Vietnam Journal of Mathematics
https://doi.org/10.1007/s10013-025-00742-1

ORIGINAL ARTICLE

Small Kissing Polytopes

Antoine Deza[1] · Zhongyuan Liu[1] · Lionel Pournin[2]

Received: 7 May 2024 / Accepted: 13 January 2025

Abstract
A lattice (d, k)-polytope is the convex hull of a set of points in $\mathbb{R}^d$ whose coordinates are integers ranging between 0 and k. We consider the smallest possible distance $\varepsilon(d, k)$ between two disjoint lattice (d, k)-polytopes. We propose an algebraic model for this distance and derive from it an explicit formula for $\varepsilon(2, k)$. Our model also allows for the computation of previously intractable values of $\varepsilon(d, k)$. In particular, we compute $\varepsilon(3, k)$ when $4 \leq k \leq 8$, $\varepsilon(4, k)$ when $2 \leq k \leq 3$, and $\varepsilon(6, 1)$.

Keywords Lattice polytopes · Facial distance · Vertex-facet distance

Mathematics Subject Classification (2010) 52C07 · 68W30 · 90C10

1 Introduction

The smallest possible distance between two disjoint lattice (d, k)-polytopes—convex hulls of sets of points with integer coordinates in $[0, k]^d$—is a natural quantity in discrete geometry. This quantity, which we refer to as $\varepsilon(d, k)$ in the sequel, is connected to the complexity of algorithms such as the linear minimization formulation by Gábor Braun, Sebastian Pokutta, and Robert Weismantel [3] of the von Neumann alternating projections algorithm [8]. It is also related to several notions that appear in optimization. For instance, the *facial distance* of a polytope P, studied by Javier Peña and Daniel Rodriguez [9] and by David Gutman and Javier Peña [6, 10], is the smallest possible distance between a face F of P and the convex hull of the vertices of P that are not contained in F. The *vertex-facet distance* of a polytope P, considered by Amir Beck and Shimrit Shtern [2], is the smallest possible distance between the affine hull of a facet F of P and a vertex of P that does not belong to F. The smallest possible

Dedicated to Tamás Terlaky on the occasion of his 70th birthday.

✉ Lionel Pournin
lionel.pournin@univ-paris13.fr

Antoine Deza
deza@mcmaster.ca

Zhongyuan Liu
liu164@mcmaster.ca

1 McMaster University, Hamilton, Ontario, Canada

2 LIPN, Université Paris 13, Villetaneuse, France

vertex-facet distance of a lattice $(d, 1)$-simplex has been estimated by Noga Alon and Văn Vũ [1]. Another such notion is the *pyramidal width* of a finite set of points, investigated by Simon Lacoste-Julien and Martin Jaggi [7] and by Luis Rademacher and Chang Shu [11], which coincides with the facial distance of the convex hull of these points [9]. Gábor Braun, Alejandro Carderera, Cyrille Combettes, Hamed Hassani, Amin Karbasi, Aryan Mokhtari, and Sebastian Pokutta provide a comprehensive overview of these notions in [4]. Lower and upper bounds on $\varepsilon(d, k)$ that are almost matching as d goes to infinity and a number of properties of this quantity as a function of d and k have been established by Shmuel Onn, Sebastian Pokutta, and two of the authors in [5]. The values of $\varepsilon(2, k)$ when $1 \leq k \leq 6$, of $\varepsilon(3, k)$ when $1 \leq k \leq 3$, of $\varepsilon(4, 1)$, and of $\varepsilon(5, 1)$ have been computed as a consequence of these properties and reported in [5]. These values are the non-bolded entries shown in Table 1. Building on the results of [5], we develop an algebraic model that allows for the computation of previously intractable values of $\varepsilon(d, k)$ in Section 2. More precisely, $\varepsilon(d, k)$ is bounded by the smallest non-zero value of a certain algebraic fraction over a subset of the lattice points contained in the hypercube $[-k, k]^{d^2}$. Using this model, we provide the following formula for $\varepsilon(2, k)$ in Section 3.

Theorem 1 *If k is greater than* 1, *then*

$$\varepsilon(2, k) = \frac{1}{\sqrt{(k-1)^2 + k^2}}.$$

We further show in Section 4 how the subset of the lattice points in the hypercube $[-k, k]^{d^2}$ over which the minimization is performed can be reduced, and discuss the computational efficiency of the resulting strategy. This makes it possible to determine values of $\varepsilon(d, k)$ whose computation was previously intractable. Using this strategy, we compute $\varepsilon(3, k)$ when $4 \leq k \leq 8$, $\varepsilon(4, k)$ when $2 \leq k \leq 3$, and $\varepsilon(6, 1)$. These values of $\varepsilon(d, k)$ are the inverse of the numbers shown in bold in Table 1. For each of the obtained values of $\varepsilon(d, k)$, we provide an explicit pair of lattice (d, k)-polytopes whose distance is precisely $\varepsilon(d, k)$. We shall refer to such a pair of polytopes as *kissing polytopes*.

2 A Least Squares Model for Polytope Distance

Let us consider two disjoint lattice (d, k)-simplices P and Q whose affine hulls are disjoint. Denote by p^0 to p^n the vertices of P and by q^0 to q^m the vertices of Q, where n and m denote the dimension of P and Q, respectively.

