Convex Optimization by Prof. Stephen Becker

Jaden Wang

January 27, 2021

Contents

Chapter 1

Theoretical Foundation

1.1 Introduction

An optimization problem looks like

$$\min_{x \in C} f(x)$$

where $f(x)$ is the **objective function** and $C \subseteq \mathbb{R}^n$ is the **constraint set**. C might look like

$$C = \{x : g_i(x) \leq 0 \ \forall \ i = 1, \ldots, m\}.$$

Remark. We can always turn a maximization problem into a minimization problem as the following:

$$\min_x f(x) = -\max_x -f(x).$$

Therefore, WLOG, we will stick with minimization.

Example. An assistant professor earns \$100 per day, and they enjoy both ice cream and cake. The optimization problem aims to maximize the utility (*e.g.* happiness) of ice cream $f_1(x_1)$ and of cake $f_2(x_2)$. The constraints we have is that $x_1 \geq 0, x_2 \geq 0$, and $x_1 + x_2 \leq 100$.

To maximize both utility, it might be natural to define

$$F(\operatorname{vec} x) = \begin{pmatrix} f_1(x_1) \\ f_2(x_2) \end{pmatrix}, \operatorname{vec} x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

and maximize F. However, this isn't a well-defined problem, because *there is no total order on* $\mathbb{R}^n$! That is, we don't have a good way to compare whether a vector is bigger than another vector, except in the cases when the same direction of inequality can be achieved for all components of two vectors and a partial order can be established. For this kind of **multi-objective** optimization problem, we can look for Pareto-optimal points in these special cases. We can also try to convert the output into a scalar as the following:

$$\min_x f_1(x) + \lambda \cdot f_2(x_2)$$

for some $\lambda > 0$ that reflects our preference for cake vs ice cream. But this can be subjective.

Thus, For the remainder of this class, we are only going to assume $f : \mathbb{R}^n \to \mathbb{R}$.

Moreover, for $f : \mathbb{R} \to \mathbb{R}$, it's very easy to solve by using root finding algorithms or grid search. So since interesting problems occur with vector inputs, we will simply use x to represent vectors.

Notation. min asks for the minimum value, whereas arg min asks for the minimizer that yields the minimum value.

1.1.1 Lipschitz continuity

Example. Let's consider a variant of the Dirichlet function, $f : \mathbb{R} \to \mathbb{R}$

$$f(x) = \begin{cases} x & \text{if } x \in \mathbb{Q} \\ 1 & \text{if } x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

Then the solution to the problem

$$\min_{x \in [0,1]} f(x) = 0$$

is $x = 0$ by observation. However, the function is not smooth and a small perturbation can yield wildly different values. Thus, it is not tractable to solve this numerically.

This requires us to add a smoothness assumption:

Definition: Lipschitz continuity

$f : \mathbb{R}^n \to \mathbb{R}$ is ***L*-Lipschitz continuous** with respect to a norm $\|\cdot\|$ if for all $x, y \in \mathbb{R}^n$,

$$|f(x) - f(y)| \leq L \cdot \|x - y\|.$$

Note. Lipschitz continuity implies continuity and uniform continuity. It is a stronger statement because it tells us *how* the function is (uniformly) continuous. However, it doesn't require differentiability.

Definition: l_p norms

For $1 \leq p < \infty$,

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}.$$

For $p = \infty$,

$$\|x\|_\infty = \max_{1 \le i \le n} |x_i|.$$

Remark. $\|x\|_1$ and $\|x\|_2^2$ have separable terms as they are sums of their components. $\|x\|_2^2$ is also differentiable which makes it the nicest norm to optimize.

Example. Let $f : \mathbb{R}^n \to \mathbb{R}$ be L-Lipschitz continuous w.r.t. $\|\cdot\|_\infty$. Let $C = [0,1]^n$, *i.e.* in $\mathbb{R}^2$, C is a square. To solve the problem

$$\min_{x \in C} f(x),$$

since we have few assumption, there is no better method (in the worst case sense) than the **uniform grid method**. The idea is that we pick $p+1$ points in each dimension, *i.e.* $\{0, \frac{1}{p}, \frac{2}{p}, \dots, 1\}$, so we would have $(p+1)^n$ points in total.

