CS 2601 Linear and Convex Optimization

Review

Bo Jiang

John Hopcroft Center for Computer Science
Shanghai Jiao Tong University

Fall 2022

Convex sets (1/2)

- definition

$C \subset \mathbb{R}^n$

$$\boldsymbol{x}, \boldsymbol{y} \in C, \theta \in [0,1] \implies \theta \boldsymbol{x} + \bar{\theta} \boldsymbol{y} \in C$$

- convex combination

$$\sum_{i=1}^{k} \theta_i \boldsymbol{x}_i, \quad \text{where } \theta_i \geq 0, \quad \sum_{i=1}^{k} \theta_i = 1$$

- convex hull of S, smallest convex set containing S, set of all convex combinations of points in S,

任意内容的凸组合

$$\operatorname{conv} S = \left\{ \sum_{i=1}^{m} \theta_i \boldsymbol{x}_i : m \in \mathbb{N}; \boldsymbol{x}_i \in S, \theta_i \geq 0, i = 1, \ldots, m; \sum_{i=1}^{m} \theta_i = 1 \right\}$$

- examples: lines, rays, line segments, hyperplanes, half ~~plane~~ space, affine space, polyhedron, norm ball, ellipsoid, simplex, positive semidefinite cone 可以直接用（除非要证）

Convex sets (2/2)

保凸运算

- convexity-preserving operations
  - intersection of convex sets 求交
  - image/preimage of convex set under affine transformation 仿射
- projection onto closed convex set

投影运算 (不.已知结论)

$$\mathcal{P}_C(\boldsymbol{x}) = \operatorname*{argmin}_{z\in C} \|\boldsymbol{x}-\boldsymbol{z}\|_2 = \operatorname*{argmin}_{z\in C} \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{z}\|_2^2$$

$\min \frac{1}{2}\|x-z\|_2^2$ 要会求

s.t. $z\in C$

- 支撑 supporting hyperplane theorem  (是什么)
- 分离 separating hyperplane theorem 

考 methods for proving convexity:

掌握

- definition 定义
- convexity-preserving operations 由例子出发 保凸运算
- 上/下水平集 sublevel/superlevel set of convex/concave functions
- epigraph/hypograph of convex/concave functions 

Convex functions (1/3)

- definition: f is convex if it has convex domain $\mathrm{dom}\, f$, and

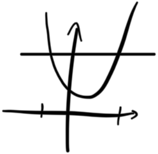
$$\boldsymbol{x}, \boldsymbol{y} \in \mathrm{dom}\, f, \theta \in (0,1) \implies f(\theta \boldsymbol{x} + \bar{\theta} \boldsymbol{y}) \leq \theta f(\boldsymbol{x}) + \bar{\theta} f(\boldsymbol{y})$$

  f is concave if $-f$ is convex. $\geq$
- affine functions are the only functions that are both convex and concave.
- strict convexity

$$\boldsymbol{x} \neq \boldsymbol{y} \in \mathrm{dom}\, f, \theta \in (0,1) \implies f(\theta \boldsymbol{x} + \bar{\theta} \boldsymbol{y}) < \theta f(\boldsymbol{x}) + \bar{\theta} f(\boldsymbol{y})$$

- strong convexity: f is m-strongly convex if $f(\boldsymbol{x}) - \frac{m}{2}\|\boldsymbol{x}\|_2^2$ is convex.
- examples: norm, negative entropy, log-sum-exp function, quadratic function with PSD quadratic term,...
- epigraph

$$\mathrm{epi} f = \{(\boldsymbol{x}, y) : \boldsymbol{x} \in \mathrm{dom}\, f,\ y \geq f(\boldsymbol{x})\}$$



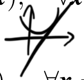
  f is a convex function iff $\mathrm{epi} f$ is a convex set.
- sublevel sets of convex functions are convex

$$C_\alpha(f) = \{\boldsymbol{x} \in \mathrm{dom}\, f : f(\boldsymbol{x}) \leq \alpha\}$$

Convex functions (2/3)

- zero-th order condition 在任一直线上凸 $g(t) = f(x+td)$
  - restriction to any line is (strictly/strongly) convex
- first-order conditions
  - convexity

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{x}), \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \operatorname{dom} f$$



  - strict convexity 太乐展开

$$f(\boldsymbol{y}) > f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{x}), \quad \forall \boldsymbol{x} \neq \boldsymbol{y} \in \operatorname{dom} f$$

