Convex Optimization in Machine Learning and Computational Finance

Lecture 1: Introduction, Convex Sets

Dr. Michel Baes, Pr. Patrick Cheridito

RiskLab / ETH Zürich

Practical information

Check the website!

How to reach me? `mbaes@math.ethz.ch`

How to pass your exam?

▷ Do your weekly exercises well

▷ Know your stuff

▷ For PhDs: the exam can be replaced by a term project.

Every engineering field

can benefit

from Convex Optimization

1. Finance: designing portfolios

An investor can put his money in n different assets, each of which with a certain expected return and a certain risk.

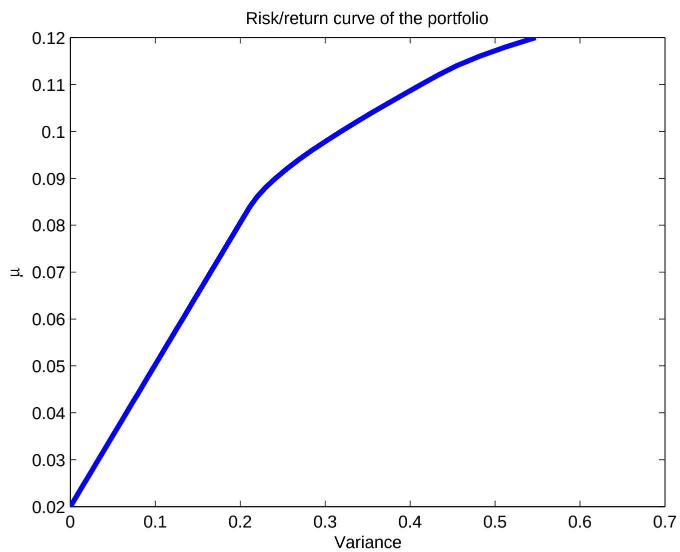
- Proportion of asset i in portfolio is x_i
- Expected return of asset i is p_i; desired return is μ
- Risk is measured by the variance of the portfolio; we know the covariance matrix Σ of the assets.

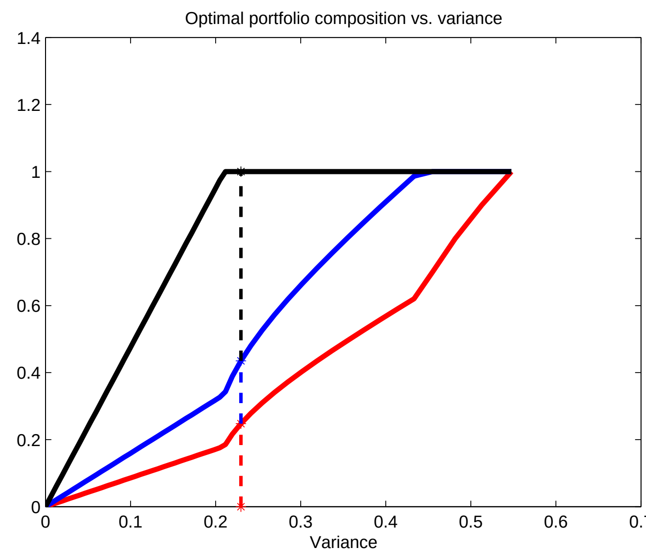
Questions: what is the minimal risk of a portfolio with return μ? How does this portfolio look like?

$$\begin{array}{ll} \text{min} & x^T \Sigma x \\ \text{s.t.} & p^T x = \mu \\ & \sum_i x_i = 1,\ x_i \geq 0 \end{array}$$

1. Finance: designing portfolios

$$\begin{array}{ll} \min & x^T \Sigma x \\ \text{s.t.} & p^T x = \mu \\ & \sum_i x_i = 1,\ x_i \geq 0 \end{array}$$





2. Finance again: Computing covariances

Fact: Computing Σ is not as obvious as it seems!