Let x^* be a global optimal point, then there exists a grid point $\widetilde{x}$ s.t.

$$\|x^* - \widetilde{x}\|_\infty \le \frac{1}{2} \cdot \frac{1}{p}.$$

Thus by Lipschitz continuity,

$$\begin{aligned} |f(x^*) - f(\widetilde{x})| &\le L \cdot \|x^* - \widetilde{x}\|_\infty \\ &\le \frac{1}{2}\frac{L}{p} \end{aligned}$$

So we can find $\widetilde{x}$ by taking the discrete minimum of all $(p+1)^n$ grid points.

In (non-discrete) optimization, we usually can't exactly find the minimizer, but rather find something very close.

Definition: epsilon-optimal solution

x is a **ε-optimal solution** to $\min_{x \in C} f(x)$ if $x \in C$ and

$$f(x) - f^* \le \varepsilon$$

where $f^* = \min_{x \in C} f(x)$.

Our uniform grid method gives us an ε-optimal solution with $\varepsilon = \frac{L}{2p}$, and requires $(p+1)^n$ function evaluations. Writing p in terms of ε, we have $p = \frac{L}{2\varepsilon}$

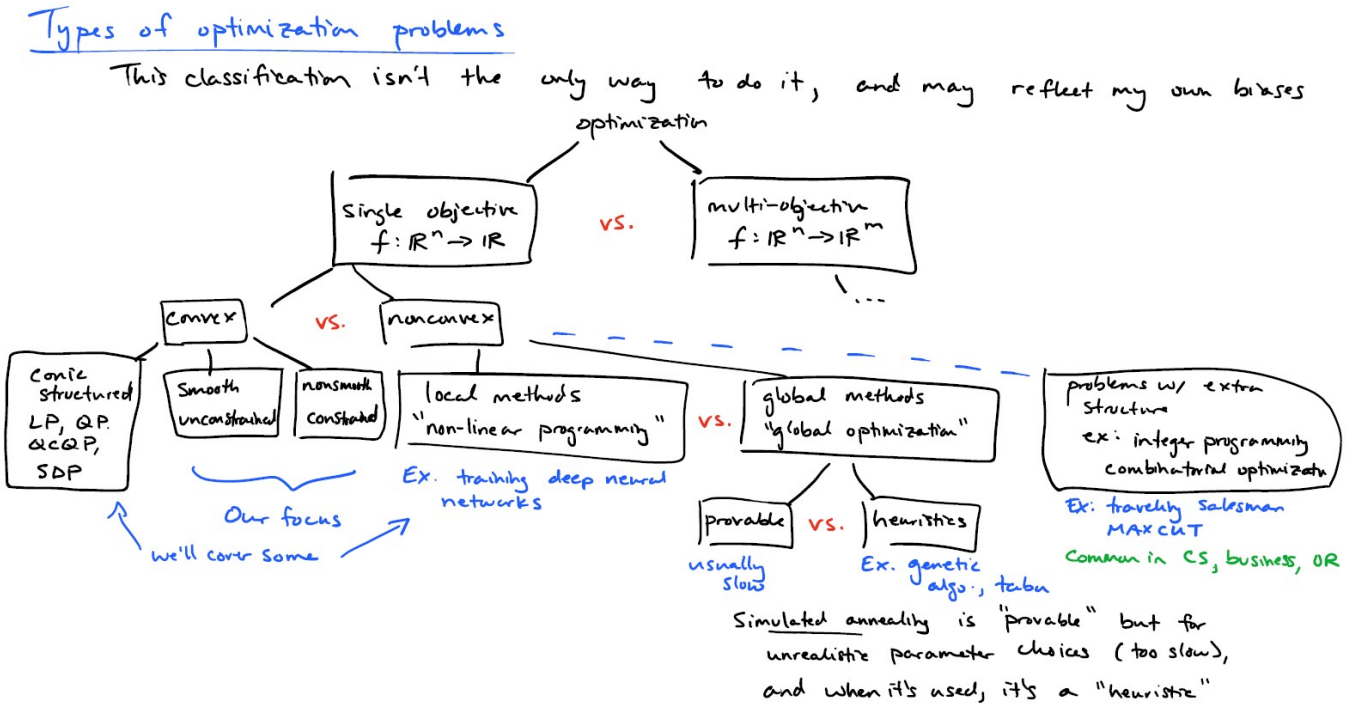
so equivalently it requires $\left(\frac{2L}{\varepsilon}+1\right)^n$ function evaluations, which approximately is ε^{-n}.

For $\varepsilon = 10^{-6}$, $n = 100$, it requires 10^{600} function evaluations. This is really bad!