Table 1 The known values of $1/\varepsilon(d, k)$

d	k							
	1	2	3	4	5	6	7	8
2	$\sqrt{2}$	$\sqrt{5}$	$\sqrt{13}$	5	$\sqrt{41}$	$\sqrt{61}$	$\mathbf{\sqrt{85}}$	$\mathbf{\sqrt{113}}$
3	$\sqrt{6}$	$5\sqrt{2}$	$\sqrt{299}$	$\mathbf{5\sqrt{42}}$	$\mathbf{\sqrt{2870}}$	$\mathbf{\sqrt{6466}}$	$\mathbf{5\sqrt{510}}$	$\mathbf{\sqrt{22826}}$
4	$3\sqrt{2}$	$\mathbf{2\sqrt{113}}$	$\mathbf{11\sqrt{71}}$					
5	$\sqrt{58}$							
6	$\mathbf{\sqrt{202}}$							

The distance of P and Q is the smallest possible value of

$$\left\|\sum_{i=0}^{n}\lambda_i p^i-\sum_{i=0}^{m}\mu_i q^i\right\|^2, \tag{1}$$

where λ_0 to λ_n and μ_0 to μ_m are two sets of non-negative numbers that each sum to 1. The constraint that each of these sets of numbers sum to 1 can be avoided by first expressing λ_0 and μ_0 as a function of the other numbers as

$$\begin{cases}\lambda_0=1-\displaystyle\sum_{i=1}^{n}\lambda_i\\ \mu_0=1-\displaystyle\sum_{i=1}^{m}\mu_i\end{cases} \tag{2}$$

and replacing them in (1) by these expressions. As a consequence,

$$d(P,Q)^2=\min_{\substack{\lambda\in\Delta_n\\ \mu\in\Delta_m}} f_{P,Q}(\lambda,\mu),$$

where

$$f_{P,Q}(\lambda,\mu)=\left\|p^0-q^0+\sum_{i=1}^{n}\lambda_i\left(p^i-p^0\right)-\sum_{i=1}^{m}\mu_i\left(q^i-q^0\right)\right\|^2 \tag{3}$$

and Δ_j denotes the j-dimensional simplex

$$\Delta_j=\left\{x\in[0,+\infty[^j:\sum_{i=1}^{j}x_i\leq 1\right\}.$$

We will consider $f_{P,Q}$ as a function from $\mathbb{R}^n\times\mathbb{R}^m$ to $[0,+\infty[$. Note that this function depends on the ordering of the vertices of P and Q but this ordering will not play a role in the sequel, and we assume that a prescribed ordering has been fixed for the vertices of each pair of polytopes P and Q. Relaxing the constraint that λ and μ should be contained in Δ_n and Δ_m provides a lower bound on $d(P,Q)$ of the form

$$d(P,Q)^2\geq\min_{\substack{\lambda\in\mathbb{R}^n\\ \mu\in\mathbb{R}^m}} f_{P,Q}(\lambda,\mu). \tag{4}$$

Note that the right-hand side of (4) is the distance between the affine hull of P and the affine hull of Q. In particular, the accuracy of this bound is related to how close the distance of P and Q is to the distance of their affine hulls.

Now consider the $d\times(d-1)$ matrix

$$A=\begin{bmatrix}p_1^1-p_1^0 & \cdots & p_1^n-p_1^0 & q_1^1-q_1^0 & \cdots & q_1^m-q_1^0\\ \vdots & & \vdots & \vdots & & \vdots\\ p_d^1-p_d^0 & \cdots & p_d^n-p_d^0 & q_d^1-q_d^0 & \cdots & q_d^m-q_d^0\end{bmatrix} \tag{5}$$

and the vector

$$b=q^0-p^0. \tag{6}$$

It will be important to keep in mind that A and b depend on P and Q. Observe that, with these notations, (3) can be rewritten into

$$f_{P,Q}(\lambda,\mu)=\|A\chi-b\|^2, \tag{7}$$

where χ is the vector such that

$$\chi^t = [\lambda_1, \ldots, \lambda_n, -\mu_1, \ldots, -\mu_m]. \tag{8}$$

Remark 1 According to (7), $f_{P,Q}(\lambda, \mu)$ is the sum of the squares of the coordinates of the vector $A\chi - b$. In particular, negating both a row of A and the corresponding coefficient of b will not change the value of that function. Likewise, negating a column of A and the corresponding row of χ will not change the value of $f_{P,Q}(\lambda, \mu)$. As a consequence, if one computes $f_{P,Q}(\lambda, \mu)$ via (7), then the right-hand side of (4) does not change when a subset of the columns of A are negated or a subset of its rows are negated together with the corresponding coefficients of b.

Let us now give an expression for the right-hand side of (4).

