Project number IST-25582

CGL

Computational Geometric Learning

Large Shadows from Sparse Inequalities

STREP

Information Society Technologies

Period covered:	November 1, 2012–October 31, 2013
Date of preparation:	November 11, 2013
Date of revision:	November 11, 2013
Start date of project:	November 1, 2010
Duration:	3 years
Project coordinator name:	Joachim Giesen (FSU)
Project coordinator organisation:	Friedrich-Schiller-Universität Jena Jena, Germany

Large Shadows from Sparse Inequalities*

Bernd Gärtner Christian Helbling
Institute of Theoretical Computer Science
ETH Zurich
Switzerland

Yoshiki Ota Takeru Takahashi
Graduate School of Information Sciences
Tohoku University
Japan

November 11, 2013

Abstract

The d-dimensional Goldfarb cube is a polytope with the property that all its 2^d vertices appear on some *shadow* of it (projection onto a 2-dimensional plane). The Goldfarb cube is the solution set of a system of $2d$ linear inequalities with at most 3 variables per inequality. We show in this paper that the d-dimensional Klee-Minty cube — constructed from inequalities with at most 2 variables per inequality — also has a shadow with 2^d vertices. In contrast, with one variable per inequality, the size of the shadow is bounded by $2d$.

1 Introduction

The study of shadows of polytopes goes back to 1955, when Gass and Saaty introduced a variant of the simplex method for solving linear programs whose objective function linearly depends on a real parameter λ [3]. For every fixed value of λ, the problem can be treated as an ordinary linear program, but the approach of Gass and Saaty was to compute the optimal value as an explicit (piecewise linear) function of λ, and afterwards simply look up the solution for any desired parameter value λ. 50 years later, this approach was rediscovered in the machine learning community, in the context of support vector machines (parameterized quadratic programs) [6].

The Gass-Saaty method. Let us assume for the discussion here that the feasible region of the linear program is a simple polytope $\mathcal{P} \subseteq \mathbb{R}^d$ with n facets (for background on polytopes and this geometric view of linear programming,

*The project CG Learning acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number: 255827

we refer to Ziegler's book [9, Section 3.2]). We also assume that the objective function is of the form $f_\lambda(\mathbf{x}) = \mathbf{c}^T\mathbf{x} + \lambda\mathbf{d}^T\mathbf{x}$, where $\mathbf{c}, \mathbf{d} \in \mathbb{R}^d$ are linearly independent and generic (non-constant on every edge of P). Then the output of the Gass-Saaty method is a sequence of vertices $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_{M-1}$ of $\mathcal{P}$, along with a sequence of real values $-\infty = \lambda_0 < \lambda_1 < \cdots < \lambda_M = \infty$, with the following property:

$$\forall k \in \{0, 1, \ldots, M\}\text{: } \mathbf{v}_k \text{ maximizes } f_\lambda \text{ over } \mathcal{P} \text{ for } \lambda \in [\lambda_k, \lambda_{k+1}].$$

Hence, for $\lambda \in [\lambda_k, \lambda_{k+1}]$, the optimal value of the linear program is $\mathbf{c}^T\mathbf{v}_k + \lambda\mathbf{d}^T\mathbf{v}_k$, so we indeed get the optimal objective function value as a pieceweise linear function in λ, with M "bends".

The two sequences are computed as follows: by solving an ordinary linear program, we initally find the vertex $\mathbf{v}_0$ that maximizes $-\mathbf{d}^T\mathbf{x}$, corresponding to parameter value $\lambda_0 = -\infty$. Now suppose that we have already computed $\mathbf{v}_k$ and a value of λ_k for which $\mathbf{v}_k$ is optimal. Starting from λ_k, we grow λ until we have a (unique) neighboring vertex v_{k+1} with $f_\lambda(\mathbf{v}_k) = f_\lambda(\mathbf{v}_{k+1})$. The corresponding value of λ will be λ_{k+1}. If λ can grow indefinitely without reaching the former equality, we have $k = M - 1$ and set $\lambda_M = \infty$. Algebraically, the computations are very simple, if the *simplex method* is used. At value λ_k, we have a certificate of optimality of $\mathbf{v}_k$, in the form of nonpositive *reduced costs* that are also linear functions in λ. Hence we can compute λ_{k+1} as the next higher value for which some reduced cost coefficient is about to become positive. At this point, a single *pivot step* will yield $\mathbf{v}_{k+1}$. For details, we refer to the original article by Gass and Saaty [3].

