

# The continuous $d$ -step conjecture for polytopes

Antoine Deza, Tamás Terlaky and Yuriy Zinchenko\*

May 29, 2008

## Abstract

The curvature of a polytope, defined as the largest possible total curvature of the associated central path, can be regarded as the continuous analogue of its diameter. We prove the analogue of the result of Klee and Walkup. Namely, we show that if the order of the curvature is less than the dimension  $d$  for all polytope defined by  $2d$  inequalities and for all  $d$ , then the order of the curvature is less than the number of inequalities for all polytopes.

**Keywords:** polytopes, diameter, Hirsch conjecture,  $d$ -step conjecture, central path, curvature.

---

\*Corresponding author, tel.: +1 905 525 9140 ext. 26866, e-mail: zinchen@mcmaster.ca, postal address: CAS, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4K1

# 1 Continuous analogue of the conjecture of Hirsch

Let  $P$  be a full dimensional convex polyhedron defined by  $n$  inequalities in dimension  $d$ . Given a pair of vertices  $v = \{v_1, v_2\}$  of  $P$ , let  $\delta^v(P)$  denote the smallest number such that  $v_1$  and  $v_2$  can be connected by a path with at most  $\delta^v(P)$  edges. Considering the maximum of  $\delta^v(P)$  over all possible pairs of vertices, we obtain the diameter  $\delta(P)$ , that is, the smallest number such that any two vertices of the polyhedron  $P$  can be connected by a path with at most  $\delta(P)$  edges. The conjecture of Hirsch, formulated in 1957 and reported in [1], states that the diameter of a polyhedron defined by  $n$  inequalities in dimension  $d$  is not greater than  $n - d$ . The conjecture does not hold for unbounded polyhedra. A polytope is a bounded polyhedron. No polynomial bound is known for the diameter of a polytope; the best one is  $2n^{\log(d)+1}$ , see Kalai and Kleitman [8].

## Conjecture 1.1. (Conjecture of Hirsch for polytopes)

*The diameter of a polytope defined by  $n$  inequalities in dimension  $d$  is not greater than  $n - d$ .*

We consider the following continuous analogue of the diameter introduced in [4]: the largest possible total curvature of the associated central path. We first recall the definitions of the central path and of the total curvature. For a polytope  $P = \{x : Ax \geq b\}$  with  $A \in \mathbb{R}^{n \times d}$ , the central path corresponding to  $\min\{c^T x : x \in P\}$  is a set of minimizers of  $\min\{c^T x + \mu f(x) : x \in P\}$  for  $\mu \in (0, \infty)$  where  $f(x)$  is the so-called strongly non-degenerate self-concordant barrier function [10] for  $P$ , and throughout this paper  $f(x) = -\sum_{i=1}^n \ln(A_i x - b_i)$  – the standard logarithmic barrier function. As a consequence of strict convexity of the logarithmic barrier function, the central path is unique. It is known that the central path is an analytic curve [12], and thus, in particular, is fully determined by any of its arbitrary small continuous segment or all of its derivatives at any a single point. Conversely, the analyticity of the central path should allow us to reconstruct the full description of a polytope  $P$  up to its orientation given by the objective vector  $c$  and possibly some equivalence relationship between the polytopes based on the central path.

Intuitively, the total curvature [13] is a measure of how far off a certain curve is from being a straight line. Let  $\psi : [\alpha, \beta] \rightarrow \mathbb{R}^d$  be a  $C^2((\alpha - \varepsilon, \beta + \varepsilon))$  map for some  $\varepsilon > 0$  with a non-zero derivative in  $[\alpha, \beta]$ . Denote its arc length by  $l(t) = \int_{\alpha}^t \|\dot{\psi}(\tau)\| d\tau$ , its parametrization by the arc length by  $\psi_{\text{arc}} = \psi \circ l^{-1} : [0, l(\beta)] \rightarrow \mathbb{R}^d$ , and its curvature at the point  $t$  by  $\kappa(t) = \ddot{\psi}_{\text{arc}}(t)$ . The total curvature is defined as  $\int_0^{l(\beta)} \|\kappa(t)\| dt$ . The requirement  $\dot{\psi} \neq 0$  insures that any given segment of the curve is traversed only once and allows to define a curvature at any point on the curve.

Let  $\lambda^c(P)$  denote the total curvature of the central path corresponding to the linear optimization problem  $\min\{c^T x : x \in P\}$ . The quantity  $\lambda^c(P)$  can be regarded as the continuous analogue of  $\delta^v(P)$ , the edge-length of the shortest path between a pair of vertices. Considering the largest  $\lambda^c(P)$  over all possible  $c$  we obtain the quantity  $\lambda(P)$ , referred to as the curvature of a polytope. The following analogue of the conjecture of Hirsch was proposed in [4].

**Conjecture 1.2. (Continuous analogue of the conjecture of Hirsch)**

*The order of the curvature of a polytope defined by  $n$  inequalities in dimension  $d$  is  $n$ .*

Holt and Klee [6] showed that, for  $n > d \geq 13$ , the conjecture of Hirsch is tight. Fritzsche and Holt [5] extended the result to  $n > d \geq 8$ . The continuous analogue would be that Conjecture 1.2 is the best possible. In [3] the authors showed that a redundant Klee-Minty  $d$ -cube  $\mathcal{C}$  satisfies  $\lambda(\mathcal{C}) \geq (\frac{3}{2})^d$ , and in [4] provided a family of  $d$ -dimensional polytopes  $\mathcal{P}$  defined by  $n > 2d$  non-redundant inequalities satisfying  $\liminf_{n \rightarrow \infty} \lambda(\mathcal{P})/n \geq \pi$  for a fixed  $d$ . In other words, the continuous analogue of the result of Holt and Klee holds.

