Supplementary Material to "Nonparametric Least Squares Estimation of a Multivariate Convex Regression Function"

Emilio Seijo and Bodhisattva Sen
Columbia University

1 Proofs of Lemmas 3.2, 3.3 and 3.4

Before we prove Lemmas 3.2 and 3.3, we need some additional results from matrix algebra. For convenience, we only state them here. The interested reader can find their proofs in Section 4 of this document.

We first introduce some notation. We write $\mathbf{e}_j \in \mathbb{R}^d$ for the vector whose components are given by $\mathbf{e}_j^k = \delta_{jk}$, where δ_{jk} is the Kronecker δ. We also write $\mathbf{e} = \mathbf{e}_1 + \ldots + \mathbf{e}_d$ for the vector of ones in $\mathbb{R}^d$. For $\alpha \in \{-1, 1\}^d$ we write

$$\mathcal{R}_\alpha = \left\{ \sum_{k=1}^{d} \theta^k \alpha^k \mathbf{e}_k : \theta \geq 0, \theta \in \mathbb{R}^d \right\}$$

for the orthant in the α direction. For any hyperplane $\mathcal{H}$ defined by the normal vector $\xi \in \mathbb{R}^d$ and the intercept $b \in \mathbb{R}$, we write $\mathcal{H} = \{x \in \mathbb{R}^d : \langle \xi, x \rangle = b\}$, $\mathcal{H}^+ = \{x \in \mathbb{R}^d : \langle \xi, x \rangle > b\}$ and $\mathcal{H}^- = \{x \in \mathbb{R}^d : \langle \xi, x \rangle < b\}$. For $r > 0$ and $x_0 \in \mathbb{R}^d$ we will write $B(x_0, r) = \{x \in \mathbb{R}^d : |x - x_0| < r\}$. We denote by $\mathbb{R}^{d \times d}$ the space of $d \times d$ matrices endowed with the topology defined by the $\|\cdot\|_2$ norm (where $\|A\|_2 = \sup_{|x| \leq 1}\{|Ax|\}$).

Lemma 1.1 *Let $r > 0$. There is a constant $R_r > 0$, depending only on r and d, such that for any $\rho_* \in (0, R_r)$ there are $\rho, \rho^* > 0$ with the property: for any $\alpha \in \{-1, 1\}^d$ and any d-tuple of vectors $\beta = \{x_1, \ldots, x_d\} \subset \mathbb{R}^d$ such that $x_j \in B(\alpha^j r \boldsymbol{e}_j, \rho)$ $\forall$ $j = 1, \ldots, d$, there is a unique pair $(\xi_{\alpha,\beta}, b_{\alpha,\beta})$, with $\xi_{\alpha,\beta} \in \mathbb{R}^d$, $|\xi_{\alpha,\beta}| = 1$ and $b_{\alpha,\beta} > 0$ for which the following statements hold:*

(i) β *forms a basis for* $\mathbb{R}^d$.

(ii) $x_1, \ldots, x_d \in \mathcal{H}_{\alpha,\beta} := \{x \in \mathbb{R}^d : \langle \xi_{\alpha,\beta}, x\rangle = b_{\alpha,\beta}\}$.

(iii) $\min\limits_{1\le j\le d}\{|\xi^j_{\alpha,\beta}|\} > 0$.

(iv) $B(0,\rho_*) \subset \mathcal{H}^-_{\alpha,\beta}$.

(v) $\{x \in \mathbb{R}^d : |x| \ge \rho^*\} \cap \mathcal{R}_\alpha \subset \mathcal{H}^+_{\alpha,\beta}$.

(vi) $B(-\alpha^j r\boldsymbol{e}_j, \rho) \subset \{x \in \mathbb{R}^d : \langle \xi_{\alpha,\beta}, x\rangle < 0\}$ *for all* $j = 1, \ldots, d$.

(vii) *For any* $w_1 \in B\left(0, \frac{\rho_*}{16\sqrt{d}}\right)$ *and* $w_2 \in B\left(\frac{3\rho_*}{8\sqrt{d}}\alpha, \frac{\rho_*}{8\sqrt{d}}\right)$ *we have*

$$\min_{1\le j\le d}\left\{\left(X_\beta^{-1}\left(w_1 + t(w_2 - w_1)\right)\right)^j\right\} > 0 \quad \forall\ t \ge 1$$

where $X_\beta = (x_1, \ldots, x_d) \in \mathbb{R}^{d\times d}$ *is the matrix whose* j*'th column is* x_j.

Figure 1a illustrates the above lemma when $d = 2$ and $\alpha = (1,1)$. The lemma states that whatever points x_1 and x_2 are taken inside the circles of radius ρ around $\alpha^1 r\mathbf{e}_1$ and $\alpha^2 r\mathbf{e}_2$, respectively, $B(0,\rho_*)$ and $\{x \in \mathbb{R}^d : |x| \ge \rho^*\} \cap \mathcal{R}_\alpha$ are contained, respectively, in the half-spaces $\mathcal{H}^-_{\alpha,\beta}$ and $\mathcal{H}^+_{\alpha,\beta}$. Assertion (vii) of the lemma implies that all the points in the half line $\{w_1 + t(w_2 - w_1)\}_{t\ge1}$ should have positive coordinates with respect to the basis β as they do with respect to the basis $\{\alpha^j\mathbf{e}_j\}_{j=1}^d$. We refer the reader to Section 4.1 of this manuscript for a complete proof of Lemma 1.1.

We now state two other useful results, namely Lemma 1.2 and Lemma 1.3. Their proofs can be found in Sections 4.2 and 4.3 of the present document, respectively.

Lemma 1.2 *Let* $r > 0$ *and consider the notation of Lemma 1.1 with the positive numbers* ρ, ρ_* *and* ρ^* *as defined there. Take 2d vectors* $\{x_{\pm1}, \ldots, x_{\pm d}\} \subset \mathbb{R}^d$ *such that* $x_{\pm j} \in B(\pm r\boldsymbol{e}_j, \rho)$ *and for* $\alpha \in \{-1,1\}^d$ *write* $\beta_\alpha = \{x_{\alpha^1 1}, x_{\alpha^2 2}, \ldots, x_{\alpha^d d}\}$, $\xi_\alpha = \xi_{\alpha,\beta_\alpha}$, $b_\alpha = b_{\alpha,\beta_\alpha}$ *and* $\mathcal{H}_\alpha = \mathcal{H}_{\alpha,\beta_\alpha}$, *all in agreement with the setting of Lemma 1.1. Then, if* $K = Conv\,(x_{\pm1}, \ldots, x_{\pm d})$ *we have:*

(i) $K = \bigcap_{\alpha\in\{-1,1\}^d}\{x \in \mathbb{R}^d : \langle \xi_\alpha, x\rangle \le b_\alpha\}$.

(ii) $K^\circ = \bigcap_{\alpha\in\{-1,1\}^d}\{x \in \mathbb{R}^d : \langle \xi_\alpha, x\rangle < b_\alpha\}$.

(iii) $\partial K = \bigcup_{\alpha\in\{-1,1\}^d} Conv\,(x_{\alpha^1 1}, \ldots, x_{\alpha^d d})$.

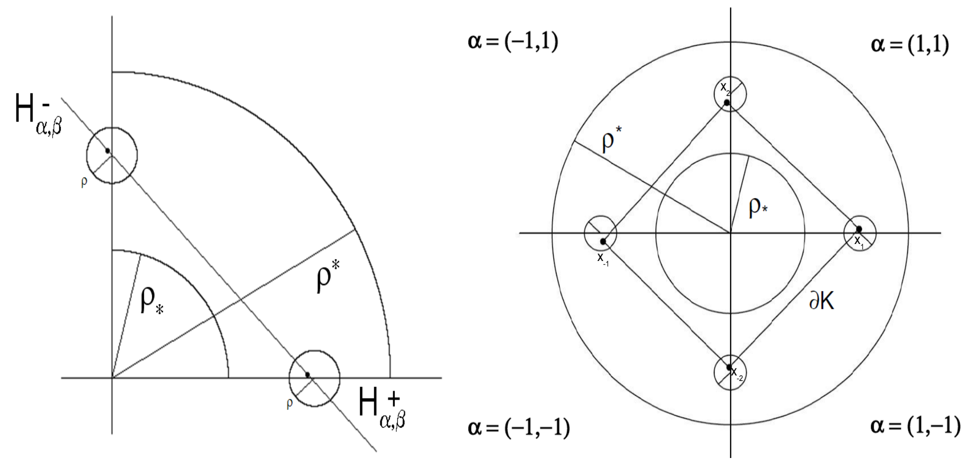

Figure 1: Explanatory diagram for (a) Lemma 1.1 (left panel); (b) Lemma 1.2 (right panel).