The shadow vertex method. The Gass-Saaty method can also be used to solve an ordinary linear program with objective function $\mathbf{c}^T\mathbf{x}$, given some initial vertex $\mathbf{v}_0$. For this, we compute an auxiliary objective function $\mathbf{d}$ which is uniquely minimized by $\mathbf{v}_0$ (this is easy); then we run the Gass-Saaty method until we get to an interval $[\lambda_k, \lambda_{k+1}]$ containing 0. The corresponding vertex $\mathbf{v}_k$ maximizes $\mathbf{c}^T\mathbf{x}$ over $\mathcal{P}$.

Shadows of polytopes. It is clear that the efficiency of the Gass-Saaty method critically depends on the number of bends M, equivalently the number of vertices in the sequence $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_{M-1}$. Each of them turns out to be a *shadow vertex*, meaning that we still "see it" when we project the polytope P to the 2-dimensional plane spanned by the vectors $\mathbf{c}$ and $\mathbf{d}$ (see Lemma 3 below).

In many cases, the number of shadow vertices is small. For example, when we project the unit cube $[0, 1]^d$ to any 2-dimensional plane, we obtain a polygon with at most $2d$ vertices (see Theorem 4 below). In the worst case, however, shadows can be large. Murty (in the dual setting) was the first to construct shadows whose size is exponential in the dimension of the polytope [8]. A more explicit primal construction was later provided by Goldfarb in form of a d-dimensional defomed cube, with all its 2^d vertices appearing on some 2-dimensional shadow [4, 5]. A further simplification of the construction is due to Amenta and Ziegler [1, Section 4.3]). The Goldfarb cube also serves as the starting point for an exponential lower bound on the complexity of a support vector machine's *regularization path* [2].

The effect of sparsity. Let us now consider an explicit inequality description the polytope $\mathcal{P}$,

$$\mathcal{P} = \{\mathbf{x} \in \mathbb{R}^d : A\mathbf{x} \leq \mathbf{b}\}, \quad A \in \mathbb{R}^{n \times d}, b \in \mathbb{R}^n.$$

We call this description *t-sparse* if every row of the matrix A has at most t nonzero entries. In other words, if every inequality contains at most t variables. The question that we ask in this paper is the following: what is the effect of sparsity on the size of the 2-dimensional shadows of $\mathcal{P}$?

The Goldfarb cube in the version of Amenta and Ziegler [1, Section 4.3]) has a 3-sparse inequality description, meaning that exponentially large shadows require at most 3 variables per inequality. On the other hand, any polytope with a 1-sparse inequality description is an axis-parallel box, with 2-dimensional shadows of size $2d$ at most (Theorem 4).

This means, the interesting case is the 2-sparse one. We were initially hoping that 2-sparsity entails small shadows as well. In this paper, we show that this is not the case, by constructing a d-dimensional polytope with a 2-sparse description by $2d$ inequalities, having a 2-dimensional shadow of size 2^d. In fact, this polytope is the well-known *Klee-Minty cube* [7], with carefully chosen projection vectors $\mathbf{c}$ and $\mathbf{d}$.

In the next section, we formally define shadows of polytopes and prove some basic properties. Section 3 deals with the 1-sparse case. The fact that 2-dimensional shadows are small in this case is well-known; we will provide a simple proof for the sake of completeness. Section 4 contains the main contribution of this paper: a 2-dimensional shadow of the d-dimensional Klee-Minty cube with 2^d vertices.

