Chapter 1

Some Basic Geometry

This chapter reviews some material we will need during the course and tries to get you acquainted with some unusual phenomena occurring in high dimensions.

1.1 Affine Geometry

We will assume that you are familiar with the basic notions of linear algebra, such as vector spaces (a.k.a. linear spaces) and linear subspaces, linear dependence/independence, dimension, linear maps, and so forth. For the most part, we will work with the d-dimensional real vector space $\mathbb{R}^d$.

If we think of $\mathbb{R}^d$ as vector space, then the *origin* $\mathbf{0} = (0, \ldots, 0)$ plays a distinguished role. If we are studying a problem that is invariant under translations, it is often more natural to work in the setting of "affine geometry".

A subset $A \subseteq \mathbb{R}^d$ is called an *affine subspace* if either $A = \emptyset$ or A is a "shifted" or "translated" linear subspace, i.e., $A = v + L$, where L is a linear subspace and $v \in \mathbb{R}^d$. Note that L is uniquely determined by A (why?), but generally v is not. The dimension $\dim A$ is defined as -1 if $A = \emptyset$ and as $\dim L$ otherwise. If $p_1, \ldots, p_n \in \mathbb{R}^d$, $n \geq 1$, and if $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ are real coefficients with $\sum_{i=1}^n \lambda_i = 1$, then $\lambda_1 p_1 + \ldots + \lambda_n p_n$ is called an *affine combination* of the p_i's. (Thus, an affine combination is a linear combination such that the coefficients sum to 1.) The *affine hull* of an arbitary subset $S \subseteq \mathbb{R}^d$ is defined as the set of all affine combinations of points in S,

$$\mathrm{aff}(S) := \{\lambda_1 p_1 + \ldots + \lambda_n p_n : n \geq 1, p_1, \ldots, p_n \in S, \lambda_1, \ldots, \lambda_n \in \mathbb{R}, \sum_{i=1}^n \lambda_i = 1\}.$$

The affine hull $\mathrm{aff}(S)$ is the smallest affine subspace containing S. An affine dependence between points $p_1, \ldots, p_n \in \mathbb{R}^d$ is a linear dependence $\alpha_1 p_1 + \ldots + \alpha_n p_n = \mathbf{0}$ (where $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$) such that $\sum_{i=1}^n \alpha_i = 0$. (Thus, in an affine combination, the coefficients sum to 1, while in an affine dependence, they sum to 0.) The affine dependence is called *nontrivial* if there is some i with $\alpha_i \neq 0$. The points $p_1, \ldots, p_n$ are called *affinely dependent* if there exists a nontrivial affine dependence between them, and *affinely independent* otherwise.

It is easy to show (do it!) that points $p_1, \ldots, p_n \in \mathbb{R}^d$ are affinely independent iff the differences $p_2 - p_1, \ldots, p_n - p_1$ between one of them and all the others are linearly independent vectors (of course, instead of p_1, we could chose any p_i as the base point).

An *affine map* $f : \mathbb{R}^d \to \mathbb{R}^k$ is one that can be expressed as the combination of a linear map and a translation. Thus, in coordinates, f can be written as $f(x) = Ax + b$, where A is a real $(k \times d)$-matrix and $b \in \mathbb{R}^k$. The composition of affine maps is again an affine map.

The space $\mathbb{R}^d$ itself leads some kind of double existence. If we think of it as a vector space, we refer to its elements as vectors, and if we think of $\mathbb{R}^d$ as an affine space, we refer to its elements as points. Often, it is suggested to use the notion of points as the primitive one and to speak of a vector when we think of the oriented difference $p - q$ between two points. At any rate, it is often convenient not to distinguish too carefully between the two viewpoints.

We remark that apart from the origin $\mathbf{0}$, there is another a special point/vector that we will use so frequently that it is worthwhile to introduce a special notation: $\mathbf{1} := (1, \ldots, 1)$, where we assume (as in the case of the origin $\mathbf{0}$) that the dimension is clear from the context.

1.2 Euclidean Space

We often associate a further piece of structure with $\mathbb{R}^d$, the *scalar product*. For $v = (v_1, \ldots, v_d)$ and $w = (w_1, \ldots, w_d) \in \mathbb{R}^d$, it is denoted by $\langle v, w \rangle$ or by $v \cdot w$ (both notations have their advantages, so we will take the liberty of sometimes using one, sometimes the other). At any rate, no matter which notation we chose, the scalar product is defined by

$$\langle v, w \rangle := v \cdot w := \sum_{i=1}^{d} v_i w_i.$$

The scalar product is *symmetric* (i.e., $\langle v, w \rangle = \langle w, v \rangle$) and *bilinear* (i.e., $\langle \alpha u + \beta v, w \rangle = \alpha \langle u, w \rangle + \beta \langle v, w \rangle$ and $\langle u, \alpha v + \beta w \rangle = \alpha \langle u, v \rangle + \beta \langle u, w \rangle$ for vectors $u, v, w \in \mathbb{R}^d$ and scalars $\alpha, \beta \in \mathbb{R}$) and *nondegenerate* (if $v \in \mathbb{R}^d$ satisfies $\langle v, w \rangle = 0$ for all $w \in \mathbb{R}^d$, then $v = \mathbf{0}$). Moreover, it is *nonnegative* in the sense that $\langle v, v \rangle \geq 0$ for all $v \in \mathbb{R}^d$, and $\langle v, v \rangle = 0$ iff $v = \mathbf{0}$. This last property implies that we can use the scalar product to define the *length* of a vector,

$$\|v\| := \sqrt{\langle v, v \rangle}, \quad v \in \mathbb{R}^d.$$

This length satisfies the following properties: For all $v, w \in \mathbb{R}^d$ and $\lambda \in \mathbb{R}$,

1. $\|v\| \geq 0$, and $\|v\| = 0$ iff $v = \mathbf{0}$.
2. $\|\lambda v\| = |\lambda|\|v\|$ (where "$|\cdot|$" denotes absolute value).
3. **Triangle Inequality.** $\|v + w\| \leq \|v\| + \|w\|$.

A measure of length of vectors that satisfies these three properties is called a *norm*. The norm defined as above using the scalar product is called *Euclidean norm* or *2-norm*. In Chapters 2 and 3, we will also study other norms on $\mathbb{R}^d$, and in order to distinguish the Euclidean norm, we will denote it by $\|v\|_2$ in those chapters. In the other chapters, however, when no confusion can arise, we use the simpler notation $\|\cdot\|$. We speak of the *d-dimensional Euclidean space* when we think of $\mathbb{R}^d$ equipped with the scalar product $\langle , \rangle$ and the induced Euclidean norm $\|\cdot\| = \|\cdot\|_2$.

The third property above is called the triangle inequality because it says that in a triangle with vertices p, q, and r, the length of any one of the sides, say $\|q - p\|$, is at most the sum of the lengths of the other two, $\|q - p\| \leq \|q - r\| + \|r - p\|$, see Figure 1.1.