(iv) $\partial K = \left(\bigcup_{\alpha\in\{-1,1\}^d}\{x \in \mathbb{R}^d : \langle \xi_\alpha, x\rangle = b_\alpha\}\right) \bigcap \left(\bigcap_{\alpha\in\{-1,1\}^d}\{x \in \mathbb{R}^d : \langle \xi_\alpha, x\rangle \le b_\alpha\}\right)$.

(v) $B(0, \rho_*) \subset K^\circ$.

(vi) $\partial B(0, \rho^*) \subset \mathrm{Ext}(K)$.

Figure 1b illustrates Lemma 1.2 for the two-dimensional case. Intuitively, the idea is that as long as the points $x_{\pm 1}$ and $x_{\pm 2}$ belong to $B(\pm r\mathbf{e}_1, \rho)$ and $B(\pm r\mathbf{e}_2, \rho)$, respectively, we will have $B(0, \rho_*)$ and $\partial B(0, \rho^*)$ as subsets of K° and $\mathrm{Ext}(K)$, respectively.

Lemma 1.3 *Let $[a, b] \subset \mathbb{R}^d$ be a compact rectangle and $r > 0$, with $r < \frac{1}{d-2}$ if $d \ge 3$. For each $\alpha \in \{-1, 1\}^d$ write $z_\alpha = a + \sum_{j=1}^d \frac{1+\alpha^j}{2}\left(b^j - a^j\right)\boldsymbol{e}_j$ so that $\{z_\alpha\}_{\alpha\in\{-1,1\}^d}$ is the set of vertices of $[a, b]$. Then, there is $\rho > 0$ such that if $x_\alpha \in B(z_\alpha + r(z_\alpha - z_{-\alpha}), \rho)$ $\forall\ \alpha \in \{-1, 1\}^d$, then*

$$[a, b] \subset Conv\left(x_\alpha : \alpha \in \{-1, 1\}^d\right)^\circ .$$

Figure 2a describes Lemma 1.3 in the two-dimensional case. As long as the points $x_{(\pm1,\pm1)}$ are chosen in the balls of radius ρ around $z_{(\pm1,\pm1)} + r(z_{(\pm1,\pm1)} - z_{(\mp1,\mp1)})$, $Conv\left(x_{(\pm1,\pm1)}\right)$ will contain $Conv\left(z_{(\pm1,\pm1)}\right)$.

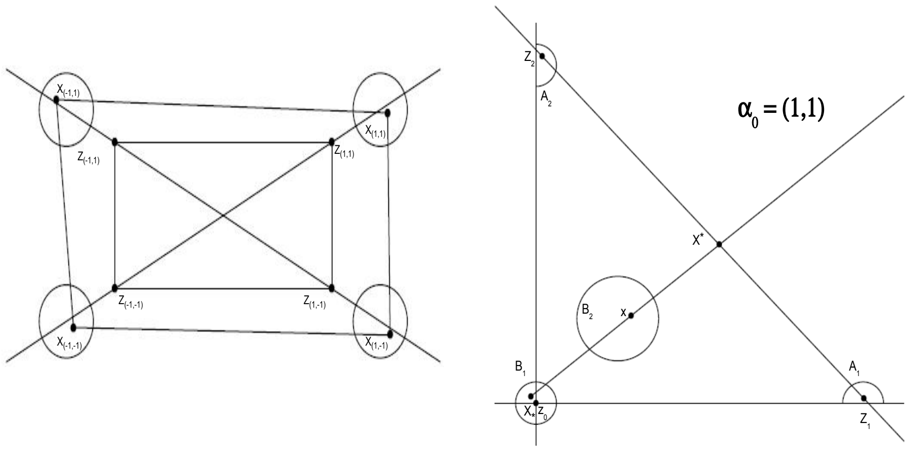

Figure 2: Explanatory diagram for (a) Lemma 1.3 (left panel); (b) Lemma 3.2 (right panel).

1.1 Proof of Lemma 3.2

Since any compact subset of $\mathfrak{X}^\circ$ is contained in a finite union of compact rectangles, it is enough to prove the result when $\mathtt{X}$ is a compact rectangle $[a, b] \subset \mathfrak{X}^\circ$. Let $r = \frac{1}{4}\min_{1\le k\le d}\{b^k - a^k\}$ and choose $\rho \in (0, \frac{1}{4}r)$, $\rho^* > 0$ and $0 < \rho_* < \frac{1}{2}r$ such that the conclusions of Lemmas 1.1 and 1.2 hold for any $\alpha \in \{-1, 1\}^d$ and any $\beta = (z_1, \dots, z_d) \in \mathbb{R}^{d\times d}$ with $z_j \in B(\alpha^j r\mathbf{e}_j, \rho)$. Take $N \in \mathbb{N}$ such that

$$\frac{1}{N}\max_{1\le k\le d}\{b^k - a^k\} < \frac{1}{32d}\rho_* \tag{1}$$

and divide $\mathtt{X}$ into N^d rectangles all of which are geometrically identical to $\frac{1}{N}[0, b-a]$. Let $\mathcal{C}$ be any one of the rectangles in the grid and choose any vertex z_0 of $\mathcal{C}$ satisfying

$$z_0 = \operatorname*{argmax}_{z\in\mathcal{C}}\left\{\max_{1\le j\le d}\left\{z^j - a^j, b^j - z^j\right\}\right\}.$$

Then, from the definition of z_0 and r, there is $\alpha_0 \in \{-1, 1\}^d$ such that

$$B(z_0, r) \cap (z_0 + \mathcal{R}_{\alpha_0}) \subset \mathtt{X}.$$

Additionally, define

$$\begin{aligned}
B_1 &= B\left(z_0, \frac{\rho_*}{16\sqrt{d}}\right), \\
B_2 &= B\left(z_0 + \frac{3\rho_*}{8\sqrt{d}}\alpha_0, \frac{\rho_*}{8\sqrt{d}}\right), \\
A_j &= B(z_0 + \alpha_0^j r\mathbf{e}_j, \rho) \cap (z_0 + \mathcal{R}_{\alpha_0}) \quad \forall\, j = 1, \dots, d, \\
A_{-j} &= B(z_0 - \alpha_0^j r\mathbf{e}_j, \rho) \quad \forall\, j = 1, \dots, d.
\end{aligned}$$

Observe that all the sets in the previous display have positive Lebesgue measure and that the A_{-j}'s are not necessarily contained in $\mathtt{X}$. Let $M_1 = \|\phi\|_{\mathtt{X}}$, $M_0 > \frac{\sigma}{\sqrt{\min\{\nu(B_1),\nu(B_2),\nu(A_1),\dots,\nu(A_d)\}}}$, $M = M_1 + M_0$ and $K_{\mathcal{C}} > 6M$. Also, notice that $\mathcal{C} \subset B_1$ because of (1). We will argue that

$$\mathbf{P}\left(\inf_{x\in\mathcal{C}}\{\hat{\phi}_n(x)\} \leq -K_{\mathcal{C}} \text{ i.o.}\right) = 0. \tag{2}$$

From Lemma 3.1, we know that

$$\mathbf{P}\left(\bigcap_{j=1}^{d}\left[\inf_{x\in A_j}\left\{\left|\hat{\phi}_n(x) - \phi(x)\right|\right\} < M_0 \text{ a.a.}\right]\right) = 1, \tag{3}$$

so there is, with probability one, $n_0 \in \mathbb{N}$ such that $\inf_{x\in A_j}\left\{\left|\hat{\phi}_n(x) - \phi(x)\right|\right\} < M_0$ for any $n \geq n_0$ and any $j = 1, \dots, d$.

Assume that the event $\left[\inf_{x\in\mathcal{C}}\{\hat{\phi}_n(x)\} \leq -K_{\mathcal{C}} \text{ i.o.}\right]$ is true. Then, there is a subsequence n_k such that $\inf_{x\in\mathcal{C}}\{\hat{\phi}_{n_k}(x)\} \leq -K_{\mathcal{C}}$ for all $k \in \mathbb{N}$. Fix any $k \geq n_0$. We know that there is $X_* \in \mathcal{C} \subset B_1$ such that $\hat{\phi}_{n_k}(X_*) \leq -K_{\mathcal{C}}$. In addition, for $j = 1, \dots, d$, there are $Z_{\alpha_0^j j} \in A_j$ such that $|\hat{\phi}_{n_k}(Z_{\alpha_0^j j}) - \phi(Z_{\alpha_0^j j})| < M_0$, which in turn implies $\hat{\phi}_{n_k}(Z_{\alpha_0^j j}) < M$. Pick any $Z_{-\alpha_0^j} \in A_{-j}$ and let $K = Conv\,(Z_{\pm 1}, \dots, Z_{\pm d}) = z_0 + Conv\,(Z_{\pm 1} - z_0, \dots, Z_{\pm d} - z_0)$.