2 Projections and Shadows

Let $\mathbf{c}, \mathbf{d} \in \mathbb{R}^d$ be linearly independent vectors. We consider the *linear projection* $\pi_{\mathbf{c},\mathbf{d}} : \mathbb{R}^d \to \mathbb{R}^2$ defined by

$$\pi_{\mathbf{c},\mathbf{d}}(\mathbf{x}) = \begin{pmatrix} \mathbf{c}^T\mathbf{x} \\ \mathbf{d}^T\mathbf{x} \end{pmatrix}. \tag{1}$$

Definition 1. *For a point set $\mathcal{P} \subseteq \mathbb{R}^d$, the* 2-dimensional shadow *(or simply* shadow*) of $\mathcal{P}$ w.r.t.* $\mathbf{c}$ *and* $\mathbf{d}$ *is*

$$\pi_{\mathbf{c},\mathbf{d}}(\mathcal{P}) := \{\pi_{\mathbf{c},\mathbf{d}}(\mathbf{x}) : \mathbf{x} \in \mathcal{P}\}.$$

If $\mathcal{P}$ is a polytope—the convex hull of its vertices [9, Proposition 2.2 (i)]—then its shadow is easily seen to be a polytope as well: the shadow is the convex hull of the projected vertices, some of which are the actual vertices of the shadow [9, Proposition 2.2 (ii)]. Hence we have the following fact.

Fact 2. *Let $\mathcal{P}$ be a polytope, and $\mathbf{w}$ be a vertex of the shadow $\pi_{\mathbf{c},\mathbf{d}}(\mathcal{P})$. Then there exists a vertex $\mathbf{v}$ of $\mathcal{P}$ such that $\mathbf{w} = \pi_{\mathbf{c},\mathbf{d}}(\mathbf{v})$.*

The next lemma provides a sufficient condition for a vertex to actually yield a shadow vertex.

Lemma 3. *Let $\mathcal{P}$ be a polytope, and let $\mathbf{v}$ be a vertex of $\mathcal{P}$. The projection $\mathbf{w} = \pi_{\mathbf{c},\mathbf{d}}(\mathbf{v})$ is a vertex of the shadow $\pi_{\mathbf{c},\mathbf{d}}(\mathcal{P})$ if there exists a linear combination $\mathbf{e}$ of $\mathbf{c}$ and $\mathbf{d}$ such that $\mathbf{v}$ is the unique maximizer of the linear function $\mathbf{e}^T\mathbf{x}$ over $\mathcal{P}$.*

Proof. For $\mathbf{a} = (a_1, a_2) \in \mathbb{R}^2$, define $\mathbf{e} = a_1\mathbf{c} + a_2\mathbf{d}$. With $\mathbf{y} = \pi_{\mathbf{c},\mathbf{d}}(\mathbf{x})$, we have

$$\mathbf{e}^T\mathbf{x} = a_1\mathbf{c}^T\mathbf{x} + a_2\mathbf{d}^T\mathbf{x} = \mathbf{a}^T\mathbf{y}. \tag{2}$$

Thus, if vertex $\mathbf{v}$ is the unique maximizer of $\mathbf{e}^T\mathbf{x}$ over $\mathcal{P}$, then $\mathbf{w} = \pi_{\mathbf{c},\mathbf{d}}(\mathbf{v})$ is the unique maximizer of $\mathbf{a}^T\mathbf{y}$ over $\pi_{\mathbf{c},\mathbf{d}}(\mathcal{P})$. This in turn means that $\mathbf{w}$ is a vertex of the shadow [9, Definition 2.1]. □

3 The 1-Sparse Case

Let us consider a system of inequalities in d variables $x_1, x_2, \ldots, x_d$, such that each inequality contains only one variable. Hence, the inequality is either an upper or a lower bound for that variable. By considering the tightest lower and upper bounds for each variable, we see that the set of solutions consists of all $\mathbf{x} \in \mathbb{R}^d$ such that

$$\ell_i \leq x_i \leq u_i, \quad i = 1, 2, \ldots, d, \tag{3}$$

for suitable numbers $\ell_i < u_i$ (we assume that the solution set is full-dimensonal and bounded). The vertices of the polytope $\mathcal{P}$—a box—defined by these inequalities are therefore all the 2^d points $\mathbf{x}$ for which $x_i \in \{\ell_i, u_i\}$ for all i.