**Proposition 1.1. (Continuous analogue of the result of Holt and Klee)**

*The conjectured order- $n$  of the curvature is attained.*

An exponential upper bound on the curvature of a polytope is provided by the result of Dedieu, Malajovich and Shub [2]. Consider a simple arrangement formed by  $n$  hyperplanes in dimension  $d$ ; we recall that an arrangement is called simple if  $n \geq d + 1$  and any  $d$  hyperplanes intersect at a unique

distinct point. The number  $I$  of bounded cells  $P_i, i = 1, \dots, I$  of the arrangement, i.e., closures of the bounded connected components of the complement to the hyperplanes, satisfies  $I = \binom{n-1}{d}$ . Dedieu, Malajovich and Shub [2] demonstrated that  $\sum_{i=1}^I \lambda^c(P_i)/I \leq 2\pi d$  for any fixed  $c$ . It immediately implies that  $\lambda^c(P) \leq 2\pi d \binom{n-1}{d} < 2\pi d n^{d-1} < 2\pi n^d$  for any  $c$ , i.e.,  $\lambda(P) \leq 2\pi n^d$ .

**Proposition 1.2. (Exponential upper bound on the curvature)**

*The curvature of a polytope defined by  $n$  inequalities in dimension  $d$  is not greater than  $2\pi n^d$ .*

The special case of  $n = 2d$  of the conjecture of Hirsch is known as the  $d$ -step conjecture, and its continuous analogue is:

**Conjecture 1.3. (Continuous analogue of the  $d$ -step conjecture)**

*The order of the curvature of a polytope defined by  $2d$  inequalities in dimension  $d$  is  $d$ .*

It has been shown by Klee and Walkup [9] that the special case  $n = 2d$  for all  $d$  is equivalent to the conjecture of Hirsch. We prove a continuous analogue of the result of Klee and Walkup, namely, we show in Section 2 that Conjecture 1.2 and Conjecture 1.3 are equivalent.

To motivate the study of the diameter and the total curvature consider solving a linear optimization problem as above. Roughly speaking, the diameter may be thought of as a lower bound on the number of pivots for a simplex method for the worst possible objective vector  $c$ , while the total curvature of the central path gives an insight into the performance of a path-following interior-point method (straight line is easy to follow). For polytopes and arrangements, respectively central path and linear optimization, we refer to the books of Grünbaum [7] and Ziegler [14], respectively Renegar [10] and Roos et al [11].

## 2 The continuous analogue of the result of Klee and Walkup

Following the analogy with the diameter, let  $\Lambda(d, n)$  be the largest total curvature  $\lambda(P)$  of the central path over all polytopes  $P$  defined by  $n$  inequalities in dimension  $d$ . Klee and Walkup [9] showed that

for the Hirsch conjecture the special case where the number of inequalities is twice the dimension is equivalent to the general case. We prove that the same holds for the continuous Hirsch conjecture.

**Theorem 2.1. (Continuous analogue of the result of Klee and Walkup)**

*The continuous Hirsch conjecture is equivalent to the continuous  $d$ -step conjecture; that is, if  $\Lambda(d, 2d) = \mathcal{O}(d)$  for all  $d$ , then  $\Lambda(d, n) = \mathcal{O}(n)$ .*

**Remark 2.1.** *In Theorem 2.1 the constant 2 may be replaced by any integer  $k > 1$ , i.e., if  $\Lambda(d, kd) = \mathcal{O}(d)$  for all  $d$ , then  $\Lambda(d, n) = \mathcal{O}(n)$ .*

The rest of Section 2 is devoted to the proof of Theorem 2.1. Suppose  $\Lambda(d, 2d) \leq 2Kd$ ,  $d \geq 2$ . Consider  $\min\{c^T x : x \in P\}$  where  $P = \{x : Ax \geq b\}$  with  $A \in \mathbb{R}^{n \times d}$  and  $\mathcal{P}$  and  $\chi$  respectively denotes the associated central path and the analytic center, i.e., the minimizer of the standard logarithmic barrier  $f(x)$  over  $P$ . The two cases  $d < n < 2d$  and  $n > 2d$  are considered separately. Denote  $\mathbf{0}$  and  $\mathbf{1}$  the vector of all zeros and all ones,  $\text{int}P$  – the interior of  $P$ . We may assume  $P$  is full-dimensional, for if not, we may reduce the problem dimension to satisfy the assumption. Note  $A$  is full-rank since  $P$  is bounded.

Case  $d < n < 2d$ : without loss of generality assume  $c = (1, 0, \dots, 0) \in \mathbb{R}^d$  and denote  $x_1^*$  the optimal value of  $\min\{c^T x : x \in P\}$ . Consider  $\min\{c^T x : x \in \tilde{P}\}$  where  $\tilde{P} = \{x : \tilde{A}x \geq \tilde{b}\}$  with  $\tilde{A} \in \mathbb{R}^{2d, d}$  and  $\tilde{b} \in \mathbb{R}^{2d}$  are given by:

$$\tilde{A}_{i,j} = \begin{cases} A_{i,j} & \text{for } i = 1, \dots, n \text{ and } j = 1, \dots, d, \\ 1 & \text{for } i = n + 1, \dots, 2d \text{ and } j = 1, \\ 0 & \text{for } i = n + 1, \dots, 2d \text{ and } j = 2, \dots, d, \end{cases} \quad \tilde{b}_i = \begin{cases} b_i & \text{for } i = 1, \dots, n, \\ x_1^* - 1 & \text{for } i = n + 1, \dots, 2d, \end{cases}$$

and  $\tilde{\mathcal{P}}$  denotes the associated central path, see Figure 1(a). Recall that the central path may be parameterized as the collection of the analytic centers of the level sets between the optimal solution and  $\chi$ ; that is,  $\mathcal{P}$  is the set of minimizers of  $f(x)$  over  $P \cap \{x : c^T x = \omega\}$  where  $\omega \in (x_1^*, \chi_1)$ . Therefore, we have  $\mathcal{P} \subseteq \tilde{\mathcal{P}}$  and consequently  $\lambda^c(P) \leq \lambda^c(\tilde{P})$ . As  $\tilde{P}$  is defined by  $2d$  inequalities in dimension  $d$ , we have  $\lambda^c(\tilde{P}) \leq 2Kd$ , and thus, for  $d < n < 2d$ ,  $\lambda^c(P) \leq 2Kd$ , that is,  $\Lambda(d, n) = \mathcal{O}(n)$ .