Take any $x \in B_2$. We will show the existence of $X^* \in Conv\left(Z_{\alpha_0^1 1}, \dots, Z_{\alpha_0^d d}\right)$ such that $x \in Conv\,(X_*, X^*)$, as shown in Figure 2b for the case $d = 2$. We will then show that the existence of such an X^* implies that

$$|\phi(x) - \hat{\phi}_{n_k}(x)| > M_0. \tag{4}$$

Consequently, since x is an arbitrary element of B_2 we will have

$$\left[\inf_{x\in\mathcal{C}}\{\hat{\phi}_n(x)\} \le -K_{\mathcal{C}} \text{ i.o.}\right] \cap \left(\bigcap_{j=1}^{d}\left[\inf_{x\in A_j}\left\{\left|\hat{\phi}_n(x)-\phi(x)\right|\right\} < M_0 \text{ a.a.}\right]\right)$$

$$\subset \left[\inf_{x\in B_2}\{|\phi(x)-\hat{\phi}_{n_k}(x)|\} \ge M_0 \text{ i.o.}\right].$$

But from Lemma 3.1, the event on the right is a null set. Taking (3) into account, we will see that (2) holds and then complete the argument by taking $K_{\mathtt{X}} = \max_{\mathcal{C}}\{K_{\mathcal{C}}\}$.

To show the existence of X^* consider the function $\psi : \mathbb{R} \to \mathbb{R}^d$ given by $\psi(t) = X_* + t(x - X_*)$. The function ψ is clearly continuous and satisfies $\psi(0) = X_*$ and $\psi(1) = x \in B_2 \subset K^\circ$. That $B_2 \subset K^\circ$ is a consequence of Lemma 1.1, (iv). The set K is bounded, so there is $T > 1$ such that $\psi(T) \in \mathrm{Ext}(K) = \mathbb{R}^d \setminus \overline{K}$. The intermediate value theorem then implies that there is $t^* \in (1, T)$ such that $X^* := \psi(t^*) \in \partial K$. Observe that by Lemma 1.2 (iii) we have

$$\partial K = \bigcup_{\alpha\in\{-1,1\}^d} Conv\left(Z_{\alpha^1 1}, \dots, Z_{\alpha^d d}\right).$$

Lemma 1.1 (i) implies that $\{Z_{\alpha_0^1 1} - z_0, \dots, Z_{\alpha_0^d d} - z_0\}$ forms a basis of $\mathbb{R}^d$ so we can write $X^* - z_0 = \sum_{j=1}^d \theta^j (Z_{\alpha_0^j j} - z_0)$. Moreover, Lemma 1.1 (vii) implies that $\theta^j > 0$ for every $j = 1, \dots, d$ as $\theta = (\theta^1, \dots, \theta^d) = (Z_{\alpha_0^1 1} - z_0, \dots, Z_{\alpha_0^d d} - z_0)^{-1}(X^* - z_0)$. Here we apply Lemma 1.1 (vii) with $w_1 = X_* \in B_1$, $w_2 = x \in B_2$ and $t^* > 1$.

For $\alpha \in \{-1, 1\}^d$ consider the pair $(\xi_\alpha, b_\alpha) \in \mathbb{R}^d \times \mathbb{R}$ as defined in Lemma 1.2 for the set of vectors $\{Z_{\pm 1} - z_0, \dots, Z_{\pm d} - z_0\}$ (here we move the origin to z_0). Observe that Lemma 1.1 (ii) implies that $\langle \xi_{\alpha_0}, Z_{\alpha_0^j j} - z_0\rangle = b_{\alpha_0}$ for all $j = 1, \dots, d$. Consequently, $\langle \xi_{\alpha_0}, X^* - z_0\rangle = b_{\alpha_0} \sum_{j=1}^d \theta^j$, but since $X^* \in \partial K$, Lemma 1.2 (iv) implies that $\langle \xi_{\alpha_0}, X^* - z_0\rangle \le b_{\alpha_0}$ and hence $\sum_{j=1}^d \theta^j \le 1$. Additionally, for $\alpha \ne \alpha_0$ we can write $\langle \xi_\alpha, X^* - z_0\rangle$ as

$$\sum_{j=1}^{d} \theta^j \langle \xi_\alpha, Z_{\alpha_0^j j} - z_0\rangle = \sum_{\alpha^j=\alpha_0^j} \theta^j b_\alpha + \sum_{\alpha^j\ne\alpha_0^j} \theta^j \langle \xi_\alpha, Z_{\alpha_0^j j} - z_0\rangle < b_\alpha \tag{5}$$

as $\langle \xi_\alpha, Z_{\alpha^j} - z_0\rangle = b_\alpha$ (by Lemma 1.1 (ii)) and $\langle \xi_\alpha, Z_{-\alpha^j} - z_0\rangle < 0$ (by Lemma 1.1 (vi)) for every $j = 1, \dots, d$. Since $\langle \xi_\alpha, w - z_0\rangle = b_\alpha$ for all $w \in Conv\left(Z_{\alpha^1 1}, \dots, Z_{\alpha^d d}\right)$ and all $\alpha \in \{-1.1\}^d$, (5) and the fact that $X^* \in \partial K$ imply that $X^* \in Conv\left(Z_{\alpha_0^1 1}, \dots, Z_{\alpha_0^d d}\right)$.

Hence $\hat{\phi}_n(X^*) \leq \sum_{j=1}^d \theta^j \hat{\phi}_{n_k}(Z_{\alpha_0^j j}) < M$. We therefore have

$$\hat{\phi}_{n_k}(X^*) < M \quad , \quad \hat{\phi}_{n_k}(X_*) < -K_{\mathcal{C}}, \tag{6}$$

$$X_* + \frac{1}{t^*}(X^* - X_*) \;=\; x. \tag{7}$$

Since $X_* \in B_1$ and $d \geq 1$ we have

$$|z_0 - X_*| < \frac{1}{8}\rho_*. \tag{8}$$

By using the triangle inequality we get the following bounds

$$\frac{1}{4}\rho_* < |z_0 - x| < \frac{1}{2}\rho_*. \tag{9}$$

And from Lemma 1.1 (iv) and the fact that $\langle \xi_{\alpha_0}, X^* \rangle = b_{\alpha_0}$ we also obtain

$$|z_0 - X^*| \geq \rho_*. \tag{10}$$

From (7) we know that $t^* = \frac{|X^* - X_*|}{|x - X_*|}$. Using the triangle inequality with (8), (9) and (10) one can find lower and upper bounds for $|X^* - X_*|$ (as $|X^* - X_*| \geq |X^* - z_0| - |z_0 - X_*|$) and $|x - X_*|$ (as $|x - X_*| \leq |x - z_0| + |z_0 - X_*|$), respectively, to obtain $t^* \geq \frac{7}{5}$. Then, (6) and (7) imply

$$\hat{\phi}_{n_k}(x) \leq \left(1 - \frac{1}{t^*}\right)\hat{\phi}_{n_k}(X_*) + \frac{1}{t^*}\hat{\phi}_{n_k}(X^*) \leq -\frac{2}{7}K_{\mathcal{C}} + \frac{5}{7}M < -M.$$

Consequently,

$$|\phi(x) - \hat{\phi}_{n_k}(x)| > M - M_1 = M_0.$$

This proves (4) and completes the proof. $\square$

1.2 Proof of Lemma 3.3

Assume without loss of generality that $\mathtt{X}$ is a compact rectangle. Let $\{z_\alpha : \alpha \in \{-1,1\}^d\}$ be the set of vertices of the rectangle. Then, there is $r \in (0,1)$ such that $B(z_\alpha, r) \subset \mathfrak{X}^\circ \ \forall \ \alpha \in \{-1,1\}^d$. Recall that from Lemma 1.3, there is $0 < \rho < \frac{1}{2}r$ such that for any $\{\eta_\alpha : \alpha \in \{-1,1\}^d\}$ if $\eta_\alpha \in B(z_\alpha + \frac{r}{2}(z_\alpha - z_{-\alpha}), \rho)$ then $\mathtt{X} \subset Conv\left(\eta_\alpha : \alpha \in \{-1,1\}^d\right)$.

