Universal Barrier is n-Self-Concordant

Yin Tat Lee[*] Man–Chung Yue[†]

Abstract

This paper shows that the self-concordance parameter of the universal barrier on any n-dimensional proper convex domain is upper bounded by n. This bound is tight and improves the previous $O(n)$ bound by Nesterov and Nemirovski. The key to our main result is a pair of new, sharp moment inequalities for s-concave distributions, which could be of independent interest.

Keywords: Universal Barrier, Self-Concordance, s-Concave Distributions, Moment Inequalities

1 Introduction

In a seminal work [10], Nesterov and Nemirovski developed a theory of interior point methods for solving general nonlinear convex constrained optimization problems. A central object of their theory is the *self-concordant barrier* for the feasible region. Roughly speaking, a self-concordant barrier on a proper convex domain[1] K is a convex function that satisfies certain differential inequalities and blows up at the boundary ∂K (see Section 2 for the precise definition). Associated with any self-concordant barrier is the *self-concordance parameter* $\nu \geq 0$. The importance of self-concordant barriers lies in the fact that the path-following interior point method developed in [10] approximately solves a convex constrained optimization problem in $O(\sqrt{\nu}\log(1/\epsilon))$ iterations if the feasible region has a ν-self-concordant barrier.

It is then natural to ask whether one can construct a self-concordant barrier for arbitrary proper convex domain and, if yes, what the self-concordance parameter ν is. The first result along this direction was given by Nesterov and Nemirovski [10]: they constructed a self-concordant barrier for general proper convex domain $K \subseteq \mathbb{R}^n$, the so-called *universal barrier*, and proved that it is $O(n)$-self-concordant. They also showed that any self-concordant barrier of n-dimensional simplex or hypercube must have self-concordance parameter at least n, see [10, Proposition 2.3.6]. Hence, their self-concordance bound is *order-optimal*.

Another self-concordant barrier, the *entropic barrier* , was recently studied by Bubeck and Eldan [3]. Exploiting the geometry of log-concave distributions and duality of exponential families, Bubeck and Eldan [3] proved that the entropic barrier satisfies the self-concordance parameter guarantee $\nu \leq n + O(\sqrt{n \log n})$ for $n \geq 80$, thus improving the result of Nesterov and Nemirovski [10].

When the proper convex domain K is a cone, the situation is clearer. Indeed, the *canonical barrier*, introduced by Hildebrand [7] and independently by Fox [4], is an n-self-concordant barrier

[*]University of Washington. E–mail: `yintat@uw.edu`

[†]Imperial College London. E–mail: `m.yue@imperial.ac.uk`

[1]A *convex domain* is a convex set with non-empty interior. A convex set is said to be *proper* if it does not contain any 1-dimensional affine subspace.

of proper convex cones with non-empty interior. Furthermore, using a result of Güler [6], Bubeck and Eldan [3] showed that both the universal barrier and the entropic barrier are also n-self-concordant on proper convex cones. These results confirmed a conjecture[2] made by Güler which asserted that, for any proper convex cone in $\mathbb{R}^n$, there always exists a self-concordant barrier whose self-concordant parameter is at most n.

This paper completes the picture by settling the same question in the more general case of proper convex domains. We show that the universal barrier is n-self-concordant on any proper convex domain $K \subseteq \mathbb{R}^n$ for $n \geq 1$. This does not only improve the results of [10] and [3] but is also tight in view of the above-mentioned lower bound on the self-concordance parameter. The key to this result is a pair of new, sharp moment inequalities for s-concave distributions (see Section 2 for the definition of s-concavity), which could be of independent interest. One of these inequalities is a generalization of [3, Lemma 2].

We should emphasize that all these bounds on the self-concordant parameters of different barriers do not immediately yield polynomial-time complexity result for convex programming problems. The iteration complexity $O(\sqrt{\nu}\log 1/\epsilon)$ counts only the number of iterations of the path-following algorithm, whereas the overall complexity depends also on the costs of computing the gradient and the Hessian of the barrier for the feasible region. The problem of constructing self-concordant barriers with (nearly) optimal self-concordance parameter and efficiently computable gradient and Hessian remains largely open. A recent breakthrough was obtained in the context of polytopes by Lee and Sidford [8]. However, our result does find applications in some online learning problems where the quality of solutions produced by certain algorithms depend on the self-concordance parameter [1, 9].

The rest of the paper is organized as follows. Section 2 collects some necessary background and preparatory results. The optimal self-concordance bound of the universal barrier, which is the main result of this paper, will be proved in Section 3. Section 4 provides the proofs of the pair of moment inequalities used for proving the main result.

1.1 Notations

We adopt the following notations throughout the paper. Given a set S, we denote by $\mathrm{cl}(S)$, $\mathrm{int}(S)$ and $\partial S = \mathrm{cl}(S)\backslash\mathrm{int}(S)$ the closure, interior and boundary of S, respectively. The indicator function of S is denoted by $\mathbb{1}_S$, *i.e.*, $\mathbb{1}_S(t) = 1$ if $t \in S$ and $\mathbb{1}_S(t) = 0$ otherwise. We denote by $\mathrm{Vol}_k(S)$ the k-dimensional Lebesgue measure of S. For any function ψ, the i-th directional derivative of ψ at x along the direction h will be denoted by $D^i\psi(x)[h,h]$. For any distribution on $\mathbb{R}$ with density p, we denote by $\mathrm{Supp}(p)$ the support of the distribution, *i.e.*, $\mathrm{Supp}(p) = \mathrm{cl}\left(\{t \in \mathbb{R} : p(t) > 0\}\right)$. The Dirac delta distribution at t will be denoted by δ_t.

2 Preliminaries

2.1 The Universal Barrier

A *convex domain* is a convex set with non-empty interior. A convex set is said to be *proper* if it does not contain any 1-dimensional affine subspace. Throughout the paper, if not specified, K will always denote a proper convex domain in $\mathbb{R}^n$. As usual, a *convex body* refers to a compact convex domain. The following definitions are standard [10].

[2] See the discussions in [3] and [4].

Definition 1. *A function $\phi : \operatorname{int}(K) \to \mathbb{R}$ is said to be a* barrier *on K if*

$$\phi(x) \to +\infty \quad as \quad x \to \partial K.$$

Definition 2. *A three times continuously differentiable convex function ϕ is said to be* self-concordant *on K if for any $x \in \operatorname{int}(K)$ and $h \in \mathbb{R}^n$,*

$$D^3\phi(x)[h,h,h,] \le 2\left(D^2\phi(x)[h,h]\right)^{\frac{3}{2}}. \tag{1}$$

If, in addition to (1), ϕ satisfies that for any $x \in \operatorname{int}(K)$ and $h \in \mathbb{R}^n$,

$$D\phi(x)[h] \le \left(\nu \cdot D^2\phi(x)[h,h]\right)^{\frac{1}{2}}, \tag{2}$$

*then ϕ is said to be ν-*self-concordant*.*

The main contribution in this paper concerns the so-called *universal barrier* introduced by Nesterov and Nemirovski [10].

Definition 3. *The* universal barrier *of K is defined as the function $\phi : \operatorname{int}(K) \to \mathbb{R}$ given by*

$$\phi(x) = \log \operatorname{Vol}_n\left(K^\circ(x)\right),$$

where $K^\circ(x) = \left\{y \in \mathbb{R}^n : y^T(z-x) \le 1, \forall z \in K\right\}$ is the polar set of K with respect to x.

It is well-known that the universal barrier is $O(n)$-self-concordant [10, Theorem 2.5.1]. As we will see in the Section 3, the bound $O(n)$ can be improved to exactly n.