Case  $n > 2d$ : without loss of generality assume  $\chi = \mathbf{0}$ . Consider  $\min\{(c, -\theta)^T(x, y) : (x, y) \in \tilde{P}\}$  where  $\tilde{P} = \{(x, y) \in \mathbb{R}^d \times \mathbb{R} : Ax - by \geq 0, y \leq 1\}$  with the associated central path  $\tilde{\mathcal{P}}$ . In particular, if the definition of  $P$  is non-redundant,  $\tilde{P}$  is the  $(d+1)$ -dimensional flipped pyramid with base  $P \times \{1\}$  and the apex  $(\chi, 0) = \mathbf{0}$ . We show that for large enough value of  $\theta \gg \|c\|$  the central path  $\mathcal{P}$  of the original optimization problem may be well approximated by a segment of the central path  $\tilde{\mathcal{P}}$  so that the total curvature of  $\mathcal{P}$  is bounded from above by the total curvature of  $\tilde{\mathcal{P}}$ . Intuitively, by choosing  $\theta$  large enough we should be able to force  $\tilde{\mathcal{P}}$  to first follow almost a straight line from the analytic center of  $\tilde{P}$  to the face containing  $P \times \{1\}$ , and once  $\tilde{\mathcal{P}}$  is forced almost onto this face the path should look just like the central path  $\mathcal{P}$  for  $\min\{c^T x : x \in P\}$  in one-less dimension, see Figure 1(b). Consequently, we argue that the total curvature of  $\mathcal{P}$  may not be less than that of  $\tilde{\mathcal{P}}$ .

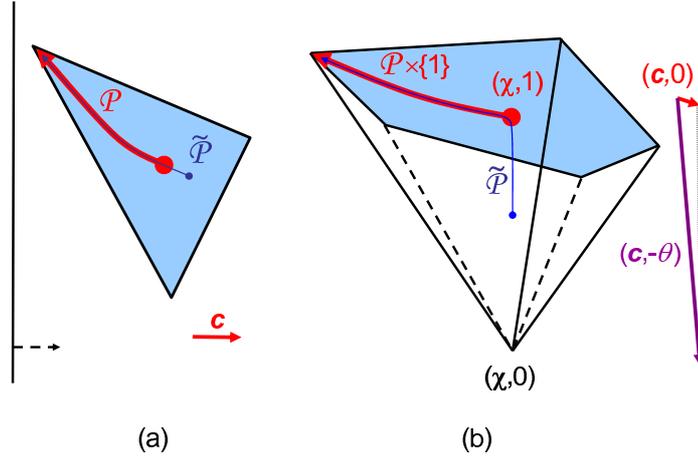


Figure 1: An illustration for the proof of Theorem 2.1 for  $d = 2$ : (a)  $2d > n = 3$ , (b)  $2d < n = 5$

Denote  $s = Ax - by$  and let  $S$  be the diagonal matrix with  $S_{ii} = s_i$ ,  $i = 1, \dots, d$ . Recall that the barrier functions  $\tilde{f}(x, y) : \text{int}\tilde{P} \rightarrow \mathbb{R}$  and  $f(x) : \text{int}P \rightarrow \mathbb{R}$  giving rise to the central paths  $\tilde{\mathcal{P}} = \{(\tilde{x}, \tilde{y})(\nu) = \arg \min_{(x, y)} \nu(c, -\theta)^T(x, y) + \tilde{f}(x, y), \nu \in [0, \infty)\}$  and  $\mathcal{P} = \{x(\nu) = \arg \min_x \nu c^T x + f(x), \nu \in [0, \infty)\}$  are  $\tilde{f}(x, y) = -\sum_{i=1}^n \ln s_i - \ln(1 - y)$  and  $f(x) = -\sum_{i=1}^n \ln s_i|_{y=1}$ . The gradient  $\tilde{g}$

and the Hessian  $\tilde{H}$  of  $\tilde{f}$  at a point  $(x, y)$  satisfy

$$\tilde{g} = \begin{pmatrix} -A^T S^{-1} \mathbf{1} \\ \frac{1}{1-y} + b^T S^{-1} \mathbf{1} \end{pmatrix}, \quad \tilde{H} = \begin{pmatrix} A^T S^{-2} A & -A^T S^{-2} b \\ (-A^T S^{-2} b)^T & \frac{1}{(1-y)^2} + b^T S^{-2} b \end{pmatrix}$$

where the Hessian inverse may be computed using Schur complement and is equal to

$$\tilde{H}^{-1} = \begin{pmatrix} (A^T S^{-2} A)^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \frac{1}{\frac{1}{(1-y)^2} - \gamma} \begin{pmatrix} (A^T S^{-2} A)^{-1} A^T S^{-2} b \\ 1 \end{pmatrix} \begin{pmatrix} (A^T S^{-2} A)^{-1} A^T S^{-2} b \\ 1 \end{pmatrix}^T$$

where  $\gamma = b^T S^{-2} b - (A^T S^{-2} b)^T (A^T S^{-2} A)^{-1} A^T S^{-2} b$ .

