CHAPTER 1

Basic concepts

Convexity

We begin by recalling some definitions and basic facts concerning convexity. For two points $a, b \in \mathbb{R}^d$ we define a *segment* $[a, b]$ joining a and b as the set $[a, b] = \{\alpha a + \beta b,\ \alpha, \beta \geq 0, \alpha + \beta = 1\}$. A set C in $\mathbb{R}^d$ is called *convex* if for any two points a and b in C, the segment $[a, b]$ is also contained in C. The intersection of convex sets is a convex set, again.

Given n points $a_1, \ldots, a_n$ in $\mathbb{R}^d$ and real coefficients $\alpha_1, \ldots, \alpha_n$, the point $a = \alpha_1 a_1 + \ldots + \alpha_n a_n$ is called their *positive combination* if all α_i are nonnegative. If in addition $\alpha_1 + \ldots + \alpha_n = 1$, then the point a is called a *convex combination* of a_i. If the coefficients α_i only satisfy the condition $\alpha_1 + \ldots + \alpha_n = 1$, then the point a is called an *affine combination* of a_i. Note the following easy observation.

Fact 1.1. *If C is a convex subset of $\mathbb{R}^d$ and $a_1, \ldots, a_n \in C$, then all convex combinations of $a_1, \ldots, a_n$ belong to C.*

For a (non-empty) subset S in $\mathbb{R}^d$ define

$$\operatorname{conv} S = \{\text{all convex combinations of elements of } S\},$$
$$\operatorname{pos} S = \{\text{all positive combinations of elements of } S\},$$
$$\operatorname{aff} S = \{\text{all affine combinations of elements of } S\}.$$

They are called the *convex* and the *positive* (or *cone*) and the *affine hull* of S, respectively. These are clearly convex sets. Moreover, $\operatorname{conv} S$ is the smallest (with respect to inclusion) convex set containing S (see Exercise 1.1), and it is also the intersection of all convex sets in $\mathbb{R}^d$ containing S.

Obviously, a segment in $\mathbb{R}^d$ is a convex set. Another canonical example of a convex set is a closed halfspace $H = \{x \in \mathbb{R}^d,\ ax \leq \alpha\}$, and another one is the open halfspace $H = \{x \in \mathbb{R}^d,\ ax < \alpha\}$. Here $a \in \mathbb{R}^d$ is a nonzero vector, and ax denotes the scalar product of $a, x \in \mathbb{R}^d$. The intersection of finitely many halfspaces is convex, and is called a *convex polyhedron*. The convex hull of finitely many points in $\mathbb{R}^d$ is a *convex polytope*. In particular, the convex hull of $k+1$ affinely independent points in $\mathbb{R}^d$ is a polytope. It is a k-dimensional *simplex*, to be denoted by Δ^k.

Given a convex set $C \subset \mathbb{R}^d$, its closure $\operatorname{cl} C$ and its interior $\operatorname{int} C$ are also convex. The set C is also convex in the affine subspace $\operatorname{aff} C$, and its interior in that space is called its *relative interior*, to be denoted by $\operatorname{relint} C$.

The *separation theorem* is a fundamental result in convexity. It says the following.

Theorem 1.2 (separation theorem). *Assume C and K are convex sets in $\mathbb{R}^d$, C is compact, and K is closed. Then $C \cap K = \emptyset$ if and only if there are closed halfspaces H_1 and H_2 with $C \subset H_1$ and $K \subset H_2$ satisfying $H_1 \cap H_2 = \emptyset$.*

The proof can be found in several books, for instance in [**66**]. Since a closed halfspace is always of the form $\{x \in \mathbb{R}^d : ax \geq \alpha\}$ or $\{x \in \mathbb{R}^d : ax \leq \alpha\}$ for some nonzero $a \in \mathbb{R}^d$ and $\alpha \in \mathbb{R}$ (ax still stands for the scalar product of $a, x \in \mathbb{R}^d$), the separation theorem can be stated as follows. Under the above conditions on C and K, $C \cap K = \emptyset$ if and only if there is a vector $a \in \mathbb{R}^d$ such that $\sup\{ax : x \in C\} < \inf\{ax : x \in K\}$.

Here is a weaker form of the separation theorem.