Take-aways from this example:

- curse-of-dimensionality: there can be trillions of variables in a Google Neural Network. It would be intractable using the grid method.
- we need more assumptions to allow us to use more powerful methods.

1.1.2 Categorization



1.1.3 Minimizers

We are given a generic problem $\min_{x\in C} f(x), C \subseteq \mathbb{R}^n$. Then a **feasible point** x means $x \in C$. A **solution** or **minimizer** or **global minimizer** x^* means

1) $x^* \in C$

2) $\forall\ y \in C, f(x^*) \leq f(y)$

It might not be unique, *i.e.* $x^* \in \arg\min_{x\in C} f(x)$.

Example.

$$\min_{x\in\mathbb{R}} f(x) \text{ where } f(x) = 0\ \forall\ x.$$

Sometimes the solution may not exist (even for convex problems).

Example.

$$\min_{x\in(0,1)} x^2.$$

x^* is a **local minimizer** if x^* is feasible and there exists an $\varepsilon > 0$ s.t. $f(x^*) \leq f(y)\quad \forall\ y \in C \cap B_\varepsilon(x^*) \coloneqq \{y : \|y - x^*\|\} < \varepsilon$. A **strict local minimizer** simply doesn't achieve equality. x^* is an **isolated local minimum** if it is a local minimum and no other local minimum are nearby. Notice that isolated implies strict but the converse is false.

Example (strict but not isolated)**.**

$$f(x) = \begin{cases} x^4 \cos\left(\frac{1}{x}\right) + 2x^4 & x \neq 0 \\ 0 & x = 0 \end{cases}$$

$x^* = 0$ is strict but not isolated due to the rapid oscillation near $x = 0$.

Notation. $f \in \mathcal{C}^3$ means f, f', f'', f''' all exist and are continuous. $f \in \mathcal{C}^3(\mathbb{R}^n)$ means $f, \nabla f, \nabla^2 f, \nabla^3 f$ all exists and are continuous.

Connections with Calculus 1

Recall that in Cal 1, we first find the stationary/critical points in the domain. Then we add the boundary points and minimize over the small (finite) set of candidates.

In high-dimension optimization, we cannot check critical points and the boundary separately because the set of points in the boundary becomes infinite. Moreover, there can be infinite critical points too.

Necessary condition: if x^* is a local or global minimizer and $C = \mathbb{R}^n$, then x^* is a **critical point**. But the converse is false.

Notation. The boundary of C is denoted as $\partial C \coloneqq \overline{C} \setminus \operatorname{int} C$.

If x^* is a critical point but is not a local or global minimizer, then it's a **saddle point**.

Theorem: Weierstrass

If f is continuous and C is compact, then f achieves its infimum over C. That is,

$$\inf_{x \in C} f(x) = \min_{x \in C} f(x).$$

Note. This is pretty much the same as the Extreme Value Theorem.

Proof

First let's prove a claim.

Claim. Every compact set K is closed and bounded.

Closed: suppose not, the compact set K doesn't contain all its limit points. That is, there exists a limit point $x \notin K$ s.t. a sequence $(x_n) \subseteq K$ converges to x. But that also means that all subsequences of (x_n) converges to $x \notin K$ as well, contradicting with the definition of compactness that for every sequence in K there exists a subsequence that converges inside K.

Bounded: suppose not, for all $M > 0$, there exists a $x \in K$ s.t. $\|x\| > M$. This allows us to find a sequence $(x_n) \subseteq K$ s.t. $x_n > n$. This way every subsequence is also unbounded and cannot converge, contradicting with the definition of sequential compactness.

Now let's begin proof proper. Since C is compact and f is continuous, the image of C under f, $f(C)$, is also compact (this follows from sequen-

tial definition of continuity). By the claim $f(C)$ is bounded and closed, meaning that it has an infimum (completeness axiom) and contains the infimum (closed). Thus, f achieves its infimum over C. □

Remark. It would be nice if our constraints C are compact. But at the very least, we want our constraint sets to be closed. For example, $\|Ax - b\| \leq \varepsilon$ instead of $\|Ax - b\| < \varepsilon$.

Several things to note about the feasible set C:

If $C = \emptyset$, the problem is infeasible. This is not always easy to spot.

In this class, C will usually be convex and not integral, *i.e.* $\mathbb{Z}^n$.

Integral constraint is problematic because the optimal integer solution might not be at all close to the optimal real solution, so we cannot obtain it by solving for the real solution first and then round it.