In what follows we adapt the notation of [10]: we write  $\|u\|_x$  for the local norm of  $v$  at  $x$  where the norm is induced by the intrinsic inner product at  $x \in D_f$  arising from the strongly non-degenerate self-concordant function  $f : D_f \rightarrow \mathbb{R}$ , namely  $\|u\| = u^T \nabla^2 f(x) u$ . First we show that for  $\theta$  large enough any segment of  $\mathcal{P}$  corresponding to  $\nu \in [\underline{\nu}, \bar{\nu}]$  may be well approximated by a suitably chosen segment of  $\tilde{\mathcal{P}}$ ; in doing so we will manufacture a good surrogate for a point  $(\tilde{x}, \tilde{y})(\nu) \in \tilde{\mathcal{P}}$  from  $x \left( \frac{1}{2}(\nu + \sqrt{\nu^2 - \nu/\theta}) \right) \in \mathcal{P}$ .

**Proposition 2.1.** *Let  $\nu \in [\underline{\nu}, \bar{\nu}]$  and  $M$  be such that  $|b^T S^{-1} \mathbf{1}|, \sqrt{|\gamma|} \leq M$  for all  $(yx(\nu), y)$ ,  $\frac{1}{2} \leq y \leq 1$ ,  $\nu \in [\underline{\nu}, \bar{\nu}]$ . If  $\theta$  is chosen large enough to satisfy  $\theta \geq \frac{1}{\underline{\nu}} \max\{1, \sqrt{82}M - 1\}$ , then for  $y = 1 - \frac{1}{1+\theta\nu}$  the point  $(yx(\nu), y)$  approximates  $(\tilde{x}, \tilde{y})(\nu/y)$ , namely*

$$\left\| \begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix} (\nu/y) - \begin{pmatrix} yx(\nu) \\ y \end{pmatrix} \right\|_{(yx(\nu), y)} \leq \frac{\sqrt{82}}{3} \frac{M}{1 + \theta\nu}.$$

**Remark 2.2.**  $M < \infty$  since the portion of the central path  $\mathcal{P}$  corresponding to any finite interval  $[\underline{\nu}, \bar{\nu}]$  lies in the interior of  $P$ .

**Remark 2.3.** With  $y = 1 - \frac{1}{1+\theta\nu}$  the function  $\nu \mapsto \frac{\nu}{y}$  is monotone increasing for  $\nu \geq 0$  and hence is invertible with its inverse being  $\nu \mapsto \frac{1}{2}(\nu + \sqrt{\nu^2 - \nu/\theta})$ .

**Remark 2.4.** Recalling that  $\tilde{f}$  is strongly non-degenerate self-concordant and, therefore, an open unit ball in the local norm at  $(x, y) \in \text{int}\tilde{P}$  satisfies  $B_{(x,y)}((x, y), 1) \subset \text{int}\tilde{P}$  where  $\tilde{P}$  is bounded by

assumption, the proposition implies that  $(\tilde{x}, \tilde{y})(\nu) \rightarrow (x(\nu), 1)$  as  $\theta \rightarrow \infty$ . Moreover, the convergence is uniform over any finite segment  $[\underline{\nu}, \bar{\nu}]$  of  $\mathcal{P}$ .

*Proof.* We rely on the fact that the Newton's method iterates exhibit local quadratic convergence to the central path for linear optimization [11], which may be rephrased in a more general setting of strongly non-degenerate self-concordant functions [10]. For concreteness we use an intermediate statement in the proof of Proposition 2.2.8 of [10]: let  $n(x) = -\nabla^2 f(x)^{-1} \nabla f(x)$  be the Newton step for strongly non-degenerate self-concordant function  $f : D_f \rightarrow \mathbb{R}$ , if  $\|n(x)\|_x \leq \frac{1}{9}$  then  $f$  has a minimizer  $z$  and  $\|x - z\|_x \leq 3\|n(x)\|_x$ .

Consider the Newton step for minimizing  $\nu'(c, -\theta)^T(x, y) + \tilde{f}(x, y)$ , and evaluate (the square of) its local norm at  $(yx(\nu), y)$  where  $y = 1 - \frac{1}{1+\theta\nu}$  and  $\nu' = \frac{\nu}{y}$ :

$$\begin{aligned} & \left| \left( \nu' \begin{pmatrix} c \\ -\theta \end{pmatrix} + \tilde{g} \right)^T \tilde{H}^{-1} \left( \nu' \begin{pmatrix} c \\ -\theta \end{pmatrix} + \tilde{g} \right) \right|_{(yx(\nu), y)} \\ &= \left| \begin{pmatrix} \frac{\nu}{y}c + \frac{1}{y}\nabla f \\ \frac{-\nu}{y}\theta + \frac{1}{1-y} + b^T S^{-1}\mathbf{1} \end{pmatrix}^T \tilde{H}^{-1} \begin{pmatrix} \frac{\nu}{y}c + \frac{1}{y}\nabla f \\ \frac{-\nu}{y}\theta + \frac{1}{1-y} + b^T S^{-1}\mathbf{1} \end{pmatrix} \right|_{(yx(\nu), y)} \end{aligned}$$

and since  $x(\nu) \in \mathcal{P}$

$$\begin{aligned} &= \left| \begin{pmatrix} 0 \\ \frac{-\nu}{y}\theta + \frac{1}{1-y} + b^T S^{-1}\mathbf{1} \end{pmatrix}^T \tilde{H}^{-1} \begin{pmatrix} 0 \\ \frac{-\nu}{y}\theta + \frac{1}{1-y} + b^T S^{-1}\mathbf{1} \end{pmatrix} \right|_{(yx(\nu), y)} \\ &= \left( \frac{-\nu}{y}\theta + \frac{1}{1-y} + b^T S^{-1}\mathbf{1} \right)^2 \frac{1}{\frac{1}{(1-y)^2} + \gamma} \Big|_{(yx(\nu), y)} \end{aligned}$$

and by the choice of  $y$  and  $\theta$

$$= \frac{(b^T S^{-1}\mathbf{1})^2}{\frac{1}{(1-y)^2} + \gamma} \Big|_{(yx(\nu), y)} \leq \frac{M^2}{(1+\theta\nu)^2 - M^2} \leq \frac{82}{81} \left( \frac{M}{1+\theta\nu} \right)^2 \leq \frac{1}{81}.$$

