Discrete Geometry II

http://wikis.fu-berlin.de/display/discgeom/Discrete+Geometry+II+SS+2014

— Preliminary Lecture Notes (without any guarantees) —

Prof. Günter M. Ziegler

Fachbereich Mathematik und Informatik, FU Berlin, 14195 Berlin
ziegler@math.fu-berlin.de

FU Berlin, Summer Term 2014

This is the second in a series of three courses on Discrete Geometry. We will get to know fascinating geometric structures such as configurations of points and lines, hyperplane arrangements, and in particular polytopes and polyhedra, and learn how to handle them using modern methods for computation and visualization and current analysis and proof techniques. A lot of this looks quite simple and concrete at first sight (and some of it is), but it also very quickly touches topics of current research.

For students with an interest in discrete mathematics and geometry, this is the starting point to specialize in discrete geometry. The topics addressed in the course supplement and deepen the understanding of discrete-geometric structures appearing in differential geometry, optimization, combinatorics, topology, and algebraic geometry. To follow the course, a solid background in linear algebra is necessary. Some knowledge of combinatorics and geometry is helpful.

–

Basic Literature

[1] Peter M. Gruber. *Convex and Discrete Geometry*, volume 336 of *Grundlehren Series*. Springer, 2007.

[2] Peter M. Gruber and Jörg Wills, editors. *Handbook of Convex Geometry*. North-Holland, Amsterdam, 1993. 2 Volumes.

[3] Branko Grünbaum. *Convex Polytopes*, volume 221 of *Graduate Texts in Math.* Springer-Verlag, New York, 2003. Second edition prepared by V. Kaibel, V. Klee and G. M. Ziegler (original edition: Interscience, London 1967).

[4] Jiří Matoušek and Bernd Gärtner. *Understanding and Using Linear Programming*. Universitext. Springer, 2007.

[5] Günter M. Ziegler. *Lectures on Polytopes*, volume 152 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995. Revised edition, 1998; seventh updated printing 2007.

Contents

A rough schedule, which we will adapt as we move along:

1. 1. Linear programming 1.1 On the diameter of polyhedra . . . . April 15
2. 1.2 Linear programming (Discrete geometry version) . . . . April 17
3. . . . and dual simplex algorithm . . . . April 22
4. 1.2.2 (Linear algebra version), LP duality . . . . April 24
5. 1.3 Further notes on linear programming . . . . April 29
6. 2. Convex bodies, volumes, and roundness 2.1 Basics . . . . May 6
7. . . . and examples. 2.2 Topological properties: boundary, interior . . . . (?/BMS) May 8
8. (cont.d) . . . . May 13
9. 2.3 Support and separation . . . . May 15
10. and Minkowski's theorem . . . . (?/NYC) May 20
11. 2.4 Spectrahedra . . . . May 22
12. 2.5 Löwner–John ellipsoids and roundness . . . . May 27
13. 2.6 Approximation of convex bodies by polytopes . . . . May 29
14. 2.7 On the difficulty of volume computation . . . . June 3
15. 2.8 Polarity and the Mahler conjecture . . . . June 5
16. 2.9 Volume, mixed volumes, and valuations . . . . June 10
17. 2.10 Geometric inequalities: Brunn–Minkowski and Alexandrov–Fenchel . . . . . . June 12
18. and applications . . . . June 17
19. 2.11 Isoperimetric inequalities . . . . June 19
20. 2.12 Measure concentration and high-dimensional effects . . . . June 24
21. 3. Geometry of Numbers 3.1 Lattices and lattice points . . . . June 26
22. 3.2 Minkowski's (first) theorem . . . . July 1
23. 3.3 Successive minima . . . . July 3
24. 3.4 Lattice points in convex bodies . . . . July 8
25. . . . . (?/SA) July 10
26. . . . . July 15
27. . . . . July 17

0 Introduction

What's the goal?

This is a second course in a large and interesting mathematical domain commonly known as "Discrete Geometry". This spans from very classical topics (such as regular polyhedra – see Euclid's *Elements*) to very current research topics (Discrete Geometry, Extremal Geometry, Computational Geometry, Convex Geometry) that are also of great industrial importance (for Computer Graphics, Visualization, Molecular Modelling, and many other topics).

My goal will be to develop these topics in a three-semester sequence of Graduate Courses in such a way that

- you get an **overview** of the field of Discrete Geometry and its manifold connections,
- you learn to understand, analyze, visualize, and confidently/competently argue about the basic **structures** of Discrete Geometry, which includes
  - point configurations/hyperplane arrangements,
  - frameworks
  - subspace arrangements, and
  - polytopes and polyhedra,
- you learn to know (and appreciate) the most important **results** in Discrete Geometry, which includes both simple & basic as well as striking key results,
- you get to learn and practice important **ideas and techniques** from Discrete Geometry (many of which are interesting also for other domains of Mathematics), and
- You learn about current **research topics** and problems treated in Discrete Geometry.

In this second course of the sequence, we will in particular treat the relationship between

- "discrete objects" (such as polytopes and polyhedra, but also lattices and lattice points) and
- "general objects" (such as convex bodies)

in terms of various notions of diameter, volume, and roundness.

This will not only be interesting *per se*, but also lead us to some major theorems and insight (e.g. on such fundamental notions as *volume*), but also to major applications (e.g. on sphere packings, which is in turn important for coding theory).

1 Linear programming and some applications

1.1 On the diameter of polyhedra

Let's consider a polyhedron of dimension d with n facets; let's call it an (d, n)-polyhedron.

Careful: Want to look at *pointed* polyhedron, $n \geq d$, which has a vertex, so the *lineality space* is trivial.