  - m-strong convexity 强↑

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{x}) + \frac{m}{2}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \operatorname{dom} f$$

优先考虑

- second-order conditions
  - convexity

$$\nabla^2 f(\boldsymbol{x}) \succeq \boldsymbol{O}, \quad \forall \boldsymbol{x} \in \operatorname{dom} f$$

$A^TA \succeq 0$ 线代知识

  - strict convexity

$A^TA \succ 0$

$$\nabla^2 f(\boldsymbol{x}) \succ \boldsymbol{O}, \quad \forall \boldsymbol{x} \in \operatorname{dom} f$$

  - m-strong convexity

$|\nabla^2 f(x) - mI|$ 正定（特征值全 $\geq m$）

$$\nabla^2 f(\boldsymbol{x}) \succeq m\boldsymbol{I}, \quad \forall \boldsymbol{x} \in \operatorname{dom} f$$

Convex functions (3/3)

- convexity 保凸 preserving operations
  - nonnegative combinations

    非负线性组合

$$f(\boldsymbol{x}) = \sum_{i=1}^{m} c_i f_i(\boldsymbol{x})$$

  - composition with affine functions

    仿射保凸

$$f(\boldsymbol{x}) = g(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b})$$

  - ( )certain composition of monotonic convex/concave functions

$$f(\boldsymbol{x}) = h(g_1(\boldsymbol{x}), \ldots, g_m(\boldsymbol{x}))$$

  - pointwise maximum/supremum

    凸函数族上确界

好用

$$f(\boldsymbol{x}) = \sup_{i \in I} f_i(\boldsymbol{x})$$

  - ( )partial minimization: for convex 凹 g and 下 convex C,

$$f(\boldsymbol{x}) = \inf_{\boldsymbol{y} \in C} g(\boldsymbol{x}, \boldsymbol{y})$$

Optimization problems

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0} \\ & \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0} \end{aligned}$$

- domain $D = \operatorname{dom} f \cap (\bigcap_i \operatorname{dom} g_i) \cap (\bigcap_j \operatorname{dom} h_j)$
- feasible set
  可行集合
  $$X = \{\boldsymbol{x} \in D : \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0},\ \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}$$
  $\boldsymbol{x}$ is feasible if $\boldsymbol{x} \in X$
- $f^* = \inf_{\boldsymbol{x} \in X} f(\boldsymbol{x})$ is the optimal value
- $\boldsymbol{x}^* \in X$ is a global minimum if $f^* = f(\boldsymbol{x}^*)$, or equivalently
  (可以不存在)
  $$f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}), \quad \forall \boldsymbol{x} \in X$$
- $\boldsymbol{x}^* \in X$ is a local minimum if for some $\delta > 0$,
  $$f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}), \quad \forall \boldsymbol{x} \in X \cap B(\boldsymbol{x}^*, \delta)$$

Convex optimization problems

$$
\begin{aligned}
\overset{①}{\min_{\boldsymbol{x}}} \quad & f(\boldsymbol{x}) \\
\text{s.t.} \quad & \boldsymbol{g}(\boldsymbol{x}) \overset{②}{\leq} \mathbf{0} \\
& \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}
\end{aligned}
$$

- f, $\boldsymbol{g}$ are convex[3], $\boldsymbol{h} = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}$ is affine[4].
- key property: local minima are global minima. （有解条件 不要求）
  - no assertion about existence; $*$ some conditions for existence
  - no assertion about uniqueness; if f is strictly convex, solution is unique if exists.
    不必惟一 (non-strict)
- examples: LP, QP, QCQP, GP
- equivalent problems: informally, solution of one problem readily yields solution to the other 等价变换
  - some simple transformation: changing variables (变量替换), eliminating equality constraints (去除等式约束), introducing slack (松弛变量) variables, transforming (改变约束) objective/constraints,...
- be able to formulate simple convex optimization problems
  写出问题的凸优化形式（会写实际问题）

Optimality conditions for smooth convex problems

最优化条件 必考

对偶
也必考

- unconstrained problem

$$\nabla f(\boldsymbol{x}^*) = \mathbf{0}$$

- constrained problem

$$\nabla f(\boldsymbol{x}^*)^T(\boldsymbol{x} - \boldsymbol{x}^*) \geq 0, \quad \forall \boldsymbol{x} \in X$$