Given observations $X_{11}, X_{12}, \cdots, X_{1N}$ of Asset 1,
$X_{21}, X_{22}, \cdots, X_{2N}$ of Asset 2,...
$X_{n1}, X_{n2}, \cdots, X_{nN}$ of Asset n,

Statistics say: if $\bar{X}_{\ell:} := \sum_{k=1}^{N} X_{\ell k}/N$, then

$$\Sigma_{ij} \approx \hat{\Sigma}_{ij} := \frac{\sum_{k=1}^{N} (X_{ik} - \bar{X}_{i:}) \left(X_{jk} - \bar{X}_{j:}\right)}{(N-1)} \text{ for all } i, j.$$

In reality, the data are not sampled at the same frequency.
The data size N might not be the same for all assets.
That might cause $\hat{\Sigma}$ not to be positive definite
and the optimal solution completely meaningless!

2. Finance again: Computing covariances

Fact: Computing Σ is not as obvious as it seems!

Note 1: When $\hat{\Sigma}$ is not positive semidefinite,
Minimizing $x^T \hat{\Sigma} x$ can yield negative values.

So, we'll try to find S positive semidefinite and close to $\hat{\Sigma}$.

Note 2: Under a multivariate normal model
with covariance Σ, assets i and j
are independent iff $[\Sigma^{-1}]_{ij} = 0$.

We identify a list Γ of asset pairs
that are thought to be independent.

$$\min_{S \text{ is psd}} ||S - \hat{\Sigma}||_F + C \sum_{\{i,j\} \in \Gamma} |[S^{-1}]_{ij}|$$

3. Reconstructing Mozart

$$y = Ax + z$$

- $y \in \mathbb{R}^n$ is a big vector that we measure
- $x \in \mathbb{R}^m$ is an unknown signal, with m **huge**
  Also, x has **lots** of zeros
- A is a known matrix
- z is a bounded noise

Examples:

1. $x \in \mathbb{R}^m$ represents an image (1 component = 1 pixel)
   A describes a wavelet transform
2. $x \in \mathbb{R}^m$ is a piece of music in the frequency domain
   A is a set of discretized sine functions
   z comes from a *quantization* (i.e. $y \in \mathbb{Z}^n$)

3. Reconstructing Mozart

$$y = Ax + z$$

$$\min\{\mathtt{nnz}(x) : ||Ax - y||_p \le \epsilon\} \qquad (\spadesuit)$$

- Non-convex, **highly unstable** objective

$$\rightsquigarrow \qquad \min\{||x||_1 : ||Ax - y||_p \le \epsilon\} \qquad (\heartsuit)$$

- When is the solution of ♡ also a solution of ♠?
- Is a solution of ♡ far from a solution of ♠?

We relate - the noise size
- the sparsity of x
- the *coherence* of A: $\max_{i \neq j} |a_i^T a_j|$

with the error made by solving ♡ instead of ♠.

Which problem is easier?

Problem **A**.
Let $\Gamma \subseteq \{(i,j) : 1 \leq i,j \leq n\}$, and

$$\begin{array}{rll} \text{max} & \sum_{i=1}^n x_i & \\ \text{s.t.} & x_i x_j = 0 & \forall\,(i,j) \in \Gamma, \\ & x_i^2 - x_i = 0 & \forall i \end{array}$$

Problem **B**.

$$\begin{array}{rl} \text{min} & 2\sum_{i=1}^n \sum_{j=1}^m c_{ij} x_{ij} + x_{00} \\ \text{s.t.} & \sum_{i=1}^n x_i = 1, \\ & \lambda_{\text{min}} \begin{pmatrix} x_1 & 0 & 0 & \sum_{j=1}^m b_{jk} x_{1j} \\ 0 & \ddots & 0 & \vdots \\ 0 & 0 & x_n & \sum_{j=1}^m b_{jk} x_{nj} \\ \sum_{j=1}^m b_{jk} x_{1j} & \cdots & \sum_{j=1}^m b_{jk} x_{nj} & x_{00} \end{pmatrix} \geq 0 \\ & \quad \text{for } 1 \leq k \leq N. \end{array}$$

Convexity is of paramount importance in optimization, albeit difficult to detect for the untrained eye

Introduction and definitions
Infimum and Minimum

Let $f : Q \subseteq \mathbb{R}^n \to \mathbb{R}$.