The *Hirsch conjecture* from 1957 is the *false* (!) statement that the edge-graph of any (d, n)-polyhedron has diameter at most $n - d$. This was disproved for unbounded polyhedra by Klee & Walkup [3] in 1967 and in general by Santos [4] in 2012. The *polynomial Hirsch conjecture* remains open: It might still be that the maximal diameter, $\Delta(d, n)$, satisfies $\Delta(d, n) \leq d(n-d)$ for all $n \geq d \geq 1$.

We will, nevertheless, see why from a "linear programming point of view" the bound $n - d$ looks natural, and even more so, why this is a relevant parameter.

Exercise 1.1. Show that $\Delta(2, n) \leq n - 2$ and $\Delta(3, n) \leq n - 3$, and that both inequalities are sharp (that is, hold with equality for $n \geq 2$ resp. $n \geq 3$).

Up to recently, the best upper bound for the diameters of polyhedra was provided by Kalai & Kleitman in a striking two page paper [2] in 1992:

$$\Delta(d, n) \leq n^{\log(d)+2},$$

which was improved only slightly by Kalai [1] to

$$\Delta(d, n) \leq n^{\log(d)+1},$$

where throughout "log" denotes the binary logarithm (i.e., base 2). However, just a few weeks ago Mike Todd (Cornell University) in a 4-page paper [5] sharpened the Kalai–Kleitman analysis to obtain

$$\Delta(d, n) \leq (n - d)^{\log(d)} = d^{\log(n-d)},$$

which indeed is sharp for $d = 1$ and $d = 2$.

In class, we will go through the arguments of Todd [5] (and thus, in particular, the idea of Kalai & Kleitman [2]).

[1] Gil Kalai. Linear programming, the simplex algorithm and simple polytopes. *Math. Programming, Ser. B*, 79:217–233, 1997. Proc. Int. Symp. Mathematical Programming (Lausanne 1997).

[2] Gil Kalai and Daniel J. Kleitman. A quasi-polynomial bound for the diameter of graphs of polyhedra. *Bulletin Amer. Math. Soc.*, 26:315–316, 1992.

[3] Victor Klee and David W. Walkup. The d-step conjecture for polyhedra of dimension $d < 6$. *Acta Math.*, 117:53–78, 1967.

[4] Francisco Santos. A counterexample to the Hirsch conjecture. *Annals of Math.*, 176:383–412, 2012.

[5] Michael J. Todd. An improved Kalai–Kleitman bound for the diameter of a polyhedron. Preprint, April 2014, 4 pages, `http://arxiv.org/abs/1402.3579`.

End of class on April 15

1.2 Geometry of linear programming and pivot rules

1.2.1 Linear programming (Discrete Geometry version)

Any system $Ax \leq b$ with $A \in \mathbb{R}^{n\times d}$, $b \in \mathbb{R}^n$ defines a polyhedron $P \subseteq \mathbb{R}^d$ with $\dim P \leq d$ and $\#$facets $\leq n$.
Without loss of generality we may assume that $\mathrm{rank} A = d$, that is the system $Ax \leq b$ has a subsystem that defines an orthant, so in particular P is either pointed (has a vertex), or is empty.
Without loss of generality (theoretically, this may be harder to compute) we may assume that $\dim P = d$, so the polyhedron is full-dimensional. Moreover, we want to get our system into the form

$$Ax \leq b,\ -x \leq 0$$

with $b > 0$ componentwise. For this we have to solve a "Phase I" problem that finds a vertex x_0 of the polyhedron, and then do a coordinate transformation that moves the vertex x_0 to 0 and transforms a system of inequalities that are tight at x_0 to the positive orthant system $x \geq 0$.
With a linear objective function we have a system of the form

$$\begin{array}{rcl} \max\ c^t x & & \\ Ax & \leq & b \\ x. & & \end{array}$$

Example:

$$\begin{array}{rrcl} \max & y & & \\ & x - y & \leq & 2 \\ & -x + y & \leq & 1 \\ & x + 2y & \leq & 7 \\ & -x & \leq & 0 \\ & y & \leq & 0. \end{array}$$

Geometric description of the polyhedron

- P is a full-dimensional polyhedron, with $\leq n$ facets, given in $\mathcal{H}$-description-
- We have a linear objective function, which might be assumed to be the last coordinate x_d, to be maximized (or in other situations: minimized).

- We assume that the polyhedron is simple, the system is in general position (this may be achieved by perturbing the right-hand sides: Exercise!).
- Any $d \times d$ full rank subsystem $A'x \leq b'$ defines a *generalized orthant*, which up to an affine transformation is equivalent to the standard positive orthant "$x \geq 0$."
- Any generalized orthant defines a point (the unique solution of $A'x = b'$) and d *rays* (by fixing all the d inequalities by one, and letting the *slack* in the last one get large).
- A generalized orthant is *feasible* if the point it defines by $A'x = b'$ is *feasible* (defines all inequalities, not only those in the subsystem). Note that this does not depend on the objective function.
- A generalized orthant is *dual feasible* if sliding along any of its rays does not improve the objective function. Note that this does not depend on the right-hand side vector b.
- A generalized orthant is *optimal* if it is both feasible and dual feasible.
- Any optimal generalized orthant defines an optimal solution of the linear program.

... and what the *primal simplex algorithm* does on it:

- We assume that after preprocessing (known as "Phase I") we have $-x \leq 0$ as a feasible generalized orthant, and in particular $x_0 = 0$ as a feasible starting vertex.
- If the generalized orthant is dual feasible, DONE with optimal solution.
- Select an improving ray, and slide along the ray. (Along the ray one inequality of the orthant is not tight any more; the objective function improves along the ray.)
- If the objective function improves without bound along the ray, DONE with optimal solution.
- Otherwise along the way we hit a bound, that is, a new facet, whose inequality completes a new feasible generalized orthant. REPEAT.