Theorem 4. *Let $\mathcal{P}$ be a box as in (3). Then $\pi_{\mathbf{c},\mathbf{d}}(\mathcal{P})$ has at most $2d$ vertices.*

Proof. We may assume that there is no i such that $c_i = d_i = 0$ (otherwise we reduce the dimension of the problem by ignoring coordinate i and obtain a bound of $2(d-1)$). We now prove that there are at most $2d$ vertices of $\mathcal{P}$ that project to some vertex $\mathbf{w}$ of $\pi_{\mathbf{c},\mathbf{d}}(\mathcal{P})$. We recall that $\mathbf{w}$ is a vertex if and only if $\mathbf{w}$ is the unique maximizer of $\mathbf{a}^T\mathbf{y}$ over $\pi_{\mathbf{c},\mathbf{d}}(\mathcal{P})$ for suitable $\mathbf{a} \in \mathbb{R}^2$ [9, Definition 2.1]. Again, we set $\mathbf{e} = a_1\mathbf{c} + a_2\mathbf{d}$.

Since $\mathbf{w}$ is the *unique* maximizer, we can slightly perturb $\mathbf{a}$ and w.l.o.g. assume that $e_i = a_1c_i + a_2d_i \neq 0$ for all i. We now claim that the sign pattern of $\mathbf{e}$ uniquely determines the preimage $\mathbf{v}$ of $\mathbf{w}$. To see this, we use (2) to argue that any preimage of $\mathbf{w}$ maximizes $\mathbf{e}^T\mathbf{x}$ over $\mathcal{P}$. But under $e_i \neq 0$ for all i, there is only one such maximizer $\mathbf{v}$, given by:

$$v_i = \begin{cases} \ell_i, & \text{if } e_i < 0 \\ u_i & \text{if } e_i > 0. \end{cases}, i = 1, 2, \ldots, d.$$

Thus, the theorem follows if we can prove that there are at most $2d$ different sign patterns that may occur in $\mathbf{e}$. For each i, we consider the line

$$L_i = \{\mathbf{y} \in \mathbb{R}^2 : y_1 c_i + y_2 d_i = 0\}$$

through the origin. The *arrangement* of all d such lines subdivides the plane into *cells* where all points $\mathbf{a}$ within a fixed cell lead to the same sign pattern of $\mathbf{e}$. The twodimensional cells correspond to the nowhere zero sign patterns of interest. It remains to observe that an arrangement of d lines through the origin induces at most $2d$ twodimensional cells. □

We remark that we have reproved a special case of a general statement that relates *zonotopes* to *arrangements of hyperplanes* [9, Corollary 7.17].

4 The 2-Sparse Case

We begin by introducing the d-dimensional Klee-Minty cube in the variant of Amenta and Ziegler [1, Section 4.1]. The original Klee-Minty cube—the first and celebrated worst-case input for the simplex method (with Dantzig's pivot rule)—differs from this variant by a suitable scaling of the inequalities [7].

Definition 5. *For fixed $0 < \varepsilon < 1/2$, the d-dimensional Klee-Minty cube is the set of solutions of the following system of $2d$ inequalities that come in d pairs, where the jth pair specifies a lower and an upper bound for variable x_j.*

$$\begin{array}{rcccl} 0 & \leq & x_1 & \leq & 1 \\ \varepsilon x_{j-1} & \leq & x_j & \leq & 1 - \varepsilon x_{j-1}, \quad j = 2, \ldots, d. \end{array} \tag{4}$$

4.1 Vertices

It is easily shown by induction that $0 \leq x_j \leq 1$ for every $\mathbf{x} = (x_1, x_2, \ldots, x_d)$ in the polyhedron (4) and every j. Hence, we are dealing with a polytope. Using $\varepsilon < 1/2$, this in turn implies that from any pair of inequalities, at most one can be tight. A vertex of the polytope (having d tight inequalities) can therefore uniquely be encoded by a bit vector $\mathbf{u} \in \{0,1\}^d$ where $u_j = 0$ means that the lower bound is tight in the jth pair of inequalities, while $u_j = 1$ means that the upper bound is tight. In fact, every bit vector $\mathbf{u}$ induces a vertex $\mathbf{x}(\mathbf{u})$ defined by selecting from each pair of inequalities the tight one according to $\mathbf{u}$.