The *infimum* of f on Q is the largest t
such that $f(x) \geq t$ for every $x \in Q$. $\inf_{x>0} 1/x = 0$.

We call the infimum a *minimum* if the infimum is attained:
$\exists x^* \in Q$ such that $f(x^*) = \inf_{x \in Q} f(x)$.

A *minimizer* is such a point x^*.
$\min_{x \in \mathbb{R}} \cos(x) = -1$,
the set of minimizers being $\{(2k+1)\pi : k \in \mathbb{Z}\}$.

The *supremum*, *maximum*, and *maximizers*
are defined similarly.

Introduction and definitions
General optimization

Let $f : Q \subseteq \mathbb{R}^n \to \mathbb{R}$

An *optimization problem* consists in finding the infimum of f over Q, and possibly some/all minimizers of f over Q.

The problem is *unconstrained* if $Q = \mathbb{R}^n$,
and *unfeasible* if Q is empty.

Usually, the set Q is given as:

$$Q = \{x \in \mathbb{R}^n : g_i(x) \leq 0, 1 \leq i \leq m, h_j(x) = 0, 1 \leq j \leq p\},$$

where $g_i, h_j : \mathbb{R}^n \to \mathbb{R}$ are given functions.
Particular cases: integer programming ($\sin(\pi x_i) = 0$), nonlinear equations ($m = 0,\ f(x) = \sum_j h_j(x)^2$).

Introduction and definitions

General optimization

Let $f : Q \subseteq \mathbb{R}^n \to \mathbb{R}$.

An *optimization problem* consists in finding the infimum of f over Q, and possibly some/all minimizers of f over Q.

- Sometimes, we have only access to a *local minimizer* x^* of f: there exists an open set $U \subseteq Q$ (see next slide), so that $f(x^*) = \min_{x \in U} f(x)$.
- Usually, we cannot compute the exact minimizer, but rather an approximation of it: for a given $\epsilon > 0$, find $\widehat{x}$ such that $f(\widehat{x}) - \min_{x \in Q} f(x) < \epsilon$.
  Or (although not well-posed in general) $||\widehat{x} - x^*|| < \epsilon$.

Introduction and definitions

Basic topology

Let E be a normed vector space, that is, a space with a norm $||\cdot||$. (e.g. $\mathbb{R}^n$, with the Euclidean norm).

Open ball centered in x and of radius $R > 0$:

$$B(x, R) := \{y \in E : ||y - x|| < R\}.$$

Closed ball: $B[x, R] := \{y \in E : ||y - x|| \leq R\}$

$U \subseteq E$ is *open* if $\forall x \in U$ there is an open ball $B(x, R_x) \subseteq U$.

$U \subseteq E$ is *closed* if $E \backslash U$ is open.

$U \subseteq E$ is *bounded* if U is contained in some open ball.

$U \subseteq E$ is *compact* if every sequence $(x_i)_{i \geq 0}$ has at least one limit point, and all its limit points are in U.

Theorem:
if $E = \mathbb{R}^n$, *compact* is equivalent to *closed and bounded*.

Introduction and definitions

Basic topology

Let E be a normed vector space, that is, a space with a norm $||\cdot||$. (e.g. $\mathbb{R}^n$, with the Euclidean norm).

Let $U \subseteq E$.
The *closure* of U, or $\mathrm{cl}(U)$ is the smallest closed set containing U. Equivalently:

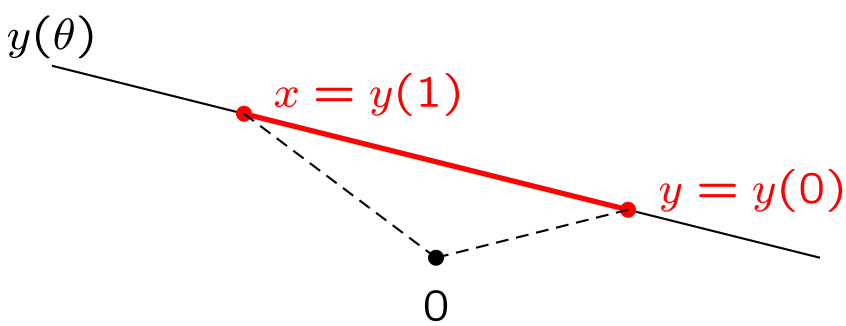
$$x \in \mathrm{cl}(U) \Leftrightarrow \forall \epsilon > 0, U \cap B(x, \epsilon) \neq \emptyset.$$

The *interior* of U, or $\mathrm{int}\,(U)$ is the largest open set contained in U. Equivalently:

$$x \in \mathrm{int}\,(U) \Leftrightarrow \exists \epsilon > 0, B(x, \epsilon) \subseteq U.$$

Affine sets

$U \subseteq E$ is *affine* if $x, y \in U$, $\theta \in \mathbb{R}$ implies $y(\theta) = \theta x + (1-\theta)y \in U$.



An affine set contains every line joining two of its points.
It is closed under affine combinations.

- If U affine and $x_0 \in U$ then $V = U - \{x_0\} = \{x - x_0 : x \in U\}$ is a linear subspace. The *dimension* of U is $\dim(V)$.

In finite dimension, every affine set can be represented as the solution set of a system of linear equations. Therefore, $U = \{x : Ax = b\}$, with $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$. Also, $V = \ker(A)$, and $b = Ax_0$ for all $x_0 \in U$.

Affine hull, dimension, and relative interior

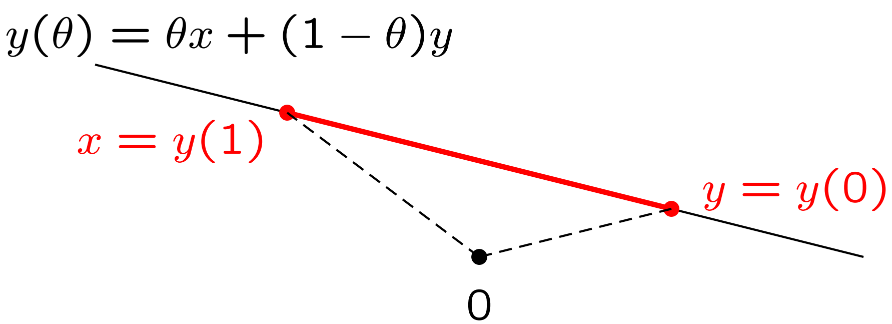
Let $U \subseteq E$, E of finite dimension.
The *affine hull* of U is the smallest affine set containing U entirely, that is:

$$\text{aff}(U) := \left\{ \sum_{i \in I} \theta_i x_i : \sum_{i \in I} \theta_i = 1, x_i \in U \ \forall i \in I, \text{ and } I \text{ finite} \right\}.$$

The *dimension* of U is $\dim(\text{aff}(U))$.
The *relative interior* of U is:

$$\text{relint}(U) = \{x \in U : \exists \epsilon > 0 \text{ with } B(x, \epsilon) \cap \text{aff}(U) \subseteq U\}.$$

The *boundary* of U is $\text{bd}(U) = \text{cl}(U) \setminus \text{relint}(U)$.



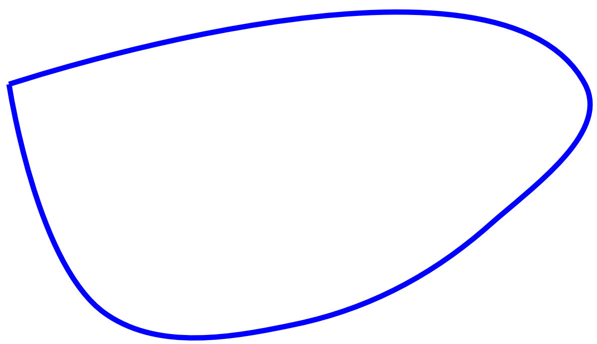
Let $U := \{y(\theta) : 0 \leq \theta \leq 1\}$.

