《First-Order Methods in Optimization》

读书笔记（I）

冯乐晨　李欢　倪元华

日期：2025 年 4 月

前 言

为什么撰写此笔记

优化理论作为现代应用数学的重要分支，广泛应用于数据科学、人工智能、控制工程等众多领域，其理论体系的发展对实际问题的建模与求解具有深远影响。正因如此，南开大学倪元华课题组于 2021 年 9 月开始组织优化理论专题讨论班，致力于从零开始较系统性地学习优化理论。随后，南开大学李欢老师、天津大学张新珍老师课题组也加入进来，讨论班的学习路线循序渐进、由浅入深。自 2022 年 9 月起，课题组师生一起学习 Yurii Nesterov 所著《Lectures on Convex Optimization》，和 Amir Beck 所著《First-Order Methods in Optimization》。

本笔记整理自冯乐晨同学于 2023 年 9 月至 2024 年 12 月期间优化讨论班上的读书笔记，内容围绕 Amir Beck 所著《First-Order Methods in Optimization》的前八章展开。该书作为凸优化领域的重要著作，以其严谨的理论体系和前沿的视角在学界享有盛誉。在讨论班的学习过程中，通过课题组成员的深入交流与反复推敲，逐步补全了许多被省略的推导过程，厘清了若干复杂论证的内在逻辑，并由冯乐晨同学整理读书笔记。

本读书笔记是课题组全体师生共同努力的结果，笔记的整理过程始于讨论班的课堂讲解。首先，由主讲同学对书中对应章节进行详细阐述，在讲解过程中，通过现场提问和集体讨论不断凝练和深化问题。每当出现疑惑或关键难点时，老师和同学积极互动，提出各自的见解和疑问，进一步激发对问题的探讨。课后，冯乐晨同学根据讨论班上的讲解、提问与讨论，对原有思路进行梳理和总结，最终整理成这份笔记，以便为我们后续的学习和研究提供参考资料。笔记力求在保留原书整体结构与符号系统基础上，补充必要的推导细节与解释，尽管在整理过程中力求严谨准确，但由于我们水平所限，笔记中仍会存在笔误及不严谨之处，恳请批评指正。

除南开大学倪元华老师、李欢老师，天津大学张新珍老师外，期间参加优化讨论班的还有南开大学人工智能学院的研究生同学：司彬彬（现于某银行工作）、岳新辉（现为某地公务员）、刘丽萍（现为某地公务员）、徐宏远（现于北京航空航天大学攻读博士学位）、张迪（现为某中学教师）、贾晖、王雨畑（现于香港理工大学攻读博士学位）、张震（将赴某研究所工作）、冯乐晨（将赴香港理工大学攻读博士学位）、李浩然（将赴香港理工大学攻读博士学位）、葛涵（将赴海外留学）、贾茹茹、刘姿含、于梅灵、李庆生、谢君、刘震、谭浩天、陈子涵、马燕琳、原增昀、丁依宁、贾顺、徐阳阳、王文远，和天津大学数学学院的研究生同学：余泉（现于湖南大学攻读博士学位）、朱琳（现于某银行工作）、郭洄园（现于海外攻读博士学位）、王一静（现于某公司工作）、许君霞（现于某公司工作）、王学友（将赴某银行工作）、许梦平、

赵千一、司文栋、徐思敏、李佳泽、丁鹤、郭艳鑫等。在此一并致谢，感谢各位的辛苦付出！

笔记主要内容

本笔记整理了《First-Order Methods in Optimization》第一章——第八章的全部内容，具体包括：

- 凸分析基本理论：向量空间理论、次梯度理论、共轭函数理论、强凸函数性质
- 近端算子理论、谱函数性质
- 原始-对偶投影次梯度算法

如何阅读此笔记

本笔记按照《First-Order Methods in Optimization》第一章——第八章顺序整理，如无特殊说明，记号均与书中对应章节一致。

作者简介

冯乐晨　2022 年毕业于南开大学，获智能科学与技术工学学士学位，现为南开大学人工智能学院控制科学与工程专业硕士研究生。研究兴趣主要包括最优控制、动力系统与优化方法，将于 2025 年 8 月将赴香港理工大学应用数学系攻读博士学位。

李欢　2019 年于北京大学获博士学位，现为南开大学人工智能学院副教授，在 JMLR、SIOPT、NIPS、ICML 等期刊与会议上发表论文多篇，研究方向为优化方法与机器学习。

倪元华　2010 年于中国科学院获运筹学与控制论博士学位，现为南开大学人工智能学院教授、博士生导师，并曾于 2014 年 4 月至 2015 年 5 月在美国加州大学圣地亚哥分校、2016 年 1 月至 2017 年 1 月在香港理工大学担任访问学者。在 Automatica、IEEE TAC、SICON 等期刊上发表论文多篇，研究方向为随机控制、最优控制、强化学习、智能博弈等，现为《System & Control Letters》期刊编委。

致谢

参与《First-Order Methods in Optimization》前八章主讲的同学有：南开大学的张震、贾茹茹、刘姿含、于梅灵、贾晖、李浩然、葛涵、李庆生、谢君、刘震、谭浩天、陈子涵、原增昀、丁依宁、贾顺、徐阳阳、王文远，和天津大学的王学友、许梦平、赵千一、司文栋、徐思敏、李佳泽、丁鹤、郭艳鑫；感谢诸位主讲同学的认真准备与精彩讲解。

感谢讨论班上诸位老师和同学课上课下的积极互动与讨论，这不仅加深了我们对知识的理解，也对本笔记的整理方向和内容完善起到了重要的推动作用。

目录

§1.5 Affine Sets and Convex Sets

集合 $S\subseteq\mathbb{E}$ 称为 affine 的，若对 $\forall x,y\in S$，$\lambda\in\mathbb{R}$，有

$$\lambda x+(1-\lambda)y\in S$$

下面证明任意仿射集 S，均可分解成 $S=p+V$，

其中 p 是一固定的向量，V 是线性子空间

证明：任取 $p\in S$，令 $V=S-p=\{y-p\mid y\in S\}$

下证 V 是线性子空间：

① 对加法封闭：$\forall x,y\in V$，则 $x+p$，$y+p\in S$.

故 $m=\frac{1}{2}(x+p)+\frac{1}{2}(y+p)=\frac{1}{2}(x+y)+p\in S$

故 $2m-p=x+y+2p-p=x+y+p\in S$

故 $x+y\in V$

② 加法的交换律、结合律，存在 0 元素是显然的

③ 存在负元素：$\forall x \in V$，则 $x+p \in S$

$-(x+p)+2p = -x+p \in S$，故 $-x \in V$

④ 对数乘封闭：$\forall x \in V$, $\lambda \in \mathbb{R}$, 则 $x+p \in S$

$(1-\lambda)p + \lambda(x+p) = \lambda x + p \in S \Rightarrow \lambda x \in V$

⑤ 对数乘满足分配律、结合律. $1x = x$ 显然.

$\boxed{/}$

反过来，线性子空间的平移是仿射的，是显然的.

§1.10 Linear Transformations

$\mathcal{A}: \mathbb{E} \to \mathbb{V}$ 是线性算子，若对 $\forall x, y \in \mathbb{E}$，$\alpha, \beta \in \mathbb{R}$

$$\mathcal{A}(\alpha x + \beta y) = \alpha \mathcal{A}(x) + \beta \mathcal{A}(y)$$

可以证明：对任意 $\mathbb{R}^n \to \mathbb{R}^m$ 的线性算子，有

$$\mathcal{A}(x) = Ax \text{ 的形式}$$

对任意 $\mathbb{R}^{m\times n} \to \mathbb{R}^k$ 的线性算子，有

$$\mathcal{A}(X) = \begin{bmatrix} \mathrm{Tr}(A_1^T X) \\ \vdots \\ \mathrm{Tr}(A_k^T X) \end{bmatrix} \text{ 的形式}$$

证明：令 $A_{ij} \in \mathbb{R}^{m\times n}$ 表示 (i,j) 处是 1，其余是 0 的矩阵，则 $\{A_{ij}\}$ 构成了 $\mathbb{R}^{m\times n}$ 中的一组基

设 $$\mathcal{A}(A_{ij}) = (e_1, \cdots, e_k)\begin{pmatrix} a_{ij}^{(1)} \\ \vdots \\ a_{ij}^{(k)} \end{pmatrix}$$

故 $$\mathcal{A}(X) = \mathcal{A}(A_{11}, \cdots, A_{ij}, \cdots, A_{mn})\,\mathrm{vec}(X)$$

$$= \underbrace{(e_1, \cdots, e_k)}_{=I}\begin{pmatrix} a_{11}^{(1)} & \cdots & a_{ij}^{(1)} & \cdots & a_{mn}^{(1)} \\ \vdots & & \vdots & & \vdots \\ a_{11}^{(k)} & \cdots & a_{ij}^{(k)} & \cdots & a_{mn}^{(k)} \end{pmatrix}\mathrm{vec}(X)$$

$$= \begin{bmatrix} \sum_{i,j} a_{ij}^{(1)} x_{ij} \\ \vdots \\ \sum_{i,j} a_{ij}^{(k)} x_{ij} \end{bmatrix},$$ 故有上面的形式

$\boxtimes$

§ 1.11 The Dual Space

$\mathbb{E}$ 上的线性泛函全体构成 $\mathbb{E}^*$

设 $\{e_1,\cdots,e_n\}$ 是 E 上一组基，则定义 $f_i\ (i=1,\cdots,n)$

$$f_i(e_j)=\delta_{ij}=\begin{cases}1, & i=j\\ 0, & i\neq j\end{cases}$$

则 $\{f_1,\cdots,f_n\}$ 是 E^* 上的一组基

证明：若 $x=a_1e_1+\cdots+a_ne_n$，则 $f_i(x)=a_i$

设 $f\in E^*$，且 $f(e_j)=b_j\ (j=1,\cdots,n)$，令

$$g=b_1f_1+\cdots+b_nf_n$$

则 $g(x)=(b_1f_1+\cdots+b_nf_n)(a_1e_1+\cdots+a_ne_n)$

$$=b_1a_1+\cdots+b_na_n$$

$$=f(x)$$

由 x 的任意性：$f=g=b_1f_1+\cdots+b_nf_n$

又由：若 $\exists c_1,\cdots,c_n$，s.t. $c_1f_1+\cdots+c_nf_n=0$

依次作用到 $e_1,\cdots,e_n$，则 $c_1=\cdots=c_n$，故 $f_1,\cdots,f_n$ 无关

$\square$

故给定 $f\in E^*$，$\exists v\in E$，s.t. $f(x)=\langle v,x\rangle$

v 就是 f 在 $f_1, \cdots, f_n$ 下的坐标

推广到 Hilbert空间：

216

定理 2.7. *(Riesz表示定理)* 设 H 为 Hilbert空间，则 $\forall f \in H^*$，$\exists !\ u_f \in H$ 使 $\|f\| = \|u_f\|$，$f(x) = (x, u_f)$，$\forall x \in H$。

反之，若 $u \in H$ 则 $f(x) = (x, u)$ 是 H 上的有界线性泛函且 $\|f\| = \|u\|$。

证明 存在性 不妨设 $f \neq \theta$，$L = Ker\, f = \{x \in H : f(x) = 0\}$ 为 H 的闭子空间，由正交分解定理 $H = L \oplus L^{\perp}$（$L \cap L^{\perp} = \{\theta\}$），显然 $L^{\perp} \neq \{\theta\}$，取 $y_0 \in L^{\perp}$ 使 $f(y_0) = 1$。

$$\forall x \in H, \quad f(x - f(x)y_0) = f(x) - f(x)f(y_0) = 0$$
$$\Longrightarrow x - f(x)y_0 \in L$$
$$\Longrightarrow (x - f(x)y_0, y_0) = (x, y_0) - f(x)(y_0, y_0) = 0,$$
$$\therefore \quad (x, y_0) = f(x)(y_0, y_0)$$

217

$$\Longrightarrow f(x) = \frac{1}{(y_0, y_0)}(x, y_0) = (x, \frac{y_0}{(y_0, y_0)}) = (x, u_f)$$

其中 $u_f = \frac{y_0}{(y_0, y_0)}$。

唯一性 设 $f(x) = (x, v)$，$\forall x \in H$，则 $0 = f(x) - f(x) = (x, u_f - v)$，$\forall x \in H \Longrightarrow u_f - v = \theta \Longrightarrow u_f = v$。

由 $|f(x)| = |(x, u_f)| \leq \|x\|\|u_f\|$（Schwarz不等式）有 $\|f\| \leq \|u_f\|$。

而 $f(u_f) = \|u_f\|^2$ 或 $f(\frac{u_f}{\|u_f\|}) = \|u_f\|$,

故 $\|f\| = \|u_f\|$。

反之部分显然。 $\square$

定义记号：$\langle f, x \rangle = f(x)$，$x \in E$，$f \in E^*$

定义 dual norm

$$\|y\|_* = \max_x \{ |\langle y, x \rangle| : \|x\| \leq 1 \}$$
$$= \max_{\|x\|=1} \{ |\langle y, x \rangle| \}$$
$$= \max_{\|x\|=1} \{ \langle y, x \rangle \}.$$

Lemma 1.4 令 $\mathbb{E}$ 是装备 $\|\cdot\|$ 的内积空间，则

$$|\langle y, x\rangle| \leq \|y\|_* \|x\|, \text{对} \forall y \in \mathbb{E}^*, x \in \mathbb{E}$$

证明：由 $\|\cdot\|_*$ 是算子范数直接可证 $\boxtimes$

Example 1.6 (Q-norm) $Q \in \mathcal{S}^n_{++}$

$\|\cdot\|_Q$ 的对偶 norm 是 $\|\cdot\|_{Q^{-1}}$

证明：$\|x\|_Q = \|Q^{\frac{1}{2}}x\|_2 \doteq \|y\|_2 \quad y = Q^{\frac{1}{2}}x$

$$\|f\|_* = \sup_{\|x\|_Q=1} |\langle f, x\rangle|$$

$$= \sup_{\|y\|_2=1} |\langle Q^{-\frac{1}{2}}f, y\rangle|$$

故 $y = \frac{Q^{-\frac{1}{2}}f}{\|Q^{-\frac{1}{2}}f\|}$ 时，取最大值，此时

$$\|f\|_* = \sqrt{f^T \Theta^{-1} f} = \|f\|_{\Theta^{-1}} \qquad \boxed{/}$$

Example 1.7 $\mathbb{E} = \mathbb{E}_1 \times \cdots \times \mathbb{E}_m$, $\|\cdot\|_{E_1}, \cdots, \|\cdot\|_{E_m}$

$$\left\langle (v_1, \cdots, v_m), (w_1, \cdots, w_m) \right\rangle = \sum_{i=1}^{m} \langle v_i, w_i \rangle$$

$$\|(u_1, \cdots, u_m)\| = \sqrt{\sum_{i=1}^{m} w_i \|u_i\|_{E_i}^2}, \quad u_i \in \mathbb{E}_i$$

令 $W = \mathrm{diag}(w_1, \cdots, w_n)$, $\vec{v} = (v_1, \cdots, v_n)$

$\vec{u} = (u_1, \cdots, u_n)$, 则令 $\vec{y} = \sqrt{W} \vec{u}$

则 $$\|\vec{v}\|_* = \sup_{\|\vec{u}\|=1} \left| \langle \vec{v}, \vec{u} \rangle \right|$$

$$= \sup_{\|\vec{y}\|_2=1} \left| \langle \sqrt{W}^{-1} \vec{v}, \sqrt{W} \vec{u} \rangle \right|$$

$$= \sup_{\|\vec{y}\|_2=1} \left| \langle \sqrt{W}^{-1} \vec{v}, \vec{y} \rangle \right|$$

$$= \|\sqrt{W^{-1}}\,\vec{v}\|_2 \, \|\vec{y}\|_2 \quad (\text{C-S不等式})$$

$$= \left\| \left(\sqrt{w_1^{-1}}\, v_1, \dots, \sqrt{w_n^{-1}}\, v_n \right) \right\|$$

$$= \sqrt{\sum_{i=1}^{n} w_i^{-1} \|v_i\|_{E_i}^2}$$

§ 2.2 Closed verus Continuity

Thm 2.8 $f: \mathbb{E} \to (-\infty, \infty]$ 是 extended real-valued function，且在 $\mathrm{dom} f$ 上连续，$\mathrm{dom} f$ 闭，则 f 闭

证明：想证 $\mathrm{epi}(f)$ 是闭的

取 $\{(x_n, y_n)\}_{n\geq 1} \subseteq \mathrm{epi}(f)$，且 $(x_n, y_n) \to (x^*, y^*)$

由 $\mathrm{dom} f$ 的闭性，$x^* \in \mathrm{dom} f$，由 f 的连续性

$f(x^*) \leq y^*$，故 $(x^*, y^*) \in \mathrm{epi}(f)$ $\boxtimes$

Corollary 2.9 $f: \mathbb{E} \to \mathbb{R}$ 连续，则 f 闭

Example 2.10 $f_\alpha: \mathbb{R} \to (-\infty, \infty]$

$$f_\alpha(x) = \begin{cases} \alpha, & x = 0 \\ x, & 0 < x \leq 1 \\ \infty, & \text{else} \end{cases}$$

则 $\alpha \le 0 \Leftrightarrow f_\alpha$ 是闭的 $\boxed{/}$

<u>Example 2.11</u>

$$f(x) = \|x\|_0 = \#\{i : x_i \neq 0\}$$

易知： $f(x) = \sum_{i=1}^{n} I(x_i)$，其中 $I(y) = \begin{cases} 0, & y=0 \\ 1, & y \neq 0 \end{cases}$

由 $\mathrm{Lev}(I, \alpha) = \begin{cases} \phi & \alpha < 0 \\ \{0\} & \alpha \in [0,1) \\ \mathbb{R} & \alpha \ge 1 \end{cases}$ 对 $\forall \alpha$ 是闭集

故 $f(x)$ 是 closed $\boxed{/}$

<u>Thm 2.12</u> $f: \mathbb{E} \to (-\infty, \infty]$ 是 proper 闭 func，设 C 是紧的，且 $C \cap \mathrm{dom}(f) \neq \phi$，则

(a) f 在 C 上下方有界

(b) f 在 C 上可达到最小值

证明:

(a) 用反证法，设 f 在 C 上是下方无界的，则 $\exists \{x_n\}_{n\geq 1} \subseteq C$,

s.t. $\lim\limits_{n\to\infty} f(x_n) = -\infty$

由于 C 是紧集，故存在子序列 $\{x_{n_k}\}_{k\geq 1}$, s.t. $x_{n_k} \to \bar{x} \in C$

由 f 闭，故 f 下半连续，从而

$$f(\bar{x}) \leq \liminf_{k\to\infty} f(x_{n_k}) = -\infty$$

由于 f 是 proper 的，故推出矛盾 $\boxtimes$

(b) 记 $f_{opt} = \inf\limits_{x\in C} f(x)$，故 $\exists \{x_n\}$, s.t. $f(x_n) \to f_{opt}$

且 $\exists$ 子列 $\{x_{n_k}\} \to \bar{x} \in C$，有

$$f(\bar{x}) \leq \liminf_{k\to\infty} f(x_{n_k}) = f_{opt}$$

故在 $\bar{x}$ 处取到最小值 $\boxtimes$

<u>Def 2.13</u> Proper func $f: \mathbb{E} \to (-\infty, +\infty]$ 是 coercive 的,

若 $\lim\limits_{\|x\|\to\infty} f(x) = \infty$

Thm 2.14 令 $f:\mathbb{E}\to(-\infty,\infty]$ 是 proper, closed, coercive func, $S\subseteq\mathbb{E}$ 是非空闭集，且 $S\cap\mathrm{dom}(f)\neq\emptyset$，则 f 在 S 上可达最小值

证明：令 $x_0\in S\cap\mathrm{dom}(f)$，由 f 的强制性

$\exists M>0$, s.t.

$$f(x)>f(x_0) \text{ 对 } \forall x, \text{ s.t. } \|x\|>M \text{ 成立}$$

由于 x^* 满足 $f(x^*)\leq f(x_0)$，故优化问题 $\Leftrightarrow$

$$\min_{x\in S\cap B_{\|\cdot\|}[0,M]} f(x)$$

由于紧集上的下半连续 func 能取到最小 $\square$

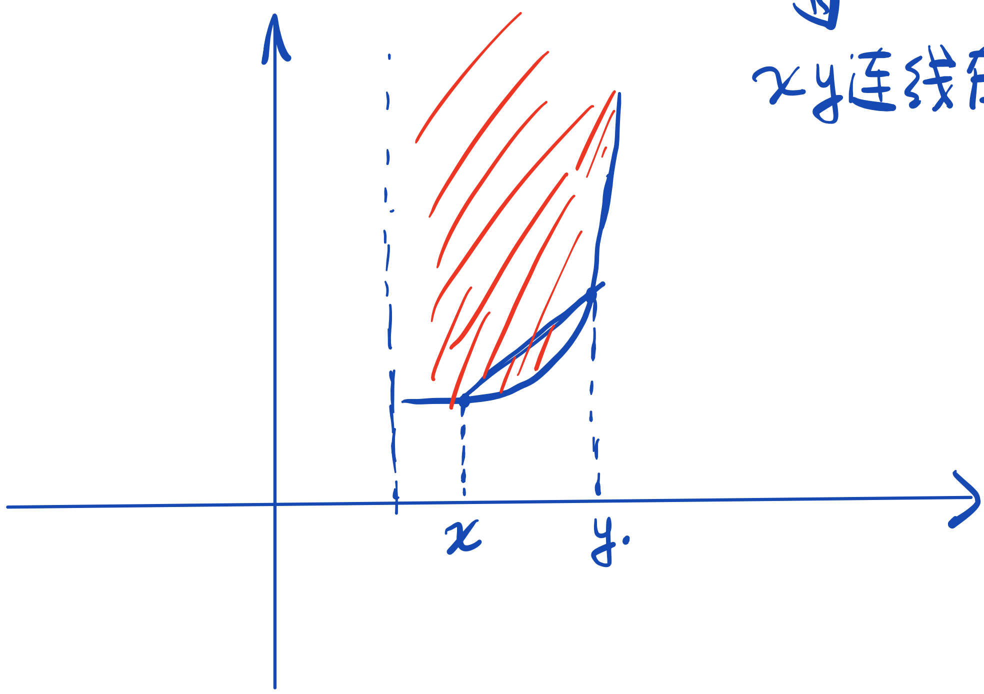
§ 2.3 Convex Function

Def 2.15 extended $f:\mathbb{E}\to[-\infty,\infty]$ 是凸的，若

$\mathrm{epi}(f)$ 是凸的

注：与 $f(\lambda x+(1-\lambda)y)\le\lambda f(x)+(1-\lambda)f(y)$ 的定义

等价 $\Longleftrightarrow$ xy 连线在 $\mathrm{epi}(f)$ 里



Thm 2.16 （保凸运算）

(a) $A: \mathbb{E}\to\mathbb{V}$ 是线性变换，$b\in\mathbb{V}$

$f:\mathbb{V}\to(-\infty,\infty]$ 是凸的，则 $g(x)=f(A(x)+b)$ 是凸的

(b) $f_1,\dots,f_m:\mathbb{E}\to(-\infty,\infty]$ 凸，$\alpha_1,\dots,\alpha_m\in\mathbb{R}_+$，则 $\sum_{i=1}^m \alpha_i f_i$ 凸

(c) $f_i:\mathbb{E}\to(-\infty,\infty]$，$i\in I$ 凸，则 $f(x)=\max_{i\in I} f_i(x)$ 凸

证明：用定义都显然 $\boxtimes$

令 $d_C(x) = \min_{y\in C} \|x-y\|$

<u>Example 2.17</u> 设 $\|\cdot\| = \sqrt{\langle\cdot,\cdot\rangle}$，令 $C \subseteq \mathbb{E}$ 非空

$$\varphi_C(x) = \frac{1}{2}\left(\|x\|^2 - d_C^2(x)\right)$$

由 $d_C^2(x) = \min_{y\in C} \|x-y\|^2 = \|x\|^2 - \max_{y\in C}\left[2\langle y,x\rangle - \|y\|^2\right]$

故 $\varphi_C(x) = \max_{y\in C}\left[\langle y,x\rangle - \frac{1}{2}\|y\|^2\right]$

由 <u>Thm 2.16</u> (c)，知 φ_C 是凸的 $\square$

<u>Thm 2.18</u> $f: \mathbb{E}\times\mathbb{V} \to (-\infty,\infty]$ 是 convex 的，且

对 $\forall x \in \mathbb{E}$，$\exists y \in \mathbb{V}$，s.t. $f(x,y) < \infty$

则 $g(x) = \min_{y\in\mathbb{V}} f(x,y)$ 凸

证明：令 $x_1, x_2 \in \mathbb{E}$，$\lambda \in [0,1]$，要证

$$g(\lambda x_1 + (1-\lambda)x_2) \leq \lambda g(x_1) + (1-\lambda) g(x_2)$$

$\lambda=0$ 或 1 时自然成立，故设 $\lambda\in(0,1)$

<u>Case I</u>：设 $g(x_1), g(x_2) > -\infty$，$\forall$ 取 $\varepsilon>0$，$\exists\, y_1, y_2\in \mathbb{V}$

$$f(x_1,y_1)\le g(x_1)+\varepsilon$$
$$f(x_2,y_2)\le g(x_2)+\varepsilon$$

$$\begin{aligned}g(\lambda x_1+(1-\lambda)x_2) &\le f(\lambda x_1+(1-\lambda)x_2, \lambda y_1+(1-\lambda)y_2)\\ &\le \lambda f(x_1,y_1)+(1-\lambda)f(x_2,y_2)\\ &\le \lambda(g(x_1)+\varepsilon)+(1-\lambda)(g(x_2)+\varepsilon)\\ &= \lambda g(x_1)+(1-\lambda)g(x_2)+\varepsilon\end{aligned}$$

<u>Case II</u>：$g(x_1)$ 或 $g(x_2)$ 是 $-\infty$，WLOG，设 $g(x_1)=-\infty$

下证 $g(\lambda x_1+(1-\lambda)x_2)=-\infty$：

$\forall$ 取 $M\in\mathbb{R}$，$\exists\, y_1\in\mathbb{V}$, s.t. $f(x_1,y_1)\le M$

又 $\exists\, y_2\in\mathbb{V}$, s.t. $f(x_2,y_2)<\infty$，故

$$\begin{aligned}f(\lambda x_1+(1-\lambda)x_2, \lambda y_1+(1-\lambda)y_2) &\le \lambda f(x_1,y_1)+(1-\lambda)f(x_2,y_2)\\ &\le \lambda M+(1-\lambda)f(x_2,y_2)\end{aligned}$$

$\Rightarrow g(\lambda x_1+(1-\lambda)x_2) \leq \lambda M + (1-\lambda) f(x_2, y_2)$ $\boxed{/}$

§ 2.3.2 Infimal Convolution

令 $h_1, h_2 : \mathbb{E} \to (-\infty, \infty]$

$$(h_1 \square h_2)(x) \equiv \min_{u \in \mathbb{E}} \{h_1(u) + h_2(x-u)\}$$

称为 h_1, h_2 的 infimal convolution

Thm 2.19 $h_1 : \mathbb{E}_1 \to (-\infty, \infty]$ 是 proper convex 的, $h_2 : \mathbb{E} \to \mathbb{R}$ 是凸的, 则 $h_1 \square h_2$ 是凸的

证明: $f(x, y) \equiv h_1(y) + h_2(x-y)$ 关于 (x, y) 凸

且对 $\forall x \in \mathbb{E}$, 取 $y \in \text{dom}(h_1)$, 有 $f(x, y) < \infty$, 故

利用 Thm 2.18 立即可证 $\boxed{/}$

Example 2.20

$$d_C(x) = \min_y \{\|x-y\|, y \in C\} = \min_y \{\delta_C(y) + \|x-y\|\}$$

$$= (\delta_C \square h_1)(x)$$

$\boxed{\checkmark}$

故 d_C 是凸的

§ 2.3.3 Continuity of Convex Functions

Thm 2.21 令 $f: \mathbb{E} \to (-\infty, \infty]$ 是 convex 的，令 $x_0 \in \text{int}(\text{dom}(f))$，则 $\exists\, \varepsilon > 0, L > 0$, s.t. $B[x_0, \varepsilon] \subseteq C$,

$$|f(x) - f(x_0)| \leq L \|x - x_0\|$$

对 $\forall x \in B[x_0, \varepsilon]$ 成立

证明：[Nesterov, Thm 3.1.11]

Thm 2.22 $f: \mathbb{R} \to (-\infty, \infty]$ proper，闭凸，则 f 在 dom f 上连续

证明：[Nesterov, Lemma 3.1.4]

§ 2.4 Support function

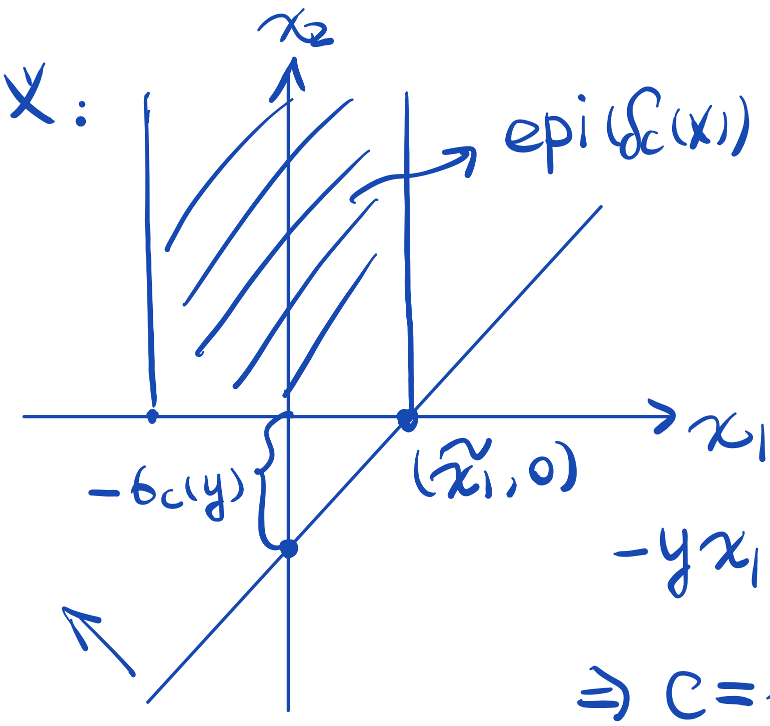
令 $C \subseteq \mathbb{E}$ 非空，则 C 的 support func $\sigma_C : \mathbb{E}^* \to (-\infty, \infty]$

$$\sigma_C(y) = \max_{x \in C} \langle y, x \rangle$$

<u>Lemma 2.23</u>　$C \subseteq \mathbb{E}$ 是非空集，则 σ_C 闭凸

证明：由 Thm 2.7(c) 和 Thm 2.16(c)，成立　$\boxtimes$

几何意义：



法向量 $(-y, 1)$

$$\overset{c}{\underset{\shortparallel}{}}$$
$-y x_1 + x_2 = c$ 过 $(\hat{x}_1, 0)$

$\Rightarrow c = -y\hat{x}_1$

Lemma 2.24

(a) 对 $\forall$ 非空集 $C \subseteq \mathbb{E}$, $y \in \mathbb{E}^*$, $\alpha \geq 0$, 有

$$\sigma_C(\alpha y) = \alpha \sigma_C(y)$$

(b) 对 $\forall$ 非空 $C \subseteq \mathbb{E}$, $y_1, y_2 \in \mathbb{E}^*$ 有

$$\sigma_C(y_1 + y_2) \leq \sigma_C(y_1) + \sigma_C(y_2)$$

(c) 对 $\forall$ 非空 $C \subseteq \mathbb{E}$, $y \in \mathbb{E}^*$, $\alpha \geq 0$

$$\sigma_{\alpha C}(y) = \alpha \sigma_C(y)$$

(d) $\forall$ 非空 $A, B \subseteq \mathbb{E}$, $y \in \mathbb{E}^*$

$$\sigma_{A+B}(y) = \sigma_A(y) + \sigma_B(y)$$

证明:

(a) $\sigma_C(\alpha y) = \max\limits_{x \in C} \langle \alpha y, x \rangle = \alpha \max\limits_{x \in C} \langle y, x \rangle = \alpha \sigma_C(y)$

(b)
$$\begin{aligned}
\sigma_C(y_1 + y_2) &= \max_{x \in C} \langle y_1 + y_2, x \rangle \\
&\leq \max_{x \in C} \langle y_1, x \rangle + \max_{x \in C} \langle y_2, x \rangle \\
&= \sigma_C(y_1) + \sigma_C(y_2)
\end{aligned}$$

(c) $\sigma_{\alpha C}(y) = \max\limits_{x \in \alpha C} \langle y, x \rangle = \max\limits_{x_1 \in C} \langle y, \alpha x_1 \rangle = \alpha\, \sigma_C(y)$

(d) $\sigma_{A+B}(y) = \max\limits_{x \in A+B} \langle y, x \rangle = \max\limits_{\substack{x_1 \in A \\ x_2 \in B}} \langle y, x_1 + x_2 \rangle$

$$= \max_{x_1 \in A} \langle y, x_1 \rangle + \max_{x_2 \in B} \langle y, x_2 \rangle = \sigma_A(y) + \sigma_B(y)$$

$\square$

Example 2.25 $\quad C = \{b_1, \cdots, b_m\}$

$$\sigma_C(y) = \max\{\langle b_1, y\rangle, \langle b_2, y\rangle, \cdots, \langle b_m, y\rangle\}$$

$\square$

Example 2.26 $\quad \sigma_K(y) = \delta_{K^\circ}(y)$

证明：

若 $y \in K^\circ$，则 $\langle y, x\rangle \leq 0 \ \forall x \in K$，而 $\langle y, \overset{\in K}{0}\rangle = 0$，故

$$\sigma_K(y) = \max_{x \in K} \langle y, x \rangle = 0$$

若 $y \notin K^\circ$，则 $\exists \tilde{x} \in K$，s.t. $\langle y, \tilde{x}\rangle > 0$，则对 $\forall \lambda > 0$，有

$$\sigma_K(y) = \max_{x \in K} \langle y, x \rangle \geq \langle y, \lambda \tilde{x} \rangle$$

取 $\lambda \to \infty$，有 $\sigma_K(y) = \infty$ $\quad\square$

Example 2.27 $\sigma_{\mathbb{R}^n_+}(y) = \delta_{\mathbb{R}^n_-}(y)$

Lemma 2.28 (Farkas' Lemma)

令 $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$, 则下面等价

A. $Ax \le 0 \Rightarrow c^T x \le 0$

B. $\exists\, y \in \mathbb{R}^m_+$, s.t. $A^T y = c$

证明:

$(B \Rightarrow A)$ $c^T x = y^T A x \le 0$

$(A \Rightarrow B)$ 只需证明逆否命题

若 $A^T y = c$, $y \ge 0$ 无解, 则 $Ax \le 0$, $c^T x > 0$ 有解

令 $K = \{x \in \mathbb{R}^n : x = A^T y, \forall y \ge 0\}$, 则

K 是闭凸锥

证明: 先证闭性, claim: 若 $x \in K$, 则 $\exists$ A^T 的一组线性无关的列向量 $(\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_r) = \tilde{A}$, s.t. $x = \tilde{A}\tilde{y}$,

且 $\tilde{y} \ge 0$

事实上，由 $x=A^Ty=\sum_{i=1}^{m}a_iy_i$，若 $\{a_i\}_{i=1}^{m}$ 线性相关，则

$\exists$ 不全为 0 的 $\{c_i\}_{i=1}^{m}$，s.t. $\sum_{i=1}^{m}c_ia_i=0$，取 $\delta>0$，有

$$\delta\sum_{i=1}^{m}c_ia_i+\sum_{i=1}^{m}a_iy_i=x$$

不失一般地，总可以设 $\exists\tilde{i}$，s.t. $c_{\tilde{i}}<0$

$$\Rightarrow\quad \sum_{i=1}^{m}(\delta c_i+y_i)a_i=0$$

则 δ 从 0^+ 逐渐增大，总 $\exists i^*$，s.t. $\delta c_{i^*}+y_{i^*}=0$

$$x=\sum_{i\neq i^*}(\underbrace{\delta c_i+y_i}_{\geq 0})a_i$$

重复此操作，claim 得证

故 $K=\bigcup_i K_i \quad K_i=\{x\mid\tilde{A}y=x,y\geq 0,\tilde{A}$ 是 A^T 的线性无关列向量排成的矩阵$\}$

K_i 的闭性是显然的（由 $(\tilde{A}^T\tilde{A})^{-1}$ 存在），故 K 闭

K 的凸性易证

故由凸集分离定理：$\exists p\in\mathbb{R}^n,\alpha\in\mathbb{R}$

$$\langle p,x\rangle\leq\alpha,\ \forall x\in K\quad \langle c,x\rangle>\alpha=0$$

断言：$\alpha=0$，显然 $\alpha\geq 0$，若 $\exists\tilde{x}\in K$, s.t.

