Convex Optimisation

Matthew Tam
(University of Melbourne)
matthew.tam@unimelb.edu.au

ACE Network
S1 2021

Lecture times

- Thursdays, 1-3pm
  (Melb. time)
- Every week except
  on April 8th (Easter break)
  i.e. UniMelb calender

Consultation

- after Lecture.
- Email me.

Lecture notes

- send out weekly/fortnightly.
- typo/corrections → email.

Course outline

Part I: Convex analysis ("Tools")

- convexity
- differentiablity of convex functions
- proximity operator
- separation theorems
- Subgradients of convex functn.

Part II: Algorithms

- Frank-Wolfe
- Proximal gradient descet
- Nesterov accelevation
- subgradient method
- Douglas-Rachford

⋮

---

* No convex duality.

Assignments

- 4 assigments (25% each)
- LaTex
- Due in: (@ 5pm)
    - Week 4 (March 25th)
    - Week 6 (April 15th)
    - Week 10 (May 13th)
    - Week 13 (June 3rd)

Census date

March 31st.

Chapter 1: Introduction

Given $f : \mathbb{R}^n \to \mathbb{R}$ and $C \subseteq \mathbb{R}^n$, a typical optimisation problem (or <u>mathematical program</u>) takes the form

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in C. \qquad (1.1)$$

Terminology:

- x = decision variable
- f = objective function
- C = constraint set.

Switch between min and max by replacing f with $-f$.

To study optimisation problems, its useful to classify them based on their math. properties.

For example:

- linear / nonlinear
- continuous / discrete
- smooth / nonsmooth (differentiability)
- <u>convex</u> / nonconvex

Problem (1.1) is convex if f is <u>convex</u> and C is <u>convex</u>.

Example 1.0.1 (LP)

Let $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$.

A linear program (LP) in standard form is given by

$$\min_{x \in \mathbb{R}^n} \langle c, x \rangle \quad \text{s.t.} \quad Ax \le b,\ x \geq 0$$

$$\langle c, x \rangle = c \cdot x = \sum_{i=1}^{n} c_i x_i$$

$f(x) = \langle c, x \rangle$

$C = \{ x \in \mathbb{R}^n : Ax \le b,\ x \geq 0 \}$

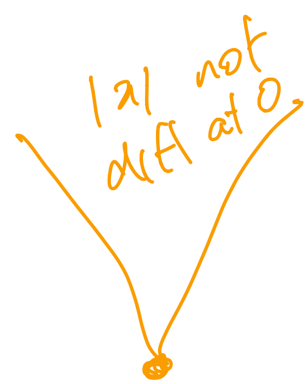
Example 1.0.2 (ℓ_1-reg. regression)

Solve $Ax = b$ where A is under-determined.

Often, we know that $x \in \mathbb{R}^n$ is sparse (i.e. many of its entries are zero).

This problem can cast as the optimisation problem

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 + \lambda \|x\|_1$$



where $\lambda > 0$. Recall:

$$\|y\|_2 = \sqrt{\sum_{i=1}^{n} y_i^2}$$

$$\|y\|_1 = \sum_{i=1}^{n} |y_i|.$$

Example 1.0.4 (portfolio opt)

Invest some quantity of money in n stocks. A portfolio is a vector $x \in \mathbb{R}^n$ st

x_i = proportion of money invested in stock i.

Let r_i denote the expected return of stock i, and Σ_{ij} denote the correlation between stocks i and j.

Then the expected return is

$$\sum_{i=1}^{n} r_i x_i$$

The risk is the variance of the portfolio given by $x^t \Sigma x$.

Finding the least risky portfolio which achieves expected return $\alpha \in \mathbb{R}$ is given by

$$\min_{x \in \mathbb{R}} \; x^T \Sigma x \quad \text{s.t.} \quad \sum_{i=1}^{n} x_i = 1, \; x \geq 0,$$

$$\sum_{i=1}^{n} r_i x_i \geq \alpha.$$

AMSI ACE Course: Convex Optimisation

Matthew Tam
University of Melbourne
matthew.tam@unimelb.edu.au

Review of Euclidean Spaces

The vector space $\mathbb{R}^n$
equipped with the dot-product.

Euclidean Spaces

\langle = <

\rangle = >

The **dot-product** on $\mathbb{R}^n$ defined by

$$x \cdot y = \langle x, y\rangle := \sum_{i=1}^{n} x_i y_i \quad \forall x, y \in \mathbb{R}^n.$$

The **(Euclidean) norm** (or 2-norm) induced by the dot-product which is given by

$$\|x\| := \sqrt{\langle x, x\rangle} = \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}} \quad \forall x \in \mathbb{R}^n.$$

$$\|x\|^2 = \langle x, x\rangle$$

Euclidean Spaces

Using the definitions, one can verifying the following properties:

1. (linearity) $\langle x + \lambda y, z\rangle = \langle x, z\rangle + \lambda\langle \overset{y,z}{\cancel{x, y}}\rangle$ for all $x, y, z \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$.
2. (symmetry) $\langle x, y\rangle = \langle y, x\rangle$ for all $x, y \in \mathbb{R}^n$.
3. (absolute homogeneity) $\|\lambda x\| = |\lambda|\|x\|$ for all $x \in \mathbb{R}^n$.

$$\begin{aligned}
\|\lambda x\| &= \left(\sum_{i=1}^{n} (\lambda x_i)^2\right)^{\frac{1}{2}} \\
&= \left(\lambda^2 \sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}} \\
&= (\lambda^2)^{\frac{1}{2}} \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}} \\
&= |\lambda| \, \|x\|.
\end{aligned}$$

Useful Identities

The following identities holds for all $x, y \in \mathbb{R}^n$.