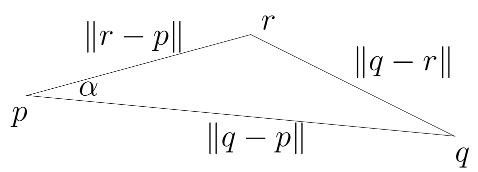

Figure 1.1: The triangle inequality and angles

In the case of the Euclidean norm, the triangle inequality follows from the fundamental

Fact 1.1 (Cauchy-Schwarz Inequality).

$$|\langle v, w\rangle| \leq \|v\|\|w\|, \quad v, w \in \mathbb{R}^d.$$

It is sometimes useful to know when equality holds in the Cauchy-Schwarz Inequality: $|\langle v, w\rangle| = \|v\|\|w\|$ iff "v and w point in the same direction", i.e., iff $v = \lambda w$ or $w = \lambda v$ for some $\lambda \geq 0$.

The Cauchy-Schwarz inequality also allows us to define the *angle* (more precisely, the "smaller angle") α between nonzero vectors $v, w \in \mathbb{R}^d$ by

$$\cos(\alpha) = \frac{\langle v, w\rangle}{\|v\|\|w\|}.$$

In the case $v = q - p$ and $w = r - p$ ee Figure 1.1. We also frequently need the

Fact 1.2 (Cosine Theorem). *For $p, q, r \in \mathbb{R}^d$ and α the angle between $q - p$ and $r - p$,*

$$\|q - r\|^2 = \|r - p\|^2 + \|q - p\|^2 - 2\|r - p\|\|q - p\| \cos(\alpha).$$

For $\alpha = \pi/2$ (or 90^o), this is Pythagoras' Theorem.

1.3 Hyperplanes

A *hyperplane* is an affine subspace of *codimension* 1 of $\mathbb{R}^d$. A hyperplane h is the solution set of one inhomogeneous linear equation,

$$h = \{x \in \mathbb{R}^d : \langle a, x \rangle = \alpha\}, \tag{1.1}$$

where $a = (a_1, \ldots, a_d) \in \mathbb{R}^d$, $a \neq \mathbf{0}$, and $\alpha \in \mathbb{R}$. We will also use the abbreviated notation

$$h = \{\langle a, x \rangle = \alpha\}.$$

(Note that for $a = \mathbf{0}$, the set of solutions to the equation $\langle a, x \rangle = \alpha$ is either all of $\mathbb{R}^d$, namely if $\alpha = 0$, or empty.) For $d = 2$, hyperplanes are *lines* (see Figure 1.2), and for $d = 3$, we get *planes*.

The vector a is the so-called *normal vector* of h. It is orthogonal to the hyperplane in the sense that

$$\langle a, p - q \rangle = 0, \qquad \text{for all } p, q \in h,$$

a fact that immediately follows from (1.1). It is not hard to show (do it!) that the distance of h to the origin is $|\alpha|/\|a\|$, attained by the unique point $\frac{\alpha}{\|a\|^2} a$. Observe that the hyperplane h is invariant under rescaling its defining equation, i.e., under multiplying both a and α by the same nonzero scalar $\lambda \neq 0$.

Any hyperplane defines a partition of $\mathbb{R}^d$ into three parts: the hyperplane h itself and two *open halfspaces*. If h is given by an equation as in (1.1), we denote these halfspaces by

$$\begin{aligned} h^+ &:= \{x \in \mathbb{R}^d : \langle a, x \rangle > \alpha\}, \\ h^- &:= \{x \in \mathbb{R}^d : \langle a, x \rangle < \alpha\}, \end{aligned}$$

and call them the *positive* and *negative open halfspace*, respectively, and if we want to stress this, we call a the *outer normal vector*. Observe that which of the halfspaces is positive and which is negative is not determined by the hyperplane but involves an additional choice, which is sometimes called a *coorientation* of h. If we rescale the defining equation by a negative scalar $\lambda < 0$, then we change the coorientation, i.e., the positive and the negative halfspace swap their roles.

We will also work with the *closed* halfspaces $\overline{h^+} := \{\langle a, x \rangle \geq \alpha\}$ and $\overline{h^-} := \{\langle a, x \rangle \leq \alpha\}$

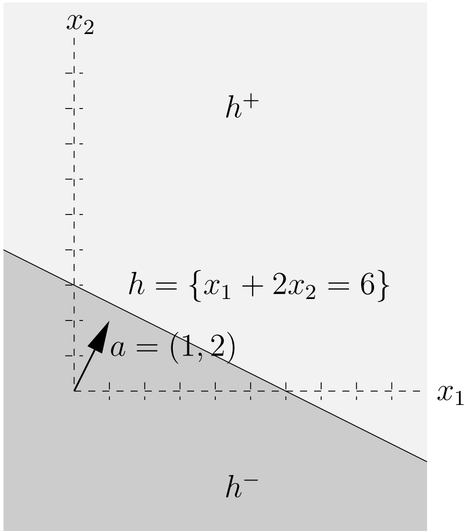

Figure 1.2: A hyperplane h in $\mathbb{R}^2$ along with its two halfspaces

Origin-avoiding hyperplanes. In the following, we will adapt the convention that for hyperplanes that do not contain the origin $\mathbf{0}$, we will choose the coorientation so that $\mathbf{0} \in h^-$. Note that $\mathbf{0} \notin h$ iff $\alpha \neq 0$, and our convention amounts assuming that $\alpha > 0$ (which we can always achieve by rescaling, if necessary).

Non-vertical hyperplanes. Sometimes it is convenient to distinguish one direction, usually the x_d-direction, as *vertical*. Hyperplanes h with $a_d \neq 0$ are called *non-vertical* and have an alternative definition in terms of only d parameters: if $h = \{a_1x_1 + \ldots + a_dx_d = \alpha\}$ with $a_d \neq 0$, then we can rewrite the defining equation as

$$x_d = -\frac{1}{a_d}(a_1x_1 + \ldots + a_{d-1}x_{d-1} - \alpha = b_1x_1 + \ldots + b_{d-1}x_{d-1} + \beta,$$

where $b_i = -a_i/a_d$, $1 \leq i \leq d-1$, and $\beta = -\alpha/a_d$. (In other words, we can view h as the graph of an affine map $\mathbb{R}^{d-1} \to \mathbb{R}$.) In this form, the line from Figure 1.2 has the equation

$$x_2 = -\frac{1}{2}x_1 + 3.$$

For non-vertical hyperplanes, we adapt the convention that the coorientation is chosen in such a way that

$$\begin{aligned}
h^+ &= \{x \in \mathbb{R}^d : x_d > \sum_{i=1}^{d-1} b_i x_i - \beta\}, \\
h^- &= \{x \in \mathbb{R}^d : x_d < \sum_{i=1}^{d-1} b_i x_i - \beta\},
\end{aligned}$$

and we say that h^+ is the halfspace *above* h, while h^- is *below* h.