Now, since the size of the Newton step for minimizing  $\nu'(c, -\theta)^T(x, y) + \tilde{f}(x, y)$  measured with respect to the local norm at  $(yx(\nu), y)$  is indeed  $\leq \frac{1}{9}$ , the statement of the proposition follows immediately.  $\square$

We present one useful proposition. Roughly speaking, it states that two similar curves might not differ greatly in their total curvatures either.

**Proposition 2.2.** *Let  $\psi : [\alpha, \beta] \rightarrow \mathbb{R}^d$  be a  $C^2((\alpha - \varepsilon, \beta + \varepsilon))$  map for some  $\varepsilon > 0$  with a non-zero derivative in  $[\alpha, \beta]$ , and  $\{\phi^j\}_{j=0,1,\dots}$  be a sequence of  $C^2((\alpha - \varepsilon, \beta + \varepsilon))$  functions with non-zero derivatives in  $[\alpha, \beta]$  that converge to  $\psi$  point-wise as  $j \rightarrow \infty$ , i.e.,  $\phi^j(t) \rightarrow \psi(t)$  for all  $t \in [\alpha, \beta]$ . Then the total curvature of  $\psi$  is bounded from above by the infimum limit of the total curvature of  $\phi^j$  over all  $j$ .*

For compactness we sketch an elementary proof of this proposition in Section 3. For a detailed exposition of a very similar argument see [3].

Next we argue that the total curvature of any finite segment  $[\underline{\nu}, \bar{\nu}]$  of the central path  $\mathcal{P}$  may not be much less than the total curvature of  $[\underline{\nu}, \bar{\nu}]$ -segment of  $\tilde{\mathcal{P}}$ . Indeed, this follows from Proposition 2.2 where the point-wise convergence of two paths follows from Remark 2.4. Note that the gradient to the central path does not vanish, since a point  $x(\nu) \in \mathcal{P}$  is the minimizer for  $\nu c^T x + f(x)$  and hence must satisfy  $\nu c + \nabla f(x) = 0$ , the derivative with respect to  $\nu$  satisfies  $\dot{x}(\nu) = -\nabla^2 f(x)^{-1} c \neq 0$ , recalling that under our assumptions  $\nabla^2 f(x)$  is non-singular. Using a similar argument one can show  $\mathcal{P}$  is  $C^2$  with respect to  $\nu$ , although a particular parametrization of  $\mathcal{P}$  is not important since it is already well-known that  $\mathcal{P}$  is real-analytic [12]. Same considerations apply to  $\tilde{\mathcal{P}}$ .

In turn, the total curvature of  $\mathcal{P}$  may be arbitrary well approximated by the total curvature of a suitably chosen finite segment of the path, letting  $\underline{\nu} \rightarrow 0$  and  $\bar{\nu} \rightarrow \infty$ . The later follows from the finiteness of the total curvature of  $\mathcal{P}$  established in [2], or intuitively from the fact that  $\mathcal{P}$  is asymptotically straight as  $\nu \rightarrow 0$  and  $\nu \rightarrow \infty$ .

Finally, we may summarize our findings in the following Lemma.

**Lemma 2.1.** *With the construction above as  $\theta \rightarrow \infty$  we have  $\liminf \lambda^{(c, -\theta)}(\tilde{P}) > \lambda^c(P)$ .*

The proof that  $\liminf \lambda^{(c, -\theta)}(\tilde{P}) \geq \lambda^c(P)$  easily follows from the Remark 2.4 and Proposition 2.2.

For strict inequality using the techniques in [3, 4] one can show that just before  $\tilde{\mathcal{P}}$  “starts to converge” to  $\mathcal{P}$ ,  $\tilde{\mathcal{P}}$  is bound to make a first sharp turn which in the limit will contribute to  $\pi/2$  additional total curvature for this path.

Now, inductively increasing the dimension and the number of inequalities by 1 and carefully using the limit argument in the above, the same construction gives a sequence of polytopes  $\tilde{\mathcal{P}} = \tilde{P}_1, \tilde{P}_2, \dots, \tilde{P}_{n-2d}$  satisfying  $\lambda^c(P) \leq \lambda^{(c, -\theta)}(\tilde{P}_1) \leq \dots \leq \lambda^{(c, -\theta, \dots)}(\tilde{P}_{n-2d})$ . Since  $\tilde{P}_{n-2d}$  is defined by  $2n - 2d$  inequalities in dimension  $n - d$ , we have  $\lambda^{(c, -\theta, \dots)}(\tilde{P}_{n-2d}) \leq 2K(n - d)$ . This implies that, for  $n > 2d$ ,  $\lambda^c(P) \leq 2K(n - d)$ , that is,  $\Lambda(d, n) = \mathcal{O}(n)$ .

