivatives of F in terms of g.

Define the function $g^{[(1)]}(x) : (a,b) \to \mathbb{R}$ by

$$g^{[(1)]}(x) = g'(x)$$

and the symmetric function $g^{[(12)]}(x_1,x_2) : (a,b)\times(a,b) \to \mathbb{R}$ by

$$(11) \qquad g^{[(12)]}(x_1,x_2) = \begin{cases} g''(x_1), & \text{if } x_1 = x_2 \\ \dfrac{g^{[(1)]}(x_1) - g^{[(1)]}(x_2)}{x_1 - x_2}, & \text{if } x_1 \neq x_2. \end{cases}$$

The integral representation $g^{[(12)]}(x_1,x_2) = \int_0^1 g''(x_2 + t(x_1 - x_2))\,dt$ shows that $g^{[(12)]}(x_1,x_2)$ is continuously differentiable. Finally, we define

$$(12) \qquad g^{[(123)]}(x_1,x_2,x_3) = \begin{cases} \frac{\partial}{\partial x_2} g^{[(12)]}(x_1,x_2), & \text{if } x_3 = x_2 \\ \dfrac{g^{[(12)]}(x_1,x_3) - g^{[(12)]}(x_1,x_2)}{x_3 - x_2}, & \text{if } x_3 \neq x_2. \end{cases}$$

and

$$(13) \qquad g^{[(132)]}(x_1,x_2,x_3) = \begin{cases} \frac{\partial}{\partial x_1} g^{[(12)]}(x_1,x_2), & \text{if } x_3 = x_1 \\ \dfrac{g^{[(12)]}(x_3,x_2) - g^{[(12)]}(x_1,x_2)}{x_3 - x_1}, & \text{if } x_3 \neq x_1. \end{cases}$$

The formula $g^{[(123)]}(x_1,x_2,x_3) = \int_0^1 \frac{\partial g^{[(12)]}}{\partial x_2}(x_1, x_3 + t(x_2 - x_3))\,dt$ shows that $g^{[(123)]}(x_1,x_2,x_3)$ is a continuous function and similarly for $g^{[(132)]}(x_1,x_2,x_3)$. With the use of these four functions we describe four tensor valued maps, as follows. Let $\mathscr{A}_{(1)} : \mathbb{R}^n \to \mathscr{T}^{1,n}$ be defined by

$$\mathscr{A}_{(1)}^{i_1}(x) = g^{[(1)]}(x_{i_1});$$

let $\mathscr{A}_{(12)} : \mathbb{R}^n \to \mathscr{T}^{2,n}$ be defined by

$$\mathscr{A}_{(12)}^{i_1 i_2}(x) = g^{[(12)]}(x_{i_1}, x_{i_2});$$

and for every permutation σ in the set $\{(123),(132)\}$ let $\mathscr{A}_\sigma : \mathbb{R}^n \to \mathscr{T}^{2,n}$ be defined by

$$\mathscr{A}_\sigma^{i_1 i_2 i_3}(x) = g^{[\sigma]}(x_{i_1}, x_{i_2}, x_{i_3}).$$

With that notation we have the following theorem, see [8].

Theorem 3.1 *Let g be a C^3 function defined on an interval (a,b) and let F be the corresponding separable spectral function. For any X in* $\operatorname{dom} F$*, let U be an orthogonal* *matrix such that* $X = U(\mathrm{Diag}\,\lambda(X))U^T$*. The first three derivatives of* F *at* X *are given by*

$$\nabla F(X) = U\Big(\mathrm{Diag}^{(1)}\mathscr{A}_{(1)}(\lambda(X))\Big)U^T, \tag{14}$$

$$\nabla^2 F(X) = U\Big(\mathrm{Diag}^{(12)}\mathscr{A}_{(12)}(\lambda(X))\Big)U^T, \tag{15}$$

$$\nabla^3 F(X) = U\Big(\mathrm{Diag}^{(123)}\mathscr{A}_{(123)}(\lambda(X)) + \mathrm{Diag}^{(132)}\mathscr{A}_{(132)}(\lambda(X))\Big)U^T. \tag{16}$$

The fact that the formulae above are well defined, no matter what orthogonal matrix U we choose in the ordered spectral decomposition, implies that

$$\nabla^k F(V^T X V) = V^T \nabla^k F(X) V, \quad k = 1,2,3,$$

for any $V \in O^n$, $X \in \mathrm{dom}\, F$. Using Theorem 2.1 we get the next corollary.

Corollary 3.2 *For any symmetric matrices* X, H_1, H_2, H_3 *and any orthogonal* V *we have*

$$\nabla F(X)[\tilde{H}_1] = \nabla F(V^T X V)[H_1],$$
$$\nabla^2 F(X)[\tilde{H}_1, \tilde{H}_2] = \nabla F(V^T X V)[H_1, H_2],$$
$$\nabla^3 F(X)[\tilde{H}_1, \tilde{H}_2, \tilde{H}_3] = \nabla F(V^T X V)[H_1, H_2, H_3],$$

where $\tilde{H}_i = V H_i V^T$ *for* $i = 1,2,3$.

Lemma 3.3 *For every permutation* σ *in* $\{(1),(12),(123),(132)\}$ *the function* $g^{[\sigma]}$ *is symmetric and thus the tensor* $\mathscr{A}_\sigma$ *is symmetric (that is, invariant under permutations of its indexes). In addition* $g^{[(123)]} = g^{[(132)]}$*, or equivalently*

$$\mathscr{A}^{i_1 i_2 i_3}_{(123)}(x) = \mathscr{A}^{i_1 i_2 i_3}_{(132)}(x). \tag{17}$$

