TCS Guide of Convex Optimization - Day 1 Overview

January 30, 2023

Theme: How fast can we find a point that minimizes f (TCS style)?

Example: Given a class of function $\mathcal{F}$ and given an oracle to access a function $f \in \mathcal{F}$ (e.g. via $f(x)$ or $\nabla f(x)$), study how many oracle call is suffices to minimize f up to certain precision.

1 Why convexity optimization?

1.1 Optimizing general functions may require brute force

Example 1. Consider $f(x) = \min(\|x - x^*\|_2, \epsilon)$. It requires $1/\Theta(\epsilon)^n$ calls of $f(x)$ to find x with $f(x) \leq \frac{\epsilon}{2}$.

1.2 Convexity enables binary search

Definition 2. (Convex)

- A set K is convex if for any $x, y \in K$, we have $tx + (1-t)y \in K$ for all $t \in [0, 1]$. [Give picture]
- A function f is convex if for any $x, y \in \mathrm{dom} f$, we have $f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$ for all $t \in [0, 1]$. [Give picture]
- A problem $\min_{x \in K} f(x)$ is convex if K and f are convex

Theorem 3 (Separation theorem)**.**

- *Given a convex set K and $y \notin K$, there is a halfspace $H = \{x : \theta^\top x \leq t\}$ such that $\theta^\top y = t$ and $K \subset H$. [Give picture proof]*
- *Given a (differentiable) convex function f, we have*

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x) \text{ for all } x, y$$

  (If f is not differentiable, ∇f needs to defined carefully.) [Give picture]

1.3 Convex Optimization are everywhere

- Linear program $\min_{Ax=b, x\geq 0} c^\top x$ where A is a matrix, b, c, x are vectors
  - In particular, maxflow, shortest path, maximum matching, minimum spanning tree, ... are convex problems
- Semidefinite program: $\min_{A_i \bullet X = b_i, X \succeq 0} C \bullet X$ where A_i, X, C are matrices, $M \bullet N = \sum_{ij} M_{ij} N_{ij}$ and $X \succeq 0$ means all eigenvalues of X is non-negative.
- Logistic regression: $\min_x \sum_{i=1}^n f(a_i^\top x)$ with $f(t) = \log(1 + e^t)$
- ...

Exercise 4. Prove that LP and logistic are convex.

2 What do we know?

2.1 Some runtimes

Here is a sequence of increasing general problems:

- $Ax = b$. (Complexity: $n^{2.38}$ time)
- Linear program. (Complexity: $n^{2.38}$ time)
- Convex Minimization: $\min_x f(x)$ with gradient oracle. (Complexity: n^3 time)
- (Generalization 1): Log-concave sampling: Sample x from $\exp(-f(x))$.
- (Generalization 2): Non-convex optimization: Find local minimum of nonconvex $f(x)$.

2.2 Some techniques

- Binary Search: convex optimization is in P
  - Each step, we eliminate part of the solution space.
- Local Method: convex optimization is in "LinearTime" if the problem is "simple" enough
  - Each step, we make progress on the function value or on the lower bound.
- Homotopy Method: convex optimization is almost as easy as general linear system
  - We find the solution by tracking the minimizer of a family of convex functions

3 Discussion

- Conversation of difficulty
- Convex optimization helps you the difficulty into two part where one part is iterative and one part is implementation.
- This allows you to reuse the idea for the iterative part.
- There are many algorithms that gives you different way to cut into two parts.

4 Tips for using convex optimization

4.1 How to check convexity

One reason convex functions/sets are abundant because we can composite convex functions/sets together via addition/intersection.

Exercise 5. Given convex functions f_i, matrices A_i, vectors b_i, and positive scalars $\lambda_i \geq 0$, the function

$$g(x) \stackrel{\text{def}}{=} \sum_i \lambda_i f_i(A_i x + b_i)$$

is convex.

The above theorem allows us to construct many convex functions from one dimensional convex functions. Here are some convex sets and functions.