Let $A_\alpha = B(z_\alpha + \frac{1}{2}r(z_\alpha - z_{-\alpha}), \frac{\rho}{2})$ and $M_0 > \frac{\sigma}{\sqrt{\min\{\nu(A_\alpha) : \alpha \in \{-1,1\}^d\}}}$ and choose

$$M_1 = \sup_{x \in Conv\left(\bigcup_{\alpha \in \{-1,1\}^d} A_\alpha\right)} \{|\phi(x)|\}.$$

Take $K_{\mathtt{X}} > M_0 + M_1$. Since

$$\mathbf{P}\left(\bigcap_{\alpha\in\{-1,1\}^d}\left[\inf_{x\in A_\alpha}\{|\hat{\phi}_n(x)-\phi(x)|\} < M_0, \text{ a.a.}\right]\right) = 1$$

by Lemma 3.1, there is, with probability one, $n_0 \in \mathbb{N}$ such that for any $n \geq n_0$ we can find $\eta_\alpha \in A_\alpha$, $\alpha \in \{-1,1\}^d$, such that $|\hat{\phi}_n(\eta_\alpha) - \phi(\eta_\alpha)| < M_0$. It follows that $\hat{\phi}_n(\eta_\alpha) \leq K_{\mathtt{X}}\ \forall\ \alpha \in \{-1,1\}^d$. Now, using Lemma 1.3 we have $\mathtt{X} \subset Conv\left(\eta_\alpha : \alpha \in \{-1,1\}^d\right)$ and the convexity of $\hat{\phi}_n$ implies that $\hat{\phi}_n(x) \leq K_{\mathtt{X}}$ for any $x \in \mathtt{X}$. $\square$

1.3 Proof of Lemma 3.4

Assume that $\mathtt{X} = [a,b]$ is a rectangle with vertices $\{z_\alpha : \alpha \in \{-1,1\}^d\}$. The function $\psi(x) = \inf_{\eta\in\overline{\mathrm{Ext}(\mathfrak{X})}}\{|x-\eta|\}$ is continuous on $\mathbb{R}^d$ so there is $x_* \in \partial\mathtt{X}$ such that $\psi(x_*) = \inf_{x\in\partial\mathtt{X}}\{\psi(x)\}$. Observe that $\psi(x_*) > 0$ because $x_* \in \partial\mathtt{X} \subset \mathfrak{X}^\circ$. By Lemma 1.3, there is a $r < \frac{1}{2}\psi(x_*)$ for which there exists $\rho < \frac{1}{4}r$ such that whenever $\eta_\alpha \in A_\alpha := B\left(z_\alpha + \frac{3}{4}r\left(\frac{z_\alpha - z_{-\alpha}}{|z_\alpha - z_{-\alpha}|}\right), \rho\right)$ for any $\alpha \in \{-1,1\}^d$ and

$$\begin{aligned} K_z &= Conv\left(z_\alpha + \frac{1}{2}r\left(\frac{z_\alpha - z_{-\alpha}}{|z_\alpha - z_{-\alpha}|}\right) : \alpha \in \{-1,1\}^d\right) \\ K_\eta &= Conv\left(\eta_\alpha : \alpha \in \{-1,1\}^d\right) \end{aligned}$$

we have

$$\mathtt{X} \subset K_z \subset K_\eta^\circ \subset K_\eta \subset \mathfrak{X}^\circ. \tag{11}$$

Let $M_0 > \frac{\sigma}{\sqrt{\min\{\nu(A_\alpha):\alpha\in\{-1,1\}^d\}}}$ and $M_1 \in \mathbb{R}$ be such that

$$\mathbf{P}\left(\inf_{x\in\mathtt{X}}\{\hat{\phi}_n(x)\} \leq -M_0 \text{ i.o.}\right) = 0 \quad \text{and} \quad M_1 = \sup_{x\in Conv\left(\bigcup_{\alpha\in\{-1,1\}^d} A_\alpha\right)}\{\phi(x)\}.$$

From Lemmas 3.1 and 3.2 we can find, with probability one, $n_0 \in \mathbb{N}$ such that $\inf_{x\in\mathtt{X}}\{\hat{\phi}_n(x)\} > -M_0$ and $\inf_{x\in A_\alpha}\{|\hat{\phi}_n(x) - \phi(x)|\} < M_0$ for any $n \geq n_0$. Define

$$\begin{aligned} M &= M_1 + M_0 \\ K_{\mathtt{X}} &= \frac{4|b-a|}{r\min_{1\leq j\leq d}\{b^j - a^j\}}M \end{aligned}$$

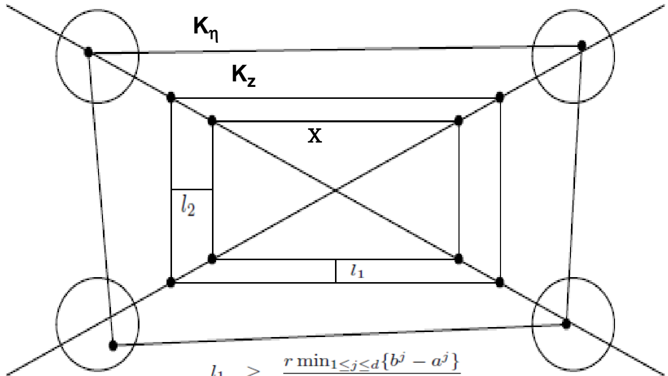

Figure 3: The smallest distance between ∂K_z and $\partial \mathtt{X}$ is at least $\frac{r \min_{1\leq j\leq d}\{b^j - a^j\}}{2|b-a|}$.

and take any $n \geq n_0$. Then, for any $\alpha \in \{-1, 1\}^d$ we can find $\eta_\alpha \in A_\alpha$ such that $|\hat{\phi}_n(\eta_\alpha) - \phi(\eta_\alpha)| < M_0$. Then, (11) implies that $\hat{\phi}_n(x) \leq M \; \forall x \in \mathtt{X}$. Take then $x \in \mathtt{X}$ and $\xi \in \partial\hat{\phi}_n(x)$. A connectedness argument, like the one used in the proof of Lemma 3.2, implies that there is $t_* > 0$ such that $x + t_*\xi \in \partial K_\eta$. But then we must have $t_* > \frac{r \min_{1\leq j\leq d}\{b^j - a^j\}}{2|\xi||b-a|}$ as a consequence of (11), since the smallest distance between ∂K_z and $\partial \mathtt{X}$ is $\frac{r \min_{1\leq j\leq d}\{b^j - a^j\}}{2|b-a|}$ and $\partial K_\eta \subset \mathrm{Ext}(K_z)$. This can be seen by taking a look at Figure 3, which shows the situation in the two dimensional case. Thus, using the definition of subgradients,

$$\frac{r \min_{1\leq j\leq d}\{b^j - a^j\}}{2|\xi||b-a|}\langle \xi, \xi\rangle \leq \langle \xi, t_*\xi\rangle \leq \hat{\phi}_n(x + t_*\xi) - \hat{\phi}_n(x) \leq 2M$$

which in turn implies $|\xi| \leq K_{\mathtt{X}}$. We have therefore shown that, with probability one, we can find $n_0 \in \mathbb{N}$ such that $|\xi| \leq K_{\mathtt{X}} \; \forall \; \xi \in \partial\hat{\phi}_n(x), \; \forall \; x \in \mathtt{X}, \; \forall \; n \geq n_0$. This completes the proof. $\square$

2 Auxiliary Lemmas

Lemma 2.1 *Let $z \in \mathbb{R}^n$, $x_1, \ldots, x_n \in \mathbb{R}^d$ and define the function $g : \mathbb{R}^d \to \overline{\mathbb{R}}$ by*

$$g(x) = \inf \left\{ \sum_{k=1}^{n} \theta^k z^k : \sum_{k=1}^{n} \theta^k = 1,\ \sum_{k=1}^{n} \theta^k x_k = x,\ \theta \geq 0,\ \theta \in \mathbb{R}^n \right\}.$$

Then, g defines a convex function whose effective domain is $Conv\,(x_1, \ldots, x_n)$. Moreover, if $\mathcal{K}_{x,z}$ is the collection of all proper convex functions ψ such that $\psi(x_j) \leq z^j$ for all $j = 1, \ldots, n$, then $g = \sup_{\psi \in \mathcal{K}_{x,z}} \{\psi\}$.