Proof. The statement is trivial when the permutation σ is (1) or (12). When $\sigma =$ (123), from (11) and (12) we obtain that

$$g^{[(123)]}(x_1,x_2,x_3) = \begin{cases} g'''(x_1), & \text{if } x_1 = x_2 = x_3 \\ \dfrac{g''(\alpha) - \frac{g'(\alpha)-g'(\beta)}{\alpha-\beta}}{\alpha - \beta}, & \text{if } \alpha \neq \beta \\ \dfrac{\frac{g'(x_1)-g'(x_3)}{x_1-x_3} - \frac{g'(x_1)-g'(x_2)}{x_1-x_2}}{x_3 - x_2}, & \text{if } x_1 \neq x_2 \neq x_3 \neq x_1, \end{cases}$$

where the second case captures all instances when two of the arguments x_1, x_2, x_3 are equal to α and the third is equal to β. The function $g^{[(132)]}(x_1,x_2,x_3)$ allows a similar representation. When x_1, x_2, and x_3 are all different we have the representation

$$g^{[(123)]}(x_1,x_2,x_3) = \frac{\begin{vmatrix} g'(x_1) & g'(x_2) & g'(x_3) \\ x_1 & x_2 & x_3 \\ 1 & 1 & 1 \end{vmatrix}}{\begin{vmatrix} x_1^2 & x_2^2 & x_3^2 \\ x_1 & x_2 & x_3 \\ 1 & 1 & 1 \end{vmatrix}} = g^{[(132)]}(x_1,x_2,x_3),$$

showing that $g^{[(123)]}$ is symmetric. We deal in a similar manner with the function $g^{[(132)]}$. Equation (17) is now easily verified. ■

4 Verification that F satisfies (3)

In this subsection we show that if g satisfies (6) then F satisfies condition (3) with parameter $n\vartheta$, that is

$$|\nabla F(X)[H]| \leq \sqrt{n\vartheta}(\nabla^2 F(X)[H,H])^{1/2}, \tag{18}$$

for any $X \in \operatorname{dom} F$ and any $H \in S^n$. Substituting (14) and (15) into (18) results in

$$\left|U\left(\operatorname{Diag}^{(1)}\mathscr{A}_{(1)}(\lambda(X))\right)U^T[H]\right| \leq \sqrt{n\vartheta}\left\{U\left(\operatorname{Diag}^{(12)}\mathscr{A}_{(12)}(\lambda(X))\right)U^T[H,H]\right\}^{1/2}.$$

It is enough to show that the inequality holds for any fixed matrix X with distinct eigenvalues and all symmetric H. By Formula (10) the last inequality is equivalent to

$$|\langle\mathscr{A}_{(1)}(\lambda(X)), \circ_{(1)}\tilde{H}\rangle| \leq \sqrt{n\vartheta}\langle\mathscr{A}_{(12)}(\lambda(X)), \tilde{H}\circ_{(12)}\tilde{H}\rangle^{1/2},$$

where $\tilde{H} = U^T H U$, which in turn is equivalent to

$$|\langle\mathscr{A}_{(1)}(x), \circ_{(1)}H\rangle| \leq \sqrt{n\vartheta}\langle\mathscr{A}_{(12)}(x), H\circ_{(12)}H\rangle^{1/2},$$

where $x \in \mathbb{R}^n_{>} := \{x \in \mathbb{R}^n \,|\, x_1 > \cdots > x_n\}$. Since $\circ_{(1)}H = \operatorname{diag} H$ and since for symmetric H we have $H\circ_{(12)}H = H\circ H$, we obtain

$$|\langle\mathscr{A}_{(1)}(x), \operatorname{diag} H\rangle| \leq \sqrt{n\vartheta}\langle\mathscr{A}_{(12)}(x), H\circ H\rangle^{1/2}. \tag{19}$$

By the definitions in Section 3, $\mathscr{A}^{i}_{(1)}(x) = g'(x_i)$ and $\mathscr{A}^{i_1 i_2}_{(12)}(x) = g^{[(12)]}(x_{i_1}, x_{i_2})$, allowing us to write both sides of (19) in more detail

$$\Big|\sum_{i_1=1}^{n} g'(x_{i_1})H^{i_1 i_1}\Big| \leq \sqrt{n\vartheta}\Big(\sum_{i_1=1}^{n} g''(x_{i_1})(H^{i_1 i_1})^2 + \sum_{\substack{i_1,i_2=1\\ i_1\neq i_2}}^{n,n} g^{[(12)]}(x_{i_1}, x_{i_2})(H^{i_1 i_2})^2\Big)^{1/2},$$

where we used $g^{[(12)]}(x_{i_1}, x_{i_2}) = g''(x_{i_1})$ when $i_1 = i_2$. The fact that g satisfies (3) shows that

$$|g'(x_{i_1})H^{i_1 i_1}| \leq \sqrt{\vartheta}\big(g''(x_{i_1})(H^{i_1 i_1})^2\big)^{1/2}.$$

The Cauchy-Schwartz inequality and the fact that $g^{[(12)]}(x_{i_1}, x_{i_2})$ is non-negative, by the convexity of g, conclude the verification that F satisfies (3) with parameter $n\vartheta$.

5 Breaking down inequality (1) for F

Our ultimate goal is to show that if g satisfies (4) and (5) then there is a constant $\mathscr{C}$ depending only on F such that

$$|\nabla^3 F(X)[H,H,H]| \leq 2\mathscr{C}(\nabla^2 F(X)[H,H])^{3/2}, \tag{20}$$

holds for any $X \in \operatorname{dom} F$ and any symmetric H. In particular, we show that $\mathscr{O} = 22$ suffices. Since F is three times continuously differentiable it is enough to prove (20) for a fixed X with distinct eigenvalues and any symmetric H. The goal in this section is to break down (20) into simpler pieces. We deal with each of these pieces in the sequel.

Let $X = U(\operatorname{Diag}\lambda(X))U^T$ be the ordered spectral decomposition of X. By Theorem 2.1 and Theorem 3.1, using the notation $\tilde{H} = U^T HU$, we obtain

$$\begin{aligned}\nabla^2 F(X)[H,H] &= U\Big(\operatorname{Diag}^{(12)}\mathscr{A}_{(12)}(\lambda(X))\Big)U^T[H,H] \\ &= \langle \mathscr{A}_{(12)}(\lambda(X)), \tilde{H} \circ_{(12)} \tilde{H}\rangle\end{aligned}$$

and

$$\begin{aligned}&\nabla^3 F(X)[H,H,H] \\ &\quad = U\Big(\operatorname{Diag}^{(123)}\mathscr{A}_{(123)}(\lambda(X)) + \operatorname{Diag}^{(132)}\mathscr{A}_{(132)}(\lambda(X))\Big)U^T[H,H,H] \\ &\quad = \langle \mathscr{A}_{(123)}(\lambda(X)), \tilde{H} \circ_{(123)} \tilde{H} \circ_{(123)} \tilde{H}\rangle + \langle \mathscr{A}_{(132)}(\lambda(X)), \tilde{H} \circ_{(132)} \tilde{H} \circ_{(132)} \tilde{H}\rangle.\end{aligned}$$