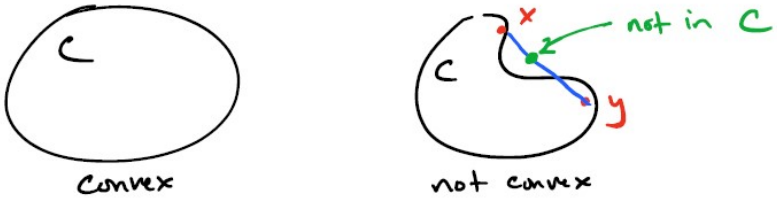
1.1.4 Convexity

Note. From now on we always assume the constraint set C is a subset of a vector space.

Definition: convex set

A set C is **convex** if for all $x, y \in C$ and for all $t \in [0, 1]$, then

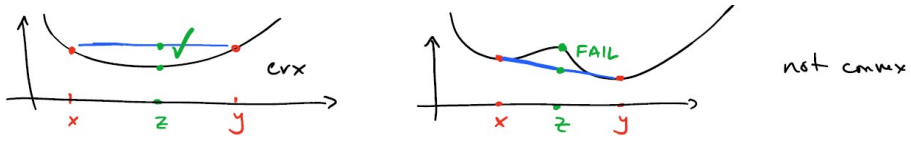
$$tx + (1 - t)y \in C.$$



Definition: convex function

$f : \mathbb{R}^n \to \mathbb{R}$ is a **convex function** if for all $x, y \in \mathbb{R}^n$ and $t \in [0, 1]$, then

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y).$$



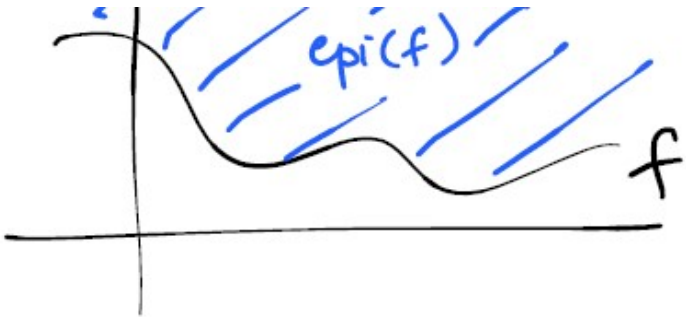
Remark. Linear or affine functions are both convex and concave.

Recall that a **graph** of a function is just the set of points we use to plot a function (now generalized to functions with domain of any dimension).

Definition: epigraph

$$\text{epi}(f) = \{(x, s) : x \in \mathbb{R}^n, s \in \mathbb{R}, s \geq f(x)\}.$$

Intuition. The epigraph of f is sort of the "upper" partition of the vector space that the graph of f resides and partitions. We use epigraph to bridge the concepts of convex sets and convex functions.



Proposition

A function f is convex if and only if the set $\text{epi}(f)$ is convex.

Theorem

If f is convex and C is convex, then any local minimizer of $\min_{x \in C} f(x)$ is in fact global. The set of global solutions is also convex and in particular connected.

This is very neat for optimization!

Proof

Given a local minimizer $x^* \in C$, let's show that it is a global minimizer. Suppose not, that is, there exists a point $x \in C$ s.t. $f(x) < f(x^*)$. Since x^* is a local minimizer, there exists an $\varepsilon > 0$ s.t. $f(x^*) \leq f(y)$ for all $y \in C$ s.t. $\|y - x^*\| < \varepsilon$. Clearly $\|x^* - x\| \geq \varepsilon$ or x^* would not be a local minimizer. Choose $t < \frac{\varepsilon}{\|x^* - x\|} \in [0, 1]$. Since C is convex, we know that the point $x_0 = tx + (1-t)x^* \in C$. Notice that

$$\begin{aligned} \|x^* - x_0\| &= \|x^* - (tx + (1-t)x^*)\| \\ &= \|t(x^* - x)\| \\ &= t\|x^* - x\| \\ &< t \cdot \frac{\varepsilon}{t} \\ &= \varepsilon \end{aligned}$$

That is, x_0 in the ε-neighborhood of x^* and it follows that $f(x^*) \leq f(x_0)$. Since f is convex,

$$\begin{aligned} f(x_0) &= f(tx + (1-t)x^*) \\ &\leq tf(x) + (1-t)f(x^*) \\ &< tf(x^*) + (1-t)f(x^*) \\ &= f(x^*) \end{aligned}$$

This contradicts with the fact that $f(x^*) \leq f(x_0)$. Hence we prove that any local minimizer x^* must also be a global minimizer.