$\langle p,\tilde{x}\rangle>0$，则由 $\forall\lambda\in\mathbb{R}_+$，$\lambda\tilde{x}\in K$，推出矛盾

故 $\langle p, A^Ty\rangle=\langle Ap, y\rangle\leq 0$ 对 $\forall y\in\mathbb{R}^m_+$

故 $Ap\leq 0$ $\boxtimes$

Example 2.29 $S=\{x\in\mathbb{R}^n: Ax\leq 0\}$

则 $\sigma_S(y)=\delta_{S^\circ}(y)$，又由 $y\in S^\circ\Leftrightarrow\langle y,x\rangle\leq 0$ $\forall x$ 满足 $Ax\leq 0$

由 Farkas's Lemma：$S^\circ=\{A^T\lambda:\lambda\in\mathbb{R}^m_+\}$

故 $\sigma_S(y)=\delta_{\{A^T\lambda:\lambda\in\mathbb{R}^m_+\}}(y)$ $\boxtimes$

Example 2.30 $C=\{x\in\mathbb{R}^n: Bx=b\}$

令 x_0 是 $Bx=b$ 的任一特解，故

$$\sigma_C(y)=\max_z\{\langle y,z\rangle+\langle y,x_0\rangle: Bz=0\}$$

$$=\langle y,x_0\rangle+\sigma_{\tilde{C}}(y)$$

$$\tilde{C} = \{x \in \mathbb{R}^n : Bx = 0\} = \{x \in \mathbb{R}^n : Ax \le 0\} \quad A = \begin{pmatrix} B \\ -B \end{pmatrix}$$

故 $\tilde{C}^\circ = \{ B^T\lambda_1 - B^T\lambda_2 : \lambda_1, \lambda_2 \in \mathbb{R}^m_+ \}$

断言：$\tilde{C}^\circ = \mathrm{Range}(B^T)$ （显然的）

故 $\sigma_C(y) = \langle y, x_0 \rangle + \delta_{\mathrm{Range}(B^T)}(y)$ $\quad\boxed{\diagup}$

注：$\sigma_C(y)$ 与 x_0 的选取无关：

若 $y \in \mathrm{Range}(B^T)$, $\exists z$, s.t. $y = B^T z$, 故 $\langle y, x_0 \rangle = \langle B^T z, x_0 \rangle$

$$= \langle z, b \rangle$$

若 $y \notin \mathrm{Range}(B^T)$, 则右边 $\equiv +\infty$

<u>Example 2.31</u>

$$\sigma_{B_{\|\cdot\|}[0,1]}(y) = \max_{\|x\| \le 1} \langle y, x \rangle = \|y\|_*$$

Example 2.32 $\quad C=\{(x_1,x_2)^T : x_1+\frac{x_2^2}{2}\leq 0\}$

则 $\sigma_C(y)=\max\limits_{x_1,x_2}\{y_1x_1+y_2x_2 : x_1+\frac{x_2^2}{2}\leq 0\}$

显然，$\sigma_C(0)=0$

当 $y\neq 0$ 时，Claim $\sigma_C(y)=\max\limits_{x_1,x_2}\{y_1x_1+y_2x_2 : x_1+\frac{x_2^2}{2}=0\}$

若 $\max\limits_{x_1,x_2}\{y_1x_1+y_2x_2 : x_1+\frac{x_2^2}{2}\leq 0\}$ 最大值在内点，则

$\nabla_x(y_1x_1+y_2x_2)=\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}=0$，与 $y\neq 0$ 矛盾

故 $\sigma_C(y)=\max\limits_{x_2}\{-\frac{y_1}{2}x_2^2+y_2x_2\}$

$$=\begin{cases} \frac{y_2^2}{2y_1}, & y_1>0 \\ 0, & y_1=y_2=0 \\ \infty, & \text{else} \end{cases}$$

由 lemma 2.23，$\sigma_C(\cdot)$ 是闭凸 func，但在 $(0,0)$ 处

不连续：对 $\forall \alpha>0$，令 $y_1(t)=\dfrac{t^2}{2\alpha}$，$y_2(t)=t \ (t>0)$

则 $(y_1(t), y_2(t)) \longrightarrow (0,0) \ (t\to 0)$，但

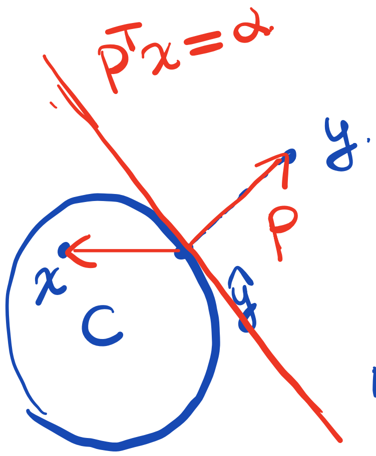
$$\sigma_C(y_1(t), y_2(t)) \equiv \alpha$$

$\boxed{/}$

Thm 2.33 令 $C \subseteq \mathbb{E}$ 是非空闭凸集，令 $y \notin C$，则 $\exists p \in \mathbb{E}^* \setminus \{0\}$，$\alpha \in \mathbb{R}$，s.t. $\langle p, y\rangle > \alpha$，且

$$\langle p, x\rangle \leq \alpha, \ \forall x \in C$$

证明：



令 $\hat{y}=\pi_C(y)$，知

$$\langle x-\hat{y}, y-\hat{y}\rangle \geq 0 \quad \forall x \in C$$

令 $p=y-\hat{y}$，$\alpha=(y-\hat{y})^T\hat{y}=p^T\hat{y}$

则对 $\forall x \in C$，有 $\langle p, x\rangle \leq \alpha$，且

$p^T y-\alpha=\|y-\hat{y}\|^2>0$，从而 $\langle p, y\rangle > \alpha$

$\boxed{/}$

Lemma 2.33 令 $A, B \subseteq \mathbb{E}$ 是非空闭凸集，则

$$A = B \Leftrightarrow \sigma_A = \sigma_B$$

证明：

$(\Rightarrow)$ $A = B \Rightarrow \sigma_A = \sigma_B$ 是显然的

$(\Leftarrow)$ 用反证法，WLOG $\exists y \in A$，但 $y \notin B$，故

$\exists p \in \mathbb{E}^* \setminus \{0\}, \alpha > 0$, s.t.

$$\langle p, x \rangle \leq \alpha < \langle p, y \rangle, \forall x \in B$$

故 $\sigma_B(p) \leq \alpha < \langle p, y \rangle \leq \sigma_A(p)$，与 $\sigma_A = \sigma_B$ 矛盾

Lemma 2.35 令 $A \subseteq \mathbb{E}$ 非空，则

(a) $\sigma_A = \sigma_{cl(A)}$

(b) $\sigma_A = \sigma_{conv(A)}$

证明：

(a) 由 $A \subseteq cl(A)$，则 $\sigma_A(y) \leq \sigma_{cl(A)}(y)$ $\forall y \in \mathbb{E}^*$

下证 $\sigma_{cl(A)}(y) \leq \sigma_A(y)$ 对 $\forall y \in \mathbb{E}^*$ 成立

对 $\forall y \in \mathbb{E}^*$，$\exists \{x^k\} \subseteq cl(A)$，s.t.

$$\langle y, x^k \rangle \to \sigma_{cl(A)}(y),\ k \to \infty$$

又由闭包的定义：$\exists \{z^k\} \subseteq A$，s.t. $\|z^k - x^k\| \leq \frac{1}{k}$，$\forall k$

$$故\ z^k - x^k \to 0,\ k \to \infty$$

由 $z^k \in A$，故

$$\sigma_A(y) \geq \langle y, z^k \rangle = \langle y, x^k \rangle + \langle y, z^k - x^k \rangle$$

令 $k \to \infty$ 知 $\sigma_A(y) \geq \sigma_{cl(A)}(y)$

(b) 由 $A \subseteq \mathrm{Conv}(A)$，有 $\sigma_A(y) \leq \sigma_{\mathrm{Conv}(A)}(y)$ $\forall y$

对 $\forall y \in \mathbb{E}^*$，则 $\exists \{x^k\} \subseteq \mathrm{Conv}(A)$，s.t.

$$\langle y, x^k \rangle \to \sigma_{\mathrm{conv}(A)}(y),\ k\to\infty$$

对 $\forall k$, $\exists z_1^k, \cdots, z_{n_k}^k \in A$, $\lambda^k \in \Delta_{n_k}$, s.t. $x^k = \sum_{i=1}^{n_k} \lambda_i^k z_i^k$

故 $\langle y, x^k \rangle = \langle y, \sum_{i=1}^{n_k} \lambda_i^k z_i^k \rangle = \sum_{i=1}^{n_k} \lambda_i^k \langle y, z_i^k \rangle$

$$\leq \sum_{i=1}^{n_k} \lambda_i^k \sigma_A(y) = \sigma_A(y)$$

取 $k\to\infty$，知 $\sigma_{\mathrm{conv}(A)}(y) \leq \sigma_A(y)$ $\boxed{/\!/}$

<u>Example 2.36</u> $\Delta_n = \mathrm{conv}\{e_1, \cdots, e_n\}$

则 $\sigma_{\Delta_n}(y) = \max\{\langle e_1, y\rangle, \cdots, \langle e_n, y\rangle\}$

$$= \max\{y_1, \cdots, y_n\}$$ $\boxed{/\!/}$

Chapter 3 Subgradient

§3.1

Def 3.1 设 $f: \mathbb{E} \to (-\infty, \infty]$ 是 proper function,

令 $x \in \mathrm{dom} f$, $g \in \mathbb{E}^*$ 是 subgradient of f 于 x, 若

$$f(y) \geq f(x) + \langle g, y-x \rangle \quad \forall y \in \mathbb{E}$$

注：只需考查：

$$f(y) \geq f(x) + \langle g, y-x \rangle, \forall y \in \mathrm{dom} f$$

Def 3.2 f 在 x 处的次微分 $\partial f(x)$ 是

$$\partial f(x) \equiv \{ g \in \mathbb{E}^* : f(y) \geq f(x) + \langle g, y-x \rangle, \forall y \in \mathbb{E} \}$$

Example 3.3 $f(x) = \|x\|$, 则

$$\partial f(0) = B_{\|\cdot\|_*}[0,1] = \{ g \in \mathbb{E}^* : \|g\|_* \le 1 \}$$

证明：$g \in \partial f(0) \Leftrightarrow \|y\| \ge \langle g, y \rangle \ \forall y \in \mathbb{E}$

下证 $\|y\| \ge \langle g, y \rangle \Leftrightarrow \|g\|_* \le 1$

$(\Leftarrow)$ $\langle g, y \rangle \le \|g\|_* \|y\| \le \|y\|$

$(\Rightarrow)$ $\|g\|_* = \max\limits_{y: \|y\| \le 1} \langle g, y \rangle \le \max\limits_{\|y\| \le 1} \|y\| = 1$

$\square$

给定$S\subseteq \mathbb{E}$, $x\in S$, 则正则锥

$$N_S(x)=\{y\in \mathbb{E}^*:\langle y,z-x\rangle\leq 0 \text{ 对 } \forall z\in S\}$$

显然$N_S(x)$是锥，且是闭凸集. 同时若$x\notin S$. 定义$N_S(x)=\emptyset$

<u>Example 3.5</u> 设$S\subseteq \mathbb{E}$, 则对$\forall x\in S$

$$y\in \partial\delta_S(x) \Leftrightarrow \delta_S(z)\geq \delta_S(x)+\langle y,z-x\rangle \quad \forall z\in S$$

$$\Leftrightarrow \langle y,z-x\rangle\leq 0 \quad \forall z\in S$$

故 $\partial\delta_S(x)=N_S(x) \quad \forall x\in S$

$x\notin S$时, 同样$\partial\delta_S(x)=N_S(x)=\emptyset$

<u>Example 3.6</u> 考虑$S=B[0,1]=\{x\in \mathbb{E}:\|x\|\leq 1\}$

故 $\partial\delta_S(x)=N_S(x)=\{y\in \mathbb{E}^*:\langle y,z-x\rangle\leq 0 \ \forall z\in S\}$

若 $x \notin S$（即 $\|x\|>1$），则 $N_S(x)=\phi$

若 $x \in S$，则 $y \in N_S(x) \Leftrightarrow$

$\langle y, z-x \rangle \leq 0$ 对 $\forall \|z\| \leq 1$ 成立

即 $\max\limits_{\|z\| \leq 1} \langle y, z \rangle \leq \langle y, x \rangle \Leftrightarrow \|y\|_* \leq \langle y, x \rangle$

故

$$\partial \delta_{B[0,1]}(x) = N_{B[0,1]}(x) = \begin{cases} \{y \in \mathbb{E}^* : \|y\|_* \leq \langle y, x \rangle\}, & \|x\| \leq 1 \\ \phi, & \|x\| > 1 \end{cases}$$

<u>Example 3.7</u> $\min\{f(x) : g(x) \leq 0, x \in X\}$

其中 $\phi \neq X \subseteq \mathbb{E}$，$f : \mathbb{E} \to \mathbb{R}$，$g : \mathbb{E} \to \mathbb{R}^m$

则 dual func: $q(\lambda) = \min\limits_{x \in X} \{L(x;\lambda) = f(x) + \lambda^T g(x)\}$

$(-q)$ 的有效域：$\mathrm{dom}(-q) = \{\lambda \in \mathbb{R}^m_+ : q(\lambda) > -\infty\}$

$\downarrow$

是凸 func

原问题不好解，故转而解 dual problem

$$\max_{\lambda\in\mathbb{R}^m}\{q(\lambda):\lambda\in\operatorname{dom}(-q)\}$$

由 $\operatorname{dom}(-q)$ 是凸集（由 $\forall\lambda_1,\lambda_2\in\operatorname{dom}(-q)$，

$$-q(\alpha\lambda_1+(1-\alpha)\lambda_2)\le\alpha(-q)(\lambda_1)+(1-\alpha)(-q)(\lambda_2)<\infty)$$

且由 Lagrange 对偶理论，q 是凹 func，故 dual

问题是凸问题

令 $\lambda_0\in\operatorname{dom}(-q)$，且设 $x_0\in\underset{x\in X}{\operatorname{Argmin}}\{f(x)+\lambda_0^Tg(x)\}$

对 $\forall\lambda\in\operatorname{dom}(-q)$：

$$\begin{aligned}q(\lambda)&=\min_{x\in X}\{f(x)+\lambda^Tg(x)\}\\&\le f(x_0)+\lambda^Tg(x_0)\\&=f(x_0)+\lambda_0^Tg(x_0)+(\lambda-\lambda_0)^Tg(x_0)\end{aligned}$$

$$= q(\lambda_0) + g(x_0)^T(\lambda - \lambda_0)$$

故 $-q(\lambda) \geq -q(\lambda_0) + (-g(x_0))^T(\lambda - \lambda_0) \quad \forall \lambda \in \mathrm{dom}(-q)$

$$-g(x_0) \in \partial(-q)(\lambda_0)$$

$\boxtimes$

<u>Example 3.8</u> $f: \mathbb{S}^n \to \mathbb{R} \quad f(X) = \lambda_{max}(X)$

设 $X \in \mathbb{S}^n$，令 v 是关于最大特征值的 eigenvector.

且 $\|v\| = 1$，则下证：$\quad vv^T \in \partial f(X)$

证明：$\forall Y \in \mathbb{S}^n$

$Y = Q^T \Lambda Q$，则 $u^T Y u = p^T \Lambda p$

$p = Qu$，且 $\|p\| = 1$

$$\lambda_{max}(Y) = \max_u \{u^T Y u : \|u\|_2 = 1\}$$

$$\geq v^T Y v$$

$$= v^T X v + v^T (Y - X) v$$

$$= \lambda_{max}(X) \|v\|_2^2 + \mathrm{Tr}(v^T (Y - X) v)$$

$$= \lambda_{max}(X) + \langle vv^T, Y - X \rangle$$

$\boxed{/}$

§3.2 Properties of the Subdifferential Set

Thm 3.9 设 $f: \mathbb{E} \to (-\infty, \infty]$ 是 proper func，则对 $\forall x \in \mathbb{E}$，$\partial f(x)$ 是闭凸集

证明：$\forall x \in \mathbb{E}$

$$\partial f(x) = \bigcap_{y \in \mathbb{E}} H_y$$

$$H_y = \{ g \in \mathbb{E}^* : f(y) \geq f(x) + \langle g, y - x \rangle \}$$

是关于 g 的半平面（闭凸集），故 $\partial f(x)$ 闭凸

$\boxed{/}$

Def 3.10 Proper func $f: \mathbb{E} \to (-\infty, \infty]$ 是 subdifferential 于 $x \in \mathrm{dom} f$，若 $\partial f(x) \neq \emptyset$

Lemma 3.11 $f: \mathbb{E} \to (-\infty, \infty]$ 是 proper func

设 $\operatorname{dom}(f)$ 是凸的，设对 $\forall x \in \operatorname{dom} f$，$\partial f(x)$ 非空.

则 f 是 convex 的

证明：设 $x, y \in \operatorname{dom} f$，$\alpha \in [0,1]$，令 $z_\alpha = \alpha x + (1-\alpha) y$

故 $z_\alpha \in \operatorname{dom} f$，从而 $\exists g \in \partial f(z_\alpha)$，s.t.

$$f(y) \geq f(z_\alpha) + \langle g, y - z_\alpha \rangle = f(z_\alpha) + (1-\alpha)\langle g, y-x \rangle$$

$$f(x) \geq f(z_\alpha) + \langle g, x - z_\alpha \rangle = f(z_\alpha) + \alpha \langle g, y-x \rangle$$

$$\Rightarrow f((1-\alpha)x + \alpha y) = f(z_\alpha) \leq (1-\alpha) f(x) + \alpha f(y)$$

$\boxed{\diagup}$

反过来不成立！如下例：

Example 3.12

$$f(x) = \begin{cases} -\sqrt{x}, & x \geq 0 \\ \infty, & \text{else} \end{cases}$$

则 $f(x)$ 是 convex 的，但在 0 处不次可微：

用反证法：设 $\exists g \in \partial f(0)$

$$\Rightarrow f(y) \geq f(0) + g(y-0) \quad \forall y \geq 0$$

$$\Longleftrightarrow -\sqrt{y} \geq gy \quad \forall y \geq 0$$

取 $y=1$ 知：$g \leq -1$

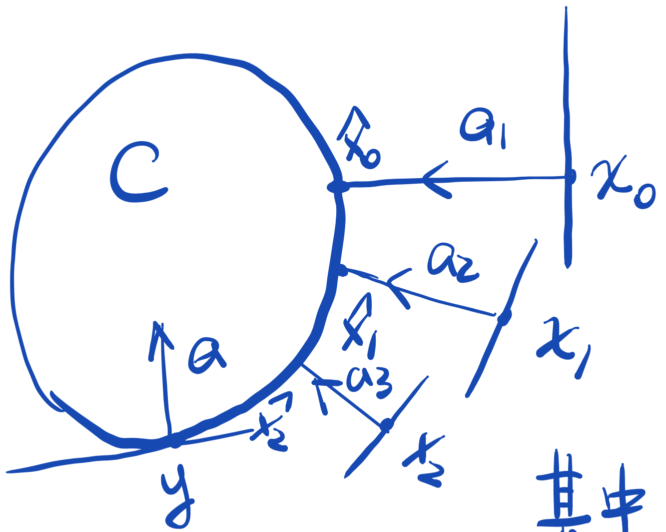
取 $y=\frac{1}{2g^2}$ 知：$-\sqrt{\frac{1}{2g^2}} \geq \frac{1}{2g} \Rightarrow \frac{1}{2g^2} \leq \frac{1}{4g^2}$ 矛盾！

$\square$

Thm 3.13 令 $\emptyset \neq C \subseteq \mathbb{E}$ 是凸集，$y \notin \mathrm{int}(C)$

则 $\exists\, 0 \neq p \in \mathbb{E}^*$, s.t. $\langle p, x\rangle \leq \langle p, y\rangle \ \forall x \in C$

证明：



取序列 $\{x_k\}$, s.t. $x_k \notin \bar{C}$ 且 $x_k \to y \in \partial\bar{C}$，令

$\hat{x}_k = \Pi_{\bar{C}}(x_k)$，有 $\forall k$

$a_k^T x \geq a_k^T x_k$, $\forall x \in \bar{C}$

其中 $a_k = \dfrac{\hat{x}_k - x_k}{\|\hat{x}_k - x_k\|}$, WLOG

设 $a_k \to a$，取 $a^T x \geq a^T y$, $\forall x \in \bar{C}$ $\boxed{/}$

注：严格说，要证 $\partial\bar{C} = \partial C$，此命题仅对C是convex时成立，Nesterov书中 Thm 3.1.14 处讨论过

Thm 3.14 令 $f:\mathbb{E}\to(-\infty,\infty]$ 是 proper convex 的

设 $\tilde{x}\in \operatorname{int}(\operatorname{dom}(f))$，则 $\partial f(\tilde{x})$ 是非空有界的

证明：在 $\mathbb{E}\times\mathbb{R}$ 空间定义内积

$$\left\langle (y_1,\beta_1),(y_2,\beta_2)\right\rangle \equiv \langle y_1,y_2\rangle+\beta_1\beta_2$$

由 $(\tilde{x}, f(\tilde{x}))$ 在 $\operatorname{epi}(f)$ 的边界上，故由 Thm 3.13

$\exists\,(p,-\alpha)\in\mathbb{E}^*\times\mathbb{R}$, s.t.

$$\langle p,\tilde{x}\rangle-\alpha f(\tilde{x})\geqslant\langle p,x\rangle-\alpha t\quad\forall (x,t)\in \operatorname{epi}(f)\ ①$$

由 $(\tilde{x}, f(\tilde{x})+1)\in\operatorname{epi}(f)$，故代入上式知

$$\langle p,\tilde{x}\rangle-\alpha f(\tilde{x})\geqslant\langle p,\tilde{x}\rangle-\alpha(f(\tilde{x})+1)\iff \alpha\geqslant 0$$

由 $\tilde{x}\in\operatorname{int}(\operatorname{dom}f)$，由 Thm 2.21，$\exists\,\varepsilon>0, L>0$

s.t. $B_{\|\cdot\|}[\tilde{x},\varepsilon]\subseteq \operatorname{dom}f$，且

$|f(x)-f(\tilde{x})| \leqslant L\|x-\tilde{x}\|, \forall x \in B_{\|\cdot\|}[\tilde{x},\varepsilon]$

由 $(x, f(x)) \in \mathrm{epi}(f)$ 对 $\forall x \in B_{\|\cdot\|}[\tilde{x},\varepsilon]$ 成立

故 $\langle p, x-\tilde{x}\rangle \leqslant \alpha(f(x)-f(\tilde{x})) \leqslant \alpha L\|x-\tilde{x}\|$ ⊛

取 $p^{\dagger} \in \mathbb{E}$, s.t. $\langle p, p^{\dagger}\rangle = \|p\|_*$, $\|p^{\dagger}\|=1$. 由

$\tilde{x}+\varepsilon p^{\dagger} \in B_{\|\cdot\|}[\tilde{x},\varepsilon]$, 故将 $x=\tilde{x}+\varepsilon p^{\dagger}$ 代入⊛

$$\varepsilon\|p\|_* = \varepsilon\langle p, p^{\dagger}\rangle \leqslant \alpha L\varepsilon\|p^{\dagger}\| = \alpha L\varepsilon$$

故 $\alpha>0$, 否则 $p=0$, $\alpha=0$, 与 Thm3.13 矛盾!

将 $t=f(x)$ 代入①式:

$$f(x) \geqslant f(\tilde{x}) + \langle g, x-\tilde{x}\rangle \text{ 对 } \forall x \in \mathrm{dom} f$$

其中 $g=\frac{p}{\alpha}$, 故 $g \in \partial f(\tilde{x})$, 下证 $\partial f(\tilde{x})$ 有界

取 $g^{\dagger} \in \mathbb{E}$, s.t. $\|g\|_* = \langle g, g^{\dagger}\rangle$, $\|g^{\dagger}\|=1$, 将

$x = \tilde{x} + \varepsilon g^{+} \in \mathrm{dom} f$ 代入上式

$$\varepsilon \|g\|_* = \varepsilon \langle g, g^{+} \rangle = \langle g, x - \tilde{x} \rangle \leq f(x) - f(\tilde{x})$$

$$\leq L \|x - \tilde{x}\| = L\varepsilon$$

故 $\|g\|_* \leq L$，有界性得证 $\boxed{/}$

注：Thm 3.14 证明了：

$$\mathrm{int}(\mathrm{dom} f) \subseteq \mathrm{dom}(\partial f)$$

Corollary 3.15 $f: \mathbb{E} \to \mathbb{R}$ 是 convex 的，则 f 在 $\mathbb{E}$ 上次可微

Thm 3.16 $f: \mathbb{E} \to (-\infty, \infty]$ 是 proper convex 的，设 $X \subseteq \mathrm{int}(\mathrm{dom} f)$ 是非空紧集，则 $Y = \bigcup_{x \in X} \partial f(x)$ 非空且有界

证明：Y显然非空，下证Y有界，用反证法：

设$\exists\{x_k\}_{k\geq 1}\subseteq X$，$g_k\in\partial f(x_k)$，s.t.

$\|g_k\|_*\to\infty\ (k\to\infty)$，对$\forall k$，令$g_k^\dagger\in \mathbb{E}$，s.t.

$\langle g_k, g_k^\dagger\rangle=\|g_k\|_*$，且$\|g_k^\dagger\|=1$，由$X$紧，$(\mathrm{int}(\mathrm{dom}f))^c$

闭，且$X\cap(\mathrm{int}(\mathrm{dom}f))^c=\phi$，故$\exists\varepsilon>0$，s.t.

$$\|x-y\|\geq\varepsilon,\ 对\ \forall x\in X,\ y\notin \mathrm{int}(\mathrm{dom}f)$$

$X \ni \qquad (\mathrm{int}\,\mathrm{dom}f)^c \ni$

上式实际显然，若不存在这样的ε，则$\exists\{x_k\}$，$\{y_k\}$

s.t. $\|x_k-y_k\|\to 0$，WLOG，设$x_k\to\bar{x}$，则$y_k\to\bar{x}$

且由闭性$\bar{x}\in X$，$\bar{x}\in(\mathrm{int}(\mathrm{dom}f))^c$，矛盾！

$g_k\in\partial f(x_k)\Rightarrow$

$$f(\underbrace{x_k+\tfrac{\varepsilon}{2}g_k^\dagger}_{\in\ \mathrm{int}(\mathrm{dom}f)})-f(x_k)\geq\frac{\varepsilon}{2}\langle g_k, g_k^\dagger\rangle=\frac{\varepsilon}{2}\|g_k\|_*$$

下证左边有界，否则 $\exists \{x_k\}_{k\in T}\ \{g_k^+\}_{k\in T}$，s.t.

$$f(x_k+\tfrac{\varepsilon}{2}g_k^+)-f(x_k)\to\infty \quad k\xrightarrow{T}\infty$$

WLOG，设 $x_k\to\bar{x}$，$g_k^+\to\bar{g}$，由 Thm2.21（凸func 在内点连续），知

$$f(x_k+\tfrac{\varepsilon}{2}g_k^+)-f(x_k)\to f(\bar{x}+\tfrac{\varepsilon}{2}\bar{g})-f(\bar{x})$$

故矛盾，从而 $\|g_k\|_*$ 有界，与假设矛盾！ $\boxed{/}$

注：设 $A\subseteq\mathbb{R}^n$，$f: A\to\mathbb{R}^m$

（Rockfeller，称为 clam P351）

称 f locally Lipschiz 连续 in $c\in A$，若

$$\exists\,\delta_c>0,\ \exists L_c\leqslant 0,\ \forall x\in A,$$

$$\|x-c\|\leqslant\delta_c\Rightarrow\|f(x)-f(c)\|\leqslant L_c\|x-c\|$$

称 f locally Lipschiz around $c\in A$，若

$\exists \delta_c > 0, \exists L_c \geq 0, \forall x, y \in B(c, \delta_c)$，有

$$\|f(x)-f(y)\| \leq L_c \|x-y\|$$

① 二者不等价，(2) ⇒ (1) 但 (1) ⇏ (2)

$$f(x) = \begin{cases} x^2 \sin\frac{1}{x^2}, & x \neq 0 \\ 0, & x = 0 \end{cases}$$

$$f'(x) = 2x\sin\frac{1}{x^2} - \frac{2}{x}\cos\frac{1}{x^2} \quad x \neq 0$$

$$f'(0) = \lim_{h\to 0} \frac{h^2 \sin\frac{1}{h^2} - 0}{h} = \lim_{h\to 0} h\sin\frac{1}{h^2} = 0$$

则 $f(x)$ 在 0 处 diam 但不 locally Lipschiz around 0

<u>Thm</u> $f: \mathbb{E} \to (-\infty, \infty]$ convex, $x_0 \in \mathrm{int}(\mathrm{dom} f)$

则 $\exists \varepsilon > 0, L > 0, B[x_0, \varepsilon] \subseteq C$, s.t.

$$|f(x) - f(y)| \leq L\|x-y\| \quad \forall x, y \in B[x_0, \varepsilon]$$

证明：f 在 x_0 处是 locally bounded (Nesterov 3.1.11)

故 $\exists\ \varepsilon>0$, $M\in\mathbb{R}$, s.t. $f(x)\le M$ 对 $\forall x\in B_{2\varepsilon}(x_0)$

取 $x_1, x_2\in B_\varepsilon(x_0)$, 令 $x_3=x_2+\frac{\varepsilon}{\alpha}(x_2-x_1)$,

$\alpha=\|x_2-x_1\|$, 则 $x_3\in N_{2\varepsilon}(x_0)$, 由

$$x_2=\frac{\varepsilon}{\alpha+\varepsilon}x_1+\frac{\alpha}{\alpha+\varepsilon}x_3$$

则 $f(x_2)\le\frac{\varepsilon}{\alpha+\varepsilon}f(x_1)+\frac{\alpha}{\alpha+\varepsilon}f(x_3)$

$$\Rightarrow f(x_2)-f(x_1)\le\frac{\alpha}{\alpha+\varepsilon}\left(f(x_3)-f(x_1)\right)$$

$$\le\frac{\alpha}{\varepsilon}|f(x_3)-f(x_1)|\le\frac{2M}{\varepsilon}\|x_2-x_1\|$$

由 x_1, x_2 的对称性

$|f(x_2)-f(x_1)|\le\frac{2M}{\varepsilon}\|x_2-x_1\|$, 故

f 在 $B_\varepsilon(x_0)$ Lipschiz

$\square$

around意义

注：设 $f: X \to \mathbb{R}^n$ 的 locally Lipschiz func，$C \subseteq X$ 紧.

则 f 在 C 上是 Lipschiz 的：

证明：用反证法，设 $\exists\, x_n, y_n \in C$, s.t.

$$\frac{\|f(x_n)-f(y_n)\|}{\|x_n-y_n\|} \to \infty$$

由 f 在 C 上有界，故 $\|x_n-y_n\| \to 0$，不妨设 $x_n \to \bar{x}$

而 f 在 $\bar{x}$ 处局部 Lipschiz 连续，故矛盾！

紧集上的局部Lipschiz func是Lipschiz的！

有了上面的讨论，Thm3.16 是显然的！

$\boxed{/\!/}$

$$ri(S)=\{x\in aff(S): B[x,\varepsilon]\cap aff(S)\subseteq S, \exists \varepsilon>0\}$$

相对内部就是相对拓扑意义下的开集！

Thm 3.17 设 $C\subseteq\mathbb{E}$ 是非空凸集，则 $ri(C)\neq\emptyset$

证明：WLOG，设 $0\in C$，则设 $\{d_i\}_{i=1}^{d}$ 是线性子空间 $aff(C)=span(C)$ 的一组基，$d_i\in C$.

设 $A:\mathbb{R}^d\to Aff(C)$ $A(\lambda)=\sum_{i=1}^{d}\lambda_i d_i$

则知 A 是线性同胚，取 $\Omega=\{\lambda\in\mathbb{R}^d_{++}:\sum_{i=1}^{d}\lambda_i<1\}$

则 $A\Omega\subseteq C\subseteq aff(C)$，由 Ω 是开集（$\mathbb{R}^d$ 下的），故由 A 是开映射（线性同胚知 A^{-1} 连续），故 $A\Omega$ 是关于 $aff(C)$ 的开集。又由 $A\Omega\subseteq C$，故 Ω 是 C 的相对内部！ $\square$

Thm 3.18 设 $f: \mathbb{E} \to (-\infty, \infty]$ 是 proper closed func. 且令 $\bar{x} \in \mathrm{ri}(\mathrm{dom} f)$, 则 $\partial f(\bar{x}) \neq \emptyset$

证明: 令 $\tilde{f} = f\big|_{\mathrm{Aff}(\mathrm{dom} f)}$, 对 $\tilde{f}$ 用 **Thm 3.14**

$\boxtimes$

注: $\mathrm{ri}(\mathrm{dom} f) \subseteq \mathrm{dom}(\partial f)$

Corollary 3.19 设 $f: \mathbb{E} \to (-\infty, \infty]$ 是 proper convex func, 则 $\exists x \in \mathrm{dom} f$, s.t. $\partial f(x) \neq \emptyset$

Thm 3.20 设 $f: \mathbb{E} \to (-\infty, \infty]$ 是 proper convex func

设 $\dim(\mathrm{dom} f) < \dim \mathbb{E}$, 设 $x \in \mathrm{dom} f$, 则 $\partial f(x)$ 无界

aff (domf) 的维度

证明：设 η 是 $\partial f(x)$ 任意向量，$\mathbb{V} \equiv \text{aff}(\text{dom} f) - \{x\}$ 是子空间，则 $\dim(\mathbb{V}) < \dim(\mathbb{E})$，则 $\exists v \in \mathbb{V}^{\perp}$

s.t. 对 $\forall \beta \in \mathbb{R}$，$y \in \text{dom} f$，有

$$f(y) \geq f(x) + \langle \eta, y-x \rangle = f(x) + \langle \eta + \beta v, y-x \rangle$$

故 $\eta + \beta v \in \partial f(x)$，从而 $\partial f(x)$ 无界 $\boxed{\angle}$

§ 3.3 方向导数

§ 3.3.1 Definition and Basic Properties

$$f'(x;d) \equiv \lim_{\alpha \to 0^+} \frac{f(x+\alpha d) - f(x)}{\alpha}$$

Thm3.21 凸func f 对任意定义域中的内点关于任意方向导数存在

证明：令 $x\in \operatorname{int}(\operatorname{dom} f)$，考虑

$$\phi(\alpha)=\frac{1}{\alpha}[f(x+\alpha p)-f(x)],\ \alpha>0$$

取 ε 足够小，s.t. $x+\varepsilon p\in \operatorname{dom} f$，令 $\beta\in(0,1]$，$\alpha\in(0,\varepsilon]$

$$f(x+\alpha\beta p)=f((1-\beta)x+\beta(x+\alpha p))$$
$$\leq(1-\beta)f(x)+\beta f(x+\alpha p)$$

故 $$\phi(\alpha\beta)=\frac{1}{\alpha\beta}[f(x+\alpha\beta p)-f(x)]$$
$$\leq\frac{1}{\alpha}[f(x+\alpha p)-f(x)]=\phi(\alpha)$$

故 $\alpha\downarrow 0$ 时，$\phi(\alpha)\downarrow$，取 $\gamma>0$ 足够小，s.t. $x-\gamma p\in \operatorname{dom} f$，

则 $x+\alpha p=x+\frac{\alpha}{\gamma}(x-(x-\gamma p))$，故由 (3.1.5)

$$f(x+\alpha p)\geq f(x)+\frac{\alpha}{\gamma}(f(x)-f(x-\gamma p))$$

$$\Rightarrow\ \phi(\alpha)\geq\frac{1}{\gamma}(f(x)-f(x-\gamma p))\quad\forall\alpha>0$$

由单调收敛Thm，极限∃

$\boxed{2}$

Lemma 3.22 $x \in \operatorname{int}(\operatorname{dom} f)$，则

(a) $d \longmapsto f'(x;d)$ 是 convex 的

(b) $\forall \lambda \geq 0, d \in \mathbb{E}$，则 $f'(x;\lambda d) = \lambda f'(x;d)$

证明：

(a) 任取 $d_1, d_2 \in \mathbb{E}$, $\lambda \in [0,1]$，则

$$f'(x;\lambda d_1 + (1-\lambda) d_2) = \lim_{\alpha \to 0^+} \frac{f(x+\alpha(\lambda d_1 + (1-\lambda) d_2)) - f(x)}{\alpha}$$

$$= \lim_{\alpha \to 0^+} \frac{f(\lambda(x+\alpha d_1) + (1-\lambda)(x+\alpha d_2)) - f(x)}{\alpha}$$

$$\leq \lim_{\alpha \to 0^+} \frac{\lambda f(x+\alpha d_1) + (1-\lambda) f(x+\alpha d_2) - f(x)}{\alpha}$$

$$= \lambda \lim_{\alpha \to 0^+} \frac{f(x+\alpha d_1) - f(x)}{\alpha} + (1-\lambda) \lim_{\alpha \to 0^+} \frac{f(x+\alpha d_2) - f(x)}{\alpha}$$

$= \lambda f'(x;d_1) + (1-\lambda) f'(x;d_2)$

(b) $\lambda = 0$ 时，显然成立，$\lambda > 0$ 时

$$f'(x;\lambda d) = \lim_{\alpha \to 0^+} \frac{f(x+\alpha\lambda d) - f(x)}{\alpha}$$

$$= \lambda \lim_{\alpha \to 0^+} \frac{f(x+\alpha\lambda d) - f(x)}{\alpha\lambda}$$

$$= \lambda f'(x;d) \qquad \square$$

Lemma 3.23　$x \in \mathrm{int}(\mathrm{dom} f)$，则

$$f(y) \geq f(x) + f'(x; y-x), \ \forall y \in \mathrm{dom} f$$

证明：$f'(x; y-x) = \lim_{\alpha \to 0^+} \frac{f(x+\alpha(y-x)) - f(x)}{\alpha}$

$$= \lim_{\alpha \to 0^+} \frac{f((1-\alpha)x + \alpha y) - f(x)}{\alpha}$$

$$\leq \lim_{\alpha\to 0^+} \frac{-\alpha f(x)+\alpha f(y)}{\alpha} = f(y)-f(x)$$

$\square$

<u>**Thm 3.24**</u> 设 $f(x)=\max\{f_1(x),\cdots,f_m(x)\}$

$f_1,\cdots,f_m: E\to(-\infty,\infty]$ 是 proper func，令

$x\in\bigcap_{i=1}^{m}\text{int}(\text{dom} f_i)$，$d\in E$，设 $f_i'(x;d)\ \exists$，$\forall i$

则 $f'(x;d)=\max\limits_{i\in I(x)} f_i'(x;d)$，

$I(x)=\{i: f_i(x)=f(x)\}$.

证明：对 $\forall i\in\{1,2,\cdots,m\}$，

$$\lim_{t\to 0^+} f_i(x+td)=\lim_{t\to 0^+}\left[t\cdot\frac{f_i(x+td)-f_i(x)}{t}+f_i(x)\right]$$

$$=0\cdot f_i'(x;d)+f_i(x)=f_i(x)$$

由 $I(x)$ 的定义：$f_i(x)>f_j(x)$ 对 $\forall i\in I(x)$，$j\notin I(x)$

故 $\exists \varepsilon>0$, s.t. $f_i(x+td)>f_j(x+td)$ 对 $\forall$ $i\in I(x)$, $j\notin I(x)$, $t\in(0,\varepsilon]$ 成立. 故对 $\forall t\in(0,\varepsilon]$

$$f(x+td)=\max_{i=1,\cdots,m} f_i(x+td)$$

$$=\max_{i\in I(x)} f_i(x+td)$$

故对 $\forall t\in(0,\varepsilon]$, 有

$$\frac{f(x+td)-f(x)}{t}=\frac{\max\limits_{i\in I(x)} f_i(x+td)-f(x)}{t}$$

$$=\max_{i\in I(x)}\frac{f_i(x+td)-f_i(x)}{t}$$

$$f'(x;d)=\lim_{t\to 0^+}\frac{f(x+td)-f(x)}{t}$$

$$=\lim_{t\to 0^+}\max_{i\in I(x)}\frac{f_i(x+td)-f_i(x)}{t}$$

$$= \max_{i \in I(x)} \lim_{t \to 0^+} \frac{f_i(x+td)-f_i(x)}{t}$$

$$= \max_{i \in I(x)} f_i'(x;d)$$

可交换：由 $\max(a,b) = \frac{a+b+|b-a|}{2}$

故 $\lim\limits_{t \to 0^+} \max\left(\frac{f_1(x+td)-f_1(x)}{t}, \cdots, \frac{f_k(x+td)-f_k(x)}{t} \right)$

$$= \max\left(\lim_{t \to 0^+} \frac{f_1(x+td)-f_1(x)}{t}, \cdots, \lim_{t \to 0^+} \frac{f_k(x+td)-f_k(x)}{t} \right)$$

Corollary 3.25 设 $f(x) = \max\{f_1(x), \cdots, f_m(x)\}$

$f_1, \cdots, f_m : \mathbb{E} \to (-\infty, \infty]$ 是 proper convex，令

$x \in \bigcap_{i=1}^{m} \operatorname{int}(\operatorname{dom} f_i)$，$d \in \mathbb{E}$，则 $f'(x;d) = \max\limits_{i \in I(x)} f_i'(x;d)$

§ 3.3.2 The Max Formula

Thm 3.26 (max formula) $f: \mathbb{E} \to (-\infty, \infty]$ 是 proper convex func，则对 $\forall x \in \operatorname{int}(\operatorname{dom} f)$，$d \in \mathbb{E}$

$$f'(x;d) = \max\{\langle g, d\rangle : g \in \partial f(x)\}$$

证明：令 $x \in \operatorname{int}(\operatorname{dom} f)$，$d \in \mathbb{E}$，则对 $\forall g \in \partial f(x)$

$$f'(x;d) = \lim_{\alpha \to 0^+} \frac{1}{\alpha}(f(x+\alpha d) - f(x))$$
$$\geq \lim_{\alpha \to 0^+} \langle g, d\rangle = \langle g, d\rangle$$

故 $f'(x;d) \geq \max\{\langle g, d\rangle : g \in \partial f(x)\}$

下证相反的方向：令 $h(w) \equiv f'(x;w)$

故由 lemma 3.22(a) 知，$h(w)$ 在 $\mathbb{E}$ 上次可微

令 $\tilde{g} \in \partial h(d)$，则对 $\forall v \in E, \alpha \geq 0$

$$\alpha f'(x;v) = f'(x;\alpha v) = h(\alpha v)$$

$$\geq h(d) + \langle \tilde{g}, \alpha v - d \rangle = f'(x;d) + \langle \tilde{g}, \alpha v - d \rangle$$

故 $\alpha (f'(x;v) - \langle \tilde{g}, v \rangle) \geq f'(x;d) - \langle \tilde{g}, d \rangle$ ⊛

α 的任意性知：

$$f'(x;v) \geq \langle \tilde{g}, v \rangle$$

故 $f(y) \geq f(x) + f'(x;y-x) \geq f(x) + \langle \tilde{g}, y-x \rangle$

故 $\tilde{g} \in \partial f(x)$，对⊛取 $\alpha = 0$ 知

$$f'(x;d) \leq \langle \tilde{g}, d \rangle \leq \max \{ \langle g, d \rangle : g \in \partial f(x) \}$$

☑

§ 3.3.3 Differentiability

Def 3.28 设 $f:\mathbb{E}\to(-\infty,\infty]$, $x\in\operatorname{int}(\operatorname{dom} f)$.

f 是可微的，若 $\exists g\in\mathbb{E}^*$, s.t.

$$\lim_{h\to 0}\frac{f(x+h)-f(x)-\langle g,h\rangle}{\|h\|}=0$$

这样唯一的 g 称为 f 的 gradient，记作 $\nabla f(x)$

唯一性是显然的

Thm 3.29 设 $f:\mathbb{E}\to(-\infty,\infty]$ 是 proper，且在 $x\in\operatorname{int}(\operatorname{dom} f)$ 处可微，则对 $\forall d\in\mathbb{E}$，有

$$f'(x;d)=\langle\nabla f(x),d\rangle$$

证明：$d=0$ 时显然成立，$d\neq 0$ 时

$$0 = \lim_{\alpha \to 0^+} \frac{f(x+\alpha d) - f(x) - \langle \nabla f(x), \alpha d \rangle}{\|\alpha d\|}$$

$$= \lim_{\alpha \to 0^+} \left[\frac{f(x+\alpha d) - f(x)}{\alpha \|d\|} - \frac{\langle \nabla f(x), d \rangle}{\|d\|} \right]$$

故 $$f'(x;d) = \lim_{\alpha \to 0^+} \frac{f(x+\alpha d) - f(x)}{\alpha}$$

$$= \lim_{\alpha \to 0^+} \left\{ \|d\| \left[\frac{f(x+\alpha d) - f(x)}{\alpha \|d\|} - \frac{\langle \nabla f(x), d \rangle}{\|d\|} \right] + \langle \nabla f(x), d \rangle \right\}$$

$$= \langle \nabla f(x), d \rangle \qquad \square$$

<u>Example 3.30</u> $f(x) = \max\limits_{i=1,\cdots,m} f_i(x)$，其中

f_i proper，设 f_i，$i=1,\cdots,m$ 在 $x \in \bigcap\limits_{i=1}^{m} \operatorname{dom} f_i$ 处可微，

则由 <u>Thm3.24</u>，<u>Thm3.29</u>

$$f'(x;d) = \max_{i \in I(x)} f_i'(x;d) = \max_{i \in I(x)} \langle \nabla f_i(x), d \rangle$$

其中 $I(x) = \{ i : f_i(x) = f(x) \}$ $\boxed{\angle}$

Example 3.31 $C \subseteq \mathbb{E}$ 非空闭凸集，令

$$\varphi(x) \equiv \frac{1}{2} d_C^2(x) = \frac{1}{2} \| x - P_C(x) \|^2$$

Claim: $\nabla \varphi_C(x) = x - P_C(x) \quad \forall x \in \mathbb{E}$

证明：固定 $x \in \mathbb{E}$，令

$$g_x(d) = \varphi_C(x+d) - \varphi_C(x) - \langle d, z_x \rangle$$

其中 $z_x = x - P_C(x)$，只需证：$\boxed{\frac{g_x(d)}{\|d\|} \to 0 \quad (d \to 0)}$ (*)

由 $\| x+d-P_C(x+d) \|^2 \le \| x+d-P_C(x) \|^2$，故 $\forall d \in \mathbb{E}$:

$$g_x(d) = \frac{1}{2}\|x+d-P_C(x+d)\|^2 - \frac{1}{2}\|x-P_C(x)\|^2 - \langle d, z_x\rangle$$

$$\leqslant \frac{1}{2}\|x+d-P_C(x)\|^2 - \frac{1}{2}\|x-P_C(x)\|^2 - \langle d, z_x\rangle$$

$$= \frac{1}{2}\|d\|^2$$

同理： $g_x(-d) \leqslant \frac{1}{2}\|d\|^2$

由 $\varphi_C(x)$ 凸，故 g_x 凸，从而由 Jensen 不等式：

$$0 = g_x(0) = g_x\left(\frac{d+(-d)}{2}\right) \leqslant \frac{1}{2}\left(g_x(d) + g_x(-d)\right)$$

$$\Rightarrow \quad g_x(d) \geqslant -g_x(-d) \geqslant -\frac{1}{2}\|d\|^2$$

故 $|g_x(d)| \leqslant \frac{1}{2}\|d\|^2$，即证 ⊛ $\square$

<u>Remark 3.32</u> (what is gradient?)