Since $\lambda(X) \in R^n_{>}$ is an arbitrary vector we substitute it by an $x \in \mathbb{R}^n_{>}$. Analogously, since H is an arbitrary symmetric matrix, we may drop the tilde on top of it. Thus, recalling that for symmetric H we have $H \circ_{(12)} H = H \circ H$, it is enough to show that the following inequality holds for any $x \in \mathbb{R}^n_{>}$ and any symmetric H:

$$\begin{aligned}&|\langle \mathscr{A}_{(123)}(x), H \circ_{(123)} H \circ_{(123)} H\rangle + \langle \mathscr{A}_{(132)}(x), H \circ_{(132)} H \circ_{(132)} H\rangle| \\ &\qquad\qquad\qquad\qquad \le 2\mathscr{O}(\langle \mathscr{A}_{(12)}(x), H \circ H\rangle)^{3/2}.\end{aligned}$$

In other words we have to show that

$$\begin{aligned}&\Big|\sum_{i_1,i_2,i_3=1}^{n,n,n} \mathscr{A}_{(123)}^{i_1 i_2 i_3}(x) H^{i_1 i_3} H^{i_2 i_1} H^{i_3 i_2} + \sum_{j_1,j_2,j_3=1}^{n,n,n} \mathscr{A}_{(132)}^{j_1 j_2 j_3}(x) H^{j_1 j_2} H^{j_2 j_3} H^{j_3 j_1}\Big| \\ &\qquad\qquad\qquad \le 2\mathscr{O}\Big(\sum_{i_1,i_2=1}^{n,n} \mathscr{A}_{(12)}^{i_1 i_2}(x) H^{i_1 i_2} H^{i_1 i_2}\Big)^{3/2}.\end{aligned}$$

Changing the order of summation in the second sum on the left-hand side according to the substitution $j_1 \to i_3$, $j_2 \to i_1$, and $j_3 \to i_2$ and using the fact that H is a symmetric matrix allows us to combine the two sums

$$\begin{aligned}&\Big|\sum_{i_1,i_2,i_3=1}^{n,n,n} \big(\mathscr{A}_{(123)}^{i_1 i_2 i_3}(x) + \mathscr{A}_{(132)}^{i_3 i_1 i_2}(x)\big) H^{i_1 i_3} H^{i_2 i_1} H^{i_3 i_2}\Big| \\ &\qquad\qquad\qquad \le 2\mathscr{O}\Big(\sum_{i_1,i_2=1}^{n,n} \mathscr{A}_{(12)}^{i_1 i_2}(x)(H^{i_1 i_2})^2\Big)^{3/2}.\end{aligned}$$

By (17) and the fact that $\mathscr{A}_\sigma$ is a symmetric tensor we get

$$\mathscr{A}_{(132)}^{i_3 i_1 i_2} = \mathscr{A}_{(123)}^{i_3 i_1 i_2} = \mathscr{A}_{(123)}^{i_1 i_2 i_3}.$$

That allows us to simplify the inequality to

$$\Big|\sum_{i,j,k=1}^{n,n,n} \mathscr{A}_{(123)}^{ijk}(x) H^{ik} H^{ji} H^{kj}\Big| \le \mathscr{O}\Big(\sum_{i,j=1}^{n,n} \mathscr{A}_{(12)}^{ij}(x)(H^{ij})^2\Big)^{3/2}. \tag{21}$$

We now identify the pieces on each side. The following five lines define two groups of quantities

$$\begin{aligned}
A &:= \sum_{i=1}^{n} a_i, &\text{where}\quad & a_i := g''(x_i)(H^{ii})^2;\\
B &:= \sum_{\substack{i,j=1\\ i\neq j}}^{n} b_{ij}, &\text{where}\quad & b_{ij} := \frac{g'(x_i)-g'(x_j)}{x_i-x_j}(H^{ij})^2;\\
C &:= \sum_{i=1}^{n} c_i, &\text{where}\quad & c_i := g'''(x_i)(H^{ii})^3;\\
D &:= \sum_{\substack{i,j=1\\ i\neq j}}^{n} d_{ij}, &\text{where}\quad & d_{ij} := \frac{g''(x_i)-\frac{g'(x_i)-g'(x_j)}{x_i-x_j}}{x_i-x_j}(H^{ij})^2 H^{ii};\ \text{and}\\
E &:= \sum_{\substack{i,j,k=1\\ i\neq j\neq k\neq i}}^{n} e_{ijk}, &\text{where}\quad & e_{ijk} := \frac{\frac{g'(x_j)-g'(x_i)}{x_j-x_i}-\frac{g'(x_k)-g'(x_j)}{x_k-x_j}}{x_k-x_i} H^{ij}H^{jk}H^{ik}.
\end{aligned} \tag{22}$$

The numbers b_{ij} and d_{ij} are not defined when $i=j$; e_{ijk} is not defined when two of the indexes are equal; a_i and b_{ij} are non-negative.

Now we recall the fact that it is enough to prove (1) only when X has distinct eigenvalues. In that case, x has distinct coordinates and the two sides of (21) can be expressed as

$$\sum_{i,j=1}^{n,n} \mathscr{A}_{(12)}^{ij}(x)(H^{ij})^2 = A+B;$$
$$\sum_{i,j,k=1}^{n,n,n} \mathscr{A}_{(123)}^{ijk}(x) H^{ik} H^{ji} H^{kj} = C+3D+E.$$

Thus we need to show that

$$|C+3D+E| \le \mathscr{O}(A+B)^{3/2},$$

or, squaring both sides, to show that

$$C^2+9D^2+E^2+6CD+6DE+2CE \le \mathscr{O}^2(A^3+3A^2B+3AB^2+B^3). \tag{23}$$

The next lemma shows that in order to establish (23) it is sufficient to establish

$$|C| \le oA^{3/2}, \tag{24}$$

$$(25) \qquad |D| \le oA^{1/2}B,$$

$$(26) \qquad |E| \le oB^{3/2},$$

for some constant o.