Lemma 2 *The function $f_{P,Q}$ admits a unique minimum over $\mathbb{R}^n \times \mathbb{R}^m$ if and only if A^tA is non-singular. Moreover, in that case,*

$$\min_{\substack{\lambda \in \mathbb{R}^n \\ \mu \in \mathbb{R}^m}} f_{P,Q}(\lambda, \mu) = \left\| A(A^tA)^{-1}A^tb - b \right\|^2. \tag{9}$$

Proof The minimum of $f_{P,Q}$ is reached at a pair (λ, μ) from $\mathbb{R}^n \times \mathbb{R}^m$ such that all the partial derivatives of $f_{P,Q}$ simultaneously vanish, that is when

$$\frac{\partial f_{P,Q}}{\partial \lambda_i}(\lambda, \mu) = 0$$

for all i satisfying $1 \leq i \leq n$ and

$$\frac{\partial f_{P,Q}}{\partial \mu_i}(\lambda, \mu) = 0$$

for all i satisfying $1 \leq i \leq m$. Since $f_{P,Q}$ is a quadratic function of λ and μ, its partial derivatives are linear. In other words, finding the minimum of $f_{P,Q}$ over $\mathbb{R}^n \times \mathbb{R}^m$ amounts to solve a least squares problem. In particular setting to 0 all of these partial derivatives results in the system of linear equalities

$$A^tA\chi = A^tb. \tag{10}$$

Since $f_{P,Q}$ is a convex quadratic function, the solutions of (10) correspond bijectively via (8) with the pairs (λ, μ) such that $f_{P,Q}$ is minimal. It immediately follows that $f_{P,Q}$ admits a unique minimum over $\mathbb{R}^n \times \mathbb{R}^m$ if and only if A^tA is non-singular. Moreover, in that case, the unique solution of (10) is

$$\chi = (A^tA)^{-1}A^tb$$

and substituting this expression of χ in (7) completes the proof. □

According to (4), Lemma 2 provides a lower bound on the distance between P and Q in the case when A^tA is non-singular. The following remark provides a necessary and sufficient condition on $(A^tA)^{-1}A^tb$ for this bound to be sharp.

Remark 2 Recall that the minimum of $f_{P,Q}$ over $\mathbb{R}^n \times \mathbb{R}^m$ is the distance between the affine hulls of P and Q. Therefore, if A^tA is non-singular, then according to (7), (8), and Lemma 2, the first n coordinates of the vector

$$\chi = (A^tA)^{-1}A^tb$$

provide an affine combination $p^\star$ of the vertices of P and its last m coordinates an affine combination $q^\star$ of the vertices of Q such that

$$d(p^\star, q^\star) = d\left(\mathrm{aff}(P), \mathrm{aff}(Q)\right).$$

In particular, if the first n coefficients of the vector χ are all non-negative and sum to at most 1 while its last m coefficients are all non-positive and sum to at least -1, then according to (2), $p^\star$ is contained in P and $q^\star$ in Q. In that case, the distance of P and Q coincides with the distance of their affine hulls. Otherwise the distance of P and Q is strictly greater than the distance of their affine hulls.

We shall now focus on certain pairs of simplices whose distance is precisely $\varepsilon(d, k)$. The following is proven in [5] (see Theorem 5.2 therein).

Theorem 3 *There exist two lattice (d, k)-polytopes P and Q such that*

(i) *$d(P, Q)$ is equal to $\varepsilon(d, k)$,*
(ii) *both P and Q are simplices,*
(iii) $\dim(P) + \dim(Q)$ *is equal to $d - 1$, and*
(iv) *the affine hulls of P and Q are disjoint.*

We shall prove that when P and Q satisfy the assertions (i) to (iv) in the statement of Theorem 3, $f_{P,Q}$ admits a unique minimum over $\mathbb{R}^n \times \mathbb{R}^m$ as a consequence of two results from [5]. The first of these results states that

$$d(P, Q) \geq \varepsilon(\dim(P \cup Q), k) \tag{11}$$

(see Lemma 4.3 in [5]) and the second that, when k is fixed, $\varepsilon(d, k)$ is a strictly decreasing function of d (see Theorem 5.1 in [5]).

Proposition 4 *If P and Q satisfy the assertions (i) to (iv) in the statement of Theorem 3, then $f_{P,Q}$ has a unique minimum over $\mathbb{R}^n \times \mathbb{R}^m$.*

Proof Denote by $\mathrm{aff}(P)$ and $\mathrm{aff}(Q)$ the affine hulls of P and Q, respectively. Consider a point $p^\star$ in $\mathrm{aff}(P)$ and a point $q^\star$ in $\mathrm{aff}(Q)$ such that

$$\|q^\star - p^\star\| = d\big(\mathrm{aff}(P), \mathrm{aff}(Q)\big). \tag{12}$$

According to the above discussion, the pair $(p^\star, q^\star)$ corresponds to a point $(\lambda^\star, \mu^\star)$ in $\mathbb{R}^n \times \mathbb{R}^m$ at which the function $f_{P,Q}$ reaches its minimum. Assume that the function $f_{P,Q}$ also reaches its minimum at a point $(\overline{\lambda}^\star, \overline{\mu}^\star)$ in $\mathbb{R}^n \times \mathbb{R}^m$ different from $(\lambda^\star, \mu^\star)$. By construction, $(\overline{\lambda}^\star, \overline{\mu}^\star)$ then provides the coefficients of an affine combination $\overline{p}^\star$ of p^0 to p^n and of an affine combination $\overline{q}^\star$ of q^0 to q^m such that

$$\|\overline{q}^\star - \overline{p}^\star\| = d\big(\mathrm{aff}(P), \mathrm{aff}(Q)\big). \tag{13}$$

Since $p^\star$ and $\overline{p}^\star$ are both contained in $\mathrm{aff}(P)$, so is their midpoint. Likewise, the midpoint of $q^\star$ and $\overline{q}^\star$ is contained in $\mathrm{aff}(Q)$. Hence,

$$d\big(\mathrm{aff}(P), \mathrm{aff}(Q)\big) \leq \left\| \frac{q^\star + \overline{q}^\star}{2} - \frac{p^\star + \overline{p}^\star}{2} \right\|.$$