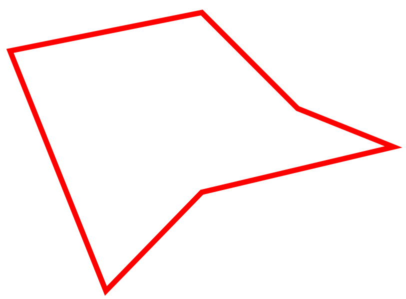
$\text{relint}(U) = \{y(\theta) : 0 < \theta < 1\}$.

$\text{bd}(U) = \{y(0), y(1)\}$.

Convex sets: definition

$U \subseteq E$ is *convex* if $x, y \in U$, $\theta \in [0,1]$
implies $y(\theta) = \theta x + (1-\theta)y \in U$.





A *convex combination* of finitely many points $(x_i)_{i \in I}$ is a point y s.t.:

$$y = \sum_{i \in I} \theta_i x_i, \quad \sum_{i \in I} \theta_i = 1, \quad \theta_i \geq 0 \text{ for all } i \in I.$$

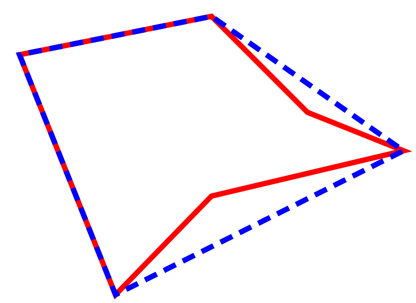
Note: Sums can be countably infinite,
but $\lim_{N\to\infty} \sum_{i=1}^{N} \theta_i x_i$ should exist.

- U is convex iff closed under convex combinations.

Convex sets: definition

The *convex hull* of U is the smallest convex set containing U entirely. One can prove:

$$\mathrm{conv}(U) = \left\{ \sum_{i \in I} \theta_i x_i : \sum_{i \in I} \theta_i = 1, \theta_i \geq 0, x_i \in U, \text{ and } I \text{ finite} \right\}.$$



Much more generally: suppose $p : \mathbb{R}^n \to \mathbb{R}$ with $p(x) \geq 0$ and $\int_U p(x)dx = 1$ where $U \subseteq \mathbb{R}^n$ is convex, then:

$$\int_U p(x)x dx \in U \text{ if the integral exists.}$$

Interpretation: X random vector in $\mathbb{R}^n$ with $\mathsf{P}[X \in U] = 1, U$ convex $\Rightarrow \mathsf{E}[X] \in U$.

Conic sets: definition

$U \subseteq E$ is a *cone* if $x \in U$, $\theta \geq 0$ implies $\theta x \in U$.

- U is a convex cone if U is a cone for which $x, y \in U$ implies $x + y \in U$.

The *conic hull* of U is the smallest convex cone containing U entirely. One can prove:

$$\text{cone}(U) := \left\{ \sum_{i \in I} \theta_i x_i : \theta_i \geq 0, x_i \in U \ \forall i \in I, \text{ and } I \text{ finite} \right\}.$$

Convex sets: elementary blocks

Examples of convex sets 1

- $\emptyset$ the empty set is affine (and thus a convex cone).
- A *singleton* $\{x_0\}$ is affine.
- A line is affine.
- A line segment is convex.
- A *ray* $\{x_0 + \theta v : \theta \geq 0\}$ where $v \neq 0, x_0$ fixed is convex, but not affine.
- A subspace is affine.
- A ball $B[x, R]$ is convex; in particular, a *Euclidean* ball $B_2[x, R] = \{y \in \mathbb{R}^n : (y - x)^T (y - x) \leq R^2\}$ is convex.