Lemma 5.1 *Inequalities (24), (25) and (26) imply (23) with $\mathscr{O} = \sqrt{\frac{13}{3}}o$.*

Proof. Suppose that (24), (25) and (26) are valid. Using the arithmetic-geometric mean inequality in the last two relationships below, we have

$$\begin{array}{rclcccccc}
C^2 & \le & o^2A^3 & & & & & & \\
9D^2 & \le & & & & & 3o^2(3AB^2) & & \\
E^2 & \le & & & & & & & o^2B^3 \\
6CD & \le & & 2o^2(3A^2B) & & & & & \\
6DE & \le & & & & & o^2(3AB^2) & + & 3o^2B^3 \\
2CE & \le & & \frac{1}{3}o^2(3A^2B) & + & & \frac{1}{3}o^2(3AB^2) & &
\end{array}$$

The lemma follows by summing up the rows and the fact that A and B are non-negative quantities. ■

Proof of Theorem 1.2. In Section 6 we will show that (24), (25) and (26) hold for $o = 10.511$. Consequently, by Lemma 5.1, inequality (23) holds for $\mathscr{O} = 22$. This establishes that if $g(x)$ is strongly 1-self-concordant then $F(X)$ is strongly 22-self-concordant.

A straightforward calculation shows that if F satisfies

$$\begin{aligned}
|\nabla F(X)[H]| &\le \sqrt{n\vartheta}\,(\nabla^2 F(X)[H,H])^{1/2}, \\
|\nabla^3 F(X)[H,H,H]| &\le 2\,\mathscr{O}(\nabla^2 F(X)[H,H])^{3/2},
\end{aligned}$$

for some positive constant $\mathscr{O}$, then $\tilde{F} := \mathscr{O}^2 F$ satisfies

$$\begin{aligned}
|\nabla \tilde{F}(X)[H]| &\le \sqrt{\mathscr{O}^2 n\vartheta}\,(\nabla^2 \tilde{F}(X)[H,H])^{1/2}, \\
|\nabla^3 \tilde{F}(X)[H,H,H]| &\le 2\,(\nabla^2 \tilde{F}(X)[H,H])^{3/2},
\end{aligned}$$

This finally establishes that if g is a ϑ-self-concordant barrier on (a,b), then $484(f \circ \lambda)$ is a $(484n\vartheta)$-self-concordant barrier on its domain. ■

6 Proofs of (24), (25) and (26)

The following lemma allows us to reduce the inequalities (24), (25) and (26) respectively to inequalities between their constituent components.

Lemma 6.1 *Let $a_i, b_{ij}, c_i, d_{ij}, e_{ijk}$ and A, B, C, D, E be as in (22). Then*

(a) *If $|c_i| \le oa_i^{3/2}$ for $i \in \{1, \ldots, n\}$ then $|C| \le oA^{3/2}$.*

(b) *If* $|d_{ij}| \le o a_i^{1/2} b_{ij}$ *for all distinct* $i,j \in \{1,\dots,n\}$ *then* $|D| \le oA^{1/2}B$.

(c) *If* $|e_{ijk}| \le o(b_{ij}b_{jk}b_{ki})^{1/2}$ *for all distinct* $i,j,k \in \{1,\dots,n\}$ *then* $|E| \le oB^{3/2}$.

By Lemma 6.1, (24), (25) and (26), are immediate consequences of the following lemma.

Lemma 6.2 *Assume* $g : \mathbb{R} \to \mathbb{R}$ *is a strongly self-concordant function with* $\operatorname{dom} g = (a,b)$. *Let* $x_i, x_j, x_k \in (a,b)$ *be different. Then with the constant* $o = 10.511$ *we have*

$$|g'''(x_i)| \le o|g''(x_i)|^{3/2}, \tag{27}$$

$$\left| \frac{g''(x_i) - \frac{g'(x_i)-g'(x_j)}{x_i-x_j}}{x_i - x_j} \right| \le o(g''(x_i))^{1/2} \left(\frac{g'(x_i) - g'(x_j)}{x_i - x_j} \right), \tag{28}$$

and

$$\left| \frac{\frac{g'(x_j)-g'(x_i)}{x_j-x_i} - \frac{g'(x_k)-g'(x_j)}{x_k-x_j}}{x_k - x_i} \right| \le o \left(\frac{g'(x_j) - g'(x_i)}{x_j - x_i} \frac{g'(x_j) - g'(x_k)}{x_j - x_k} \frac{g'(x_k) - g'(x_i)}{x_k - x_i} \right)^{1/2}. \tag{29}$$

We next present the proofs of Lemma 6.1 and Lemma 6.2.

6.1 Proof of Lemma 6.1

(a) Recall the formulae

$$|C|^2 = \sum_{i=1}^n c_i^2 + 2\sum_{i<j} c_i c_j \quad \text{and}$$
$$A^3 = \sum_{i=1}^n a_i^3 + 3\sum_{i<j}(a_i^2 a_j + a_i a_j^2) + 6 \sum_{i<j<k} a_i a_j a_k.$$

By the assumptions we have $\sum_{i=1}^n c_i^2 \le o^2(\sum_{i=1}^n a_i^3)$. The arithmetic-geometric mean inequality shows

$$o^2(a_i^2 a_j + a_i a_j^2) \ge 2o^2(a_i^3 a_j^3)^{1/2} \ge 2|c_i c_j|,$$

for any $i < j$. The lemma follows since the a's are non-negative.

(b) Let $a = \max\{a_i \,|\, i = 1,\dots,n\}$. Then

$$|D| = \Big| \sum_{\substack{i,j=1 \\ i \ne j}}^n d_{ij} \Big| \le \sum_{\substack{i,j=1 \\ i \ne j}}^n |d_{ij}| \le o \sum_{\substack{i,j=1 \\ i \ne j}}^n a_i^{1/2} b_{ij} \le o\, a^{1/2} \sum_{\substack{i,j=1 \\ i \ne j}}^n b_{ij} \le oA^{1/2}B.$$

(c) To ease notation assume $o = 1$; the argument for general $o \geq 1$ is analogous. First, we prove the following inequality ($m \leq n$)

$$\sqrt{2}\Big|\sum_{\substack{i,j=1\\ i<j}}^{m-1} e_{ijm}\Big| \leq \Big(\sum_{\substack{i,j=1\\ i<j}}^{m-1} b_{ij}\Big)^{1/2}\Big(\sum_{i=1}^{m-1} b_{im}\Big). \tag{30}$$

To show (30), square the left-hand side and bound it as follows:

$$\begin{aligned} 2\Big(\sum_{\substack{i,j=1\\ i<j}}^{m-1} e_{ijm}\Big)^2 &\leq 2\Big(\sum_{\substack{i,j=1\\ i<j}}^{m-1} \sqrt{b_{ij}}\sqrt{b_{im}b_{jm}}\Big)^2 \leq 2\Big(\sum_{\substack{i,j=1\\ i<j}}^{m-1} b_{ij}\Big)\Big(\sum_{\substack{i,j=1\\ i<j}}^{m-1} b_{im}b_{jm}\Big) \\ &\leq \Big(\sum_{\substack{i,j=1\\ i<j}}^{m-1} b_{ij}\Big)\Big(\sum_{i=1}^{m-1} b_{im}\Big)^2. \end{aligned}$$

In the first inequality we used the given relationship between the numbers e_{ijk} and b_{ij}. In the second, we applied the Cauchy-Schwarz inequality.