Proof: To see that g defines a convex function, for any $x \in \mathbb{R}^d$ write

$$A_x = \left\{ \theta \in \mathbb{R}^n : \sum_{k=1}^{n} \theta^k = 1,\ \sum_{k=1}^{n} \theta^k x_k = x,\ \theta \geq 0 \right\}$$

and observe that for any $x, y \in \mathbb{R}^d$, $t \in (0,1)$, $\vartheta \in A_y$ and $\theta \in A_x$ we have $t\theta + (1-t)\vartheta \in A_{tx+(1-t)y}$ and hence

$$\frac{g\,(tx + (1-t)y) - (1-t)\sum_{k=1}^{n} \vartheta^k z^k}{t} \leq \sum_{k=1}^{n} \theta^k z^k.$$

Taking infimum over A_x and rearranging terms, we get

$$\frac{g\,(tx + (1-t)y) - tg(x)}{1-t} \leq \sum_{k=1}^{n} \vartheta^k z^k$$

and taking now the infimum over A_y gives the desired convexity. The convention that $\inf(\emptyset) = +\infty$ shows that the effective domain is precisely the convex hull of $x_1, \ldots, x_n$. Finally, for any $\psi \in \mathcal{K}_{x,z}$ and $x \in Conv\,(x_1, \ldots, x_n)$ we have, for $\theta \in \mathbb{R}^n$ with $\theta \geq 0$, $x = \sum_{j=1}^{n} \theta^j x_j$ and $\sum_{j=1}^{n} \theta^j = 1$,

$$\psi(x) \leq \sum_{j=1}^{n} \theta^j \psi(x_j) \leq \sum_{j=1}^{n} \theta^j z^j$$

since $\psi(x_j) \leq z^j$ for any $j = 1, \ldots, n$. The definition of g as an infimum then implies that $\psi(x) \leq g(x)\ \forall\ \psi \in \mathcal{K}_{x,z}$, $x \in Conv\,(x_1, \ldots, x_n)$. The result then follows from the fact that $g \in \mathcal{K}_{x,z}$. $\square$

Lemma 2.2 *Let $\hat{\phi}_n$ be the least squares estimator obtained from the sample $(X_1, Y_1), \ldots, (X_n, Y_n)$. Then,*

(i) $\sum_{k=1}^{n} (\psi(X_k) - \hat{\phi}_n(X_k))(Y_k - \hat{\phi}_n(X_k)) \leq 0$ *for any convex function ψ which is finite on Conv $(X_1, \ldots, X_n)$;*

(ii) $\sum_{k=1}^{n} \hat{\phi}_n(X_k)(Y_k - \hat{\phi}_n(X_k)) = 0$*;*

(iii) $\sum_{k=1}^{n} Y_k = \sum_{k=1}^{n} \hat{\phi}_n(X_k)$*;*

(iv) the set on which $\hat{\phi}_n < \infty$ is Conv $(X_1, \ldots, X_n)$;

(v) for any $x \in \mathbb{R}^d$ the map $(X_1, \ldots, X_n, Y_1, \ldots, Y_n) \hookrightarrow \hat{\phi}_n(x)$ is a Borel-measurable function from $\mathbb{R}^{n(d+1)}$ into $\mathbb{R}$.

Proof: Property (i) follows from Moreau's decomposition theorem, which can be stated as:
Consider a closed convex set $\mathcal{C}$ on a Hilbert space $\mathcal{H}$ with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. Then, for any $x \in \mathcal{H}$ there is only one vector $x_{\mathcal{C}} \in \mathcal{C}$ satisfying $\|x - x_{\mathcal{C}}\| =$ argmin$_{\xi \in \mathcal{C}} \{\|x - \xi\|\}$. The vector $x_{\mathcal{C}}$ is characterized by being the only element of $\mathcal{C}$ for which the inequality $\langle \xi - x_{\mathcal{C}}, x - x_{\mathcal{C}} \rangle \leq 0$ holds for every $\xi \in \mathcal{C}$ (see Moreau (1962), Zeidler (1985) or Song and Zhengjun (2004)).

Taking ψ to be $\kappa \hat{\phi}_n$ and letting κ vary through $(0, \infty)$ gives (ii) from (i). Similarly, (iii) follows from (i) by letting ψ to be $\hat{\phi}_n \pm 1$. Property (iv) is obvious from the definition of $\hat{\phi}_n$.

To see why (v) holds, we first argue that the map $(X_1, \ldots, X_n, Y_1, \ldots, Y_n) \hookrightarrow Z_n$ is measurable. This follows from the fact that Z_n is the solution to a convex quadratic program and thus can be found as a limit of sequences whose elements come from arithmetic operations with $(X_1, \ldots, X_n, Y_1, \ldots, Y_n)$. Examples of such sequences are the ones produced by active set methods, e.g, see Boland (1997); or by interior-point methods (see Kapoor and Vaidya (1986) or Mehrotra and Sun (1990)). The measurability of $\hat{\phi}_n(x)$ follows from a similar argument, since it is the optimal value of a linear program whose solution can be obtained from arithmetic operations involving just $(X_1, \ldots, X_n, Y_1, \ldots, Y_n)$ and Z_n (e.g., via the well-known simplex method; see Nocedal and Wright (1999), page 372 or Luenberger (1984), page 30). $\square$

3 Proof of Lemma 3.10

Throughout this proof we will denote by $\mathbf{B}$ the unit ball (w.r.t. the euclidian norm) in $\mathbb{R}^d$. From Theorem 25.5, page 246 on Rockafellar (1970) we know that f is continuously differentiable on $\mathcal{C}$. Let

$$h_* = \inf_{\xi \in \mathtt{X}, \eta \in \mathbb{R}^d \setminus \mathcal{C}} \{|\xi - \eta|\} > 0.$$

Pick $\epsilon > 0$. We will first show that there is $n_\epsilon \in \mathbb{N}$ such that

$$\langle \xi, \eta \rangle \leq \langle \nabla f(x), \eta \rangle + \epsilon, \quad \forall\, \xi \in \partial f_n(x),\ \forall\, x \in \mathtt{X},\ \forall\, \eta \in \mathbf{B},\ \forall\, n \geq n_\epsilon. \tag{12}$$

Suppose that such an n_ϵ does not exist. Then, there is an increasing sequence $(m_n)_{n=1}^\infty$ such that for any $n \in \mathbb{N}$ we can find $x_{m_n} \in \mathtt{X}$, $\xi_{m_n} \in \partial f_{m_n}(x_{m_n})$, $\eta_{m_n} \in \mathbf{B}$ satisfying $\langle \xi_{m_n}, \eta_{m_n} \rangle > \langle \nabla f(x_{m_n}), \eta_{m_n} \rangle + \epsilon$. But $\mathtt{X}$ and $\mathbf{B}$ are both compact, so there are $x_* \in \mathtt{X}$, $\eta_* \in \mathbf{B}$ and a subsequence $(k_n)_{n=1}^\infty$ of $(m_n)_{n=1}^\infty$ such that $x_{k_n} \to x_*$ and $\eta_{k_n} \to \eta_*$. Then, for any $0 < h < h_*$ we have

$$\frac{f_{k_n}(x_{k_n} + h\eta_{k_n}) - f_{k_n}(x_{k_n})}{h} \geq \langle \xi_{k_n}, \eta_{k_n} \rangle > \langle \nabla f(x_{m_n}), \eta_{k_n} \rangle + \epsilon \quad \forall\, n \in \mathbb{N},$$

and therefore

$$\varliminf_{n \to \infty} \lim_{h \downarrow 0} \frac{f_{k_n}(x_{k_n} + h\eta_{k_n}) - f_{k_n}(x_{k_n})}{h} \geq \langle \nabla f(x_*), \eta_* \rangle + \epsilon.$$

But this is impossible in view of Theorem 24.5, page 233 on Rockafellar (1970). It follows that we can choose some $n_\epsilon \in \mathbb{N}$ with the property described in (12). By noting that $-\mathbf{B} = \mathbf{B}$, we can conclude from (12) that

$$|\langle \xi, \eta \rangle - \langle \nabla f(x), \eta \rangle| \leq \epsilon \quad \forall\, \xi \in \partial f_n(x),\ \forall\, x \in \mathtt{X},\ \forall\, \eta\ \in \mathbf{B},\ \forall\, n \geq n_\epsilon.$$

By taking $\eta_\xi = \frac{\xi - \nabla f(x)}{|\xi - \nabla f(x)|}$ when $\xi \neq \nabla f(x)$ we get

$$\sup_{\substack{x \in \mathtt{X} \\ \xi \in \partial f_n(x)}} \{|\xi - \nabla f(x)|\} \leq \epsilon \quad \forall\, n \geq n_\epsilon.$$

Since $\epsilon > 0$ was arbitrarily chosen, this completes the proof. $\square$

4 Results from matrix algebra

Before proving Lemma 1.1, we need the following result.

Lemma 4.1 *Let $j \in \{1, \ldots, d\}$, $\alpha \in \{-1, 1\}^d$ and $\rho_* > 0$. Then, the optimal value of the optimization problem*

$$\begin{array}{ll} \min & \langle \alpha^j \boldsymbol{e}_j, w_2 - w_1 \rangle \\ s.t. & \left| w_2 - \frac{3\rho_*}{8\sqrt{d}}\alpha \right| \leq \frac{\rho_*}{8\sqrt{d}} \\ & |w_1| \leq \frac{\rho_*}{16\sqrt{d}} \\ & w_1, w_2 \in \mathbb{R}^d \end{array}$$

is $\frac{3}{16\sqrt{d}}\rho_*$ and it is attained at $w_1^* = \frac{\rho_*}{16\sqrt{d}}\alpha^j \boldsymbol{e}_j$ and $w_2^* = \frac{3\rho_*}{8\sqrt{d}}\alpha - \frac{\rho_*}{8\sqrt{d}}\alpha^j \boldsymbol{e}_j$.