2.2 Probabilistic Tools

Since all distributions considered in this paper are absolutely continuous with respect to the Lebesgue measure, we identify a distribution with its density. For any distribution p on $\mathbb{R}$, we denote its mean by $\mu_1(p)$ and the second and third moments about the mean by $\mu_2^2(p)$ and $\mu_3^3(p)$, respectively, *i.e.*,

$$\begin{aligned}
\mu_1(p) &= \int_{-\infty}^{\infty} tp(t)dt, \\
\mu_2^2(p) &= \int_{-\infty}^{\infty} (t-\mu_1(p))^2 p(t)dt \quad \text{and} \\
\mu_3^3(p) &= \int_{-\infty}^{\infty} (t-\mu_1(p))^3 p(t)dt.
\end{aligned}$$

The following type of distributions on $\mathbb{R}$ is particularly important in this paper.

Definition 4. *Let $L \subseteq \mathbb{R}^n$ be any convex body and $h \in \mathbb{R}^n$. The* marginal distribution of the convex body L along the direction h*, denoted by $p(L,h;\cdot)$, is the distribution on $\mathbb{R}$ given by, for any $t \in \mathbb{R}$,*

$$p(L,h;t) = \frac{\operatorname{Vol}_{n-1}\left(\left\{y \in L : y^T h = t\right\}\right)}{\operatorname{Vol}_n(L)}.$$

Note that the polar set $K^\circ(x)$ with respect to any $x \in \operatorname{int}(K)$ is a convex body. Therefore, we can talk about its marginal distributions. Interestingly, the directional derivatives of the universal barrier on K at x can be expressed in terms of moments of the marginal distribution of the polar set $K^\circ(x)$. The following formulas can be found in [10, p. 52].

Lemma 1. *Let $x \in \operatorname{int}(K)$ and $h \in \mathbb{R}^n$ be given. Let $p = p\left(K^\circ(x), h; \cdot\right)$. Then we have that*

$$\begin{aligned} D\phi(x)[h] &= -(n+1)\mu_1(p), \\ D^2\phi(x)[h,h] &= (n+1)(n+2)\mu_2^2(p) + (n+1)\mu_1^2(p) \text{ and} \\ D^3\phi(x)[h,h,h] &= -(n+1)(n+2)(n+3)\mu_3^3(p) - 6(n+1)(n+2)\mu_2^2(p)\mu_1(p) - 2(n+1)\mu_1^3(p). \end{aligned}$$

Next, we recall the definition of s-concave distributions [2].

Definition 5. *A distribution p on $\mathbb{R}$ is said to be* s-concave *if for any $\lambda \in [0,1]$ and $t_1, t_2 \in \mathbb{R}$, it holds that*

$$p(\lambda t_1 + (1-\lambda)t_2) \geq \left(\lambda \left(p(t_1)\right)^s + (1-\lambda)\left(p(t_2)\right)^s\right)^{\frac{1}{s}}. \tag{3}$$

For the case $s = 0$, $s = -\infty$ and $s = +\infty$, the right-hand side of (3) becomes $\left(p(t_1)\right)^\lambda \left(p(t_2)\right)^{1-\lambda}$, $\min\{p(t_1), p(t_2)\}$ and $\max\{p(t_1), p(t_2)\}$.

Note that 0-concave distributions are nothing but log-concave distributions.

We pause to provide some intuitions for the $O(n)$ bound on the self-concordance parameter of Nesterov and Nemirovski [10] and explain why improvement is possible. First, it is a fact in convex geometry that the width of a convex body $L \subseteq \mathbb{R}^n$ along any direction h is of the order $O\left(n \cdot \mu_2^2\left(p\left(L, h; \cdot\right)\right)\right)$. Lemma 1 then implies that ϕ satisfies inequality (2) with $\nu = O(n)$. Second, the Prékopa-Leindler inequality implies that $p(L, h; \cdot)$ is a log-concave distribution. Combining this with another convex-geometric fact that the third moment of any log-concave distribution is bounded by its second moment, we can deduce inequality (1) from Lemma 1. Our improvement is made possible by the observation that any marginal distribution $p(L, h; \cdot)$ is actually $\frac{1}{n-1}$-concave, a stronger property than the log-concavity. This observation follows immediately from the *Brunn's concavity principle.*

Theorem 1 (Brunn's Concavity Principle, [2, Theorem 1.2.2])**.** *Let L be a convex body in $\mathbb{R}^n$ and F be a k-dimensional subspace. Then, the function $r : F^\perp \to \mathbb{R}$ defined by*

$$r(x) = \operatorname{vol}_k(L \cap (F + x))$$

is $\frac{1}{k}$-concave on its support.

The crux to the proof of our main result is the following improved moments inequalities whose proof is postponed to Section 4.

Proposition 1. *Let $k \geq 1$ be an integer and p be a $\frac{1}{k-1}$-concave distribution on $\mathbb{R}$. Then we have that*

$$\mu_3^3(p) \leq 2\sqrt{\frac{k+2}{k}}\frac{k-1}{k+3}\mu_2^3(p). \tag{4}$$

Suppose furthermore that $0 \in \operatorname{Supp}(p)$. Then we have that

$$\mu_1^2(p) \leq k(k+2)\mu_2^2(p) \tag{5}$$

Remark 1. *As we will see in the proof of Proposition 1, inequalities (4) and (5) are both sharp. By assuming p to be centered (i.e., $\mu_1(p) = 0$) and letting $k \to +\infty$, inequality (4) recovers [3, Lemma 2]. Also, the condition that $0 \in \operatorname{Supp}(p)$ for inequality (5) is necessary. This can be seen by substituting, for example, $p = \delta_t$ for any $t \neq 0$ into (5).*

3 Self-Concordance of the Universal Barrier

Now we have enough tools at our disposal to prove the main result of this paper.

Theorem 2. *For any $n \geq 1$ and proper convex domain $K \subseteq \mathbb{R}^n$, the universal barrier ϕ is an n-self-concordant barrier for K.*

Proof. That ϕ is a barrier on K follows from [10, Theorem 2.5.1]. It remains to show that ϕ satisfies the differential inequalities (1) and (2) with $\nu = n$.

Let any $x \in \operatorname{int}(K)$ and $h \in \mathbb{R}^n$ be given. Then $K^\circ(x)$ is a convex body containing the origin. Also, let p be the marginal distribution of $K^\circ(x)$ along h, *i.e.*, $p = p\left(K^\circ(x), h; \cdot\right)$. Since $K^\circ(x)$ contains the original, we have that $\operatorname{Supp}(p)$ is a non-degenerate closed interval and $0 \in \operatorname{Supp}(p)$. Furthermore, Theorem 1 shows that p is a $\frac{1}{n-1}$-concave distribution on $\mathbb{R}$. Hence, by Proposition 1, we have that

$$\mu_3^3 \leq 2\sqrt{\frac{n+2}{n}}\frac{n-1}{n+3}\mu_2^3 \tag{6}$$

and that

$$\mu_1^2 \leq n(n+2)\mu_2^2. \tag{7}$$

Here we write μ_i instead of $\mu_i(p)$ for $i = 1, 2, 3$. Using Lemma 1 and inequality (7), we have

$$\begin{aligned}
\frac{D\phi(x)[h]}{\sqrt{D^2\phi(x)[h,h]}} &\leq \frac{|(n+1)\mu_1|}{\sqrt{(n+1)(n+2)\mu_2^2 + (n+1)\mu_1^2}} \\
&\leq \frac{(n+1)\,|\mu_1|}{\sqrt{(n+1)(n+2)\frac{\mu_1^2}{n(n+2)} + (n+1)\mu_1^2}} \\
&= \sqrt{n}.
\end{aligned}$$

This shows that ϕ satisfies inequality (2) with $\nu = n$.