Example. Convex sets: polyhedron $\{x : Ax \leq b\}$, polytope $\text{convhull}(\{v_1, \ldots, v_m\})$ with $v_1, \ldots, v_m \in \mathbb{R}^n$, ellipsoid $\{x : x^\top Ax \leq 1\}$ with $A \succeq 0$, positive semidefinite cone $\{X \in \mathbb{R}^{n\times n} : X \succeq 0\}$, norm ball $\{x : \|x\|_p \leq 1\}$ for all $p \geq 1$.

Example. Convex functions: x, $\max(x, 0)$, e^x, $|x|^a$ for $a \geq 1$, $-\log(x)$, $x \log x$, $\|x\|_p$ for $p \geq 1$, $(x, y) \to \frac{x^2}{y}$ (for $y > 0$), $A \to -\log\det A$ over PSD matrices A, $(x, Y) \to x^\top Y^{-1} x$ (for $Y \succ 0$), $\log \sum_i e^{x_i}$, $(\prod_i^n x_i)^{\frac{1}{n}}$.

4.2 How to write down a convex problem

I will write use a problem to illustrate the importance of how to write down the problem

Consider the compressive sensing problem

$$\min_x \frac{1}{2}\|Ax-b\|_2^2 + \|x\|_1$$

where $A \in \mathbb{R}^{r\times d}$ with $r \ll d$. Here r is # of measurement and d is the # of variables. Imagine $r = 100$ and d is 1 million.

Question: Is there any way to reduce # of variables to r?

Idea: If $\|Ax-b\|_2^2$ term does not exist, the problem is easy.

Using

$$\frac{1}{2}\|u\|_2^2 = \max_s s^\top u - \frac{1}{2}\|s\|_2^2,$$

we can rewrite the problem by

$$\min_x \frac{1}{2}\|Ax-b\|_2^2 + \|x\|_1 = \min_x \max_s s^\top(Ax-b) - \frac{1}{2}\|s\|_2^2 + \|x\|_1.$$

Why this is simpler? If s is given, the problem is very simple:

$$\begin{aligned}
&\min_x s^\top(Ax-b) - \frac{1}{2}\|s\|_2^2 + \|x\|_1 \\
=& -b^\top s - \frac{1}{2}\|s\|_2^2 + \min_x s^\top Ax + \|x\|_1 \\
=& -b^\top s - \frac{1}{2}\|s\|_2^2 + \min_x \sum_{i=1}^d (A^\top s)_i x_i + |x_i| \\
=& -b^\top s - \frac{1}{2}\|s\|_2^2 + \sum_{i=1}^d \min_x (A^\top s)_i x_i + |x_i| \\
=& \begin{cases} -\infty & \text{if } \|A^\top s\|_\infty > 1 \\ -b^\top s - \frac{1}{2}\|s\|_2^2 & \text{otherwises} \end{cases}.
\end{aligned}$$

So, if s is given, both the OPT value and the minimizer is given by some explicit formula.

Theorem 6 (Minimax theorem)**.** *(Roughly) Given a function $f(x, s)$ which is convex in x and concave in s. Then,*

$$\min_x \max_s f(x,s) = \max_s \min_x f(x,s).$$

Going back to the original problem

$$\begin{aligned}
\min_x \frac{1}{2}\|Ax-b\|_2^2 + \|x\|_1 &= \min_x \max_s s^\top(Ax-b) - \frac{1}{2}\|s\|_2^2 + \|x\|_1 \\
(\text{minimax}) &= \max_s \min_x s^\top(Ax-b) - \frac{1}{2}\|s\|_2^2 + \|x\|_1 \\
&= \max_{\|A^\top s\|_\infty \le 1} -b^\top s - \frac{1}{2}\|s\|_2^2
\end{aligned}$$

Note that LHS has d variables and RHS has r variables.

Exercise 7. Suggest some way to cover the exact solution on x given the exact solution on s. (Feel free to make extra assumptions on the problem.)