梯度与内积的选取相关！

设 $\mathbb{E}=\mathbb{R}^n$，标准的内积：

$$\nabla f(x) = D_f(x) \equiv \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{pmatrix}$$

设 $\mathbb{E}=\mathbb{R}^n$，$\langle x, y\rangle = x^T H y$，$H \in S^n_{++}$

注意到方向导数的定义并不依赖内积，故

$$f'(x;d) = D_f(x)^T d = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x) d_i \qquad (3.30)$$

$$\begin{aligned}(\nabla f(x))_i &= \nabla f(x)^T e_i = \nabla f(x)^T H (H^{-1} e_i) \\ &= \langle \nabla f(x), H^{-1} e_i \rangle \\ &= f'(x; H^{-1} e_i) \qquad (3.22) \\ &= D_f(x)^T H^{-1} e_i \qquad (3.30)\end{aligned}$$

此时 $\nabla f(x) = H^{-1} D_f(x)$

<u>Thm 3.33</u> $f: \mathbb{E} \to (-\infty, \infty]$ 是 proper convex,

$x \in \operatorname{int}(\operatorname{dom} f)$，若 f 在 x 处可微，则 $\partial f(x) = \{\nabla f(x)\}$

反之，若 f 在 x 处有唯一次梯度，则在 x 处可微

证明：设 $x \in \operatorname{int}(\operatorname{dom} f)$，且 f 在 x 处可微，

$$f'(x;d) = \langle \nabla f(x), d \rangle \quad \forall d \in \mathbb{E}$$

设 $g \in \partial f(x)$，下证 $g = \nabla f(x)$，由 max formula:

$$\langle \nabla f(x), d \rangle = f'(x;d) \geq \langle g, d \rangle，故$$

$$\langle g - \nabla f(x), d \rangle \leq 0$$

$$\Rightarrow \|g-\nabla f(x)\|_* = \max_{\|d\|\le 1} \langle g-\nabla f(x), d\rangle \le 0$$

$\Rightarrow g=\nabla f(x)$，从而 $\partial f(x)=\{\nabla f(x)\}$

($\Leftarrow$)：设 f 在 $x\in \operatorname{int}(\operatorname{dom} f)$ 处有唯一次梯度 g

令 $h(u)\equiv f(x+u)-f(x)-\langle g,u\rangle$，只需证

$$\lim_{u\to 0}\frac{h(u)}{\|u\|}=0$$

显然，$0\in \operatorname{int}(\operatorname{dom} h)$，且 0 是 h 在 0 处唯一次梯度

故 $\forall d\in \mathbb{E}$，$h'(0;d)=\sigma_{\partial h(0)}(d)=0$

$$0=h'(0;d)=\lim_{\alpha\to 0^+}\frac{h(\alpha d)-h(0)}{\alpha}=\lim_{\alpha\to 0^+}\frac{h(\alpha d)}{\alpha}$$

令 $\{v_1,\cdots,v_k\}$ 是 $\mathbb{E}$ 上一组正交基，由 $0\in \operatorname{int}(\operatorname{dom} h)$

$\exists\, \varepsilon \in (0,1)$, s.t. $\varepsilon v_i, -\varepsilon v_i \in \mathrm{dom}\, h$ 对 $\forall i = 1, \cdots, k$

由 dom h 凸，故 $D = \mathrm{conv}\left(\{\pm \varepsilon v_i\}_{i=1}^{k}\right) \subseteq \mathrm{dom}\, h$

令 $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$，故 $B_{\|\cdot\|}[0, r] \subseteq D$，$r = \frac{\varepsilon}{k}$

令 $w \in B_{\|\cdot\|}[0, r]$，则 $w = \sum_{i=1}^{k} \langle w, v_i \rangle v_i$

$$\|w\|^2 = \sum_{i=1}^{k} \langle w, v_i \rangle^2$$

由 $\|w\|^2 \leq r^2$，故 $|\langle w, v_i \rangle| \leq r$，故

$$w = \sum_{i=1}^{k} \langle w, v_i \rangle v_i$$

$$= \sum_{i=1}^{k} \frac{|\langle w, v_i \rangle|}{\varepsilon} [\mathrm{sgn}(\langle w, v_i \rangle) \varepsilon v_i]$$

$$+ \left(1 - \sum_{i=1}^{k} \frac{|\langle w, v_i \rangle|}{\varepsilon}\right) \cdot 0 \in D$$

"$\in D$" 由于 $0, \pm \varepsilon v_i \in D$，$\sum_{i=1}^{k} \frac{|\langle w, v_i \rangle|}{\varepsilon} \leq \frac{kr}{\varepsilon} = 1$

故 $B_{\|\cdot\|}[0,r] \subseteq D$

记 $\{\pm\varepsilon v_i\}_{i=1}^{k}$ 为 $z_1,\cdots,z_{2k}$，取 $0\neq u\in B_{\|\cdot\|}[0,r^2]$

故 $r\frac{u}{\|u\|}\in B_{\|\cdot\|}[0,r]\subseteq D$，故 $\exists\lambda\in\Delta_{2k}$, s.t.

$$r\frac{u}{\|u\|}=\sum_{i=1}^{2k}\lambda_i z_i$$

故

$$\frac{h(u)}{\|u\|}=\frac{h\left(\frac{\|u\|}{r}r\frac{u}{\|u\|}\right)}{\|u\|}$$

$$=\frac{h\left(\sum_{i=1}^{2k}\lambda_i\frac{\|u\|}{r}z_i\right)}{\|u\|}$$

$$\leq\sum_{i=1}^{2k}\lambda_i\frac{h\left(\|u\|\frac{z_i}{r}\right)}{\|u\|}$$

$$\leq\max_{i=1,\cdots,2k}\left\{\frac{h\left(\|u\|\frac{z_i}{r}\right)}{\|u\|}\right\}$$

又由 $\lim\limits_{u\to 0}\frac{h\left(\|u\|\frac{z_i}{r}\right)}{\|u\|}=\lim\limits_{\|u\|\to 0}\frac{h\left(\|u\|\frac{z_i}{r}\right)}{\|u\|}$

$$= \lim_{\alpha \to 0^+} \frac{h(\alpha \frac{z_i}{r})}{\alpha} = 0$$

$\boxed{/}$

Example 3.34

$f(x) = \|x\|_2$

$$\partial f(x) = \begin{cases} \left\{ \frac{x}{\|x\|_2} \right\}, & x \neq 0 \\ B_{\|\cdot\|_2}[0,1], & x = 0 \end{cases}$$

§ 3.4 Computing Subgradient

§ 3.4.1 Multiplication by a Positive Scalar

<u>Thm 3.35</u> 设 $f: \mathbb{E} \to (-\infty, \infty]$ 是 proper, $\alpha > 0$, 则 $\forall x \in \operatorname{dom} f$, 有 $\partial(\alpha f)(x) = \alpha \partial f(x)$

证明: 利用定义显然. $\boxed{/}$

§ 3.4.2 Summation

<u>Thm 3.36</u> 设 $f_1, f_2: \mathbb{E} \to (-\infty, \infty]$ 是 proper convex

$x \in \operatorname{dom} f_1 \cap \operatorname{dom} f_2$, 则

(a) $\partial f_1(x) + \partial f_2(x) \subseteq \partial(f_1 + f_2)(x)$

(b) 若 $x \in \operatorname{int}(\operatorname{dom} f_1) \cap \operatorname{int}(\operatorname{dom} f_2)$, 则

$$\partial(f_1 + f_2)(x) = \partial f_1(x) + \partial f_2(x)$$

证明

(a) 设 $g \in \partial f_1(x) + \partial f_2(x)$，则 $\exists\, g_1 \in \partial f_1(x),\ g_2 \in \partial f_2(x)$

s.t. $g = g_1 + g_2$，由 g_1, g_2 定义，对 $\forall y \in \mathrm{dom} f_1 \cap \mathrm{dom} f_2$

$$f_1(y) \geqslant f_1(x) + \langle g_1, y-x \rangle$$
$$f_2(y) \geqslant f_2(x) + \langle g_2, y-x \rangle$$

$$\Rightarrow f_1(y) + f_2(y) \geqslant f_1(x) + f_2(x) + \langle g_1 + g_2, y-x \rangle$$

故 $g \in \partial(f_1+f_2)(x)$

(b) 设 $d \in \mathbb{E}$，$f \equiv f_1 + f_2$，由 $x \in \mathrm{int}(\mathrm{dom} f)$

$$\sigma_{\partial f(x)}(d) = \max\{\langle g, d\rangle : g \in \partial f(x)\}$$
$$= f'(x;d)$$
$$= f_1'(x;d) + f_2'(x;d)$$
$$= \max\{\langle g_1, d\rangle : g_1 \in \partial f_1(x)\}$$
$$+ \max\{\langle g_2, d\rangle : g_2 \in \partial f_2(x)\}$$

$$= \max\left\{ \langle g_1+g_2, d\rangle : g_1 \in \partial f_1(x), g_2 \in \partial f_2(x) \right\}$$

$$= \sigma_{\partial f_1(x)+\partial f_2(x)}(d)$$

由Thm 3.9, 3.14，$\partial f(x), \partial f_1(x), \partial f_2(x)$ 非空紧凸集

故 $\partial f_1(x)+\partial f_2(x)$ 是非空紧凸集，由lemma 2.34

$$\partial f(x) = \partial f_1(x) + \partial f_2(x) \qquad \boxtimes$$

Thm 3.40 设 $f_1, \dots, f_m : \mathbb{E} \to (-\infty, \infty]$ 是 proper convex func，设 $\bigcap_{i=1}^m \mathrm{ri}(\mathrm{dom} f_i) \neq \emptyset$，则对 $\forall x \in \mathbb{E}$

$$\partial\left(\sum_{i=1}^m f_i\right)(x) = \sum_{i=1}^m \partial f_i(x)$$

证明：设 $x^* = x_1^* + \dots + x_m^*$，$x_i^* \in \partial f_i(x)$，则 $\forall z$

$$f(z) = f_1(z) + \dots + f_m(z)$$

$$\geq f_1(x)+\langle x_1^*, z-x\rangle+\cdots+f_m(x)+\langle x_m^*, z-x\rangle$$

$$=f(x)+\langle z-x, x^*\rangle$$

故 $x^*\in\partial f(x)$，设 $\bigcap_{i=1}^m ri(\mathrm{dom} f_i)\neq\phi$

Lemma 1 设 $f_1,\cdots,f_m$ 是 proper, convex 于 $\mathbb{R}^n$.

$\bigcap_{i=1}^m ri(\mathrm{dom} f_i)\neq\phi$，则

$$(f_1+\cdots+f_m)^*(x^*)=\inf\{f_1^*(x_1^*)+\cdots+f_m^*(x_m^*) \mid x_1^*+\cdots+x_m^*=x^*\}$$

其中对 $\forall x^*$，inf 可达

Lemma 2：$\forall$ proper, convex f, $\forall x$

$$x^*\in\partial f(x)\Leftrightarrow f(x)+f^*(x^*)=\langle x, x^*\rangle$$

故 $x^*\in\partial f(x)\Leftrightarrow$

$$\langle x, x^* \rangle = f_1(x) + \cdots + f_m(x) + \inf\{ f_1^*(x_1^*) + \cdots + f_m^*(x_m^*) \mid x_1^* + \cdots + x_m^* = x^* \}$$

其中inf是可达的，故 $\forall x^* \in \partial f(x)$ 均可分解成

$x_1^* + \cdots + x_m^* = x^*$, s.t.

$$\langle x, x_1^* \rangle + \cdots + \langle x, x_m^* \rangle = f_1(x) + \cdots + f_m(x) + f_1^*(x_1^*) + \cdots + f_m^*(x_m^*)$$

但 $\langle x, x_i^* \rangle \le f_i(x) + f_i^*(x_i^*)$，等号成立 $\Leftrightarrow x_i^* \in \partial f_i(x)$

故 $\partial f(x) = \partial f_1(x) + \cdots + \partial f_m(x)$ $\square$

<u>Example 3.41</u> $f: \mathbb{R}^n \to \mathbb{R}$, $f(x) = \|x\|_1 = \sum_{i=1}^n |x_i|$

则 $f = \sum_{i=1}^n f_i$, $f_i \equiv |x_i|$

$$\partial f_i(x) = \begin{cases} \{\operatorname{sgn}(x_i) e_i\}, & x_i \neq 0 \\ [-e_i, e_i], & x_i = 0 \end{cases}$$

由 Corollary 3.39

$$\partial f(x) = \sum_{i=1}^{n} \partial f_i(x) = \sum_{i \in I_{\neq}(x)} \operatorname{sgn}(x_i) e_i + \sum_{i \in I_0(x)} [-e_i, e_i]$$

$$I_{\neq}(x) = \{i : x_i \neq 0\}, \quad I_0(x) = \{i : x_i = 0\}$$

故 $$\partial f(x) = \left\{ z \in \mathbb{R}^n : z_i = \operatorname{sgn}(x_i), i \in I_{\neq}(x), |z_j| \leq 1, j \in I_0(x) \right\}$$

注：$\operatorname{sgn}(x) \in \partial f(x)$ ◪

§3.4.3 Affine Transformation

Thm 3.43 $f: \mathbb{E} \to (-\infty, \infty]$ proper, convex

$A: \mathbb{V} \to \mathbb{E}$ linear，令 $h(x) = f(A(x) + b)$，h proper

即 $\operatorname{dom} h = \{x \in \mathbb{V} : A(x) + b \in \operatorname{dom} f\} \neq \emptyset$

(a) $\forall x \in \operatorname{dom} h \quad A^T(\partial f(A(x)+b)) \subseteq \partial h(x)$

(b) 若 $x \in \operatorname{int}(\operatorname{dom} h)$，$A(x)+b \in \operatorname{int}(\operatorname{dom} f)$

$$\partial h(x) = A^T(\partial f(A(x)+b))$$

证明

(a) $x \in \operatorname{dom} h$，$g \in A^T(\partial f(A(x)+b))$，则 $\exists d \in \mathbb{E}^*$

s.t. $g = A^T(d)$，其中 $d \in \partial f(A(x)+b)$

对 $\forall y \in \operatorname{dom} h$，有 $A(y)+b \in \operatorname{dom} f$，故

$$f(A(y)+b) \geq f(A(x)+b) + \langle d, A(y-x) \rangle$$

$$\Leftrightarrow \quad h(y) \geq h(x) + \langle A^T(d), y-x \rangle \quad \forall y \in \operatorname{dom} h$$

故 $g = A^T(d) \in \partial h(x)$，从而 (a) 得证 $\square$

(b) 由 $x \in \operatorname{int}(\operatorname{dom} h)$，故 $h'(x;d) = \sigma_{\partial h(x)}(d) \quad \forall d$

又由 $\quad h'(x;d) = \lim\limits_{\alpha \to 0^+} \dfrac{h(x+\alpha d) - h(x)}{\alpha}$

$$= \lim_{\alpha \to 0^+} \frac{f(A(x)+b+\alpha A(d)) - f(A(x)+b)}{\alpha}$$

$$= f'(A(x)+b; \quad A(d))$$

故 $\sigma_{\partial h(x)}(d) = f'(A(x)+b; \quad A(d))$

$$= \max_g \left\{ \langle g, A(d) \rangle : g \in \partial f(A(x)+b) \right\}$$

$$= \max_g \left\{ \langle A^T(g), d \rangle : g \in \partial f(A(x)+b) \right\}$$

$$= \max_{\tilde{g}} \left\{ \langle \tilde{g}, d \rangle : \tilde{g} \in A^T \partial f(A(x)+b) \right\}$$

$$= \sigma_{A^T(\partial f(A(x)+b))}(d)$$

非空紧凸集

故由 lemma 2.34 $\partial h(x) = A^T(\partial f(A(x)+b))$ $\boxtimes$

Example 3.44

$$f(x) = \|Ax+b\|_1$$

$$\doteq g(Ax+b)$$

故 $\partial f(x) = A^T \partial g(Ax+b)$

$$= \sum_{i \in I_{\neq}(x)} \operatorname{sgn}(a_i^T x + b_i) A^T e_i + \sum_{i \in I_0(x)} [-A^T e_i, A^T e_i]$$

$$= \sum_{i \in I_{\neq}(x)} \operatorname{sgn}(a_i^T x + b_i) a_i + \sum_{i \in I_0(x)} [-a_i, a_i]$$

Example 3.45

$$f(x) = \|Ax+b\|_2$$

则 $\partial f(x) = A^T \partial g(Ax+b)$

$$= \begin{cases} \dfrac{A^T(Ax+b)}{\|Ax+b\|_2}, & Ax+b \neq 0 \\ A^T B_{\|\cdot\|_2}[0,1], & Ax+b = 0 \end{cases}$$

注：对 x，满足 $Ax+b=0$，则

$$\partial f(x) = A^T B_{\|\cdot\|_2}[0,1] = \{A^T y : \|y\|_2 \le 1\}$$

则此时 $0 \in \partial f(x)$

§ 3.4.4 Composition

<u>Thm 3.46</u> 设 f 在 $[a,b]$ 连续，且 $f'_+(a)$ 存在

设 g 是定义在开区间 I，且 $I \supseteq \mathrm{Range}(f)$，设 g 在 $f(a)$ 处可微，则 $h(t) = g(f(t))$ $(a \le t \le b)$ 是于 $t=a$ 处右可微，且 $h'_+(a) = g'(f(a)) f'_+(a)$

证明：

$$h'_+(a) = \lim_{t \to a^+} \frac{g(f(t)) - g(f(a))}{t-a}$$

$$= \lim_{t \to a^+} \frac{g(f(t)) - g(f(a))}{f(t) - f(a)} \cdot \frac{f(t) - f(a)}{t-a}$$

$$= g'(f(a)) f'_+(a)$$

$\square$

<u>Thm 3.47</u> 设 $f:\mathbb{E}\to\mathbb{R}$ convex, $g:\mathbb{R}\to\mathbb{R}$

是非降凸 func，设 $x\in\mathbb{E}$，且 g 在 $f(x)$ 可微

设 $h=g\circ f$，则 $\partial h(x)=g'(f(x))\partial f(x)$

证明：$\forall d\in\mathbb{E}$，定义一维 func：

$$f_{x,d}(t)=f(x+td),\ t\in\mathbb{R}$$

$$h_{x,d}(t)=h(x+td),\ t\in\mathbb{R}$$

则 $h_{x,d}(t)=g(f_{x,d}(t))$，$\forall t\in\mathbb{R}$

由 f 凸，g 是非减凸，故 $h=f\circ g$ 是 convex，由 <u>Thm 3.21</u>

对任意方向，f、h 的方向导数存在，由定义：

$$(f_{x,d})'_+(0)=f'(x;d)$$

$$(h_{x,d})'_+(0)=h'(x;d)$$

由 $h_{x,d}=g\circ f_{x,d}$，$f_{x,d}$ 在 0 处右可微，g 在 $f_{x,d}(0)=$

$f(x)$处可微，故由 Thm 3.46

$$(h_{x,d})'_+(0) = g'(f(x))(f_{x,d})'_+(0)$$

故 $$h'(x;d) = g'(f(x))\, f'(x;d)$$

由 $x \in \operatorname{int}(\operatorname{dom} f) = \operatorname{int}(\operatorname{dom} h) = \mathbb{E}$

$$h'(x;d) = \sigma_{\partial h(x)}(d)\ ;\quad f'(x;d) = \sigma_{\partial f(x)}(d)$$

故 $$\begin{aligned}\sigma_{\partial h(x)}(d) &= h'(x;d)\\ &= g'(f(x))\, f'(x;d)\\ &= g'(f(x))\, \sigma_{\partial f(x)}(d)\\ &\overset{\text{lemma 2.24(c)}}{=} \sigma_{g'(f(x))\partial f(x)}(d)\end{aligned}$$

由 $\partial h(x)$，$\partial f(x)$ 非空闭凸，故

$$\partial h(x) = g'(f(x))\, \partial f(x) \qquad \boxed{/\!/}$$

Example 3.48 $\quad h(x) = \|x\|_1^2$

令 $f(x) = \|x\|_1$，$g(t) = [t]_+^2 = \max\{t, 0\}$

故 f, g 是 convex 的，g 是非减的，且 g 在 $\mathbb{R}$ 上可微，$g'(t) = 2[t]_+$

$$\partial h(x) = g'(f(x))\partial f(x) = 2[\|x\|_1]_+ \partial f(x)$$

$$= 2\|x\|_1 \partial f(x)$$

$$= 2\|x\|_1 \left\{ z \in \mathbb{R}^n : z_i = \operatorname{sgn}(x_i),\ i \in I_{\neq}(x),\ |z_j| \le 1,\ j \in I_0(x) \right\}$$

Example 3.49 设 $C \subseteq \mathbb{E}$ 非空闭凸集

$$d_C(x) = \min_y \{ \|x-y\|,\ y \in C \}$$

则
$$\partial d_C(x) = \begin{cases} \left\{ \dfrac{x - P_C(x)}{d_C(x)} \right\}, & x \notin C \\ N_C(x) \cap B[0,1], & x \in C \end{cases}$$

证明：由 Example 3.31 $\varphi_C(x) = \frac{1}{2} d_C^2(x)$ 可微

$$\partial \varphi_C(x) = \{ x - P_C(x) \} \quad \forall x \in \mathbb{E}$$

由 $\varphi_c = g\circ d_c$，其中 $g(t)=\frac{1}{2}[t]_+^2$

$$\partial\varphi_c(x) = g'(d_c(x))\,\partial d_c(x) \;=\; d_c(x)\,\partial d_c(x)$$

- 设 $x\notin C$，则 $d_c(x)\neq 0$ 故

$$\partial d_c(x) = \left\{\frac{x-P_c(x)}{d_c(x)}\right\},\ \forall x\notin C$$

此时，$d_c(x)$ 可微

- 设 $x\in C$，下证 $\partial d_c(x) = N_c(x)\cap B[0,1]$

若 $d\in\partial d_c(x)$，则 $\quad d_c(y)\geqslant\langle d, y-x\rangle,\ \forall y\in E$

故对 $\forall y\in C$，$\langle d, y-x\rangle\leqslant 0 \iff d\in N_c(x)$

取 $y=x+d$，有

$$\|d\|^2 = \langle d, x+d-x\rangle \leqslant d_c(x+d)\leqslant\|x+d-x\| = \|d\|$$

$\Rightarrow \|d\|\leqslant 1$，故 $\partial d_c(x)\subseteq N_c(x)\cap B[0,1]$

反之，$\forall d \in N_C(x) \cap B[0,1]$，则对 $\forall y \in \mathbb{E}$

$$\langle d, y-x \rangle = \underbrace{\langle d, y-P_C(y) \rangle}_{①} + \langle d, P_C(y)-x \rangle$$

由 $P_C(y) \in C$，$d \in N_C(x)$，$① \leq 0$

$$\begin{aligned}\langle d, y-x \rangle &\leq \langle d, P_C(y)-x \rangle \\ &\leqslant \|d\| \|y-P_C(y)\| \\ &\leqslant \|y-P_C(y)\| = d_C(y)\end{aligned}$$

故 $d \in \partial d_C(x)$ $\boxed{/\!/}$

§ 3.4.5 Maximization

<u>Thm 3.50</u> 设 $f_1, \cdots, f_m : \mathbb{E} \to (-\infty, \infty]$ 是 proper convex

设 $f(x) = \max\{f_1(x), \cdots, f_m(x)\}$，$x \in \bigcap_{i=1}^{m} \operatorname{int}(\operatorname{dom} f_i)$

则 $\partial f(x) = \operatorname{conv}\left\{\bigcup_{i\in I(x)} \partial f_i(x)\right\}$, $I(x) = \{i \in \{1,\cdots,m\} : f_i(x) = f(x)\}$

证明：f 凸（由 Thm 2.16(c)），且由 Corollary 3.25

$\forall d \in \mathbb{E}$, $f'(x;d) = \max\limits_{i\in I(x)} f_i'(x;d)$

为了记号的简单，设 $I(x) = \{1, 2, \cdots, k\}$，对

$k \in \{1, 2, \cdots, m\}$，故由 Thm 3.26

$$f'(x;d) = \max_{i=1,\cdots,k} f_i'(x;d)$$

$$= \max_{i=1,\cdots,k} \max_{g_i\in\partial f_i(x)} \langle g_i, d\rangle$$

由 $\max\{a_1,\cdots,a_k\} = \max\limits_{\lambda\in\Delta_k} \sum\limits_{i=1}^{k} \lambda_i a_i$

$$f'(x;d) = \max_{\lambda\in\Delta_k}\left\{\sum_{i=1}^{k}\lambda_i \max\left\{\langle g_i, d\rangle : g_i \in \partial f_i(x)\right\}\right\}$$

$$= \max\left\{\left\langle \sum_{i=1}^{k}\lambda_i g_i, d\right\rangle : g_i\in\partial f_i(x), \lambda\in\Delta_k\right\}$$

$$= \max\left\{\langle g,d\rangle : g\in \operatorname{conv}\left(\bigcup_{i=1}^{k}\partial f_i(x)\right)\right\}$$

$$= \sigma_A(d)$$

其中 $A=\operatorname{conv}\left(\bigcup_{i=1}^{k}\partial f_i(x)\right)$，由 $x\in \operatorname{int}(\operatorname{dom} f)$

$$f'(x;d) = \sigma_{\partial f(x)}(d)$$

故 $\sigma_A(d) = \sigma_{\partial f(x)}(d)$，对 $\forall d\in \mathbb{E}$

由 $\partial f_i(x)$ 非空闭凸，故 $\bigcup_{i=1}^{k}\partial f_i(x)$ 非空紧，

从而 $\operatorname{conv}\left(\bigcup_{i=1}^{k}\partial f_i(x)\right)$ 非空紧凸，故

$$\partial f(x) = A \qquad \boxtimes$$

<u>Example 3.51</u> $\quad f(x)=\max\{x_1,\cdots,x_n\}$

则 $f(x)=\max\{f_1(x),\cdots,f_n(x)\}$，$f_i(x)=x_i$

故 $\partial f_i(x)=\{e_i\}$，记 $I(x)=\{i: f(x)=x_i\}$

故 $\partial f(x) = \operatorname{conv}\left(\bigcup_{i\in I(x)} \partial f_i(x)\right)$

$= \operatorname{conv}\left(\bigcup_{i\in I(x)} \{e_i\}\right)$

$= \left\{\sum_{i\in I(x)} \lambda_i e_i : \sum_{i\in I(x)} \lambda_i = 1, \lambda_j \geq 0, j\in I(x)\right\}$

特别地：$\partial f(\alpha e) = \Delta_n$

Example 3.52 $\quad f(x) = \|x\|_\infty$

- 当 $x=0$ 时，由 Example 3.3

$\partial f(0) = B_{\|\cdot\|_1}[0,1] = \{x\in\mathbb{R}^n : \|x\|_1 \leq 1\}$

- 当 $x\neq 0$ 时，$f(x) = \max\{f_1(x),\dots,f_n(x)\}$，$f_i(x) = |x_i|$

记 $I(x) = \{i : |x_i| = \|x\|_\infty\}$

由 $x\neq 0$，知 $\forall i\in I(x)$，有 $x_i\neq 0$，$\partial f_i(x) = \{\operatorname{sgn}(x_i)e_i\}$

故 $\partial f(x) = \operatorname{conv}\left(\bigcup_{i\in I(x)} \partial f_i(x)\right)$

$$= \operatorname{conv}\left(\bigcup_{i\in I(x)} \{\operatorname{sgn}(x_i)e_i\}\right)$$

$$= \left\{\sum_{i\in I(x)} \lambda_i \operatorname{sgn}(x_i)e_i : \sum_{i\in I(x)} \lambda_i = 1,\ \lambda_j \geqslant 0, j\in I(x)\right\}$$

$\boxed{/}$

Example 3.53 $\quad f(x) = \max\limits_{i=1,\cdots,m} \{a_i^T x + b_i\}$

$a_i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$，则 $f(x) = \max\limits_{i=1,\cdots,m}\{f_1(x),\cdots,f_m(x)\}$

其中 $f_i(x) = a_i^T x + b_i$，$\partial f_i(x) = \{a_i\}$

故 $\quad \partial f(x) = \left\{\sum_{i\in I(x)} \lambda_i a_i : \sum_{i\in I(x)} \lambda_i = 1,\ \lambda_j \geqslant 0, j\in I(x)\right\}$

其中 $I(x) = \{i : f(x) = a_i^T x + b_i\}$ $\quad\boxed{/}$

Example 3.54 $\quad f(x) = \|Ax+b\|_\infty$

$A\in\mathbb{R}^{m\times n}$，$b\in\mathbb{R}^m$，则 $f(x) = g(Ax+b)$

其中 $g(y)=\|y\|_\infty$，故由 Example 3.52

$$\partial g(y)=\begin{cases} B_{\|\cdot\|_1}[0,1], & y=0 \\ \left\{\sum\limits_{i\in I(y)}\lambda_i \operatorname{sgn}(y_i)e_i : \sum\limits_{i\in I(y)}\lambda_i=1, \lambda_j\geq 0, j\in I(y)\right\} & y\neq 0 \end{cases}$$

其中 $I(y)=\left\{i\in\{1,2,\cdots,m\} : |y_i|=\|y\|_\infty\right\}$

由 $\partial f(x)=A^T\partial g(Ax+b)$

$$\partial f(x)=\begin{cases} A^TB_{\|\cdot\|_1}[0,1], & Ax+b=0 \\ \left\{\sum\limits_{i\in I_x}\lambda_i \operatorname{sgn}(a_i^Tx+b_i)a_i : \sum\limits_{i\in I_x}\lambda_i=1, \lambda_j\geq 0, j\in I_x\right\}, & Ax+b\neq 0 \end{cases}$$

其中 $a_1^T,\cdots,a_m^T$ 是 A 的行，$I_x=I(Ax+b)$ $\boxed{/}$

Thm 3.55 设 I 是任意，proper convex func

$f_i:\mathbb{E}\to(-\infty,\infty]$，设 $f(x)=\max\limits_{i\in I}f_i(x)$

则 $\forall x\in \operatorname{dom}f$，有

$$\operatorname{conv}\left(\bigcup_{i\in I(x)}\partial f_i(x)\right)\subseteq \partial f(x)$$，其中 $I(x)=\{i\in I: f(x)=f_i(x)\}$

证明：设 $x\in \operatorname{dom} f$，则 $\forall z\in \operatorname{dom} f$，$i\in I(x)$，$g\in\partial f_i(x)$

$$f(z)\geq f_i(z)\geq f_i(x)+\langle g, z-x\rangle$$
$$= f(x)+\langle g, z-x\rangle$$

故 $g\in\partial f(x)$，故 $\partial f_i(x)\subseteq\partial f(x)$，由 $\partial f(x)$ 的凸性

知 $\operatorname{conv}\left(\bigcup_{i\in I(x)}\partial f_i(x)\right)\subseteq\partial f(x)$ $\boxed{/}$

<u>Example 3.56</u> 设 $A_0, A_1, \dots, A_m\in \mathbb{S}^n$

$$\mathcal{A}(x)=A_0+\sum_{i=1}^m x_i A_i,\quad \forall x\in\mathbb{R}^m$$

$f(x)=\lambda_{\max}(\mathcal{A}(x))$，由对 $\forall x\in\mathbb{R}^m$

$$f(x)=\max_{y\in\mathbb{R}^n:\|y\|_2=1} y^T\mathcal{A}(x)y$$

且由 $f_y(x)\equiv y^T\mathcal{A}(x)y=y^TA_0y+\sum_{i=1}^m (y^TA_iy)x_i$

故 f_y 是 proper convex func，设

$$\tilde{y} \in \operatorname*{Argmax}_{\|y\|_2=1} y^T A(x) y$$

则 $(\tilde{y}^T A_1 \tilde{y}, \tilde{y}^T A_2 \tilde{y}, \cdots, \tilde{y}^T A_m \tilde{y})^T \in \partial f(x)$

$\boxed{\text{方法二}}$ $f(x) = \lambda_{max}(A(x)) = g(B(x)+A_0)$

$B(x) = \sum_{i=1}^{m} x_i A_i$, $g(X) = \lambda_{max}(X)$

故 $\partial f(x) = B^T(\partial g(B(x)+A_0))$

由 Example 3.8，知 $\tilde{y}\tilde{y}^T \in \partial g(B(x)+A_0)$，故

$$B^T(\tilde{y}\tilde{y}^T) \in \partial f(x)$$

故 $B^T(\tilde{y}\tilde{y}^T) = (\mathrm{Tr}(A_1\tilde{y}\tilde{y}^T), \cdots, \mathrm{Tr}(A_m\tilde{y}\tilde{y}^T))^T$

$$= (\tilde{y}^T A_1 \tilde{y}, \cdots, \tilde{y}^T A_m \tilde{y})^T$$

$\square$

§ 3.5 The Value Function

$$f_{opt} = \min_{x\in X} \{ f(x) : g_i(x) \leq 0, i=1,\cdots,m, Ax+b=0 \} \quad (3.56)$$

$f, g_1, \cdots, g_m : \mathbb{E} \to (-\infty, \infty]$，$X \subseteq \mathbb{E}$ 非空

$A \in \mathbb{R}^{p\times n}$，$b \in \mathbb{R}^p$，定义

$$g(x) \equiv (g_1(x), \cdots, g_m(x))^T$$

则 (3.56) 等价于

$$\min_{x\in X} \{ f(x) : g(x) \leq 0, Ax+b=0 \}$$

定义 value function $v : \mathbb{R}^m \times \mathbb{R}^p \to [-\infty, \infty]$

$$v(u,t) = \min_{x\in X} \{ f(x) : g(x) \leq u, Ax+b=t \}$$

记 $C(u,t) = \{ x\in X : g(x) \leq u, Ax+b=t \}$

则 $v(u,t) = \min \{ f(x) : x\in C(u,t) \}$

Lemma 3.58 设 $f, g_1, \cdots, g_m: \mathbb{E} \to (-\infty, \infty]$, $X \subseteq \mathbb{E}$ 非空, $A \in \mathbb{R}^{p\times n}$, $b \in \mathbb{R}^p$, 则

$v(u,t) \geq v(w,t) \quad \forall u \leq w$

证明: 由 $C(u,t) \subseteq C(w,t)$, 故结论显然. $\boxed{/}$

Lemma 3.58 设 $f, g_1, \cdots, g_m: \mathbb{E} \to (-\infty, \infty]$ convex $X \subseteq \mathbb{E}$ 非空, 凸, $A \in \mathbb{R}^{p\times n}$, $b \in \mathbb{R}^p$, 设 v proper, 则

v 在 $\mathbb{R}^m \times \mathbb{R}^p$ 上凸

证明: 设 $(u,t), (w,s) \in \mathrm{dom}\, v$, $\lambda \in [0,1]$

只需证: $v\big(\lambda u + (1-\lambda)w, \lambda t + (1-\lambda)s\big) \leq \lambda v(u,t) + (1-\lambda) v(w,s)$

由 v proper, 则 $\exists \{x_k\}_{k\geq 1}, \{y_k\}_{k\geq 1}$

$\underbrace{x_k \in C(u,t), y_k \in C(w,s)}$, $f(x_k) \to v(u,t)$, $f(y_k) \to v(w,s)$

$\Downarrow$

$g(x_k) \leq u$, $g(y_k) \leq w$

故 $g(\lambda x_k+(1-\lambda)y_k)\le \lambda g(x_k)+(1-\lambda)g(y_k)$

$$\le \lambda u+(1-\lambda)w$$

且 $A(\lambda x_k+(1-\lambda)y_k)+b=\lambda(Ax_k+b)+(1-\lambda)(Ay_k+b)$

$$=\lambda t+(1-\lambda)s$$

综上：$\lambda x_k+(1-\lambda)y_k\in C(\lambda u+(1-\lambda)w,\ \lambda t+(1-\lambda)s)$

由 f 的凸性：

$$f(\lambda x_k+(1-\lambda)y_k)\le \lambda f(x_k)+(1-\lambda)f(y_k)$$

令 $k\to\infty$，则

$$\liminf_{k\to\infty} f(\lambda x_k+(1-\lambda)y_k)\le \lambda v(u,t)+(1-\lambda)v(w,s)$$

由 v 的定义知：

$$v\big(\lambda u+(1-\lambda)w,\ \lambda t+(1-\lambda)s\big)\le f(\lambda x_k+(1-\lambda)y_k)$$

$$\Rightarrow\ v\big(\lambda u+(1-\lambda)w,\ \lambda t+(1-\lambda)s\big)\le \liminf_{k\to\infty} f(\lambda x_k+(1-\lambda)y_k)$$

$$\leq \lambda v(u,t) + (1-\lambda) v(w,s) \qquad \boxed{/}$$

考虑(3.56)的 dual func: $q: \mathbb{R}^m \times \mathbb{R}^q \to [-\infty, \infty)$

$$q(y,z) = \min_{x \in X} \left\{ L(x,y,z) = f(x) + y^T g(x) + z^T (Ax+b) \right\}$$

则 $$q_{opt} = \max_{y \in \mathbb{R}^m_+,\, z \in \mathbb{R}^p} \left\{ q(y,z) : (y,z) \in \operatorname{dom}(-q) \right\}$$

其中 $\operatorname{dom}(-q) = \left\{ (y,z) \in \mathbb{R}^m_+ \times \mathbb{R}^p : q(y,z) > -\infty \right\}$

问：何时 $f_{opt} = q_{opt}$?

Thm A.1 (strong duality theorem)

$$
\begin{aligned}
f_{opt} = \min\ & f(x) \\
\text{s.t. }\ & g_i(x) \leq 0,\ i = 1, \cdots, m \\
& h_j(x) \leq 0,\ j = 1, \cdots, p \qquad (A.1) \\
& s_k(x) = 0,\ k = 1, \cdots, q \\
& x \in X
\end{aligned}
$$

其中 $X = P \cap C$，$P \subseteq \mathbb{E}$ 是 convex polyhedral set

$C \subseteq \mathbb{E}$ convex；$f, g_i, i=1,\cdots,m : \mathbb{E} \to (-\infty, \infty]$ 凸

且 $X \subseteq \mathrm{dom} f$，$X \subseteq \mathrm{dom}\, g_i$． h_j，s_k 是 affine func

设以下约束规格成立：

(i) ∃可行解 $\bar{x}$，s.t. $g(\bar{x}) < 0$

(ii) ∃ $\hat{x}$，s.t. $h_j(x) \le 0$，$s_k(x) = 0$，且 $\hat{x} \in P \cap \mathrm{ri}(C)$

若 (A.1) 有有限的，则 dual 问题

$$q_{opt} = \max \left\{ q(\lambda, \eta, \mu) : (\lambda, \eta, \mu) \in \mathrm{dom}(-q) \right\}$$

其中 $q : \mathbb{R}_+^m \times \mathbb{R}_+^p \times \mathbb{R}^q \to \mathbb{R} \cup \{-\infty\}$

$$q(\lambda, \eta, \mu) = \min_{x \in X} L(x, \lambda, \eta, \mu)$$

$$= \min_{x \in X} \left[f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \eta_j h_j(x) + \sum_{k=1}^{q} \mu_k s_k(x) \right]$$

最优解是可达的，且 $f_{opt} = q_{opt}$

证明：WLOG，$h(x)=Cx+d$，$s(x)=Ax+b$

$D=\mathbb{R}^n$

<u>case1</u>：$\hat{x}\in \operatorname{int} C$，$\begin{pmatrix}C\\A\end{pmatrix}$行满秩

$$S=\left\{(p_1,p_2,q,r)\ \middle|\ \begin{array}{l}\exists x\in C,\ s.t.\ g(x)\le p_1,\\ h(x)\le p_2,\ s(x)=q,\ f(x)\le r\end{array}\right\}$$

claim：S是凸集，且$(0,0,0,f_{opt})\in\partial S$

该claim易证

故由<u>Thm3.13</u>，$\exists (u_1,u_2,v,w)\ne 0$，s.t.

$$p_1^T u_1+p_2^T u_2+q^T v+rw\le wf_{opt}\quad \forall (p_1,p_2,q,r)\in S$$

故$\forall x\in C$，有

$$u_1^T g(x)+u_2^T h(x)+v^T s(x)+wf(x)\le wf_{opt}\quad (*)$$

易知$u_1\le 0$，$u_2\le 0$，$w\le 0$，下证$w<0$

若不然，则 $w=0$，此时 ⊛ $\Leftrightarrow$ 对 $\forall x\in C$

$$u_1^T g(x)+u_2^T h(x)+v^T s(x)\le 0$$

取 $x=\bar{x}$，则 $s(\bar{x})=0$，且 WLOG，设 $h(\bar{x})=0$

否则将 $h_i(x)<0$ 的分量吸收到 $g(x)$ 中

则 $u_1^T g(\bar{x})\le 0$，又由 $u_1\le 0$，$g(\bar{x})<0\Rightarrow u_1=0$

此时，⊛ $\Leftrightarrow$ 对 $\forall x\in C$，有

$$u_2^T h(x)+v^T s(x)\le 0$$

取 $x=\hat{x}$，则由 $s(\hat{x})=0$ 知 $u_2^T h(\hat{x})\le 0$，但

$u_2\le 0$，$h(\hat{x})\le 0$，故 $u_2^T h(\hat{x})=0$

故 $u_2^T(h(x)-h(\hat{x}))+v^T(s(x)-s(\hat{x}))\le 0$

即 $u_2^T C(x-\hat{x})+v^T A(x-\hat{x})\le 0 \quad \forall x\in C$

又由 $\hat{x}\in \operatorname{int} C$，故 $u_2^T C+v^T A=0$

由 $\begin{pmatrix} C \\ A \end{pmatrix}$ 行满秩，知 $u_2=0$，$v=0$，故矛盾

从而 $w<0$，㊟两边同除 w：

$$\left(\frac{u_1}{w}\right)^T g(x)+\left(\frac{u_2}{w}\right)^T h(x)+\left(\frac{v}{w}\right)^T s(x)+f(x)\geqslant f_{opt}$$

$\Leftrightarrow \quad L(x; u_1^*, u_2^*, v^*)\geqslant f_{opt}$，其中

$u_1^*=\frac{u_1}{w}\geqslant 0$，$u_2^*=\frac{u_2}{w}\geqslant 0$，$v^*=\frac{v}{w}$

对 x 取 min $\Rightarrow g(u_1^*, u_2^*, v^*)\geqslant f_{opt}$

由 $g_{opt}\geqslant g(u_1^*, u_2^*, v^*)\geqslant f_{opt}\underset{\text{weak duality}}{\geqslant} g_{opt}$

故 $f_{opt}=g_{opt}$

<u>case2</u>：$\hat{x}\in ri(C)$，$\begin{pmatrix} C \\ A \end{pmatrix}$ 行满秩

将优化问题看作在 aff(C) 上，此时 $\hat{x}\in \operatorname{int}\widetilde{C}$.

case2 可以退化到 case1

case3: $\hat{x} \in \mathrm{ri}(C)$, $\begin{pmatrix} C \\ A \end{pmatrix}$ 不一定行满秩

约束 $Cx+d \le 0$ 可以转化成 $\begin{cases} Cx+d+s=0 \\ s \geq 0 \end{cases}$

故 (A.1) 转化成

$$\min_{\tilde{x}=\binom{x}{s}} \quad f(\tilde{x})$$

$$\text{s.t.} \quad g(x) \le 0$$

$$\boxed{\begin{aligned} Cx+s+d&=0 \\ Ax+b&=0 \end{aligned}} \quad \begin{bmatrix} C & I \\ A & 0 \end{bmatrix}\tilde{x}+\begin{bmatrix} d \\ b \end{bmatrix}=0$$

$$\tilde{x} \in C \times \{s : s \geq 0\}$$

对等式约束 $\begin{bmatrix} C & I \\ A & 0 \end{bmatrix}\tilde{x}+\begin{bmatrix} d \\ b \end{bmatrix}=0$，自然可以

不失一般地设 $\begin{bmatrix} C & I \\ A & 0 \end{bmatrix}$ 行满秩 $\Leftrightarrow$ $\begin{bmatrix} C \\ A \end{bmatrix}$ 行满秩

$\square$

Thm3.59 设 $f, g_1, \cdots, g_m: \mathbb{E} \to (-\infty, \infty]$ 是 convex 的,

$X \subseteq \mathbb{E}$ 是非空凸集, $A \in \mathbb{R}^{p\times n}$, $b \in \mathbb{R}^p$, 设 v 是 value func

且 $f_{opt} = q_{opt} \in (-\infty, \infty)$, 且对偶问题的 optimal set

非空, 则

(a) v 是 proper, convex 的

(b) (y, z) 是 (3.63) 的最优解 $\Leftrightarrow$ $-(y, z) \in \partial v(0,0)$

证明: 设 (y, z) 是 dual problem 的最优解

则 $$L(x; y, z) \geq \min_{x \in X} L(x, y, z) = q(y, z)$$

$$= q_{opt} = f_{opt} = v(0,0) \quad \forall x \in X$$

故对 $\forall x \in C(u, t)$

$$v(0,0) - y^T u - z^T t \leq L(x; y, z) - y^T u - z^T t$$

$$= f(x) + y^T (g(x) - u) + z^T (Ax + b - t)$$

$\leqslant f(x)$ （由 $y \geqslant 0$, $g(x) \leq u$）

故 $v(u,t) \geqslant v(0,0) - y^T u - z^T t \quad \forall u,t$

故 $-(y,z) \in \partial v(0,0)$

从而 $v(u,t) \geqslant f_{opt} - y^T u - z^T t > -\infty$

结合 lemma3.58，(a)和(b)的 $\Rightarrow$ 方向得证

下证 (b)($\Leftarrow$)方向：设 $-(y,z) \in \partial v(0,0)$

$$v(u,t) \geqslant v(0,0) - y^T u - z^T t \quad \forall (u,t) \in \mathbb{R}^m \times \mathbb{R}^p \quad (3.65)$$

则设 $x \in X$

$$f(x) \geqslant v(g(x), Ax+b)$$

$$\geqslant v(0,0) - y^T g(x) - z^T (Ax+b)$$

故 $v(0,0) \leqslant L(x;y,z)$ 对 $\forall x \in X$

$$\Rightarrow v(0,0) \leqslant \min_{x \in X} L(x;y,z) = q(y,z)$$

设 $j\in\{1,2,\cdots,m\}$，令 $u=e_j$，$t=0$ 代入 (3.65)

有 $y_j \geqslant v(0,0)-v(e_j,0) \geqslant 0$

Lemma 3.57

故 $y\geqslant 0$，从而

$$q_{opt}=f_{opt}=v(0,0)\leqslant q(y,z)\leqslant q_{opt}$$

$$\Rightarrow \quad q(y,z)=q_{opt} \qquad \boxed{\angle}$$

Thm 3.60 设 $f,g_1,\cdots,g_m:\mathbb{E}\to(-\infty,\infty]$ 是 convex

$X\subseteq\mathbb{E}$ 是非空凸集，$A\in\mathbb{R}^{p\times n}$，$b\in\mathbb{R}^p$，设 f_{opt} 和 q_{opt}

是 (3.56) 和 (3.63) 最优值，且 $f_{opt}=q_{opt}\in(-\infty,\infty)$.