The process stops in finite time, since in every step we improve the objective function (no cycles) and there are only finitely many orthants — not more than $\binom{n}{d}$. (A better bound is obtained from the upper bound theorem — need a version for unbounded polyhedra: Exercise!)

End of class on April 17

Alternatively, here is what the *dual simplex algorithm* does on a linear program:

- We assume that after preprocessing (known as "Phase I") we have found a dual feasible generalized orthant, which in particular defines a current solution (vertex of the system, but not necessarily of the polyhedron).
- If the generalized orthant is feasible, DONE with optimal solution.
- Select an inequality violated by the current solution.
- If the violating inequality hits none of the rays of the current generalized orthant, then DONE with proof that the system is infasible.
- Otherwise construct a new dual feasible generalized orthant whose current solution gives a better upper bound on the maximum of the system. REPEAT.

The process stops in finite time, if we take care that in every step we improve the current upper bound on the objective function values on the polyhedron (no cycles) and there are only finitely many generalized orthants.

End of class on April 22

1.2.2 Linear programming (Numerical Linear Algebra version)

We write down two linear programs, in the following form.
The *primal linear program* is

$$
\begin{array}{llrcl}
(P) & \quad \max\ c^t x & & & \\
 & & Ax & \leq & b \\
 & & x & \geq & 0.
\end{array}
$$

The associated *dual linear program* is

$$
\begin{array}{llrcl}
(D) & \quad \min\ b^t y & & & \\
 & & A^t y & \geq & c \\
 & & y & \geq & 0.
\end{array}
$$

Lemma 1.2 (Weak Duality Theorem)**.** *If for a primal-dual pair of linear programs x_0 is a feasible solution for the primal (P) and y_0 is a feasible solution for the dual (D), then*

$$c^t x_0 \leq b^t y_0.$$

In particular, the maximum of (P) is smaller or equals to the minimum of (D).

Proof. We compute

$$c^t x_0 \leq (A^t y_0)^t x_0 = y_0^t (A x_0) \leq y_0^t b = b^t y_0.$$

□

The linear programs are then, by introduction of *slack variables*, converted into systems of linear equations, to be solved in non-negative variables.
Thus the primal linear program becomes

$$
\begin{array}{lrcl}
(P) & \quad \max\ c^t x + 0^t \hat{x} & = & \gamma \\
 & Ax + I_n \hat{x} & = & b \\
 & x \geq 0,\ \hat{x} \geq 0 & &
\end{array}
$$

This system has an "obvious" current solution, given by $x \equiv 0$ (the "non-basic variables" are set to 0: these correspond to the inequalities that define the current generalized orthant), $\hat{x} = b$ (the "basic variables" are uniquely determined). This starting solution has the value $\gamma = 0$.
These systems are manipulated by *row operations*, which do not change the solution space. Thus after a number of steps we still have the system in the form

$$
\begin{array}{lrcl}
(P) & \quad \max\ \bar{c}^t x_N + 0^t x_B & = & \bar{\gamma} \\
 & \bar{A}_N x_N + I_n x_B & = & \bar{b} \\
 & x_N \geq 0,\ x_B \geq 0 & &
\end{array}
$$

Here the columns have been resorted, to keep the "basic variables" and the "non-basic variables" together, that is, the index sets B and N together give the set of all columns labelled by $B \cup N =$

$\{1, 2, \ldots, d{+}n\}$. The coefficients in the system are $\bar{A}_N = A_B^{-1} A_N$, and $\bar{b} = A_B^{-1} b$. The objective function has been rewritten in terms of the non-basic variables. Its coefficients

$$\bar{c}_N^t = c_N^t - c_B^t A_B^{-1} A_N$$

are known as the *reduced costs*: in the geometric interpretation they give the slopes of the rays of the current generalized orthant.
The current solution is given by $x_N \equiv 0$, which uniquely determines the non-basic variables to be $x_B = \bar{b} = A_B^{-1} b$.
Thus the (current solution of the) system is *feasible* if $\bar{b} \geq 0$, and it is *dual feasible* if $\bar{c}_N \leq 0$.
A similar treatment/computation can be done for the dual system (D).

Lemma 1.3. *For any pair of primal linear program (P) and its dual program (D) in the equation form given above,*

- *the bases B for the system (P) are in bijection with the non-bases N of the system (D);*
- *the feasible bases for (P) are in bijection with the dual-feasible non-bases for (D);*
- *etc.*

Proof. This rests on the observation that in the $(n+d) \times (n+d)$ matrix

$$\begin{pmatrix} A & I_n \\ -I_d & A^t \end{pmatrix}$$

the row space spanned by the first n rows is the orthogonal complement of the space spanned by the last d rows. □

Theorem 1.4 (Duality Theorem for Linear Programming)**.** *If a primal linear program (P) and its dual (D) are both feasible, then they have optimal solutions x^* and $y*$, and these have the same optimal value.*
If one of the programs is not feasible, then the other one is either infeasible as well, or it is unbounded.

Proof. The optimal solutions exist, since the Simplex Algorithm will find it! □

From the geometry of an optimal basis/optimal generalized orthant, we also get *complementary slackness*: If in the optimal solution an inequality is not tight, then the corresponding variable in the dual program is zero; if a variable is positive, then the corresponding dual inequality has to be tight. This can also be seen from analysis of the inequalities in the proof of the Weak Duality Theorem.
The optimal solution to a linear program can be computed *efficiently*:

In Practice there are commercial, as well as non-commercial, software libraries for linear programming, which include implementations of the Primal Simplex Algorithm, the Dual Simplex Algorithm, as well as other methods (such as *Interior Point Methods*) which will solve to optimality practically every linear program that appears in practice.