Finally, we prove that ϕ satisfies inequality (1). Towards that end, we first observe that $\mu_2 > 0$, for otherwise it would contradict to the non-degeneracy of $\operatorname{Supp}(p)$. Therefore,

$$D^2\phi(x)[h,h] = (n+1)\left((n+2)\mu_2^2 + \mu_1^2\right) > 0.$$

Using Lemma 1 and inequality (6), we have

$$\begin{aligned}
\frac{D^3\phi(x)[h,h,h]}{\left(D^2\phi(x)[h,h]\right)^{\frac{3}{2}}} &= \frac{-(n+2)(n+3)\mu_3^3 - 6(n+2)\mu_2^2\mu_1 - 2\mu_1^3}{\sqrt{n+1}\left((n+2)\mu_2^2 + \mu_1^2\right)^{\frac{3}{2}}} \\
&\leq \frac{(n+2)(n+3)\left(2\sqrt{\frac{n+2}{n}}\frac{n-1}{n+3}\mu_2^3\right) - 6(n+2)\mu_2^2\mu_1 - 2\mu_1^3}{\sqrt{n+1}\left((n+2)\mu_2^2 + \mu_1^2\right)^{\frac{3}{2}}} \\
&= \frac{1}{\sqrt{n+1}}\frac{\frac{2(n-1)}{\sqrt{n}} - 6\tau - 2\tau^3}{(1+\tau^2)^{\frac{3}{2}}},
\end{aligned} \tag{8}$$

where we set $\tau = \frac{\mu_1}{\mu_2\sqrt{n+2}}$. Let $c_n = \frac{(n-1)}{2\sqrt{n}}$ and $\ell : \mathbb{R} \to \mathbb{R}$ be the function defined by, for any $t \in \mathbb{R}$,

$$\ell(t) = \frac{4c_n - 6t - 2t^3}{(1+t^2)^{\frac{3}{2}}}.$$

Then,

$$\ell'(t) = \frac{6(t^2 - 2c_n t - 1)}{(1+t^2)^{\frac{5}{2}}}.$$

The stationary points are $t = c_n \pm \sqrt{c_n^2 + 1} = -\frac{1}{\sqrt{n}}$ or $\sqrt{n}$. Hence,

$$\begin{aligned}
\ell(t) &\leq \max\left\{\lim_{t\to-\infty} \ell(t), \ell\left(-\tfrac{1}{\sqrt{n}}\right), \ell\left(\sqrt{n}\right), \lim_{t\to\infty} \ell(t)\right\} \\
&= \max\left\{2, 2\sqrt{n+1}, -2\sqrt{\frac{n+1}{n}}, -2\right\} \\
&= 2\sqrt{n+1}. \qquad (9)
\end{aligned}$$

Combining inequalities (8) and (9), we get

$$\frac{D^3\phi(x)[h,h,h]}{(D^2\phi(x)[h,h])^{\frac{3}{2}}} \leq \frac{1}{\sqrt{n+1}} \cdot 2\sqrt{n+1} = 2.$$

This completes the proof. □

4 Proof of Proposition 1

The goal of this section is to prove Proposition 1. We first handle the case $k = 1$, *i.e.*, p is ∞-concave. We claim that $S := \{t \in \mathbb{R} : p(t) > 0\}$ is convex. We argue this by contradiction. Suppose S is non-convex. Then there exist $t_1, t_2 \in S$ and $\lambda \in (0,1)$ such that $\lambda t_1 + (1-\lambda)t_2 \notin S$, which implies the contradiction that $0 = p(\lambda t_1 + (1-\lambda)t_2) \geq \max\{p(t_1), p(t_2)\} > 0$. Next, we claim that p is constant on S. Again we prove this by contradiction. Suppose there exist $t_1, t_2 \in S$ such that $p(t_1) > p(t_2)$. Then,

$$p(t_2) = \lim_{\lambda\to 0} p(\lambda t_1 + (1-\lambda)t_2) \geq \lim_{\lambda\to 0} p(t_1) = p(t_1) > p(t_2),$$

which is a contradiction. So p is either a uniform distribution on an interval or a Dirac delta distribution. Inequalities (4) and (5) are evident in both possibilities.

It remains to prove Proposition 1 for $k \geq 2$. We will first prove inequality (5) in Section 4.1 and then inequality (4) in Section 4.2. Before doing that, let us provide a brief overview of the proofs. Each of the proofs start with a sequence of reductions and approximations. This is to modify the distribution class and turn the inequality into an equivalent variational formulation so that we can apply the following *localization lemma*[3]:

Theorem 3 (Localization Lemma [5, Theorem 2]). *Let $m \geq 1$, $H \subseteq \mathbb{R}^m$ be a compact convex set, $s \in [-1,1]$ and $f : H \to \mathbb{R}$ an upper semi-continuous function. Let $\mathcal{M}(H)$ be the set of measures with support contained in H and $\mathcal{P}_f \subset \mathcal{M}(H)$ be the subset of measures φ that has a s-concave density and satisfies $\int f d\varphi \geq 0$. If $\Pi : \mathcal{M}(H) \to \mathbb{R}$ is a convex upper semi-continuous function, then $\sup_{\varphi\in\mathcal{P}_f} \Pi(\varphi)$ is achieved at either a Dirac delta distribution δ_u with $f(u) \geq 0$ or a measure with density q such that*

(i) Supp(q) an interval contained in H,

[3]Note that our notation s is the γ in the paper [5].

(ii) q^s *(or* $\log q$ *if* $s = 0$*) is affine on Supp*(q)*,*

(iii) $\int f(u)q(u)du = 0$*, and*

(iv) $\int_a^t f(u)q(u)du > 0$ *for all* $t \in (a, b)$ *or* $\int_t^b f(u)q(u)du > 0$ *for all* $t \in (a, b)$*.*

The localization lemma will allow us to restrict our attention to $\frac{1}{k-1}$-affine distributions on $\mathbb{R}$, *i.e.*, distributions of the form

$$\frac{(\alpha t + \beta)^{k-1} \cdot \mathbb{1}_{[a,b]}(t)}{\int_a^b (\alpha u + \beta)^{k-1} du}, \tag{10}$$

where $a \leq b$ and $\alpha t + \beta \geq 0$ for any $t \in [a, b]$. Substituting an arbitrary $\frac{1}{k-1}$-affine distribution into the desired inequality, the task is further reduced to proving an algebraic inequality. Finally, the proof is completed by proving the algebraic inequality using simple calculus.

4.1 Proof of Inequality (5)

Let $\mathcal{P}$ be the set of $\frac{1}{k-1}$-concave distributions and $\bar{\mathcal{P}} \subseteq \mathcal{P}$ be the subset of distributions $p \in \mathcal{P}$ with $0 \in \text{Supp}(p)$. Also, the set of $\frac{1}{k-1}$-affine distributions on $\mathbb{R}$ (*i.e.*, (10)) will be denoted by $\mathcal{Q}$. We note that inequality (5) is equivalent to

$$\left(\int_{-\infty}^{\infty} tp(t)dt\right)^2 \leq \frac{k(k+2)}{(k+1)^2} \int_{-\infty}^{\infty} t^2 p(t)dt. \tag{11}$$

So we will prove inequality (11) instead.

4.1.1 To Distributions with Bounded Support

We first show that it suffices to prove inequality (11) for $p \in \bar{\mathcal{P}}$ with bounded support. This can be done by limiting arguments: Let any $p \in \bar{\mathcal{P}}$ and $\epsilon > 0$ be given. By continuity, there exists a real number $M > 0$ such that

$$\left| \left(\int_{-\infty}^{\infty} tp(t)dt\right)^2 - \left(\frac{\int_{-M}^{M} tp(t)dt}{\int_{-M}^{M} p(u)du}\right)^2 \right| \leq \frac{\epsilon}{2}, \tag{12}$$

and

$$\frac{k(k+2)}{(k+1)^2} \left| \int_{-\infty}^{\infty} t^2 p(t)dt - \frac{\int_{-M}^{M} t^2 p(t)dt}{\int_{-M}^{M} p(u)du} \right| \leq \frac{\epsilon}{2}. \tag{13}$$

Note that the distribution

$$\frac{p(t)\mathbb{1}_{[-M,M]}}{\int_{-M}^{M} p(u)du} \in \bar{\mathcal{P}}$$

has a bounded support. Therefore, if inequality (11) holds for any distribution in $\bar{\mathcal{P}}$ with bounded support, then from inequalities (12) and (13), we have

$$\left(\int_{-\infty}^{\infty} tp(t)dt\right)^2 \leq \frac{k(k+2)}{(k+1)^2} \int_{-\infty}^{\infty} t^2 p(t)dt + \epsilon.$$

Since the above inequality holds for any small $\epsilon > 0$, inequality (11) follows by taking limiting $\epsilon \searrow 0$.