### 3 Sketch of the proof of Proposition 2.2

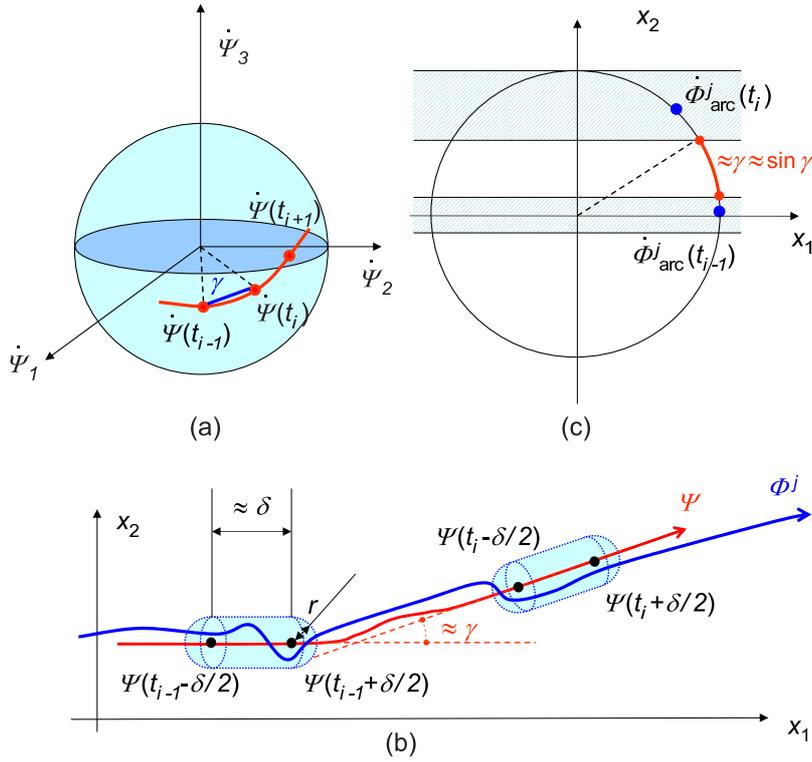


Figure 2: Illustration of the proof of Proposition 2.2

Observe that without loss of generality we may assume  $\psi$  is parameterized by the arc length: a short computation shows that if  $\dot{\psi} \neq 0$  then the second derivative of the arc length parametrization of the curve is continuous since  $\dot{\psi}_{\text{arc}}(t) = \psi(l^{-1}(t))'_t = \dot{\psi}(l^{-1}(t)) \cdot (l^{-1}(t))'_t = \frac{\dot{\psi}(l^{-1}(t))}{\|\dot{\psi}(l^{-1}(t))\|}$ , and consequently  $\ddot{\psi}_{\text{arc}} \circ l = \frac{\ddot{\psi}\|\dot{\psi}\|^2 - \dot{\psi}(\ddot{\psi}^T \dot{\psi})}{\|\dot{\psi}\|^3}$ . In what follows we remove the arc subscript from  $\psi$  to shorten the notation and write  $\alpha$  for  $l(\alpha)$  and  $\beta$  for  $l(\beta)$ . Note that the total curvature of  $\psi$  corresponds to the length of the gradient curve  $\dot{\psi}$  between the points  $\dot{\psi}(\alpha)$  and  $\dot{\psi}(\beta)$  which in turn belong to the unit  $n$ -sphere.

First we argue that the total curvature of  $\psi$  may be arbitrarily well approximated by the sum of chordal distances between pairs  $(\dot{\psi}(t_{i-1}), \dot{\psi}(t_i))$ ,  $\sum_{i=1}^N \|\dot{\psi}(t_i) - \dot{\psi}(t_{i-1})\|$ , where  $t_0 = \alpha < t_1 < t_2 < \dots < t_N = \beta$  as long as  $\max_{i=1, \dots, N} (t_i - t_{i-1}) \rightarrow 0$ , see Figure 2(a). Trivially  $\int_{\alpha}^{\beta} \|\ddot{\psi}(t)\| dt = \sum_{i=1}^N \int_{t_{i-1}}^{t_i} \|\ddot{\psi}(t)\| dt$  and observe

$$\begin{aligned} & \left| \int_{t_{i-1}}^{t_i} \|\ddot{\psi}(t)\| dt - \int_{t_{i-1}}^{t_i} \frac{\|\dot{\psi}(t_i) - \dot{\psi}(t_{i-1})\|}{t_i - t_{i-1}} dt \right| \leq \int_{t_{i-1}}^{t_i} \left| \|\ddot{\psi}(t)\| - \frac{\|\dot{\psi}(t_i) - \dot{\psi}(t_{i-1})\|}{t_i - t_{i-1}} \right| dt \\ & \leq \int_{t_{i-1}}^{t_i} \left\| \ddot{\psi}(t) - \frac{\dot{\psi}(t_i) - \dot{\psi}(t_{i-1})}{t_i - t_{i-1}} \right\| dt = \int_{t_{i-1}}^{t_i} \left\| \ddot{\psi}(t) - \begin{pmatrix} \ddot{\psi}_1(\bar{t}_1) \\ \ddot{\psi}_2(\bar{t}_2) \\ \vdots \\ \ddot{\psi}_n(\bar{t}_n) \end{pmatrix} \right\| dt \leq \int_{t_{i-1}}^{t_i} M dt = \frac{M(t_i - t_{i-1})^2}{2} \end{aligned}$$

with  $t_{i-1} \leq \bar{t}_i \leq t_i$  for all  $i$ , where the second inequality follows from the triangle inequality, the first equality follows from the Intermediate-Value Theorem, and the last inequality is implied by the Lipschitz continuity of the second derivative of  $\psi$  on the compact set where the Lipschitz constant is denoted by  $M$ . So, up to a quadratic error term the curvature contribution over the segment  $[t_{i-1}, t_i]$  may be approximated by the length of the linear segment connecting the starting point  $\dot{\psi}(t_{i-1})$  and the ending point  $\dot{\psi}(t_i)$  with both points on the unit sphere, implying a linear error term for the total curvature approximation over  $[\alpha, \beta]$ . In particular, we may consider the partitioning  $t_0 = \alpha < t_1 < t_2 < \dots < t_N = \beta$  where each  $\|\dot{\psi}(t_i) - \dot{\psi}(t_{i+1})\| = \gamma$  for some small fixed  $\gamma$ .