对偶问题的 optimal set 非空，设 (y^*,z^*) 是 dual 问题

的 optimal solution，设 $\tilde{x}\in X$ 满足：

$$f(\tilde{x})-f_{opt}+\rho_1\|[g(\tilde{x})]_+\|_2+\rho_2\|A\tilde{x}+b\|_2\leqslant\delta$$

其中 $\delta > 0$，$\rho_1 \geq 2\|y^*\|_2$，$\rho_2 \geq 2\|z^*\|_2$，则

$$f(\tilde{x}) - f_{opt} \leq \delta$$

$$\|[g(\tilde{x})]_+\|_2 \leq \frac{2}{\rho_1}\delta$$

$$\|A\tilde{x}+b\|_2 \leq \frac{2}{\rho_2}\delta$$

证明：$f(\tilde{x}) - f_{opt} \leq \delta$ 自然成立

设 $v(u,t) = \min\limits_{x\in X}\{f(x) : g(x) \leq u,\ Ax+b=t\}$

由 (y^*, z^*) 是 dual 问题最优解，则 $(-y^*, -z^*) \in \partial v(0,0)$

对 $\forall (u,t) \in \operatorname{dom} v$，有

$$v(u,t) - v(0,0) \geq \langle -y^*, u\rangle + \langle -z^*, t\rangle$$

将 $u = \tilde{u} \equiv [g(\tilde{x})]_+$，$t = \tilde{t} \equiv A\tilde{x}+b$ 代入

$$(\rho_1 - \|y^*\|_2)\|\tilde{u}\|_2 + (\rho_2 - \|z^*\|_2)\|\tilde{t}\|_2$$

$$= -\|y^*\|_2\|\tilde{u}\|_2 - \|z^*\|_2\|\tilde{t}\|_2 + \rho_1\|\tilde{u}\|_2 + \rho_2\|\tilde{t}\|_2$$

$$\leq \langle -y^*, \tilde{u}\rangle + \langle -z^*, \tilde{t}\rangle + \rho_1\|\tilde{u}\|_2 + \rho_2\|\tilde{t}\|_2$$

$$\leq v(\tilde{u}, \tilde{t}) - v(0,0) + \rho_1\|\tilde{u}\|_2 + \rho_2\|\tilde{t}\|_2$$

$$\leq f(\tilde{x}) - f_{opt} + \rho_1\|\tilde{u}\|_2 + \rho_2\|\tilde{t}\|_2 \leq \delta$$

故 $\|[g(\tilde{x})]_+\|_2 = \|\tilde{u}\|_2 \leq \dfrac{\delta}{\rho_1 - \|y^*\|_2} \leq \dfrac{2}{\rho_1}\delta$

$\|A\tilde{x}+b\|_2 = \|\tilde{t}\|_2 \leq \dfrac{\delta}{\rho_2 - \|z^*\|_2} \leq \dfrac{2}{\rho_2}\delta$

$\boxtimes$

§3.6 Lipschiz 连续和 subgradient 的界

<u>Thm3.61</u> 设 $f: \mathbb{E} \to (-\infty, \infty]$ 是 proper, convex func

$X \subseteq \text{int}(\text{dom} f)$,

(i) $|f(x) - f(y)| \leq L\|x - y\| \quad \forall x, y \in X$

(ii) $\|g\|_* \leq L$, 对 $\forall g \in \partial f(x), x \in X$

则 (a) (ii) $\Rightarrow$ (i)

(b) 若 X 开，则 (i) $\Leftrightarrow$ (ii)

证明：

(a) 设 (ii) 成立，$x, y \in X$，$g_x \in \partial f(x)$，$g_y \in \partial f(y)$

则 $f(x) - f(y) \le \langle g_x, x-y \rangle \le L\|x-y\|$

$f(y) - f(x) \le \langle g_y, y-x \rangle \le L\|x-y\|$

故 (i) 成立

(b) (ii) $\Rightarrow$ (i) 已经证明，下证 (i) $\Rightarrow$ (ii)

取 $x \in X$，$g \in \partial f(x)$，下证 $\|g\|_* \le L$

定义 $g^+ \in \mathbb{E}$，s.t. $\|g^+\| = 1$ 且 $\langle g^+, g \rangle = \|g\|_*$

取 ε 足够小，s.t. $x + \varepsilon g^+ \in X$，则

$$f(x+\varepsilon g^+) \ge f(x) + \langle g, \varepsilon g^+ \rangle$$

$$\Rightarrow \varepsilon \|g\|_* = \langle g, \varepsilon g^\dagger \rangle \leq f(x + \varepsilon g^\dagger) - f(x)$$

$$\leq L\varepsilon$$

$$\Rightarrow \|g\|_* \leq L \qquad \boxed{/}$$

§3.7 Optimality Condition

§3.7.1 Fermat's Optimality Condition

Thm3.63 设 $f:\mathbb{E}\to(-\infty,\infty]$ 是 proper, convex func

则 $x^* \in \arg\min\{f(x): x\in\mathbb{E}\} \iff 0\in\partial f(x^*)$

证明: $x^* \in \arg\min\{f(x): x\in\mathbb{E}\} \iff$

$$f(x)\geq f(x^*)+\langle 0, x-x^*\rangle \quad \forall x\in \operatorname{dom} f$$

$\iff 0\in\partial f(x^*)$ $\boxtimes$

Example3.64 $\min\limits_{x\in\mathbb{R}^n}\left[f(x)=\max\limits_{i=1,\dots,m}\{a_i^T x+b_i\}\right]$

记 $I(x)=\{i: f(x)=a_i^T x+b_i\}$

由 Example3.53

$$\partial f(x)=\left\{\sum_{i\in I(x)}\lambda_i a_i : \sum_{i\in I(x)}\lambda_i=1, \lambda_j\geq 0, j\in I(x)\right\}$$

由Fermet最优条件：x^*是(3.71)最优解 $\Leftrightarrow$ $\exists \lambda \in \Delta_m$

$$\text{s.t.} \quad 0=\sum_{i=1}^{m} \lambda_i a_i, \ \lambda_j = 0 \text{ 对 } \forall j \notin I(x^*)$$

记 $A=\begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix} \in \mathbb{R}^{m\times n}$，则上式 $\Leftrightarrow$

$\exists \lambda \in \Delta_m$, s.t. $A^T\lambda = 0$, $\lambda_j(a_j^T x^* + b_j - f(x^*)) = 0$, $j=1,\cdots,m$

$\boxed{}$

Example 3.65 (medians) 给定n个不同数 $a_1 < \cdots < a_n$

记 $A=\{a_1,\cdots,a_n\} \subseteq \mathbb{R}$，A的medians β满足

$$\#\{i: a_i \le \beta\} \geqslant \frac{n}{2}, \quad \#\{i: a_i \geqslant \beta\} \geqslant \frac{n}{2}$$

易知 $$\text{median}(A) = \begin{cases} a_{\frac{n+1}{2}}, & n \text{ odd} \\ [a_{\frac{n}{2}}, a_{\frac{n}{2}+1}], & n \text{ even} \end{cases}$$

下证上式是优化问题

$$\min\left\{ f(x) \equiv \sum_{i=1}^{n} |x_i - a_i| \right\}$$

的最优解：记 $f_i(x) = |x-a_i|$，$f(x) = f_1(x)+\cdots+f_n(x)$

对 $\forall i \in \{1,2,\cdots,n\}$

$$\partial f_i(x) = \begin{cases} 1, & x>a_i \\ -1, & x<a_i \\ [-1,1], & x=a_i \end{cases}$$

由 Thm3.40

$$\partial f(x) = \partial f_1(x) + \cdots + \partial f_n(x)$$

$$= \begin{cases} \#\{i: a_i<x\} - \#\{i: a_i>x\}, & x \notin A \\ \#\{i: a_i<x\} - \#\{i: a_i>x\} + [-1,1], & x \in A \end{cases}$$

$$= \begin{cases} 2i-n, & x \in (a_i, a_{i+1}) \\ 2i-1-n+[-1,1], & x = a_i \\ -n, & x < a_1 \\ n, & x > a_n \end{cases}$$

设 $i \in \{1,2,\cdots,n\}$，$0 \in \partial f(a_i) \Leftrightarrow |2i-1-n| \le 1$

$\Leftrightarrow \frac{n}{2} \le i \le \frac{n}{2}+1$

对 $x\in(a_i,a_{i+1})$, $0\in\partial f(x) \Leftrightarrow i=\frac{n}{2}$

- 若 n 是 odd: $x^*=a_{\frac{n+1}{2}}$
- 若 n 是 even: $x^*\in[a_{\frac{n}{2}},a_{\frac{n}{2}+1}]$ $\square$

Example 3.66 (Fermat-Weber)

给定 $\mathbb{R}^d$ 中 m 个不同点, $A=\{a_1,\cdots,a_m\}$

$$\text{(FW)}\quad \min_{x\in\mathbb{R}^d}\left\{f(x)\equiv\sum_{i=1}^{m}w_i\|x-a_i\|_2\right\}$$

记 $f_i(x)=w_i g_i(x)$, 其中 $g_i(x)\equiv\|x-a_i\|_2$, 则

$$\partial f_i(x)=\begin{cases} w_i\dfrac{x-a_i}{\|x-a_i\|_2}, & x\neq a_i \\ B_{\|\cdot\|_2}[0,w_i], & x=a_i \end{cases}$$

则 $\partial f(x)=\sum_{i=1}^{m}\partial f_i(x)$

$$= \begin{cases} \sum_{i=1}^{m} w_i \dfrac{x-a_i}{\|x-a_i\|_2}, & x \notin A \\ \sum_{\substack{i=1 \\ i\neq j}}^{m} w_i \dfrac{x-a_i}{\|x-a_i\|_2} + B[0, w_j], & x = a_j \ (j=1, \cdots, m) \end{cases}$$

故 x^* 是 (FW) 的最优解 $\Leftrightarrow$

或 $x^* \notin A$, $\sum_{i=1}^{m} w_i \dfrac{x^*-a_i}{\|x^*-a_i\|_2} = 0$

或对 $j \in \{1, \cdots, m\}$, $x^* = a_j$, 且 $\left\| \sum_{i=1, i\neq j}^{m} w_i \dfrac{x^*-a_i}{\|x^*-a_i\|_2} \right\| \leq w_j$

$\square$

§ 3.7.2 Convex Constrained Optimization

$$\min \{ f(x) : x \in C \}$$

f 是 extended 实值 convex func, $C \subseteq \mathbb{E}$ 是凸集

<u>Thm 3.67</u> 设 $f: \mathbb{E} \to (-\infty, \infty]$ 是 proper, convex func

$C \subseteq \mathbb{E}$ 是 convex，且 $\mathrm{ri}(\mathrm{dom} f) \cap \mathrm{ri}(C) \neq \phi$，则 $x^* \in C$

是 (3.75) 的最优解 $\Leftrightarrow$

$\exists g \in \partial f(x^*)$，$-g \in N_C(x^*)$

证明：(3.75) $\Leftrightarrow$ $\min\limits_{x \in \mathbb{E}} f(x) + \delta_C(x)$

由 $\mathrm{ri}(\mathrm{dom} f) \cap \mathrm{ri}(C) \neq \phi$，故由 Thm3.40 ：$\forall x \in C$

$$\begin{aligned} \partial(f+\delta_C)(x) &= \partial f(x) + \partial \delta_C(x) \\ &= \partial f(x) + N_C(x) \end{aligned}$$

由 Fermat 最优性条件：$x^* \in C$ 是 (3.75) 最优解 $\Leftrightarrow$

$$0 \in \partial f(x^*) + N_C(x^*)$$

$\Leftrightarrow$ $(-\partial f(x^*)) \cap N_C(x^*) \neq \phi$ $\square$

Corollary 3.68 设 $f: \mathbb{E} \to (-\infty, \infty]$ 是 proper convex func

C 是凸集，s.t. $\mathrm{ri}(\mathrm{dom} f) \cap \mathrm{ri}(C) \neq \phi$，则 $x^* \in C$

是 (3.75) 的最优解 $\Leftrightarrow$

$\exists g \in \partial f(x^*)$, s.t. $\langle g, x - x^* \rangle \geqslant 0 \quad \forall x \in C$

<u>Example 3.69</u> 设推论 3.68 条件成立，$C = \Delta_n, \mathbb{E} = \mathbb{R}^n$

给定 $x^* \in \Delta_n$，下证

$$(\mathrm{I}) \quad g^T (x - x^*) \geqslant 0 , \quad \forall x \in \Delta_n$$

$$\Leftrightarrow \quad (\mathrm{II}) \quad \exists \mu \in \mathbb{R}, \text{ s.t. } g_i \begin{cases} = \mu , & x_i^* > 0 \\ \geqslant \mu , & x_i^* = 0 \end{cases}$$

证明:

($\Leftarrow$) 对 $\forall x \in \Delta_n$

$$g^T (x - x^*) = \sum_{i=1}^{n} g_i (x_i - x_i^*)$$

$$= \sum_{i: x_i^* > 0} g_i (x_i - x_i^*) + \sum_{i: x_i^* = 0} g_i x_i$$

$$\geqslant \sum_{i: x_i^* > 0} \mu (x_i - x_i^*) + \mu \sum_{i: x_i^* = 0} x_i$$

$$= \mu \sum_{i=1}^{n} x_i - \sum_{i: x_i^* > 0} x_i^* \cdot \mu = \mu - \mu = 0$$

($\Rightarrow$) 设 i,j 两不同指标，$x_i^*>0$，定义 $x\in\Delta_n$，

$$x_k=\begin{cases} x_k^*, & k\notin\{i,j\} \\ x_i^*-\frac{x_i^*}{2}, & k=i \\ x_j^*+\frac{x_i^*}{2}, & k=j \end{cases}$$

则 $g^T(x-x^*)\geq 0 \Rightarrow -\frac{x_i^*}{2}g_i+\frac{x_i^*}{2}g_j\geq 0$

由 $x_i^*>0$ 知：$g_i\leq g_j$

故对任意两指标 $i\neq j$，s.t. $x_i^*>0$，$x_j^*>0$，则

由上述过程知：$g_i\leq g_j$，$g_j\leq g_i \Rightarrow g_i=g_j\doteq\mu$

对任意指标 j，s.t. $x_j^*=0$，有 $g_j\geq\mu$. $\square$

Corollary 3.70 设 $f:\mathbb{E}\to(-\infty,\infty]$ proper, convex

设 $\text{ri}(\Delta_n)\cap\text{ri}(\text{dom}f)\neq\emptyset$，则 $x^*\in\Delta_n$ 是下优化问题

最优解：$\min\{f(x):x\in\Delta_n\}\Leftrightarrow\exists g\in\partial f(x^*)$，$\mu\in\mathbb{R}$，

s.t.

$$g_i \begin{cases} = \mu, & x_i^* > 0 \\ \geq \mu, & x_i^* = 0 \end{cases}$$

Example 3.71

$$\min_x \left\{ \sum_{i=1}^n x_i \log x_i - \sum_{i=1}^n y_i x_i : x \in \Delta_n \right\}$$

其中 $y \in \mathbb{R}^n$ 是给定向量，(3.79) $\Leftrightarrow \min\{f(x) : x \in \Delta_n\}$

$$f(x) = \begin{cases} \sum_{i=1}^n x_i \log x_i - \sum_{i=1}^n y_i x_i, & x \geq 0 \\ \infty & , \text{else} \end{cases}$$

设 $\exists$ 最优解 x^*, s.t. $x^* > 0$，则由推论3.70，且 f 于任意 $x \geq 0$ 处可微，知 $\exists \mu \in \mathbb{R}$, s.t. 对 $\forall i$

$$\frac{\partial f}{\partial x_i}(x^*) = \log x_i^* + 1 - y_i = \mu$$

$$\Rightarrow \quad x_i^* = e^{\mu - 1 + y_i} = \alpha e^{y_i}, \quad i = 1, \dots, n$$

其中 $\alpha = e^{\mu - 1}$，由 $\sum_{i=1}^n x_i^* = 1$，知 $\alpha = \dfrac{1}{\sum_{j=1}^n e^{y_j}}$

$$x_i^* = \frac{e^{y_i}}{\sum_{j=1}^{n} e^{y_j}}, \quad i=1,\cdots,n$$

$\boxed{/}$

§3.7.3 The Nonconvex Composite Model

Thm3.72 设 $f: \mathbb{E} \to (-\infty, \infty]$ proper, $g: \mathbb{E} \to (-\infty, \infty]$ 是 proper, convex, s.t. $\text{dom}\, g \subseteq \text{int}(\text{dom} f)$, 考虑优化问题:

$$(P) \quad \min_{x \in \mathbb{E}} f(x) + g(x)$$

(a) 若 $x^* \in \text{dom}\, g$ 是 (P) 的 local 最优, 且 f 在 x^* 处可微

则 $-\nabla f(x^*) \in \partial g(x^*) \qquad (3.80)$

(b) 设 f 是 convex, 若 f 在 $x^* \in \text{dom}\, g$ 处可微, 则 x^* 是 (P) 的全局 optimal $\Leftrightarrow$ (3.80) 成立

证明: (a) 设 $y \in \text{dom}\, g$, 对 $\forall \lambda \in (0,1)$, 则

$$x_\lambda = (1-\lambda) x^* + \lambda y \in \text{dom}\, g$$

由 x^* 的local最优性：对充分小的 λ

$$f(x_\lambda)+g(x_\lambda)\geqslant f(x^*)+g(x^*)$$

$$\Leftrightarrow\quad f\big((1-\lambda)x^*+\lambda y\big)+g\big((1-\lambda)x^*+\lambda y\big)\geqslant f(x^*)+g(x^*)$$

由 g 的凸性：

$$f\big((1-\lambda)x^*+\lambda y\big)+(1-\lambda)g(x^*)+\lambda g(y)\geqslant f(x^*)+g(x^*)$$

$$\Leftrightarrow\quad \frac{f\big((1-\lambda)x^*+\lambda y\big)-f(x^*)}{\lambda}\geqslant g(x^*)-g(y)$$

取 $\lambda\downarrow 0$ $\quad f'(x^*;\ y-x^*)\geqslant g(x^*)-g(y)$

在 x^* 处可微，方向导数∃

由Thm3.29 ：$f'(x^*;y-x^*)=\langle\nabla f(x^*),\ y-x^*\rangle$

故对 $\forall\, y\in \mathrm{dom}\,g$: $g(y)\geqslant g(x^*)+\langle-\nabla f(x^*),y-x^*\rangle$

$$\Rightarrow\quad -\nabla f(x^*)\in\partial g(x^*)$$

(b) $(\Rightarrow)$ 若 x^* 是(P)的最优解，自然是 local optimal 的

故 (3.80) 成立

(⇐) 设 (3.80) 成立，则对 $\forall\, y \in \mathrm{dom}\, g$

$$g(y) \geq g(x^*) + \langle -\nabla f(x^*), y - x^* \rangle$$

又由 f 的凸性： $f(y) \geq f(x^*) + \langle \nabla f(x^*), y - x^* \rangle$

$$\Rightarrow \quad f(y) + g(y) \geq f(x^*) + g(x^*) \qquad \boxed{\diagup}$$

<u>Def 3.73</u> 设 $f: \mathbb{E} \to (-\infty, \infty]$ 是 proper, $g: \mathbb{E} \to (-\infty, \infty]$ 是 proper, convex 的, s.t. $\mathrm{dom}\, g \subseteq \mathrm{int}(\mathrm{dom} f)$, 考虑

$$(P) \quad \min_{x \in \mathbb{E}} f(x) + g(x)$$

则 x^* (f 在 x^* 处可微) 是 (P) 的 stationary point, 若

$$-\nabla f(x^*) \in \partial g(x^*)$$

<u>Example 3.74</u> 当 $g = \delta_C$, C 是非空凸集，(P) 变成

$$\min \{ f(x) : x \in C \}$$

则 $x^* \in C$ 是 (P) 的 stationary point, 若 f 在 x^* 处可微

则 $-\nabla f(x^*) \in \partial \delta_C(x^*) = N_C(x^*)$

$\Leftrightarrow \quad \langle \nabla f(x^*), x - x^* \rangle \geq 0 \quad \forall x \in C$ $\quad \boxed{/}$

Example 3.75 考虑问题

$$\min_{x \in \mathbb{R}^n} f(x) + \lambda \|x\|_1 \qquad (3.84)$$

其中 $f: \mathbb{R}^n \to (-\infty, \infty]$, $x^* \in \operatorname{int}(\operatorname{dom} f)$, f 在 x^* 是可微的

是 (3.84) 的 stationary point, 若

$$-\nabla f(x^*) \in \lambda \partial g(x^*)$$

$$\Leftrightarrow \quad \frac{\partial f(x^*)}{\partial x_i} \begin{cases} = -\lambda, & x_i^* > 0 \\ = \lambda, & x_i^* < 0 \\ \in [-\lambda, \lambda], & x_i^* = 0 \end{cases}$$ $\quad \boxed{/}$

§ 3.7.4 KKT condition

Lemma 3.76 设 $f, g_1, \cdots, g_m : \mathbb{E} \to \mathbb{R}$，考虑问题：

$$\begin{aligned} \min \ & f(x) \\ \text{s.t.} \ & g_i(x) \le 0, \ i = 1, \cdots, m \end{aligned} \tag{3.86}$$

设(3.86)的最优值 $\bar{f}$ 有限，设

$$F(x) \equiv \max\left\{ f(x) - \bar{f}, g_1(x), \cdots, g_m(x) \right\}$$

则(3.86)的最优集与 $F(x)$ 的最小值集相同

证明：设 X^* 是(3.86)的 optimal set，只需证：

(i) $F(x) > 0, \ \forall x \notin X^*$

(ii) $F(x) = 0, \ \forall x \in X^*$

(i)：设 $x \notin X^*$，则 x 可能不可行，此时 $\exists i$, s.t. $g_i(x) > 0$

故 $F(x) > 0$，x 也可能可行，此时 $f(x) > \bar{f}$，故 $F(x) > 0$

(ii)：设 $x \in X^*$，则 $g_i(x) \le 0, \forall i$，$f(x) = \bar{f}$，故 $F(x) = 0$

Thm3.77 (Fritz-John 最优性条件) 考虑问题：

$$\begin{aligned} \min\ & f(x) \\ \text{s.t.}\ & g_i(x) \le 0,\ i=1,\cdots,m \end{aligned} \qquad (3.89)$$

其中 $f, g_1, \cdots g_m : \mathbb{E} \to \mathbb{R}$ 是 convex. 设 x^* 是 (3.89) 最优解，

则 $\exists\ \lambda_0, \cdots, \lambda_m \geq 0$，且不全为 0，s.t.

$$0 \in \lambda_0 \partial f(x^*) + \sum_{i=1}^{m} \lambda_i \partial g_i(x^*)$$

$$\lambda_i g_i(x^*) = 0 \quad , i=1,\cdots,m$$

证明：$\tilde{f} = f(x^*)$，由 lemma3.76，x^* 是优化问题：

$$\min_{x \in \mathbb{E}} \{ F(x) \equiv \max\{ g_0(x), \cdots, g_m(x) \} \}$$

的最优解，其中 $g_0(x) \equiv f(x) - \tilde{f}$. 由 $f, g_1, \cdots, g_m$ 的凸性

知 F 是convex的，由 Fermat's 最优性条件：

$$0 \in \partial F(x^*)$$

由 **Thm3.50** $\quad \partial F(x^*) = \operatorname{conv}\left(\bigcup_{i \in I(x^*)} \partial g_i(x^*) \right)$

其中 $I(x^*)=\{i\in\{0,1,\cdots,m\}:g_i(x^*)=0\}$，故 $\exists\lambda_i\geq 0$，

$i\in I(x^*)$，s.t. $\sum_{i\in I(x^*)}\lambda_i=1$，且

$$0\in\sum_{i\in I(x^*)}\lambda_i\partial g_i(x^*)$$

由 $g_0(x^*)=f(x^*)-\bar{f}=0$，故 $0\in I(x^*)$，故

$$0\in\lambda_0\partial f(x^*)+\sum_{i\in I(x^*)\setminus\{0\}}\lambda_i\partial g_i(x^*)$$

记 $\lambda_i=0$，$\forall i\in\{1,\cdots,m\}\setminus I(x^*)$，即证 $\square$

Slater 条件:
$\exists\bar{x}\in\mathbb{E}$，s.t. $g_i(\bar{x})<0$，$i=1,\cdots,m$

<u>Thm 3.78</u> (KKT 条件) 考虑问题

$$\begin{aligned}\min\ & f(x) \\ \text{s.t.}\ & g(x)\leq 0,\ i=1,\cdots,m\end{aligned}\qquad(3.96)$$

其中 $f,g_1,\cdots,g_m:\mathbb{E}\to\mathbb{R}$ 是 convex 的

(a) 设 x^* 是 (3.96) 的最优解，且设 Slater条件成立，则

$\exists \lambda_1, \cdots, \lambda_m \geq 0$, s.t.

$$0 \in \partial f(x^*) + \sum_{i=1}^{m} \lambda_i \partial g_i(x^*) \qquad (3.97)$$

$$\lambda_i g_i(x^*) = 0, \quad i=1,\cdots,m \qquad (3.98)$$

(b) 若 $x^* \in E$ 满足 (3.97), (3.98) 对 $\lambda_1, \cdots, \lambda_m \geq 0$，则 x^* 是 (3.96) 的最优解

证明：(a) 由Fritz-John条件，$\exists \tilde{\lambda}_0, \cdots, \tilde{\lambda}_m \geq 0$，不全为0，

s.t. $$0 \in \tilde{\lambda}_0 \partial f(x^*) + \sum_{i=1}^{m} \tilde{\lambda}_i \partial g_i(x^*)$$

$$\tilde{\lambda}_i g_i(x^*) = 0, \quad i=1,\cdots,m$$

下证 $\tilde{\lambda}_0 \neq 0$，用反证法设 $\tilde{\lambda}_0 = 0$，则 $\tilde{\lambda}_1, \cdots, \tilde{\lambda}_m$ 不全为0，

则由 (3.99) $$0 \in \sum_{i=1}^{m} \tilde{\lambda}_i \partial g_i(x^*)$$

则 $\exists \xi_i \in \partial g_i(x^*)$, $i=1,\cdots,m$, s.t. $\sum_{i=1}^{m} \tilde{\lambda}_i \xi_i = 0$

设$\bar{x}$是满足Slater条件的点，由对$\forall i=1,\cdots,m$

$$g_i(x^*)+\langle \xi_i, \bar{x}-x^*\rangle \leq g_i(\bar{x})$$

$$\Rightarrow \sum_{i=1}^{m}\tilde{\lambda}_i g_i(x^*)+\left\langle \sum_{i=1}^{m}\tilde{\lambda}_i\xi_i, \bar{x}-x^*\right\rangle \leq \sum_{i=1}^{m}\tilde{\lambda}_i g_i(\bar{x}),\ i=1,\cdots,m$$

故 $\sum_{i=1}^{m}\tilde{\lambda}_i g_i(\bar{x})\geq 0$，又由$\tilde{\lambda}_i\geq 0$, $g_i(\bar{x})<0$，知

$\tilde{\lambda}_1,\cdots,\tilde{\lambda}_m=0$，矛盾！故$\tilde{\lambda}_0>0$，取$\lambda_i=\frac{\tilde{\lambda}_i}{\tilde{\lambda}_0}$, $i=1,\cdots,m$

即证！

(b) 加条件：x^*是可行的

设$\hat{x}$是(3.96)的可行点，即$g_i(\hat{x})\leq 0, i=1,\cdots,m$

设 $$h(x)=f(x)+\sum_{i=1}^{m}\lambda_i g_i(x)$$

由(3.97)知 $0\in\partial h(x^*)$

由Fermat's最优性条件：x^*是h的最优解，故

$$f(x^*)=f(x^*)+\sum_{i=1}^{m}\lambda_i g_i(x^*)=h(x^*)\leq h(\hat{x})$$

$$= f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) \le f(x)$$

$\square$

§4 Conjugate Functions

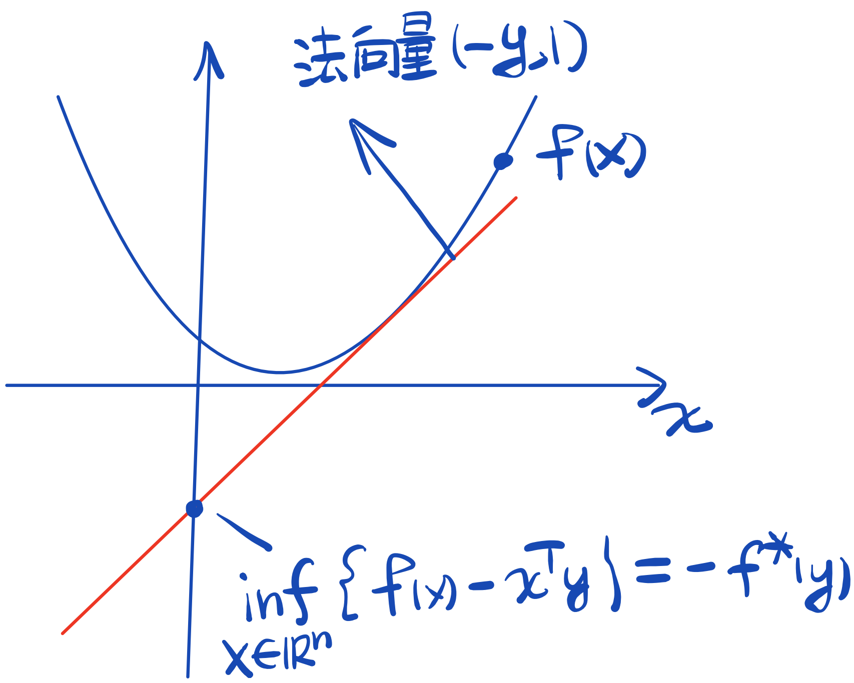
§4.1 Definition and Basic Properties

Def 4.1 (共轭func) 设 $f: \mathbb{E} \to [-\infty, \infty]$ 是 extended real-value func, 则 $f^*: \mathbb{E}^* \to [-\infty, \infty]$ 是:

$$f^*(y) = \max_{x \in \mathbb{E}} \{ \langle y, x \rangle - f(x) \}$$

称为共轭func.

几何解释



Example 4.2 设 $f=\delta_C$，其中 $C\subseteq \mathbb{E}$ 非空，则对 $\forall y \in \mathbb{E}^*$，有

$$f^*(y)=\max_{x\in\mathbb{E}}\{\langle y,x\rangle-\delta_C(x)\}=\max_{x\in C}\langle y,x\rangle=\sigma_C(y)$$

Thm 4.3 设 $f:\mathbb{E}\to(-\infty,\infty]$ 是 extended 实值 func. 则

f^* 是闭凸的

证明：f^* 是对仿射 func 的逐点取 max，由

Thm 2.16(c)，Thm 2.7(c)，f^* 是闭凸的 $\square$

Example 4.4 $f(x)=\frac{1}{2}\|x\|^2+\delta_C(x)$，则由 Example 2.17 知，$f^*(y)=\frac{1}{2}\|y\|^2-\frac{1}{2}d_C^2(y)$ $\square$

Thm 4.5 设 $f:\mathbb{E}\to(-\infty,\infty]$ 是 proper convex 的，则

f^*是proper的

证明：由f proper，则$\exists \hat{x}\in\mathbb{E}$，s.t. $f(\hat{x})<\infty$，由共轭func的定义，对$\forall y\in\mathbb{E}^*$，有

$$f^*(y)\geq\langle y,\hat{x}\rangle-f(\hat{x})>-\infty$$

下只需证$\exists g\in\mathbb{E}^*$，s.t. $f^*(g)<\infty$，由Corollary 3.19 $\exists x\in \mathrm{dom} f$，s.t. $\partial f(x)\neq\emptyset$，任取$g\in\partial f(x)$，则$\forall z\in\mathbb{E}$

$$f(z)\geq f(x)+\langle g,z-x\rangle$$

故 $$f^*(g)=\max_{z\in\mathbb{E}}\{\langle g,z\rangle-f(z)\}$$

$$\leq\langle g,x\rangle-f(x)<\infty \qquad \square$$

<u>Thm4.6</u> 设$f:\mathbb{E}\to(-\infty,\infty]$是proper的，则

$$f(x)+f^*(y)\geq\langle y,x\rangle,\ \forall x\in\mathbb{E},\ y\in\mathbb{E}^*$$

证明：由共轭func的定义：对$\forall x\in\mathbb{E}$，$y\in\mathbb{E}^*$

$$f^*(y) \geq \langle y, x\rangle - f(x)$$

由 f 的 proper 性质，$f^*(y) > -\infty$，$f(x) > -\infty$，故左右两边加 $f(x)$ 即证

避免出现 $f^*(y) + f(x)$ 是 $\infty - \infty$ 的未定义的形式

但若证明的结论是：$f^*(y) \geq \langle y, x\rangle - f(x)$

则不再需要 proper 的条件

$\boxtimes$

§4.2 The Biconjugate

$$f^{**}(x) = \max_{y \in \mathbb{E}^*} \{ \langle x, y\rangle - f^*(y) \}, \quad x \in \mathbb{E}$$

Lemma 4.7 ($f^{**} \leq f$) 设 $f: \mathbb{E} \to [-\infty, \infty]$，则 $f(x) \geq f^{**}(x)$

对 $\forall x \in \mathbb{E}$ 成立

证明：由共轭 func 的定义：$\forall x \in \mathbb{E}$, $y \in \mathbb{E}^*$

$$f^*(y) \geq \langle y, x\rangle - f(x)$$

故 $\quad f(x) \geq \langle y,x\rangle - f^*(y)$

故 $\quad f(x) \geq \max\limits_{y\in\mathbb{E}^*}\{\langle y,x\rangle - f^*(y)\} = f^{**}(y)$

<u>Thm4.8</u> $f:\mathbb{E}\to(-\infty,\infty]$ 是 proper, 闭凸func, 则 $f^{**}=f$

证明：由lemma4.7, $f^{**}\leq f$, 下证 $f^{**}\geq f$; 用反证法, 设 $\exists x\in\mathbb{E}$, s.t. $f^{**}(x)<f(x)$, 故

$$(x, f^{**}(x)) \notin \operatorname{epi}(f) \subseteq \mathbb{E}\times\mathbb{R}$$

在内积空间 $\mathbb{V}=\mathbb{E}\times\mathbb{R}$ 装备内积

$$\langle (u,s),(v,t)\rangle_{\mathbb{V}} = \langle u,v\rangle + st \quad .$$

由 f 是 proper 闭凸func, 故 $\operatorname{epi}(f)$ 是非空闭凸集, 故由 <u>Thm2.33</u>, $\exists a\in\mathbb{E}^*$, $b, c_1, c_2\in\mathbb{R}$, s.t.

$$\langle a,z\rangle + bs \leq c_1 < c_2 \leq \langle a,x\rangle + bf^{**}(x),\ \forall (z,s)\in\operatorname{epi}f$$

故 $\langle a, z-x\rangle + b(s-f^{**}(x)) \leq c_1-c_2 \equiv c < 0,\ \forall (z,s)\in\operatorname{epi}f$

b一定非正，则

- 若 $b<0$，则取 $y=-\frac{a}{b}$，有

$$\langle y, z-x\rangle - s + f^{**}(x) \le \frac{c}{-b} < 0, \ \forall (z,s)\in \mathrm{epi}(f)$$

取 $s=f(z)$，有

$$\langle y, z\rangle - f(z) - \langle y, x\rangle + f^{**}(x) \le \frac{c}{-b} < 0$$

关于 z 取 max：

$$f^{*}(y) - \langle y, x\rangle + f^{**}(x) \le \frac{c}{-b} < 0$$

5 Fenchel's 不等式

- 若 $b=0$，则取 $\hat{y}\in \mathrm{dom} f^{*}$（由 Thm 4.5，结合 f 的 proper 和凸性，这样的 $\hat{y}$ 是存在的）

设 $\varepsilon>0$，定义 $\hat{a}=a+\varepsilon\hat{y}$，$\hat{b}=-\varepsilon$，则 $\forall z\in \mathrm{dom} f$

$$\langle \hat{a}, z-x\rangle + \hat{b}(f(z) - f^{**}(x))$$

$$= \langle a, z-x\rangle + \varepsilon\left[\langle \hat{y}, z\rangle - f(z) + f^{**}(x) - \langle \hat{y}, x\rangle\right]$$

$$\leq c + \varepsilon\left[\langle y, z\rangle - f(z) + f^{**}(x) - \langle y, x\rangle\right]$$

$$\leq c + \varepsilon\left[f^{*}(y) - \langle y, x\rangle + f^{**}(x)\right]$$

故 $\langle \hat{a}, z-x\rangle + \hat{b}(f(z) - f^{**}(x)) \leq \hat{c}$

其中 $\hat{c} \equiv c+\varepsilon[f^{*}(y) - \langle y, x\rangle + f^{**}(x)]$，由 $c<0$，可取 $\varepsilon>0$ 充分小，s.t. $\hat{c}<0$，取 $\tilde{y} = -\frac{1}{\hat{b}}\hat{a}$.

$$\langle \tilde{y}, z\rangle - f(z) - \langle \tilde{y}, x\rangle + f^{**}(x) \leq -\frac{\hat{c}}{\hat{b}} < 0 \quad \forall z \in \operatorname{dom} f$$

对 z 取极大：

$$f^{*}(\tilde{y}) - \langle \tilde{y}, x\rangle + f^{**}(x) \leq \frac{\hat{c}}{-\hat{b}} < 0$$

与 Fenchel's 不等式 $\square$

Example 4.9 设 $C \subseteq \mathbb{E}$ 是给定非空集，由 $cl(convC)$ 是闭凸集，故

$$\sigma^*_{cl(convC)} = \left(\delta^*_{cl(convC)}\right)^* = \delta^{**}_{cl(convC)} = \delta_{cl(convC)}$$

又由 lemma 2.35：$\sigma_C = \sigma_{cl(convC)}$，故

$$\sigma^*_C = \delta_{cl(convC)} \qquad \boxed{/}$$

Example 4.10 $f(x) = \max\{x_1, \cdots, x_n\}$，则 $\forall x \in \mathbb{R}^n$

$$\max\{x_1, \cdots, x_n\} = \max_{y \in \Delta_n} y^T x = \sigma_{\Delta_n}(x)$$

故由 Example 4.9，且 Δ_n 的凸性，知

$$f^* = \delta_{\Delta_n} \qquad \boxed{/}$$

Example 4.11 设 C 是非空闭凸集，$f(x) = \frac{1}{2}\|x\|^2 - \frac{1}{2}d_C^2(x)$

由 Example 4.4，$f = g^*$，其中 $g(y) = \frac{1}{2}\|y\|^2 + \delta_C(y)$

由 C 的凸性，知 $g(y)$ 是 proper 闭凸集，知

$$f^*(y) = g^{**}(y) = g(y) = \frac{1}{2}\|y\|^2 + \delta_C(y) \qquad \boxtimes$$

§4.3 Conjugate Calculus Rules

Thm 4.12 设 $g: \mathbb{E}_1 \times \cdots \times \mathbb{E}_p \to (-\infty, \infty]$

$g(x_1, \cdots, x_p) = \sum_{i=1}^{p} f_i(x_i)$，其中 f_i 是 proper 的，则

$$g^*(y_1, \cdots, y_p) = \sum_{i=1}^{p} f_i^*(y_i), \quad \forall y_i \in \mathbb{E}_i^*, i = 1, \cdots, p$$

证明：$\forall (y_1, \cdots, y_p) \in \mathbb{E}_1^* \times \cdots \times \mathbb{E}_p^*$，成立

$$g^*(y_1, \cdots, y_p) = \max_{x_1, \cdots, x_p} \left\{ \langle (y_1, \cdots, y_p), (x_1, \cdots, x_p) \rangle - g(x_1, \cdots, x_p) \right\}$$

$$= \max_{x_1, \cdots, x_p} \left\{ \sum_{i=1}^{p} \langle y_i, x_i \rangle - \sum_{i=1}^{p} f_i(x_i) \right\}$$

$$= \sum_{i=1}^{p} \max_{x_i} \{ \langle y_i, x_i \rangle - f_i(x_i) \}$$

$$= \sum_{i=1}^{p} f_i^*(y_i) \qquad \boxed{/}$$

<u>Thm4.13</u> $f: \mathbb{E} \to (-\infty, \infty]$，$A: \mathbb{V} \to \mathbb{E}$是可逆线性映射，$a \in \mathbb{V}$，$b \in \mathbb{V}^*$，$c \in \mathbb{R}$，则设

$$g(x) = f(A(x-a)) + \langle b, x \rangle + c$$

有 $g^*(y) = f^*((A^T)^{-1}(y-b)) + \langle a, y \rangle - c - \langle a, b \rangle$

证明：设 $z = A(x-a)$，$\Rightarrow x = A^{-1}(z) + a$

$$g^*(y) = \max_{x} \{ \langle y, x \rangle - g(x) \}$$

$$= \max_{x} \{ \langle y, x \rangle - f(A(x-a)) - \langle b, x \rangle - c \}$$

$$= \max_{z} \{ \langle y, A^{-1}(z) + a \rangle - f(z) - \langle b, A^{-1}(z) + a \rangle - c \}$$

$$= \max_{z} \left\{ \langle y-b, A^{-1}(z) \rangle - f(z) + \langle a, y \rangle - \langle a, b \rangle - c \right\}$$

$$= f^*((A^T)^{-1}(y-b)) + \langle a, y \rangle - c - \langle a, b \rangle \qquad \boxed{/\!/}$$

Thm 4.14 $f: \mathbb{E} \to (-\infty, \infty]$, $\alpha \in \mathbb{R}_{++}$

(a) 设 $g(x) = \alpha f(x)$，则 $g^*(y) = \alpha f^*(\frac{y}{\alpha})$

(b) 设 $h(x) = \alpha f(\frac{x}{\alpha})$，则 $h^*(y) = \alpha f^*(y)$

证明:

(a)
$$g^*(y) = \max_{x} \left\{ \langle y, x \rangle - g(x) \right\}$$
$$= \max_{x} \left\{ \langle y, x \rangle - \alpha f(x) \right\}$$
$$= \alpha \max_{x} \left\{ \langle \frac{y}{\alpha}, x \rangle - f(x) \right\}$$
$$= \alpha f^*(\frac{y}{\alpha})$$