Definition 6. *For $\mathbf{u} \in \{0,1\}^d$, we ley $\mathbf{x}(\mathbf{u}) \in \mathbb{R}^d$ be the vector recursively defined by*

$$x_j(\mathbf{u}) := (1 - u_j)\varepsilon x_{j-1}(\mathbf{u}) + u_j(1 - \varepsilon x_{j-1}(\mathbf{u})) = u_j + (1 - 2u_j)\varepsilon x_{j-1}(\mathbf{u}), \tag{5}$$

for $j = 1, \ldots, d$, where we use the convention that $x_0(\mathbf{u}) = 0$. In particular, $\mathbf{x}(\mathbf{u})$ is one of 2^d vertices of the Klee-Minty cube (4).

4.2 Edges and Edge Directions

Two vertices $\mathbf{u}, \mathbf{u}'$ are *neighbors* if and only if their convex hull is an edge (having $d-1$ tight inequalities). This in turn is the case if and only if $\mathbf{u}'$ is of the form $\mathbf{u} \oplus \{\ell\}$ (the bit vector obtained from $\mathbf{u}$ by flipping the ℓ-th component). A general result [9, Lemma 3.6] entails the following key fact.

Fact 7. *Let $\mathbf{e} \in \mathbb{R}^d$, $\mathbf{u} \in \{0,1\}^d$. The following two statements are equivalent.*

(i) $\mathbf{e}^T\mathbf{x}(\mathbf{u}) > \mathbf{e}^T\mathbf{x}$ *for all* $\mathbf{x}$ *in (4)*

(ii) $\mathbf{e}^T\mathbf{x}(\mathbf{u}) > \mathbf{e}^T\mathbf{x}(\mathbf{u} \oplus \{\ell\})$ *for all* $\ell \in \{1, 2, \ldots, d\}$.

Below we will use this fact together with Lemma 3 to prove that $\mathbf{x}(\mathbf{u})$ yields a shadow vertex for all $\mathbf{u}$. In order to arrive at vectors $\mathbf{c}, \mathbf{d}$ that define a suitable shadow, we need a little more notation.

Definition 8. *For $\mathbf{u} \in \{0,1\}^d$ and $\ell \in \{1, 2, \ldots, d\}$, we define*

$$q^{(\ell)}(\mathbf{u}) = x_\ell(\mathbf{u} \oplus \{\ell\}) - x_\ell(\mathbf{u}), \quad \ell = 1, 2, \ldots, d. \tag{6}$$

Furthermore, for $i, j \in \{1, 2, \ldots, d\}$, we set

$$p_i^j(\mathbf{u}) = \prod_{k=i}^{j} (1 - 2u_k) \in \{-1, 1\}. \tag{7}$$

Note that $p_i^j(\mathbf{u})$ simply encodes the parity of the bit vector $u_i, u_{i+1}, \ldots, u_j$. In order to apply Fact 7, we need to compute the edge directions.

Lemma 9. *Let $\mathbf{u} \in \{0,1\}^d$ and $\ell \in \{1, 2, \ldots, d\}$. Then*

$$x_j(\mathbf{u} \oplus \{\ell\}) - x_j(\mathbf{u}) = \begin{cases} 0, & \text{if } j < \ell, \\ p_{\ell+1}^j(\mathbf{u})\varepsilon^{j-\ell} q^{(\ell)}(\mathbf{u}) & \text{if } j \geq \ell. \end{cases}$$

Proof. By induction on j. For $j < \ell$, the values $x_j(\mathbf{u} \oplus \{\ell\})$ and $x_j(\mathbf{u})$ agree, since by (5), they only depend on bits $u_i, i \leq j < \ell$. For $j = \ell$, we recover (6). For $j > \ell$, we use (5) to compute

$$x_j(\mathbf{u} \oplus \{\ell\}) - x_j(\mathbf{u}) = (1 - 2u_j) \cdot \varepsilon(x_{j-1}(\mathbf{u} \oplus \{\ell\}) - x_{j-1}(\mathbf{u})),$$

and the statement follows from the inductive hypothesis. □

The previous lemma shows that all components of $\mathbf{x}(\mathbf{u} \oplus \{\ell\})) - \mathbf{x}(\mathbf{u})$ are multiples of $q^{(\ell)}(\mathbf{u})$, and it will be convenient to take out this factor.