Note that any partitioning of  $[\alpha, \beta]$  trivially gives a lower bound of  $\sum_{i=1}^N \|\dot{\phi}^j(t_i) - \dot{\phi}^j(t_{i-1})\|$  on the total curvature of  $\phi^j$  for all  $j$ . Also, note that since  $\phi^j(t) \rightarrow \psi(t)$  as  $j \rightarrow \infty$  on a compact set

$[\alpha, \beta]$ , the convergence is uniform.

Next we claim that for  $j$  large enough the changes in the first derivative of  $\phi^j$  in the *neighboring segment* to  $\psi([t_{i-1}, t_i])$  is at least almost as large as the change in the derivative of  $\psi$  itself over  $[t_{i-1}, t_i]$ , namely is bounded below by  $\approx \|\dot{\psi}(t_i) - \dot{\psi}(t_{i+1})\| = \gamma$ . If this is true, then since  $\sum_{i=1}^N \|\dot{\phi}^j(t_i) - \dot{\phi}^j(t_{i-1})\|$  gives a lower bound on the total curvature of  $\phi^j$ , the argument would be complete. To show that the former is indeed the case, consider  $\delta < 1/2 \max_{i=1, \dots, N} (t_i - t_{i-1})$  and small enough so that  $\dot{\psi}(t_i) \approx \frac{\psi(t_i + \delta/2) - \psi(t_i - \delta/2)}{\delta}$  where the approximation error is quadratic in  $\delta$  and uniform for all  $i$  (again, this may be achieved since  $\psi$  is  $C^2$  over a compact). Since the convergence  $\phi^j(t) \rightarrow \psi(t)$  is uniform, for any  $r > 0$  there exists  $J$  so that for all  $j \geq J$  we have  $|\phi^j(t) - \psi(t)| < r$  for all  $t \in [\alpha, \beta]$ . Pick  $J$  so that  $r \ll \delta \sin \gamma$ . Consider two cylindrical tubes with rounded bases around segments  $[\psi(t_{i-1} - \delta/2), \psi(t_{i-1} + \delta/2)]$  and  $[\psi(t_i - \delta/2), \psi(t_i + \delta/2)]$  – each tube is a union of a cylinder of height  $\approx \delta$  and radius  $r$  with base centers at  $\psi(t_k - \delta/2)$  and  $\psi(t_k + \delta/2)$ ,  $k = i-1$  and  $i$  respectively, and two  $r$ -balls with the same centers. Align the coordinate system so that the first coordinate is collinear with the vector  $(\psi(t_{i-1} - \delta/2), \psi(t_{i-1} + \delta/2))$  and the second coordinate chosen so that together with the first coordinate axis it spans a hyperplane parallel to  $[\psi(t_i - \delta/2), \psi(t_i + \delta/2)]$ , with the first two coordinates of  $\psi(t_i)$  positive, see Figure 2(b). The following two inequalities should be interpreted as true up to higher order error terms. Note that by the Intermediate-Value Theorem there exists  $\underline{\tau}$  such that  $\left| \left( \dot{\phi}_{\text{arc}}^j(\underline{\tau}) \right)_2 \right| \leq \frac{2r}{\delta - 2r}$ : considering the tube around  $[\psi(t_{i-1} - \delta/2), \psi(t_{i-1} + \delta/2)]$  note that for  $\dot{\phi}_{\text{arc}}^j$  to traverse the tube will take at least  $\delta - 2r$  change of the arc length parameter, while at the same time its second coordinate will change by at most  $2r$ . Similarly, considering the second tube we conclude that there exists  $\bar{\tau}$  such that  $\left| \left( \dot{\phi}_{\text{arc}}^j(\bar{\tau}) \right)_2 \right| \geq \frac{\delta \sin \gamma - 2r}{\delta + 2r}$ . Clearly, with  $r$  small enough, or equivalently,  $j$  large enough,  $\|\dot{\phi}_{\text{arc}}^j(t_i) - \dot{\phi}_{\text{arc}}^j(t_{i-1})\|$  is at least  $\gamma$  up to the higher order terms, since we consider the shortest distance between two points  $\dot{\phi}_{\text{arc}}^j(\underline{\tau})$  and  $\dot{\phi}_{\text{arc}}^j(\bar{\tau})$  on the  $n$ -unit sphere, see Figure 2(c).

**Acknowledgments** Research supported by an NSERC Discovery grant, by a MITACS grant and by the Canada Research Chair program.

## References

- [1] G. Dantzig: *Linear Programming and Extensions*. Princeton University Press (1963).
- [2] J.-P. Dedieu, G. Malajovich and M. Shub: On the curvature of the central path of linear programming theory. *Foundations of Computational Mathematics* 5 (2005) 145–171.
- [3] A. Deza, T. Terlaky and Y. Zinchenko: Central path curvature and iteration-complexity for redundant Klee-Minty cubes. In: D. Gao and H. Sherali (eds.) *Advances in Mechanics and Mathematics* III (2007) 215–248.
- [4] A. Deza, T. Terlaky and Y. Zinchenko: Polytopes and Arrangements: Diameter and Curvature. *Operations Research Letters* 36-2 (2008) 215–222.
- [5] K. Fritzsche and F. Holt: More polytopes meeting the conjectured Hirsch bound. *Discrete Mathematics* 205 (1999) 77–84.
- [6] F. Holt and V. Klee: Many polytopes meeting the conjectured Hirsch bound. *Discrete and Computational Geometry* 20 (1998) 1–17.
- [7] B. Grünbaum: *Convex Polytopes*. V. Kaibel, V. Klee and G. Ziegler (eds.), Graduate Texts in Mathematics 221, Springer-Verlag (2003).
- [8] G. Kalai and D. Kleitman: A quasi-polynomial bound for the diameter of graphs of polyhedra. *Bulletin of the American Mathematical Society* 26 (1992) 315–316.
- [9] V. Klee and D. Walkup: The  $d$ -step conjecture for polyhedra of dimension  $d < 6$ . *Acta Mathematica* 133 (1967) 53–78.