(b)
$$h^*(y) = \max_{x} \left\{ \langle y, x \rangle - h(x) \right\}$$
$$= \max_{x} \left\{ \langle y, x \rangle - \alpha f(\frac{x}{\alpha}) \right\}$$

$$= \alpha \max_x \left\{ \langle y, \frac{x}{\alpha} \rangle - f\left(\frac{x}{\alpha}\right) \right\}$$

$$= \alpha \max_z \left\{ \langle y, z \rangle - f(z) \right\}$$

$$= \alpha f^*(y)$$

§4.4 Example

§4.4.1 Exponent $f(x) = e^x$，定义 $0\log 0 \equiv 0$

则 $f^*(y) = \max\limits_x \{ xy - e^x \}$

$$= \begin{cases} y\log y - y, & y \geq 0 \\ \infty, & \text{else} \end{cases}$$

§4.4.2 Negative Log

$$f(x) = \begin{cases} -\log x, & x > 0 \\ \infty, & x \leq 0 \end{cases}$$

则对 $\forall y \in \mathbb{R}$

$$f^*(y) = \max_{x>0} \{xy + \log x\}$$

$$= \begin{cases} -1-\log(-y), & y<0 \\ \infty, & y \geqslant 0 \end{cases}$$

§ 4.4.3 Hinge Loss

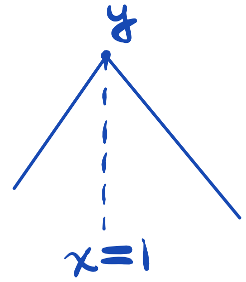
$$f(x) = \max\{1-x, 0\}$$

$$f^*(y) = \max_{x} \left[yx - \max\{1-x, 0\} \right]$$

$$= \max_{x} \left[\min\{(1+y)x-1, yx\} \right]$$

又由 $\min\{(1+y)x-1, yx\} = \begin{cases} (1+y)x-1, & x<1 \\ yx, & x \geqslant 1 \end{cases}$

• $\begin{cases} 1+y \geqslant 0 \\ y \leq 0 \end{cases}$ 时



$f^*(y) = y$

• $\begin{cases} 1+y \leq 0 \\ y \leq 0 \end{cases}$ 或 $\begin{cases} 1+y \geqslant 0 \\ y \geqslant 0 \end{cases}$ 时, $f^*(y) = \infty$

综上 $f^*(y) = y + \delta_{[-1,0]}(y)$

§ 4.4.4 $f(x) = \frac{1}{p}|x|^p \quad (p>1)$

易证 $f^*(y) = \frac{1}{q}|x|^q$, 其中 $\frac{1}{p} + \frac{1}{q} = 1$

§ 4.4.5

$$f(x) = \begin{cases} -\frac{x^p}{p}, & x \geq 0 \\ \infty, & x < 0 \end{cases}$$

同理, $f^*(y) = \begin{cases} -\frac{(-y)^q}{q}, & y < 0 \\ \infty, & \text{else} \end{cases}$

§ 4.4.6 $f(x) = \frac{1}{2}x^T A x + b^T x + c$, $A \in S_{++}^n$

$$f^*(y) = \frac{1}{2}(y-b)^T A^{-1}(y-b) - c$$

§ 4.4.7 $f(x) = \frac{1}{2}x^T A x + b^T x + c$, $A \in S_+^n$

则对 $\forall y\in\mathbb{R}^n$

$$f^*(y)=\max_x\left\{y^Tx-f(x)\right\}$$

$$=\max_x\left\{g(x)\equiv -\frac{1}{2}x^TAx+(y-b)^Tx-c\right\}$$

由 g 的 concave 和可微性，则 $\nabla g(x)=0$ 时取最小

$$Ax=y-b \quad (4.9)$$

• 当 $y\in b+\text{Range}(A)$，可以取 (4.9) 的一个解

$$\tilde{x}=A^+(y-b)$$

则 $f^*(y)=-\frac{1}{2}\tilde{x}^TA\tilde{x}-(b-y)^T\tilde{x}-c$

$$=-\frac{1}{2}(y-b)A^+AA^+(y-b)-(b-y)^TA^+(y-b)-c$$

$$=\frac{1}{2}(y-b)^TA^+(y-b)-c$$

• 当 $y-b\notin\text{Range}(A)$ 时，下证此时 $f^*(y)=\infty$

由 $\text{Range}(A)=\text{Null}(A)^\perp$，故 $y-b\notin\text{Null}(A)^\perp$，故

$\exists\, v \in Null(A)$, s.t. $(y-b)^T v > 0$, 由 $\forall \alpha \in \mathbb{R}$

$$g(\alpha v) = \alpha (y-b)^T v - c$$

故 $g(\alpha v) \to \infty \ (\alpha \to \infty)$

综上 $$f^*(y) = \begin{cases} \frac{1}{2}(y-b)^T A^+ (y-b) - c, & y \in b + Range(A). \\ \infty, & else \end{cases}$$

§4.4.8 Negative Entropy

$$f(x) = \begin{cases} \sum_{i=1}^n x_i \log x_i, & x \geq 0 \\ \infty, & else \end{cases}$$

易证 $f^*(y) = \sum_{i=1}^n e^{y_i - 1}$ ☑

§4.4.9 Negative Sum of Logs

$$f(x) = \begin{cases} -\sum_{i=1}^n \log x_i, & x > 0 \\ \infty, & else \end{cases}$$

则 $$f^*(y)=\begin{cases} -n-\sum_{i=1}^{n}\log(-y_i), & y<0 \\ \infty, & \text{else} \end{cases}$$

§ 4.4.10 Negative Entropy over the Unit Simplex

$$f(x)=\begin{cases} \sum_{i=1}^{n} x_i \log x_i, & x\in\Delta_n \\ \infty, & \text{else} \end{cases}$$

则 $f^*(y)=\log\left(\sum_{j=1}^{n} e^{y_j}\right)$

§ 4.4.11 log-sum-exp

$$g(x)=\log\left(\sum_{j=1}^{n} e^{x_j}\right)$$

则由§4.4.10知 $g^*(y)=\begin{cases} \sum_{i=1}^{n} y_i \log y_i, & y\in\Delta_n \\ \infty, & \text{else} \end{cases}$

§ 4.4.12 $f(x)=\|x\|$

则 $f^*(y)=\delta_{B_{\|\cdot\|_*}[0,1]}(y)=\begin{cases}0, & \|y_*\|\le 1\\ \infty, & \text{else}\end{cases}$

§ 4.4.13 Ball-Pen

设 $f(x)=\begin{cases}-\sqrt{1-\|x\|^2}, & \|x\|\le 1\\ \infty, & \text{else}\end{cases}$

$$f^*(y)=\max_x\left\{\langle y,x\rangle+\sqrt{1-\|x\|^2}:\|x\|\le 1\right\}$$

$$=\max_{\alpha\in[0,1]}\max_{x:\|x\|=\alpha}\left\{\langle y,x\rangle+\sqrt{1-\alpha^2}\right\}$$

$$=\max_{\alpha\in[0,1]}\left\{g(\alpha)\equiv\alpha\|y\|_*+\sqrt{1-\alpha^2}\right\}$$

易知 $\tilde{\alpha}=\dfrac{\|y\|_*}{\sqrt{\|y\|_*^2+1}}$ 时，$g(\alpha)$ 取 max，从而

$$f^*(y)=\sqrt{\|y\|_*^2+1}$$

一般地，$f_\alpha(x)=\begin{cases}-\sqrt{\alpha^2-\|x\|^2}, & \|x\|\le\alpha\\ \infty, & \text{else}\end{cases}$

由 $f_\alpha(x)=\alpha f(\frac{x}{\alpha})$ 知

$$f_\alpha^*(y)=\alpha f^*(y)=\alpha\sqrt{1+\|y\|_*^2} \qquad \boxed{/}$$

§ 4.4.14 $g_\alpha(x)=\sqrt{\alpha^2+\|x\|^2}, \alpha>0$

易证 $g_\alpha^*(y)=\begin{cases}-\alpha\sqrt{1-\|y\|_*^2}, & \|y\|_*\le 1\\ \infty, & \text{else}\end{cases} \qquad \boxed{/}$

§ 4.4.15 Squared Norm

$f(x)=\frac{1}{2}\|x\|^2$ 则 $f^*(y)=\frac{1}{2}\|y\|_*^2 \qquad \boxed{/}$

§ 4.4.17 Fenchel's Duality Theorem

$$(P) \quad \min_{x\in\mathbb{E}} f(x)+g(x)$$

则 (P) 等价于: $\min\limits_{x,z\in\mathbb{E}}\{f(x)+g(z): x=z\}$

Lagrangian:

$$L(x,z;y) = f(x) + g(z) + \langle y, z-x \rangle$$
$$= -[\langle y,x\rangle - f(x)] - [\langle -y,z\rangle - g(z)]$$

dual function:

$$q(y) = \min_{x,z} L(x,z;y) = -f^*(y) - g^*(-y)$$

Fenchel's dual 问题:

$$(D) \quad \max_{y\in\mathbb{E}^*} \{-f^*(y) - g^*(-y)\}$$

Thm 4.15 (Fenchel's duality theorem)

设 $f, g: \mathbb{E} \to (-\infty, \infty]$ 是 proper convex 函数, 若 $\mathrm{ri}(\mathrm{dom} f) \cap \mathrm{ri}(\mathrm{dom}\, g) \neq \emptyset$, 则

$$\min_{x\in\mathbb{E}} \{f(x) + g(x)\} = \max_{y\in\mathbb{E}^*} \{-f^*(y) - g^*(-y)\}$$

且当右侧优化问题有限时，maximum 是可达的

证明：此定理可以看作 Thm A.1 的简单版本

$\square$

§ 4.5 Infimal Convolution and Conjugacy

Thm 4.16 (conjugate of infimal convolution)

对 proper func $h_1, h_2 : \mathbb{E} \to (-\infty, \infty]$，有

$$(h_1 \square h_2)^* = h_1^* + h_2^*$$

证明：对 $\forall y \in \mathbb{E}^*$，有

$$(h_1 \square h_2)^*(y) = \max_{x \in \mathbb{E}} \left\{ \langle y, x \rangle - (h_1 \square h_2)(x) \right\}$$

$$= \max_{x \in \mathbb{E}} \left\{ \langle y, x \rangle - \min_{u \in \mathbb{E}} \left\{ h_1(u) + h_2(x-u) \right\} \right\}$$

$$= \max_{x \in \mathbb{E}} \max_{u \in \mathbb{E}} \left\{ \langle y, x \rangle - h_1(u) - h_2(x-u) \right\}$$

$$= \max_{x \in \mathbb{E}} \max_{u \in \mathbb{E}} \left\{ \langle y, x-u \rangle + \langle y, u \rangle - h_1(u) - h_2(x-u) \right\}$$

$$= \max_{u \in \mathbb{E}} \max_{x \in \mathbb{E}} \left\{ \langle y, x-u \rangle - h_2(x-u) + \langle y, u \rangle - h_1(u) \right\}$$

$$= \max_{u\in\mathbb{E}} \left\{ h_2^*(y) + \langle y, u\rangle - h_1(u) \right\}$$

$$= h_1^*(y) + h_2^*(y)$$ $\boxed{/}$

Thm 4.17 (conjugate of sum)

设 $h_1: \mathbb{E} \to (-\infty, \infty]$ 是 proper, convex 的，$h_2: \mathbb{E} \to \mathbb{R}$ 是实值凸的，则 $(h_1+h_2)^* = h_1^* \,\square\, h_2^*$

证明：$\forall\, y \in \mathbb{E}^*$

$$(h_1+h_2)^*(y) = \max_{x\in\mathbb{E}} \left\{ \langle y, x\rangle - h_1(x) - h_2(x) \right\}$$

$$= -\min_{x\in\mathbb{E}} \left\{ h_1(x) + g(x) \right\}$$

其中 $g(x) \equiv h_2(x) - \langle y, x\rangle$，注意到

$$\mathrm{ri}(\mathrm{dom} h_1) \cap \mathrm{ri}(\mathrm{dom} g) = \mathrm{ri}(\mathrm{dom} h_1) \cap \mathbb{E} = \mathrm{ri}(\mathrm{dom} h_1) \overset{\text{Thm 3.17}}{\neq} \emptyset$$

故由 Fenchel's 不等式：

$$\min_{x\in\mathbb{E}} \left\{ h_1(x) + g(x) \right\} = \max_{z\in\mathbb{E}^*} \left\{ -h_1^*(z) - g^*(-z) \right\}$$

$$= \max_{z\in\mathbb{E}^*}\left\{-h_1^*(z) - \max_{x\in\mathbb{E}}\left\{\langle -z, x\rangle - g(x)\right\}\right\}$$

$$= \max_{z\in\mathbb{E}^*}\left\{-h_1^*(z) - \max_{x\in\mathbb{E}}\left\{\langle y-z, x\rangle - h_2(x)\right\}\right\}$$

$$= \max_{z\in\mathbb{E}^*}\left\{-h_1^*(z) - h_2^*(y-z)\right\}$$

故 $(h_1+h_2)^*(y) = \min_{z\in\mathbb{E}^*}\left\{h_1^*(z) + h_2^*(y-z)\right\}$

$$= (h_1^* \square h_2^*)(y) \qquad \boxed{/}$$

<u>Corollary 4.18</u> 设 h_1 是 proper 闭凸的，h_2 是实值凸的，

则 $h_1 + h_2 = (h_1^* \square h_2^*)^*$

证明：由 <u>Thm 2.21</u>，h_2 是连续的，故 h_2 闭，从而

h_1+h_2 是 proper，闭凸 func，由 <u>Thm 4.8</u>，<u>Thm 4.17</u>：

$$h_1+h_2 = (h_1+h_2)^{**} = (h_1^* \square h_2^*)^* \qquad \boxed{/}$$

Thm 4.19 设 h_1 proper, convex, h_2 实值的 func, 设 $h_1 \square h_2$ 是实值 func, 则 $h_1 \square h_2 = (h_1^* + h_2^*)^*$

证明: 由 Thm 4.16 $(h_1 \square h_2)^* = h_1^* + h_2^*$

由 Thm 2.19, $h_1 \square h_2$ 是凸的, 又由 $h_1 \square h_2$ 是实值的.

故 $h_1 \square h_2$ 是连续的, 从而 $h_1 \square h_2$ 是闭凸的, 利用

Thm 4.8 知, $(h_1 \square h_2)^{**} = h_1 \square h_2$, 故得证 $\boxtimes$

§4.6 Subdifferential of Conjugate functions

Thm 4.20 设 $f: \mathbb{E} \to (-\infty, \infty]$ 是 proper, convex 的

对 $\forall x \in \mathbb{E}$, $y \in \mathbb{E}^*$, 以下等价

(i) $\langle x, y \rangle = f(x) + f^*(y)$

(ii) $y \in \partial f(x)$

若 f 是闭的，则 (i)(ii) 与 (iii) 等价

(iii) $x \in \partial f^*(y)$.

证明：$y \in \partial f(x) \Leftrightarrow$

$$f(z) \geqslant f(x) + \langle y, z-x \rangle \quad \forall z \in \mathbb{E} \quad \Leftrightarrow$$

$$\langle y, x \rangle - f(x) \geqslant \langle y, z \rangle - f(z), \ \forall z \in \mathbb{E}$$

对右边对 z 取 max，上式 $\Leftrightarrow$

$$\langle y, x \rangle - f(x) \geqslant \max_{z \in \mathbb{E}} \langle y, z \rangle - f(z) = f^*(y)$$

又由 Fenchel's 不等式，知上式 $\Leftrightarrow$

$$\langle x, y \rangle = f(x) + f^*(y)$$

以下设 f 是闭的，由 Thm4.8，$f^{**} = f$，取 $g = f^*$，则

(i) $\Leftrightarrow$ $\langle x, y \rangle = g(y) + g^*(x)$

由 (i) $\Leftrightarrow$ (ii) 知，上式 $\Leftrightarrow$ $x \in \partial g(y) = \partial f^*(y)$ $\quad\square$

Corollary 4.21 设 $f:\mathbb{E}\to(-\infty,\infty]$ 是 proper，闭凸的

则对 $\forall x\in\mathbb{E}, y\in\mathbb{E}^*$，有

$$\partial f(x)=\underset{\tilde{y}\in\mathbb{E}^*}{\operatorname{argmax}}\left\{\langle x,\tilde{y}\rangle-f^*(\tilde{y})\right\}$$

$$\partial f^*(y)=\underset{\tilde{x}\in\mathbb{E}}{\operatorname{argmax}}\left\{\langle y,\tilde{x}\rangle-f(\tilde{x})\right\}.$$

证明：Thm 4.20 (i) $\Leftrightarrow$

$x\in\underset{\tilde{x}\in\mathbb{E}}{\operatorname{argmax}}\left\{\langle y,\tilde{x}\rangle-f(\tilde{x})\right\}$，且由 (i) (iii) 等价

知 $\partial f(x)=\underset{\tilde{x}\in\mathbb{E}}{\operatorname{argmax}}\left\{\langle y,\tilde{x}\rangle-f(\tilde{x})\right\}$

对 f^* 做相同的操作，注意到 $f=f^{**}$，即证 $\square$

Example 4.22 $f:\mathbb{E}\to\mathbb{R}$ $f(x)=\|x\|$

由 Example 2.31 知 $f=\sigma_{B_{\|\cdot\|_*}[0,1]}$

故 $f^* = \delta_{B_{\|\cdot\|_*}[0,1]}$，则

$$\partial f(0) = \underset{y\in\mathbb{E}^*}{\operatorname{argmin}} f^*(y) = \underset{y\in\mathbb{E}^*}{\operatorname{argmin}} \delta_{B_{\|\cdot\|_*}[0,1]} = B_{\|\cdot\|_*}[0,1]$$

$\square$

<u>Thm 4.23</u> $f:\mathbb{E}\to\mathbb{R}$ 凸，对给定 $L>0$，以下等价：

(i) $|f(x)-f(y)| \leq L\|x-y\|$，$\forall x,y\in\mathbb{E}$

(ii) $\|g\|_* \leq L$，对 $\forall g\in\partial f(x)$，$x\in\mathbb{E}$

(iii) $\operatorname{dom} f^* \subseteq B_{\|\cdot\|_*}[0,L]$

证明：<u>Thm 3.61</u> 证明了 (i) (ii) 等价

先证 (iii) $\Rightarrow$ (ii)，由 Corollary 4.21：对 $\forall x\in\mathbb{E}$

$$\partial f(x) = \underset{y\in\mathbb{E}^*}{\operatorname{argmax}} \{\langle x,y\rangle - f^*(y)\}$$

故 $\partial f(x) \subseteq \operatorname{dom} f^*$，从而 $\partial f(x) \subseteq B_{\|\cdot\|_*}[0,L]$ $\forall x$

从而 $\forall g \in \partial f(x)$，有 $\|g\|_* \le L$，(ii) 得证

再证 (i) $\Rightarrow$ (iii)，由 (i) 成立，故

$$f(x) - f(0) \le |f(x) - f(0)| \le L\|x\|$$

故 $-f(x) \geqslant -f(0) - L\|x\|$，对 $\forall y \in \mathbb{E}^*$

$$f^*(y) = \max_{x \in \mathbb{E}} \{\langle x, y\rangle - f(x)\}$$

$$\geqslant \max_{x \in \mathbb{E}} \{\langle x, y\rangle - f(0) - L\|x\|\}$$

$\forall$ 取 $\tilde{y} \in \mathbb{E}^*$, s.t. $\|\tilde{y}\|_* > L$，下证 $\tilde{y} \notin \operatorname{dom} f^*$

取 $y^\dagger \in \mathbb{E}$, s.t. $\|y^\dagger\| = 1$，且 $\langle \tilde{y}, y^\dagger\rangle = \|\tilde{y}\|_*$

令 $C = \{\alpha y^\dagger : \alpha \geqslant 0\} \subseteq \mathbb{E}$，有

$$f^*(\tilde{y}) \geqslant \max_{x \in \mathbb{E}} \{\langle x, \tilde{y}\rangle - f(0) - L\|x\|\}$$

$$\geqslant \max_{x \in C} \{\langle x, \tilde{y}\rangle - f(0) - L\|x\|\}$$

$$= \max_{\alpha \geqslant 0} \{\langle \alpha \tilde{y}, y^\dagger\rangle - f(0) - L\alpha\|y^\dagger\|\}$$

$$= \max_{\alpha \geq 0} \{\alpha \|\tilde{y}\|_* - f(0) - L\alpha\}$$

$$= \max_{\alpha \geq 0} \{\alpha(\|\tilde{y}\|_* - L) - f(0)\}$$

$$= \infty$$

故 $\tilde{y} \in \mathrm{dom} f^*$ $\boxed{\checkmark}$

§5.1 L-smooth function

<u>**Def 5.1**</u> 令 $L \geq 0$, $f: \mathbb{E} \to (-\infty, \infty]$ 是 L-smooth 的于 D.

若 f 可微, 且 $\|\nabla f(x) - \nabla f(y)\|_* \leq L\|x-y\|$, $\forall x, y \in D$

<u>**Example 5.2**</u> $f(x) = \frac{1}{2} x^T A x + b^T x + c$, $A \in \mathbb{S}^n$

$$\|\nabla f(x) - \nabla f(y)\|_q = \|Ax - Ay\|_q \leq \|A\|_{p,q} \|x-y\|_p$$

其中 $\|A\|_{p,q} = \max\{\|Ax\|_q : \|x\|_p \leq 1\}$

其中 $q\in[1,\infty]$, s.t. $\frac{1}{p}+\frac{1}{q}=1$，故 f 是 $\|A\|_{p,q}$-smooth

下证 $\|A\|_{p,q}$ 是最小的smoothness参数：

设 f 是 L-smooth的，取 $\tilde{x}$，s.t. $\|\tilde{x}\|_p=1$，$\|A\tilde{x}\|_q=\|A\|_{p,q}$

则：$\|A\|_{p,q}=\|A\tilde{x}\|_q=\|\nabla f(\tilde{x})-\nabla f(0)\|_q\le L\|\tilde{x}\|_p=L$

从而 $L\ge\|A\|_{p,q}$ $\square$

Example 5.3 $\quad f(x)=\langle b,x\rangle+c$

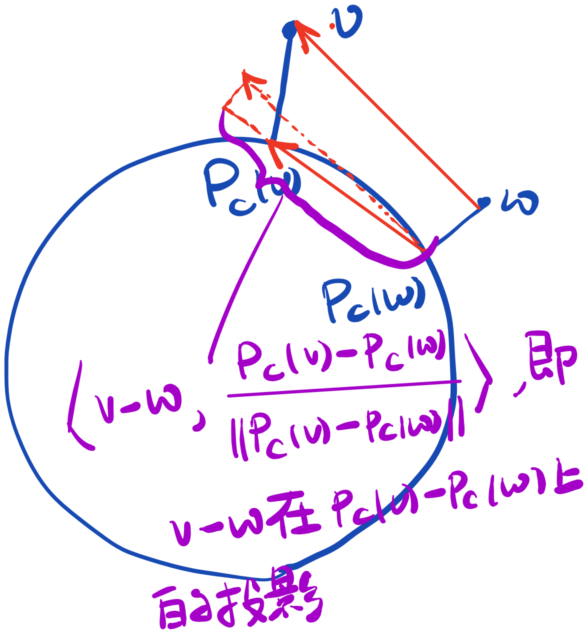
则 f 是 0-smooth的

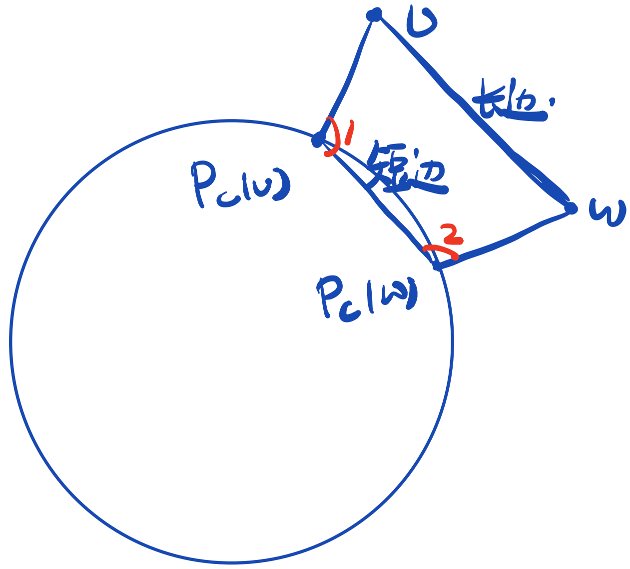
Thm 5.4 设 $C\subseteq E$ 非空闭凸集

(a) $\langle P_C(v)-P_C(w),v-w\rangle\ge\|P_C(v)-P_C(w)\|^2$

(b) $\quad\|P_C(v)-P_C(w)\|\le\|v-w\|$

证明：定理证明见Nesterov §2.2





注：几何上看定理是显然的，由∠1，∠2都是钝角！

Example5.5 $C \subseteq \mathbb{E}$ 非空闭凸，$\varphi_C(x)=\frac{1}{2}d_C^2(x)$

Example3.31 证明了：$\nabla \varphi_C(x) = x - P_C(x)$

下证：φ_C 是1-smooth的：对$\forall x, y \in \mathbb{E}$

$$\|\nabla \varphi_C(x) - \nabla \varphi_C(y)\|^2 = \|x-y-P_C(x)+P_C(y)\|^2$$

$$= \|x-y\|^2 - 2\langle P_C(x)-P_C(y), y-y\rangle + \|P_C(x)-P_C(y)\|^2$$

$$\leq \|x-y\|^2 - 2\|P_C(x)-P_C(y)\|^2 + \|P_C(x)-P_C(y)\|^2$$

$$\leq \|x-y\|^2 \quad \square$$

§ 5.1.1 The descent Lemma

<u>Lemma 5.7</u> $f: \mathbb{E} \to (-\infty, \infty]$ 是 L-smooth 的 于凸集 D，则对 $\forall x, y \in D$，有

$$f(y) \leq f(x) + \langle \nabla f(x), y-x \rangle + \frac{L}{2} \|x-y\|^2$$

证明：由 N-L 定理：

$$f(y) - f(x) = \int_0^1 \langle \nabla f(x + t(y-x)), y-x \rangle dt$$

故

$$f(y) - f(x) = \langle \nabla f(x), y-x \rangle + \int_0^1 \langle \nabla f(x+t(y-x)) - \nabla f(x), y-x \rangle dt$$

$$\Rightarrow \quad \left| f(y) - f(x) - \langle \nabla f(x), y-x \rangle \right|$$

$$= \left| \int_0^1 \langle \nabla f(x+t(y-x)) - \nabla f(x), y-x \rangle dt \right|$$

$$\leq \int_0^1 \left| \langle \nabla f(x+t(y-x)) - \nabla f(x), y-x \rangle \right| dt$$

$$\le \int_0^1 \|\nabla f(x+t(y-x))-\nabla f(x)\|_* \|y-x\| \, dt$$

$$\le \int_0^1 tL\|y-x\|^2 dt = \frac{L}{2}\|y-x\|^2 \qquad \boxed{/\!/}$$

§5.1.2 Characterize of L-smooth func

<u>Thm 5.8</u> $f: \mathbb{E} \to \mathbb{R}$ 凸，可微，$L>0$，以下等价：

(i) f L-smooth

(ii) $f(y) \le f(x) + \langle \nabla f(x), y-x \rangle + \frac{L}{2}\|y-x\|^2 \quad \forall x, y$

(iii) $f(y) \ge f(x) + \langle \nabla f(x), y-x \rangle + \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|_*^2 \quad \forall x, y$

(iv) $\langle \nabla f(x) - \nabla f(y), x-y \rangle \ge \frac{1}{L}\|\nabla f(x) - \nabla f(y)\|_*^2 \quad \forall x, y$

(v) $f(\lambda x + (1-\lambda)y) \ge \lambda f(x) + (1-\lambda) f(y) - \frac{L}{2}\lambda(1-\lambda)\|x-y\|^2$
$\forall x, y$

证明：

(i) $\Rightarrow$ (ii) Lemma 5.7 已证

(ii) $\Rightarrow$ (iii) 设 $\nabla f(y) \ne \nabla f(x)$，否则 (iii) 是自然的（由 f 凸）

固定 $x\in\mathbb{E}$，考虑

$$g_x(y)=f(y)-f(x)-\langle\nabla f(x),y-x\rangle,\quad y\in\mathbb{E}$$

则对 $\forall y,z\in\mathbb{E}$

$$\begin{aligned}
g_x(z)&=f(z)-f(x)-\langle\nabla f(x),z-x\rangle\\
&\le f(y)+\langle\nabla f(y),z-y\rangle+\frac{L}{2}\|z-y\|^2-f(x)-\langle\nabla f(x),z-x\rangle\\
&=f(y)-f(x)-\langle\nabla f(x),y-x\rangle+\langle\nabla f(y)-\nabla f(x),z-y\rangle+\frac{L}{2}\|z-y\|^2\\
&=g_x(y)+\langle\nabla g_x(y).\,y-x\rangle+\frac{L}{2}\|z-y\|^2
\end{aligned}$$

即 g_x 满足性质(ii)，由 g_x 凸，且 $\nabla g_x(x)=0$，由Fermat：

$$g_x(x)\le g_x(z)\quad\forall z\in\mathbb{E}$$

设 $y\in\mathbb{E}$，$v\in\mathbb{E}$，s.t. $\|v\|=1$ 且 $\langle\nabla g_x(y),v\rangle=\|\nabla g_x(y)\|_*$

$$令\ z=y-\frac{\|\nabla g_x(y)\|_*}{L}v$$

代入上式：$0=g_x(x)\le g_x\left(y-\dfrac{\|\nabla g_x(y)\|_*}{L}v\right)$

$$\leq g_x(y) - \frac{\|\nabla g_x(y)\|_*}{L}\langle \nabla g_x(y), v\rangle + \frac{1}{2L}\|\nabla g_x(y)\|_*^2\|v\|^2$$

$$= g_x(y) - \frac{1}{2L}\|\nabla g_x(y)\|_*$$

$$= f(y) - f(x) - \langle \nabla f(x), y-x\rangle - \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_*^2$$

(iii) $\Rightarrow$ (iv) 由 (iii)

$$f(y) \geqslant f(x) + \langle \nabla f(x), y-x\rangle + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_*^2 \quad ①$$

$$f(x) \geqslant f(y) + \langle \nabla f(y), x-y\rangle + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_*^2 \quad ②$$

① + ② $\Rightarrow$ (iv)

(iv) $\Rightarrow$ (i) Lipschitz 条件：

$$\|\nabla f(x) - \nabla f(y)\|_* \leq L\|x-y\|$$

当 $\nabla f(x) = \nabla f(y)$ 时，是自然的，下设 $\nabla f(x) \neq \nabla f(y)$

由 (iv) + CS 不等式：$\forall x, y \in \mathbb{E}$

$$\|\nabla f(x) - \nabla f(y)\|_*\|x-y\| \geqslant \langle \nabla f(x) - \nabla f(y), x-y\rangle$$

$$\overset{(iv)}{\geqslant} \frac{1}{L}\|\nabla f(x) - \nabla f(y)\|_*^2$$

$\Rightarrow \quad \|\nabla f(x)-\nabla f(y)\|_* \le L\|x-y\|$

证明：(i)⇒(ii)⇒(iii)⇒(iv)⇒(i)，下证 (ii)⇔(v)

(ii)⇒(v) 设 $x, y\in E$, $\lambda\in[0,1]$，记 $x_\lambda=\lambda x+(1-\lambda)y$，(ii)⇒

$$f(x)\le f(x_\lambda)+\langle\nabla f(x_\lambda), x-x_\lambda\rangle+\frac{L}{2}\|x-x_\lambda\|^2$$

$$f(y)\le f(x_\lambda)+\langle\nabla f(x_\lambda), y-x_\lambda\rangle+\frac{L}{2}\|y-x_\lambda\|^2$$

$\Leftrightarrow$

$$f(x)\le f(x_\lambda)+(1-\lambda)\langle\nabla f(x_\lambda), x-y\rangle+\frac{L(1-\lambda)^2}{2}\|x-y\|^2 \quad ③$$

$$f(y)\le f(x_\lambda)+\lambda\langle\nabla f(x_\lambda), y-x\rangle+\frac{L\lambda^2}{2}\|x-y\|^2 \quad ④$$

③$\cdot\lambda$+④$\cdot(1-\lambda)\cdot\Rightarrow$ (v)

(v)⇒(ii) (v) ⇔

$$f(y)\le f(x)+\frac{f(x+(1-\lambda)(y-x))-f(x)}{1-\lambda}+\frac{L}{2}\lambda\|x-y\|^2$$

取 $\lambda\to1^-$： $f(y)\le f(x)+f'(x;y-x)+\frac{L}{2}\|x-y\|^2$

$$=f(x)+\langle\nabla f(x), y-x\rangle+\frac{L}{2}\|y-x\|^2 \quad \square$$

Remark 5.9：Thm 5.8 对 f 凸性的假设是必要的.

如 $f(x)=-\frac{1}{2}\|x\|^2$ 是 1-smooth 但不是 L-smooth $(L<1)$

但由 f 的凹性，有：

$$f(y)\leq f(x)+\langle\nabla f(x),y-x\rangle$$

故对 f 凸性假设是必要的

Thm 5.10 设 $f:U\to\mathbb{R}$ 是 C^2 的在开集 $U\subseteq\mathbb{R}^n$，令 $x\in U$，$r>0$，s.t. $B(x,r)\subseteq U$，则对 $\forall y\in B(x,r)$，$\exists\xi\in[x,y]$，

s.t. $f(y)=f(x)+\nabla f(x)^T(y-x)+\frac{1}{2}(y-x)^T\nabla^2 f(\xi)(y-x)$

证明：对 $\forall x\in U$，$\forall y\in B(x,r)\subseteq U$，由 U 开，知 $\exists\delta>0$

$x+t(y-x)\in U$，对 $\forall t\in(-\delta,1+\delta)$ 成立

定义 $\varphi:\underbrace{(-\delta,1+\delta)}^{J}\to\mathbb{R}$，$\varphi(t)=f(x+t(y-x))$ 于 J 可微

$$\varphi'(t)=\nabla f(x+t(y-x))^T(y-x)$$

且对 $\forall t\in J$, $\varphi''(t)$存在,且

$$\varphi''(t)=\lim_{s\to 0}\frac{1}{s}\left[\nabla f(x+(t+s)(y-x))^T(y-x)-\nabla f(x+t(y-x))^T(y-x)\right]$$

$$=(y-x)^T\nabla^2 f(x+t(y-x))(y-x)$$

综上: $\varphi:[0,1]\to\mathbb{R}$ 在$(0,1)$上 C^2, φ, φ' 在$[0,1]$连续.

用一维func的中值定理:

$$f(y)-f(x)-\nabla f(x)^T(y-x)=\varphi(1)-\varphi(0)-\varphi'(0)$$

$$=\frac{1}{2}(y-x)^T\nabla^2 f(x+t(y-x))(y-x)$$

对某个 $t\in(0,1)$ 成立 $\square$

<u>Example 5.11</u> $f(x)=\frac{1}{2}\|x\|_p^2=\frac{1}{2}\left(\sum_{i=1}^n|x_i|^p\right)^{\frac{2}{p}}$, $p\in[2,\infty)$

下证: f 是 $(p-1)$-smooth w.r.t. ℓ_p-norm

- $p=2$时, Example 5.2 已经证明

· $p>2$时：

当 $x\neq 0$ 时

$$\frac{\partial f}{\partial x_i}(x)=\frac{1}{p}\left(\sum_{i=1}^{n}|x_i|^p\right)^{\frac{2-p}{p}}\cdot \operatorname{sgn}(x_i)\cdot p|x_i|^{p-1}$$

$$=\operatorname{sgn}(x_i)\frac{|x_i|^{p-1}}{\|x\|_p^{p-2}}$$

当 $x=0$ 时

$$\frac{\partial f}{\partial x_i}(0)=\lim_{\substack{s_i\to 0\\ s_j=0,j\neq i}}\frac{\frac{1}{2}\|s\|_p^2}{s_i}=\lim_{\substack{s_i\to 0\\ s_j=0,j\neq i}}\frac{\frac{1}{2}|s_i|^2}{s_i}=0$$

故 $$\frac{\partial f}{\partial x_i}(x)=\begin{cases}\operatorname{sgn}(x_i)\dfrac{|x_i|^{p-1}}{\|x\|_p^{p-2}}, & x\neq 0\\ 0, & x=0\end{cases}$$

故偏导数在 $\mathbb{R}^n$ 连续 $\Rightarrow f$ 在 $\mathbb{R}^n$ 上可微

对 $\forall x\neq 0$，有

$$\frac{\partial^2 f}{\partial x_i\partial x_j}(x)=\begin{cases}(2-p)\operatorname{sgn}(x_i)\operatorname{sgn}(x_j)\dfrac{|x_i|^{p-1}|x_j|^{p-1}}{\|x\|_p^{2p-2}}, & i\neq j\\ (p-1)\dfrac{|x_i|^{p-2}}{\|x\|_p^{p-2}}+(2-p)\dfrac{|x_i|^{2p-2}}{\|x\|_p^{2p-2}}, & i=j\end{cases}$$

设 $x, y \in \mathbb{R}^n$，且 $0 \notin [x, y]$，由 Thm 5.10，取 $U \supseteq [x, y]$ 开

且 $0 \notin U$，则 $\exists \xi \in [x, y]$. s.t.

$$f(y) = f(x) + \nabla f(x)^T (y-x) + \frac{1}{2}(y-x)^T \nabla^2 f(\xi)(y-x)$$

下证：对 $\forall d \in \mathbb{R}^n$，有 $d^T \nabla^2 f(\xi) d \leq (p-1)\|d\|_p^2$

注意到：对 $\forall t \in \mathbb{R}$. $\nabla^2 f(t\xi) = \nabla^2 f(\xi)$，故 W.L.O.G.