Theorem 1.3 (separation theorem, weaker form). *Assume C and K are convex sets in $\mathbb{R}^d$. Then $C \cap K = \emptyset$ implies that there is a non-zero vector $a \in \mathbb{R}^d$ such that $\sup\{ax : x \in C\} \leq \inf\{ax : x \in K\}$.*

Given a non-empty and closed convex set $C \subset \mathbb{R}^d$, the *nearest point map* $\text{pr}_C = \text{pr} : \mathbb{R}^d \to C$ is defined as the nearest point in C to $x \in \mathbb{R}^d$. So $\text{pr}(x) = y$ means that $y \in C$ and

$$\|x - y\| = \min\{\|x - z\| : z \in C\}.$$

Here and throughout the book $\|x\|$ denotes the Euclidean length of the vector $x \in \mathbb{R}^d$. The existence and uniqueness of $y = \text{pr}_C(x)$ is well-known and is the content of Exercise 1.6.

The *distance* between $x \in \mathbb{R}^d$ and a convex set C is defined as

$$\text{dist}(x, C) = \inf\{\|x - z\| : z \in C\},$$

which is the same as $\|x - \text{pr}_C(x)\|$ when C is closed.

Assume next that $C \subset \mathbb{R}^d$ is a convex set with non-empty interior and $z \in C$ is a point on the boundary $\text{bd}\, C$ of C. The separation theorem implies the following.

Theorem 1.4 (supporting hyperplane theorem). *Under the above conditions there is a closed halfspace H such that $C \subset H$ and $z \in \text{bd}\, H$.*

Another equivalent form is that for every $z \in \text{bd}\, C$ there is a *supporting hyperplane* h to C at z, meaning that C is on one side of h and $z \in h$. One should simply take $h = \text{bd}\, H$ in the previous theorem.

Farkas lemma and linear programming

The Farkas lemma [**35**] gives a necessary and sufficient condition for the existence of a solution to the system of linear inequalities of the form

$$a_i x \leq b_i,\ i \in [n], \tag{1.1}$$

where $a_i \in \mathbb{R}^d$ for $i \in [n]$ and $b = (b_1, \dots, b_n) \in \mathbb{R}^n$. Here and throughout this book $[n]$ stands for the set $\{1, \dots, n\}$. One more piece of notation: For a vector a in $\mathbb{R}^d$ and a real number b we write $\begin{pmatrix} a \\ b \end{pmatrix}$ for the vector in $\mathbb{R}^{d+1}$ whose first d coordinates coincide with those of a and whose last one is b.

Theorem 1.5 (Farkas lemma)**.** *The linear system (1.1) has no solution if and only if*

$$\begin{pmatrix} 0 \\ -1 \end{pmatrix} \in \text{pos}\left\{ \begin{pmatrix} a_i \\ b_i \end{pmatrix} : i \in [n] \right\}.$$

One direction is easy, namely, when $\begin{pmatrix} 0 \\ -1 \end{pmatrix}$ equals $\sum_{i \in [n]} \gamma_i \begin{pmatrix} a_i \\ b_i \end{pmatrix}$ with coefficients $\gamma_i \geq 0$ for all i, then multiplying the ith inequality in (1.1) by γ_i and summing gives $0 \leq -1$, so the system can't have a solution. The other direction is an application of the separation theorem.

The Farkas lemma is also closely related to linear programming. Linear programs come in various forms. We will need the following one where A is a $d \times n$ matrix, $b \in \mathbb{R}^d$, $c \in \mathbb{R}^n$ is fixed, and $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^d$ are variables. Define

$$P = \{x \in \mathbb{R}^n : Ax \geq b,\ x \geq 0\} \text{ and } D = \{y \in \mathbb{R}^d : y^T A \leq c,\ y \geq 0\},$$

where y^T denotes the transpose of the vector y. The primal form of the *linear program* is this:

$$\text{minimize } cx \ \text{ subject to } x \in P. \tag{1.2}$$

The dual linear program is the following:

$$\text{maximize } by \ \text{ subject to } y \in D. \tag{1.3}$$

Theorem 1.6 (duality theorem of linear programming)**.** *The minimum of the primal program is always at least as large as the maximum of the dual program, and they are equal if both P and D are non-empty.*

Graphs and hypergraphs

Graphs and hypergraphs are an important tool in combinatorial convexity. A graph is a pair $G = G(V, E)$, where V is a finite set and E is a collection of some 2-element subsets of V. Here V is the vertex set of the graph G, and E is its edge set. So an edge of G is just the set $\{u, v\} \in E$ where $u, v \in V$; normally $u \neq v$. The complete graph on n vertices is the graph $G(V, E)$ with $|V| = n$ and E consisting of all pairs $\{u, v\}$ contained in V, $u \neq v$. The standard notation is K_n. The *edge density* of a graph $G(V, E)$ with $|V| = n$ is defined as

$$\delta(G) = \frac{|E|}{\binom{n}{2}}.$$

A path in a graph $G(V, E)$ is a sequence $u_0, e_1, u_2, e_2, \ldots, e_k, u_k$ where each $u_i \in V$ and $e_i = \{u_{i-1}, u_i\} \in E$ for all $i \in [k]$. The path *connects* the vertices u_0 and u_k. This path is a *circuit* if $u_0 = u_k$. G is *connected* if any two vertices of G are connected by a path. A *tree* is a connected graph containing no circuits.