Next, observe that

$$|E| = 6\Big|\sum_{\substack{i,j,k=1\\ i<j<k}}^{n} e_{ijk}\Big| \text{ and } B^{3/2} = \sqrt{8}\Big(\sum_{\substack{i,j=1\\ i<j}}^{n} b_{ij}\Big)^{3/2}.$$

Therefore the inequality $|E| \leq |B|^{3/2}$ is equivalent to

$$\Big(\sum_{\substack{i,j,k=1\\ i<j<k}}^{n} e_{ijk}\Big)^2 \leq \frac{2}{9}\Big(\sum_{\substack{i,j=1\\ i<j}}^{n} b_{ij}\Big)^3. \tag{31}$$

We next prove (31) by induction on n. The base case $n = 3$ is

$$e_{123}^2 \leq \frac{2}{9}(b_{12}+b_{13}+b_{23})^3,$$

which is verified easily with the arithmetic-geometric mean inequality. Assume that (31) holds for $n = m-1$. Then, for $n = m$ the left-hand side in (31) expands to

$$\Big(\sum_{\substack{i,j,k=1\\ i<j<k}}^{m} e_{ijk}\Big)^2 = \Big(\sum_{\substack{i,j,k=1\\ i<j<k}}^{m-1} e_{ijk}\Big)^2 + 2\Big(\sum_{\substack{i,j,k=1\\ i<j<k}}^{m-1} e_{ijk}\Big)\Big(\sum_{\substack{i,j=1\\ i<j}}^{m-1} e_{ijm}\Big) + \Big(\sum_{\substack{i,j=1\\ i<j}}^{m-1} e_{ijm}\Big)^2.$$

On the other hand the right-hand side of (31) is

$$\begin{aligned} \frac{2}{9}\Big(\sum_{\substack{i,j=1\\ i<j}}^{m} b_{ij}\Big)^3 &= \frac{2}{9}\Big(\sum_{\substack{i,j=1\\ i<j}}^{m-1} b_{ij}\Big)^3 + \frac{2}{3}\Big(\sum_{\substack{i,j=1\\ i<j}}^{m-1} b_{ij}\Big)^2\Big(\sum_{i=1}^{m-1} b_{im}\Big) \\ &\quad + \frac{2}{3}\Big(\sum_{\substack{i,j=1\\ i<j}}^{m-1} b_{ij}\Big)\Big(\sum_{i=1}^{m-1} b_{im}\Big)^2 + \frac{2}{9}\Big(\sum_{i=1}^{m-1} b_{im}\Big)^3. \end{aligned}$$

Using the induction hypothesis between the first terms in those expansions, we see that it is enough to show:

$$2\Big(\sum_{\substack{i,j,k=1\\ i<j<k}}^{m-1} e_{ijk}\Big)\Big(\sum_{\substack{i,j=1\\ i<j}}^{m-1} e_{ijm}\Big)+\Big(\sum_{\substack{i,j=1\\ i<j}}^{m-1} e_{ijm}\Big)^2 \\ \le \frac{2}{3}\Big(\sum_{\substack{i,j=1\\ i<j}}^{m-1} b_{ij}\Big)^2\Big(\sum_{i=1}^{m-1} b_{im}\Big)+\frac{2}{3}\Big(\sum_{\substack{i,j=1\\ i<j}}^{m-1} b_{ij}\Big)\Big(\sum_{i=1}^{m-1} b_{im}\Big)^2+\frac{2}{9}\Big(\sum_{i=1}^{m-1} b_{im}\Big)^3.$$

By the induction hypothesis and (30) we can bound the left-hand side as follows:

$$2\Big(\sum_{\substack{i,j,k=1\\ i<j<k}}^{m-1} e_{ijk}\Big)\Big(\sum_{\substack{i,j=1\\ i<j}}^{m-1} e_{ijm}\Big)+\Big(\sum_{\substack{i,j=1\\ i<j}}^{m-1} e_{ijm}\Big)^2 \\ \le \frac{2}{3}\Big(\sum_{\substack{i,j=1\\ i<j}}^{m-1} b_{ij}\Big)^{3/2}\Big(\sum_{\substack{i,j=1\\ i<j}}^{m-1} b_{ij}\Big)^{1/2}\Big(\sum_{i=1}^{m-1} b_{im}\Big)+\frac{1}{2}\Big(\sum_{\substack{i,j=1\\ i<j}}^{m-1} b_{ij}\Big)\Big(\sum_{i=1}^{m-1} b_{im}\Big)^2.$$

The result follows using the non-negativity of the numbers b_{ij}. ■

6.2 Proof of Lemma 6.2

First of all, observe that by letting $x_i \to x_j$ inequality (29) implies (28) after a suitable relabeling of indices. In addition, since g is self-concordant it satisfies (4) and thus inequality (27) holds as long as $o \ge 2$. Consequently, to prove Lemma 6.2, it suffices to prove (29) for $o = 10.511$.