However, by the triangle inequality,

$$\left\| \frac{q^\star + \overline{q}^\star}{2} - \frac{p^\star + \overline{p}^\star}{2} \right\| \leq \frac{1}{2}\|q^\star - p^\star\| + \frac{1}{2}\|\overline{q}^\star - \overline{p}^\star\|$$

with equality if and only if $q^\star - p^\star$ is a multiple of $\overline{q}^\star - \overline{p}^\star$ by a positive coefficient or one of these vectors is equal to 0. Under the assumption that assertion (iv) from the statement of Theorem 3 holds, these vectors are both non-zero. Hence, $q^\star - p^\star$ is a multiple of $\overline{q}^\star - \overline{p}^\star$ by a positive coefficient and according to (12) and (13), these vectors must therefore be equal. It immediately follows that the vectors $\overline{p}^\star - p^\star$ and $\overline{q}^\star - q^\star$ also coincide.

Now recall that $(\lambda^\star, \mu^\star)$ and $(\overline{\lambda}^\star, \overline{\mu}^\star)$ are different points. As a consequence, so are the pairs $(p^\star, q^\star)$ and $(\overline{p}^\star, \overline{q}^\star)$. Since the vectors $\overline{p}^\star - p^\star$ and $\overline{q}^\star - q^\star$ coincide they must therefore be non-zero. Hence, the translates of $\mathrm{aff}(P)$ and of $\mathrm{aff}(Q)$ through the origin of $\mathbb{R}^d$ intersect in a non-zero vector. Therefore,

$$\dim(P \cup Q) \leq \dim(P) + \dim(Q).$$

Under the assumption that P and Q satisfy the assertion (iii) in the statement of Theorem 3, it follows that $P \cup Q$ has dimension at most $d-1$. By (11), this implies that the distance between P and Q is at least $\varepsilon(d-1, k)$. Hence, if the assertion (i) in the statement of Theorem 3 holds for P and Q, then one obtains that

$$\varepsilon(d, k) \geq \varepsilon(d-1, k).$$

However, Theorem 5.1 in [5] states that $\varepsilon(d, k)$ is less than $\varepsilon(d-1, k)$. By this contradiction, $f_{P,Q}$ has a unique minimum over $\mathbb{R}^n \times \mathbb{R}^m$. □

Combining Proposition 4, Lemma 2, and Theorem 3, one obtains a lower bound on $\varepsilon(d, k)$ from (4) of the form

$$\varepsilon(d, k) \geq \min_{P,Q} \left\{ \left\| A(A^tA)^{-1}A^tb - b \right\| \right\}, \tag{14}$$

where the minimum ranges over the pairs of lattice (d, k)-simplices P and Q whose dimensions sum to $d-1$, for which the matrix A obtained from (5) is such that A^tA is non-singular and the vector b obtained from (6) satisfies

$$A(A^tA)^{-1}A^tb \neq b.$$

3 The 2-dimensional Case

In this section, we give a formula for $\varepsilon(2, k)$ using the model described in Section 1. Consider two disjoint lattice $(2, k)$-polytopes P and Q that satisfy assertions (i) to (iv) from the statement of Theorem 3. Since the dimensions of P and Q sum to 1, one of these polytopes has dimension 0 and the other has dimension 1. We assume that P is a line segment and that Q is made of a single point by exchanging these two polytopes if needed.

Let us first observe that according to (5) and (6),

$$\begin{cases} A = p^1 - p^0, \\ b = q^0 - p^0. \end{cases}$$

As a consequence, (14) simplifies into

$$\varepsilon(2, k) \geq \frac{|(p_2^1 - p_2^0)(q_1^0 - p_1^0) - (p_1^1 - p_1^0)(q_2^0 - p_2^0)|}{\sqrt{(p_1^1 - p_1^0)^2 + (p_2^1 - p_2^0)^2}}.$$

It follows that $\varepsilon(2, k)$ is at least the smallest possible value of

$$\frac{|x_2x_3 - x_1x_4|}{\sqrt{x_1^2 + x_2^2}} \tag{15}$$

over all the lattice points x contained in the hypercube $[-k, k]^4$ such that neither $x_1^2 + x_2^2$ nor $x_2x_3 - x_1x_4$ is equal to 0. We bound (15) as follows.

Lemma 5 *If k is greater than 1 then, for every lattice point x in the hypercube $[-k, k]^4$ such that $x_1^2 + x_2^2$ and $x_2x_3 - x_1x_4$ are both non-zero,*

$$\frac{|x_2x_3 - x_1x_4|}{\sqrt{x_1^2 + x_2^2}} \geq \frac{1}{\sqrt{(k-1)^2 + k^2}}. \tag{16}$$

Proof Consider a lattice point x in $[-k, k]^4$ such that neither $x_1^2 + x_2^2$ nor $x_2x_3 - x_1x_4$ is equal to 0. We assume without loss of generality that x_1 and x_2 are non-negative thanks to the symmetries of $[-k, k]^4$. We review two cases.

First assume that x_1 and x_2 coincide. In that case,

$$\frac{|x_2x_3 - x_1x_4|}{\sqrt{x_1^2 + x_2^2}} = \frac{|x_3 - x_4|}{\sqrt{2}}. \tag{17}$$

Since $x_2x_3 - x_1x_4$ is not equal to 0, x_3 and x_4 cannot coincide and the right-hand side of (17) is at least $1/\sqrt{2}$. As k is greater than 1,

$$\frac{1}{\sqrt{(k-1)^2 + k^2}} \leq \frac{1}{\sqrt{2}}$$

and the lemma follows in this case.