(a) (Cauchy–Schwarz inequality)

$$|\langle x, y \rangle| \leq \|x\|\|y\|.$$

(b) (Triangle inequality)

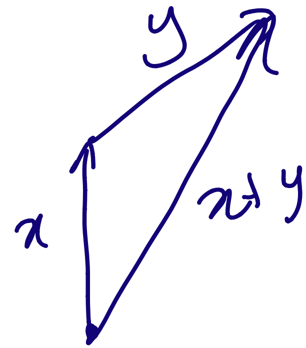
$$\|x + y\| \leq \|x\| + \|y\|.$$

(c) (Polarisation identity)

$$\begin{aligned}\langle x, y \rangle &= \frac{1}{2}\left(\|x+y\|^2 - \|x\|^2 - \|y\|^2\right) \\ &= \frac{1}{2}\left(\|x\|^2 + \|y\|^2 - \|x-y\|^2\right) \\ &= \frac{1}{2}\left(\|x+y\|^2 - \|x-y\|^2\right).\end{aligned}$$

(d) (Parallelogram law)

$$2\|x\|^2 + 2\|y\|^2 = \|x+y\|^2 + \|x-y\|^2.$$



Example: Proof of the polarisation identity

$$\boxed{\langle x, y\rangle = \frac{1}{2}\left(\|x+y\|^2 - \|x\|^2 - \|y\|^2\right)}$$

$$\begin{aligned}
\|x+y\|^2 &= \langle x+y, x+y\rangle \\
&= \langle x, x+y\rangle + \langle y, x+y\rangle \\
&= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle \\
&= \|x\|^2 + 2\langle x, y\rangle + \|y\|^2
\end{aligned}$$

$$\Updownarrow$$

$$\frac{1}{2}\left(\|x+y\|^2 - \|x\|^2 - \|y\|^2\right) = \cancel{2}\langle x, y\rangle$$

Cauchy Sequences and Completeness

Definition (Cauchy sequence)

A sequence $(x_k) \subseteq \mathbb{R}^n$ is a **Cauchy sequence** if

$$\|x_k - x_m\| \to 0 \text{ as } k, m \to \infty.$$

Theorem

The space $\mathbb{R}^n$ equipped with the Euclidean norm $\|\cdot\|$ is *complete*, that is, every Cauchy sequence is a convergent sequence.

Recall (x_n) converges to $x \in \mathbb{R}$ if

$$\|x_k - x\| \to 0 \text{ as } k \to \infty.$$

Topological properties of sets

Definition (open and closed sets)

Let $C \subseteq \mathbb{R}^n$.

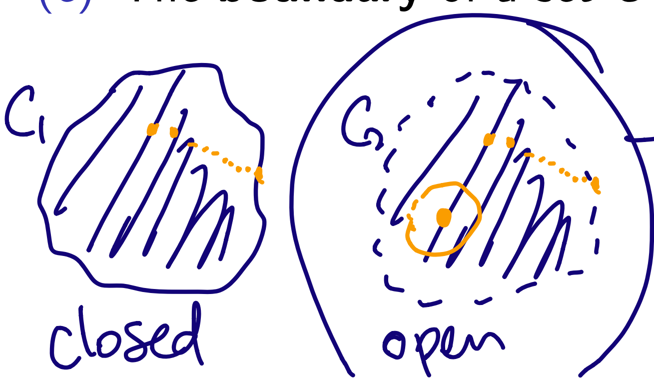
(a) C is **closed** if $(x_k) \subseteq C$ with $x_k \to x$ implies $x \in C$.

(b) C is **open**, if for every $x \in C$, there exists a $\delta > 0$ such that $\mathbb{B}_\delta(x) \subseteq C$.

(c) The **closure** of C, denoted $\operatorname{cl} C$, is the smallest closed set containing C.

(d) The **interior** of C, denoted $\operatorname{int} C$, is the largest open set contained in C.

(e) The **boundary** of a set C is given by $\operatorname{bdry} C := \operatorname{cl} C \setminus \operatorname{int} C$.



$\operatorname{cl} C_2 = C_1$

$\operatorname{int} C_1 = C_2$

Definition (compactness)

A set $C \subseteq \mathbb{R}^n$ is **compact** if every sequence (x_k) contained in C contains a convergent subsequence.

Theorem (Bolzano–Weierstrass)

Every bounded sequence in $\mathbb{R}^n$ contains a convergent subsequence. A nonempty set in $\mathbb{R}^n$ is compact if and only if it is closed and bounded.

Recall

- (x_n) is bdded if $\exists M > 0$ st. $\|x_k\| \leq M \quad \forall k \in \mathbb{N}$.
- C is bdded if $\exists M > 0$ st. $\|c\| \leq M \quad \forall c \in C$.

$$\Delta = \left\{ x \in \mathbb{R}^n : \sum_{i=1}^{n} x_i = 1, \ x \geq 0 \right\}$$

This is a compact set.