1.4 Duality

In a sense, points and hyperplanes behave in the same way. Even if it is not clear what exactly this means, the statement may appear surprising at first sight. Here are two *duality transforms* that map points to hyperplanes and vice versa, in such a way that relative positions of points w.r.t. hyperplanes are preserved.

The origin-avoiding case. For $p = (p_1, \ldots, p_d) \in \mathbb{R}^d \setminus \{\mathbf{0}\}$, the origin-avoiding hyperplane

$$p^* = \{x \in \mathbb{R}^d \ \langle p, x\rangle = 1\} \tag{1.2}$$

is called the hyperplane *dual to* p. Vice versa, for an origin-avoiding hyperplane $h = \{x \in \mathbb{R}^d : \langle a, x\rangle = \alpha\}$, $\alpha \neq 0$, the point

$$h^* = \left(\frac{a_1}{\alpha}, \ldots, \frac{a_d}{\alpha}\right) \in \mathbb{R}^d \setminus \{\mathbf{0}\} \tag{1.3}$$

is called the point *dual to* h. We get $(p^*)^* = p$ and $(h^*)^* = h$, so this duality transform is an *involution* (a mapping satisfying $f(f(x)) = x$ for all x).

It follows from the above facts about hyperplanes that p^* is orthogonal to p and has distance $1/\|p\|$ from the origin. Thus, points close to the origin are mapped to hyperplanes far away, and vice versa. p is actually on p^* if and only if $\|p\| = 1$, i.e. if p is on the so-called *unit sphere*, see Figure 1.3.

The important fact about the duality transform is that relative positions of points w.r.t. hyperplanes are maintained.

Lemma 1.3. *For all points $p \neq \mathbf{0}$ and all origin-avoiding hyperplanes h, we have*

$$p \in \left\{ \begin{array}{c} h^+ \\ h^- \\ h \end{array} \right. \quad \Leftrightarrow \quad h^* \in \left\{ \begin{array}{c} (p^*)^+ \\ (p^*)^- \\ p^* \end{array} \right.$$

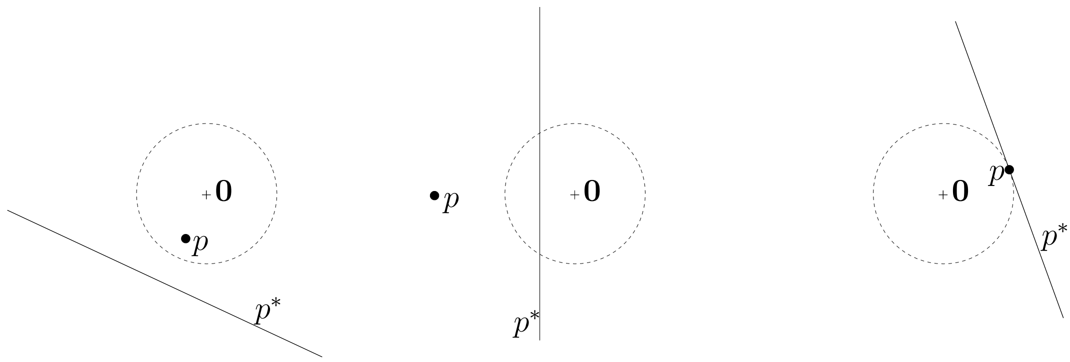

Figure 1.3: Duality in the origin-avoiding case

Proof. Really boring, but still useful in order to see what happens (or rather, that nothing happens). Let's look at h^+, the other cases are the same.

$$p \in h^+ \Leftrightarrow \sum_{i=1}^{d} a_i p_i > \alpha \Leftrightarrow \sum_{i=1}^{d} p_i \frac{a_i}{\alpha} \geq 1 \Leftrightarrow: h^* \in (p^*)^+.$$

□

The non-vertical case. The previous duality has two kinds of singularities: it does not work for the point $p = \mathbf{0}$, and it does not work for hyperplanes containing $\mathbf{0}$. The following duality has only one kind of singularity: it does not work for vertical hyperplanes, but it works for *all* points.

For $p = (p_1, \ldots, p_d) \in \mathbb{R}^d$, the non-vertical hyperplane

$$p^* = \{x \in \mathbb{R}^d : x_d = \sum_{i=1}^{d-1} p_i x_i - p_d\} \tag{1.4}$$

is called the hyperplane *dual to* p.[1] Vice versa, given a non-vertical hyperplane $h = \{x_d = \sum_{i=1}^{d-1} b_i x_i - \beta\}$, the point

$$h^* = (b_1, \ldots, b_{d-1}, \beta) \tag{1.5}$$

is called the point *dual to* h. Here is the analogue of Lemma 1.3.

Lemma 1.4. *For all points p and all non-vertical hyperplanes h, we have*

$$p \in \begin{cases} h^+ \\ h^- \\ h \end{cases} \Leftrightarrow \quad h^* \in \begin{cases} (p^*)^+ \\ (p^*)^- \\ p^* \end{cases}$$

[1] We could use another symbol to distinguish this from the previous duality, but since we never mix both dualities, it will always be clear to which one we refer.

We leave the proof as an exercise. It turns out that this duality has a geometric interpretation involving the *unit paraboloid* instead of the unit sphere [3]. Which of the two duality transforms is more useful depends on the application.

Duality allows us to translate statements about hyperplanes into statements about points, and vice versa. Sometimes, the statement is easier to understand after such a translation. Exercise 6 gives a nontrivial example. Here is one very easy translation in the non-vertical case. In the origin-avoiding case, the essence is the same, but the precise statement is slightly different (Exercise 7).

Observation 1.5. *Let p, q, r be points in $\mathbb{R}^2$. The following statements are equivalent, see Figure 1.4.*

(i) The points p, q, r are collinear *(lie on the common line ℓ).*

(ii) The lines p^*, q^*, r^* are concurrent *(go through the common point ℓ^*), or are* parallel *to each other, if ℓ is vertical).*

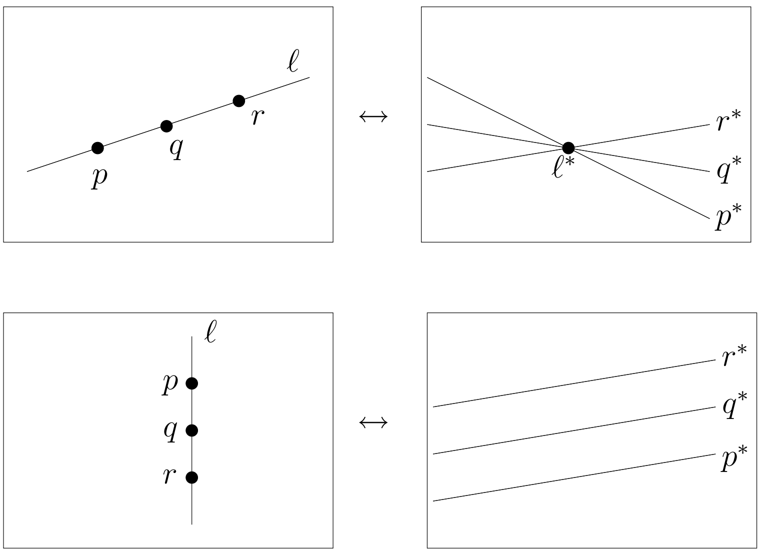

Figure 1.4: Duality: collinear points translate to concurrent lines (top) or parallel lines (bottom)