A *bipartite* graph $G(V, E)$ is a graph where the vertex set is partitioned into two (non-empty) sets, P and Q say, $V = P \cup Q$ and $P \cap Q = \emptyset$, so that every edge $e \in E$ has one point in P and the other in Q. The complete bipartite graph with partition classes P and Q is the graph $G(P \cup Q, E)$ where E consists of all pairs $\{p, q\}$ with $p \in P$ and $q \in Q$; the standard notation is $K(n, m)$ when $|P| = n$ and $|Q| = m$. In Exercise 1.7 you are asked to show the following fact.

Fact 1.7. *Every graph $G(V, E)$ with edge density $\alpha > 0$ contains a bipartite subgraph $G'(V, E')$ whose edge density is at least $\alpha/2$.*

In Chapter 29 we will use Turán's famous theorem; see for instance [**62**]. It is about graphs containing no complete subgraph of size $k + 1$.

Theorem 1.8 (Turán's theorem)**.** *If a graph $G(V, E)$ with $|V| = n$ contains no copy of K_{k+1} as a subgraph, then it has at most $\binom{n}{2} - k\binom{n/k}{2}$ edges.*

A *directed graph* is a pair $G = G(V, A)$ where V is the set of vertices of G and A is the set of its *arcs* where arc $a \in A$ is an ordered pair (u, v) with both $u, v \in V$. The head of arc a is v and its tail is u. A *directed path* in a directed graph $G = G(V, A)$ is a sequence $u_0, a_1, u_2, a_2, \ldots, a_k, u_k$ where each $u_i \in V$ and u_{i-1} is the tail and u_i is the head of the arc $a_i \in A$ for all $i \in [k]$. A *directed cycle* in G is a directed path with $u_0 = u_k$.

A hypergraph $\mathcal{H}$ on ground set V is a collection of subsets of V. Usually V is finite and is called the *vertex set* of $\mathcal{H}$, and the element of $\mathcal{H}$ are the edges of the hypergraph. A hypergraph is *uniform* if all edges have the same size, and it is *k-uniform* if this same size is $k \in \mathbb{N}$, $k \geq 2$. In particular, a graph $G(V, E)$ is in fact a 2-uniform hypergraph with vertex set V.

A *matching* in a hypergraph $\mathcal{H}$ is a collection of disjoint edges, and the *matching number* of $\mathcal{H}$ is the maximum cardinality of a matching in $\mathcal{H}$. The standard notation for the matching number is $\nu(\mathcal{H})$. A *transversal* of $\mathcal{H}$ is a set of vertices containing at least one vertex from every edge of $\mathcal{H}$. The *transversal number* of $\mathcal{H}$, to be denoted by $\tau(\mathcal{H})$, is the minimum cardinality of a transversal. It is evident that

$$\nu(\mathcal{H}) \leq \tau(\mathcal{H}).$$

Let $x_e \geq 0$ be a number associated with edge $e \in \mathcal{H}$. The *fractional matching number* of $\mathcal{H}$ is the solution of the linear program

$$\text{maximize} \sum_{e \in \mathcal{H}} x_e \text{ subject to } \sum_{e \ni v} x_e \leq 1 \text{ for all } v \in V. \tag{1.4}$$

The standard notation for the maximum is $\nu^*(\mathcal{H})$. It is clear that $\nu^*(\mathcal{H})$ is always finite. It is also evident that $\nu(\mathcal{H})$ is the solution of (1.4) when the variables $x_e \geq 0$ are restricted to integers. It follows that $\nu(\mathcal{H}) \leq \nu^*(\mathcal{H})$.

The dual notion is the *fractional transversal number* of $\mathcal{H}$. It is denoted by $\tau^*(\mathcal{H})$ and is defined by variables $y_v \geq 0$ (for every $v \in V$) via

$$\text{minimize} \sum_{v \in V} y_v \text{ subject to } \sum_{v \in e} y_v \geq 1 \text{ for all } e \in \mathcal{H}. \tag{1.5}$$

Again, $\tau(\mathcal{H})$ is the solution of (1.5) when the variables $y_v \geq 0$ have to take integer values. It is clear that $\tau^*(\mathcal{H}) \leq \tau(\mathcal{H})$.