- equality constrained problem: Lagrange condition

$$\begin{cases} \nabla f(\boldsymbol{x}^*) + \boldsymbol{A}^T \boldsymbol{\lambda}^* = \mathbf{0} \\ \boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{b} \end{cases}$$

- inequality constrained problem: KKT conditions (necessary at regular point; sufficient)
  - primal feasibility: $\boldsymbol{h}(\boldsymbol{x}^*) = \mathbf{0}$, $\boldsymbol{g}(\boldsymbol{x}^*) \leq \mathbf{0}$
  - dual feasibility: $\boldsymbol{\mu}^* \geq \mathbf{0}$
  - stationarity: $\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \mathbf{0}$ 讨论 2nd
  - complementary slackness: $\mu_j^* g_j(\boldsymbol{x}^*) = 0$, $\forall j$

Lagrange duality (1/2)

- general primal problem,

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & \boldsymbol{g}(\boldsymbol{x}) \leq \mathbf{0} \\ & \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0} \end{aligned}$$

- Lagrangian

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \sum_{i=1}^{k} \lambda_i h_i(\boldsymbol{x}) + \sum_{j=1}^{m} \mu_j g_j(\boldsymbol{x})$$

  - $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \leq f(\boldsymbol{x})$ for feasible $\boldsymbol{x}$ and $\boldsymbol{\mu} \geq \mathbf{0}$

- dual function

$$\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\boldsymbol{x} \in D} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$$ 会求

  - always concave
  - domain: $\{(\boldsymbol{\lambda}, \boldsymbol{\mu}) : \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) > -\infty\}$
  - lower bound property $\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) \leq \phi^* \leq f^* \leq f(\boldsymbol{x})$ for $\boldsymbol{\mu} \geq \mathbf{0}$, $\boldsymbol{x} \in X$

Lagrange duality (2/2)

- dual problem

$$\begin{aligned} \max_{\boldsymbol{\lambda},\boldsymbol{\mu}} \quad & \phi(\boldsymbol{\lambda},\boldsymbol{\mu}) \\ \text{s.t.} \quad & \boldsymbol{\mu} \geq \mathbf{0} \end{aligned}$$

  - always a convex optimization problem
  - dual LP that makes constraints explicit
- weak duality: $\phi^* \leq f^*$
  - optimal duality gap $f^* - \phi^*$
- strong duality: $\phi^* = f^*$
  - (refined) Slater's condition for convex problems
  - strong duality almost always holds for LP
- KKT conditions and strong duality for convex problems

  成立　　取到　　　　判断强对偶

  KKT $\iff$ strong duality + primal optimality + dual optimality

Algorithms (1/2)

unconstrained problems

- smooth f
  - descent method: $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + t_k \boldsymbol{d}_k$
  - descent direction
    - negative gradient: $\boldsymbol{d}_k = -\nabla f(\boldsymbol{x}_k)$
    - Newton direction: $\boldsymbol{d}_k = -[\nabla^2 f(\boldsymbol{x}_k)]^{-1} \nabla f(\boldsymbol{x}_k)$
  - step size
    - constant
    - exact line search
    - backtracking line search (Armijo's rule)
  - $*$ condition number
  - $*$ convergence analysis
- $*$ smooth f + nonsmooth h
  - proximal gradient descent

$$\boldsymbol{x}_{k+1} = \operatorname{prox}_{t_k h}(\boldsymbol{x}_k - t_k \nabla f(\boldsymbol{x}_k))$$

$$\operatorname{prox}_h(\boldsymbol{x}) = \operatorname*{argmin}_{\boldsymbol{z}} \left\{ \frac{1}{2} \|\boldsymbol{z} - \boldsymbol{x}\|_2^2 + h(\boldsymbol{z}) \right\}$$

Algorithms (2/2)

constrained problems

- equality constraints
  - constraint elimination
  - Newton's method
    - KKT system for finding descent direction

$$\begin{bmatrix} \nabla^2 f(\boldsymbol{x}_k) & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{O} \end{bmatrix} \begin{bmatrix} \boldsymbol{d} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} -\nabla f(\boldsymbol{x}_k) \\ \boldsymbol{0} \end{bmatrix}$$

- $*$ inequality constraints
  - projected gradient descent

$$\boldsymbol{x}_{k+1} = \mathcal{P}_X(\boldsymbol{x}_k - t_k \nabla f(\boldsymbol{x}_k))$$

  - barrier method

Good Luck with Finals!