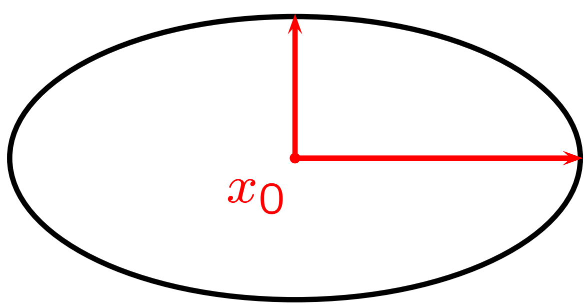
Convex sets: more elementary blocks

Examples of convex sets 2

- An *ellipsoid* $\mathcal{E} := \{Ax + x_0 : ||x||_2 \leq 1\}$, where $A \in \mathbb{R}^{n \times n}$ is symmetric and *positive definite* (i.e. $x^T Ax > 0$ when $x \neq 0$) is convex.

Its center is x_0, and the lengths of the semi-axes of $\mathcal{E}$ are given by $\lambda_i(A)$. Alternatively, by taking $E := A^2$, we have:

$$\mathcal{E} = \{x : (x - x_0)^T E^{-1} (x - x_0) \leq 1\}.$$



Convex sets: more elementary blocks

Examples of convex sets 3

- A *hyperplane* of $\mathbb{R}^n$, that is $H := \{x : a^T x = b\}$, where $0 \neq a \in \mathbb{R}^n$, $b \in \mathbb{R}$, is affine.

a is the *normal vector*: let $x_0 \in \mathbb{R}^n$ be such that $a^T x_0 = b$. Then:

$$H = \{x : a^T(x - x_0) = 0\} = x_0 + a^\perp, \text{ where } a^\perp = \{v : a^T v = 0\}.$$

H divides $\mathbb{R}^n$ into 2 closed *halfspaces*:

$$H^- = \{x : a^T x \leq b\}, \quad H^+ = \{x : a^T x \geq b\}.$$

- Halfspaces are convex.

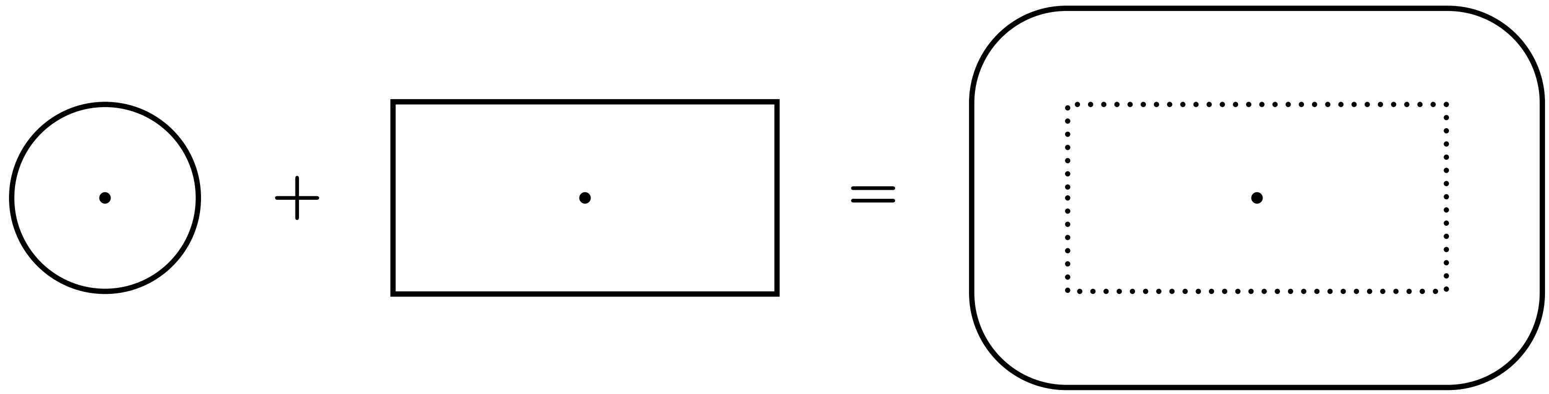
Convex sets: construction

- The intersection of convex sets is convex.

Note: Works (also) for uncountable intersections:
if U_α is convex for every $\alpha \in \mathbb{R}$, then $\cap_{\alpha \in \mathbb{R}} U_\alpha$ is convex.