To show that the set of global minimizers is connected, it suffices to prove that it is path-connected. The path we check is of course the line segment in the definition of convex set:

$$g(t) = ta + (1-t)b, \qquad a, b \in \operatorname{argmin}_{x \in C} f(x).$$

It's easy to see that $f(g(t)) \leq tf(a) + (1-t)f(b) = \min_{x \in C} f(x)$ for all $t \in [0,1]$. It follows that $f(g(t))$ must equal to the global minimum for all $t \in [0,1]$. This makes $g(t) \in \operatorname{argmin}_{x \in C} f(x)\ \forall\ t \in [0,1]$. Thus, the continuous function $g(t)$ is the path we seek. Since a, b are arbitrary global minimizers, we kill two birds in one stone and show that the set of global minimizers is 1. convex and 2. path-connected and therefore connected. □

1.2 Convex Sets

1.2.1 Convex, affine, and cone

Definition

Let $x, y \in \mathbb{R}^n$ (or any vector space), then

1) $tx + (1-t)y, t \in [0,1]$ is a **convex combination** (of x, y).
2) $tx + (1-t)y, t \in \mathbb{R}$ is a **linear combination**.
3) $tx, t \geq 0$ is a **conic combination**.

Definition

A set $C \subseteq \mathbb{R}^n$ is

1) **convex** if for all $x, y \in C$, it contains all convex combinations of x, y.
2) **affine** if for all $x, y \in C$, it contains all linear combinations of x, y.
3) a **cone** if for all $x \in C$, it contains all conic combinations of x.
4) a **convex cone** if it's convex and a cone. That is, for all $x, y \in C, t, x \geq 0$, $tx + sy \in C$.

Note. Affine implies convex based on definition.

Remark. An affine set/subspace is like a subspace except it is possible shifted (may not include 0). Think inhomogenous equation from differential equations. It's also analogous to cosets.

Recall from analysis, the **closure** of A, $\overline{A}$, is the union of A and all its limit points. We can also characterize $\overline{A}$ as the smallest closed set containing A or equivalently the intersection of all closed sets containing A. We can do something similar here.

Definition: affine hull

The **affine hull** of C, $\text{aff}(C)$, is the smallest affine set containing C.

The **affine dimension** of C is $\dim(\text{aff}(C))$. For example, although the unit circle in $\mathbb{R}^2$ has dimension 1, its affine hull is all of $\mathbb{R}^2$ so its affine dimension is 2.

Definition: convex hull

The **convex hull** of C, $\text{conv}(C)$, is the smallest convex set containing C. It is equivalent to the set

$$\left\{\sum_{i=1}^{k} t_i x_i : x_i \in C, t_i \leq 0, \sum_{i=1}^{k} t_i = 1\right\}.$$

Intuition. Given an arbitrary set C, we wrap a rubber band around it and the region enclosed by the rubber band is $\text{conv}(C)$.

Definition: relative interior

The **relative interior** of a set C is the set

$$\text{ri}(C) = \{x \in C : \exists\, \varepsilon > 0, B_\varepsilon(x) \cap \text{aff}(C) \subseteq C\}$$

Note. This is really useful for studying symmetric matrices.

Example. Let $C = [0,1] \subseteq \mathbb{R}$, then $\text{int}(C) = \text{ri}(C) = (0,1)$.

However, if $C = [0,1] \times \{0\}$ which has the same shape but is a embedding in $\mathbb{R}^2$, then $\text{int}(C) = \emptyset$ because for every $x \in C$, $B_\varepsilon(x)$ goes outside C along the

second dimension. But $\text{ri}(C) = (0, 1)$ because $\text{aff}(C) = \mathbb{R} \times \{0\} \cong \mathbb{R}$.

1.2.2 Important examples

Definition: hyperplane

For $a \in \mathbb{R}^n, b \in \mathbb{R}$, the **hyperplane** would be the affine, $n-1$ dimensional set

$$\{x \in \mathbb{R}^n : a^T x = b\}.$$

Alternatively,

$$\{x \in \mathbb{R}^n : a^T (x - x_0) = 0\}.$$

Note. Hyperplanes are convex and affine with $n - 1$ dimension. In 2D, it's a line. In 3D it's an actual plane. Also recall from cal 3 that a is the normal vector of the hyperplane.