Proof: Writing $w = (w_1; w_2)$ with $w_1, w_2 \in \mathbb{R}^d$ for any $w \in \mathbb{R}^{2d}$, consider $f, g_1, g_2 : \mathbb{R}^{2d} \to \mathbb{R}$ defined by:

$$\begin{aligned} f(w) &= \langle \alpha^j \mathbf{e}_j, w_2 - w_1 \rangle, \\ g_1(w) &= \frac{1}{2}\left(\left(\frac{\rho_*}{16\sqrt{d}} \right)^2 - |w_1|^2 \right), \\ g_2(w) &= \frac{1}{2}\left(\left(\frac{\rho_*}{8\sqrt{d}} \right)^2 - \left| w_2 - \frac{3\rho_*}{8\sqrt{d}}\alpha \right|^2 \right). \end{aligned}$$

Then, f, g_1, g_2 are twice continuously differentiable on $\mathbb{R}^{2d}$ and the optimization problem can be re-written as minimizing $f(w)$ over the set $\{w \in \mathbb{R}^{2d} : g_1(w) \geq 0, g_2(w) \geq 0\}$. The proof now follows by noting that the vector $w^* = (w_1^*; w_2^*) \in \mathbb{R}^{2d}$ and the Lagrange multipliers $\lambda_1^* = \frac{16\sqrt{d}}{\rho_*}$ and $\lambda_2^* = \frac{8\sqrt{d}}{\rho_*}$ are the only ones which satisfy the Karush-Kuhn-Tucker second order necessary and sufficient conditions for a strict local solution to this problem as stated in Theorem 12.5, page 343 and Theorem 12.6, page 345 in Nocedal and Wright (1999). $\square$

4.1 Proof of Lemma 1.1

Without loss of generality, we may assume that $r = 1$. Let R_r be $\frac{1}{\sqrt{d}}$ and pick $\delta \in \left(0, \frac{1}{\sqrt{d}}\right)$, $\rho_* = \frac{1}{\sqrt{d}} - \delta$ and $\rho^* = \frac{2d}{1-\delta\sqrt{d}}$. Consider a matrix $Z = (z_1, \ldots, z_d) \in \mathbb{R}^{d \times d}$

with columns $z_1, \ldots, z_d \in \mathbb{R}^d$ and define the function $\tilde{\xi} : \mathbb{R}^{d\times d} \to \mathbb{R}^d$ as

$$\tilde{\xi}(Z) = \begin{vmatrix} \mathbf{e}_1 & z_2^1 - z_1^1 & \cdots & z_d^1 - z_1^1 \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{e}_d & z_2^d - z_1^d & \cdots & z_d^d - z_1^d \end{vmatrix}$$

where the bars denote the determinant and the equation is written symbolically to express that $\tilde{\xi}(Z)$ is a linear combination of the vectors $\{\mathbf{e}_j\}_{1\le j\le d}$ with the cofactor corresponding to the $(j, 1)$'th position as the coefficient of $\mathbf{e}_j$. This is a common notation for "generalized vector products"; see, for instance, Courant and John (1999), Section 2.4.b, page 187 for more details. Since the determinant and all cofactors can be seen as a continuous function on $\mathbb{R}^{d\times d}$, it follows that $\tilde{\xi}$ is continuous on $\mathbb{R}^{d\times d}$. Now choose $\alpha \in \{-1, 1\}^d$ and observe that

$$\begin{aligned} \tilde{\xi}(\alpha^1\mathbf{e}_1, \ldots, \alpha^d\mathbf{e}_d) &= \left(\prod_{j=1}^d \alpha^j\right)\alpha, \\ \left|\tilde{\xi}(\alpha^1\mathbf{e}_1, \ldots, \alpha^d\mathbf{e}_d)\right| &= \sqrt{d}, \\ \langle\tilde{\xi}(\alpha^1\mathbf{e}_1, \ldots, \alpha^d\mathbf{e}_d), \alpha^j\mathbf{e}_j\rangle &= \prod_{k=1}^d \alpha^k \quad \forall\, j = 1, \ldots, d. \end{aligned}$$

Since $\mathbb{R}^{d\times d}$ has the product topology of the d-fold topological product of $\mathbb{R}^d$ with itself, the continuity of $\tilde{\xi}$ and of $\langle\cdot,\cdot\rangle$ imply that we can find $\rho_\alpha \in \left(0, \frac{1}{\sqrt{d}} - \delta\right)$ such that if $x_j \in B(\alpha^j\mathbf{e}_j, \rho_\alpha)$ for any $j = 1, \ldots, d$, $\beta = \{x_1, \ldots, x_d\}$ and $X_\beta = (x_1, \ldots, x_d)$, then

$$\left|\left|\tilde{\xi}(X_\beta)\right| - \sqrt{d}\right| < \delta,$$

$$\left|\frac{\tilde{\xi}(X_\beta)}{|\tilde{\xi}(X_\beta)|} - \frac{\prod_{1\le j\le d}\alpha^j}{\sqrt{d}}\alpha\right| < \delta, \tag{13}$$

$$\left|\left\langle\frac{\tilde{\xi}(X_\beta)}{|\tilde{\xi}(X_\beta)|}, x_j\right\rangle - \frac{\prod_{k=1}^d \alpha^k}{\sqrt{d}}\right| < \delta \quad \forall\, j = 1, \ldots, d. \tag{14}$$

Taking this into account, define

$$\xi_{\alpha,\beta} = \left(\prod_{j=1}^d \alpha^j\right)\frac{\tilde{\xi}(X_\beta)}{|\tilde{\xi}(X_\beta)|}, \text{ and } b_{\alpha,\beta} = \langle\xi_{\alpha,\beta}, x_1\rangle.$$

From the definition of the function $\tilde{\xi}$ it is straight forward to see that $\langle \xi_{\alpha,\beta}, x_j - x_1 \rangle = 0$ $\forall j \in \{1, \dots, d\}$, so we in fact have

$$x_1, \dots, x_d \in \mathcal{H}_{\alpha,\beta} := \{x \in \mathbb{R}^d : \langle \xi_{\alpha,\beta}, x \rangle = b_{\alpha,\beta}\}.$$

Moreover, (13) and (14) imply

$$\begin{aligned} \frac{1}{\sqrt{d}} + \delta > b_{\alpha,\beta} > \frac{1}{\sqrt{d}} - \delta > 0, \\ \min_{1 \le j \le d} \left\{ |\xi^j_{\alpha,\beta}| \right\} > \frac{1}{\sqrt{d}} - \delta > 0. \end{aligned}$$

For simplicity, and without loss of generality (the other cases follow from symmetry), we now assume that $\alpha = \mathbf{e}$, the vector of ones. By solving the corresponding quadratic programming problems, it is not difficult to see that

$$\begin{aligned} \rho_* = \frac{1}{\sqrt{d}} - \delta < b_{\alpha,\beta} &= \inf_{\langle \xi_{\alpha,\beta}, x \rangle \ge b_{\alpha,\beta}} \{|x|\} \\ \rho^* = \frac{2d}{1 - \delta\sqrt{d}} > \frac{b_{\alpha,\beta}}{\min_{1 \le j \le d}\{|\xi^j_{\alpha,\beta}|\}} &= \sup_{\substack{\langle \xi_{\alpha,\beta}, x \rangle \le b_{\alpha,\beta} \\ x \ge 0}} \{|x|\}. \end{aligned}$$

For the first inequality see, for instance, Exercise 16.2, page 484 of Nocedal and Wright (1999). For the second one, one must notice that $2\sqrt{d} > \frac{1}{\sqrt{d}} + \delta > b_{\alpha,\beta}$ and that the optimal value of the optimization problem must be attained at one of the vertices of the polytope $\{x \in \mathbb{R}^d_+ : \langle \xi_{\alpha,\beta}, x \rangle \le b_{\alpha,\beta}\}$. The latter statement can be derived from the Karush-Kuhn-Tucker conditions of the problem.

The inequalities in the last display imply that $B(0, \rho_*) \subset \mathcal{H}^-_{\alpha,\beta}$ and $\{x \in \mathbb{R}^d : |x| \ge \rho^*\} \cap \mathcal{R}_\alpha \subset \mathcal{H}^+_{\alpha,\beta}$.