1.5 Convex Sets

A set $K \subseteq \mathbb{R}^d$ is called *convex* if for all $p, q \in K$ and for all $\lambda \in [0, 1]$, we also have

$$(1 - \lambda)p + \lambda q \in K.$$

Geometrically, this means that for any two points in K, the connecting *line segment* is completely in K, see Figure 1.5.

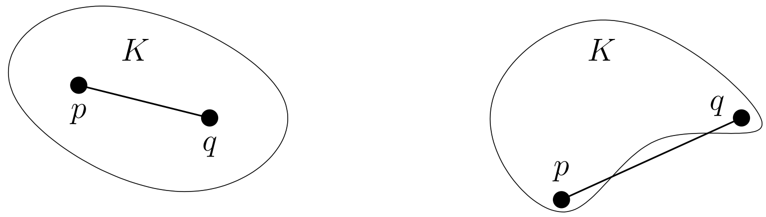

Figure 1.5: A convex set (left) and a non-convex set (right)

It immediately follows that the intersection of an arbitrary collection of convex sets is convex. Convex sets are "nice" sets in many respects, and we often consider the *convex hull* of a set.

Definition 1.6 (Convex Hull). *Let X be an arbitrary subset of $\mathbb{R}^d$. The* convex hull *of X is defined as the intersection of all convex sets containing X,*

$$\operatorname{conv} X := \bigcap_{\substack{C \supseteq X \\ C \text{ convex}}} C.$$

The convex hull can also be characterized in terms of convex combinations: If $p_1, \ldots, p_n \in \mathbb{R}^d$, a *convex combination* of the points is a linear combination $\lambda_1 p_1 + \ldots + \lambda_n p_n$ such that all $\lambda_i \geq 0$ and $\sum_{i=1}^n \lambda_i = 1$.

Lemma 1.7. *Let $X \subseteq \mathbb{R}^d$. The convex hull of X equals the set of all finite convex combinations of points in X,*

$$\operatorname{conv}(X) = \{\sum_{x \in S} \lambda_x x \mid S \subset X \textit{ finite}, \sum_{x \in S} \lambda_x = 1, \textit{ and } \lambda_x \geq 0 \textit{ for all } x \in S\}.$$

The proof is left as an exercise.

Of particular interest for us are convex hulls of finite point sets, see Figure 1.6 for an illustration in $\mathbb{R}^2$. For these, and more generally for closed sets X, the convex hull can also be characterized as the intersection of all halfspaces containing X.

Lemma 1.8. *If $X \subseteq \mathbb{R}^d$ is finite, then*

$$\operatorname{conv}(X) = \bigcap_{\substack{H \supseteq X \\ H \text{ closedhalfspace}}} H.$$

The proof of this lemma is the subject of Exercise 5. We remark that instead of closed halfspaces, one could also take open halfspaces or all halfspaces. In its present form, however, the lemma immediately yields the corollary that the convex hull of a closed point set is closed, which is sometimes useful to know.

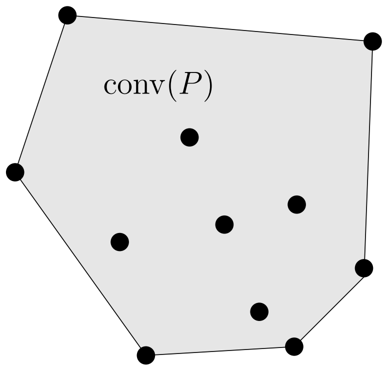

Figure 1.6: The convex hull of a point cloud $P \subseteq \mathbb{R}^2$

Theorem 1.9 (Separation Theorem). *Let $C, D \subseteq \mathbb{R}^d$ be convex sets, $C \cap D = \emptyset$. Then there exists a hyperplane $h = \{\langle a, x\rangle = \alpha\}$ that* weakly separates *the two two sets, in the sense that they lie in opposite closed halfspaces, i.e., $C \subseteq \overline{h^+}$ and $D \subseteq \overline{h^-}$.*

If both C and D are closed and at least one of them is bounded (hence compact), then they can be strictly separated, *i.e., h can be chosen such that $C \subseteq h^+$ and $D \subseteq h^-$.*

The proof of strict separability for two compact convex sets is the subject of Exercise 4. The general case follows by a limiting argument, which we omit, see [7, Chapter 1] or [2] for a proof. The generalization of the Separation Theorem to infinite-dimensional vector spaces, the *Hahn-Banach Theorem*, is one of the basic theorems in functional analysis, see, for instance, [4].

Another fundamental fact about convex sets is

Lemma 1.10 (Radon's Lemma). *Let $S \subseteq \mathbb{R}^d$, $|S| \geq d+2$. Then there exist two disjoint subsets $A, B \subseteq S$ such that $\operatorname{conv}(A) \cap \operatorname{conv}(B) \neq \emptyset$.*

For instance, if $S = \{p_1, p_2, p_3, p_4\}$ is a set of 4 points in the plane $\mathbb{R}^2$, and if we assume that no three of the points are collinear, there are exactly two possibilities what such a *Radon Partition* can look like: Either the points of S form the vertices of a convex quadrilateral, say numbered in counterclockwise order, in which case the diagonals intersect and we can take $A = \{p_1, p_3\}$ and $B = \{p_2, p_4\}$. Or one of of the points, say p_4, is contained in the convex hull of the other three, in which case $A = \{p_4\}$ and $B = \{p_1, p_2, p_3\}$ (or vice versa).