Definition 10. *For $\mathbf{u} \in \{0,1\}^d$ and $\ell \in \{1, 2, \ldots, d\}$, let $\mathbf{y}^{(\ell)}(\mathbf{u})$ be the vector defined by*

$$y_j^{(\ell)}(\mathbf{u}) = \begin{cases} 0, & \text{if } j < \ell, \\ p_{\ell+1}^j(\mathbf{u})\varepsilon^{j-\ell} & \text{if } j \geq \ell. \end{cases} \tag{8}$$

We thus have

$$\mathbf{x}(\mathbf{u} \oplus \{\ell\}) - \mathbf{x}(\mathbf{u}) = \mathbf{y}^{(\ell)}(\mathbf{u}) \cdot q^{(\ell)}(\mathbf{u}). \tag{9}$$

4.3 The Shadow

Let $\mathcal{P}$ be the Klee-Minty cube as defined in (4). We want to construct vectors $\mathbf{c}$ and $\mathbf{d}$ such that the shadow $\pi_{\mathbf{c},\mathbf{d}}(\mathcal{P})$ has the maximum of 2^d vertices. Our approach is as follows. With a suitable $\mathbf{c}$, we use $\mathbf{d} = (0,\dots,0,1)$. For every $\mathbf{u} \in \{0,1\}^d$, we find a multiple $\mathbf{d}(\mathbf{u})$ of $\mathbf{d}$ such that the vertex $\mathbf{x}(\mathbf{u})$ is the unique maximizer of the linear function $\mathbf{e}(\mathbf{u})^T\mathbf{x} := (\mathbf{c}+\mathbf{d}(\mathbf{u}))^T\mathbf{x}$ over (4). With Lemma 3, we conclude that $\pi_{\mathbf{c},\mathbf{d}}(\mathbf{x}(\mathbf{u}))$ is a shadow vertex.

Definition 11. *For* $\mathbf{u} \in \{0,1\}^d$, *let*

$$\mathbf{c} := (\varepsilon^{3(d-1)}, \varepsilon^{3(d-2)}, \dots, \varepsilon^3, 0) \in \mathbb{R}^d$$

and

$$\mathbf{d}(u) := (0,0,\dots,0,-\sum_{j=0}^{d-1} p_{j+1}^d(\mathbf{u})\varepsilon^{2(d-j)}) \in \mathbb{R}^d.$$

Lemma 12. *Let* $\mathbf{u} \in \{0,1\}^d, \ell \in \{1,2\dots,d\}, \mathbf{e}(\mathbf{u}) := \mathbf{c} + \mathbf{d}(u)$. *For* $\varepsilon < 1/2$,

$$\mathbf{e}(\mathbf{u})^T(\mathbf{x}(\mathbf{u} \oplus \{\ell\}) - \mathbf{x}(\mathbf{u})) < 0,$$

meaning that $\mathbf{x}(\mathbf{u})$ *has larger* $\mathbf{e}(\mathbf{u})^T\mathbf{x}$*-value than all its neighbors.*

According to Fact 7, $\mathbf{x}(\mathbf{u})$ then uniquely maximizes $\mathbf{e}(\mathbf{u})^T\mathbf{x}$ over the Klee-Minty cube (4) and thus contributes to the shadow by Lemma 3. It only remains to prove Lemma 12.

Proof. Making use of (9), we first compute $\mathbf{e}(\mathbf{u})^T\mathbf{y}(u) = \mathbf{c}^T\mathbf{y}(\mathbf{u}) + \mathbf{d}(\mathbf{u})^T\mathbf{y}(\mathbf{u})$. We have

$$\mathbf{c}^T\mathbf{y}(\mathbf{u}) = \sum_{j=\ell}^{d-1} \varepsilon^{3(d-j)} p_{\ell+1}^j(\mathbf{u})\varepsilon^{j-\ell} = \sum_{j=\ell}^{d-1} p_{\ell+1}^j(\mathbf{u})\varepsilon^{(3d-\ell)-2j} \tag{10}$$

and

$$\begin{aligned} \mathbf{d}(\mathbf{u})^T\mathbf{y}(\mathbf{u}) &= -\sum_{j=0}^{d-1} p_{j+1}^d(\mathbf{u})\varepsilon^{2(d-j)} p_{\ell+1}^d(\mathbf{u})\varepsilon^{d-\ell} \\ &= -\sum_{j=0}^{d-1} p_{j+1}^d(\mathbf{u}) p_{\ell+1}^d(\mathbf{u})\varepsilon^{(3d-\ell)-2j}. \end{aligned} \tag{11}$$