4.1.2 To Distributions with Non-negative Support

Inequality (11) is equivalent to

$$\Phi(p) := \frac{\left(\int_{-\infty}^{\infty} tp(t)dt\right)^2}{\int_{-\infty}^{\infty} t^2 p(t)dt} \leq \frac{k(k+2)}{(k+1)^2}.$$

Since Φ is unchanged if we flip the distribution p horizontally about $t = 0$, we can assume without loss that the mean $\mu_1(p)$ is non-negative. By the above reduction, we could also assume that $\mathrm{Supp}(p) = [M_1, M_2]$ for some $M_1, M_2 \in \mathbb{R}$ with $M_1 < M_2$. Let $p_u(t) = p(t-u)$ be the distribution obtained by shifting p to the right by u units. Then, for any $u \geq 0$,

$$\begin{aligned}
&\frac{d\Phi(p_u)}{du} = \frac{d}{du}\frac{\left(\int_{M_1+u}^{M_2+u} tp_u(t)dt\right)^2}{\int_{M_1+u}^{M_2+u} t^2 p_u(t)dt} \\
&= \frac{\left(\int_{M_1+u}^{M_2+u} t^2 p_u(t)dt\right)\frac{d}{du}\left(\int_{M_1}^{M_2}(t+u)p(t)dt\right)^2}{\left(\int_{M_1+u}^{M_2+u} t^2 p_u(t)dt\right)^2} - \frac{\left(\int_{M_1+u}^{M_2+u} tp_u(t)dt\right)^2 \frac{d}{du}\left(\int_{M_1}^{M_2}(t+u)^2 p(t)dt\right)}{\left(\int_{M_1+u}^{M_2+u} t^2 p_u(t)dt\right)^2} \\
&= \frac{2\left(\int_{M_1+u}^{M_2+u} t^2 p_u(t)dt\right)\left(\int_{M_1}^{M_2}(t+u)p(t)dt\right)}{\left(\int_{M_1+u}^{M_2+u} t^2 p_u(t)dt\right)^2} - \frac{2\left(\int_{M_1+u}^{M_2+u} tp_u(t)dt\right)^2\left(\int_{M_1}^{M_2}(t+u)p(t)dt\right)}{\left(\int_{M_1+u}^{M_2+u} t^2 p_u(t)dt\right)^2} \\
&= \frac{2\mu_1(p_u)}{\left(\int_{M_1+u}^{M_2+u} t^2 p_u(t)dt\right)^2}\left(\int_{M_1+u}^{M_2+u} t^2 p_u(t)dt - \left(\int_{M_1+u}^{M_2+u} tp_u(t)dt\right)^2\right) \geq 0,
\end{aligned}$$

where the last inequality follows from that $\mu_1(p_u) = \mu_1(p) + u \geq 0$. This shows that shifting p rightwards can only (monotonically) increase the value of Φ. Therefore, we can assume that $\mathrm{Supp}(p) = [0, M_3]$ for some $M_3 > 0$.

4.1.3 To Distributions with $p(0) > 0$

Here we show that it suffices to focus on distributions $p \in \bar{\mathcal{P}}$ with $p(0) > 0$. From the above reductions, we can focus on $p \in \bar{\mathcal{P}}$ such that $\mathrm{Supp}(p) = [0, M_3]$ for some $M_3 > 0$. Let $\epsilon \in (0, M_3)$. By definition, we have $p(\epsilon) > 0$. Consider the distribution $p_{\text{-}\epsilon}$ obtained by shifting p to the left by ϵ units. We can bound the changes in the integrals in (11) as follows:

$$\left|\int_0^{M_3} tp(t)dt - \int_{-\epsilon}^{M_3-\epsilon} tp_{\text{-}\epsilon}(t)dt\right| = \left|\int_0^{M_3} tp(t)dt - \int_0^{M_3}(u-\epsilon)p(u)du\right| = \epsilon, \tag{14}$$

and

$$\begin{aligned}
&\left|\int_0^{M_3} t^2 p(t)dt - \int_{-\epsilon}^{M_3-\epsilon} t^2 p_{\text{-}\epsilon}(t)dt\right| \\
&= \left|\int_0^{M_3} t^2 p(t)dt - \int_0^{M_3}(u-\epsilon)^2 p(u)du\right| = \left|2\epsilon\int_0^{M_3} up(u)du - \epsilon^2\right| = O(\epsilon).
\end{aligned} \tag{15}$$

Although $p_{\text{-}\epsilon}(0) = p(\epsilon) > 0$, the support $\mathrm{Supp}(p_{\text{-}\epsilon})$ of $p_{\text{-}\epsilon}$ is not non-negative. To remedy this, we consider the truncated distribution

$$p^{\epsilon}(t) = \frac{p(t+\epsilon)\mathbb{1}_{[0,M_3-\epsilon]}}{\int_0^{M_3-\epsilon} p(u)du}.$$

The distribution $p^\epsilon \in \bar{\mathcal{P}}$ retains all the desirable properties: $p^\epsilon(0) > 0$ and has a non-negative bounded support. Furthermore, combining inequalities (14) and (15) with arguments similar to those in Section 4.1.1, we can easily show that $\Phi(p^\epsilon) \to \Phi(p)$ as $\epsilon \searrow 0$. Hence, we could assume without loss of generality that $p(0) > 0$.

4.1.4 To Distributions Supported in $[0, 1]$

Let $p \in \bar{\mathcal{P}}$. Due to above reductions, we may assume that $\mathrm{Supp}(p) = [0, M_3]$ for some $M_3 > 0$ and that $p(0) > 0$. Consider the transformation $\tilde{p}(x) = M_3 \cdot p(M_3 x)$. One can easily check that $\tilde{p}$ is a probability distribution in $\bar{\mathcal{P}}$ with $\tilde{p}(0) > 0$ and $\mathrm{Supp}(\tilde{p}) = [0, 1]$. Furthermore, we have that

$$\Phi\left(\tilde{p}\right) = \frac{\left(\int_{-\infty}^{\infty} t\tilde{p}(t)dt\right)^2}{\int_{-\infty}^{\infty} t^2\tilde{p}(t)dt} = \frac{M_3\left(\int_0^1 tp(M_3 t)dt\right)^2}{\int_0^1 t^2 p(M_3 t)dt} = \frac{\left(\int_0^{M_3} u\cdot p(u)du\right)^2}{\int_0^{M_3} u^2\cdot p(u)du} = \Phi(p).$$

Therefore, it suffices henceforth to focus on the subset of distributions $p \in \bar{\mathcal{P}}$ with $p(0) > 0$ and $\mathrm{Supp}(p) = [0, 1]$.