Proof. By passing to a subset of S, if necessary, we may assume that $S = \{p_1, \ldots, p_{d+2}\}$ is a finite set that contains exactly $d+2$ points. Since the maximum size of an affinely independent set in $\mathbb{R}^d$ is $d+1$, there is a nontrivial affine dependence $\sum_{i=1}^{d+2} \alpha_i p_i = \mathbf{0}$, $\sum_{i=1}^{d+2} \alpha_i = 0$, not all $\alpha_i = 0$. We group the indices according to the signs of the α_i's: $P := \{i : \alpha_i \geq 0\}$ and $N := \{i : \alpha_i < 0\}$. Now, by bringing the terms with negative coefficients on on side, we conclude

$\lambda := \sum_{i\in P} \alpha_i = \sum_{i\in N}(-\alpha_i)$ and $\lambda \neq 0$ (otherwise all $\alpha_i = 0$). Moreover, $\sum_{i\in P} \alpha_i p_i = \sum_{i\in N}(-\alpha_i)p_i$. Now, the coefficients on both sides of the last equation are nonnegative and sum up to λ. Thus, dividing by λ, we see that the convex hulls of $A := \{p_i : i \in P\}$ and $B := \{p_i : i \in N\}$ intersect. □

A nontrivial and very important consequence of Radon's Lemma is

Fact 1.11 (Carathéodory's Theorem). *If $S \subseteq \mathbb{R}^d$ and $p \in \operatorname{conv}(S)$, then there exists a subset $A \subseteq S$, $|A| \leq d+1$, such that $p \in \operatorname{conv}(A)$.*

Again, we omit the proof and refer to [2] for the details. Another very important statement about convex sets is

Theorem 1.12 (Helly's Theorem). *Let $C_1, \ldots, C_n \subseteq \mathbb{R}^d$ be convex sets, $n \geq d+1$. If every $d+1$ of the sets C_i have a non-empty common intersection, the common intersection $\bigcap_{i=1}^n C_i$ of all sets is nonempty.*

For an application, see Exercise 6.

Proof. Fix d. We proceed by induction on n. If $n = d+1$, there is nothing to prove, so we may assume $n \geq d+2$. For each index i, $1 \leq i \leq n$, the family of C_j's with $j \neq i$ also satisfies the assumptions of Helly's Theorem, so by induction, their common intersection is nonempty, i.e., there exists some point $p_i \in \bigcap_{j\neq i} C_j \neq \emptyset$. If $p_k = p_l$ for some $k \neq l$, then $p_k \in C_j$ for all $j \neq k$, and also $p_k = p_l \in C_k$ because $k \neq l$, so $p_k \in \bigcap_{i=1}^n C_i$ as desired. Thus, we can assume that all the points p_i, $1 \leq i \leq n$ are distinct. Since there are $n \geq d+2$ of these points, by Radon's Lemma there are disjoint subsets J and K of $[n] := \{1, \ldots, n\}$ such that $\operatorname{conv}\{p_j : j \in J\} \cap \operatorname{conv}\{p_k : k \in K\} \neq \emptyset$. Let us pick a point q in the intersection of these two convex hulls. We claim that $q \in \bigcap_{i=1}^n C_i$. For consider any index i. Since J and K are disjoint, i cannot belong to both of them, say $i \notin J$ But this means that for all $j \in J$, $p_j \in C_i$ (by choice of the p_j's). Consequently, $q \in \operatorname{conv}\{p_j : j \in J\} \subseteq C_i$. The case $i \notin K$ is symmetric, so we have shown that q indeed belongs to every c_i. □

Remark 1.13. *There is also an "infinite version" of Helly's Theorem: If $\mathcal{C}$ is an infinite family of* compact *convex sets in $\mathbb{R}^d$, and if any $d+1$ of the sets in $\mathcal{C}$ intersect, then $\bigcap_{C\in\mathcal{C}} C \neq \emptyset$. Recall that a subset $K \subseteq \mathbb{R}^d$ is compact iff it is* closed *(if $\{a_n\}_{n\in\mathbb{N}} \subseteq K$ and if $a = \lim_{n\to\infty} a_n$ exists in $\mathbb{R}^d$, then $a \in K$) and* bounded *(i.e., there exists some constant C such that $\|x\| \leq C$ for all $x \in K$). If one of these conditions is dropped, then the infinite version of Helly's Theorem fails, see Exercise 8.*

1.6 Balls and Boxes

Here are basic types of convex sets in $\mathbb{R}^d$ (see also Exercise 2).

Definition 1.14. *Fix $d \in \mathbb{N}, d \geq 1$.*

(i) Let $a = (a_1, \dots, a_d) \in \mathbb{R}^d$ *and* $b = (b_1, \dots, b_d)$ *be two* d*-tuples such that* $a_i \leq b_i$ *for* $i = 1, \dots, d$*. The* box $Q_d(a, b)$ *is the* d*-fold Cartesian product*

$$Q_d(a, b) := \prod_{i=1}^{d} [a_i, b_i] \subseteq \mathbb{R}^d.$$

(ii) $Q_d := Q_d(\mathbf{0}, \mathbf{1})$ *is the* unit box*, see Figure 1.7 (left).*

(iii) Let $c \in \mathbb{R}^d, \rho \in \mathbb{R}^+$*. The* ball $B_d(c, \rho)$ *is the set*

$$B_d(c, \rho) = \{x \in \mathbb{R}^d \mid \|x - c\| \leq \rho\}.$$

(iv) $B_d := B_d(\mathbf{0}, 1)$ *is the* unit ball*, see Figure 1.7 (right).*

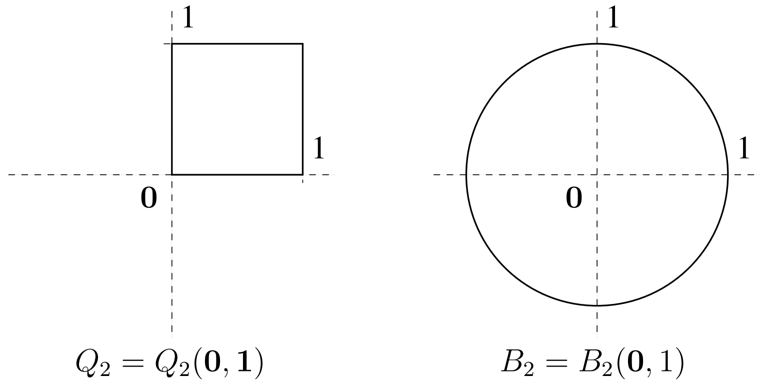

Figure 1.7: The unit box (left) and the unit ball (right)

While we have a good intuition concerning balls and boxes in dimensions 2 and 3, this intuition does not capture the behavior in higher dimensions. Let us discuss a few counterintuitive phenomena.

Diameter. The *diameter* of a compact[2] set $X \subseteq \mathbb{R}^d$ is defined as

$$\mathrm{diam}(X) = \max_{x,y \in X} \|x - y\|.$$

What can we say about the diameters of balls and boxes?