Definition: half-space

A hyperplane partition $\mathbb{R}^n$ into two **half-spaces**. They have the form

$$\{x \in \mathbb{R}^n : a^T x \leq b\}.$$

Note. Half spaces are convex but not affine.

Definition: Euclidean ball

Open ball: $B_\varepsilon(x) = \{y \in \mathbb{R}^n : \|y - x\| < \varepsilon\}$.

Closed ball: $\overline{B_\varepsilon}(x) = \{y \in \mathbb{R}^n, \|y - x\| \leq \varepsilon\}$.

Note. Balls are convex but not affine.

Definition: ellipsoid

An **ellipsoid** has the form

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

for some matrix $P \succ 0$.

Notation. $A \succ 0$ in this course means A is symmetric and positive definite.

Note. Ellipsoid, like ball, is convex but not affine. If we choose $P = \varepsilon^2 I$, then we get an ε-ball.

Intuition. This is a generalization of Cal 3 ellipsoid using quadratic form. Recall that the P^{-1} in the middle of $x^T x$ is giving us a *weighted* sum. Since $y^T y = \|y\|^2 \leq 1$ is a unit ball, a weighted norm would help us transform an ellipsoid into the unit ball. We use the inverse of P in the definition because the image of this quadratic form is sort of a unit ball, but we are more interested in knowing how to go from the unit ball to the ellipsoid, and P encodes this transformation. Also since $P \succ 0$, using Spectral Theorem we can find the principle axises and length of the ellipsoid.

Example (cones)**.**

- positive orthant in $\mathbb{R}^n$, *e.g.* first quadrant in $\mathbb{R}^2$.

  $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n : x_i \geq 0\}$ is a cone.

  However, $\mathbb{R}^n_{++} = \{x \in \mathbb{R}^n : x_i > 0\}$ is not a cone as a cone must include the additive identity 0 to be closed under non-negative scalar multiplication.

- Lorentz cone/2nd order cone/"ice cream cone"

$$C = \{(x, t) \in \mathbb{R}^{n+1}, x \in \mathbb{R}^n, t \in \mathbb{R} : \|x\|_2 \leq t\}.$$

- the set of positive semidefinite matrices (PSD): this is the most important nonploy hedral cone. We assume PSDs are Hermitian and are denoted by $A \succeq 0$.

  Notation. $\mathbb{S}^n$ denotes the set of symmetric $n \times n$ matrices. Similar to the reals, we use $\mathbb{S}^n_+$ to denote the set of symmetric positive semidefinite

matrices.

Definition: polyhedron

A **polyhedron** $\mathcal{P} \subseteq \mathbb{R}^n$ is a set of the intersection of a *finite* number of half-spaces and hyperplanes. That is,

$$\mathcal{P} = \{x \in \mathbb{R}^n : a_j^T x \leq b_j, j = 1, \ldots, m; c_j^T x = d_j, j = 1, \ldots, p\}.$$

Note. Intersection of infinite number of half-spaces and hyperplanes are not necessarily a polyhedron (in the intuitive sense) because it can "smooth out" the edges and turn it into for example a ball, such as

$$\overline{B_1}(0) = \{x \in \mathbb{R}^n : a_j^T x \leq 1, a_j \in \text{ unit circle}\}.$$

Note. Polyhedra are always convex since it's finite intersection of convex sets.

Note. The terms "polygon", "polyhedron", and "polytope" will be used interchangeably in this course.

Recall that if a set of points are *linearly independent*, then their linear combinations can equal zero only if all coefficients are zero. Moreover, in an n-dimensional vector space, there can at most be n linearly independent points.

Definition: affinely independent

A set of points $\{x_i\}_{i=0}^n$ is **affinely independent** if

$$\sum_{i=1}^{n} t_i(x_i - x_0) = 0 \Rightarrow t_i = 0.$$

Note. It doesn't matter which x_i we choose to be x_0. They are all equivalent. This is because an affine space can be think of as a translation of a vector space. That is, every element in the affine space is an element from a vector space offset by the same translation vector. When we subtract any two elements from the affine space, the translation vector cancels out and leaves us an element from the vector space so we go back to linear independence in the vector space. Again we can think of the solutions to inhomogeneous differential equations for a concrete example.