Finally, for $x \in B(-\alpha^j \mathbf{e}_j, \frac{1}{2}\rho_\alpha)$ we have $|x + x_j| < \rho_\alpha$ and therefore $\langle \xi_{\alpha,\beta}, x \rangle < -\langle \xi_{\alpha,\beta}, x_j \rangle + \rho_\alpha < \delta - \frac{1}{\sqrt{d}} + \rho_\alpha < 0$. We can then take any $\rho \le \frac{1}{2} \min_{\alpha \in \{-1,1\}^d} \{\rho_\alpha\}$ to make (i)-(vi) be true. We'll now argue that by making ρ smaller, if required, (vii) also holds.

Let $B_1 = B\left(0, \frac{\rho_*}{16\sqrt{d}}\right)$, $B_2 = B\left(\frac{3\rho_*}{8\sqrt{d}}\alpha, \frac{\rho_*}{8\sqrt{d}}\right)$ and consider the functions $\varphi, \psi : \mathbb{R}^{d \times d} \to \mathbb{R}$ given by

$$\begin{aligned} \varphi(X) &= \inf_{w_1 \in B_1, w_2 \in B_2} \left\{ \min_{1 \le j \le d} \left\{ (X(w_2 - w_1))^j \right\} \right\}, \\ \psi(X) &= \sup_{w_1 \in B_1} \left\{ \max_{1 \le j \le d} \left\{ (X w_1)^j \right\} \right\}. \end{aligned}$$

Both of these functions are Lipschitz continuous with the metric induced by the $\|\cdot\|_2$-norm on $\mathbb{R}^{d\times d}$ with Lipschitz constants smaller than ρ_*. To see this, observe that

$$|X(w_2-w_1)-Y(w_2-w_1)| \le \|X-Y\|_2|w_2-w_1| \le \frac{9}{16}\rho_*\|X-Y\|_2$$

for all $w_1 \in B_1$, $w_2 \in B_2$ and $X, Y \in \mathbb{R}^{d\times d}$. Also, simple algebra shows that $|\min_{1\le j\le d}\{x^j\} - \min_{1\le j\le d}\{y^j\}| \le |x-y| \ \forall\ x, y \in \mathbb{R}^d$. From these assertions, one immediately gets the Lipschitz continuity of φ. Similar arguments show the same for ψ.

Let $\mathcal{I}_\alpha \in \mathbb{R}^{d\times d}$ be the diagonal matrix whose j'th diagonal element is precisely α^j. From Lemma 4.1 it is seen that $\varphi(\mathcal{I}_\alpha) = \frac{3\rho_*}{16\sqrt{d}}$. On the other hand, it is immediately obvious that $\psi(\mathcal{I}_\alpha) = \frac{\rho_*}{16\sqrt{d}}$. Using one more time the continuity of ψ and φ and that the topology in $\mathbb{R}^{d\times d}$ is the same as the topology of the d-fold topological product of $\mathbb{R}^d$, for each $\alpha \in \{-1,1\}^d$ we can find r_α for which $X_\beta = (x_1,\ldots,x_d) \in \mathbb{R}^{d\times d}$ and $|x_j - \alpha^j\mathbf{e}_j| < r_\alpha$ for all $j = 1,\ldots,d$ imply $|\psi(X_\beta^{-1}) - \frac{\rho_*}{16\sqrt{d}}| < \frac{\rho_*}{32\sqrt{d}}$ and $|\varphi(X_\beta^{-1}) - \frac{3\rho_*}{16\sqrt{d}}| < \frac{\rho_*}{16\sqrt{d}}$. It follows that

$$\begin{aligned}
&\inf_{\substack{t\ge 1\\ w_1\in B_1, w_2\in B_2}} \left\{\min_{1\le j\le d}\left\{\left(X_\beta^{-1}(w_1+t(w_2-w_1))\right)^j\right\}\right\} \\
\ge\ &\inf_{\substack{t\ge 1\\ w_1\in B_1, w_2\in B_2}} \left\{\min_{1\le j\le d}\left\{\left(tX_\beta^{-1}(w_2-w_1)\right)^j\right\}\right\} - \sup_{w_1\in B_1}\left\{\max_{1\le j\le d}\left\{\left(X_\beta^{-1}w_1\right)^j\right\}\right\} \\
\ge\ &\varphi(X_\beta^{-1}) - \psi(X_\beta^{-1}) \ > \ \frac{\rho_*}{8\sqrt{d}} - \frac{3\rho_*}{32\sqrt{d}} \ = \ \frac{\rho_*}{32\sqrt{d}} > 0.
\end{aligned}$$

The proof is then finished by taking $\rho \le \min_{\alpha\in\{-1,1\}^d}\left\{r_\alpha \wedge \frac{\rho_\alpha}{2}\right\}$. $\square$

4.2 Proof of Lemma 1.2

Assume again, without loss of generality, that $r = 1$. Lemma 1.1 (ii) and (vi) imply that $x_{\alpha^j j}, x_{-\alpha^j j} \in \{x \in \mathbb{R}^d : \langle x, \xi_\alpha\rangle \le b_\alpha\}$ for any $j = 1,\ldots,n$ and any $\alpha \in \{-1,1\}^d$. It follows that, in addition to being convex, $\cap_{\alpha\in\{-1,1\}^d}\{x \in \mathbb{R}^d : \langle \xi_\alpha, x\rangle \le b_\alpha\}$ contains $\{x_{\pm 1},\ldots,x_{\pm d}\}$ and hence it must contain K. For the other contention, take $x \in \cap_{\alpha\in\{-1,1\}^d}\{w \in \mathbb{R}^d : \langle \xi_\alpha, w\rangle \le b_\alpha\}$ with $x \ne 0$ and any $\alpha \in \{-1,1\}^d$ for which $x \in \mathcal{R}_\alpha$. Then, $\langle \xi_\alpha, x\rangle > 0$ for otherwise we would have

$$\kappa x \in \mathcal{R}_\alpha \setminus \mathcal{H}_\alpha^+ \ \ \forall\ \kappa \ge 0$$

which is impossible by (v) in Lemma 1.1. Thus, $\mathcal{J}_x = \{\alpha \in \{-1,1\}^d : \langle \xi_\alpha, x\rangle > 0\} \neq \emptyset$ and we can define

$$r_x = \min_{\alpha\in\mathcal{J}_x}\left\{\frac{b_\alpha}{\langle \xi_\alpha, x\rangle}\right\} \quad \text{and } \alpha_x = \operatorname*{argmin}_{\alpha\in\mathcal{J}_x}\left\{\frac{b_\alpha}{\langle \xi_\alpha, x\rangle}\right\}.$$

Note that $r_x \geq 1$. Since β_{α_x} is a basis, there is $\theta \in \mathbb{R}^d$ such that $r_x x = \theta^1 x_{\alpha_x^1 1} + \ldots + \theta^d x_{\alpha_x^d d}$. But then,

$$b_{\alpha_x} = \langle r_x x, \xi_{\alpha_x}\rangle = \sum_{k=1}^d \theta^k \langle x_{\alpha_x^k k}, \xi_{\alpha_x}\rangle = b_{\alpha_x}\sum_{k=1}^d \theta^k$$

where the last equality follows from (ii) of Lemma 1.1 and therefore $\theta^1 + \ldots + \theta^d = 1$. Now assume that $\theta^j < 0$ for some $j \in \{1,\ldots,d\}$ and set $\gamma_x \in \{-1,1\}^d$ with $\gamma_x^k = \alpha_x^k$ for $k \neq j$ and $\gamma_x^j = -\alpha_x^j$. But then, $\sum_{k\neq j}\theta^k = 1 - \theta^j > 1$, $\langle x_{\alpha_x^k k}, \xi_{\gamma_x}\rangle = b_{\gamma_x}$ for $k \neq j$ and $\langle x_{\alpha_j^j j}, \xi_{\gamma_x}\rangle < 0$ by (ii) and (vi) in Lemma 1.1. Therefore,

$$\begin{aligned}\langle r_x x, \xi_{\gamma_x}\rangle &= \theta^j \langle x_{-\alpha_x^j j}, \xi_{\gamma_x}\rangle + \sum_{k\neq j}\theta^k\langle x_{\alpha_x^k k}, \xi_{\gamma_x}\rangle && (15)\\ &> \sum_{k\neq j}\theta^k\langle x_{\alpha_x^k k}, \xi_{\gamma_x}\rangle > b_{\gamma_x} && (16)\end{aligned}$$

which is impossible because it contradicts the definition of r_x. Hence, $\theta \geq 0$ and we have $r_x x \in Conv(\beta_{\alpha_x})$. Note that since 0 belongs in the interior of $\cap_{\alpha\in\{-1,1\}^d}\{w \in \mathbb{R}^d : \langle \xi_\alpha, w\rangle \leq b_\alpha\}$, there there is $\kappa > 0$ such that $-\kappa x \in \cap_{\alpha\in\{-1,1\}^d}\{w \in \mathbb{R}^d : \langle \xi_\alpha, w\rangle \leq b_\alpha\}$. Applying the same arguments as before to $-\kappa x$ instead of x, we can find $\tilde{r}_x > 0$ and $\tilde{\alpha}_x \in \{-1,1\}^d$ such that $-\tilde{r}_x x \in Conv(\beta_{\tilde{\alpha}_x})$. It follows that $-\tilde{r}_x x, r_x x \in K$ and therefore $0, x \in K$ since $r_x \geq 1$. Hence, we have proved (i).