Lemma 1.15. *For* $d \in \mathbb{N}, d \geq 1$*,*

(i) $\mathrm{diam}(Q_d) = \sqrt{d}$*, and*

(ii) $\mathrm{diam}(B_d) = 2$*.*

[2]a set that is closed and bounded

Proof. This is not difficult, but it is instructive to derive it using the material we have. For $x, y \in Q_d$, we have $|x_i - y_i| \leq 1$ for $i = 1, \ldots, d$, from which

$$\|x - y\|^2 = (x - y) \cdot (x - y) = \sum_{i=1}^{d} (x_i - y_i)^2 \leq d$$

follows, with equality for $x = \mathbf{0}, y = \mathbf{1}$. This gives (i). For (ii), we consider $x, y \in B_d$ and use the triangle inequality to obtain

$$\|x - y\| \leq \|x - \mathbf{0}\| + \|\mathbf{0} - y\| = \|x\| + \|y\| \leq 2,$$

with equality for $x = (1, 0, \ldots, 0), y = (-1, 0, \ldots, 0)$. This is (ii). $\square$

The counterintuitive phenomenon is that the unit box contains points which are arbitrarily far apart, if d only gets large enough. For example, if our unit of measurement is cm (meaning that the unit box has side length $1\mathrm{cm}$), we find that $Q_{10{,}000}$ has two opposite corners which are $1\mathrm{m}$ apart; for $Q_{10^{10}}$, the diameter is already $1\mathrm{km}$.

1.7 Volume and Surface Area

We will use the notation vol (or by vol_d, if we want to stress the dimension) for the *d-dimensional volume* or *Lebesgue measure*. An exact definition requires a certain amount of measure and integration theory, which we will not discuss here. In particular, we will not discuss the issue of *non-measurable sets*, but adopt the convention that whenever we speak of the volume of a set A, it will be implicitly assumed that A is measurable. A few key properties that the d-dimensional volume enjoys are the following:

1. The volume of a d-dimensional box equals $\mathrm{vol}_d(Q_d(a, b)) = \prod_{i=1}^{d}(b_i - a_i)$. In particular, $\mathrm{vol}_d(Q_d) = 1$.

2. Volume is *translation-invariant*, i.e., $\mathrm{vol}(A) = \mathrm{vol}(A + x)$ for all $x \in \mathbb{R}^d$.

3. Volume is invariant under orthogonal maps (rotations and reflections. More generally, if $T : \mathbb{R}^d \to \mathbb{R}^d$ is a linear transformation, then $\mathrm{vol}_d(T(A)) = |\det T| \,\mathrm{vol}(A)$.

Volume is also closely related to integration. If one prefers the latter as a primitive notion, one can also consider the equation

$$\mathrm{vol}(X) = \int_{\mathbb{R}^d} \mathbf{1}_X(x) dx,$$

as a definition of the volume of a bounded (and measurable) subset $X \subset \mathbb{R}^d$, where $\mathbf{1}_X$ is the *characteristic function* of X,

$$\mathbf{1}_X(x) = \begin{cases} 1, & \text{if } x \in X, \\ 0, & \text{otherwise.} \end{cases}$$

The Unit Sphere and Surface Area. The boundary of the unit ball B_d in $\mathbb{R}^d$, i.e., the set of all vectors of (Euclidean) norm 1, is the *unit sphere*

$$\mathbb{S}^{d-1} = \{u \in \mathbb{R}^d : \|u\|_2 = 1\}.$$

We will also need the notion of $(d-1)$-dimensional *surface area* for (measurable) subsets $E \subseteq \mathbb{S}^{d-1}$. If one takes the notion of volume as given, one can define the surface area $\sigma = \sigma_{d-1}$ by

$$\sigma_{d-1}(E) := \frac{1}{d} \operatorname{vol}_d(\operatorname{cone}(E, \mathbf{0})),$$

where $\operatorname{cone}(E, \mathbf{0}) := \{tx : 0 \leq t \leq 1, x \in E\}$.

In particular, we note the following facts about the volume of the unit ball and the surface area of the unit sphere:

Fact 1.16. *Let* $d \in \mathbb{N}, d \geq 1$.

(i) $\operatorname{vol}_d(B_d) = \frac{\pi^{d/2}}{(d/2)!}$.

(ii) $\sigma_{d-1}(\mathbb{S}^{d-1}) = \frac{2\pi^{d/2}}{(d/2-1)!}$.

Here, for a real number $\alpha > -1$, the *generalized factorial* $\alpha!$ (also often called the Gamma Function $\Gamma(\alpha+1)$) is defined by $\alpha! := \int_0^\infty t^\alpha e^{-t} dt$. This function obeys the familiar law $(\alpha+1)! = (\alpha+1)\alpha!$. In particular, it coincides with the usual recursively defined factorial for integers, and for half-integers we have

$$(d/2)! = \sqrt{\pi} \prod_{m=0}^{(d-1)/2} \left(m + \frac{1}{2}\right), \quad \text{for odd } d.$$

We recall the following important approximation:

Fact 1.17 (Stirling's Formula). $\alpha! \sim \frac{\alpha^\alpha}{e^\alpha}\sqrt{2\pi\alpha}$ *as* $\alpha \to \infty$ *(where* $f \sim g$ *means* $f/g \to 1$*).*

We skip the proofs of Lemma 1.16 and of Stirling's formula, because they take us too far into measure-theoretic and analytic territory; here is just a sketch of a possible approach for Part (i) of the lemma: *Cavalieri's principle* says that the volume of a compact set in $\mathbb{R}^d$ can be calculated by integrating over the $(d-1)$-dimensional volumes of its *slices*, obtained by cutting the set orthogonal to some fixed direction. In case of a ball, these slices are balls again, so we can use induction to reduce the problem in $\mathbb{R}^d$ to the problem in $\mathbb{R}^{d-1}$.

Let us discuss the counterintuitive implication of Lemma 1.16. The intuition tells us that the unit ball is larger than the unit box, and for $d = 2$, Figure 1.7 clearly confirms this. B_2 is larger than Q_2 by a factor of π (the volume of B_2). You might recall (or derive from the lemma) that

$$\mathrm{vol}(B_3) = \frac{4}{3}\pi,$$

meaning that B_3 is larger than Q_3 by a factor of more than four. Next we get

$$\mathrm{vol}(B_4) \approx 4.93, \quad \mathrm{vol}(B_5) \approx 5.26,$$

so $\mathrm{vol}(B_d)/\,\mathrm{vol}(Q_d)$ seems to grow with d. Calculating

$$\mathrm{vol}(B_6) \approx 5.17$$

makes us sceptical, though, and once we get to

$$\mathrm{vol}(B_{13}) \approx 0.91,$$

we have to admit that the unit ball in dimension 13 is in fact *smaller* than the unit box. From this point on, the ball volume rapidly decreases (Table 1.1, see also Figure 1.8), and in the limit, it even vanishes:

$$\lim_{d\to\infty} \mathrm{vol}(B_d) = 0,$$

because $\Gamma(d/2 + 1)$ grows faster than $\pi^{d/2}$.

d	13	14	15	16	17	$\cdots$	20
$\mathrm{vol}(B_d)$	0.91	0.6	0.38	0.24	0.14	$\cdots$	0.026

Table 1.1: Unit ball volumes

1.8 Measure Concentration

Suppose the surface of the earth is completely covered with grass, and your task is to mow it. You have a giant lawn mower able to mow a strip that spans a spherical angle of α, say (where α is small in order not to make your task too easy). What percentage of the grass have you mowed after you have gone around the equator once? See Figure 1.9 (left) for an illustration of the situation.