By Lemma 3.3, both sides are symmetric with respect to x_i, x_j and x_k. Thus, we may assume that $x_i < x_j < x_k$. In addition, substituting $g(x)$ with $g(-x)$ if necessary, we may assume that $x_j - x_i \ge x_k - x_j$. By the Mean-Value Theorem, there exists $z \in [x_j, x_k]$ such that

$$\frac{g'(x_k)-g'(x_j)}{x_k-x_j} = g''(z).$$

Further adjustment of the points x_i, x_j and x_k is achieved by defining the function

$$\tilde{g}(t) := g((1-t)x_i + tx_k) - g'(x_i)(x_k - x_i)t - g(x_i).$$

Notice that $\tilde{g}$ is convex and satisfies (5) and (4) if and only if g does. Thus, it is enough to prove (29) after substituting g by $\tilde{g}$. Next, since $x_i < x_j < x_k$ and $x_j - x_i \ge x_k - x_j$ there are numbers t_2 and t_z in $[1/2, 1]$ such that $t_2 \le t_z$ and

$$x_j = (1-t_2)x_i + t_2 x_k,\\ z = (1-t_z)x_i + t_z x_k.$$

Since $\tilde{g}(0) = 0$ and $\tilde{g}'(0) = 0$, the function $\tilde{g}(t)$ achieves its minimum 0 at $t = 0$. In addition, we have $\tilde{g}''(t) > 0$ and therefore $0 \le \tilde{g}'(t_2) \le \tilde{g}'(1)$. Utilizing the relationships

$$\begin{aligned}
\tilde{g}'(1) &= (g'(x_k) - g'(x_i))(x_k - x_i),\\
\tilde{g}'(t_2) &= (g'(x_j) - g'(x_i))(x_k - x_i),\\
\tilde{g}''(1) &= g''(x_k)(x_k - x_i)^2,\\
\tilde{g}''(t_z) &= g''(z)(x_k - x_i)^2,
\end{aligned}$$

we can simplify the left-hand side of (29):

$$\left| \frac{\frac{g'(x_j)-g'(x_i)}{x_j-x_i} - \frac{g'(x_k)-g'(x_j)}{x_k-x_j}}{x_k - x_i} \right| = \left| \frac{\frac{\tilde{g}'(t_2)}{(x_k-x_i)(x_j-x_i)} - g''(z)}{x_k - x_i} \right| = \left| \frac{\frac{\tilde{g}'(t_2)}{t_2(x_k-x_i)^2} - \frac{\tilde{g}''(t_z)}{(x_k-x_i)^2}}{x_k - x_i} \right|$$

$$= \frac{|\tilde{g}'(t_2) - t_2\tilde{g}''(t_z)|}{t_2(x_k - x_i)^3}.$$

Similarly, the right-hand side of (29) simplifies to

$$\left(\frac{g'(x_j) - g'(x_i)}{x_j - x_i} \frac{g'(x_j) - g'(x_k)}{x_j - x_k} \frac{g'(x_k) - g'(x_i)}{x_k - x_i} \right)^{1/2}$$

$$= \left(\frac{\tilde{g}'(t_2)}{t_2(x_k - x_i)^2} \frac{\tilde{g}''(t_z)}{(x_k - x_i)^2} \frac{\tilde{g}'(1)}{(x_k - x_i)^2} \right)^{1/2} = \frac{\sqrt{\tilde{g}'(t_2)\tilde{g}''(t_z)\tilde{g}'(1)}}{t_2^{1/2}(x_k - x_i)^3}.$$

Thus, in terms of $\tilde{g}$, inequality (29) is equivalent to

$$\text{(32)} \qquad \left| t_2\tilde{g}''(t_z) - \tilde{g}'(t_2) \right| \le o\big(t_2\tilde{g}'(t_2)\tilde{g}''(t_z)\tilde{g}'(1)\big)^{1/2},$$

where $1/2 \le t_2 \le t_z \le 1$.

The remainder of this section is devoted to proving (32), which in turn proves Lemma 6.2. To that end, we will rely on the next two lemmas. The first is [6, Theorem 2.2.5] together with [6, Theorem 2.5.3].

Given a nondegenerate self-concordant function $g : \mathbb{R} \to \mathbb{R}$ and x in $\operatorname{dom} g$, let

$$n(x) := g'(x)/(g''(x))^{1/2}.$$

Lemma 6.3 *Suppose that g is a nondegenerate strongly 1-self-concordant function on interval (a,b). If $n(x) \le 1/4$ then g has a minimizer x^* and*

$$\text{(33)} \qquad |x^* - x|(g''(x))^{1/2} \le n(x) + \frac{3n^2(x)}{(1 - n(x))^3}.$$

Lemma 6.4 *Suppose that g is a strongly 1-self-concordant function on interval (a,b). For every $x \in \operatorname{dom} g$, if $\alpha_x(y) := |x - y|(g''(x))^{1/2} < 1$, then $y \in \operatorname{dom} g$ and*

$$\text{(34)} \qquad (1 - \alpha_x(y))^2 g''(x) \le g''(y) \le \frac{g''(x)}{(1 - \alpha_x(y))^2};$$

(35) $$|g''(x) - g''(y)| \leq \frac{g''(x)}{(1-\alpha_x(y))^2} - g''(x).$$

Proof. Inequality (34) is a well known result, see [5, Theorem 2.1.1]. Inequality (35) is a consequence from (34). Indeed, if $g''(x) \leq g''(y)$ it is the right-hand side of (34), while in the case $g''(x) \geq g''(y)$ it follows from the left-hand side of (34) and the arithmetic-geometric mean inequality. ∎

We will often refer to the following function

$$h(x) := x + 3x^2/(1-x)^3.$$

Observe that $h(x)$ is strictly increasing on the interval $[0,1)$.

Lemma 6.5 *There is a unique solution c^* to the optimization problem*

$$\min\max\left\{\frac{3-c}{\sqrt{2}(1-3c+c^2)}, \frac{1}{h^{-1}(c/(2+c))}\right\},$$

where the minimization is taken over values of c such that $c \in (0,(3-\sqrt{5})/2)$. The optimal solution is $c^* \approx 0.3035902890$ and

$$\frac{3-c^*}{\sqrt{2}(1-3c^*+(c^*)^2)} = \frac{1}{h^{-1}(c^*/(2+c^*))}.$$

Denoting that common value by o we have

$$2 \leq o \leq 10.511.$$

Proof. The functions in the optimization problem are well-defined for none of the denominators becomes zero. Observe that

$$\operatorname{argmin}\max\left\{\frac{3-c}{\sqrt{2}(1-3c+c^2)}, \frac{1}{h^{-1}(c/(2+c))}\right\} = \operatorname{argmax}\min\left\{\frac{\sqrt{2}(1-3c+c^2)}{3-c}, h^{-1}(c/(2+c))\right\}.$$

In addition, the function

(36) $$x \mapsto \frac{\sqrt{2}(1-3x+x^2)}{3-x}$$

is decreasing, while the function

$$x \mapsto h^{-1}\left(\frac{x}{2+x}\right)$$

is increasing over the interval $[0,(3-\sqrt{5})/2)$. It is easy to see that at $x=0$ the first function is bigger than the second, while at $x=(3-\sqrt{5})/2$ the situation is reversed.