To prove (ii), note that $A := \cap_{\alpha\in\{-1,1\}^d}\{w \in \mathbb{R}^d : \langle \xi_\alpha, w\rangle < b_\alpha\}$ is open and, by (i), it is contained in K. Thus, $A \subset K^\circ$. That $K^\circ \subset A$ follows from the fact that if $x \in K \setminus A$, then $\langle \xi_\alpha, x\rangle = b_\alpha$ for some $\alpha \in \{-1,1\}^d$, which implies that $B(x,\tau) \cap \operatorname{Ext}(K) \neq \emptyset$ for all $\tau > 0$ and hence $x \notin K^\circ$.

It is then obvious that (iv) follows from the identity $\partial K = \overline{K} \setminus K^\circ$ and the fact that K is closed.

Pick any $\alpha \in \{-1,1\}^d$ and observe that (ii) and (vi) from Lemma 1.1 imply that for any $\gamma \in \{-1,1\}^d$ we have

$$\langle \xi_\gamma, x_{\alpha^k k}\rangle \begin{cases} = b_\gamma & \text{if } \gamma^k = \alpha^k \\ < 0 \leq b_\gamma & \text{if } \gamma^k = -\alpha^k \end{cases}$$

which by (iv) of this lemma show that

$$x_{\alpha^j j} \in \{w \in \mathbb{R}^d : \langle \xi_\alpha, w\rangle = b_\alpha\} \cap \left(\cap_{\gamma\in\{-1,1\}^d}\{w \in \mathbb{R}^d : \langle \xi_\gamma, w\rangle \le b_\gamma\}\right)$$

for all $\alpha \in \{-1,1\}^d$ and $j = 1,\ldots,d$. Since the sets on the right-hand side of the last display are all convex we can conclude that

$$Conv\left(x_{\alpha^1 1},\ldots,x_{\alpha^j j}\right) \subset \{w \in \mathbb{R}^d : \langle \xi_\alpha, w\rangle = b_\alpha\} \cap \left(\cap_{\gamma\in\{-1,1\}^d}\{w \in \mathbb{R}^d : \langle \xi_\gamma, w\rangle \le b_\gamma\}\right)$$

for all $\alpha \in \{-1,1\}^d$. Thus, $\bigcup_{\alpha\in\{-1,1\}^d} Conv\left(x_{\alpha^1 1},\ldots,x_{\alpha^j j}\right) \subset \partial K$. Finally, take $x \in \partial K$. Then, there is $\alpha_x \in \{-1,1\}^d$ such that $\langle \xi_{\alpha_x}, x\rangle = b_{\alpha_x}$. Since β_{α_x} is a basis we can again find $\theta \in \mathbb{R}^d$ such that $x = \theta^1 x_{\alpha_x^1 1} + \ldots + \theta^d x_{\alpha_x^d d}$. Just as before, $\langle \xi_{\alpha_x}, x_{\alpha_x^j j}\rangle = b_{\alpha_x}$ implies that $\sum \theta^j = 1$. And again, if $\theta^j < 0$ for some j, we can take $\gamma_x \in \{-1,1\}^d$ with $\gamma_x^k = \alpha_x^k$ for $k \neq j$ and $\gamma_x^j = -\alpha_x^j$ and arrive at a contradiction with similar arguments to those used in (15) and (16). This shows that $x \in Conv\left(\beta_{\alpha_x}\right)$ and completes the proof as (v) and (vi) are direct consequences of $(i)-(iv)$ and Lemma 1.1. □

4.3 Proof of Lemma 1.3

Let $r \in (0, \frac{1}{d-2})$ if $d \ge 3$ and $r > 0$ if $d \le 2$. Since the geometric properties of any rectangle depend only on the direction and magnitude of the diagonal, we may assume without loss of generality that $b > 0$ and that $a = \frac{r}{1+r}b$. This is because we can define $\tilde{b} = (1+r)(b-a) > 0$ and $\tilde{a} = a - r(b-a)$ to obtain $[a,b] = \tilde{a} + \left[\frac{r}{r+1}\tilde{b}, \tilde{b}\right]$. For any $\alpha \in \{-1,1\}^d$, define $\alpha_j = \alpha - 2\alpha^j \mathbf{e}_j \in \mathbb{R}^d$ and $w_\alpha = z_\alpha + r(z_\alpha - z_{-\alpha})$. Additionally, define the functions $\psi_\alpha, \varphi_\alpha : \mathbb{R}^{d\times d} \times \mathbb{R}^d \to \mathbb{R}$ by

$$\begin{aligned}\psi_\alpha(\Theta,\theta) &= \langle \mathbf{e}, \Theta(z_\alpha - \theta)\rangle \\ \varphi_\alpha(\Theta,\theta) &= \min_{1\le j\le d}\left\{(\Theta(z_\alpha - \theta))^j\right\}.\end{aligned}$$

Considering $\mathbb{R}^{d\times d}$ with the topology generated be the $\|\cdot\|_2$ norm and $\mathbb{R}^{d\times d}\times\mathbb{R}^d$ with the product topology, it is easily seen that both functions defined in the last display are continuous. Now, let $W_\alpha \in \mathbb{R}^{d\times d}$ be the matrix whose j'th column is precisely $w_{\alpha_j} - w_\alpha$. It is not difficult to see that $\psi_\alpha(W_\alpha^{-1}, w_\alpha) = \frac{dr}{1+2r} < 1$ and $\varphi_\alpha(W_\alpha^{-1}, w_\alpha) = \frac{r}{1+2r} > 0$. For instance, one can check that for $\alpha = -\mathbf{e}$, one has $w_\alpha = 0$ and $w_{\alpha_j} = \frac{1+2r}{1+r}b^j \mathbf{e}_j$ and the result is now evident. By symmetry, the same is true for any $\alpha \in \{-1,1\}^d$. Therefore, for any $\alpha \in \{-1,1\}^d$ there is ρ_α such that whenever $|x_{\alpha_j} - w_{\alpha_j}| < \rho_\alpha \; \forall \; j = 1, \ldots, n$ and X_α is the matrix whose j'th column is $x_{\alpha_j} - x_\alpha$, we get

$$\psi_\alpha(X_\alpha^{-1}, x_\alpha) \;<\; 1, \tag{17}$$
$$\varphi_\alpha(X_\alpha^{-1}, x_\alpha) \;>\; 0. \tag{18}$$

Letting $\rho = \min_{\alpha \in \{-1,1\}^d} \{\rho_\alpha\}$ completes the proof as (17) and (18) imply $z_\alpha \in Conv\,(x_\alpha, x_{\alpha_1}, \ldots, x_{\alpha_d})^\circ$. □

References

Boland, N. L. (1997). A dual-active-set algorithm for positive semi-definite quadratic programming. *Math. Program.*, 78:1–27.

Courant, R. and John, F. (1999). *Introduction to Calculus and Analysis, Vol. II/1.* Springer, New York, NY, USA.

Kapoor, S. and Vaidya, P. M. (1986). Fast algorithms for convex quadratic programming and multicommodity flows. *Proceedings of the eighteenth annual ACM symposium on Theory of computing*, pages 147–159.

Luenberger, D. (1984). *Linear and Nonlinear Programming.* Addison-Wesley Publishing Company, Reading, MA, USA.

Mehrotra, S. and Sun, J. (1990). An algorithm for convex quadratic programming that requires $o(n^{3.5}l)$ arithmetic operations. *Math. Oper. Res.*, 15:342–363.

Moreau, J. (1962). Decomposition orthogonal d'un espace hilbertien selon deux cones mutuellement polaires. *C. R. Acad. Sci. Paris*, 255:238–240.

Nocedal, J. and Wright, S. (1999). *Numerical Optimization.* Springer, New York, NY, USA.

Rockafellar, T. R. (1970). *Convex Analysis.* Princeton University Press, Princeton, NJ, USA.

Song, W. and Zhengjun, C. (2004). The generalized decomposition theorem in banach spaces and its applications. *J. Approx. Theory*, 129:167–181.

Zeidler, E. (1985). *Nonlinear Functional Analysis and its Applications III.* Springer-Verlag, New York, NY.