In Theory there are two different computational models:

In the bit model the "Ellipsoid Method" (which will appear later in this course) is a polynomial time method for solving linear programs, whose running time is polynomial in the bit-size of the input. This method is theoretically very important, but has not been implemented in practice.

In the unit cost model the Simplex Algorithm with a suitable choice of variable selections ("pivot rule") may be polynomial — but this has not been proven. Indeed, we do not even know whether in general there is any short (i.e. polynomially many edges) path from a given starting vertex of the program to the optimal vertex. The best upper bound is the $n^{\log_2 d}$ upper bound discussed at the beginning of this course — and this bound is *not* a polynomial in n and d. An upper bound of the type $d(n-d)$ might exist, but has not been proven.

Thus the complexity of Linear Programming, and in particular of the Simplex Algorithm, is a major open problem both for Optimization, and for Discrete Geometry!

End of class on April 24

1.3 Further Notes on Linear Programming

Let's step away from the simplex algorithm, and let's look at the problem itsself — and let's assume we have a solution method (algorithm, perhaps software) that solves the problem, but which we can treat as a "black box." This is the *oracle* view, which has become popular in optimization, with grave consequences for (computational) discrete and convex geometry: well-defined input, well-defined output; estimate complexity

Examples:

LP-OPTIMIZATION problem/oracle:

INPUT: $d \geq 1$, $n \geq 1$, $A \in \mathbb{Q}^{n\times d}$, $b \in \mathbb{Q}^n$, $c \in \mathbb{Q}^d$

TASK: $\max c^t x : \ Ax \leq b,\ x \geq 0$

OUTPUT: optimal solution $x^* \in \mathbb{Q}^d$, with certificate (basis)
or information that problem is infeasible, with certificate (basis & inequality),
or information that problem is unbounded, with certificate (basis & ray).

LP-FEASIBILITY problem/algorithms/oracle:

INPUT: $d \geq 1$, $n \geq 1$, $A \in \mathbb{Q}^{n\times d}$, $b \in \mathbb{Q}^n$

TASK: find $x : \ Ax \leq b,\ x \geq 0$

OUTPUT: feasible solution $x^* \in \mathbb{Q}^d$, with certificate (basis)
or information that problem is infeasible, with certificate (basis & inequality).

Note: Any algorithm for solving **LP-OPTIMIZATION** can be used to solve **LP-FEASIBILITY**. We will see that the other direction "works as well."

Note: Two algorithms we know/could work out for **LP-OPTIMIZATION**: Fourier–Motzkin elimination (see Discrete Geometry I), and the Simplex Algorithm.

1.3.1 Complexity issues

Could it be that the solution exists, but it is too large (or too small) to write down in reasonable time?

Real input/solutions don't make sense, or need work to make sense of.

Recommended reading: Lovász' lecture notes [**?**].

Could get answer from Fourier–Motzkin elimination.

Here: get answer from simplex and Cramer's rule and Hadamard inequality.

Lemma 1.5 (Hadamard inequality)**.** *Let $A \in \mathbb{R}^{n\times n}$ be a matrix with columns $A = (A_1, \ldots, A_n)$. Then*

$$|\det A| \ \le \ |A_1| \cdots |A_n|.$$

Lemma 1.6 (The Cramer's rule estimate)**.** *Let $A \in \mathbb{Z}^{n\times n}$, $b \in \mathbb{Z}^n$, $\det A \neq 0$ (integer data!). Then the (rational!) solution for the system of equations $Ax = b$ satisfies*

$$|x_i| \le |A_1| \cdots |A_n| \cdot |b|.$$

Proof. Cramer's rule, together with the observation that the denominator, $\det A$, is an integer, so its absolute value is at least 1. The same is true for the length of each column $|A_i|$. □

1.3.2 Feasibility

First, we should discuss the problem how to find a feasible generalized orthant for $Ax \le b$, $x \ge 0$, in order to even *start* the simplex algorithm. Here are two solutions to that problem:

- Use the complexity estimates to get explicit upper bounds for the variables, and thus have a starting basis for the dual simplex algorithm (that is, a feasible basis for the simplex algorithm applied to the dual program).
- Phase I: Write down an artificial OPTIMIZATION program, which is feasible, and whose optimal solution (basis) will give a feasible solution (and a feasible basis!) for the FEASIBILITY problem: For example

$$\min x_0 : \ Ax - x_0\mathbf{1} \le b, \ x \ge 0, x_0 \ge 0.$$

It is trivial that if we can solve **LP-OPTIMIZATION** then we can solve **LP-FEASIBILITY**, in a way that is completely independent of the the specific algorithm used to "implement" **LP-OPTIMIZATION**; that is, we can use any **LP-OPTIMIZATION** *oracle* to "simulate" an **LP-FEASIBILITY** algorithm; in other words, we can program a (fast) algorithm for **LP-FEASIBILITY** if we can use a (fast) subroutine for **LP-OPTIMIZATION** (e.g. by putting objective function zero).

However, note that the converse is also true: If we know how to solve **LP-FEASIBILITY**, then we can also solve **LP-OPTIMIZATION**, that is,

LP-FEASIBILITY $\Longrightarrow$ **LP-OPTIMIZATION**.

For this, note that any *feasible* solution (x, y) for the primal-dual program

$$\begin{array}{llll} & & c^t x \ge b^t y & \\ (PD) & Ax \le b & & A^t y \ge c \\ & x \ge 0 & & y \ge 0 \end{array}$$

1.3.3 Modelling issues

Conversion of programs from equality form to inequality form, and conversely. See the Exercises.

1.3.4 Perturbation techniques

If we replace the right-hand sides b_i by $b_i + \varepsilon^i$, for a suitably small ε, then

- the perturbed problem will be feasible if *and only if* the original problem is feasible,
- the perturbed problem will be primally *non-degenerate*, that is, it describes a simple polyhedron, and at any generalized orthant (basis), no extra inequalities are tight (that is, the non-basic variables are non-zero).