$\|\xi\|_p = 1$，则对 $\forall d \in \mathbb{R}^n$

$$d^T \nabla^2 f(\xi) d = (2-p)\|\xi\|_p^{2-2p} \left(\sum_{i=1}^n |\xi_i|^{p-1} \operatorname{sgn}(\xi_i) d_i \right)^2$$

$$+ (p-1)\|\xi\|_p^{2-p} \sum_{i=1}^n |\xi_i|^{p-2} d_i^2$$

$$\leq (p-1)\|\xi\|_p^{2-p} \sum_{i=1}^n |\xi_i|^{p-2} d_i^2$$

由 Hölder 不等式 $\left(\sum_{k=1}^n |x_k y_k| \leq \left(\sum_{k=1}^n |x_k|^p\right)^{\frac{1}{p}} \left(\sum_{k=1}^n |y_k|^q\right)^{\frac{1}{q}} \right)$

$$\sum_{i=1}^n |\xi_i|^{p-2} d_i^2 \leq \left(\sum_{i=1}^n \left(|\xi_i|^{p-2}\right)^{\frac{p}{p-2}} \right)^{\frac{p-2}{p}} \left(\sum_{i=1}^n (d_i^2)^{\frac{p}{2}} \right)^{\frac{2}{p}}$$

$= \|\xi\|_p^{p-2} \|d\|_p^2 = \|d\|_p^2$

故对 $\forall d \in \mathbb{R}^n$：$d^T \nabla^2 f(\xi) d \leq (p-1) \|d\|_p^2$

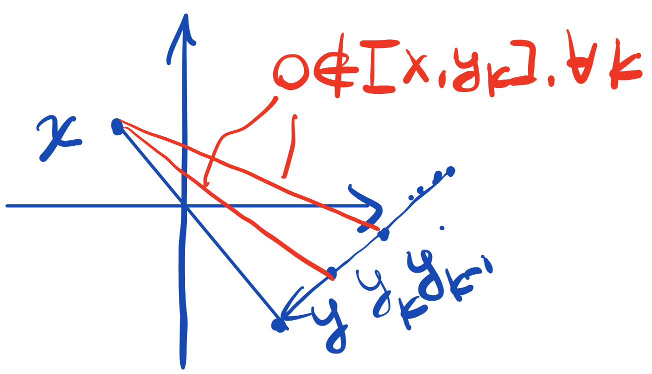
取 $d = x-y$：

$f(y) \leq f(x) + \nabla f(x)^T (y-x) + \frac{p-1}{2} \|x-y\|_p^2$

以下设 $0 \in [x, y]$

• 当 $n=1$（即 $x, y \in \mathbb{R}$ 时）$f(x) = \frac{1}{2} x^2$ $(p-1)$-smooth 自然。

• 当 $n \geq 2$ 时，总 $\exists \{y_k\}$, s.t. $y_k \to y$，且 $0 \notin [x, y_k]$



故对 $\forall k \geq 0$

$f(y_k) \leq f(x) + \nabla f(x) (y_k - x) + \frac{p-1}{2} \|x - y_k\|_p^2$

由 f 的连续性，对 $k \to \infty$ 即证 $\boxed{/}$

§ 5.1.3 Second-order Characterization

Thm 5.12 $f: \mathbb{R}^n \to \mathbb{R}$ 是 $C^2(\mathbb{R}^n)$ 的，则给定 $L \geq 0$，以下等价：(i) f 是 L-smooth 的 w.r.t. ℓ_p-norm ($p \in [1, \infty]$)

(ii) $\|\nabla^2 f(x)\|_{p,q} \leq L$，$\forall x \in \mathbb{R}^n$，$q \in [1, \infty]$，s.t. $\frac{1}{p} + \frac{1}{q} = 1$

证明：

(ii) $\Rightarrow$ (i) 由微积分基本定理：

$$\nabla f(y) = \nabla f(x) + \int_0^1 \nabla^2 f(x + t(y-x))(y-x)\,dt$$

$$= \nabla f(x) + \left(\int_0^1 \nabla^2 f(x+t(y-x))\,dt\right)(y-x)$$

$$\Rightarrow \|\nabla f(y) - \nabla f(x)\|_q = \left\|\left(\int_0^1 \nabla^2 f(x+t(y-x))\,dt\right)(y-x)\right\|_q$$

$$\leq \left\|\int_0^1 \nabla^2 f(x+t(y-x))\,dt\right\|_{p,q} \|y-x\|_p$$

$$\leq \int_0^1 \|\nabla^2 f(x+t(y-x))\|_{p,q}\,dt \;\|y-x\|_p \leq L\|y-x\|_p$$

(i) $\Rightarrow$ (ii) 对 $\forall d \in \mathbb{R}^n, \alpha > 0$, 有

$$\nabla f(x+\alpha d) - \nabla f(x) = \int_0^\alpha \nabla^2 f(x+td)\, d\, dt$$

$$\Rightarrow \left\| \int_0^\alpha \nabla^2 f(x+td)\, dt\, d \right\|_q = \| \nabla f(x+\alpha d) - \nabla f(x) \|_q$$

$$\leq \alpha L \|d\|_p$$

$$\Rightarrow \left\| \frac{1}{\alpha} \int_0^\alpha \nabla^2 f(x+td)\, dt\ d \right\| \leq L \|d\|_p$$

令 $\alpha \to 0^+$, 由积分中值定理:

$$\|\nabla^2 f(x) d\|_q \leq L \|d\|_p \quad , \forall d$$

故 $\|\nabla^2 f(x)\|_{p,q} \leq L$ $\square$

<u>Corollary 5.13</u> $f: \mathbb{R}^n \to \mathbb{R}$ 是 C^2 convex 的, 则 f 是 L-smooth 的 w.r.t. ℓ_2-norm $\Leftrightarrow \lambda_{max}(\nabla^2 f(x)) \leq L \ \forall x$

证明: 由 f 凸, $\nabla^2 f(x) \succeq 0 \ \forall x$, 故

$$\|\nabla^2 f(x)\|_{2,2} = \sqrt{\lambda_{max}(\nabla^2 f(x)^2)} = \lambda_{max}(\nabla^2 f(x))$$

结合 Thm5.12 即证 $\boxed{\angle}$

Example 5.14 $f(x)=\sqrt{1+\|x\|_2^2}$，则 f 1-smooth

w.r.t ℓ_2-norm：对 $\forall x\in\mathbb{R}^n$，有

$$\nabla f(x)=\frac{x}{\sqrt{\|x\|_2^2+1}}$$

且 $$\nabla^2 f(x)=\frac{1}{\sqrt{\|x\|_2^2+1}}I-\frac{xx^T}{(\|x\|_2^2+1)^{\frac{3}{2}}}$$

$$\preceq\frac{1}{\sqrt{\|x\|_2^2+1}}I\preceq I$$

故 $\lambda_{max}(\nabla^2 f(x))\leq 1$，故由 Corollary 5.13 即证 $\boxed{\angle}$

Example 5.15 $f(x)=\log(e^{x_1}+\cdots+e^{x_n})$

下证 f 1-smooth w.r.t. ℓ_2-norm：

$$\frac{\partial f}{\partial x_i}(x)=e^{x_i}\Big/\sum_{k=1}^{n}e^{x_k}\quad,\ i=1,\cdots,n$$

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \begin{cases} -\dfrac{e^{x_i}e^{x_j}}{\left(\sum_{k=1}^{n} e^{x_k}\right)^2}, & i \neq j \\ -\dfrac{e^{x_i}e^{x_i}}{\left(\sum_{k=1}^{n} e^{x_k}\right)^2} + \dfrac{e^{x_i}}{\sum_{k=1}^{n} e^{x_k}}, & i = j \end{cases}$$

$$= \mathrm{diag}(w) - ww^T$$

其中 $w_i = \dfrac{e^{x_i}}{\sum_{k=1}^{n} e^{x_k}}$，故 $\forall x$

$$\nabla^2 f(x) = \mathrm{diag}(w) - ww^T \preceq \mathrm{diag}(w) \preceq I$$

故 $\lambda_{max}(\nabla^2 f(x)) \leq 1$，且 f 是 convex 的，故 f 1-smooth w.r.t. ℓ_2-norm

下证 f 是 1-smooth 的 w.r.t. ℓ_∞-norm

先证 $\forall d \in \mathbb{R}^n$，$d^T \nabla^2 f(x) d \leq \|d\|_\infty^2$

由 $d^T \nabla^2 f(x) d = d^T(\mathrm{diag}(w) - ww^T)d$

$$= d^T \mathrm{diag}(w) d - (w^T d)^2$$

$$\le d^T \mathrm{diag}(w) d$$

$$= \sum_{i=1}^{n} w_i d_i^2$$

$$\le \|d\|_\infty^2 \sum_{i=1}^{n} w_i = \|d\|_\infty^2$$

故由 Thm 5.10，对 $\forall x, y$，$\exists \xi \in [x, y]$

$$f(y) = f(x) + \nabla f(x)^T (y-x) + \frac{1}{2}(y-x)^T \nabla^2 f(\xi)(y-x)$$

$$\le f(x) + \nabla f(x)^T (y-x) + \frac{1}{2}\|y-x\|_\infty^2$$

即 f 是 1-smooth 的 w.r.t. ℓ_∞-norm $\boxed{\diagup}$

§ 5.2 Strong Convexity

Def 5.16 $f: \mathbb{E} \to (-\infty, \infty]$ 是 σ-strongly convex 的：

若 dom$f$ 凸，且对 $\forall x, y \in \mathrm{dom} f$，$\lambda \in [0,1]$：

$$f(\lambda x + (1-\lambda) y) \le \lambda f(x) + (1-\lambda) f(y) - \frac{\sigma}{2} \lambda (1-\lambda) \|x-y\|^2$$

Thm 5.17 设 $\mathbb{E}$ 是 Euclidean 空间, $f: \mathbb{E} \to (-\infty, \infty]$ σ-strongly convex $\Leftrightarrow$ $\wedge$

$f(\cdot) - \frac{\sigma}{2}\|\cdot\|^2$ 是 convex

证明: 令 $g(\cdot) \equiv f(\cdot) - \frac{\sigma}{2}\|\cdot\|^2$, 则 g 凸 $\Leftrightarrow$

$\operatorname{dom} g = \operatorname{dom} f$ 凸, 且对 $\forall x, y \in \operatorname{dom} f$, $\lambda \in [0,1]$:

$$f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y) + \frac{\sigma}{2}\left[\|\lambda x + (1-\lambda) y\|^2 - \lambda\|x\|^2 - (1-\lambda)\|y\|^2\right]$$

$$= \lambda f(x) + (1-\lambda) f(y) - \frac{\sigma}{2}\lambda(1-\lambda)\|x-y\|^2$$

$\boxed{/}$

Remark 5.18 $\mathbb{E}$ 是 Euclidean 是本质的, 设

$$f(x) = \begin{cases} \sum_{i=1}^{n} x_i \log x_i, & x \in \Delta_n, \\ \infty, & \text{else} \end{cases}$$

断言(见例15.27): f 是 1-strongly convex w.r.t ℓ_1-norm

但 $g(x)=f(x)-\alpha\|x\|_1^2$ 对 $\forall\alpha>0$ 是凸的于 dom f

（由 Δ_n 上，$\|x\|_1\equiv 1$），f 显然不会对 $\forall\alpha>0$，有 α-

strongly convex 的性质 $\boxed{/}$

Example 5.19 $\mathbb{E}=\mathbb{R}^n$ 有 ℓ_2-norm

$$f(x)=\frac{1}{2}x^TAx+b^Tx+c$$

易证 f 强凸 $\Leftrightarrow A\in S_{++}^n$，且 $\lambda_{min}(A)$ 是最大的强凸系数 $\boxed{/}$

Lemma 5.20 $f:\mathbb{E}\to(-\infty,\infty]$ 是 σ-strongly convex 的

$g:\mathbb{E}\to(-\infty,\infty]$ 凸，则 $f+g$ σ-strongly convex.

$\boxed{/}$

证明：证明是显然的

Example 5.21 设 $\mathbb{E}$ 是 Euclidean 空间，$C\subseteq\mathbb{E}$ 是非空凸的.

则 $\frac{1}{2}\|x\|^2+\delta_C(x)$ 是 1-strongly convex 的 $\boxed{/}$

Lemma 5.22 $f:\mathbb{R}\to(-\infty,\infty]$ 闭凸，设 $[a,b]\subseteq \mathrm{dom}f$

则 $f(b)-f(a)=\int_a^b h(t)dt$，其中 $h(t):(a,b)\to\mathbb{R}$ 有：

$h(t)\in\partial f(t)$，$\forall t\in(a,b)$.

证明：由 Thm2.22，f 在 $[a,b]$ 上连续，且 f 是 Lipschitz 连续的，故由 N-L 公式，$\exists\, g$ 可积，s.t.

$$f(y)=f(x)+\int_x^y g(t)dt \quad (a\le x<y\le b). \quad (*)$$

下证：$\exists c\in[a,b]$，s.t. $\dfrac{f(b)-f(a)}{b-a}\in\partial f(c)$

令 $g(x)=f(x)-f(a)-\dfrac{f(b)-f(a)}{b-a}(x-a)$

则 $g\in C^0[a,b]$，$g(a)=g(b)=0$，则 $\exists c\in(a,b)$，s.t.

$g(c)$ 是 g 在 $[a,b]$ 上的 minimal

故由 Fermat' lemma：

$$0 \in \partial g(c) = \partial f(c) - \frac{f(b)-f(a)}{b-a}.$$

上面证明过程对 $[x,y]$ $\forall a \le x < y \le b$ 均成立，结合

⊗即证 $\square$

<u>lemma 5.23</u> 设 C 是凸集，$x \in ri(C)$，$y \in cl(C)$，$\lambda \in (0,1]$，则 $\lambda x + (1-\lambda) y \in ri(C)$

证明：WLOG，设 $\dim C = n$，i.e.，$ri\, C = int\, C$

只需证：$\exists \varepsilon > 0$，s.t. $\lambda x + (1-\lambda) y + \varepsilon B \subseteq C$

由 $y \in cl(C)$，知 $y \in C + \varepsilon B$ $\forall \varepsilon > 0$，故对 $\forall \varepsilon > 0$

$$\lambda x + (1-\lambda) y + \varepsilon B \subset \lambda x + (1-\lambda)(C + \varepsilon B) + \varepsilon B$$

$$= \lambda \left[x + \varepsilon \frac{2-\lambda}{\lambda} B \right] + (1-\lambda) C$$

由 $x \in int\, C$，对充分小 ε，有 $x + \varepsilon \frac{2-\lambda}{\lambda} B \subset C$

故 $\lambda \left[x + \varepsilon \frac{2-\lambda}{\lambda} B \right] + (1-\lambda) C \subseteq \lambda C + (1-\lambda) C = C$ $\square$

Thm 5.24 设 $f:\mathbb{E}\to(-\infty,\infty]$ proper, 闭凸函数, 给定 $\sigma>0$

以下等价:

(i) f σ-strongly convex

(ii) $f(y)\geqslant f(x)+\langle g, y-x\rangle+\frac{\sigma}{2}\|y-x\|^2$

$\forall x\in \mathrm{dom}(\partial f)$, $y\in \mathrm{dom} f$, $g\in\partial f(x)$

(iii) $\langle g_x-g_y, x-y\rangle\geqslant\sigma\|x-y\|^2$

$\forall x,y\in\mathrm{dom}(\partial f)$, $g_x\in\partial f(x)$, $g_y\in\partial f(y)$

证明:

(ii) $\Rightarrow$ (i) 取 $x,y\in\mathrm{dom} f$, $\lambda\in(0,1)$, $z\in\mathrm{ri}(\mathrm{dom} f)$

由 Lemma 5.23: 对 $\forall\alpha\in(0,1]$, $\tilde{x}=(1-\alpha)x+\alpha z$
$\in\mathrm{ri}(\mathrm{dom} f)$

固定 α, 记 $x_\lambda=\lambda\tilde{x}+(1-\lambda)y$, 则再用 lemma 5.23

对 $\forall\lambda\in(0,1)$, 有 $x_\lambda\in\mathrm{ri}(\mathrm{dom} f)$, 由 Thm 3.18, $x_\lambda\in\mathrm{dom}(\partial f)$

取 $g\in\partial f(x_\lambda)$，由(ii)

$$f(\tilde{x})\geqslant f(x_\lambda)+\langle g,\tilde{x}-x_\lambda\rangle+\frac{\sigma}{2}\|\tilde{x}-x_\lambda\|^2$$

$$\Leftrightarrow\ f(\tilde{x})\geqslant f(x_\lambda)+(1-\lambda)\langle g,\tilde{x}-y\rangle+\frac{\sigma(1-\lambda)^2}{2}\|y-\tilde{x}\|^2 \quad ①$$

同理：$f(y)\geqslant f(x_\lambda)+\lambda\langle g,y-\tilde{x}\rangle+\frac{\sigma\lambda^2}{2}\|y-\tilde{x}\|^2 \quad ②$

$①\cdot\lambda+②\cdot(1-\lambda)\Rightarrow$

$$f(\lambda\tilde{x}+(1-\lambda)y)\leqslant\lambda f(\tilde{x})+(1-\lambda)f(y)-\frac{\sigma\lambda(1-\lambda)}{2}\|\tilde{x}-y\|^2$$

$$\Leftrightarrow\ g_1(\alpha)\leqslant\lambda g_2(\alpha)+(1-\lambda)f(y)-\frac{\sigma\lambda(1-\lambda)}{2}\|(1-\alpha)x+\alpha z-y\|^2$$

其中 $g_1(\alpha)\equiv f(\lambda(1-\alpha)x+(1-\lambda)y+\lambda\alpha z)$

$g_2(\alpha)\equiv f((1-\alpha)x+\alpha z)$

g_1,g_2 均是1维 proper 闭凸 func，由 Thm2.22，g_1,g_2

在有效域上连续，故取 $\alpha\to0^+$：

$$g_1(0)\leqslant\lambda g_2(0)+(1-\lambda)f(y)-\frac{\sigma\lambda(1-\lambda)}{2}\|x-y\|^2$$

由 $g_1(0)=f(\lambda x+(1-\lambda)y)$，$g_2(0)=f(x)$，即证 (i)

(i) $\Rightarrow$ (iii) 设 $x, y\in \mathrm{dom}(\partial f)$，$g_x\in\partial f(x)$，$g_y\in\partial f(y)$

令 $\lambda\in[0,1)$. 记 $x_\lambda=\lambda x+(1-\lambda)y$，由 (i)

$$f(x_\lambda)\leq \lambda f(x)+(1-\lambda)f(y)-\frac{\sigma}{2}\lambda(1-\lambda)\|x-y\|^2$$

$$\Leftrightarrow \quad \frac{f(x_\lambda)-f(x)}{1-\lambda}\leq f(y)-f(x)-\frac{\sigma}{2}\lambda\|x-y\|^2$$

由 $g_x\in\partial f(x)$：

$$\frac{f(x_\lambda)-f(x)}{1-\lambda}\geq\frac{\langle g_x, x_\lambda-x\rangle}{1-\lambda}=\langle g_x, y-x\rangle$$

故 $$\langle g_x, y-x\rangle\leq f(y)-f(x)-\frac{\sigma\lambda}{2}\|x-y\|^2$$

上式对 $\forall\lambda\in[0,1)$ 成立，取 $\lambda\to 1^-$

$$\langle g_x, y-x\rangle\leq f(y)-f(x)-\frac{\sigma}{2}\|x-y\|^2$$

由 x, y 的对称性：

$\langle g_y, x-y\rangle \le f(x)-f(y)-\frac{6}{2}\|y-x\|^2$

$\Rightarrow \langle g_x-g_y, x-y\rangle \geq 6\|x-y\|^2$, 即证 (iii)

(iii) $\Rightarrow$ (ii) 设 $x\in \operatorname{dom}\partial f$, $y\in \operatorname{dom} f$, $g\in\partial f(x)$

z 是 $ri(\operatorname{dom} f)$ 中任一向量, 令 $\tilde{y}=(1-\alpha)y+\alpha z$, $\alpha\in(0,1)$

由 <u>lemma 5.23</u>, $\tilde{y}\in ri(\operatorname{dom} f)$, 考虑 1 维 func:

$$\varphi(\lambda)=f(x_\lambda),\ \lambda\in[0,1]$$

其中 $x_\lambda=(1-\lambda)x+\lambda\tilde{y}$, 对 $\forall\lambda\in(0,1)$, 由 $x_\lambda\in ri(\operatorname{dom} f)$,

知 $\partial f(x_\lambda)$ 非空. 取 $g_\lambda\in\partial f(x_\lambda)$, 则 $\langle g_\lambda, \tilde{y}-x\rangle\in\partial\varphi(\lambda)$

故由 <u>lemma 5.22</u>

$f(\tilde{y})-f(x)=\varphi(1)-\varphi(0)=\int_0^1\langle g_\lambda, \tilde{y}-x\rangle d\lambda$

由 (iii) $\langle g_\lambda-g, x_\lambda-x\rangle \geq 6\|x_\lambda-x\|^2$

$\Leftrightarrow \langle g_\lambda, \tilde{y}-x\rangle \geq \langle g, \tilde{y}-x\rangle + 6\lambda\|\tilde{y}-x\|^2$

故

$$f(\tilde{y})-f(x)\geqslant\int_0^1\left[\langle g,\tilde{y}-x\rangle+\sigma\lambda\|\tilde{y}-x\|^2\right]d\lambda$$

$$=\langle g,\tilde{y}-x\rangle+\frac{\sigma}{2}\|\tilde{y}-x\|^2$$

$\Leftrightarrow$

$$f((1-\alpha)y+\alpha z)\geqslant f(x)+\langle g,(1-\alpha)y+\alpha z-x\rangle+\frac{\sigma}{2}\|(1-\alpha)y+\alpha z-x\|^2$$

取 $\alpha\to0^+$，再由 $\alpha\mapsto f((1-\alpha)y+\alpha z)$ 的连续性即证 (ii)

$\boxed{/\!/}$

Thm 5.25 设 $f: \mathbb{E} \to (-\infty, \infty]$ 是 proper, closed, σ-强凸

则 (a) f 有唯一最小值点.

(b) $f(x) - f(x^*) \geq \frac{\sigma}{2}\|x - x^*\|^2$, $\forall x \in \mathrm{dom} f$

证明:

(a) 由 dom f 非空, 凸, 故由 Thm 3.17, $\exists x_0 \in \mathrm{ri}(\mathrm{dom} f)$

由 Thm 3.18, $\partial f(x_0) \neq \emptyset$, 设 $g \in \partial f(x_0)$, 由 Thm 5.24 (ii)

$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle + \frac{\sigma}{2}\|x - x_0\|^2, \quad \forall x \in \mathbb{E}$$

由有限维空间 norm 等价: $\exists C > 0$, s.t.

$$\|y\| \geq \sqrt{C}\|y\|_a, \quad \|\cdot\|_a = \sqrt{\langle \cdot, \cdot \rangle}$$

引入 $\|\cdot\|_a$ norm 为了能和 $\langle g, x - x_0 \rangle$ 配方

故 $f(x) \geq f(x_0) + \langle g, x - x_0 \rangle + \frac{C\sigma}{2}\|x - x_0\|_a^2$

$$= f(x_0) - \frac{1}{2C\sigma}\|g\|_a^2 + \frac{C\sigma}{2}\left\|x - \left(x_0 - \frac{1}{C\sigma}g\right)\right\|_a^2, \quad \forall x \in \mathbb{E}$$

$$\Rightarrow \mathrm{Lev}(f, f(x_0)) \subseteq B_{\|\cdot\|_a}\left[x_0 - \frac{1}{C\sigma}g, \frac{1}{C\sigma}\|g\|_a\right]$$

由 f 的闭性知，$\mathrm{Lev}(f, f(x_0))$ 闭，综合上式知：$\mathrm{Lev}(f, f(x_0))$ 紧，故 $\min_{x\in \mathrm{dom} f} f(x) \Leftrightarrow \min_{x\in \mathrm{Lev}(f, f(x_0))} f(x)$.

由 Thm 2.12，f 在 $\mathrm{Lev}(f, f(x_0))$ 上 minimizer $\exists$.

下证唯一性：设 $\tilde{x}, \hat{x}$ 都是 f 的 minimizers，$\Rightarrow$

$f(\tilde{x}) = f(\hat{x}) = f_{opt}$，由 σ-强凸性

$$f_{opt} \le f\left(\tfrac{1}{2}\tilde{x} + \tfrac{1}{2}\hat{x}\right) \le \tfrac{1}{2} f(\tilde{x}) + \tfrac{1}{2} f(\hat{x}) - \tfrac{\sigma}{8}\|\tilde{x} - \hat{x}\|^2$$

$$= f_{opt} - \tfrac{\sigma}{8}\|\tilde{x} - \hat{x}\|^2$$

故 $\tilde{x} = \hat{x}$

(b) 设 x^* 是 f 的唯一极小值点，由 Fermet's 最优性条件：

$0 \in \partial f(x^*)$，故由 Thm 5.24 (ii)：

$$f(x) - f(x^*) \geq \langle 0, x - x^* \rangle + \tfrac{\sigma}{2}\|x - x^*\|^2 = \tfrac{\sigma}{2}\|x - x^*\|^2$$

$\boxed{/}$

§ 5.3 Smoothness and Strong Convexity Correspondence.

Thm 5.26 设 $\sigma > 0$，则

(a) 若 $f: \mathbb{E} \to \mathbb{R}$ 是 $\frac{1}{\sigma}$-smooth convex func, 则 f^* 是 σ-strongly convex w.r.t. $\|\cdot\|_*$

(b) 若 $f: \mathbb{E} \to (-\infty, \infty]$ 是 proper 闭 σ-强凸的, 则 f^*: $\mathbb{E}^* \to \mathbb{R}$ 是 $\frac{1}{\sigma}$-smooth的

证明:

(a) 设 $f: \mathbb{E} \to \mathbb{R}$ 是 $\frac{1}{\sigma}$-smooth convex func

下证: f^* σ-强凸; 取 $y_1, y_2 \in \mathrm{dom}(\partial f^*)$, $v_1 \in \partial f^*(y_1)$, $v_2 \in \partial f^*(y_2)$, 故由 Thm 4.20 结合 f 的 proper 闭凸性:

$y_1 \in \partial f(v_1)$, $y_2 \in \partial f(v_2)$, 由 f 的可微性, 知 $y_1 = \nabla f(v_1)$, $y_2 = \nabla f(v_2)$, 由 Thm 5.8 (i) (iv) $\Rightarrow$

$$\langle y_1 - y_2, v_1 - v_2 \rangle \geq \sigma \|y_1 - y_2\|_*^2$$

由 Thm 5.24 (iii) 知, f^* 是 σ-强凸的.

(b) 设 f 是 proper, 闭, σ-强凸的, 由 Coro 4.21:

$$\partial f^*(y) = \underset{x \in \mathbb{E}}{\operatorname{argmax}} \left\{ \langle x, y \rangle - f(x) \right\}, \quad \forall y \in \mathbb{E}^*$$

关于x强凸

由 f 的强凸性与 Thm5.25(a)，$\partial f^*(y)$ 对 $\forall y \in \mathbb{E}^*$ 是 singleton，故由 Thm3.33 知，f^* 在 $\mathbb{E}^*$ 上可微

取 $y_1, y_2 \in \mathbb{E}^*$，记 $v_1 = \nabla f^*(y_1)$，$v_2 = \nabla f^*(y_2)$，由 Thm4.20 知：$y_1 \in \partial f(v_1)$，$y_2 \in \partial f(v_2)$，故由 Thm5.24 (i) (iii)

$$\langle y_1 - y_2, v_1 - v_2 \rangle \geq \sigma \|v_1 - v_2\|^2,$$

即 $\langle y_1 - y_2, \nabla f^*(y_1) - \nabla f^*(y_2) \rangle \geq \sigma \|\nabla f^*(y_1) - \nabla f^*(y_2)\|^2$

又由 C-S 不等式：

$$\|\nabla f^*(y_1) - \nabla f^*(y_2)\| \leq \frac{1}{\sigma} \|y_1 - y_2\|_*$$

即证：f^* 是 $\frac{1}{\sigma}$-光滑 $\boxtimes$

§5.3.2 Examples of strongly Convex func

Example5.27 $f(x) = \begin{cases} \sum_{i=1}^n x_i \log x_i, & x \in \Delta_n \\ \infty, & \text{else} \end{cases}$

下证 f 的强凸性：由 §4.4.10，$f^*(y) = \log\left(\sum_{i=1}^n e^{y_i}\right)$

由 Example 5.15 知：f^* 是 1-smooth 的 w.r.t. ℓ_∞, ℓ_2-norm

故由 Thm 5.26 知：f 是 1-强凸的 w.r.t. ℓ_1, ℓ_2-norm.

Example 5.28 $\quad f(x)=\frac{1}{2}\|x\|_p^2 \quad p\in(1,2]$

由 §4.4.15 知，$f^*(y)=\frac{1}{2}\|y\|_q^2$ ($q\geq 2$, 且 $\frac{1}{p}+\frac{1}{q}=1$). 由

Example 5.11, f^* 是 $(q-1)$-smooth 的 w.r.t. ℓ_q-norm

故由 Thm 5.26, f 是 $\frac{1}{q-1}=p-1$-强凸 func

Example 5.29 $\quad f(x)=\begin{cases} -\sqrt{1-\|x\|_2^2} & , \|x\|_2\leq 1 \\ \infty & , \text{else} \end{cases}$

由 §4.4.13, $f^*(y)=\sqrt{\|y\|_2^2+1}$, 由 Example 5.14

f^* 是 1-smooth w.r.t. ℓ_2-norm, 故 f 是 1-强凸 w.r.t. ℓ_2-norm

§5.3.3 Smoothness and Differentiability of the Infimal Convolution

<u>Thm 5.30</u> 设 $f: \mathbb{E} \to (-\infty, \infty]$ 是 proper, closed func, $\omega: \mathbb{E} \to \mathbb{R}$ 是 L-smooth convex func, 设 $f \square \omega$ 实值, 则

(a) $f \square \omega$ 是 L-smooth 的

(b) 设 $x \in \mathbb{E}$, $u(x)$ 是 $\min_u \{f(u) + \omega(x-u)\}$ 的 minimizer

则 $\nabla (f \square \omega)(x) = \nabla \omega(x - u(x))$

证明:

(a) 由 <u>Thm4.19</u>, $f \square \omega = (f^* + \omega^*)^*$, 由 f, ω 是 proper 闭凸 func, 则由 <u>Thm4.3</u>, f^*, ω^* 是闭凸的

由 <u>Thm5.26</u>, ω^* 是 $\frac{1}{L}$-强凸的, 故由 <u>lemma 5.20</u>, $f^* + \omega^*$ 是 $\frac{1}{L}$-强凸的, closed 的 func; 由 <u>Thm4.16</u>

$$(f \square \omega)^* = f^* + \omega^*$$

由 $f \square \omega$ 是 convex 的 (<u>Thm2.19</u>) 且 proper, 故由 <u>Thm4.5</u> $f^* + \omega^*$ 是 proper 的. 综上: $f^* + \omega^*$ 是 proper, closed, $\frac{1}{L}$-强凸的 func, 由 <u>Thm5.26</u> $f \square \omega = (f^* + \omega^*)^*$ 是 L-smooth 的.

(b) 设 $x\in\mathbb{E}$, s.t. $u(x)$ 是 $\min_u\{f(u)+\omega(x-u)\}$

的 minimizer, 定义 $z\equiv\nabla\omega(x-u(x))$, 下证:

$\nabla(f\square\omega)(x)=z$, 等价于证明: $\forall\xi\in\mathbb{E}$,

$$\lim_{\|\xi\|\to0}\frac{|\phi(\xi)|}{\|\xi\|}=0,\text{ 其中 } ⊛$$

$\phi(\xi)\equiv(f\square\omega)(x+\xi)-(f\square\omega)(x)-\langle\xi,z\rangle$. 由卷积定义:

$$(f\square\omega)(x+\xi)\leq f(u(x))+\omega(x+\xi-u(x))$$

故 $\phi(\xi)=(f\square\omega)(x+\xi)-(f\square\omega)(x)-\langle\xi,z\rangle$

$\leq\omega(x+\xi-u(x))-\omega(x-u(x))-\langle\xi,z\rangle$

$\leq\langle\xi,\nabla\omega(x+\xi-u(x))\rangle-\langle\xi,z\rangle$

$=\langle\xi,\nabla\omega(x+\xi-u(x))-\nabla\omega(x-u(x))\rangle$

$\leq\|\xi\|\cdot\|\nabla\omega(x+\xi-u)-\nabla\omega(x-u(x))\|_*$

$\leq L\|\xi\|^2$

又由 $f\square\omega$ 的凸性, 知 $\phi(\xi)$ 是 convex 的, 故

$\phi(0)=\phi\left(\frac{1}{2}\xi+\frac{1}{2}(-\xi)\right)\le\frac{1}{2}\phi(\xi)+\frac{1}{2}\phi(-\xi)$，故由 $\phi(0)=0$

$\phi(\xi)\ge-\phi(-\xi)\ge-L\|\xi\|^2$，即证(*)式 $\boxed{\checkmark}$

<u>Example 5.31</u> $\varphi_C(x)=\frac{1}{2}d_C^2(x)$，$C\subseteq\mathbb{E}$ 闭凸集

由 $\varphi_C=\delta_C\,\square\,h$，$h(x)=\frac{1}{2}\|x\|^2$，由 h 是 1-smooth convex 的

δ_C 是 proper 闭凸 func，故由 <u>Thm 5.30</u>，φ_C 是 1-smooth 的.

Chapter 6 The Proximal Operator

§6.1 Def, Existence, and Uniqueness

Def 6.1 给定 $f: \mathbb{E} \to (-\infty, \infty]$，f的 proximal mapping:

$$\operatorname{prox}_f(x) = \underset{u \in \mathbb{E}}{\operatorname{argmin}} \left\{ f(u) + \frac{1}{2}\|u-x\|^2 \right\}, \ \forall x \in \mathbb{E}$$

Example 6.2 $g_1(x) \equiv 0$

$$g_2(x) = \begin{cases} 0, & x \neq 0 \\ -\lambda, & x = 0 \end{cases} \qquad g_3(x) = \begin{cases} 0 & x \neq 0 \\ \lambda, & x = 0 \end{cases} \qquad (\lambda > 0)$$

则 $\operatorname{prox}_{g_1}(x) = \underset{u}{\operatorname{argmin}} \left\{ g_1(x) + \frac{1}{2}(u-x)^2 \right\} = \{x\}$

$\operatorname{prox}_{g_2}(x) = \underset{u}{\operatorname{argmin}}\ \tilde{g}_2(u,x)$，其中

$$\tilde{g}_2(u,x) = g_2(u) + \frac{1}{2}(u-x)^2 = \begin{cases} -\lambda + \frac{x^2}{2}, & u = 0 \\ \frac{1}{2}(u-x)^2, & u \neq 0 \end{cases}$$

当 $x\neq 0$ 时：$\frac{1}{2}(u-x)^2$ 在 $u=x$ 处取 min，且最小值为 0

故 $0>-\lambda+\frac{x^2}{2}$ 时，$\tilde{g}_2(\cdot,x)$ 的 minimizer 是 $u=0$

$0<-\lambda+\frac{x^2}{2}$ 时，$\tilde{g}_2(\cdot,x)$ 的 minimizer 是 $u=x$

$0=-\lambda+\frac{x^2}{2}$ 时，$\tilde{g}_2(\cdot,x)$ 的 minimizer 是 $u=x$ 和 $u=0$

当 $x=0$ 时，显然 $\tilde{g}_2(\cdot,0)$ 的 minimizer 是 $u=0$

综上：
$$\operatorname{prox}_{g_2}(x)=\begin{cases}\{0\}, & |x|<\sqrt{2\lambda}\\ \{x\}, & |x|>\sqrt{2\lambda}\\ \{0,x\}, & |x|=\sqrt{2\lambda}\end{cases}$$

$\operatorname{prox}_{g_3}(x)=\underset{u}{\operatorname{argmin}}\ \tilde{g}_3(u,x)$，其中

$$\tilde{g}_3(u,x)=\begin{cases}\lambda+\frac{1}{2}x^2, & u=0\\ \frac{1}{2}(u-x)^2, & u\neq 0\end{cases}$$

当 $x\neq 0$ 时 $\frac{1}{2}(u-x)^2$ 在 $u=x$ 处取 min，为 0，且 $0<\lambda+\frac{1}{2}x^2$

当 $x=0$ 时，由 $\lambda>0$ 知，minimizer 不存在

综上：$\mathrm{prox}_{g_3}(x)=\begin{cases}\{x\}, & x\neq 0\\ \phi, & x=0\end{cases}$

Thm 6.3 设 $f:\mathbb{E}\to(-\infty,\infty]$ 是 proper 闭凸 func, 则

$\mathrm{prox}_f(x)$ 是 singleton. 对 $\forall x\in\mathbb{E}$

证明：对 $\forall x\in\mathbb{E}$,

$$\mathrm{prox}_f(x)=\underset{u\in\mathbb{E}}{\mathrm{argmin}}\ \tilde{f}(u,x)$$

其中 $\tilde{f}(u,x)\equiv f(u)+\frac{1}{2}\|u-x\|^2$，由 Lemma 5.20, Thm 2.7

知 $\tilde{f}(\cdot,x)$ 是 closed，强凸 func，且由 f 是 proper 的

故 $\tilde{f}(\cdot,x)$ 是 proper 的，故由 Thm 5.25(a) 知，Thm 6.3

成立 $\square$

Thm 6.4 设 $f:\mathbb{E}\to(-\infty,\infty]$ 是 proper, closed func,

设 $u\mapsto f(u)+\frac{1}{2}\|u-x\|^2$ 对 $\forall x\in\mathbb{E}$ 是 coercive 的.

则 $\mathrm{prox}_f(x)$ 对 $\forall x\in\mathbb{E}$ 非空

证明：对 $\forall x \in \mathbb{E}$，$h(u) \equiv f(u) + \frac{1}{2}\|u-x\|^2$ 是 proper closed，coercive func，由 Thm2.14，Thm 6.4 成立 $\boxtimes$

§ 6.2 First Set of Examples of Proximal Mapping

§ 6.2.1 $f \equiv c$

$$\operatorname{prox}_f(x) = \underset{u \in \mathbb{E}}{\operatorname{argmin}} \left\{ c + \frac{1}{2}\|u-x\|^2 \right\} = x$$

§ 6.2.2 $f(x) = \langle a, x\rangle + b$

$$\operatorname{prox}_f(x) = \underset{u \in \mathbb{E}}{\operatorname{argmin}} \left\{ \langle a, u\rangle + b + \frac{1}{2}\|u-x\|^2 \right\}$$

$$= \underset{u \in \mathbb{E}}{\operatorname{argmin}} \left\{ \langle a, x\rangle + b - \frac{1}{2}\|a\|^2 + \frac{1}{2}\|u-(x-a)\|^2 \right\}$$

$$= x - a$$

§ 6.2.3 $f(x)=\frac{1}{2}x^TAx+b^Tx+c$, $A\in S_+^n$

则 $\operatorname{prox}_f(x)$ 是下优化问题的 minimizer

$$\min_{u\in\mathbb{E}}\left\{\frac{1}{2}u^TAu+b^Tu+c+\frac{1}{2}\|u-x\|^2\right\}$$

$$\Rightarrow\quad Au+b+(u-x)=0 \Rightarrow u=(A+I)^{-1}(x-b).$$

$$\operatorname{prox}_f(x)=(A+I)^{-1}(x-b)$$

$\square$

§ 6.2.4 One-dimensional Examples

Lemma 6.5

$$g_1(x)=\begin{cases}\mu x, & x\geq 0\\ \infty, & x<0\end{cases}\qquad \operatorname{prox}_{g_1}(x)=[x-\mu]_+$$

$$g_2(x)=\lambda|x|,\qquad \operatorname{prox}_{g_2}(x)=[\,|x|-\lambda]_+\operatorname{sgn}(x)$$

$$g_3(x)=\begin{cases}\lambda x^3, & x\geq 0\\ \infty, & x<0\end{cases}\qquad \operatorname{prox}_{g_3}(x)=\frac{-1+\sqrt{1+12\lambda[x]_+}}{6\lambda}$$

$$g_4(x)=\begin{cases}-\lambda\log x, & x>0\\ \infty, & x\le 0\end{cases},\quad \text{prox}_{g_4}(x)=\frac{x+\sqrt{x^2+4\lambda}}{2}$$

$$g_5(x)=\delta_{[0,\eta]\cap\mathbb{R}}(x),\quad \text{prox}_{g_5}(x)=\min\{\max\{x,0\},\eta\}$$

其中$\lambda\in\mathbb{R}_+$, $\eta\in[0,\infty]$, $\mu\in\mathbb{R}$

证明:

[prox of g_1] $\text{prox}_{g_1}(x)$是以下func的minimizer

$$f(u)=\begin{cases}\infty, & u<0\\ f_1(u), & u\geq 0\end{cases}$$

其中$f_1(u)=\mu u+\frac{1}{2}(u-x)^2$, $f_1'(u)=0 \Leftrightarrow u=x-\mu$

- 若$x>\mu$时, $\text{prox}_{g_1}(x)=x-\mu$
- 若$x\le\mu$时, $\text{prox}_{g_1}(x)=0$ ← 唯一的不可微点于dom f

[prox of g_2] $\text{prox}_{g_2}(x)$是以下func的minimizer

$$h(u)=\begin{cases} h_1(u)\equiv \lambda u+\frac{1}{2}(u-x)^2, & u>0 \\ h_2(u)\equiv -\lambda u+\frac{1}{2}(u-x)^2, & u\leq 0 \end{cases}$$

• 若 minimizer 在 $u>0$ 处取，取 $0=h_1'(u) \Leftrightarrow u=x-\lambda$

故若 $x>\lambda$，则有 $\mathrm{prox}_{g_2}(x)=x-\lambda$

• 同理若 $x<\lambda$，则 $\mathrm{prox}_{g_2}(x)=x+\lambda$

• 若 $|x|\leq\lambda$ 时，$\mathrm{prox}_{g_2}(x)$ 一定在唯一的不可微点取

即 $\mathrm{prox}_{g_2}(x)=0$

[prox of g_3] $\mathrm{prox}_{g_3}(x)$ 是以下 func 的 minimizer

$$s(u)=\begin{cases} \lambda u^3+\frac{1}{2}(u-x)^2, & u\geqslant 0 \\ \infty, & u<0 \end{cases}$$

• 若 minimizer 在 $u>0$ 处取，则 $\tilde{u}=\mathrm{prox}_{g_3}(x)$ 满足：

$s'(\tilde{u})=0$, i.e., $3\lambda\tilde{u}^2+\tilde{u}-x=0$

上面的方程有正根 $\Leftrightarrow x>0$，此时唯一正根即为

$$\mathrm{prox}_{g_3}(x) = \tilde{u} = \frac{-1+\sqrt{1+12\lambda x}}{6\lambda}$$

- 若 $x \le 0$，则 s 的minimizer在唯一不可微点处取到，

即 $\mathrm{prox}_{g_3}(x) = 0$

[prox of g_4] $\tilde{u} = \mathrm{prox}_{g_4}(x)$ 是以下 func 在 $\mathbb{R}_{++}$ 上的 minimizer：

$$t(u) = -\lambda \log u + \frac{1}{2}(x-u)^2$$

令 $t'(\tilde{u}) = 0 \Rightarrow -\frac{\lambda}{\tilde{u}} + (\tilde{u}-x) = 0 \iff \tilde{u}^2 - x\tilde{u} - \lambda = 0$

故 $\mathrm{prox}_{g_4}(x) = \tilde{u} = \dfrac{x+\sqrt{x^2+4\lambda}}{2}$

[prox of g_5] 先设 $\eta < \infty$，则 $\tilde{u} = \mathrm{prox}_{g_5}(x)$ 是

$w(u) = \frac{1}{2}(u-x)^2$ 在 $[0,\eta]$ 上的 minimizer

w 在 $\mathbb{R}$ 上的minimizer在 $u=x$ 处取到，故

- $0 \le x \le \eta$ 时，$\tilde{u} = x$
- $x < 0$ 时，$w\uparrow : [0,\eta]$，故 $\tilde{u} = 0$

· $x>\eta$时，$\omega\downarrow$: $[0,\eta]$. 故 $\tilde{u}=\eta$

综上：$\text{prox}_{g_5}(x)=\tilde{u}=\begin{cases} x, & 0\le x\le\eta \\ 0, & x<0 \\ \eta, & x>\eta \end{cases} = \min\{\max\{x,0\},\eta\}$.