Now assume that x_1 and x_2 are different. Since $|x_2x_3 - x_1x_4|$ is at least 1,

$$\frac{|x_2x_3 - x_1x_4|}{\sqrt{x_1^2 + x_2^2}} \geq \frac{1}{\sqrt{x_1^2 + x_2^2}}. \tag{18}$$

Recall that x_1 and x_2 are integers contained in $[0, k]$. Since they are different, one of them is at most $k - 1$. As a consequence,

$$\frac{1}{\sqrt{x_1^2 + x_2^2}} \geq \frac{1}{\sqrt{(k-1)^2 + k^2}}$$

and combining this with (18) completes the proof. □

By Lemma 5, and the preceding discussion, $\varepsilon(2, k)$ is at least the right-hand side of (16) when k is at least 2. Theorem 1 states that this is sharp.

Proof of Theorem 1 It suffices to exhibit a lattice point P and a lattice segment Q, both contained in the square $[0, k]^2$ satisfying

$$d(P, Q) = \frac{1}{\sqrt{(k-1)^2 + k^2}}.$$

Such an example is obtained by taking for P the lattice point whose two coordinates are equal to 1 and for Q any of the two line segments that are incident to the origin and whose other vertex has coordinates k and $k - 1$. This point and one of these line segments are represented in Fig. 1 when k is equal to 4. □

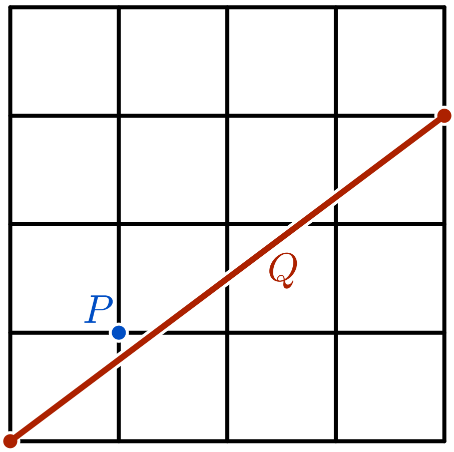

Fig. 1 A pair of kissing lattice (2, 4)-polytopes

Remark 3 The strategy exposed in this section in the 2-dimensional case can be generalized to any higher dimension. In particular, a quotient similar to (15) can be explicitly computed for any fixed dimension d that depends on a lattice point x contained in the hypercube $[-k, k]^{d^2}$. The first $d(d-1)$ coordinates of the point x are the entries of A and its last d coordinates are the coordinates of b. The minimum of that quotient under the constraint that its numerator and denominator are positive provides a lower bound on $\varepsilon(d, k)$. For instance, when d is equal to 3, the minimal value of the ratio

$$\frac{|x_1(x_6x_8 - x_5x_9) + x_2(x_4x_9 - x_6x_7) + x_3(x_5x_7 - x_4x_8)|}{\sqrt{(x_1x_5 - x_2x_4)^2 + (x_1x_6 - x_3x_4)^2 + (x_2x_6 - x_3x_5)^2}} \tag{19}$$

over all the lattice points x in the hypercube $[-k, k]^9$ such that the numerator and the denominator of (19) are positive is a lower bound on $\varepsilon(3, k)$. However, the expression for this quotient gets complicated as the dimension increases and solving the corresponding integer minimization problem becomes involved.

4 The Computation of $\varepsilon(d, k)$

According to the discussion in Section 1, a lower bound on $\varepsilon(d, k)$ can be obtained by considering all the sets of $d+1$ pairwise distinct points from $\{0, \dots, k\}^d$ and for each such set $\mathcal{S}$, all the partitions of $\mathcal{S}$ into two subsets $\{p^0, \dots, p^n\}$ and $\{q^0, \dots, q^m\}$, where $n+m$ is equal to $d-1$. For each such partition, one can build a matrix A and a vector b according to (5) and (6). The smallest possible non-zero value of the right-hand side of (14) over all the obtained pairs (A, b) such that A^tA is non-singular will then be a lower bound on $\varepsilon(d, k)$. However, this strategy requires to consider

$$N = (2^{d+1} - 2)\binom{(k+1)^d}{d+1}$$

pairs (A, b). Note that while this number would decrease to at best

$$\frac{2^{d+1} - 2}{2^d d!}\binom{(k+1)^d}{d+1}$$

if the enumeration could be performed up to the symmetries of the d-dimensional hypercube. However, these symmetries are not all easy to handle in practice as one still needs to enumerate all N pairs (A, b) just to check for them.

We adopt a different strategy in order to significantly decrease the search space without having to handle symmetries. The main idea is to do the enumeration coordinate-wise in order to build a list $\mathcal{L}$ of the possible rows for the pair (A, b) for each n and m that sum to $d-1$ and such that $n \leq m$, and then building (A, b) back by selecting d pairwise different rows from $\mathcal{L}$. By a *row of* (A, b), we mean a vector r from $\mathbb{R}^d$ whose first $d-1$ entries form a row of A and whose last entry is the corresponding coordinate of b. Note that our requirement that the rows of $\mathcal{L}$ selected to build a given pair (A, b) are pairwise distinct is without loss of generality. Indeed, if two of these rows would coincide, a pair of columns of $A^t A$ would be multiples of one another and that matrix would then be singular. We shall see that the size of $\mathcal{L}$ does not depend on n or m.