(see Exercise).

Moreover,

- similarly, by perturbing the objective function the program can be made *dually non-degenerate*, so that in particular the optimal solution is unique (if it exists), and
- the suitable $\varepsilon > 0$ can be estimated explicitly.

1.3.5 Integral solutions? An example

In general, the optimal solutions will not be integral, although many applications ask for integral solutions. Even if we find the best integral solution, this will come without a certificate, as there may be not dual constraints that are tight at the best integer solution.

However, in many combinatorial situations, we are lucky. Here is one example.

Example 1.7 (Network flows). If the bounds on each arc are integral, then the optimal solution will be integral.

(This may be seen from an algorithm by successive improvement, or from a matrix argument, see exercise.)

Interpretation of dual solutions: Max cut!

Max-Flow-Min-Cut theorem!

Exercise 1.8. Let $A \in \{0, 1, -1\}^{n\times n}$ be a $0/\pm 1$ matrix. Show that

(i) The determinant of a $0/1$-matrix A can be large, even if there are only two 1s per row.

(ii) The determinant of A is not large if there is at most one 1 and at most one -1 per row.

(iii) Use the Hadamard inequality to give an upper bound on $|\det A|$

(iv) For $A \in \{0, 1\}^{n\times n}$ give a much better upper bound, by

— Multiplying the matrix by 2,

— Adding a column of 0's and then a row of 1's,

— subtracting the first row from all others

and then applying Hadamard to the resulting ± 1-matrix.

(v) Give examples where this bound is tight.

End of class on April 29

2 Convex Bodies, Volumes, and Roundness

> "Although convexity is a simple property to formulate, convex bodies possess a surprisingly rich structure" (Keith Ball [?])

Archimedes book "On the sphere and the cylinder"

2.1 Some basic definitions and examples

Definition 2.1 (Linear/affine/conical/convex hulls)**.** Define in $\mathbb{R}^n$:

- *Linear subspace*, *linear hull*
- *Affine subspace* (possibly empty), *affine hull*
- *Conical subspace* (= *convex cone*, or simply *cone*), *conical hull*
- *Convex hull*, *convex set*

Definition 2.2 (Convex set, line-free, bounded, convex body)**.** Define in $\mathbb{R}^n$: A convex set is

- *line free*: does not contain an affine line
- *bounded*: does not contain a ray
- *convex body*: a closed, bounded (that is, compact) full-dimensional convex set
- *strictly convex*: if $\lambda x + (1-\lambda)y \in \mathrm{int}C$ for $0 < \lambda < 1$ and $x \neq y$.

Examples:

- linear and affine subspaces
- convex polygons in the plane
- regular polyhedra in 3-space

Example 2.3. The unit ball of $\mathbb{R}^d$ with ℓ_2 norm is a centrally-symmetric proper convex body.

Indeed, convexity follows from the triangle inequality

Thus the "theory of finite-dimensional Banach spaces" is equivalent to the "theory of centrally-symmetric convex bodies."

For example, Dvoretzky's theorem, which says that every centrally symmetric convex body in $\mathbb{R}^n$ has a central section of dimension roughly $\log n$ that is linearly approximately equivalent to some $\mathbb{R}^d$ with the Euclidean norm, is a theorem about centrally symmetric convex bodies, which have sections that are roughly ellipsoids. (Indeed, concentration of measure implies that a random subspace will do . . .)

End of class on May 6

Example 2.4. The set PSD_n of positive semi-definite $(n \times n)$-matrices is a closed convex cone in $\mathbb{R}^{n\times n}$ of dimension $\binom{n+1}{2}$.

Example 2.5. If identify the N-dimensional vector space $\mathbb{R}[x_1, \dots, x_d]_{\leq 2k}$ of real polynomials in d variables of degree less than $2k$ with $\mathbb{R}^N$. Then the set

$$\mathcal{P}_{d,2k} = \{p\,:\; p \in \mathbb{R}[x_1, \dots, x_d]_{\leq 2k} \text{ and } p(x) \;\geq\; 0 \text{ for all } x \in \mathbb{R}^d\}$$

of positive polynomials is a closed, unbounded convex cone in $\mathbb{R}^N$. Similarly the set

$$\Sigma_{d,2k} = \{p\,:\; \exists h_1, \dots, h_n \in \mathbb{R}[x_1, \dots, x_d]_{\leq k} \text{ such that } p(x) = h_1^2(x) + \dots h_n^2(x)\}$$

of sums of squares (SOS) is a closed, unbounded convex cone in $\mathbb{R}^N$.

Exercise: The dimension of the vector space $\mathbb{R}[x_1, \dots, x_d]_{\leq 2k}$ is $\binom{2k+d}{d}$.

Theorem 2.6 (Gauß–Lucas Theorem)**.** *Let $p \in \mathbb{C}[z]$ be a complex polynomial in one variable with roots $r_1, \dots, r_d$, then the roots of the derivative p' of p are contained in the convex hull* $\operatorname{conv}\{r_1, \dots, r_d\}$.

Proof. If r_1 is a zero of p as well as of p', then $r_1 = 1r_1 + 0r_2 + \cdots + 0r_d$ is a convex combination. Assume z is a zero of p' but not of p. Write p and p' in terms of their roots (they factor over $\mathbb{C}$) and look at $\frac{p(z)}{p'(z)}$ to get a convex combination of the r_i. □

If p has only real roots, then the above result is a consequence of the Rolle's theorem (or the mean value theorem).