4.1.5 To $\frac{1}{k-1}$-Affine Distributions

Let $\Psi : \bar{\mathcal{P}} \to \mathbb{R}$ be the function defined as

$$\Psi(q) = (k+1)^2\left(\int_0^1 tq(t)dt\right)^2 - k(k+2)\int_0^1 t^2 q(t)dt.$$

For any $\epsilon > 0$, consider the problem

$$\begin{array}{ll} \Psi^\epsilon := \sup\limits_q & \Psi(q) \\ \text{subjectto} & \frac{1}{\epsilon}\int_0^\epsilon q(t)dt \geq \epsilon, \\ & q \in \bar{\mathcal{P}}', \end{array} \tag{P_ϵ}$$

where $\bar{\mathcal{P}}'$ is the subset of distributions $q \in \mathcal{P}$ with $\mathrm{Supp}(q) \subseteq [0, 1]$. Note that Ψ^ϵ is finite. By Theorem 3, the supremum Ψ^ϵ is achieved by either

1. $\delta_{\bar{t}}$ for some $\bar{t} \in [0, \epsilon]$; or
2. some $q^\epsilon \in \mathcal{Q}$ with $\mathrm{Supp}(q^\epsilon) \subseteq [0, 1]$ and
$$\frac{1}{\epsilon}\int_0^\epsilon q^\epsilon(t)dt = \epsilon. \tag{16}$$

To prove inequality (11), it suffices to prove that as $\epsilon \searrow 0$,

$$\Psi^\epsilon \leq o_\epsilon(1). \tag{17}$$

Indeed, since $p(0) > 0$, there exists an $\bar{\epsilon} > 0$ such that for any $\epsilon \in (0, \bar{\epsilon})$, we have $\frac{1}{\epsilon}\int_0^\epsilon p(t)dt \geq \epsilon$. Hence, p must be a feasible solution to problem (P_ϵ) and hence $\Psi(p) \leq \Psi^\epsilon$. Taking limit $\epsilon \searrow 0$ yields $\Psi(p) \leq 0$, which is equivalent to inequality (11) by the above reductions.

We first consider Case 1:

$$\begin{aligned} \Psi(\delta_{\bar{t}}) &= (k+1)^2\left(\int_0^1 t\delta_{\bar{t}}(t)dt\right)^2 - k(k+2)\int_0^1 t^2\delta_{\bar{t}}(t)dt \\ &= (k+1)^2\bar{t}^2 - k(k+2)\bar{t}^2 = \bar{t}^2 \leq \epsilon^2, \end{aligned}$$

which implies inequality (17). So it remains to prove inequality (17) for Case 2: the $\frac{1}{k-1}$-affine distribution $q^\epsilon \in \mathcal{Q}$.

4.1.6 To an Algebraic Inequality

Since $q^\epsilon \in \mathcal{Q}$ and $\mathrm{Supp}(q^\epsilon) \subseteq [0,1]$, there exist constants $a, b \in [0,1]$ and $\alpha, \beta \in \mathbb{R}$ such that $a < b$, $\alpha t + \beta \geq 0$ on $[a,b]$ and

$$q^\epsilon(t) = \frac{(\alpha t + \beta)^{k-1} \cdot \mathbb{1}_{[a,b]}(t)}{\int_a^b (\alpha u + \beta)^{k-1} du}.$$

We claim that without loss of generality, we can assume that $a = 0$ and $\beta \geq 0$. To prove the claim, from the equality (16), we get

$$\int_{[0,\epsilon]\cap[a,b]} (\alpha t + \beta)^{k-1} dt = \epsilon^2 \int_a^b (\alpha u + \beta)^{k-1} du,$$

which shows that $\epsilon \in (a,b)$. Consider the shifted distribution $q^\epsilon_{\text{-}a}$. Following the same arguments for deriving (14) and (15), we can show that

$$|\Psi(q^\epsilon) - \Psi(q^\epsilon_{\text{-}a})| = O(\epsilon).$$

Therefore, we can assume without loss that $a = 0$ when proving (17) for q^ϵ, which in turn implies $\beta \geq 0$.

Next, we claim that without loss of generality, we can also assume that $\alpha > 0$. Suppose $\alpha = 0$. In such a case, q^ϵ is a uniform distribution, *i.e.*, $q^\epsilon(t) = \frac{1}{b}$ for all $t \in [0,b]$. We therefore have

$$\begin{aligned}
\Psi(q^\epsilon) &= \frac{(k+1)^2}{b^2}\left(\int_0^b t dt\right)^2 - \frac{k(k+2)}{b}\int_0^b t^2 dt \\
&= \frac{(k+1)^2 b^2}{4} - \frac{k(k+2)b^2}{3} \\
&= \frac{b^2}{12}\left(-k^2 - 2k + 3\right) \leq 0.
\end{aligned}$$

For $\alpha < 0$, we consider the distribution $\tilde{q}^\epsilon(t) = q^\epsilon(b - t)$, the distribution obtained by flipping q^ϵ horizontally about $t = \frac{b}{2}$. In other words,

$$\tilde{q}^\epsilon(t) = \frac{(-\alpha t + \alpha b + \beta)^{k-1} \cdot \mathbb{1}_{[0,b]}(t)}{\int_0^b (\alpha u + \beta)^{k-1} du},$$

which is again supported on $[0,b]$ and $\frac{1}{k-1}$-affine. In addition, $\alpha b + \beta \geq 0$ and $-\alpha > 0$. Therefore, the claim would follow if we can prove that

$$\Psi(\tilde{q}^\epsilon) \geq \Psi(q^\epsilon),$$

which can be easily shown to be equivalent to

$$\frac{\int_0^b t(\alpha t + \beta)^{k-1} dt}{\int_0^b (\alpha t + \beta)^{k-1} dt} \leq \frac{b}{2}. \tag{18}$$

Inequality (18) follows immediately from the next lemma.

Lemma 2. *Let q be a non-increasing distribution supported on $[0,1]$. Then the mean of q is at most $\frac{1}{2}$.*

Proof. For any $t \in [0, \frac{1}{2}]$, we $q(t) \geq q(1-t)$ and hence

$$tq(t) + (1-t)q(1-t) \leq (1-t)q(t) + tq(1-t).$$

Integrating both sides, we get

$$\begin{aligned}
\int_0^{\frac{1}{2}} tq(t) + (1-t)q(1-t)dt \leq \int_0^{\frac{1}{2}} (1-t)q(t) + tq(1-t)dt,\\
\int_0^1 tq(t)dt \leq \int_0^1 (1-t)q(t)dt,\\
\int_0^1 tq(t)dt \leq \frac{1}{2}.
\end{aligned}$$

□

Therefore, we can safely ignore the case of $\alpha \leq 0$.

It is obvious that inequality (17) is implied by

$$(k+1)^2 \left(\int_0^b tq^\epsilon(t)dt \right)^2 \leq k(k+2) \int_0^b t^2 q^\epsilon(t)dt. \tag{19}$$

Let $\kappa = \frac{\beta}{b\alpha} \geq 0$. We compute the integrals

$$\begin{aligned}
\int_0^b tq^\epsilon(t)dt &= \frac{\int_0^b t(\alpha t+\beta)^{k-1}dt}{\int_0^b (\alpha t+\beta)^{k-1}dt} = \frac{b\int_0^1 t(t+\frac{\beta}{b\alpha})^{k-1}dt}{\int_0^1 (t+\frac{\beta}{b\alpha})^{k-1}dt}\\
&= b\left(\frac{\int_0^1 (t+\kappa)^k dt}{\int_0^1 (t+\kappa)^{k-1}dt} - \kappa \right) = b\left(\frac{(1+\kappa)^{k+1} - \kappa^{k+1}}{(1+\kappa)^k - \kappa^k} \frac{k}{k+1} - \kappa \right),
\end{aligned} \tag{20}$$

and

$$\begin{aligned}
\int_0^b t^2 q^\epsilon(t)dt &= \frac{\int_0^b t^2(\alpha t+\beta)^{k-1}dt}{\int_0^b (\alpha t+\beta)^{k-1}dt} = \frac{b^2\int_0^1 t^2(t+\frac{\beta}{b\alpha})^{k-1}dt}{\int_0^1 (t+\frac{\beta}{b\alpha})^{k-1}dt}\\
&= b^2\left(\frac{\int_0^1 (t+\kappa)^{k+1}dt}{\int_0^1 (t+\kappa)^{k-1}dt} - \frac{2\kappa\int_0^1 (t+\kappa)^k dt}{\int_0^1 (t+\kappa)^{k-1}dt} + \kappa^2 \right)\\
&= b^2\left(\frac{(1+\kappa)^{k+2} - \kappa^{k+2}}{(1+\kappa)^k - \kappa^k} \frac{k}{k+2} - 2\kappa \frac{(1+\kappa)^{k+1} - \kappa^{k+1}}{(1+\kappa)^k - \kappa^k} \frac{k}{k+1} + \kappa^2 \right).
\end{aligned} \tag{21}$$