Therefore, they must intersect each other. Using MAPLE, we find that this happens at $c^* \approx 0.3035902890$. Thus $o \approx 10.51096786 \le 10.511$. ■

Let c^* and o be the constants defined in Lemma 6.5. We prove (32) by considering two cases depending on the size of $\tilde{g}''(t_z)$ relatively to $(c^*)^2$.

Case 1. Suppose that $\tilde{g}''(t_z) \le (c^*)^2$. Since $c^* < 1$ we have

$$\alpha_{t_z}(t) = (t_z - t)\sqrt{\tilde{g}''(t_z)} < 1 \text{ for all } t \in [0, t_2].$$

Thus, (35) shows that

$$\left|\tilde{g}''(t_z) - \tilde{g}''(t)\right| \le \frac{\tilde{g}''(t_z)}{(1-(t_z-t)\sqrt{\tilde{g}''(t_z)})^2} - \tilde{g}''(t_z),$$

for all $t \in [0, t_2]$. Integrating the last inequality we get

$$\begin{aligned}
\left|t_2\tilde{g}''(t_z) - \tilde{g}'(t_2)\right| &\le \int_0^{t_2} |\tilde{g}''(t_z) - \tilde{g}''(t)|\, dt \\
&\le \int_0^{t_2} \Big(\frac{\tilde{g}''(t_z)}{(1-(t_z-t)\sqrt{\tilde{g}''(t_z)})^2} - \tilde{g}''(t_z)\Big)\, dt \qquad (37)\\
&= t_2\tilde{g}''(t_z)\Big(\frac{1}{(1-(t_z-t_2)\sqrt{\tilde{g}''(t_z)})(1-t_z\sqrt{\tilde{g}''(t_z)})} - 1\Big).
\end{aligned}$$

To simplify the notation, define

$$\beta := (1-(t_z-t_2)\sqrt{\tilde{g}''(t_z)})(1-t_z\sqrt{\tilde{g}''(t_z)}) > 0.$$

Recalling that $\tilde{g}'(t_2) \le \tilde{g}'(1)$, we see that in order to prove (32), it is enough to show that

$$\sqrt{t_2\tilde{g}''(t_z)}\Big(\frac{1}{\beta} - 1\Big) \le o\tilde{g}'(t_2). \qquad (38)$$

Utilizing (37) again we find that

$$t_2\tilde{g}''(t_z)\Big(2 - \frac{1}{\beta}\Big) \le \tilde{g}'(t_2),$$

which shows that in order to prove (38) it is sufficient to show

$$(1-\beta) \le o\sqrt{t_2\tilde{g}''(t_z)}(2\beta - 1). \qquad (39)$$

Using the fact that $1/2 \le t_2 \le t_z \le 1$ we estimate

$$\beta \ge (1 - 0.5\sqrt{\tilde{g}''(t_z)})(1 - \sqrt{\tilde{g}''(t_z)}).$$

Substituting that estimate for β in (39), using the fact that $1/2 \le t_2$ and simplifying we find that it is sufficient to prove

$$3 - \sqrt{\tilde{g}''(t_z)} \le o\sqrt{2}(1 - 3\sqrt{\tilde{g}''(t_z)} + \tilde{g}''(t_z)).$$

By Lemma 6.5, $\sqrt{\tilde{g}''(t_z)} \leq c^* \in (0,(3-\sqrt{5})/2)$, implying that $(1-3\sqrt{\tilde{g}''(t_z)}+\tilde{g}''(t_z)) > 0$. Since the reciprocal of function (36) is increasing over the interval $(0,(3-\sqrt{5})/2)$ we conclude

$$\frac{3-\sqrt{\tilde{g}''(t_z)}}{\sqrt{2}\left(1-3\sqrt{\tilde{g}''(t_z)}+\tilde{g}''(t_z)\right)} \leq \frac{3-c^*}{\sqrt{2}\left(1-3c^*+(c^*)^2\right)} = o,$$

by the definition of o in Lemma 6.5.

Case 2. Suppose now that $\tilde{g}''(t_z) \geq (c^*)^2$. Using that assumption together with facts that $\tilde{g}'(t_2) \leq \tilde{g}'(1)$ and $t_2 \geq 1/2$ we estimate

$$o\sqrt{t_2\tilde{g}'(t_2)\tilde{g}'(1)\tilde{g}''(t_z)} \geq o\frac{c^*}{\sqrt{2}}\tilde{g}'(t_2) \geq \tilde{g}'(t_2),$$

where we used that by the definition of o and c^* in Lemma 6.5 we have

$$o \geq \frac{\sqrt{2}}{c^*}.$$

Thus, in order to completely establish (32) it suffices to show that

$$o\sqrt{\tilde{g}'(t_2)\tilde{g}'(1)} \geq \sqrt{t_2\tilde{g}''(t_z)}. \tag{40}$$

Since we have

$$\tilde{g}''(t_z) = \frac{\tilde{g}'(1)-\tilde{g}'(t_2)}{1-t_2},$$

substituting into (40) and squaring both sides, we can rewrite the inequality as

$$o^2\Big(\frac{1-t_2}{t_2}\Big) \geq \frac{\tilde{g}'(1)-\tilde{g}'(t_2)}{\tilde{g}'(1)\tilde{g}'(t_2)}.$$

Since $t_2 \leq 1$ we will be done if we show that

$$o^2(1-t_2) \geq \frac{\tilde{g}'(1)-\tilde{g}'(t_2)}{\tilde{g}'(1)\tilde{g}'(t_2)}. \tag{41}$$

Since

$$\frac{\tilde{g}'(1)-\tilde{g}'(t_2)}{\tilde{g}'(1)\tilde{g}'(t_2)} = \int_{t_2}^{1} \frac{\tilde{g}''(\tau)}{(\tilde{g}'(\tau))^2}\,d\tau,$$

and $1/2 \leq t_2 \leq 1$, to prove (41) it suffices to show

$$o^2 \geq \frac{\tilde{g}''(\tau)}{(\tilde{g}'(\tau))^2} \equiv \frac{1}{(\tilde{n}(\tau))^2}, \text{ for all } \tau \in [1/2,1]. \tag{42}$$