2.2 Topological properties

Theorem 2.7 (Carathéodory)**.** *Let $A \subseteq \mathbb{R}^d$ be a set and $x \in \operatorname{conv}(A)$ a point in the convex hull of A. Then there are $d + 1$ points $p_1, \dots, p_{d+1}$ in A such that $x \in \operatorname{conv}\{p_1, \dots, p_{d+1}\}$.*

Proof. Write $x \in \operatorname{conv}(A)$ as convex combination of a *minimal* number of points $p_1, \dots, p_n \in A$. If n is more than $d + 1$, then there is an affine dependency, where one of the coefficients is positive. Subtract a multiple of this dependency to kill one of the coefficients of the convex combination. □

Corollary 2.8. *If $A \subset \mathbb{R}^d$ is compact, then so is $\operatorname{conv}(A)$.*

Proof. Let $\Delta_d = \operatorname{conv}\{e_1, \dots, e_{d+1}\}$ be the standard d-simplex in $\mathbb{R}^{d+1}$. Consider a map from $A^{d+1} \times \Delta_d \to A$ given by $(p_1, \dots, p_{d+1}, \lambda) \mapsto \sum \lambda_i p_i$. Clearly its image is contained in $\operatorname{conv}(A)$. The converse is true by Carathéodory's theorem. Hence $\operatorname{conv}(A)$ is the image of a compact set under a continuous map. □

End of class on May 8

Definition 2.9 (interior points, interior, boundary points, boundary)**.**

Definition 2.10 (relative interior, relative boundary)**.**

Proposition 2.11. *If K is convex, then $\operatorname{relint} K$ is also convex.*

Lemma 2.12. *If $x_0, \dots, x_k$ are affinely independent, $P := \operatorname{conv}\{x_0, \dots, x_k\}$, then $x \in \operatorname{relint} P$ if and only if $x = \lambda_0 x_0 + \cdots + \lambda_k x_k$ with all $\lambda_i > 0$.*

Corollary 2.13. *K convex, not empty, then $\operatorname{relint} K \neq \emptyset$.*

Definition 2.14 (dimension of a convex set)**.**

Theorem 2.15 (Carathéodory's Theorem — ambient space free version)**.**

Corollary 2.16. *Characterization of relative interior of a polytope $P := \operatorname{conv}\{x_0, \dots, x_n\}$: $x = \lambda_0 x_0 + \cdots + \lambda_k x_k$ with all $\lambda_i > 0$.*

Definition 2.17 (extreme points)**.**

Theorem 2.18 (Minkowski)**.** *K closed and bounded convex set, then $K = \operatorname{conv}(\operatorname{ext} K)$.*

End of class on May 13

2.3 Support and separation

Definition 2.19 (Nearest point map)**.** Let $A \subset \mathbb{R}^d$ be a non-empty closed convex set. Then the *nearest-point map* of A is the map $\pi_A : \mathbb{R}^d \to A$ which assigns to each $x \in \mathbb{R}^d$ the point on A with the smallest (Euclidean) distance from x.

Proposition 2.20. *The map π_A of "Definition" 2.19 exists (that is, the nearest point exists, lies in A, and is unique) and the map is contractive:*

$$\|\pi_A(x) - \pi_A(y)\| \leq \|x - y\|$$

and thus in particular continuous.

Exercise 2.21. There is a converse: If for a closed set A the nearest point $\pi_A(x)$ is unique for all x, then A is convex.

Notation: H hyperplane, H^+, H^- half spaces: They are closed convex sets, their interiors are $\mathrm{inter}(H^+) = \mathbb{R}^d \setminus H^-$ and $\mathrm{inter}(H^-) = \mathbb{R}^d \setminus H^+$, their boundary is $\partial H^+ = \partial H^- = H$.

Definition 2.22 (separates)**.** If $A \subseteq \mathbb{R}^d$ is a convex set and $p \in \mathbb{R}^n$, then a hyperplane H *separates* p from A if $p \in H^+$ and $A \subseteq \mathrm{inter}(H^-)$, that is, $A \cap H^+ = \emptyset$. The hyperplane H *strictly separates* A and p if $A \subseteq \mathrm{inter}(H^-)$ and $p \in \mathrm{inter}(H^+)$.
If $A, B \subseteq \mathbb{R}^d$ are convex sets, then a hyperplane H *separates* B from A if $B \subset H^+$ and $A \subseteq \mathrm{inter}(H^-)$. The hyperplane H *strictly separates* A and B if $A \subseteq \mathrm{inter}(H^-)$ and $B \subseteq \mathrm{inter}(H^+)$.

Note that if separation implies that the sets are disjoint, and strict separation implies weak separation. However, separation is not symmetric: There may be a hyperplane that separates B from A, but none that separates A from B.

Theorem 2.23 (Separation Theorem)**.** *Let $A \subseteq \mathbb{R}^d$ be a non-empty closed convex set and $p \notin A$, then there is a hyperplane that strictly separates p and A.*

Proof. Set $q := \pi_A(p)$, $c := p - q$, $H := \{x \in \mathbb{R}^d : c^t x = c^t q\}$ and $H_{1/2} := \{x \in \mathbb{R}^d : c^t x = c^t q\}$.
An elementary geometric argument shows that $A \subset H^-$, while $p \notin H^-$, such that H separates p from A with $q \in H$, and $H_{1/2}$strictly separates p and A. □

Definition 2.24 (supporting hyperplane)**.** A *supporting hyperplane* H for a convex set A satisfies $A \subseteq H^-$ and $A \cap H \neq \emptyset$.

... so this exists by the (proof of the) Separation Theorem.

Corollary 2.25. *closed convex set is intersection of half spaces given by supporting hyperplanes*

Corollary 2.26. *If A is a convex body, then for each direction $c \neq 0$ there is a unique supporting hyperplane $H = \{x : c^t x = \delta\}$.*

Definition 2.27 (support function)**.** Convex body A, define $h_A : \mathbb{R}^d \to \mathbb{R}$ by $h_A(c) := \max\{c^t x : x \in A\}$.