Substituting (20) and (21) into (19) yields

$$\begin{aligned}
&(k+1)^2 \left(\frac{(1+\kappa)^{k+1} - \kappa^{k+1}}{(1+\kappa)^k - \kappa^k} \frac{k}{k+1} - \kappa \right)^2\\
\leq\, &k(k+2) \left(\frac{(1+\kappa)^{k+2} - \kappa^{k+2}}{(1+\kappa)^k - \kappa^k} \frac{k}{k+2} - 2\kappa \frac{(1+\kappa)^{k+1} - \kappa^{k+1}}{(1+\kappa)^k - \kappa^k} \frac{k}{k+1} + \kappa^2 \right).
\end{aligned}$$

Setting $\gamma = \frac{1+\kappa}{\kappa} \geq 1$, it suffices to prove the following algebraic inequality

$$\begin{aligned}
&(k+1)^2 \left(\frac{\gamma^{k+1} - 1}{\gamma^k - 1} \frac{k}{k+1} - 1 \right)^2\\
\leq\, &k(k+2) \left(\frac{\gamma^{k+2} - 1}{\gamma^k - 1} \frac{k}{k+2} - 2 \cdot \frac{\gamma^{k+1} - 1}{\gamma^k - 1} \frac{k}{k+1} + 1 \right).
\end{aligned} \tag{22}$$

4.1.7 Proving the Algebraic Inequality (22)

We now conclude the proof of inequality (11) by proving (22). First, multiplying both sides by $(k+1)(\gamma^k-1)^2$, we see that inequality (22) becomes

$$\begin{aligned}0 \le k(\gamma^k-1)\Big(&k(k+1)(\gamma^{k+2}-1)-2k(k+2)(\gamma^{k+1}-1)+(k+1)(k+2)(\gamma^k-1)\Big)\\ &-(k+1)\Big(k(\gamma^{k+1}-1)-(k+1)(\gamma^k-1)\Big)^2 := f_0(\gamma).\end{aligned}$$

The function f_0 can be simplified into

$$f_0(\gamma)=2k\gamma^{2k+1}-(k+1)\gamma^{2k}-k^2(k+1)\gamma^{k+2}+2k(k^2+k-1)\gamma^{k+1}-(k^3+k^2-2)\gamma^k+(k-1).$$

Observing that $f_0(1)=0$, it suffices to show that $f_0'(\gamma)\ge 0$ for any $\gamma\ge 1$. By simple calculation,

$$\begin{aligned}f_0'(\gamma)=&2k(2k+1)\gamma^{2k}-2k(k+1)\gamma^{2k-1}-k^2(k+1)(k+2)\gamma^{k+1}\\ &+2k(k+1)(k^2+k-1)\gamma^k-k(k^3+k^2-2)\gamma^{k-1}\\ =&k\gamma^{k-1}f_1(\gamma),\end{aligned}$$

where

$$f_1(\gamma)=2(2k+1)\gamma^{k+1}-2(k+1)\gamma^k-k(k+1)(k+2)\gamma^2+2(k+1)(k^2+k-1)\gamma-(k^3+k^2-2).$$

Since $f_1(1)=0$, it suffices to show that $f_1'(\gamma)\ge 0$ for any $\gamma\ge 1$. Again by simple calculation,

$$\begin{aligned}f_1'(\gamma)&=2(k+1)(2k+1)\gamma^k-2k(k+1)\gamma^{k-1}-2k(k+1)(k+2)\gamma+2(k+1)(k^2+k-1)\\ &=2(k+1)f_2(\gamma),\end{aligned}$$

where

$$f_2(\gamma)=(2k+1)\gamma^k-k\gamma^{k-1}-k(k+2)\gamma+(k^2+k-1).$$

Since $f_2(1)=0$, it suffices to show that $f_2'(\gamma)\ge 0$ for any $\gamma\ge 1$. Finally,

$$\begin{aligned}f_2'(\gamma)&=k(2k+1)\gamma^{k-1}-k(k-1)\gamma^{k-2}-k(k+2)\\ &=k\gamma^{k-2}\left[(2k+1)\gamma-(k-1)\right]-k(k+2)\\ &\ge k\left[(2k+1)-(k-1)-(k+2)\right]\\ &=0,\end{aligned}$$

where the inequality follows from the fact that $\gamma\ge 1$. This shows that $f_0(\gamma)\ge 0$ for any $\gamma\ge 1$ and hence completes the proof of inequality (11).

4.2 Proof of Inequality (4)

Inequality (4) is trivial for distributions $p\in\mathcal{P}$ with $\mu_2(p)=0$. Therefore, we assume that $\mu_2(p)>0$. We will need the following notations. For any distribution $p\in\mathcal{P}$, we let

$$\eta(p)=\frac{\mu_1(p)}{\mu_2(p)},\quad \Xi(p)=\left|\frac{\mu_3^3(p)}{\mu_2^3(p)}\right| \quad\text{and}\quad \Xi_k=\sup_{p\in\mathcal{P}}\Xi(p).$$

Then, inequality (4) is equivalent to

$$\Xi_k\le 2\sqrt{\frac{k+2}{k}\frac{k-1}{k+3}}.$$

The following observation will be helpful:

$$\sup_{p\in\mathcal{P}} \Xi(p) = \sup_{p\in\mathcal{P}} \frac{\mu_3^3(p)}{\mu_2^3(p)}. \tag{23}$$

To prove it, for any $p \in \mathcal{P}$ such that

$$\frac{\mu_3^3(p)}{\mu_2^3(p)} < 0,$$

we define $\tilde{p}(x) = p(-x)$. Then we have that $\tilde{p} \in \mathcal{P}$ and that

$$\frac{\mu_3^3(\tilde{p})}{\mu_2^3(\tilde{p})} = -\frac{\mu_3^3(p)}{\mu_2^3(p)} > 0.$$

4.2.1 To $\frac{1}{k-1}$-Affine Distributions

Using the formulas

$$\int_{-\infty}^{\infty} t^2 p(t)dt = \mu_2^2(p) + \mu_1^2(p)$$

and

$$\int_{-\infty}^{\infty} t^3 p(t)dt = \mu_3^3(p) + 3\mu_1(p)\mu_2^2(p) + \mu_1^3(p),$$

we get

$$\frac{\int_{-\infty}^{\infty} t^3 p(t)dt}{\left(\int_{-\infty}^{\infty} t^2 p(t)dt\right)^{\frac{3}{2}}} = \frac{\frac{\mu_3^3(p)}{\mu_2^3(p)} + 3\eta(p) + \eta^3(p)}{(1+\eta^2(p))^{\frac{3}{2}}}. \tag{24}$$

Since $\mu_2(p)$ and $\mu_3(p)$ are invariant to horizontal shift of the distribution p,

$$\begin{aligned} \sigma := \sup_{p\in\mathcal{P}} \frac{\int_{-\infty}^{\infty} t^3 p(t)dt}{\left(\int_{-\infty}^{\infty} t^2 p(t)dt\right)^{\frac{3}{2}}} &= \sup_{\eta\in\mathbb{R}} \sup_{p\in\mathcal{P}} \frac{\frac{\mu_3^3(p)}{\mu_2^3(p)} + 3\eta + \eta^3}{(1+\eta^2)^{\frac{3}{2}}} \\ = \sup_{\eta\in\mathbb{R}} \frac{\left(\sup_{p\in\mathcal{P}} \frac{\mu_3^3(p)}{\mu_2^3(p)}\right) + 3\eta + \eta^3}{(1+\eta^2)^{\frac{3}{2}}} &= \sup_{\eta\in\mathbb{R}} \frac{\Xi_k + 3\eta + \eta^3}{(1+\eta^2)^{\frac{3}{2}}}, \end{aligned} \tag{25}$$

where the last equality follows from (23).