Note. In a n-dimensional vector space, at most $n+1$ points can be affinely independent. The "+1" comes from bringing the 0 in the vector space up to x_0, elevating the n-dimensional vector space to an n-dimensional embedding in a $n+1$-dimensional vector space. For example, $\mathbb{R}^n$ requires at least n points to fully describe it via the span. We can visualize $\mathbb{R}^n$ as an origin-containing plane embedded in $\mathbb{R}^{n+1}$. If we translate $\mathbb{R}^n$ by a vector x_0, we get an n-dimensional affine space that now requires at least $n+1$ points to describe.

Definition: simplex

For any set of $k+1$ affinely independent points $\{x_i\}_{i=0}^n$ in $\mathbb{R}^n$, they determine a **simplex**

$$C = \text{conv}\left(\{x_i\}_{i=0}^k\right) = \left\{x = \sum_{i=0}^k t_i x_i : t_i \geq 0, \sum_{i=0}^k t_i = 1\right\}.$$

Note. The affine dimension of $k+1$ point simplex is k, so we call it a k-dim simplex.

Intuition. Due to affine independence, we can think of the convex hull of the points as having "all its fat trimmed". Using the rubber band visualization, we can see why the following examples are true.

Example.

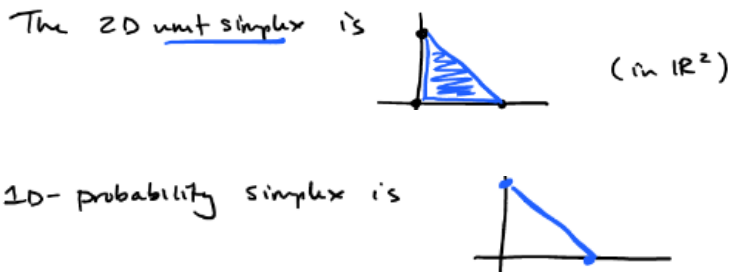
- 1-dim simplex $= \text{conv}(\{x_0, x_1\})$ is a line segment in $\mathbb{R}^n, n \geq 1$.
- 2-dim simplex $= \text{conv}(\{x_0, x_1, x_2\})$ is a filled triangle in $\mathbb{R}^n, n \geq 2$.
- 3-dim simplex is a tetrahedron in $\mathbb{R}^n, n \geq 3$. Clearly if $n = 2$, 4 points cannot be affinely independent and thus cannot generate a simplex.
- In $\mathbb{R}^n$, the **unit simplex** is the simplex generated by $\{0, e_1, e_2, \ldots, e_n\}$, where e_i is the standard basis. It has affine dimension of n. This can be expressed as

$$\left\{x \in \mathbb{R}^n : x \geq 0, \sum_{i=1}^n x_i \leq 1\right\} \text{ or } \{x \in \mathbb{R}^n : x \geq 0, \mathbb{1}^T x \leq 1\}.$$

- In $\mathbb{R}^n$, the $(n-1)$-dim **probability simplex** is generated by $\{e_1, e_2, \ldots, e_n\}$

(basically unit simplex without 0). It can be expressed as

$$\{x \in \mathbb{R}^n : x \geq 0, \mathbb{1}^T x = 1\}.$$



Remark. Simplex is generated by finite points. *Atomic norm* generalizes this to infinite points and uses gauge functions for signal processing.

1.3 Operations that preserve convexity

1) Cartesian products: If $C_1 \subseteq \mathbb{R}^{n_1}$ and $C_2 \subseteq \mathbb{R}^{n_2}$ both convex, then $C_1 \times C_2 \subseteq \mathbb{R}^{n_1+n_2}$ is convex.

2) Arbitrary intersections (even uncountable): If C_1, C_2 are convex, then $C_1 \cap C_2$ is convex. *This is not true for unions.*

3) Image and preimage of an affine function $f(x) = Ax + b$:

   - $f(C)$ is convex if $C \subseteq \mathbb{R}^n$ is.
   - $f^{-1}(C)$ is convex if $C \subseteq \mathbb{R}^m$ is.

   This implies that scaling, translation, rotation, and projection all preserve convexity.

   So does **Minkowski sum**:

$$C_1 + C_2 \coloneqq \{x + y : x \in C_1, y \in C_2\}.$$

   We should think of it like a convolution, where each element in C_1 is convolved with the entire C_2.