The duality theorem of linear programming shows that $\nu^*(\mathcal{H}) = \tau^*(\mathcal{H})$. We record the important inequalities

$$\nu(\mathcal{H}) \leq \nu^*(\mathcal{H}) = \tau^*(\mathcal{H}) \leq \tau(\mathcal{H}). \tag{1.6}$$

The *edge density* of a k-uniform hypergraph $\mathcal{H}$ with vertex set V, $|V| = n$, is defined as

$$\delta(\mathcal{H}) = \frac{|\mathcal{H}|}{\binom{n}{k}}.$$

The *complete k-uniform* hypergraph $\mathcal{H}$ with vertex set V is the collection of all k tuples from V. The notation is $\binom{V}{k}$. So if $|V| = n$, then $|\mathcal{H}| = \binom{n}{k}$. Of course $n \geq k$ is assumed.

A k-uniform hypergraph $\mathcal{H}$ is called *k-partite* if the vertex set of $\mathcal{H}$ has a partition into k parts: $V = V_1 \cup V_2 \cup \ldots \cup V_k$ with $V_i \neq \emptyset$ for each $i \in [k]$ such that $|E \cap V_i| = 1$ for every edge $E \in \mathcal{H}$ and every $i \in [k]$. Here again $[k] = \{1, 2, \ldots, k\}$ is a very useful piece of notation. So an edge of a k-partite hypergraph is of the form $E = \{v_1, \ldots, v_k\}$ where $v_i \in V_i$ for all $i \in [k]$. See Figure 1.1 where some edges have different colours in order to distinguish them. We will often call such a k tuple a *transversal* of the set system $V_1, \ldots, V_k$.

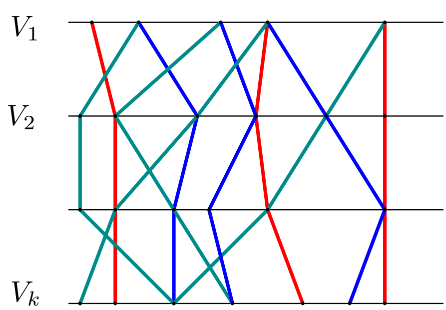

FIGURE 1.1. A k-partite hypergraph.

The *complete k-partite* hypergraph with vertex set partition $V = V_1 \cup \ldots \cup V_k$ is the hypergraph containing all k tuples $\{v_1, \ldots, v_k\}$ satisfying $v_i \in V_i$ for all $i \in [k]$. The standard notation for the complete k-partite hypergraph is $K(t_1,\dots,t_k)$ where $|V_i|=t_i$. We will use the notation $K(k\times t)$ when all classes have size t, that is, when $t=t_1=\dots=t_k$. So $K(k\times t)$ is the complete k-uniform k-partite hypergraph with each class of size t.

A far reaching generalization of Fact 1.7 is the following theorem of Erdős and Simonovits [**33**] that has several applications in combinatorial convexity. A *copy* of $K(k\times t)$ in the hypergraph $\mathcal{H}$ with vertex set V is a complete k-partite subhypergraph of $\mathcal{H}$ with partition classes of size t each. Thus such a copy means k disjoint subsets $U_1,\dots,U_k$ of V, each of size t such that every transversal of the system $U_1,\dots,U_k$ is an edge in $\mathcal{H}$.

It is convenient to introduce here the $\ll$ notation: For positive functions $f,g:\mathbb{N}\to\mathbb{R}$ the notation $f(n)\ll g(n)$ means that there is a constant $c>0$ such that $f(n)\le cg(n)$ for every $n\in\mathbb{N}$; here the constant does not depend on n. It may depend on some parameter, say p, in which case we can write $f(n)\ll_p g(n)$. The notation $\gg$ is similar: $f(n)\gg g(n)$ means that there is a constant $c>0$ such that $f(n)\ge cg(n)$ for every $n\in\mathbb{N}$. The constant c is again independent of n.

Theorem 1.9 (Erdős-Simonovits theorem)**.** *Assume $\mathcal{H}$ is a k-uniform hypergraph on n vertices and its edge density is $\delta(\mathcal{H})\ge p$, where the parameter p satisfies $n^{-t^{-k+1}}\ll_t p\le 1$. Then $\mathcal{H}$ contains at least*

$$bp^{t^k}n^{kt}$$

copies of $K(k\times t)$, where $b>0$ is a constant that only depends on k and t.