Corollary 2.28. *Convex body is determined by its support function.*

Definition 2.29 (Minkowski sum)**.** The *Minkowski sum* of two sets $A, B \subseteq \mathbb{R}^d$ is

$$A + B := \{x + y : x \in A,\ y \in B\}.$$

Lemma 2.30. *If A and B are convex, then so is $A + B = B + A$.*

Lemma 2.31. *K, L, M convex bodies. Then*
(i) $h_{K+L} = h_K + h_L$.
(ii) $K + M = L + M$ implies $K = L$.

Remark 2.32. We have just established that the set of convex bodies $\mathcal{K}_d$ is a cancellative commutative monoid.

End of class on May 15

Theorem 2.33 (Supporting Hyperplane Theorem)**.** *Let $A \subset \mathbb{R}^d$ be a closed and convex set. Then for every point p in the (relative) boundary ∂A of A there is a supporting hyperplane $H_A(p)$ for A at p, that is, $A \subseteq H_A^-(p)$ and $p \in H \cap A$.*

Proof. If A is not full-dimensional, replace the ambient space by an affine subspace of dimension $\dim(A)$. A supporting hyperlane in this subspace lies inside some (actually many) hyperplanes in $\mathbb{R}^d$, all of which are supporting. So assume A is full-dimensional. Via the nearest point map π_A we get a supporting hyperplane for A at each point $\pi(y)$ for $y \in \mathbb{R}^d \setminus A$ with unit normal vector $\frac{y - \pi_A(y)}{\|y - \pi_A(y)\|_2}$. Take a series $(y_n) \subset \mathbb{R}^d \setminus A$ that converges to p. The corresponding sequence of unit normal vectors $u_n := \frac{y_n - \pi_A(y_n)}{\|y_n - \pi_A(y_n)\|_2}$ for the supporting hyperplanes at $\pi_A(y_n)$ has, by compactness of the unit sphere, a subsequence converging to $u \in S^{d-1}$. There is a corresponding subsequence of (y_n) that also converges to p. Using convergence of the sequences and continuity of the inner product argue that $H_A(p) := \{x \in \mathbb{R}^d : u^t x = u^t p\}$ is a supporting for A at p. □

Proof of Minkowski's Theorem 2.18. The inclusion $K \supseteq \operatorname{conv}(\operatorname{ext} K)$ is trivial. For the other inclusion argue by induction on $d = \dim(C)$. The cases $d = 0, 1$ are trivial. Assume the theorem holds for all compact and convex sets of dimension less than d. Assume K has dimension d. Let $p \in \partial K$. Then, by Theorem 2.33 above, there is a supporting hyperplane $H_K(p)$ for K at p. The "face" $F := K \cap H_K(p)$ is of lower dimension and hence $p \in \operatorname{conv}(\operatorname{ext} F)$. By the homework assignment $\operatorname{ext} F \subseteq \operatorname{ext} K$ and hence $p \in \operatorname{conv}(\operatorname{ext} K)$. If $p \in \operatorname{relint}(K)$ take a line through p that intersects ∂A in two points. Argue using faces that these points are in $\operatorname{conv}(\operatorname{ext} K)$, so p must be in $\operatorname{conv}(K)$ as well. □

2.4 Spectrahedra

Definition 2.34. A *spectrahedron* S is the intersection of the cone PSD_n of symmetric positive-semidefinite matrices with a d-dimensional affine subspace V (of the space of symmetric $n \times n$ matrices). If A is positive semi-definite we write $A \succeq 0$.

Proposition 2.35. *A spectrahedron S is convex and closed. It can be written as*

$$S = \{(x_1, \dots, x_d) \in \mathbb{R}^d \colon A_0 + x_1 A_1 + \dots x_d A_d \succeq 0\},$$

for suitable symmetric matrices $A_0, \dots, A_d$ of size $n \times n$. Let $A(x) := A_0 + x_1 A_1 + \dots x_d A_d$ denote the (symmetric) matrix valued function from $\mathbb{R}^d \to \mathbb{R}^{n \times n}$.

Example 2.36. The cylinder

$$C := \{(x, y, z) \in \mathbb{R}^3 \colon x^2 + y^2 \leq 1, -1 \leq z \leq 1\}$$

is a spectrahedron. Consider the points $(x, y, z) \in \mathbb{R}^3$ such that the sum

$$A_0 + xA_1 + yA_2 + zA_3 = \begin{pmatrix} 1+x & y & 0 & 0 \\ y & 1-x & 0 & 0 \\ 0 & 0 & 1+z & 0 \\ 0 & 0 & 0 & 1-z \end{pmatrix} \succeq 0.$$

Here A_0 is the identity matrix. A_1 has a 1 in position $(1, 1)$ and a -1 at $(2, 2)$ and otherwise zeros. A_2 is zero except for 1s at $(1, 2)$ and $(2, 1)$. Finally, A_3 is zero except for a 1 at $(3, 3)$ and a -1 at $(4, 4)$. It turns out that C is the set of all points $w = (x, y, z)$ that satisfy $A(w) \succeq 0$. The cylinder C can also be viewed as the intersection of PSD_4 with the affine subspace $A_0 + \mathrm{span}\{A_1, A_2, A_3\}$.

Proposition 2.37. *Any polyhedron P is a spectrahedron.*

Proof commented, since it is a current exercise. □

Example 2.38. Any univariate sum of squares (SOS) polynomial $p \in \mathbb{R}[t]$ of degree $2n$ that can be written as

$$p = (1, t, t^2, \dots, t^n)^t \begin{pmatrix} 1 & 0 & a \\ 0 & 1-2a & 0 \\ a & 0 & 1 \end{pmatrix} (1, t, t^2, \dots, t^n)$$

defines a spectrahedron S, where S is given by all a such that the matrix is positive semi-definite. Actually $S = [-1, 1/2]$. This extends to polynomials of higher degree that can be written as $\mathbf{t}^t A \, \mathbf{t}$ for positive semi-definite A.