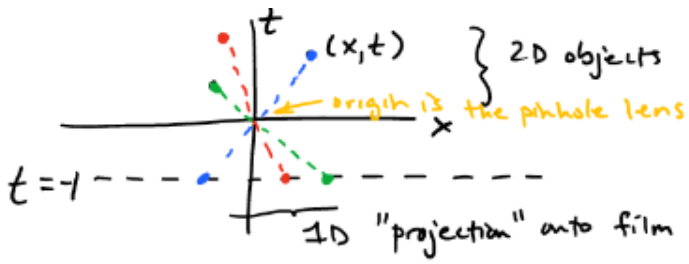
1.3.1 Linear-fractional and perspective functions

Definition: perspective function

A **perspective function** is a function $P : \mathbb{R}^{n+1} = \mathbb{R}^n \times R_{++} \to \mathbb{R}^n$ s.t. for $z \in \mathbb{R}^n$ and $t \in \mathbb{R}_{++}$,

$$P(z,t) = \frac{z}{t}.$$

Intuition. We can think of it as normalizing by t, $(\frac{z}{t}, 1)$, and then projecting to $\mathbb{R}^n$ (or equivalently dropping the last component). Geometrically we can think of it as a "pin-hole" camera that projects 3D points in the t-positive half of $\mathbb{R}^3$ through the pin-hole at origin onto the 2D film at $t = -1$. This gives us $(-\frac{z}{t}, -1)$ which is the negative perspective. Then since all the points are on $t = -1$ we simply drop it.



If P is the perspective function, we can conclude that

- If $C \in \mathbb{R}^{n+1}$ is convex, then $P(C)$ is convex in $\mathbb{R}^n$.
- If $C = \mathbb{R}^n$ is convex, then $P^{-1}(C)$ is convex in $\mathbb{R}^{n+1}$.

Definition: linear-fractional function

A **linear-fractional function** is $f : \mathbb{R}^n \to \mathbb{R}^m$ that composes the perspective P with an affine function g, where

$$g(x) = \begin{pmatrix} A \\ c^T \end{pmatrix} x + \begin{pmatrix} b \\ d \end{pmatrix} : \mathbb{R}^n \to \mathbb{R}^{m+1}, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, c \in \mathbb{R}^n, d \in \mathbb{R}.$$

That is, $f = P \circ g$, and

$$f(x) = \frac{Ax + b}{c^T x + d}, \text{ with domain } \{x : c^T x + d > 0\}.$$

Note. Since the image and preimage of both affine and perspective functions preserve convexity, the image and preimage of a linear-fractional function again preserve convexity.

1.3.2 Generalized inequalities

Definition: proper cone

A cone $K \subseteq \mathbb{R}^n$ is called a **proper cone** if it satisfies the following:

1) convex
2) closed
3) solid or nonempty interior
4) pointed or contains no line or $x \in K, -x \in K \Rightarrow x = 0$.

Proposition

Any proper cone K induces a partial order:

$x \preceq_K y$ or simply $x \leq y$ if $y - x \in K$,

$x \prec_K y$ or simply $x < y$ if $y - x \in \text{int}(K)$.

Definition: dual cone

If K is a set, its **dual cone** is

$$K^* = \{y : \langle x, y \rangle \geq 0 \ \forall \ x \in K\}.$$

Note. The larger K is, the more restricted K^* becomes.

Properties of dual cones

1) K^* is a cone even if K isn't a cone.
2) K^* is convex even if K isn't convex.
3) $K_1 \subseteq K_2 \Rightarrow K_2^* \subseteq K_1^*$.
4) $K^{**} = K$ iff K is a proper cone.

Example. If K is a subspace, then $K^* = K^\perp$. This is because $-x \in K$ so equality is achieved in the definition.

Example (self-dual)**.** $\mathbb{R}^n_+ = \left(\mathbb{R}^n_+\right)^*$ because it is a proper cone.

Example (PSD matrices)**.** $K = S^n_+$, and $x \in K \Rightarrow X = GG^T$ (Cholesky).

Then

$$K^* = \{Y \in S^n : \langle Y, X \rangle \geq 0 \ \forall \ X \succeq 0\}.$$

Recall that

$$\begin{aligned}
\langle Y, X \rangle &= \text{tr}(Y^T X) \\
&= \text{tr}(YX) \text{ since } Y = Y^T \\
&= \text{tr}(YGG^T) \\
&= \text{tr}(G^T YG) \text{ by cyclic property of trace}
\end{aligned}$$

The last expression is ≥ 0 for all matrices G iff $Y \succeq 0$. Hence we show that S^n_+ is self-dual.