As a consequence, this alternative strategy only considers

$$\left\lfloor \frac{d+1}{2} \right\rfloor \binom{|\mathcal{L}|}{d}$$

pairs (A, b). For each of these pairs such that $A^t A$ is non-singular, the right-hand side of (14) is evaluated, and the smallest non-zero value obtained for this quantity over all the considered pairs (A, b) is the desired lower bound on $\varepsilon(d, k)$. It should be noted that the efficiency of this strategy depends on how large $\mathcal{L}$ is.

Let us get into more details about how we build $\mathcal{L}$. For a given pair of positive integers n and m that sum to $d-1$, we generate all the possible rows of the pair (A, b) as

$$(x_1 - x_0, \ldots, x_n - x_0, y_1 - y_0, \ldots, y_m - y_0, y_0 - x_0), \tag{20}$$

where x is a point from $\{0, \ldots, k\}^n$, y is a point from $\{0, \ldots, k\}^m$ such that x and y are not both equal to 0, and x_0 and y_0 are two integers from $\{0, \ldots, k\}$.

The list obtained from this procedure contains at most $(k-1)^{d+1}$ rows. Its size can be reduced using the following property.

Proposition 6 *Consider a $d \times (d-1)$ matrix A with integer entries such that $A^t A$ is non-singular. Futher consider a vector b contained in $\mathbb{Z}^d$. If the pair $(\overline{A}, \overline{b})$ is obtained by dividing each row of the pair (A, b) by the greatest common divisor of its coordinates and by negating a subset of the resulting rows, then*

$$\left\| A(A^t A)^{-1} A^t b - b \right\| \geq \left\| \overline{A}(\overline{A}^t \overline{A})^{-1} \overline{A}^t \overline{b} - \overline{b} \right\| \tag{21}$$

and the two sides of this inequality are either both zero or both positive.

Proof Pick two non-negative integers n and m that sum to $d-1$. Denote b by q^0 and the origin of $\mathbb{R}^d$ by p^0. Further denote by p^1 to p^n the first n columns of A and consider the points q^1 to q^m from $\mathbb{Z}^d$ such that $q^1 - q^0$ to $q^m - q^0$ are the last m columns of A. According to the construction described in Section 1,

$$\left\| A(A^t A)^{-1} A^t b - b \right\|^2 = \min_{\substack{\lambda \in \mathbb{R}^n \\ \mu \in \mathbb{R}^m}} f_{P,Q}(\lambda, \mu), \tag{22}$$

where P is the convex hull of p^0 to p^n and Q that of q^0 to q^m.

Further denote by $(\overline{A}, \overline{b})$ the pair obtained by dividing each row of (A, b) by the greatest common divisor of its coordinates and by negating a fixed (but otherwise arbitrary) subset of the resulting rows. As above,

$$\left\| \overline{A}(\overline{A}^t\overline{A})^{-1}\overline{A}^t\overline{b} - \overline{b} \right\|^2 = \min_{\substack{\lambda \in \mathbb{R}^n \\ \mu \in \mathbb{R}^m}} f_{\overline{P},\overline{Q}}(\lambda, \mu), \tag{23}$$

where $\overline{P}$ and $\overline{Q}$ are the convex hulls of the points $\overline{p}^0$ to $\overline{p}^n$ and $\overline{q}^0$ to $\overline{q}^m$ that are extracted from $(\overline{A}, \overline{b})$ just as the points p^i and q^i are extracted from (A, b).

By construction, for any pair (λ, μ) of vectors in $\mathbb{R}^n \times \mathbb{R}^m$,

$$\begin{aligned} &\sum_{j=1}^{d} \left(-q_j^0 + \sum_{i=1}^{n} \lambda_i p_j^i - \sum_{i=1}^{m} \mu_i \left(q_j^i - q_j^0 \right) \right)^2 \\ &\quad = \sum_{j=1}^{d} r_j^2 \left(-\overline{q}_j^0 + \sum_{i=1}^{n} \lambda_i \overline{p}_j^i - \sum_{i=1}^{m} \mu_i \left(\overline{q}_j^i - \overline{q}_j^0 \right) \right)^2, \end{aligned} \tag{24}$$

where r_1 to r_d denote the greatest common divisors of the rows of (A, b). Observe that according to (3), the left-hand side of this equality is precisely $f_{P,Q}(\lambda, \mu)$. Since the numbers r_1 to r_d are not less than 1, its right-hand side is at least $f_{\overline{P},\overline{Q}}(\lambda, \mu)$ and it follows that, for every point (λ, μ) contained in $\mathbb{R}^n \times \mathbb{R}^m$,

$$f_{P,Q}(\lambda, \mu) \geq f_{\overline{P},\overline{Q}}(\lambda, \mu).$$

In turn, by (22) and (23), the desired inequality holds. It remains to show that if the right-hand side of (21) is equal to 0, then so is its left-hand side.