For $j \geq \ell$, (7) and the subsequent parity interpretation of p yields

$$p_{j+1}^d(\mathbf{u})p_{\ell+1}^d(\mathbf{u}) = p_{\ell+1}^j(\mathbf{u}),$$

meaning that the terms for $j = \ell,\dots,d-1$ in (10) and (11) cancel, and we get

$$\mathbf{e}(\mathbf{u})^T\mathbf{y}(\mathbf{u}) = \mathbf{c}^T\mathbf{y}(\mathbf{u}) + \mathbf{d}(\mathbf{u})^T\mathbf{y}(\mathbf{u}) = -\sum_{j=0}^{\ell-1} p_{j+1}^d(\mathbf{u})p_{\ell+1}^d(\mathbf{u})\varepsilon^{(3d-\ell)-2j}.$$

This expression is a polynomial in ε whose nonzero coefficients are in $\{-1, 1\}$. Hence, for $\varepsilon < 1/2$, the sign of this polynomial is determined by the coefficient for $j = \ell - 1$ which is

$$-p_\ell^d(\mathbf{u})p_{\ell+1}^d(\mathbf{u}) = -p_\ell^\ell(\mathbf{u}) = -(1 - 2u_\ell).$$

Now using (9), our actual expression of interest $\mathbf{e}(\mathbf{u})^T(\mathbf{x}(\mathbf{u} \oplus \{\ell\}) - \mathbf{x}(\mathbf{u}))$ has the same sign as

$$\begin{aligned} -(1 - 2u_\ell)q^{(\ell)}(\mathbf{u}) &= -(1 - 2u_\ell)(x_\ell(\mathbf{u} \oplus \{\ell\}) - x_\ell(\mathbf{u})) \\ &\overset{(5)}{=} -(1 - 2u_\ell)^2(1 - 2\varepsilon x_{\ell-1}(\mathbf{u})). \end{aligned}$$

By $\varepsilon < 1/2$ and $x_{\ell-1}(\mathbf{u}) \leq 1$, the sign of this expression is negative, as desired. □

References

[1] Nina Amenta and Günter M. Ziegler. Deformed products and maximal shadows of polytopes. In B. Chazelle, J. Goodman, and R. Pollack, editors, *Advances in Discrete and Computational Geometry*, volume 223 of *Contemporary Mathematics*, pages 57–90. Amer. Math. Soc., Providence, RI, 1998.

[2] Bernd Gärtner, Martin Jaggi, and Clément Maria. An exponential lower bound on the complexity of regularization paths. *Journal of Computational Geometry*, 3(1):168–195, 2012.

[3] Saul Gass and Thomas Saaty. The computational algorithm for the parametric objective function. *Naval Research Logistics Quarterly*, 2(1-2):39–45, 1955.

[4] Donald Goldfarb. Worst case complexity of the shadow vertex simplex algorithm. Technical report, Columbia University, 1983.

[5] Donald Goldfarb. On the Complexity of the Simplex Method. In *Advances in optimization and numerical analysis, Proc. 6th Workshop on Optimization and Numerical Analysis, January 1992*, pages 25–38, Oaxaca, Mexico, 1994.

[6] Trevor Hastie, Saharon Rosset, Robert Tibshirani, and Ji Zhu. The Entire Regularization Path for the Support Vector Machine. *The Journal of Machine Learning Research*, 5:1391–1415, 2004.

[7] Victor Klee and George J. Minty. How good is the simplex algorithm? In O. Shisha, editor, *Inequalities*, volume III, pages 159–175. Academic Press, New York, 1972.

[8] Katta G Murty. Computational complexity of parametric linear programming. *Mathematical Programming*, 19(1):213–219, 1980.

[9] Günter M. Ziegler. *Lectures on polytopes*. Springer-Verlag, New York, 1995.