Similar to Section 4.1.1, we can approximate the supremum σ by truncating the distribution. In particular, one can prove that for any $\epsilon > 0$, there is a real number $M > 0$ such that $\sigma \le \sigma_M + \epsilon$, where

$$\begin{array}{rl} \sigma_M := \sup\limits_{p} & \int_{-\infty}^{\infty} t^3 p(t)dt \\ \text{subjectto} & \int_{-\infty}^{\infty} t^2 p(t)dt \le 1, \\ & p \in \mathcal{P}_M. \end{array} \tag{26}$$

and $\mathcal{P}_M \subseteq \mathcal{P}$ is the set of distributions $p \in \mathcal{P}$ with $\mathrm{Supp}(p) \subseteq [-M, M]$. By Theorem 3, the supremum σ_M of problem (26) is achieved by some distribution $q \in \mathcal{Q}$. Thus, we arrive at the following relation:

$$\sup_{p\in\mathcal{P}} \frac{\int_{-\infty}^{\infty} t^3 p(t)dt}{\left(\int_{-\infty}^{\infty} t^2 p(t)dt\right)^{\frac{3}{2}}} \le \sigma_M + 2\epsilon \le \int_{-\infty}^{\infty} t^3 q(t)dt + 2\epsilon$$
$$\le \frac{\int_{-\infty}^{\infty} t^3 q(t)dt}{\left(\int_{-\infty}^{\infty} t^2 q(t)dt\right)^{\frac{3}{2}}} + 2\epsilon \le \sup_{p\in\mathcal{Q}} \frac{\int_{-\infty}^{\infty} t^3 p(t)dt}{\left(\int_{-\infty}^{\infty} t^2 p(t)dt\right)^{\frac{3}{2}}} + 2\epsilon.$$

Taking limiting $\epsilon \searrow 0$ yields

$$\sup_{p\in\mathcal{P}} \frac{\int_{-\infty}^{\infty} t^3 p(t)dt}{\left(\int_{-\infty}^{\infty} t^2 p(t)dt\right)^{\frac{3}{2}}} \le \sup_{p\in\mathcal{Q}} \frac{\int_{-\infty}^{\infty} t^3 p(t)dt}{\left(\int_{-\infty}^{\infty} t^2 p(t)dt\right)^{\frac{3}{2}}}. \tag{27}$$

Combining (25) and (27) gives that

$$\sup_{\eta\in\mathbb{R}} \frac{\Xi_k + 3\eta + \eta^3}{(1+\eta^2)^{\frac{3}{2}}} \le \sup_{p\in\mathcal{Q}} \frac{\int_{-\infty}^{\infty} t^3 p(t)dt}{\left(\int_{-\infty}^{\infty} t^2 p(t)dt\right)^{\frac{3}{2}}} := \sigma'. \tag{28}$$

To bound Ξ_k, we consider two cases. Case 1: $\sigma' \le \sqrt{2}$. Putting $\eta = 1$ in (28) gives

$$\frac{\Xi_k + 4}{2^{\frac{3}{2}}} \le \sqrt{2},$$

which implies $\Xi_k \le 0$. Case 2: $\sigma' > \sqrt{2}$. Let $\bar{q} \in \mathcal{Q}$ be an ϵ-approximate maximizer of σ' with $\epsilon < 0.01$. Then,

$$\begin{aligned} \frac{\Xi_k + 3\eta(\bar{q}) + \eta^3(\bar{q})}{(1+\eta^2(\bar{q}))^{\frac{3}{2}}} &\le \sup_{\eta\in\mathbb{R}} \frac{\Xi_k + 3\eta + \eta^3}{(1+\eta^2)^{\frac{3}{2}}} = \sup_{p\in\mathcal{P}} \frac{\int_{-\infty}^{\infty} t^3 p(t)dt}{\left(\int_{-\infty}^{\infty} t^2 p(t)dt\right)^{\frac{3}{2}}} \\ &\le \frac{\int_{-\infty}^{\infty} t^3 \bar{q}(t)dt}{\left(\int_{-\infty}^{\infty} t^2 \bar{q}(t)dt\right)^{\frac{3}{2}}} + \epsilon \le \frac{\Xi(\bar{q}) + 3\eta(\bar{q}) + \eta^3(\bar{q})}{(1+\eta^2(\bar{q}))^{\frac{3}{2}}} + \epsilon, \end{aligned} \tag{29}$$

where the equality follows from (25), the second inequality from (27) and the last inequality from (24). On the other hand, we have

$$1.4 \le \frac{\int_{-\infty}^{\infty} t^3 \bar{q}(t)dt}{\left(\int_{-\infty}^{\infty} t^2 \bar{q}(t)dt\right)^{\frac{3}{2}}} \le \frac{\Xi(\bar{q}) + 3\eta(\bar{q}) + \eta^3(\bar{q})}{(1+\eta^2(\bar{q}))^{\frac{3}{2}}}. \tag{30}$$

Rearranging inequality (30) and then using AM-GM inequality, we have

$$0.4 \cdot \eta^3(\bar{q}) \le \Xi(\bar{q}) + 3 \cdot \eta(\bar{q}) \le \Xi(\bar{q}) + \frac{\eta(\bar{q})^3}{3} + \frac{2 \cdot 3^{\frac{3}{2}}}{3}.$$

Hence, we have $\eta^3(\bar{q}) \le 15 \cdot \Xi(\bar{q}) + 52$. Using this and inequality (29), we get

$$\Xi_k \le (1 + O(\epsilon)) \cdot \Xi(\bar{q}) + O(\epsilon).$$

Since ϵ is arbitrarily small, it suffices to show that

$$\Xi(\bar{q}) \le 2\sqrt{\frac{k+2}{k}\frac{k-1}{k+3}}. \tag{31}$$

4.2.2 To an Algebraic Inequality

Now we prove inequality (31). The case of $\alpha = 0$ or $a = b$ is trivial. So we assume that $\alpha \neq 0$ and $a < b$. We state without proof the simple observation that Ξ is invariant under translation and scaling.

Lemma 3. *Let $p \in \mathcal{Q}$ and $\tilde{p}(t) = |\tilde{\alpha}| \cdot p(\tilde{\alpha} t + \tilde{\beta})$, where $\tilde{\alpha}, \tilde{\beta} \in \mathbb{R}$ are real numbers with $\tilde{\alpha} \neq 0$. Then $\tilde{p} \in \mathcal{Q}$ and $\Xi(p) = \Xi(\tilde{p})$.*

By Lemma 3, instead of $\bar{q}$, it suffices to consider the distribution

$$\tilde{q}(t) = \frac{t^{k-1} \cdot \mathbb{1}_{[\alpha a+\beta, \alpha b+\beta]}(t)}{\int_{\alpha a+\beta}^{\alpha b+\beta} u^{k-1} du}.$$

Case 1: $\alpha a + \beta = 0$. Since $\alpha b + \beta > \alpha a + \beta = 0$, Lemma 3 allows us to simplify $\tilde{q}$ further to

$$\hat{q}(t) = \frac{t^{k-1} \cdot \mathbb{1}_{[0,1]}(t)}{\int_0^1 u^{k-1} du} = k \cdot t^{k-1} \cdot \mathbb{1}_{[0,1]}(t).$$