- [10] J. Renegar: *A Mathematical View of Interior-Point Methods in Convex Optimization*. SIAM (2001).
- [11] C. Roos, T. Terlaky and J.-Ph. Vial: *Interior Point Methods for Linear Optimization*. Springer (2006).
- [12] G. Sonnevend, J. Stoer and G. Zhao: On the complexity of following the central path of linear programs by linear extrapolation II. *Mathematical Programming* 52 (1-3) (1991) 527–553.
- [13] M. Spivak: *Comprehensive Introduction to Differential Geometry*. Publish or Perish (1990).
- [14] G. Ziegler: *Lectures on Polytopes*. Graduate Texts in Mathematics 152, Springer-Verlag (1995).

Antoine Deza, Tamás Terlaky, Yuriy Zinchenko

ADVANCED OPTIMIZATION LABORATORY, DEPARTMENT OF COMPUTING AND SOFTWARE,  
SCHOOL OF COMPUTATIONAL ENGINEERING AND SCIENCE,  
MCMASTER UNIVERSITY, HAMILTON, ONTARIO, CANADA.

*Email:* deza, terlaky, zinchen@mcmaster.ca

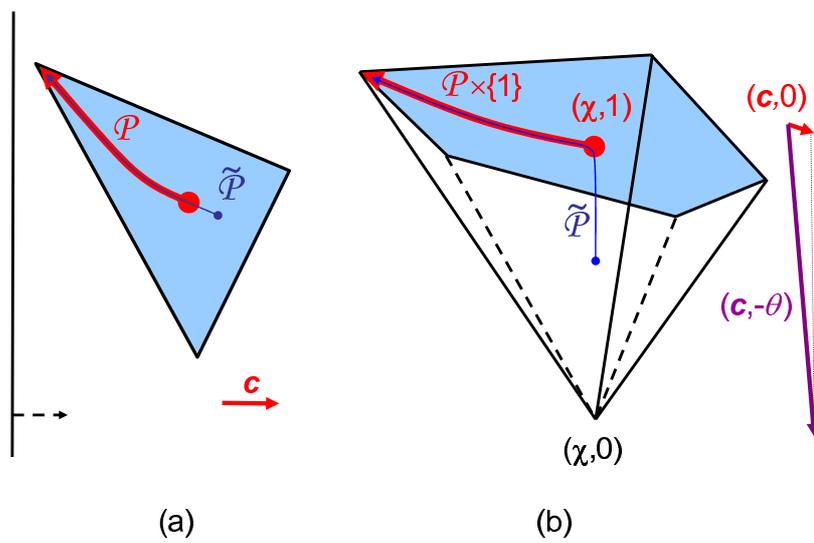


Figure 1: An illustration for the proof of Theorem 2.1 for  $d = 2$ : (a)  $2d > n = 3$ , (b)  $2d < n = 5$

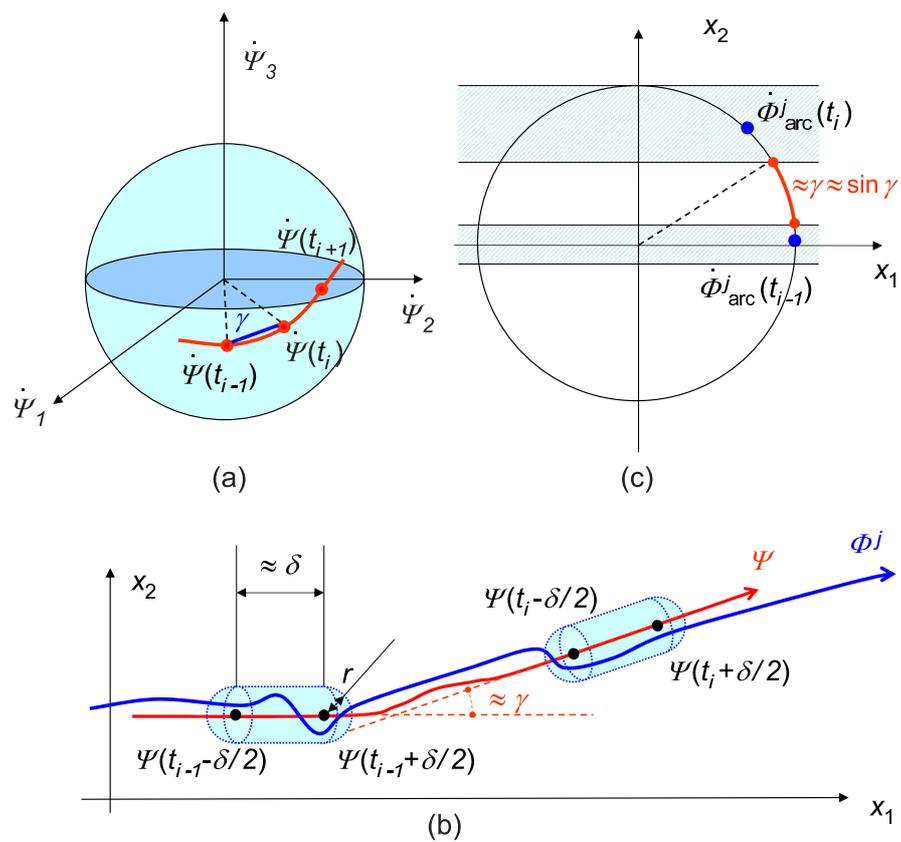


Figure 2: Illustration of the proof of Proposition 2.2