Assume that the right-hand side of (21) is equal to 0. In that case, there exists a pair (λ, μ) in $\mathbb{R}^n \times \mathbb{R}^m$ such that $f_{\overline{P},\overline{Q}}(\lambda, \mu)$ is equal to 0. By (3), $f_{\overline{P},\overline{Q}}(\lambda, \mu)$ is the squared norm of a vector and since it is equal to 0, all of the coordinates of that vector must be equal to 0. In other words, for every integer j satisfying $1 \leq j \leq d$,

$$-\overline{q}_j^0 + \sum_{i=1}^{n} \lambda_i \overline{p}_j^i - \sum_{i=1}^{m} \mu_i \left(\overline{q}_j^i - \overline{q}_j^0 \right) = 0$$

and it follows from (3) and (24) that $f_{P,Q}(\lambda, \mu)$ must vanish. According to (22), the left-hand side of (21) is then equal to 0, as desired. □

By Proposition 6, we can assume that when several of the generated rows are multiples of one another, only the one among them whose coordinates are relatively prime and whose first non-zero coordinate is positive is included in $\mathcal{L}$. It should be observed that, before $\mathcal{L}$ is reduced this way, its size does not depend on n or m. As announced, this property still holds once $\mathcal{L}$ has been reduced. Indeed, observe that the rows generated by (20) with the same two points x and y and the same scalars x_0 and y_0 but with different values of n and m can be recovered from one another by adding $y_0 - x_0$ to (or subtracting this quantity from) certain of their coordinates. Hence, all of these rows have the same greatest common divisor for their coordinates. We report in Table 2 as a function of d and k the number of rows contained in $\mathcal{L}$ after this procedure has been carried out. Note that, when k is equal to 1, d to 3, n to 1, and m to 1, there are only 6 rows in $\mathcal{L}$:

$$\mathcal{L} = \{(1,0,0),\ (0,1,0),\ (1,-1,0),\ (0,1,-1),\ (1,1,-1),\ (1,0,-1)\}.$$

Only twenty pairs (A, b) are generated from this list of rows.

Table 2 The number of rows in $\mathcal{L}$ as a function of d and k

d	k									
	1	2	3	4	5	6	7	8	9	10
3	6	24	72	144	288	432	720	1008	1440	1872
4	14	89	359	929	2189	4019	7469	11969		
5	30	300	1620	5400	15120					
6	62	965	6971							
7	126	3024								

When k is equal to 3 and d to 4, the number of pairs (A, b) that have to be considered with the approach outlined at the beginning of the section is

$$(2^5 - 2)\binom{4^4}{5} = 264\,286\,471\,680,$$

which would shrink down to at best

$$\frac{2^5 - 2}{2^4 4!}\binom{4^4}{5} = 688\,246\,020,$$

if these pairs could be enumerated up to the symmetries of the hypercube. However, there is no easy way to test for these symmetries without reviewing all of the 264 286 471 680 pairs. With our approach, the number of rows contained in $|\mathcal{L}|$ is equal to 359 in that case as shown in Table 2 and we only need to consider

$$2\binom{359}{4} = 1\,361\,176\,502$$

pairs (A, b) in order to compute our lower bound on $\varepsilon(4, 3)$.

This strategy does not only provide a lower bound on $\varepsilon(d, k)$ but also pairs (A, b) that achieve this lower bound. Keeping track of the points x and y and of the scalars x_0 and y_0 that are used to build each row in $\mathcal{L}$ according to (20), one can recover two lattice (d, k)-polytopes P and Q such that the obtained lower bound on $\varepsilon(d, k)$ is precisely the distance of the affine hulls of P and Q. If the distance between these affine hulls coincides with the distance between P and Q, which can easily be checked from the pair (A, b) according to Remark 2, then this lower bound is sharp. Interestingly, using this observation, all the lower bounds on $\varepsilon(d, k)$ that we have obtained using the presented strategy have turned out to be the precise value of $\varepsilon(d, k)$. In Table 1, the bolded entries correspond to the values of $\varepsilon(d, k)$ obtained in this article and the non-bolded entries are the ones computed in [5]. Theorem 1 further provides the values of $\varepsilon(2, k)$ that are not shown in the table and its proof gives a pair of kissing polytopes corresponding to these values of $\varepsilon(2, k)$. When d is at least 3, all the known values of $\varepsilon(d, k)$ are shown in the table. Let us now provide, for each of these values of $\varepsilon(d, k)$, a corresponding pair P and Q of kissing polytopes.

If k is equal to 2, 4, 5, 6, 7, or 8, then $\varepsilon(3, k)$ is achieved by the line segment P with vertices $(0, 0, 0)$ and $(k - 1, k, k)$ and the line segment Q with vertices $(k, 1, 2)$ and $(0, k, k - 1)$. These two line segments are shown in Fig. 2 in the special case when k is equal to 4. Note that for all k, the distance between these two segments is