It is important to note that the parameter p may depend on n as the condition $p\gg n^{-t^{-k+1}}$ shows. The proof can be found in [**33**] and is based on induction and averaging. The interested reader may check that if $\mathcal{H}$ is a k-uniform hypergraph on n vertices where edges are chosen randomly and independently with probability p, then the expected number of copies of $K(k\times t)$ is of the order $p^{t^k}n^{kt}$.

Szemerédi's regularity lemma [**90**] is a result of central importance graph theory. It says roughly that for a given $\varepsilon>0$, the vertices of a sufficiently large graph can be partitioned into a certain number of parts (of almost equal sizes) such that the bipartite graph between two parts of the partition behaves like a random bipartite graph up to a small error term bounded by ε, for most of the pairs of the partition. Here the "certain number" of parts can be specified, and so can the statement "most of the pairs". Szemerédi's regularity lemma is not used in this book so we give no further details about it. But we will need a weaker and hypergraph version of it which is due to Komlós and Simonovits [**56**], see also [**74**]. It is about a k-partite hypergraph $\mathcal{H}$ with classes $V_1,\dots,V_k$. The edge density of a k-partite hypergraph is, by definition,

$$\delta(\mathcal{H})=\frac{|\mathcal{H}|}{|V_1|\cdot\ldots\cdot|V_k|}.$$

Note that this definition is specific to k-partite hypergraphs and is different from the non-partite case. But in both the number of edges is divided by the number of the underlying complete hypergraph, partite and non-partite alike. Subsets $U_i \subset V_i$ for every $i \in [k]$ define a *subhypergraph* which is again k-partite with classes $U_1, \dots, U_k$ and whose edges are exactly the edges $\{v_1, \dots, v_k\}$ of $\mathcal{H}$ satisfying $v_i \in U_i$ for all i. This subhypergraph is of course k-partite and its edge density is well-defined. Here comes the weaker hypergraph version of Szemerédi's regularity lemma.

Theorem 1.10 (weak hypergraph regularity lemma). *Let $\mathcal{H}$ be a k-partite hypergraph with classes $V_1, \dots, V_k$ with $|V_i| = n$ for all i, $\delta(\mathcal{H}) \geq \beta$ for some $\beta > 0$, and $\varepsilon \in (0, 1/2)$. We assume that n is large enough in terms of k, β, ε. Then there exist subsets U_i of V_i of size $|U_i| = m \geq \beta^{1/\varepsilon^2}$ for all $i \in [k]$ such that:*
(i) *the edge density of the subhypergraph on vertices $U_1, \dots, U_k$ is at least β, and*
(ii) *for every choice of subsets $W_i \subset U_i$ with $|W_i| \geq \varepsilon m$ for all $i \in [k]$, the subhypergraph on $U_1, \dots, U_k$ contains at least one edge.*

Exercises

Exercise 1.1. Show that $\operatorname{conv} S$ is the smallest (with respect to inclusion) convex set containing $S \subset \mathbb{R}^d$.

Exercise 1.2. Show that every polytope is a polyhedron.

Exercise 1.3. Give an example of a polyhedron in $\mathbb{R}^d$ that is not a polytope.

Exercise 1.4. Show that the set of all positive semi-definite matrices of a fixed size is convex.

Exercise 1.5. Let f be a polynomial with complex coefficients which is not constant. Show that the roots of f' lie in the convex hull of the roots of f.

Exercise 1.6. Prove that, given a convex (closed) set $C \subset \mathbb{R}^d$, the nearest point map $\operatorname{pr}_C = \operatorname{pr} : \mathbb{R}^d \to C$ is well defined.

Exercise 1.7. Show that every graph $G(V, E)$ with edge density $\alpha > 0$ contains a bipartite subgraph on the same vertex set whose edge density is at least $\alpha/2$.

Exercise 1.8. Prove the original (and more precise) form of Turán's theorem: Assume that the graph $G(V, E)$ with $|V| = n$ contains no copy of K_{k+1} as a subgraph and $n = pk + q$ where $n \geq k + 1$ and p, q are integers with $0 \leq q \leq k - 1$. Then G has at most $\binom{n}{2} - \frac{1}{2}(n - k + q)p$ edges.

Exercise 1.9. Show the baby version of the Erdős-Simonovits theorem: If the edge density of a graph $G(V, E)$ is $\delta > 0$, then it contains many copies of $K(1, d)$ and $K(2, d)$ for every fixed d (when n is large).