- The *Minkovski sum* of two convex sets U_1, U_2 is convex:

$$U_1 + U_2 = \{u_1 + u_2 : u_1 \in U_1, u_2 \in U_2\}.$$



Conic sets: construction

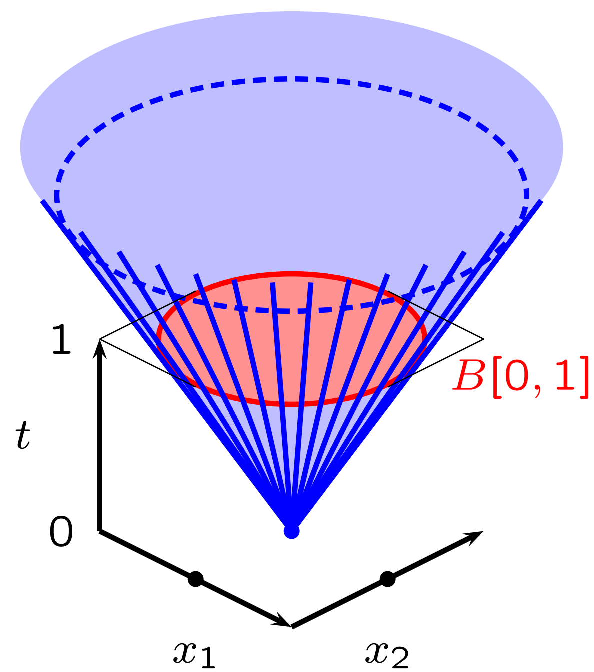
- Let $U \subseteq \mathbb{R}^n$ be convex.

The *light cone* generated by U is:

$$\{(t, x) \in \mathbb{R} \times \mathbb{R}^n : x \in tU \text{ for a } t > 0\} \cup \{0\}$$

is a convex cone.

The light cone generated by $B_2[0, 1]$ is the *second-order cone*, or *ice-cream cone*, or *Lorentz cone*.



Convex sets: more elementary blocks

Examples of convex sets 4

- A *polyhedron*

$$P = \{x : a_i^T x \leq b_i, 1 \leq i \leq m, c_j^T x = d_j, 1 \leq j \leq p\}$$

is convex (intersection of convex sets).

$$P = \{x \mid \underbrace{Ax \leq b}_{\text{vector inequality}}, Cx = d\}, A = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix}, C = \begin{pmatrix} c_1 \\ \vdots \\ c_p \end{pmatrix}.$$

- When $b = 0$ and $p = 0$, P is a convex cone.

Convex sets: more elementary blocks

Examples of convex sets 5

Let $v_0, \ldots, v_k$ be *affinely independent* vectors in $\mathbb{R}^n$, i.e. $v_1 - v_0, \ldots, v_k - v_0$ are linearly independent.

- The *k-simplex* of $v_0, \ldots, v_k$ is the convex hull of $\{v_0, \ldots, v_k\}$.

Note: A k-simplex is a polyhedron.

Proof: Let $C = \mathbf{conv}\{v_0, \ldots, v_k\}$ and $\mathbf{1} = (1, \ldots, 1)^T \in \mathbb{R}^k$.
Then $x \in C$ iff $x = v_0 + By$, $y \geq 0$, $\mathbf{1}^T y \leq 1$ with $B := [v_1 - v_0, \ldots, v_k - v_0]$.
As the rank of B is k, there exists an invertible matrix A such that $AB = \begin{pmatrix} I_k \\ 0 \end{pmatrix}$. Let $A =: \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}$, $A_1 \in \mathbb{R}^{k \times n}$. Multiplying the above representation of C by A and simplifying, we get:

$$C = \{x : A_1 x \geq A_1 v_0,\ \mathbf{1}^T A_1 x \leq 1 + \mathbf{1}^T A_1 v_0,\ A_2 x = A_2 v_0\}.$$

For next week

- **Convex sets from matrices**:
  towards semidefinite programming.

- **What can we actually do with these convex sets**:
  the Hahn-Banach separation Theorem.