$$d(P, Q) = \frac{1}{\sqrt{2(2k^2 - 4k + 5)(2k^2 - 2k + 1)}} \tag{25}$$

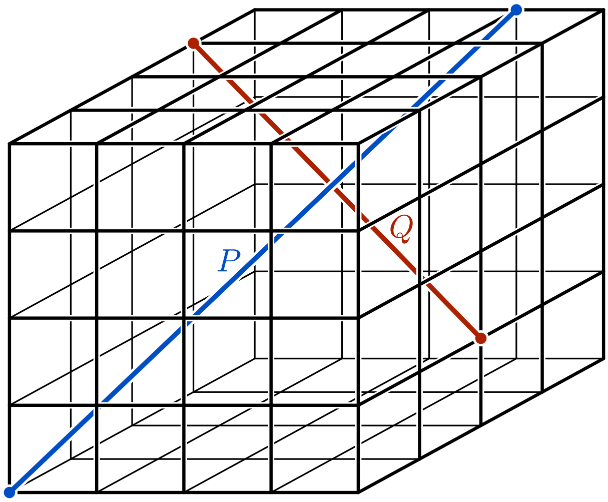

Fig. 2 A pair of kissing lattice (3, 4)-polytopes

and it is tempting to ask whether $\varepsilon(3, k)$ is equal to this value for every integer k greater than 8. In any case, note that the right-hand side of (25) provides an upper bound on $\varepsilon(3, k)$ that decreases like $1/(2\sqrt{2}k^2)$ as k goes to infinity. In the remaining two cases, when k is equal to 1 or 3, the line segments that achieve $\varepsilon(3, k)$ do not follow the pattern we have just described. Indeed, $\varepsilon(3, 1)$ is the distance between a diagonal P of the cube $[0, 1]^3$ and a diagonal Q of one of its square faces such that P and Q are disjoint (see [5]) while $\varepsilon(3, 3)$ is the distance between the line segment with vertices $(0, 0, 0)$ and $(2, 3, 3)$ and the line segment with vertices $(3, 2, 0)$ and $(0, 1, 2)$.

The values of $\varepsilon(4, k)$ reported in Table 1 are always achieved by a line segment P and a triangle Q as follows. When k is equal to 1, the vertices of P are $(0, 0, 0, 0)$ and $(1, 1, 1, 1)$ while the vertices of Q are $(1, 0, 0, 0)$, $(0, 1, 1, 0)$, and $(0, 1, 0, 1)$. When k is equal to 2, the vertices of P are $(0, 0, 0, 0)$ and $(1, 2, 1, 2)$ and those of Q are $(2, 2, 1, 0)$, $(0, 1, 0, 2)$, and $(0, 0, 2, 1)$. When k is equal to 3, the vertices of P are $(0, 0, 1, 0)$ and $(2, 3, 3, 3)$ and the vertices of Q are $(3, 0, 3, 2)$, $(0, 2, 0, 3)$, and $(0, 3, 3, 0)$.

The unique value of $\varepsilon(5, k)$ reported in Table 1 is $\varepsilon(5, 1)$. This value is the distance between the diagonal of the hypercube $[0, 1]^5$ incident to the origin of $\mathbb{R}^5$ and the tetrahedron with vertices $(1, 1, 0, 0, 0)$, $(0, 1, 0, 1, 1)$, $(0, 0, 1, 0, 1)$, and $(0, 0, 1, 1, 0)$. Finally, $\varepsilon(6, 1)$ is equal to the distance between the diagonal of the hypercube $[0, 1]^6$ incident to the origin of $\mathbb{R}^6$ and the 5-dimensional simplex with vertices $(1, 0, 1, 1, 0, 0)$, $(1, 0, 0, 0, 1, 1)$, $(0, 1, 1, 0, 1, 1)$, $(0, 1, 0, 1, 0, 1)$, and $(0, 1, 0, 1, 1, 0)$.

Remark 4 It is noteworthy that, while an expression of $\varepsilon(2, k)$ can be guessed from the pairs of kissing polytopes obtained for the first few values of k, this is not the case for $\varepsilon(d, k)$ when k is fixed. Even when k is equal to 1, the pairs of kissing polytopes known for the first few values of d do not exhibit a clear pattern.

Acknowledgements The authors thank the anonymous referees for providing a number of valuable comments and suggestions. This work was partially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant program number RGPIN-2020-06846.

References

1. Alon, N., Vũ, V.H.: Anti-Hadamard matrices, coin weighing, threshold gates, and indecomposable hypergraphs. J. Comb. Theory Ser. A **79**, 133–160 (1997)
2. Beck, A., Shtern, S.: Linearly convergent away-step conditional gradient for non-strongly convex functions. Math. Program. **164**, 1–27 (2017)
3. Braun, G., Pokutta, S., Weismantel, R.: Alternating linear minimization: revisiting von Neumann's alternating projections. arXiv:2212.02933 (2022)
4. Braun, G., Carderera, A., Combettes, C.W., Hassani, H., Karbasi, A., Mokhtari, A., Pokutta, S.: Conditional gradient methods. arXiv:2211.14103 (2022)
5. Deza, A., Onn, S., Pokutta, S., Pournin, L.: Kissing polytopes. SIAM J. Discrete Math. **38**, 2643–2664 (2024)
6. Gutman, D.H., Peña, J.F.: The condition number of a function relative to a set. Math. Program. **188**, 255–294 (2021)
7. Lacoste-Julien, S., Jaggi, M.: On the global linear convergence of Frank–Wolfe optimization variants. In: Proceedings of the 28th International Conference on Neural Information Processing Systems (NIPS), pp. 496–504 (2015)
8. von Neumann, J.: On rings of operators. Reduction theory. Ann. Math. **50**, 401–485 (1949)
9. Peña, J.F., Rodríguez, D.: Polytope conditioning and linear convergence of the Frank-Wolfe algorithm. Math. Oper. Res. **44**, 1–18 (2019)
10. Peña, J.F.: Generalized conditional subgradient and generalized mirror descent: duality, convergence, and symmetry. arXiv:1903.00459 (2019)
11. Rademacher, L., Shu, C.: The smoothed complexity of Frank-Wolfe methods via conditioning of random matrices and polytopes. Math. Stat. Learn. **5**, 273–310 (2022)