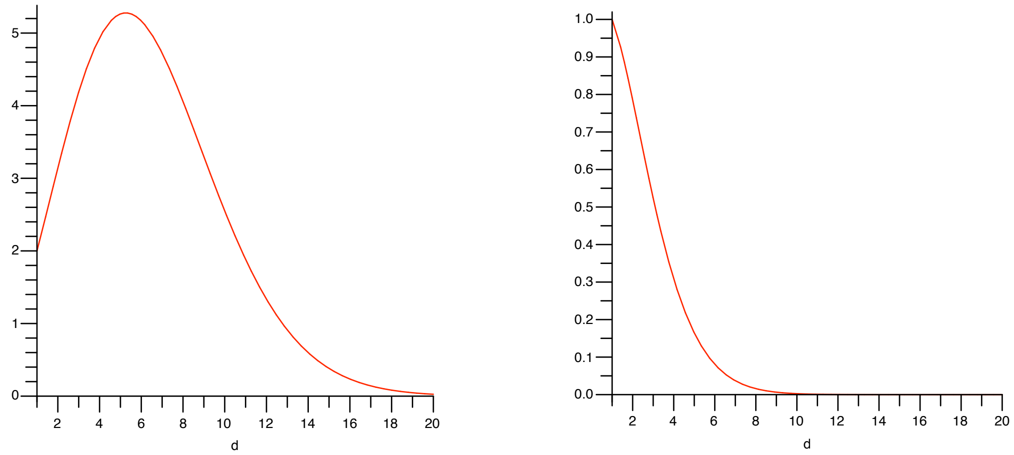

Figure 1.8: Plots of $\operatorname{vol}_d(B_d)$ (on the left) and of $\operatorname{vol}_d(B_d)/\operatorname{vol}_d(Q_d(-\mathbf{1},\mathbf{1}))$ (on the right), for $d = 1,\ldots,20$.

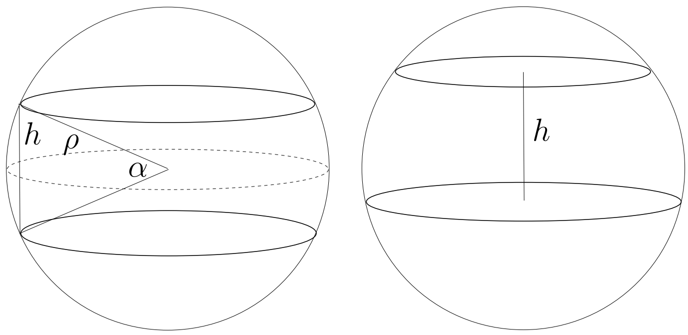

Figure 1.9: The giant lawn mower covers a strip of width h

The lawn mower question is easy to solve using a known fact: The area covered by pushing the lawn mower around the equator is

$$2\pi\rho h,$$

with ρ the radius of the ball (in our case, $\rho \approx 6,378\text{km}$) and h the width of the strip. Interestingly, this area does not depend on the strip being centered around the equator—any strip of width h has area $2\pi\rho h$, see Figure 1.9 (right). For $h = 2\rho$, the strip covers the whole surface, and the area is $4\pi\rho^2$, which follows from the above formula for the surface area of $\mathbb{S}^2$ by scaling by a factor of ρ.

Because the strip spans spherical angle α, we get $h = 2\rho\sin(\alpha/2)$, meaning that the fraction of the earth's surface you have mowed is

$$\frac{2\pi\rho h}{4\pi\rho^2} = \frac{h}{2\rho} = \sin(\alpha/2).$$

If $\alpha = 10^o$, for example (a pretty big mower, the strip is more than $1,000\mathrm{km}$ wide), the fraction covered is 8.7%.

Measure Concentration on the Sphere. The counterintuitive phenomenon is that your task would be much simpler if the earth were of higher dimension. For sufficiently large dimension, one round with your 10^o-mower (or any α-mower, for fixed α) covers 90% (or any desired percentage) of the surface. This means, the surface area of B_d is concentrated around the equator for large d. Not only that: by symmetry of B_d, the surface area is concentrated around *any* equator.

In fact, there is a much more general result of this kind. Since we are interested in relative surface area, we might as well renormalize the surface area measure so that the whole sphere gets measure 1. Thus, we define, for $E \subseteq \mathbb{S}^{d-1}$,

$$P(E) := \frac{\sigma_{d-1}(E)}{\sigma_{d-1}(\mathbb{S}^{d-1})}.$$

In other words, P is the uniform probability measure on $\mathbb{S}^{d-1}$.

Furthermore, we need the following notion: For $A \subseteq \mathbb{S}^{d-1}$ and a real number $t > 0$, let $A_t := \{x \in \mathbb{S}^d : \mathrm{dist}(x, A) \leq t\}$ be the set of points at (Euclidean) distance at most t from A, where, formally, $\mathrm{dist}(x, A) := \inf_{y \in A} \|x - y\|$. With this notation, we can state the theorem about measure concentration on the sphere:

Theorem 1.18. *Let $A \subseteq \mathbb{S}^{d-1}$, $P(A) \geq 1/2$. Then for any real $t > 0$,*

$$1 - P(A_t) \leq 2e^{-t^2 d/2}.$$

For the aforementioned question of the relative surface measure of a strip around the equator, we can apply the theorem twice, once taking A to be the northern hemisphere, and once the southern hemisphere. In fact, there is an even stronger statement (which is referred to as the *isoperimetric inequality for the sphere*) which states that among all sets A with $P(A) = 1/2$, $P(A_t)$ is minimized, simultaneously for all t, if A is a hemisphere.

Figure 1.10 shows the (width of the) strip around the equator that contains 90% of the area, for three values of d, see Matoušek's book [7]).

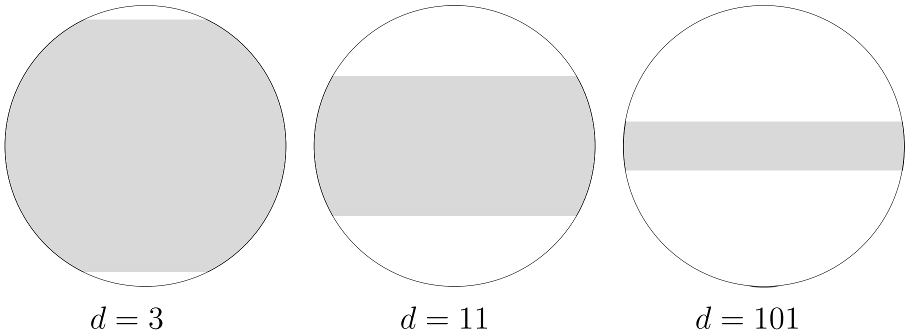

Figure 1.10: strip around the equator containing 90% of the area

Measure Concentration on the Discrete Cube. Without seeing the connection yet, you already know a similar phenomenon involving the unit box Q_d. Let the "equator" of Q_d be the set

$$\{x \in Q_d \mid \sum_{i=1}^{d} x_i = \frac{d}{2}\},$$

see Figure 1.11 (left) for a picture in dimension 3.