To prove (42) proceed by contradiction: Assume on the contrary that there exists $\tau \in [1/2,1]$ such that

$$\tilde{n}(\tau) < \frac{1}{o} = h^{-1}(c^*/(2+c^*)) \leq \frac{1}{4},$$

where we used the definition of o in Lemma 6.5 and the fact that $c^*/(2+c^*) \le h(1/4) = 25/36$. Then, by Lemma 6.3 with $x^* = 0$ and $x = \tau$, we have

$$\begin{aligned}\frac{1}{2}(\tilde{g}''(\tau))^{1/2} \le \tau(\tilde{g}''(\tau))^{1/2} \le \tilde{n}(\tau) + \frac{3\tilde{n}^2(\tau)}{(1-\tilde{n}(\tau))^3} &= h(\tilde{n}(\tau)) \\ &< h(h^{-1}(c^*/(2+c^*))) = c^*/(2+c^*).\end{aligned}$$

Thus for $t \in [1/2, 1]$

$$\alpha_\tau(t) = |\tau - t| g''(\tau)^{1/2} \le \frac{1}{2} g''(\tau)^{1/2} < c^*/(2+c^*) < 1.$$

Consequently, by Lemma 6.4, for all $t \in [1/2, 1]$

$$\tilde{g}''(t) \le \frac{\tilde{g}''(\tau)}{(1-\alpha_\tau(t))^2} < \frac{4(c^*)^2/(2+c^*)^2}{(1-c^*/(2+c^*))^2} = (c^*)^2,$$

which contradicts the assumption $\tilde{g}''(t_z) \ge (c^*)^2$. Therefore (42) must hold.

7 Examples

In the following examples let $\mathscr{O} = 22$.

Example 7.1 Let $g : \mathbb{R} \to \mathbb{R}$ be an even function defined on the interval $(-a, a)$, where $a > 0$. Define $F(X) = (f \circ \lambda)(X)$, where $f(x) = g(x_1) + \cdots + g(x_n)$. Restricting $F(X)$ to the subspace of matrices

$$\left\{ \begin{pmatrix} 0 & x \\ x^T & 0 \end{pmatrix} \,|\, x \in \mathbb{R}^n \right\},$$

results in the function $2g(\|x\|) + (n-2)g(0)$ defined on $\{x \in \mathbb{R}^n \,|\, \|x\| < a\}$. (The eigenvalues of the matrices in the subspace are $\{\|x\|, 0, ..., 0, -\|x\|\}$.) Thus, if g is a strongly self-concordant function (resp. self-concordant barrier) then $g(\|x\|)$ (resp. $2\mathscr{O}^2 g(\|x\|)$) is a strongly self-concordant function (resp. self-concordant barrier) on the set $\{x \in \mathbb{R}^n \,|\, \|x\| < a\}$, see [5, Proposition 5.1.5].

Considering the space of 2×2 symmetric matrices $\begin{pmatrix} 2x & y \\ y & 2z \end{pmatrix}$ leads to the next example.

Example 7.2 Let $g : \mathbb{R} \to \mathbb{R}$ be a strongly self-concordant function (resp. a self-concordant barrier) on its domain. Define

$$F(x,y,z) = g\big(x + z + \sqrt{x^2 - 2xz + z^2 + y^2}\big) + g\big(x + z - \sqrt{x^2 - 2xz + z^2 + y^2}\big).$$

(a) If $\operatorname{dom} g = (0,1)$ then F (resp. $\mathscr{O}^2 F$) is a strongly self-concordant function (resp. self-concordant barrier) on the set

$$\{(x,y,z) \in \mathbb{R}^3 \,|\, 4xz > y^2,\ 1 + 4xz > y^2 + 2z + 2x\}.$$

(b) If $\operatorname{dom} g = (0, +\infty)$ then F (resp. $\mathcal{O}^2 F$) is a strongly self-concordant function (resp. self-concordant barrier) on the set

$$\{(x,y,z) \in \mathbb{R}^3 \,|\, 4xz > y^2\}.$$

(b.1) Restricting F to the affine space

$$\left\{ \begin{pmatrix} 2x+1 & 1 \\ 1 & 1 \end{pmatrix} |\, x \in \mathbb{R} \right\}$$

shows that

$$x \in \mathbb{R} \mapsto g\big(x+1+\sqrt{x^2+1}\big) + g\big(x+1-\sqrt{x^2+1}\big)$$

is a strongly self-concordant function on $(0, +\infty)$. Repeating the argument we obtain that

$$\begin{aligned} g\Big(x+2+\sqrt{x^2+1}+\sqrt{2x^2+2x+2x\sqrt{x^2+1}+3+2\sqrt{x^2+1}}\Big) \\ +g\Big(x+2+\sqrt{x^2+1}-\sqrt{2x^2+2x+2x\sqrt{x^2+1}+3+2\sqrt{x^2+1}}\Big) \\ +g\Big(x+2-\sqrt{x^2+1}+\sqrt{2x^2+2x-2x\sqrt{x^2+1}+3-2\sqrt{x^2+1}}\Big) \\ +g\Big(x+2-\sqrt{x^2+1}-\sqrt{2x^2+2x-2x\sqrt{x^2+1}+3-2\sqrt{x^2+1}}\Big) \end{aligned}$$

is a strongly self-concordant function on $(0, +\infty)$.

References

[1] Bhatia R.: Matrix Analysis, Springer-Verlag, New York (1997)

[2] Lewis A.S., Overton M.L.: Eigenvalue optimization, *Acta Numer.* **5**, 149–190 (1996)

[3] Nesterov Yu.: Introductory Lectures on Convex Optimization: A Basic Course, Kluwer Academic Publishers, (2004)

[4] Nesterov Yu.: Constructing self-concordant barriers for convex cones, *CORE Discussion Paper*, #2006/30 (2006)

[5] Nesterov Yu., Nemirovskii A.S.: Interior-Point Polynomial Algorithms in Convex Programming, SIAM, Philadelphia, SIAM Studies in Applied Mathematics, vol. 13, (1994)

[6] Renegar J.: A Mathematical View of Interior-Point Methods in Convex Optimization, Society for Industrial and Applied Mathematics, Philadelphia, MPS-SIAM Series on Optimization, (2001)

[7] Sendov H.S.: Generalized Hadamard product and the derivatives of spectral functions, *SIAM J. Matrix Anal. Appl.*, **28**, 667–681 (2006)

[8] Sendov H.S.: The higher-order derivatives of spectral functions, *Linear Algebra Appl.*, vol. 424, 240–281 (2007)