Example 2.39 (Non-example). Consider the (linear) projection of the cylinder C into the plane given by $x + 2z = 0$. What we get is the convex hull C' of two non-intersecting ellipses in the plane. Recalling that a matrix is positive semidefinite if the determinants of all of its diagonal minors are non-negative, we can conclude that any spectrahedron must be a so-called *basic semialgebraic set*, that is, a set of points satisfying finitely many polynomial inequalities where the polynomials are of finite degree. Using the fact that infinitely many points determine a polynomial of finite degree one can argue that C' is not basic semialgebraic, hence implying that C' is not a spectrahedron.

End of class on May 20 + May 22

2.5 Löwner–John ellipsoids and roundness

Definition 2.40. An *ellipsoid* $E \subseteq \mathbb{R}^d$ is the image $f(B^d)$ of the unit ball under an invertible affine transformation $f : \mathbb{R}^d \to \mathbb{R}^d$.

If the transformation is $f : x \mapsto Ax + c$, then

$$\begin{aligned} f(B^d) &= \{x \in \mathbb{R}^d : \langle A^{-1}(x-c), A^{-1}(x-c)\rangle \leq 1\} \\ &= \{x \in \mathbb{R}^d : \langle Q(x-c), x-c\rangle \leq 1\} \end{aligned}$$

for $Q = A^* A^{-1} = (AA^t)^{-1}$ positive-definite.

Lemma 2.41. *The volume of E is* $|\det A| \mathrm{vol} B^d = \frac{\mathrm{vol} B^d}{\sqrt{\det Q}}$.

Exercise 2.42. If $E = \{x \in \mathbb{R}^d : \langle Qx, x\rangle \leq 1\}$, show that the polar is $E^* = \{x \in \mathbb{R}^d : \langle Q^{-1}x, x\rangle \leq 1\}$. Deduce that $(\mathrm{vol} E)(\mathrm{vol} E^*) = (\mathrm{vol} B^d)^2$.

Exercise 2.43. If $g : \mathbb{R}^d \to \mathbb{R}^d$ is a surjective linear map, and $E \subset \mathbb{R}^d$ is an ellipsoid, then $g(E)$ is an ellipsoid in $\mathbb{R}^k$.

Lemma 2.44. *Every ellipsoid $E \subset \mathbb{R}^d$ can be written in the form $E = S(B^d) + c$, where S is a positive-definite (symmetric) matrix.*

Proof. Use the (left) polar decomposition: every invertible A can be written as $A = P'U$, where $U = A\sqrt{A^tA}^{-1}$ is a unitary matrix, and $P' = AU^{-1} = \sqrt{AA^t}$ is positive-definite. Then $A(B^d) = S(B^d)$. □

Lemma 2.45. *If X, Y are positive-definite (symmetric square) matrices, then*

$$\det\left(\frac{X+Y}{2}\right) \geq \sqrt{\det(X)\det(Y)},$$

with equality if and only $X = Y$.

Proof. We can write $X = U^t D^2 U$ for unitary U and non-negative diagonal D, and with this $Y = U^t DY'DU$. With this we obtain that without loss of generality $X = I_d$.
Furthermore, the resulting Y' can be diagonalized, and without loss of generality Y is diagonal. Then things reduce to simple inequalities of the form $\frac{1+\lambda_i}{2} \geq \sqrt{\lambda_i}$ for certain positive eigenvalues λ_i. □

Theorem 2.46 (Löwner–John). *If $K \subset \mathbb{R}^d$ is a convex body, then the maximum-volume ellipsoid $E \subseteq K$ exists and is unique.*

Proof. For the existence, consider the set

$$X := \{(S, c) : S \text{ positive semidefinite},\ c \in \mathbb{R}^d, S(B^d) + c \subseteq K\}.$$

By Lemma 2.44, every ellipsoid in K is represented by a pair (S, c) in X. As K is bounded, we get that X is bounded. It is also closed, so it is compact. Moreover, the volume function on X, given by $\det(S)\mathrm{vol}(B^d)$, is continuous, so the maximum exists.

To show that it is unique, first note that from any two ellipsoids of the same maximum volume $E_1 = S_1(B^d) + c_1$ and $E_2 = S_2(B^d) + c_2$ we can construct a new one $\frac{1}{2}(E_1 + E_2)$ given by $S := \frac{1}{2}(S_1 + S_2)$ and $c := c_1 + c_2$. Lemma 2.45 now yields that if both E_1 and E_2 have maximal volume, then $S_1 = S_2$.

To see $c_1 = c_2$, we may now after a coordinate transformation assume that $S_1 = S_2 = I$ is a unit ball. So we just have to show that the convex hull of the union of two distinct unit balls contains an ellipsoid of larger volume. □

Theorem 2.47. *The minimal volume ellipsoid that contains a given convex body K is also unique.*

Theorem 2.48. *Let $K \subset \mathbb{R}^d$ be a convex body and let $E \subseteq K$ be the maximal volume ellipsoid in K, where we assume that its center is the origin 0. Then*

$$E \subseteq K \subset dE.$$

Proof. Elementary calculation. □

Theorem 2.49. *Let $K = -K \subset \mathbb{R}^d$ be a centrally-symmetric convex body and let $E \subseteq K$ be the maximal volume ellipsoid in K. Then*

$$E \subseteq K \subset \sqrt{d}E.$$

Proof. Elementary calculation. □

End of class on May 27