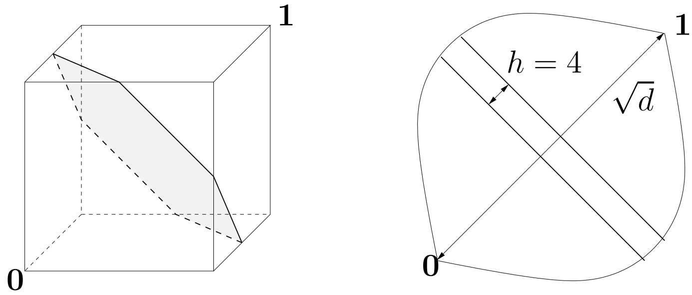

Figure 1.11: The "equator" of the unit box (left); symbolic drawing of a strip of width h around the equator that contains 96.3% of all box corners (right)

Note that the equator is the intersection of Q_d with a hyperplane. Motivated by Figure 1.10, we plan to prove now that the strip around the equator containing 90% of the 2^d box corners becomes thinner and thinner (compared to the diameter of the box), as d grows. It turns out that this is nothing else than the well-known

Fact 1.19 (Chernoff Bound, 0/1-Version). *Let $X_1, \ldots, X_d$ be independent random variables with $\Pr[X_i = 0] = \Pr[X_i = 1] = 1/2$ for all i, and let $X = X_1 + \cdots + X_d$. For all $t > 0$, we have*

$$\Pr\left[\left|X - \frac{d}{2}\right| > t\right] < 2e^{-2t^2/d}.$$

We will prove a more general theorem in the next section. Setting $t = 2\sqrt{d}$, for example, yields

$$\text{prob}\left(\left|X - \frac{d}{2}\right| > 2\sqrt{d}\right) < 2\exp(-4) \approx 0.037.$$

Because X is the sum of coordinates of a randomly chosen unit box corner, it follows that a fraction of no more than 3.7% of all box corners is outside the strip

$$\{x \in \mathbb{R}^d \mid \frac{d}{2} - 2\sqrt{d} \leq \sum_{i=1}^{d} x_i \leq \frac{d}{2} + 2\sqrt{d}\}$$

around the equator. It follows from our earlier material on hyperplanes (calculation of distance to the origin) that the width of this strip is

$$\frac{d/2 + 2\sqrt{d}}{\sqrt{d}} - \frac{d/2 - 2\sqrt{d}}{\sqrt{d}} = 4,$$

a constant! As d gets larger, the strip therefore becomes thinner and thinner compared to the strip

$$\{x \in \mathbb{R}^d \mid 0 \leq \sum_{i=1}^{d} x_i \leq d\}$$

of width $\sqrt{d}$ containing the whole unit box, see Figure 1.11 (right) for a symbolic picture.

1.9 Exercises

Exercise 1. *Prove that if $P \subset \mathbb{R}^d$ is an affinely independent point set with $|P| = d$, then there exists a unique hyperplane containing all points in P. (This generalizes the statement that there is a unique line through any two distinct points.)*

Exercise 2. *Prove that all boxes $Q_d(a, b)$ and all balls $B(c, \rho)$ are convex sets.*

Exercise 3. *(a) Show that if $\mathcal{C}$ is an arbitrary collection of convex sets in $\mathbb{R}^d$, then $\bigcap_{C \in \mathcal{C}} C$ is again a convex set.*

(b) Prove Lemma 1.7.

Exercise 4. *Let C, D be nonempty compact convex sets in $\mathbb{R}^d$, $C \cap D = \emptyset$.*

(a) Show that there exist points $p \in C$ and $q \in D$ such that, for all $x \in C$ and all $y \in D$, $\|p - q\| \leq \|x - y\|$. (Hint: *You may use the fact that $C \times D$ is also compact; which theorems about continuous functions on compact sets do you remember from analysis?)*

(b) *Let h be the hyperplane with normal vector $p - q$ and passing through the point $m := (p+q)/2$ (the midpoint of the segment pq; what is the equation of this hyperplane?). Show that h* separates *C and D, i.e., that $C \subseteq h^+$ and $D \subseteq h^-$. (We could let the hyperplane pass through any point in the interior of the segment pq instead of the midpoint and the statement would still be true.)*

Exercise 5. *Prove Lemma 1.8. Can you give a counterexample if X is not closed?*

Hint. If $p \notin \operatorname{conv} X$, argue first that $\operatorname{dist}(p, X) := \inf\{\|x - p\| : x \in X\} > 0$. Then use the Separation Theorem to obtain a weakly separating hyperplane, and argue by induction on the dimension.

Exercise 6. *Let S be a set of vertical* line segments[3] *in $\mathbb{R}^2$, see Figure 1.12. Prove the following statement: if for every three of the line segments, there is a line that intersects all three segments, then there is a line that intersects* all *segments.*

Can you give a counterexample in the case of non-vertical segments?

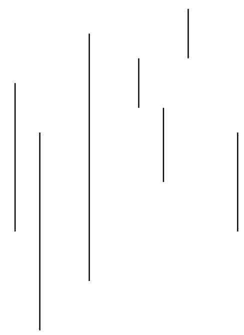

Figure 1.12: A set of vertical line segments in $\mathbb{R}^2$

Hint. Use the duality transform (non-vertical case) and Helly's Theorem. For this, you need to understand the following: (i) what is the set of lines dual to the set of points on a (vertical) segment? (ii) if a line intersects the segment, what can we say about the point dual to this line?

Exercise 7. *State and prove the analogue to Observation 1.5 for the origin-avoiding case.*

Exercise 8. *Show that without the additional compactness assumption, the infinite version of Helly's Theorem is generally not true. That is, give an example, for some dimension d of your choice, of an infinite family $\mathcal{C}$ of (noncompact) convex sets such that*

(i) *any $d+1$ of the sets in $\mathcal{C}$ have a nonempty intersection,*

[3] a line segment is the convex hull of a set of two points

(ii) *but* $\bigcap_{C \in \mathcal{C}} C = \emptyset$.

Exercise 9. *In order to generate a random point p in B_d, we could proceed as follows: first generate a random point p in $Q_d(-\mathbf{1}, \mathbf{1})$ (this is easy, because it can be done coordinatewise); if $p \in B_d$, we are done, and if not, we repeat the choice of p until $p \in B_d$ holds. Explain why this is not necessarily a good idea. For $d = 20$, what is the expected number of trials necessary before the event '$p \in B_d$' happens?*