当 $\eta=\infty$，则 $g_5(x)=\delta_{[0,\infty)}(x)$，则 g_5 与 g_1 相同 $(\mu=0)$，

故此时 $\text{prox}_{g_5}(x)=[x]_+$，也可以写成如下形式：

$$\text{prox}_{g_5}(x)=\min\{\max\{x,0\},\infty\}.$$

$\boxed{\angle}$

§6.3 Prox Calculus Rules

<u>Thm 6.6</u> $f: \mathbb{E}_1\times\mathbb{E}_2\times\cdots\times\mathbb{E}_m\to(-\infty,\infty]$

$$f(x_1,\cdots,x_m)=\sum_{i=1}^{m} f_i(x_i) \text{ 对 } \forall x_i\in\mathbb{E}_i,\ i=1,\cdots,m$$

则对 $\forall x_1\in\mathbb{E}_1,\cdots,x_m\in\mathbb{E}_m$. 有

$$\text{prox}_f(x_1,\cdots,x_m)=\text{prox}_{f_1}(x_1)\times\cdots\times\text{prox}_{f_m}(x_m)$$

证明：

$$\operatorname{prox}_f(x_1,\cdots,x_m)=\underset{y_1,\cdots,y_m}{\operatorname{argmin}}\sum_{i=1}^{m}\left[\frac{1}{2}\|y_i-x_i\|^2+f_i(y_i)\right]$$

$$=\prod_{i=1}^{m}\underset{y_i}{\operatorname{argmin}}\left[\frac{1}{2}\|y_i-x_i\|^2+f_i(y_i)\right]$$

$$=\prod_{i=1}^{m}\operatorname{prox}_{f_i}(x_i)$$

$\boxed{/\!/}$

<u>**Example 6.8**</u> (ℓ_1-norm) $g(x)=\lambda\|x\|_1$

则 $g(x)=\sum_{i=1}^{n}\varphi(x_i)$，其中 $\varphi(t)=\lambda|t|$，由 <u>lemma 6.5</u>

$\operatorname{prox}_\varphi(s)=T_\lambda(s)$，其中 T_λ 是 soft thresholding func

$$T_\lambda(y)=[|y|-\lambda]_+\operatorname{sgn}(y)=\begin{cases}y-\lambda, & y\geq\lambda\\ 0, & |y|<\lambda\\ y+\lambda, & y\leq-\lambda\end{cases}$$

故由 <u>Thm 6.6</u> $\operatorname{prox}_g(x)=(T_\lambda(x_j))_{j=1}^{n}$

拓展 T_λ 的定义范围：

$$T_\lambda(x)\equiv(T_\lambda(x_j))_{j=1}^{n}=[|x|-\lambda e]_+\odot\operatorname{sgn}(x)$$

$\boxed{/\!/}$

Example 6.9 $g(x)=\begin{cases}-\lambda\sum_{j=1}^{n}\log x_j, & x>0\\ \infty, & \text{else}\end{cases}$ $(\lambda>0)$

则 $g(x)=\sum_{i=1}^{n}\varphi(x_i)$，其中 $\varphi(t)=\begin{cases}-\lambda\log t, & t>0\\ \infty, & t<0\end{cases}$

由 lemma 6.5 $\operatorname{prox}_{\varphi}(s)=\dfrac{s+\sqrt{s^2+4\lambda}}{2}$

则由 Thm 6.6 $\operatorname{prox}_g(x)=\left(\dfrac{x_j+\sqrt{x_j^2+4\lambda}}{2}\right)_{j=1}^{n}$ $\square$

Example 6.10 $f(x)=\lambda\|x\|_0$，$\lambda>0$

则 $f(x)=\sum_{i=1}^{n}I(x_i)$，其中 $I(t)=\begin{cases}\lambda, & t\neq 0\\ 0, & t=0\end{cases}$

注意到 $I(\cdot)=J(\cdot)+\lambda$，其中 $J(t)=\begin{cases}0, & t\neq 0\\ -\lambda, & t=0\end{cases}$

$$\operatorname{prox}_J(s)=\begin{cases}\{0\}, & |s|<\sqrt{2\lambda}\\ \{s\}, & |s|>\sqrt{2\lambda}\\ \{0,s\}, & |s|=\sqrt{2\lambda}\end{cases}\doteq H_{\sqrt{2\lambda}}(s)$$

其中 H_α 是 hard thresholding 算子

$$H_\alpha(s)=\begin{cases}\{0\}, & |s|<\alpha\\ \{s\}, & |s|>\alpha\\ \{0,s\}, & |s|=\alpha\end{cases}$$

则 $\text{prox}_I(s)=\underset{t}{\text{argmin}}\{I(t)+\frac{1}{2}(t-s)^2\}$

$$=\underset{t}{\text{argmin}}\{J(t)+\lambda+\frac{1}{2}(t-s)^2\}$$

$$=\underset{t}{\text{argmin}}\{J(t)+\frac{1}{2}(t-s)^2\}$$

$$=\text{prox}_J(s)$$

故由 Thm 6.6 知:

$$\text{prox}_g(x)=H_{\sqrt{2\lambda}}(x_1)\times\cdots\times H_{\sqrt{2\lambda}}(x_n)$$ $\boxed{}$

Thm 6.11 设 $g:\mathbb{E}\to(-\infty,\infty]$ 是 proper func, 设 $\lambda\neq 0$, $a\in\mathbb{E}$, 令 $f(x)=g(\lambda x+a)$. 则

$$\text{prox}_f(x)=\frac{1}{\lambda}[\text{prox}_{\lambda^2 g}(\lambda x+a)-a]$$

证明:

$$\mathrm{prox}_f(x) = \arg\min_u \left\{ f(u) + \frac{1}{2}\|u-x\|^2 \right\}$$

$$= \arg\min_u \left\{ g(\lambda u + a) + \frac{1}{2}\|u-x\|^2 \right\}.$$

令 $z = \lambda u + a$，则待优化目标 func 为：

$$g(z) + \frac{1}{2}\left\|\frac{1}{\lambda}(z-a) - x\right\|^2 = \frac{1}{\lambda^2}\left[\lambda^2 g(z) + \frac{1}{2}\|z - (\lambda x + a)\|^2\right]$$

故 $z = \mathrm{prox}_{\lambda^2 g}(\lambda x + a)$，故由 $z = \lambda u + a \Rightarrow$

$$\mathrm{prox}_f(x) = \frac{1}{\lambda}\left[\mathrm{prox}_{\lambda^2 g}(\lambda x + a) - a\right] \qquad \boxed{\checkmark}$$

Thm 6.12 设 $g: \mathbb{E} \to (-\infty, \infty]$ 是 proper 的，$\lambda \neq 0$，定义

$f(x) = \lambda g(x/\lambda)$，则 $\text{prox}_f(x) = \lambda \, \text{prox}_{g/\lambda}(x/\lambda)$

证明：

$$\text{prox}_f(x) = \underset{u}{\text{argmin}} \left\{ f(u) + \frac{1}{2}\|u - x\|^2 \right\}$$

$$= \underset{u}{\text{argmin}} \left\{ \lambda g\left(\frac{u}{\lambda}\right) + \frac{1}{2}\|u - x\|^2 \right\}$$

$$\overset{z = \frac{u}{\lambda}}{=} \lambda \, \underset{z}{\text{argmin}} \left\{ \lambda g(z) + \frac{1}{2}\|\lambda z - x\|^2 \right\}$$

$$= \lambda \, \underset{z}{\text{argmin}} \left\{ \lambda^2 \left[\frac{g(z)}{\lambda} + \frac{1}{2}\left\|z - \frac{x}{\lambda}\right\|^2 \right] \right\}$$

$$= \lambda \, \underset{z}{\text{argmin}} \left\{ \frac{g(z)}{\lambda} + \frac{1}{2}\left\|z - \frac{x}{\lambda}\right\|^2 \right\}$$

$$= \lambda \, \text{prox}_{g/\lambda}\left(\frac{x}{\lambda}\right)$$

$\boxed{/}$

Thm 6.13 设 $g: \mathbb{E} \to (-\infty, \infty]$ 是 proper，$f(x) = g(x) + \frac{c}{2}\|x\|^2 + \langle a, x \rangle + r$，其中 $c > 0$，$a \in \mathbb{E}$，$r \in \mathbb{R}$，则

$$\mathrm{prox}_f(x) = \mathrm{prox}_{\frac{1}{c+1}g}\left(\frac{x-a}{c+1}\right)$$

证明:

$$\mathrm{prox}_f(x) = \underset{u}{\mathrm{argmin}}\left\{ f(u) + \frac{1}{2}\|u-x\|^2 \right\}$$

$$= \underset{u}{\mathrm{argmin}}\left\{ g(u) + \frac{c}{2}\|u\|^2 + \langle a, u\rangle + r + \frac{1}{2}\|u-x\|^2 \right\}$$

$$= \underset{u}{\mathrm{argmin}}\left\{ g(u) + \frac{c+1}{2}\left\| u - \left(\frac{x-a}{c+1}\right)\right\|^2 \right\}$$

$$= \mathrm{prox}_{\frac{1}{c+1}g}\left(\frac{x-a}{c+1}\right) \qquad \square$$

Example 6.14 $\quad f(x) = \begin{cases} \mu x, & 0 \le x \le \alpha \\ \infty, & \text{else} \end{cases}, \quad \mu \in \mathbb{R}, \alpha \in [0, \infty].$

注意到: $f(x) = \delta_{[0,\alpha]\cap\mathbb{R}}(x) + \mu x$

由 lemma 6.5, $\mathrm{prox}_{\delta_{[0,\alpha]\cap\mathbb{R}}}(x) = \min\{\max\{x, 0\}, \alpha\}$.

由 Thm 6.13 ($c=0, a=\mu, r=0$), 知

$$\mathrm{prox}_f(x) = \mathrm{prox}_g(x-\mu) = \min\{\max\{x-\mu, 0\}, \alpha\} \qquad \square$$

Thm 6.15 设 $g:\mathbb{R}^m\to(-\infty,\infty]$ 是 proper, closed, convex func

设 $f(x)=g(\mathcal{A}(x)+b)$, $b\in\mathbb{R}^m$, $\mathcal{A}:\mathbb{V}\to\mathbb{R}^m$ 是线性映射.

满足 $\mathcal{A}\circ\mathcal{A}^T=\alpha I$ 对某 $\alpha>0$, 则对 $\forall x\in\mathbb{V}$

$$\operatorname{prox}_f(x)=x+\frac{1}{\alpha}\mathcal{A}^T\left(\operatorname{prox}_{\alpha g}(\mathcal{A}(x)+b)-\mathcal{A}(x)-b\right)$$

证明: $\operatorname{prox}_f(x)$ 是以下问题 optimal solution

$$\min_{u\in\mathbb{V}}\left\{f(u)+\frac{1}{2}\|u-x\|^2\right\}$$

等价于 $\displaystyle\min_{u\in\mathbb{V}}\left\{g(\mathcal{A}(u)+b)+\frac{1}{2}\|u-x\|^2\right\}$

$$\Leftrightarrow\quad\min_{u\in\mathbb{V},\,z\in\mathbb{R}^m}\ g(z)+\frac{1}{2}\|u-x\|^2$$

$$\text{s.t.}\quad z=\mathcal{A}(u)+b$$

由 Thm 5.25 知, 上优化问题最优解 $\exists$, 设为 $(\tilde{z},\tilde{u})$, 注意到 $\tilde{u}=\operatorname{prox}_f(x)$, 固定 $z=\tilde{z}$, 知 $\tilde{u}$ 是下优化问题最优解:

$$\min_{u\in\mathbb{V}}\ \frac{1}{2}\|u-x\|^2$$

$$\text{s.t.}\quad A(u) = \tilde{z} - b$$

由 Thm A.1，上问题有 strong duality，由 Thm A.2，$\exists\, y \in \mathbb{R}^m$：

$$\tilde{u} \in \underset{u \in \mathbb{V}}{\operatorname{argmin}} \left\{ \frac{1}{2}\|u-x\|^2 + \langle y, A(u) - \tilde{z} + b \rangle \right\} \qquad (6.12)$$

$$A(\tilde{u}) = \tilde{z} - b \qquad (6.13)$$

由 (6.12) 知 $\tilde{u} = x - A^T(y)$，代入 (6.13)

$$A(x - A^T(y)) = \tilde{z} - b$$

由假设 $A \circ A^T = \alpha I$ 知，

$$\alpha y = A(x) + b - \tilde{z}\text{，故}$$

$$\operatorname{prox}_f(x) = \tilde{u} = x + \frac{1}{\alpha} A^T(\tilde{z} - A(x) - b) \qquad (6.15)$$

将 $u = \tilde{u}$ 代入极小化问题 (6.10)，知

$$\tilde{z} = \underset{z \in \mathbb{R}^m}{\operatorname{argmin}} \left\{ g(z) + \frac{1}{2}\left\| x + \frac{1}{\alpha} A^T(z - A(x) - b) - x \right\|^2 \right\}$$

$$= \underset{z \in \mathbb{R}^m}{\operatorname{argmin}} \left\{ g(z) + \frac{1}{2\alpha^2} \| A^T(z - A(x) - b) \|^2 \right\}$$

$$\langle A^T(z-A(x)-b), A^T(z-A(x)-b)\rangle$$

$$= \alpha \|z-A(x)-b\|^2$$

$$= \underset{z\in\mathbb{R}^m}{\operatorname{argmin}} \left\{ \alpha g(z) + \frac{1}{2}\|z-A(x)-b\|^2 \right\}$$

$$= \operatorname{prox}_{\alpha g}(A(x)+b)$$

再代回(6.15)即证 $\boxed{\diagup}$

<u>Example 6.16</u> 设 $g:\mathbb{E}\to(-\infty,\infty]$ 是 proper, closed convex

$\mathbb{E}=\mathbb{R}^d$, 设 $f:\mathbb{E}^m\to(-\infty,\infty]$ 如下定义:

$$f(x_1,\cdots,x_m) = g(x_1+\cdots+x_m)$$

则 f 可以写作 $f(x_1,\cdots,x_m)=g(A(x_1,\cdots,x_m))$, 其中 A:

$$A(x_1,\cdots,x_m) = x_1+\cdots+x_m$$

显然 $A^T:\mathbb{E}\to\mathbb{E}^m$ 是 $A^T(x)=(x,\cdots,x)$ (用定义即可)

且 $\forall x\in\mathbb{E}$, $A(A^T(x))=mx$, 故用 <u>Thm 6.15</u>

$$\operatorname{prox}_f(x_1,\cdots,x_m)_j = x_j + \frac{1}{m}\left(\operatorname{prox}_{mg}\left(\sum_{i=1}^m x_i\right) - \sum_{i=1}^m x_i\right), \quad j=1,\cdots,m$$

Example 6.17 $f:\mathbb{R}^n\to\mathbb{R}$, $f(x)=|a^Tx|$, $a\in\mathbb{R}^n\setminus\{0\}$

记 $f(x)=g(a^Tx)$，其中 $g(t)=|t|$，由 lemma 6.5，

$\mathrm{prox}_{\lambda g}=T_\lambda$，其中 $T_\lambda(x)=[|x|-\lambda]_+\mathrm{sgn}(x)$，用 Thm 6.15：

$$\mathrm{prox}_f(x)=x+\frac{1}{\|a\|^2}\left(T_{\|a\|^2}(a^Tx)-a^Tx\right)a$$ $\square$

Thm 6.18 设 $f:\mathbb{E}\to\mathbb{R}$ $f(x)=g(\|x\|)$，其中 $g:\mathbb{R}\to(-\infty,\infty]$ 是 proper, closed, convex func 满足 $\mathrm{dom}\,g\subseteq[0,\infty)$，则

$$\mathrm{prox}_f(x)=\begin{cases}\mathrm{prox}_g(\|x\|)\dfrac{x}{\|x\|}, & x\neq 0\\ \{u\in\mathbb{E}:\|u\|=\mathrm{prox}_g(0)\}, & x=0\end{cases}$$

证明：由定义，$\mathrm{prox}_f(0)$ 是以下问题的极小点集：

$$\min_{u\in\mathbb{E}}\left\{f(u)+\frac{1}{2}\|u\|^2\right\}=\min_{u\in\mathbb{E}}\left\{g(\|u\|)+\frac{1}{2}\|u\|^2\right\}$$

令 $w=\|u\|$，结合 $\mathrm{dom}\,g\subseteq[0,\infty)$，知上问题 $\Leftrightarrow$

$$\min_{w\in\mathbb{R}}\left\{g(w)+\frac{1}{2}w^2\right\}$$

上面的优化问题最优点集是 $\mathrm{prox}_g(0)$，故

$$\mathrm{prox}_f(0) = \{u \in \mathbb{E} : \|u\| = \mathrm{prox}_g(0)\}$$

下计算 $x \neq 0$ 时 $\mathrm{prox}_f(x)$，等价于求以下优化问题

$$\min_{u \in \mathbb{E}} \left\{ g(\|u\|) + \frac{1}{2}\|u - x\|^2 \right\}$$

$$= \min_{u \in \mathbb{E}} \left\{ g(\|u\|) + \frac{1}{2}\|u\|^2 - \langle u, x\rangle + \frac{1}{2}\|x\|^2 \right\}$$

$$= \min_{\alpha \in \mathbb{R}_+} \min_{\substack{u \in \mathbb{E} \\ \|u\| = \alpha}} \left\{ g(\alpha) + \frac{1}{2}\alpha^2 - \langle u, x\rangle + \frac{1}{2}\|x\|^2 \right\}$$

由C-S不等式，易证 inner minimization 的 minimizer 是

$$u = \alpha \frac{x}{\|x\|}$$

对应的optimal value是

$$g(\alpha) + \frac{1}{2}\alpha^2 - \alpha\|x\| + \frac{1}{2}\|x\|^2 = g(\alpha) + \frac{1}{2}(\alpha - \|x\|)^2$$

则 $\alpha = \underset{\alpha \in \mathbb{R}_+}{\operatorname{argmin}} \left\{ g(\alpha) + \frac{1}{2}(\alpha - \|x\|)^2 \right\}$

$$= \underset{\alpha \in \mathbb{R}}{\operatorname{argmin}} \left\{ g(\alpha) + \frac{1}{2}(\alpha - \|x\|)^2 \right\}$$

$$= \mathrm{prox}_g(\|x\|)$$

故 $\operatorname{prox}_f(x) = \operatorname{prox}_g(\|x\|)\frac{x}{\|x\|}$ $\boxed{/\!/}$

<u>**Example 6.19**</u> $f: \mathbb{E} \to \mathbb{R}$, $f(x) = \lambda\|x\|$, $\lambda > 0$, $\|\cdot\|$ 是 Euclidean norm, 则 $f(x) = g(\|x\|)$

$$g(t) = \begin{cases} \lambda t, & t \geq 0 \\ \infty, & t < 0 \end{cases}$$

则由 <u>Thm 6.18</u>，对 $\forall x \in \mathbb{E}$

$$\operatorname{prox}_f(x) = \begin{cases} \operatorname{prox}_g(\|x\|)\frac{x}{\|x\|} & , x \neq 0 \\ \{u \in \mathbb{E} : \|u\| = \operatorname{prox}_g(0)\} & , x = 0 \end{cases}$$

由 <u>Lemma 6.5</u>，$\operatorname{prox}_g(t) = [t-\lambda]_+$，故 $\operatorname{prox}_g(0) = 0$，

$\operatorname{prox}_g(\|x\|) = [\|x\| - \lambda]_+$，故

$$\operatorname{prox}_f(x) = \begin{cases} [\|x\| - \lambda]_+ \frac{x}{\|x\|}, & x \neq 0 \\ 0, & x = 0 \end{cases}$$

可以写作下更紧凑的形式：

$$\operatorname{prox}_{\lambda\|\cdot\|}(x) = \left(1 - \frac{\lambda}{\max\{\|x\|, \lambda\}}\right)x .$$

Example 6.20 设 $f(x)=\lambda\|x\|^3$，其中 $\lambda>0$，则 $f(x)=\lambda g(\|x\|)$

$$g(t)=\begin{cases} t^3, & t\geq 0 \\ \infty, & t<0 \end{cases}$$

由 Thm 6.18

$$\operatorname{prox}_f(x)=\begin{cases} \operatorname{prox}_g(\|x\|)\dfrac{x}{\|x\|}, & x\neq 0 \\ \{u\in\mathbb{E}:\|u\|=\operatorname{prox}_g(0)\}, & x=0 \end{cases}$$

由 Lemma 6.5，$\operatorname{prox}_g(t)=\dfrac{-1+\sqrt{1+12\lambda[t]_+}}{6\lambda}$，故

$$\operatorname{prox}_f(x)=\begin{cases} \dfrac{-1+\sqrt{1+12\lambda\|x\|}}{6\lambda}\dfrac{x}{\|x\|}, & x\neq 0 \\ 0, & x=0 \end{cases}$$

故 $\operatorname{prox}_{\lambda\|\cdot\|^3}(x)=\dfrac{2}{1+\sqrt{1+12\lambda\|x\|}}x$ $\boxed{/\!/}$

Example 6.21 $f(x)=-\lambda\|x\|$，$\lambda>0$

与 Example 6.19 同理知：

$$\operatorname{prox}_{-\lambda\|\cdot\|}(x)=\begin{cases} \left(1+\dfrac{\lambda}{\|x\|}\right)x, & x\neq 0 \\ \{u:\|u\|=\lambda\}, & x=0 \end{cases}$$

Example 6.22 $f(x)=\begin{cases}\lambda|x|, & |x|\le\alpha\\ \infty, & \text{else}\end{cases}$, $\lambda,\alpha\in[0,\infty)$

则 $f(x)=g(|x|)$，其中 $g(x)=\begin{cases}\lambda x, & 0\le x\le\alpha\\ \infty, & \text{else}\end{cases}$

由 Thm 6.18，Example 6.14 知

$$\text{prox}_{\lambda|\cdot|+\delta_{[-\alpha,\alpha]}}(x)=\min\{\max\{|x|-\lambda,0\},\alpha\}\,\text{sgn}(x) \qquad \boxed{/\!/}$$

Example 6.23 $f(x)=\begin{cases}\sum_{i=1}^{n}w_i|x_i|, & -\alpha\le x\le\alpha\\ \infty, & \text{else}\end{cases}$

$w\in\mathbb{R}_+^n$，$\alpha\in[0,\infty]^n$. 则 $f=\sum_{i=1}^{n}f_i$，其中

$$f_i(x)=\begin{cases}w_i|x|, & -\alpha_i\le x\le\alpha_i\\ \infty, & \text{else}\end{cases}$$

由 Example 6.22，Thm 6.6 知

$$\text{prox}_f(x)=\left(\min\{\max\{|x_i|-w_i,0\},\alpha_i\}\,\text{sgn}(x_i)\right)_{i=1}^{n} \qquad \boxed{/\!/}$$

§ 6.4 Prox of Indicators - Orthogonal Projection

§6.4.1 The first projection theorem

$g(x)=\delta_C(x)$，C是非空集，则

$$\operatorname{prox}_g(x)=\underset{u\in\mathbb{E}}{\operatorname{argmin}}\left\{\delta_C(u)+\frac{1}{2}\|u-x\|^2\right\}=P_C(x)$$

Thm 6.24 设$C\subseteq\mathbb{E}$非空，则$\operatorname{prox}_{\delta_C}(x)=P_C(x)\quad\forall x\in\mathbb{E}$

Thm 6.25 $C\subseteq\mathbb{E}$非空闭凸集，则$P_C(x)$对$x\in\mathbb{E}$是单点集

证明：由$\delta_C(\cdot)$是proper闭凸函数，结合Thm 6.3即证 $\boxed{/}$

§6.4.2 First Example in $\mathbb{R}^n$

Lemma 6.26

nonnegative orthant $\quad C_1=\mathbb{R}^n_+ \quad [x]_+$

box $\quad C_2=\mathrm{Box}[\ell,u] \quad \left(\min\{\max\{x_i,\ell_i\},u_i\}\right)_{i=1}^n$

affine set $\quad C_3=\{x\in\mathbb{R}^n: Ax=b\} \quad x-A^T(AA^T)^{-1}(Ax-b)$

ℓ_2 ball $\quad C_4=B_{\|\cdot\|_2}[c,r] \quad c+\dfrac{r}{\max\{\|x-c\|_2,r\}}(x-c)$

half-space $\quad C_5=\{x: a^Tx\le\alpha\} \quad x-\dfrac{[a^Tx-\alpha]_+}{\|a\|^2}a$

其中 $\ell \in [-\infty,\infty)^n$，$u \in (-\infty,\infty]^n$．$\ell \le u$，$A \in \mathbb{R}^{m\times n}$ full row rank，$b \in \mathbb{R}^m$，$c \in \mathbb{R}^n$，$r>0$，$a \in \mathbb{R}^n \setminus \{0\}$，$\alpha \in \mathbb{R}$

证明：(i) (ii) 是显然的，下证 (iii)

(iii) $\Leftrightarrow$ 优化问题
$$\min_u \frac{1}{2}\|u-x\|^2$$
$$\text{s.t. } Au=b$$

则 $L(u,\lambda)=\frac{1}{2}\|u-x\|^2+\langle \lambda, Au-b\rangle$

则由KKT条件

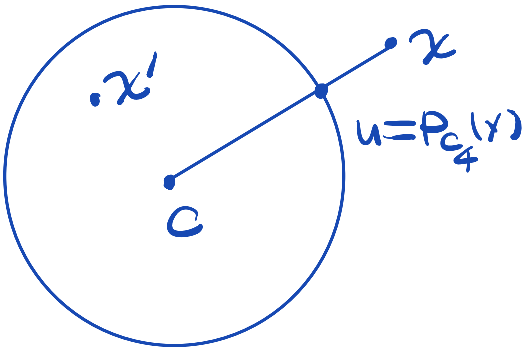
$$\begin{cases} \nabla_u L(u,\lambda) = (u-x)+A^T\lambda = 0 & ① \\ Au=b & ② \end{cases}$$

$$b-Ax+A\cdot A^T\lambda=0$$

$\Rightarrow$ $\lambda=(AA^T)^{-1}(Ax-b)$ 再代入①

$$u=x-A^T(AA^T)^{-1}(Ax-b)$$

(iv)



此时

$$u = c + \frac{r}{\|x-c\|}(x-c)$$

$C_4 = B_{\|\cdot\|_2}[c, r]$

否则 $x' \in C_4$，则 $P_{C_4}(x') = x'$. 综上

$$u = c + \frac{r}{\max\{\|x-c\|_2, r\}}(x-c)$$

(v)

$C_5 = \{x : a^T x \le \alpha\}$.

$x \notin C_5$ 时

$$\min_u \ \frac{1}{2}\|u-x\|^2$$

$$\text{s.t.} \quad a^T u \le \alpha$$

则 $L(u,\lambda) = \frac{1}{2}\|u-x\|^2 + \langle \lambda, a^T u - \alpha \rangle$

其中 $\lambda \in \mathbb{R}_+$

$$\nabla_u L(u,\lambda) = u - x + \lambda a = 0$$

$$a^T u = \alpha$$

$$\Rightarrow \quad \alpha - a^T x + \lambda \|a\|^2 = 0$$

$$\lambda = \frac{a^T x - \alpha}{\|a\|^2}$$

故 $u = x - \lambda a = x - \dfrac{a^T x - \alpha}{\|a\|^2} a$

当 $x \in C_5$ 时，显然 $P_{C_5}(x) = x$，综上

$$P_{C_5}(x) = x - \frac{[a^T x - \alpha]_+}{\|a\|^2} a \qquad \boxtimes$$

§ 6.4.3 Projection onto the Intersection of a Hyperplane and a Box

Thm 6.27 设 $C\subseteq\mathbb{R}^n$，且

$$C=H_{a,b}\cap \mathrm{Box}[l,u]=\{x\in\mathbb{R}^n: a^Tx=b,\ l\le x\le u\}$$

其中 $a\in\mathbb{R}^n\setminus\{0\}$，$b\in\mathbb{R}$，$l\in[-\infty,\infty)^n$，$u\in(-\infty,\infty]^n$

设 $C\neq\phi$，则 $P_C(x)=P_{\mathrm{Box}[l,u]}(x-\mu^*a)$，其中

$\mathrm{Box}[l,u]=\{y\in\mathbb{R}^n: l_i\le y_i\le u_i,\ i=1,\cdots,n\}$，且 μ^* 是

方程 $a^TP_{\mathrm{Box}[l,u]}(x-\mu a)=b$ 的解

证明：$P_C(x)$ 等价于如下优化问题

$$\min_y\left\{\frac{1}{2}\|y-x\|_2^2: a^Ty=b,\ l\le y\le u\right\} \quad (6.20)$$

则 Lagrangian 定义为：

$$\mathcal{L}(y;\mu)=\frac{1}{2}\|y-x\|_2^2+\mu(a^Ty-b)$$

$$=\frac{1}{2}\|y-(x-\mu a)\|_2^2-\frac{\mu^2}{2}\|a\|_2^2+\mu(a^Tx-b)$$

由 Thm A.1，(6.20)是 strong duality 的

因为不涉及非线性不等式约束

故由 Thm A.2 知，y^*是(6.20)的 optimal solution $\Leftrightarrow$

$\exists \mu^* \in \mathbb{R}$, s.t.
$$\begin{cases} y^* \in \underset{\ell \le y \le u}{\operatorname{argmin}}\ \mathcal{L}(y;\mu^*) & (6.22) \\ a^Ty^*=b & (6.23) \end{cases}$$

(6.22) $\Leftrightarrow$ $y^*=P_{\mathrm{Box}[\ell,u]}(x-\mu^*a)$

则可行性条件 $\Leftrightarrow$ $a^TP_{\mathrm{Box}[\ell,u]}(x-\mu^*a)=b$ $\square$

Remark 6.28 记 $\varphi(\mu)=a^TP_{\mathrm{Box}[\ell,u]}(x-\mu a)-b$,

则以上 Thm 6.27 涉及 $\varphi(\mu)$ 的求根问题，由

$$\varphi(\mu) = \sum_{i=1}^{n} a_i \min\{\max\{x_i - \mu a_i, t_i\}, u_i\} - b$$

由 $P_{Box[t,u]}(\cdot)$ 可以显式表示

$$\begin{cases} a_i > 0 \text{时}: x_i - \mu a_i \downarrow, \min\{\max\{x_i - \mu a_i, t_i\}, u_i\} \downarrow, \\ a_i < 0 \text{时}: x_i - \mu a_i \uparrow, \min\{\max\{x_i - \mu a_i, t_i\}, u_i\} \uparrow, \end{cases}$$

故 $\mu \mapsto a_i \min\{\max\{x_i - \mu a_i, t_i\}, u_i\} \downarrow$，故

$\varphi(\mu)$是不增的，可以用二分法求根 $\square$

<u>Corollary 6.29</u> 对 $\forall x \in \mathbb{R}^n$，$P_{\Delta_n}(x) = [x - \mu^* e]_+$

其中 μ^* 是方程 $e^T[x - \mu^* e]_+ - 1 = 0$ 的根

§6.4.4 Projection onto Level Set

Thm 6.30 设 $C = \mathrm{Lev}(f, \alpha) = \{x \in \mathbb{E} : f(x) \le \alpha\}$

$f: \mathbb{E} \to (-\infty, \infty]$ 是 proper, closed, convex, 且 $\alpha \in \mathbb{R}$. 设

$\exists \hat{x} \in \mathbb{E}$, s.t. $f(\hat{x}) < \alpha$. 则

$$P_C(x) = \begin{cases} P_{\mathrm{dom} f}(x), & f(P_{\mathrm{dom} f}(x)) \le \alpha \\ \mathrm{prox}_{\lambda^* f}(x), & \text{else.} \end{cases}$$

其中 λ^* 是以下方程任意正根:

$$\varphi(\lambda) \equiv f(\mathrm{prox}_{\lambda f}(x)) - \alpha = 0$$

更多地, $\varphi(\lambda)$ 不增

证明: $P_C(x)$ 等价于以下优化问题:

$$\min_{y \in \mathbb{E}} \left\{ \frac{1}{2}\|y - x\|^2 : f(y) \le \alpha,\ y \in X \right\}, \quad X = \mathrm{dom} f$$

则 Lagrangian: $\lambda \geq 0$,

$$L(y; \lambda) = \frac{1}{2}\|y - x\|^2 + \lambda f(y) - \alpha \lambda$$

由 $\hat{x}$ 的存在性知优化问题有 slater 条件，故 strong duality 成立，且由 Thm A.2，y^* 是优化问题最优解

$\Leftrightarrow \exists \lambda^* \in \mathbb{R}_+$, s.t.

$$y^* \in \underset{y \in X}{\operatorname{argmin}}\, L(y; \lambda^*)$$

$$f(y^*) \le \alpha$$

$$\lambda^* (f(y^*) - \alpha) = 0$$

- 若 $P_X(x)$ 存在，且 $f(P_X(x)) \le \alpha$，则 $y^* = P_X(x)$

  $\lambda^* = 0$ 是以上 KKT system 的解

- 若 $P_X(x)$ 不存在 or $f(P_X(x)) > \alpha$，则 $\lambda^* > 0$.

  否则 $y^* = P_X(x)$ 推出矛盾，此时 KKT system $\Leftrightarrow$

$$y^* = \operatorname{prox}_{\lambda^* f}(x), \quad f(\operatorname{prox}_{\lambda^* f}(x)) = \alpha$$

下证 φ 的非增性：

$$\operatorname{prox}_{\lambda f}(x) = \operatorname*{argmin}_{y\in X}\left\{\frac{1}{2}\|y-x\|^2 + \lambda(f(y)-\alpha)\right\}$$

取 $0\leqslant\lambda_1<\lambda_2$，记 $v_1=\operatorname{prox}_{\lambda_1 f}(x)$，$v_2=\operatorname{prox}_{\lambda_2 f}(x)$，有

$$\begin{aligned}
&\frac{1}{2}\|v_2-x\|^2+\lambda_2(f(v_2)-\alpha)\\
&=\frac{1}{2}\|v_2-x\|^2+\lambda_1(f(v_2)-\alpha)+(\lambda_2-\lambda_1)(f(v_2)-\alpha)\\
&\geqslant\frac{1}{2}\|v_1-x\|^2+\lambda_1(f(v_1)-\alpha)+(\lambda_2-\lambda_1)(f(v_2)-\alpha)\\
&=\frac{1}{2}\|v_1-x\|^2+\lambda_2(f(v_1)-\alpha)+(\lambda_2-\lambda_1)(f(v_2)-f(v_1))\\
&\geqslant\frac{1}{2}\|v_2-x\|^2+\lambda_2(f(v_2)-\alpha)+(\lambda_2-\lambda_1)(f(v_2)-f(v_1))
\end{aligned}$$

$\Rightarrow (\lambda_2-\lambda_1)(f(v_2)-f(v_1))\leqslant 0$，由 $\lambda_2>\lambda_1$，故

$f(v_2)\leqslant f(v_1)$，从而

$$\varphi(\lambda_2)=f(v_2)-\alpha\leqslant f(v_1)-\alpha=\varphi(\lambda_1) \qquad \square$$

Remark 6.31 注意到 Thm 6.30 中的 f 是闭的，但

dom$f$不一定是闭的（如 $f(x)=\begin{cases}\frac{1}{x}, & x>0\\ \infty, & \text{otherwise}\end{cases}$），

当dom$f$不是闭的，$P_{\mathrm{dom}f}(x)$有可能不存在，故此时

$$P_C(x)=\mathrm{prox}_{\lambda^* f}(x)$$ $\boxed{\angle}$

Example 6.32 设

$$C=H^-_{a,b}\cap \mathrm{Box}[\ell,u]=\{x\in\mathbb{R}^n: a^Tx\le b,\ \ell\le x\le u\}$$

其中 $a\in\mathbb{R}^n\setminus\{0\}$，$b\in\mathbb{R}$，$\ell\in[-\infty,\infty)^n$，$u\in(-\infty,\infty]^n$

设 $C\ne\phi$，则 $C=\mathrm{Lev}(f,b)$，$f=a^Tx+\delta_{\mathrm{Box}[\ell,u]}(x)$.

对 $\forall\lambda>0$，

$$\mathrm{prox}_{\lambda f}(x)=\mathrm{prox}_{\lambda a^T(\cdot)+\delta_{\mathrm{Box}[\ell,u]}(\cdot)}(x)$$

$$\overset{\text{Thm 6.13}}{=}\mathrm{prox}_{\delta_{\mathrm{Box}[\ell,u]}}(x-\lambda a)$$

$$=P_{\mathrm{Box}[\ell,u]}(x-\lambda a)$$

再由Thm 6.30知：

$$P_C(x)=\begin{cases} P_{Box[l,u]}(x), & a^T P_{Box[l,u]}(x)\le b \\ P_{Box[l,u]}(x-\lambda^* a), & a^T P_{Box[l,u]}(x)>b \end{cases}$$

其中 λ^* 是以下非增func的任一正根

$$\varphi(\lambda)=a^T P_{Box[l,u]}(x-\lambda a)-b \qquad \boxed{/}$$

Example 6.33 $C=B_{\|\cdot\|_1}[0,\alpha]=\{x\in\mathbb{R}^n:\|x\|_1<\alpha\}$ $(\alpha>0)$

则 $C=Lev(f,\alpha)$. $f(x)=\|x\|_1$，由Example 6.8

$$\mathrm{prox}_{\lambda f}(x)=T_\lambda(x),\ \forall x\in\mathbb{R}^n$$

其中 $T_\lambda(x)=[x-\lambda e]_+\odot \mathrm{sgn}(x)$，由Thm 6.30

$$P_{B_{\|\cdot\|_1}[0,\alpha]}(x)=\begin{cases} x, & \|x\|\le\alpha \\ T_{\lambda^*}(x), & \|x\|>\alpha \end{cases}$$

λ^* 是方程 $\varphi(\lambda)=\|T_\lambda(x)\|_1-\alpha$ 的任一正根. $\boxed{/}$

定义双边 soft thresholding operator：$\forall a,b\in(-\infty,\infty]^n$

$$S_{a,b}(x)=\left(\min\{\max\{|x_i|-a_i,0\},b_i\}\operatorname{sgn}(x_i)\right)_{i=1}^n$$

<u>Example 6.34</u> $C=\left\{x\in\mathbb{R}^n:\sum_{i=1}^n w_i|x_i|\le\beta,\ -\alpha\le x\le\alpha\right\}$

其中 $w\in\mathbb{R}_+^n$，$\alpha\in[0,\infty]^n$，$\beta\in\mathbb{R}_{++}$，则 $C=\operatorname{Lev}(f,\beta)$

$$f(x)=w^T|x|+\delta_{\operatorname{Box}[-\alpha,\alpha]}(x)=\begin{cases}\sum_{i=1}^n w_i|x_i|, & -\alpha\le x\le\alpha\\ \infty, & \text{else}\end{cases}$$

由<u>Example 6.23</u>知：

$$\operatorname{prox}_{\lambda f}(x)=\left(\min\{\max\{|x_i|-\lambda w_i,0\},\alpha_i\}\operatorname{sgn}(x_i)\right)_{i=1}^n$$

$$=S_{\lambda w,\alpha}(x).$$

故由<u>Thm 6.30</u>，有

$$P_C(x)=\begin{cases}P_{\operatorname{Box}[-\alpha,\alpha]}(x), & w^T|P_{\operatorname{Box}[-\alpha,\alpha]}(x)|\le\beta\\ S_{\lambda^* w,\alpha}(x), & w^T|P_{\operatorname{Box}[-\alpha,\alpha]}(x)|>\beta\end{cases}$$

其中λ^*是以下非增func的任一正根：

$$\varphi(\lambda) = w^T \left| S_{\lambda w, \alpha}(x) \right| - \beta$$

<u>Example 6.35</u> 设 $C = \left\{ x \in \mathbb{R}^n_{++} : \prod_{i=1}^{n} x_i \geq \alpha \right\}$

则 $C = \left\{ x \in \mathbb{R}^n_{++} : -\sum_{i=1}^{n} \log x_i \leq -\log \alpha \right\}$

故 $C = \mathrm{Lev}(f, -\log \alpha)$，其中

$$f(x) = \begin{cases} -\sum_{i=1}^{n} \log x_i, & x \in \mathbb{R}^n_{++} \\ \infty, & \text{else} \end{cases}$$

由<u>Example 6.9</u>，有

$$\mathrm{prox}_{\lambda f}(x) = \left(\frac{x_j + \sqrt{x_j^2 + 4\lambda}}{2} \right)_{j=1}^{n}$$

- 若 $x \notin \mathbb{R}^n_{++}$，则 $P_{\mathbb{R}^n_{++}}(x)$ 不存在，此时 $P_C(x) = \mathrm{prox}_{\lambda^* f}(x)$

综上：

$$P_C(x) = \begin{cases} x & x \in C \\ \left(\dfrac{x_j + \sqrt{x_j^2 + 4\lambda^*}}{2} \right)_{j=1}^{n}, & x \notin C \end{cases}$$

其中λ^*是以下方程任一正根：

$$\varphi(\lambda) = -\sum_{j=1}^{n} \log\left(\frac{x_j + \sqrt{x_j^2 + 4\lambda}}{2} \right) + \log \alpha \qquad \square$$

§ 6.4.5 Projection onto Epigraphs

Thm 6.36 令 $C=\mathrm{epi}(g)=\{(x,t)\in\mathbb{E}\times\mathbb{R}: g(x)\le t\}$

$g:\mathbb{E}\to\mathbb{R}$ convex，则

$$P_C((x,s))=\begin{cases}(x,s) & , g(x)\le s\\ (\mathrm{prox}_{\lambda^* g}(x), s+\lambda^*) & , g(x)>s\end{cases}$$

其中 λ^* 是以下func的任一正根：

$$\psi(\lambda)=g\left(\mathrm{prox}_{\lambda g}(x)\right)-\lambda-s$$

且 ψ 是非增的

证明：定义 $f:\mathbb{E}\times\mathbb{R}\to\mathbb{R}$ $f(x,t)=g(x)-t$

$$\mathrm{prox}_{\lambda f}(x,s)=\underset{y,t}{\mathrm{argmin}}\left\{\frac{1}{2}\|y-x\|^2+\frac{1}{2}(t-s)^2+\lambda f(y,t)\right\}$$

$$= \underset{y,t}{\operatorname{argmin}} \left\{ \frac{1}{2}\|y-x\|^2 + \frac{1}{2}(t-s)^2 + \lambda g(y) - \lambda t \right\}$$

上面优化问题关于 y,t 是可分的

$$\operatorname{prox}_{\lambda f}(x,s) = \left(\underset{y}{\operatorname{argmin}} \left\{ \frac{1}{2}\|y-x\|^2 + \lambda g(y) \right\}, \right.$$

$$\left. \underset{t}{\operatorname{argmin}} \left\{ \frac{1}{2}(t-s)^2 - \lambda t \right\} \right)$$

$$= \left(\operatorname{prox}_{\lambda g}(x), \operatorname{prox}_{\lambda h}(s) \right)$$

其中 $h(t) \equiv -t$. $\operatorname{prox}_{\lambda h}(\cdot)$ 是二次 func，易知 $\operatorname{prox}_{\lambda h}(z) = z + \lambda$

故 $\operatorname{prox}_{\lambda f}(x,s) = \left(\operatorname{prox}_{\lambda g}(x), s + \lambda \right)$

由 $\operatorname{epi}(g) = \operatorname{Lev}(f,0)$，故由 Thm 6.30：

$$P_C((x,s)) = \begin{cases} (x,s), & g(x) \le s \\ \left(\operatorname{prox}_{\lambda^* g}(x), s + \lambda^* \right), & g(x) > s \end{cases}$$

其中 λ^* 是以下 func 的任一正根：

$$\psi(\lambda) = g\left(\operatorname{prox}_{\lambda g}(x) \right) - \lambda - s$$

由 Thm 6.30，上 φ 非增 $\boxed{/}$

Example 6.37 Lorentz cone: $L^n = \{(x,t)\in\mathbb{R}^n\times\mathbb{R} : \|x\|_2 \le t\}$

下证：对 $\forall (x,s)\in\mathbb{R}^n\times\mathbb{R}$

$$P_{L^n}(x,s) = \begin{cases} \left(\dfrac{\|x\|_2+s}{2\|x\|_2}x, \dfrac{\|x\|_2+s}{2}\right), & \|x\|_2 \geq |s| \\ (0,0) & , s < \|x\|_2 < -s \\ (x,s) & , \|x\|_2 \le s \end{cases}$$

证明：由 Thm 6.36：

$$P_{L^n}((x,s)) = \begin{cases} (x,s) & , \|x\|_2 \le s \\ (\mathrm{prox}_{\lambda^*\|\cdot\|_2}(x), s+\lambda^*) & , \|x\|_2 > s \end{cases}$$

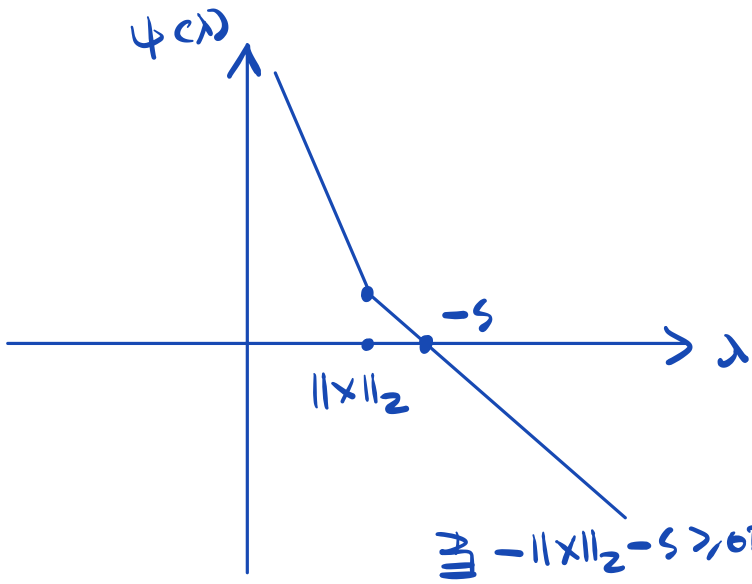
其中 λ^* 是以下 func 的唯一正根：

$$\psi(\lambda) = \|\mathrm{prox}_{\lambda^*\|\cdot\|_2}(x)\|_2 - \lambda - s$$

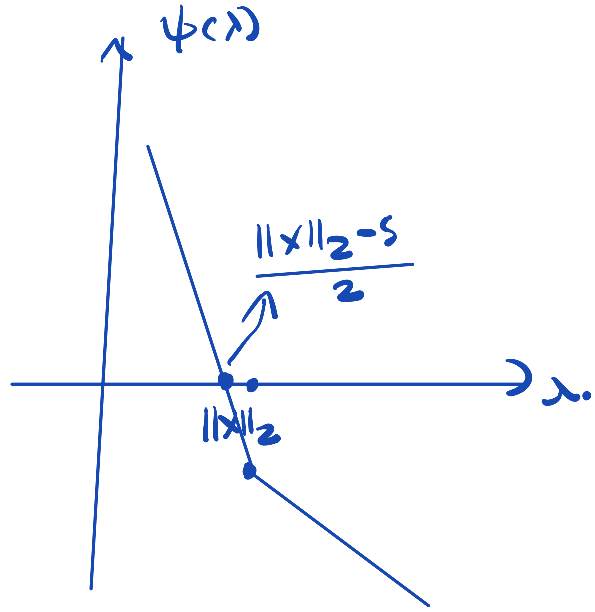
设 $(x,s)\in\mathbb{R}^n\times\mathbb{R}$, s.t. $\|x\|_2 > s$, 由 Example 6.19

$$\text{prox}_{\lambda\|\cdot\|_2}(x)=\left[1-\frac{\lambda}{\max\{\|x\|_2,\lambda\}}\right]_+ x$$

故 $$\psi(\lambda)=\begin{cases}\|x\|_2-2\lambda-s, & \lambda\le\|x\|_2\\ -\lambda-s, & \lambda\geq\|x\|_2\end{cases}$$