We compute

$$\mu_2^2(\hat{q}) = \int_0^1 t^2 \hat{q}(t) dt - \left(\int_0^1 t \hat{q}(t) dt \right)^2 = \frac{k}{k+2} - \left(\frac{k}{k+1} \right)^2 = \frac{k}{(k+1)^2 (k+2)},$$

and

$$\begin{aligned} \mu_3^3(\hat{q}) &= \int_0^1 t^3 \hat{q}(t) dt - 3\mu_1(\hat{q}) \mu_2^2(\hat{q}) - \mu_1^3(\hat{q}) \\ &= \frac{k}{k+3} - 3 \cdot \frac{k}{k+1} \cdot \frac{k}{(k+1)^2(k+2)} - \left(\frac{k}{k+1} \right)^3 \\ &= \frac{-2k(k-1)}{(k+1)^3 (k+2)(k+3)}. \end{aligned}$$

Therefore,

$$\Xi(\bar{q}) = \Xi(\hat{q}) = \frac{2k(k-1)}{(k+1)^3(k+2)(k+3)} \cdot \frac{(k+1)^3 (k+2)^{\frac{3}{2}}}{k^{\frac{3}{2}}} = 2\sqrt{\frac{k+2}{k}} \frac{k-1}{k+3}.$$

Case 2: $\alpha a + \beta > 0$. Again using Lemma 3, instead of $\tilde{q}$, it suffices to consider

$$\check{q}(t) = \frac{t^{k-1} \cdot \mathbb{1}_{[1,\gamma]}(t)}{\int_1^\gamma u^{k-1} du},$$

where $\gamma = \frac{\alpha b+\beta}{\alpha a+\beta} > 1$. Then it suffices to show that

$$
\begin{aligned}
&\frac{4(k+2)(k-1)^2}{k(k+3)^2} \geq (\Xi(\check{q}))^2 \\
&= \frac{\left(\int_1^\gamma t^3\check{q}(t)dt - 3\left(\int_1^\gamma t\check{q}(t)dt\right)\left(\int_1^\gamma t^2\check{q}(t)dt\right) + 2\left(\int_1^\gamma t\check{q}(t)dt\right)^3\right)^2}{\left(\int_1^\gamma t^2\check{q}(t)dt - \left(\int_1^\gamma t\check{q}(t)dt\right)^2\right)^3} \\
&= \frac{\left(\left(\int_1^\gamma t^{k-1}dt\right)^2\left(\int_1^\gamma t^{k+2}dt\right) - 3\left(\int_1^\gamma t^{k-1}dt\right)\left(\int_1^\gamma t^{k}dt\right)\left(\int_1^\gamma t^{k+1}dt\right) + 2\left(\int_1^\gamma t^{k}dt\right)^3\right)^2}{\left(\left(\int_1^\gamma t^{k-1}dt\right)\left(\int_1^\gamma t^{k+1}dt\right) - \left(\int_1^\gamma t^{k}dt\right)^2\right)^3} \\
&= \frac{\left(\left(\frac{\gamma^k-1}{k}\right)^2\left(\frac{\gamma^{k+3}-1}{k+3}\right) - 3\left(\frac{\gamma^k-1}{k}\right)\left(\frac{\gamma^{k+1}-1}{k+1}\right)\left(\frac{\gamma^{k+2}-1}{k+2}\right) + 2\left(\frac{\gamma^{k+1}-1}{k+1}\right)^3\right)^2}{\left(\left(\frac{\gamma^k-1}{k}\right)\left(\frac{\gamma^{k+2}-1}{k+2}\right) - \left(\frac{\gamma^{k+1}-1}{k+1}\right)^2\right)^3}.
\end{aligned}
$$

Upon rearranging terms, the above inequality is equivalent to

$$
\begin{aligned}
0 \leq\ & 4(k-1)^2\left((k+1)^2\left(\gamma^k-1\right)\left(\gamma^{k+2}-1\right) - k(k+2)\left(\gamma^{k+1}-1\right)^2\right)^3 \\
& - \Bigg(2k^2(k+2)(k+3)\left(\gamma^{k+1}-1\right)^3 + (k+2)(k+1)^3\left(\gamma^k-1\right)^2\left(\gamma^{k+3}-1\right) \\
& - 3k(k+3)(k+1)^2\left(\gamma^k-1\right)\left(\gamma^{k+1}-1\right)\left(\gamma^{k+2}-1\right)\Bigg)^2 := g(\gamma).
\end{aligned}
\tag{32}
$$

We prove inequality (32) with the help of computer algebra systems (*e.g.*, MATHEMATICA). The procedures are the same as those for proving (22), see Section 4.1.7. Specifically, we first check that $g(1) \geq 0$. So it suffices to show that g' is non-decreasing on $\gamma \geq 1$. We then compute the derivative g' and extract non-negative factors. We found that again g' is non-negative at $\gamma = 1$. Therefore, it suffices to show that the second derivative $g^{(2)}$ is non-decreasing on $\gamma \geq 1$. Repeating these procedures 17 times (checking $g^{(i)}(1) \geq 0$, extracting non-negative factors and differentiating), we prove that inequality (32) indeed holds for any $\gamma \geq 1$ and integer $k \geq 1$. Since the proof does not provide much insight and the derivatives of g are tedious, we omit the proof here. Interested readers can download the program for generating the proof from the personal websites of the authors.

5 Conclusion

This paper showed that the universal barrier of Nesterov and Nemirovski [10] is n-self-concordant on any proper convex domain in $\mathbb{R}^n$. The key to the proof of this result is a pair of new, sharp moment inequalities for s-concave distributions, which could be of independent interest. Currently, our inequalities concern only the first three moments. An interesting research question would be to generalize them to higher moments. Also, it would be ideal to have a simpler proof without the assistance of computer algebra system.

Acknowledgement

We thank Sébastien Bubeck and Santosh S. Vempala for helpful discussions. This work was supported in part by NSF awards CCF-1717349 and CCF-1740551.

References

[1] J. Abernethy and A. Rakhlin. Beating the Adaptive Bandit with High Probability. In *Information Theory and Applications Workshop, 2009*, pages 280–289. IEEE, 2009.

[2] S. Brazitikos, A. Giannopoulos, P. Valettas, and B.-H. Vritsiou. *Geometry of Isotropic Convex Bodies*, volume 196 of *Mathematical Surveys and Monographs*. American Mathematical Society, 2014.

[3] S. Bubeck and R. Eldan. The Entropic Barrier: A Simple and Optimal Universal Self-concordant Barrier. In *Conference on Learning Theory*, pages 279–279, 2015.

[4] D. J. Fox. A Schwarz Lemma for Kähler Affine Metrics and the Canonical Potential of a Proper Convex Cone. *Annali di Matematica Pura ed Applicata (1923-)*, 194(1):1–42, 2015.

[5] M. Fradelizi and O. Guédon. The Extreme Points of Subsets of s-Concave Probabilities and a Geometric Localization Theorem. *Discrete & Computational Geometry*, 31(2):327–335, 2004.

[6] O. Güler. Barrier Functions in Interior Point Methods. *Mathematics of Operations Research*, 21(4):860–885, 1996.

[7] R. Hildebrand. Canonical Barriers on Convex Cones. *Mathematics of operations research*, 39(3):841–850, 2014.

[8] Y. T. Lee and A. Sidford. Path Finding Methods for Linear Programming: Solving Linear Programs in $\tilde{O}(\sqrt{\text{rank}})$ Iterations and Faster Algorithms for Maximum flow. In *Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on*, pages 424–433. IEEE, 2014.

[9] H. Narayanan and A. Rakhlin. Random Walk Approach to Regret Minimization. In *Advances in Neural Information Processing Systems*, pages 1777–1785, 2010.

[10] Y. Nesterov and A. Nemirovskii. *Interior-point Polynomial Algorithms in Convex Programming*, volume 13. Siam, 1994.