当 $-\|x\|_2-s\geq 0$ 时



当 $-\|x\|_2-s\le 0$ 时

综上：$$\lambda^*=\begin{cases}\dfrac{\|x\|_2-s}{2}, & \|x\|_2\geq -s\\ -s, & \|x\|_2< -s\end{cases}$$

故 $\|x\|_2>s$ 时

$$\left(\text{prox}_{\lambda^*\|\cdot\|_2}(x),\ s+\lambda^*\right)$$

$$=\left(\left[1-\frac{\lambda^*}{\max\{\|x\|_2,\lambda^*\}}\right]_+ x,\ s+\lambda^*\right)$$

- 当 $\|x\|_2 \geq -s$ 时，$\lambda^* = \frac{\|x\|_2 - s}{2} \leq \|x\|_2$，故 $\max\{\|x\|_2, \lambda^*\} = \|x\|_2$
- 当 $\|x\|_2 < -s$ 时，$\lambda^* = -s > \|x\|_2$，故 $\max\{\|x\|_2, \lambda^*\} = \lambda^*$

$$= \begin{cases} \left(\left[1 - \frac{\|x\|_2 - s}{2\|x\|_2}\right]_+ x, \frac{\|x\|_2 + s}{2} \right), & \|x\|_2 \geq -s \\ (0,0), & \|x\|_2 < -s \end{cases}$$

$$= \begin{cases} \left(\frac{\|x\|_2 + s}{2\|x\|_2} x, \frac{\|x\|_2 + s}{2} \right), & \|x\|_2 \geq -s \\ (0,0), & \|x\|_2 < -s \end{cases}$$

故 $P_{\perp n}(x,s) = (0,0)$ 当 $s < \|x\|_2 < -s$；当 $\begin{cases} \|x\|_2 > s \\ \|x\|_2 \geq -s \end{cases}$ 时

$$P_{\perp n}(x,s) = \left(\frac{\|x\|_2 + s}{2\|x\|_2} x, \frac{\|x\|_2 + s}{2} \right)$$

又由：$\{(x,s) : \|x\|_2 \geq |s|\} = \{(x,s) : \|x\|_2 > s, \|x\|_2 \geq -s\}$

$$\cup \{(x,s) : \|x\|_2 = s\}$$

而 $\|x\|_2=s$ 时，$P_{L^n}(x,s)=(x,s)=\left(\frac{\|x\|_2+s}{2\|x\|_2}x,\ \frac{\|x\|_2+s}{2}\right)$

综上即证 $\square$

Example 6.38 设 $C=\{(y,t)\in\mathbb{R}^n\times\mathbb{R}:\|y\|_1\le t\}$

$$P_C((x,s))=\begin{cases}(x,s) & ,\ \|x\|_1\le s\\ (\mathcal{T}_{\lambda^*}(x),\ s+\lambda^*) & ,\ \|x\|_1>s\end{cases}$$

其中 λ^* 是以下func的任一正根：

$$\varphi(\lambda)=\|\mathcal{T}_\lambda(x)\|_1-\lambda-s$$ $\square$

§6.5 The Second Prox Theorem

Thm 6.39 设 $f:\mathbb{E}\to(-\infty,\infty]$ 是proper闭凸func，则

对 $\forall x,u\in\mathbb{E}$，以下3点等价：

(i) $u=\operatorname{prox}_f(x)$

(ii) $x-u\in\partial f(u)$

(iii) $\langle x-u, y-u\rangle \le f(y)-f(u), \forall y \in \mathbb{E}$

注：$\|x\|^2=\langle x,x\rangle$，则有

$$\langle \nabla\|x\|^2, h\rangle := \lim_{s\to 0}\frac{1}{s}\left[\|x+sh\|^2-\|x\|^2\right]$$

$$=\lim_{s\to 0}\frac{1}{s}\left[2s\langle x,h\rangle+s^2\langle h,h\rangle\right]$$

$$=\langle 2x, h\rangle$$

由对 $\forall$ 方向 h 以上成立，则由数学分析知 $\nabla\|x\|^2=2x$

证明：由定义，$u=\mathrm{prox}_f(x) \iff u$ 是下优化问题 minimizer

$$\min_v\left\{f(v)+\frac{1}{2}\|v-x\|^2\right\}$$

由 Fermat's 引理，以上 $\iff 0\in\partial f(u)+u-x$

故 (i) (ii) 等价。(ii) (iii) 的等价性由次梯度的定义知 $\boxtimes$

Corollary 6.40 设 f 是 proper 闭凸 func，则 x 是 f 的 minimizer $\iff x=\mathrm{prox}_f(x)$

证明：$x\in\underset{u}{\operatorname{argmin}} f(u) \Leftrightarrow 0\in\partial f(x)$

$\Leftrightarrow x-x\in\partial f(x) \Leftrightarrow x=\operatorname{prox}_f(x)$ $\boxed{/}$

<u>Thm 6.41</u> 设 $C\subseteq E$ 非空闭凸，设 $u\in C$，则 $u=P_C(x)$ iff

$\langle x-u, y-u\rangle\leq 0$ 对 $\forall y\in C$

证明：在 <u>Thm 6.39</u> 取 $f=\delta_C$，取由 (i) (iii) 等价

立即可证 $\boxed{/}$

<u>Thm 6.42</u> 设 f 是 proper 闭凸的，$\forall x,y\in E$

(a) $\langle x-y, \operatorname{prox}_f(x)-\operatorname{prox}_f(y)\rangle \geq \|\operatorname{prox}_f(x)-\operatorname{prox}_f(y)\|^2$

(b) $\|\operatorname{prox}_f(x)-\operatorname{prox}_f(y)\|\leq\|x-y\|$

证明：

(a) 记 $u=\operatorname{prox}_f(x)$，$v=\operatorname{prox}_f(y)$

由 Thm 6.39：$x-u\in\partial f(u)$，$y-v\in\partial f(v)$

由次梯度定义：$\begin{cases} f(v)\geq f(u)+\langle x-u, v-u\rangle \\ f(u)\geq f(v)+\langle y-v, u-v\rangle \end{cases}$

$\Rightarrow \quad 0\geq\langle y-x+u-v, u-v\rangle$

$\Leftrightarrow \quad \langle x-y, u-v\rangle\geq\|u-v\|^2$

(b) 若 $\text{prox}_f(x)=\text{prox}_f(y)$，则显然成立

设 $\text{prox}_f(x)\neq\text{prox}_f(y)$，则

$$\|\text{prox}_f(x)-\text{prox}_f(y)\|^2\leq\langle\text{prox}_f(x)-\text{prox}_f(y), x-y\rangle$$

$$\leq\|\text{prox}_f(x)-\text{prox}_f(y)\|\,\|x-y\|$$

两边同除 $\|\text{prox}_f(x)-\text{prox}_f(y)\|$ 即证 $\boxed{\angle}$

Lemma 6.43 设 $C\subseteq E$ 非空闭凸，$\lambda>0$，则 $\forall x\in E$

$$\operatorname{prox}_{\lambda d_C}(x)=\begin{cases}(1-\theta)x+\theta P_C(x), & d_C(x)>\lambda\\ P_C(x), & d_C(x)\le\lambda\end{cases}$$

其中 $\theta=\dfrac{\lambda}{d_C(x)}$

证明：设 $u=\operatorname{prox}_{\lambda d_C}(x)$，则由 Thm 6.39：

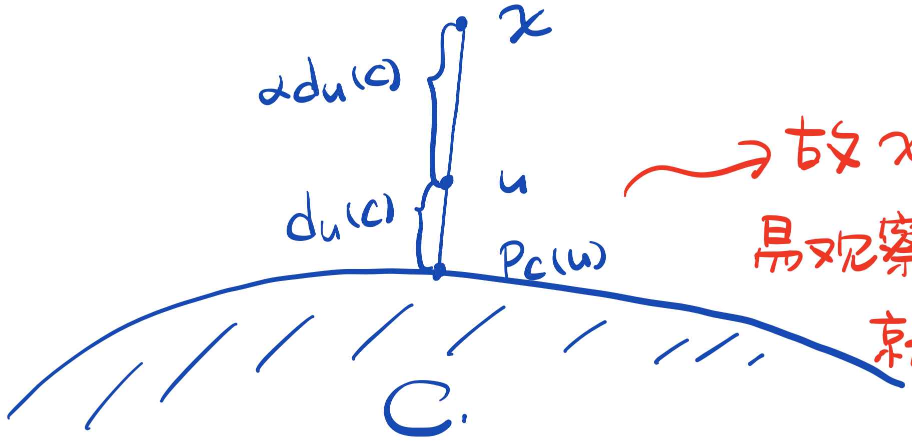
$$x-u\in\lambda\,\partial d_C(u) \qquad (6.34)$$

Case I：$u\notin C$，此时 (6.34) 等价于：

$$x-u=\lambda\,\frac{u-P_C(u)}{d_C(u)}$$

记 $\alpha=\dfrac{\lambda}{d_C(u)}$，则上式 $\Leftrightarrow$

$$u=\frac{1}{\alpha+1}x+\frac{\alpha}{\alpha+1}P_C(u)$$



故 $x, u, P_C(u)$ 共线；且易观察 $P_C(u)=P_C(x)$。下面就是证明这个观察.

由Thm6.41，证明 $P_C(u)=P_C(x) \Leftrightarrow$

$$\langle x-P_C(u), y-P_C(u)\rangle \le 0, \ \forall y\in C$$

$$\Rightarrow (\alpha+1)\langle u-P_C(u), y-P_C(u)\rangle \le 0 \quad \forall y\in C$$

由投影的性质，上式成立，则 $P_C(u)=P_C(x)$

$$\Rightarrow u=\frac{1}{\alpha+1}x+\frac{\alpha}{\alpha+1}P_C(x)$$ 但 α 与 u 相关

且 $d_C(x)=(\alpha+1)d_C(u)=d_C(u)+\lambda$

$$\frac{1}{\alpha+1}=\frac{d_C(u)}{\lambda+d_C(u)}=\frac{d_C(x)-\lambda}{d_C(x)}=1-\theta$$

$$\Rightarrow u=(1-\theta)x+\theta P_C(x)$$

Case II：若 $u\in C$，则 $u=P_C(x)$，下证明这个事实：

设 $v\in C$，由 $u=\text{prox}_{\lambda d_C}(x)$，则：

$$\lambda d_C(u)+\frac{1}{2}\|u-x\|^2\le\lambda d_C(v)+\frac{1}{2}\|v-x\|^2$$

由 $d_C(u)=d_C(v)=0$，知：

$$u=\operatorname*{argmin}_{v\in C}\|v-x\|=P_C(x)$$

由<u>Example 3.49</u> 知：最优性条件 (6.34) $\Leftrightarrow$

$$\frac{x-P_C(x)}{\lambda}\in N_C(u)\cap B[0,1]$$

$$\Rightarrow\quad \left\|\frac{x-P_C(x)}{\lambda}\right\|\leq 1 \quad\Leftrightarrow\quad d_C(x)=\|x-P_C(x)\|\leq\lambda$$

由 case I（即 $u\notin C$ 时）可以推出 $d_C(x)>\lambda$

而 case II（即 $u\in C$ 时）可以推出 $d_C(x)\leq\lambda$

故知可以将 $d_C(x)>\lambda$ / $d_C(x)\leq\lambda$ 作为区分 case I, II 的准则，故

$$\operatorname{prox}_{\lambda d_C}(x)=\begin{cases}(1-\theta)x+\theta P_C(x), & d_C(x)>\lambda\\ P_C(x), & d_C(x)\leq\lambda\end{cases}$$

$\boxtimes$

§6.6 Moreau Decomposition

Thm 6.44 设 $f: \mathbb{E} \to (-\infty, \infty]$ 是 proper, convex, closed

对 $\forall x \in \mathbb{E}$： $\operatorname{prox}_f(x) + \operatorname{prox}_{f^*}(x) = x$

证明：设 $x \in \mathbb{E}$，记 $u = \operatorname{prox}_f(x)$，由 Thm 6.39：

$$x - u \in \partial f(u) \qquad ①$$

由 Thm 4.20（结合 f 的闭凸性），知 ① $\Leftrightarrow$

$$u \in \partial f^*(x-u) \Leftrightarrow x - (x-u) \in \partial f^*(x-u)$$

再由 Thm 6.39，$x - u = \operatorname{prox}_{f^*}(x)$，故：

$$\operatorname{prox}_f(x) + \operatorname{prox}_f(x) = x \qquad \square$$

Thm 6.45：设 $f: \mathbb{E} \to (-\infty, \infty]$ 是 proper, closed, convex,

设 $\lambda > 0$，则对 $\forall x \in \mathbb{E}$：

$$\operatorname{prox}_{\lambda f}(x) + \lambda \operatorname{prox}_{\lambda^{-1} f^*}(x/\lambda) = x$$

证明：由 Moreau 分解：

$$\mathrm{prox}_{\lambda f}(x) = x - \mathrm{prox}_{(\lambda f)^*}(x) \overset{\text{Thm 4.14(a)}}{=} x - \mathrm{prox}_{\lambda f^*(\cdot/\lambda)}(x)$$

$$\overset{\text{Thm 6.12}}{=} x - \lambda \mathrm{prox}_{\lambda^{-1} f^*}\left(\frac{x}{\lambda}\right) \qquad \boxed{/}$$

§6.6.1 Support functions

<u>Thm 6.46</u> 设 $C \subseteq \mathbb{E}$ 是非空闭凸集，设 $\lambda > 0$，则 $\forall x \in \mathbb{E}$

$$\mathrm{prox}_{\lambda \sigma_C}(x) = x - \lambda P_C\left(\frac{x}{\lambda}\right)$$

证明：由 <u>Thm 6.45</u>：

$$\mathrm{prox}_{\lambda \sigma_C}(x) = x - \lambda \mathrm{prox}_{\lambda^{-1} \sigma_C^*}\left(\frac{x}{\lambda}\right)$$

$$= x - \lambda \mathrm{prox}_{\lambda^{-1} \delta_C}\left(\frac{x}{\lambda}\right)$$

$$= x - \lambda \operatorname*{argmin}_u \left\{ \lambda^{-1} \delta_C(u) + \frac{1}{2}\left\| \frac{x}{\lambda} - u \right\|^2 \right\}$$

$$= x - \lambda P_C\left(\frac{x}{\lambda}\right) \qquad \boxed{/}$$

Example 6.47 设 $f: \mathbb{E} \to \mathbb{R}$ 是 $f(x) = \lambda \|x\|_\alpha$, $\lambda > 0$

由例 2.31 知 $\|x\|_\alpha = \sigma_C(x)$

其中 $C = B_{\|\cdot\|_{\alpha,*}}[0,1] = \{x \in \mathbb{E} : \|x\|_{\alpha,*} \le 1\}$

（由有限维空间都是自反的）

故 $\operatorname{prox}_{\lambda\|\cdot\|_\alpha}(x) = x - \lambda P_{B_{\|\cdot\|_{\alpha,*}}[0,1]}\left(\frac{x}{\lambda}\right)$ $\square$

Example 6.48 $\|\cdot\|_\alpha$ 取 $\|\cdot\|_\infty$ 时：

$$\operatorname{prox}_{\lambda\|\cdot\|_\infty}(x) = x - \lambda P_{B_{\|\cdot\|_1}[0,1]}\left(\frac{x}{\lambda}\right) \quad \square$$

Example 6.49 $g(x) = \max(x) = \max\{x_1, \cdots, x_n\}$

则由 $\max(x) = \sigma_{\Delta_n}(x)$ 知：

$$\operatorname{prox}_{\lambda\max(\cdot)}(x) = x - \lambda P_{\Delta_n}\left(\frac{x}{\lambda}\right) \quad \square$$

Example 6.50 $f(x) = x_{[1]} + \cdots + x_{[k]}$.

其中 $k \in \{1, 2, \cdots, n\}$, $x_{[i]}$ 表示 x 中第 i 大的元素:

claim: $f = \sigma_C$, $C = \{y \in \mathbb{R}^n : e^T y = k,\ 0 \leq y \leq e\}$

证明:

① $f \leq \sigma_C$ 是显然的, 只需取 $\tilde{y} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \vdots \\ 1 \\ \vdots \end{bmatrix}$, 取1项是 $x_{[i]}$ 对应的坐标 $(i = 1, \cdots, k)$ 即证:

$$\sigma_C(x) \geq \langle \tilde{y}, x \rangle = f(x)$$

$y_{[i]}$ 不能理解成 y 的 i-largest 分量!

② 下证 $f \geq \sigma_C$: $\forall x \in \mathbb{R}^n$, $\forall y \in C$

这里 $[i]$ 表示 x i-largest 的坐标, 所以

$$\sum_{i=1}^n x_i y_i = \sum_{i=1}^k x_{[i]} y_{[i]} + \sum_{i=k+1}^n x_{[i]} y_{[i]}$$

$$\leq \sum_{i=1}^k x_{[i]} y_{[i]} + x_{[k+1]} \underbrace{\sum_{i=k+1}^n y_{[i]}}_{=t}$$

$$= k \left[\sum_{i=1}^k x_{[i]} \left(\frac{1}{k} y_{[i]} \right) + \frac{t}{k} x_{[k+1]} \right]$$

下证: $\sum_{i=1}^{k} x_{[i]}\left(\frac{1}{k} y_{[i]}\right) + \frac{t}{k} x_{[k+1]} \le \frac{1}{k}\sum_{i=1}^{k} x_{[i]}$

$$\Leftrightarrow \frac{1}{k}\sum_{i=1}^{k} x_{[i]}(y_{[i]}-1) \le -\frac{t}{k} x_{[k+1]}$$

由 $\sum_{i=1}^{k} x_{[i]} \frac{1-y_{[i]}}{k} \ge x_{[k]} \sum_{i=1}^{k} \frac{1-y_{[i]}}{k}$

$$= x_{[k]} \frac{k-(k-t)}{k} = x_{[k]}\frac{t}{k} \ge x_{[k+1]}\frac{t}{k}$$

即证 $\sigma_C \le f$

故由 Thm 6.46: $\operatorname{prox}_{\lambda f}(x) = x - \lambda P_C\left(\frac{x}{\lambda}\right)$ $\square$

Example 6.51 $f(x) = \sum_{i=1}^{k} |x_{\langle i\rangle}|$

$k\in\{1,\cdots,n\}$, $x_{\langle i\rangle}$ 是 x 中 ith largest 绝对值, i.e.,

$|x_{\langle 1\rangle}| \ge |x_{\langle 2\rangle}| \ge \cdots \ge |x_{\langle n\rangle}|$, 则

$$f(x) = \max\left\{\sum_{i=1}^{n} z_i x_i : \|z\|_1 \le k, -e \le z \le e\right\}$$

（证明思路与例 6.50 完全一致）

故 $f=\sigma_C$，$C=\{z\in\mathbb{R}^n: \|z\|_1\le k, -e\le z\le e\}$

$\mathrm{prox}_{\lambda f}(x)=x-\lambda P_C(\frac{x}{\lambda})$ $\square$

§ 6.7 The Moreau Envelope

Def：给定 proper, closed, convex f，$\mu>0$，f 的 Moreau Envelope 定义为：$M_f^\mu(x)=\min\limits_{u\in\mathbb{E}}\{f(u)+\frac{1}{2\mu}\|x-u\|^2\}$

注：μ 称为 smoothing parameter，且由 Thm 6.3，上优化问题有唯一 optimal solution：$\mathrm{prox}_{\mu f}(x)$，i.e.,

$$M_f^\mu(x)=f(\mathrm{prox}_{\mu f}(x))+\frac{1}{2\mu}\|x-\mathrm{prox}_{\mu f}(x)\|^2$$

Example 6.53：$f=\delta_C(x)$，$C\subseteq\mathbb{E}$ 非空闭凸。

则由 Thm 6.24，$\mathrm{prox}_{\mu f}(x)=P_C(x)$，故

$$M_f^\mu(x)=\delta_C(P_C(x))+\frac{1}{2\mu}\|x-P_C(x)\|^2=\frac{1}{2\mu}d_C^2(x)$$ $\square$

Example 6.54 (Huber func) $f(x)=\|x\|$

由例 6.19：对 $\forall x\in \mathbb{E}$, $\mu>0$

$$\mathrm{prox}_{\mu f}(x)=\left(1-\frac{\mu}{\max\{\|x\|,\mu\}}\right)x$$

故 $M_f^{\mu}(x)=\|\mathrm{prox}_{\mu f}(x)\|+\frac{1}{2\mu}\|x-\mathrm{prox}_{\mu f}(x)\|^2$

$$=\begin{cases}\frac{1}{2\mu}\|x\|^2, & \|x\|\le\mu\\ \|x\|-\frac{\mu}{2}, & \|x\|>\mu\end{cases}$$

$$\doteq H_\mu(x) \quad (\text{Huber func})$$

$\boxtimes$

定义：$\omega_\mu(x)=\frac{1}{2\mu}\|x\|^2$，则 $M_f^{\mu}=f\,\square\,\omega_\mu$

Thm 6.55 f 是 proper, closed, convex func，则

(a) $M_f^{\mu}=f\,\square\,\omega_\mu$;

(b) $M_f^{\mu}:\mathbb{E}\to\mathbb{R}$ 是实值 convex 的

证明：$M_f^\mu = f \square \omega_\mu$ 由定义即证，由 f proper convex，ω_μ 是实值凸 func，故 M_f^μ 是 convex 的 $\square$

<u>Corollary 6.56</u> 设 f 是 proper, closed, convex 的，则

$$(M_f^\mu)^* = f^* + \omega_{\frac{1}{\mu}}$$

证明：由 <u>Thm 4.16</u> 即证 $\square$

<u>Lemma 6.57</u> 设 f 是 proper, closed, convex 的，$\lambda, \mu > 0$，则

$$\forall x \in \mathbb{E},\ \lambda M_f^\mu(x) = M_{\lambda f}^{\mu/\lambda}(x)$$

证明：对 $\forall x \in \mathbb{E}$

$$\begin{aligned}
\lambda M_f^\mu(x) &= \lambda \min_u \left\{ f(u) + \frac{1}{2\mu}\|x-u\|^2 \right\} \\
&= \min_u \left\{ \lambda f(u) + \frac{1}{2\mu/\lambda}\|x-u\|^2 \right\} \\
&= M_{\lambda f}^{\mu/\lambda}(x)
\end{aligned}$$

$\square$

<u>Thm 6.58</u> $\mathbb{E} = \mathbb{E}_1 \times \cdots \times \mathbb{E}_m$，且

$$f(x_1,\cdots,x_m)=\sum_{i=1}^{m}f_i(x_i),\quad x_i\in\mathbb{E}_i,\ i=1,\cdots,m$$

其中 f_i 是 proper, closed, convex 的，则 $\forall\mu>0$

$$M_f^{\mu}(x_1,\cdots,x_m)=\sum_{i=1}^{m}M_{f_i}^{\mu}(x_i)$$

证明:

$$M_f^{\mu}(x)=\min_{\substack{u_i\in\mathbb{E}_i\\ i=1,\cdots,m}}\left\{f(u_1,\cdots,u_m)+\frac{1}{2\mu}\|(u_1,\cdots,u_m)-x\|^2\right\}$$

$$=\min_{\substack{u_i\in\mathbb{E}_i\\ i=1,\cdots,m}}\left\{\sum_{i=1}^{m}f_i(u_i)+\frac{1}{2\mu}\sum_{i=1}^{m}\|u_i-x_i\|^2\right\}$$

$$=\sum_{i=1}^{m}\min_{u_i\in\mathbb{E}}\left\{f_i(u_i)+\frac{1}{2\mu}\|u_i-x_i\|^2\right\}$$

$$=\sum_{i=1}^{m}M_{f_i}^{\mu}(x_i)\qquad\square$$

Example 6.59 $f(x)=\|x\|_1=\sum_{i=1}^{n}g(x_i)$，其中 $g(t)=|t|$

$$M_f^{\mu}(x)=\sum_{i=1}^{n}M_g^{\mu}(x_i)=\sum_{i=1}^{n}H_{\mu}(x_i)\qquad\square$$

§ 6.7.2 Differentiability of the Moreau Envelope

Thm 6.60 f 是 proper, closed, convex func，设 $\mu>0$，则

M_f^μ 是 $\frac{1}{\mu}$-smooth 的 func，且对 $\forall x\in\mathbb{E}$

$$\nabla M_f^\mu(x)=\frac{1}{\mu}(x-\mathrm{prox}_{\mu f}(x))$$

证明：Thm 6.55(a) $\Rightarrow$ $M_f^\mu=f\,\square\,\omega_\mu$，故由 Thm 5.30 知

M_f^μ 是 $\frac{1}{\mu}$-smooth 的，且由 Thm 5.30 (b) $(u(x)=\mathrm{prox}_{\mu f}(x))$

$$\nabla M_f^\mu(x)=\nabla\omega_\mu(x-u(x))=\frac{1}{\mu}(x-\mathrm{prox}_{\mu f}(x)) \qquad \square$$

Example 6.61 设 $C\subseteq\mathbb{E}$ 非空闭凸集，由例 6.53，

$\frac{1}{2}d_C^2=M_{\delta_C}^1$，故 $\frac{1}{2}d_C^2$ 是 1-smooth 的，且

$$\nabla\left(\frac{1}{2}d_C^2\right)(x)=x-\mathrm{prox}_{\delta_C}(x)=x-P_C(x) \qquad \square$$

Example 6.62 Huber func:

$$H_\mu(x)=\begin{cases}\frac{1}{2\mu}\|x\|^2, & \|x\|\le\mu\\ \|x\|-\frac{\mu}{2}, & \|x\|>\mu\end{cases}$$

由例6.54，$H_\mu = M_f^\mu$，其中 $f(x)=\|x\|$，故 H_μ 是 $\frac{1}{\mu}$-smooth

$$\nabla H_\mu(x) = \frac{1}{\mu}\left(x - \mathrm{prox}_{\mu f}(x)\right)$$

$$\overset{\text{例6.19}}{=} \frac{1}{\mu}\left(x - \left(1 - \frac{\mu}{\max\{\|x\|, \mu\}}\right)x\right)$$

$$= \begin{cases} \frac{1}{\mu}x, & \|x\| \le \mu \\ \frac{x}{\|x\|}, & \|x\| > \mu \end{cases} \qquad \square$$

§ 6.7.3 Prox of the Moreau Envelope

<u>Thm 6.63</u> 设 $f:\mathbb{E}\to(-\infty,\infty]$ 是 proper，闭凸 func. $\mu>0$

则对 $\forall x\in\mathbb{E}$

$$\operatorname{prox}_{M_f^{\mu}}(x)=x+\frac{1}{\mu+1}\left(\operatorname{prox}_{(\mu+1)f}(x)-x\right)$$

证明：

$$\min_u\left\{M_f^{\mu}(u)+\frac{1}{2}\|u-x\|^2\right\}=\min_u\min_y\left\{f(y)+\frac{1}{2\mu}\|u-y\|^2+\frac{1}{2}\|u-x\|^2\right\}$$

由 $\min\limits_u\min\limits_y$ 的交换性：等价于下优化问题：

$$\min_y\min_u\left\{f(y)+\frac{1}{2\mu}\|u-y\|^2+\frac{1}{2}\|u-x\|^2\right\}\quad(6.47)$$

inner 优化问题最优 $\Leftrightarrow$ $\frac{1}{\mu}(u-y)+(u-x)=0$

$$u=u_\mu\equiv\frac{\mu x+y}{\mu+1}$$，代入 (6.47)

$$f(y)+\frac{1}{2\mu}\|u_\mu-y\|^2+\frac{1}{2}\|u_\mu-x\|^2$$

$$= f(y) + \frac{1}{2\mu}\left\|\frac{\mu(x-y)}{\mu+1}\right\|^2 + \frac{1}{2}\left\|\frac{y-x}{\mu+1}\right\|^2$$

$$= f(y) + \frac{1}{2(\mu+1)}\|x-y\|^2$$

外层优化 $\Leftrightarrow \min\limits_{y}\left\{f(y) + \frac{1}{2(\mu+1)}\|x-y\|^2\right\}$

$\Leftrightarrow y = \mathrm{prox}_{(\mu+1)f}(x)$

故 $\mathrm{prox}_{M_f^{\mu}}(x) = \frac{1}{\mu+1}\left(\mu x + \mathrm{prox}_{(\mu+1)f}(x)\right)$ $\square$

注：claim：$\forall X, Y$ 非空，$f: X\times Y \to \mathbb{R}$，有

$$\inf_{(x,y)\in X\times Y} f(x,y) = \inf_{x\in X}\left(\inf_{y\in Y} f(x,y)\right)$$

证明：对 $\forall x\in X$，有 $LHS \leq \inf\limits_{y\in Y} f(x,y)$

故 $LHS \leq \inf\limits_{x}\inf\limits_{y} f(x,y) = RHS$

下证 $RHS \leq LHS$，当 $LHS \neq -\infty$ 时，对 $\forall \varepsilon > 0$，$\exists$

$(x_0, y_0) \in X\times Y$，s.t. $f(x_0,y_0) < LHS + \varepsilon$

$$\text{RHS} = \inf_x \inf_y f(x,y) \le \inf_y f(x_0, y_0)$$

$$\le f(x_0, y_0) < \text{LHS} + \varepsilon$$

由 ε 的任意性, $\text{RHS} \le \text{LHS}$

若 $\text{LHS} = -\infty$ 时, 对 $\forall\, l \in \mathbb{R}$, $\exists\, (x_0, y_0) \in X \times Y$

s.t. $f(x_0, y_0) < l$, 故 $\text{RHS} \le f(x_0, y_0) < l$, 故 $\text{RHS} = -\infty$

$\square$

<u>Corollary 6.64</u> f proper 闭凸, $\lambda, \mu > 0$, 则 $\forall x \in \mathbb{E}$

$$\operatorname{prox}_{\lambda M_f^{\mu}}(x) = x + \frac{\lambda}{\mu+\lambda}\left(\operatorname{prox}_{(\mu+\lambda)f}(x) - x\right)$$

证明:

$$\operatorname{prox}_{\lambda M_f^{\mu}}(x) = \operatorname{prox}_{M_{\lambda f}^{\mu/\lambda}}(x)$$

$$= x + \frac{\lambda}{\mu+\lambda}\left(\operatorname{prox}_{(\mu+\lambda)f}(x) - x\right)$$

$\square$

Example 6.65 $C\subseteq\mathbb{E}$ 非空闭凸, $\lambda>0$, $f=\frac{1}{2}d_C^2$

$$\operatorname{prox}_{\lambda f}(x)=\operatorname{prox}_{\lambda M_g^1}(x) \quad (\text{Example } 6.53)$$

$$=x+\frac{\lambda}{\lambda+1}\left(\operatorname{prox}_{(\lambda+1)g}(x)-x\right)$$

$$=x+\frac{\lambda}{\lambda+1}\left(P_C(x)-x\right) \qquad \square$$

Example 6.66 $f(x)=\lambda H_\mu(x)$

$$\operatorname{prox}_{\lambda H_\mu}(x)=\operatorname{prox}_{\lambda M_g^\mu}(x)$$

$$=x+\frac{\lambda}{\lambda+\mu}\left(\operatorname{prox}_{(\mu+\lambda)g}(x)-x\right)$$

$$=x+\frac{\lambda}{\lambda+\mu}\left[\left(1-\frac{\mu+\lambda}{\max\{\|x\|,\mu+\lambda\}}\right)x-x\right]$$

$$=\left(1-\frac{\lambda}{\max\{\|x\|,\mu+\lambda\}}\right)x \qquad \square$$

Thm 6.67 f proper 闭凸, $\mu>0$, 则对 $\forall x\in\mathbb{E}$

$$M_f^{\mu}(x) + M_{f^*}^{1/\mu}(x/\mu) = \frac{1}{2\mu}\|x\|^2$$

证明：对 $\forall x \in \mathbb{E}$

$$M_f^{\mu}(x) = \min_u \{f(u) + \psi(u)\}$$

其中 $\psi(u) \equiv \frac{1}{2\mu}\|u-x\|^2$，由 <u>Thm 4.15</u>

$$M_f^{\mu}(x) = \max_{v\in\mathbb{E}} \{-f^*(v) - \psi^*(-v)\}$$

$$= -\min_{v\in\mathbb{E}} \{f^*(v) + \psi^*(-v)\}$$

$$= -\min_{v\in\mathbb{E}} \{f^*(v) + \frac{\mu}{2}\|v\|^2 - \langle x, v\rangle\}$$

$$= -\min_{v\in\mathbb{E}} \{f^*(v) + \frac{\mu}{2}\|v - x/\mu\|^2 - \frac{1}{2\mu}\|x\|^2\}$$

$$= \frac{1}{2\mu}\|x\|^2 - M_{f^*}^{1/\mu}(x/\mu) \qquad \square$$

§ 6.8 Miscellaneous Prox Computations

§ 6.8.1 Norm of a Linear Transformation over $\mathbb{R}^n$

Lemma 6.68 $f(x)=\|Ax\|_2$, $A\in\mathbb{R}^{m\times n}$ full row rank.

$\lambda>0$, 则:

$$\mathrm{prox}_{\lambda f}(x)=\begin{cases} x-A^T(AA^T)^{-1}Ax, & \|(AA^T)^{-1}Ax\|_2\le\lambda \\ x-A^T(AA^T+\alpha^*I)^{-1}Ax, & \|(AA^T)^{-1}Ax\|_2>\lambda \end{cases}$$

α^* 是下 decreasing func 唯一正根:

$$g(\alpha)=\|(AA^T+\alpha I)^{-1}Ax\|_2^2-\lambda^2$$

证明: $\mathrm{prox}_{\lambda f}(x)$ 等价于下优化问题

$$\min_u\left\{\lambda\|Au\|_2+\frac{1}{2}\|u-x\|_2^2\right\}$$

又等价于:

$$\min_{u\in\mathbb{R}^n,z\in\mathbb{R}^m}\left\{\frac{1}{2}\|u-x\|_2^2+\lambda\|z\|_2 \ : z=Au\right\}$$

定义 Lagrangian:

$$\begin{aligned} L(u,z;y)&=\frac{1}{2}\|u-x\|_2^2+\lambda\|z\|_2+y^T(z-Au) \\ &=\left[\frac{1}{2}\|u-x\|_2^2-(A^Ty)^Tu\right]+\left[\lambda\|z\|_2+y^Tz\right] \end{aligned}$$

故 $\min\limits_{u,z} f(u,z;y)$

$$= \underbrace{\min_{u}\left[\frac{1}{2}\|u-x\|_2^2-(A^Ty)^Tu\right]}_{①}+\underbrace{\min_{z}\left[\lambda\|z\|_2+y^Tz\right]}_{②}$$

优化问题①的minimizer是 $\tilde{u}=x+A^Ty$

故 $\min\limits_{u}\left[\frac{1}{2}\|u-x\|_2^2-(A^Ty)^Tu\right]$

$$=\frac{1}{2}\|\tilde{u}-x\|_2^2-(A^Ty)^T\tilde{u}$$

$$=-\frac{1}{2}y^TAA^Ty-(Ax)^Ty$$

对优化问题②:

$$\min_{z}\left[\lambda\|z\|_2+y^Tz\right]=-\max_{z}\left[(-y)^Tz-\lambda\|z\|_2\right]$$

$$=-g^*(-y)$$

其中 $g(\cdot)=\lambda\|\cdot\|_2$，由 $g^*(w)=\lambda\delta_{B_{\|\cdot\|_2}[0,1]}(w/\lambda)$

$$=\delta_{B_{\|\cdot\|_2}[0,\lambda]}(w)$$

故 $\min_{z}[\lambda\|z\|_2+y^Tz]=\begin{cases}0, & \|y\|_2\le\lambda\\ -\infty, & \|y\|_2>\lambda\end{cases}$

故对偶问题：

$$\max_{y\in\mathbb{R}^m}\left\{-\frac{1}{2}y^TAA^Ty-(Ax)^Ty:\ \|y\|_2\le\lambda\right\} \quad (6.52)$$

由 Thm A.1，是 strong duality 的 primal-dual pair

$(6.52)\Leftrightarrow$

$$\max_{y\in\mathbb{R}^m}\left\{-\frac{1}{2}y^TAA^Ty-(Ax)^Ty:\ \|y\|_2^2\le\lambda^2\right\}$$

则 $\text{prox}_{\lambda f}(x)=x+A^Ty$，其中 y 是上问题最优解

y 是 optimal solution $\Leftrightarrow\exists\alpha^*$, s.t.

$$(AA^T+\alpha^*I)y+Ax=0$$

$$\alpha^*(\|y\|_2^2-\lambda^2)=0$$

$$\|y\|_2^2\le\lambda^2$$

$$\alpha^*\geqslant 0$$

① $\alpha^*=0$时，$y=-(AA^T)^{-1}Ax$

只需 $\|(AA^T)^{-1}Ax\|_2\le\lambda$ 即可，此时

$$\mathrm{prox}_{\lambda f}(x)=x-A^T(AA^T)^{-1}Ax$$

② 若 $\|(AA^T)^{-1}Ax\|_2>\lambda$，则 $\alpha^*>0$，故

$$\|y\|_2^2=\lambda^2 \qquad (6.60)$$

故 $y=-(AA^T+\alpha^*I)^{-1}Ax$

由(6.60)知：α^*是下function的正根

$$g(\alpha)=\|(AA^T+\alpha I)^{-1}Ax\|_2^2-\lambda^2$$

$$g'(\alpha)=-2x^TA^T(AA^T+\alpha I)^{-3}\underbrace{Ax}_{\neq 0}<0 \qquad \square$$

§ 6.8.2 平方 ℓ_1-norm

$$\varphi(s,t)=\begin{cases}\frac{s^2}{t}, & t>0\\ 0, & s=t=0\\ \infty, & \text{else}\end{cases}$$

Lemma 6.69

$$\min_{\lambda\in\Delta_n}\sum_{j=1}^{n}\varphi(x_j,\lambda_j)=\|x\|_1^2$$

且 optimal solution 是 $\tilde{\lambda}_j=\begin{cases}\frac{|x_j|}{\|x\|_1}, & x\neq 0\\ \frac{1}{n}, & x=0\end{cases}$, $j=1,\cdots,n$

证明：(6.62) 是闭凸 func 在紧集的 min，故由 Thm 2.12

$\exists$ optimal solution $\lambda^*\in\Delta_n$。定义：

$$I_0=\{i\in\{1,2,\cdots,n\}:\lambda_i^*=0\}$$

$$I_1=\{i\in\{1,2,\cdots,n\}:\lambda_i^*>0\}$$

则 $\sum_{j\in I_1}\lambda_j^*=\sum_{j=1}^{n}\lambda_j^*=1$；易证 $x_j=0$ 对 $\forall j\in I_0$，否则

$\varphi(x_j, \lambda_j^*) = \infty$，与 λ^* 的最优性矛盾. 由 C-S 不等式：

$$\sum_{j=1}^{n} |x_j| = \sum_{j\in I_1} |x_j| = \sum_{j\in I_1} \frac{|x_j|}{\sqrt{\lambda_j^*}} \sqrt{\lambda_j^*} \le \sqrt{\sum_{j\in I_1} \frac{x_j^2}{\lambda_j^*}} \cdot \sqrt{\sum_{j\in I_1} \lambda_j^*}$$

$$= \sqrt{\sum_{j\in I_1} \frac{x_j^2}{\lambda_j^*}}$$

$$\sum_{j=1}^{n} \varphi(x_j, \lambda_j^*) = \sum_{j\in I_1} \varphi(x_j, \lambda_j^*) = \sum_{j\in I_1} \frac{x_j^2}{\lambda_j^*} \ge \|x\|_1^2 \qquad ①$$

另一方面，$\tilde{\lambda} \in \Delta_n$，故

$$\sum_{j=1}^{n} \varphi(x_j, \lambda_j^*) \le \sum_{j=1}^{n} \varphi(x_j, \tilde{\lambda}_j) = \|x\|_1^2 \qquad ②$$

结合 ① ② 即证 $\square$

<u>Lemma 6.70</u>

$$\mathrm{prox}_{\rho f}(x) = \begin{cases} \left(\dfrac{\lambda_i x_i}{\lambda_i + 2\rho}\right)_{i=1}^{n}, & x \ne 0 \\ 0, & x = 0 \end{cases}$$

其中 $\lambda_i = \left[\dfrac{\sqrt{\rho}\,|x_i|}{\sqrt{\mu^*}} - 2\rho\right]_+$，$\mu^*$ 是以下 func 的任一正根：

$$\psi(\mu)=\sum_{i=1}^{n}\left[\frac{\sqrt{\rho}|x_i|}{\sqrt{\mu}}-2\rho\right]_+-1$$

证明：$x=0$时显然，下设$x\neq 0$

$u=\text{prox}_{\rho f}(x) \Leftrightarrow$

$$\min_{u\in\mathbb{R}^n,\lambda\in\Delta_n}\left\{\frac{1}{2}\|u-x\|_2^2+\rho\sum_{i=1}^{n}\varphi(u_i,\lambda_i)\right\}$$

先对u做min（二次func）$\Rightarrow u_i=\dfrac{\lambda_i x_i}{\lambda_i+2\rho}$；代入上式

$$\min_{\lambda}\ \sum_{i=1}^{n}\frac{\rho x_i^2}{\lambda_i+2\rho}$$

$$\text{s.t.}\quad e^T\lambda=1$$

$$\lambda\geq 0$$

以上问题 strong duality，Lagrangian:

$$L(\lambda;\mu)=\sum_{i=1}^{n}\left(\frac{\rho x_i^2}{\lambda_i+2\rho}+\lambda_i\mu\right)-\mu$$

$$\lambda^*\text{是optimal}\Leftrightarrow\begin{cases}\lambda^*\in\underset{\lambda\geq 0}{\text{argmin}}\ L(\lambda;\mu^*)\\ e^T\lambda^*=1\end{cases};\ \exists\mu^*$$

由 $x \neq 0 \Rightarrow \mu^* > 0$，且 $y_i^* = \left[\frac{\sqrt{\rho}\,|x_i|}{\sqrt{\mu}} - 2\rho\right]_+$

dual optimal μ^* 满足：

$$\sum_{i=1}^{n} \left[\frac{\sqrt{\rho}\,|x_i|}{\sqrt{\mu}} - 2\rho\right]_+ = 1 \qquad \square$$

§ 6.8.3 Projection onto the Set of s-Sparse Vectors

Lemma 6.71 $s \in \{1, 2, \cdots, n\}$，则

$$P_{C_s}(x) = \left\{ U_S x_S : |S| = s,\ S \subseteq \{1, 2, \cdots, n\},\ \sum_{i \in S} |x_i| = \sum_{i=1}^{s} |x_{\langle i \rangle}| \right\}$$

证明：$C_s = \bigcup_{|S|=s} A_S$，$A_S = \{x \in \mathbb{R}^n : x_{S^c} = 0\}$

故 $P_{C_s}(y) = \bigcup_{|S|=s} \{P_{A_S}(x)\}$

$P_{C_s}(y)$ 是 $P_{A_S}(y)$，$\forall |S| = s$ 中 $\|P_{A_S}(x) - x\|^2$ 最小的 vector

$P_{A_S}(x)$ 是以下优化问题最优解：

$$\min_{y\in\mathbb{R}^n}\left\{\|y-x\|_2^2 : y_{S^c}=0\right\}$$

$$\Leftrightarrow \quad \min_{y\in\mathbb{R}^n}\left\{\|y_S-x_S\|_2^2+\|x_{S^c}\|_2^2 : y_{S^c}=0\right\}$$

则 $y_S = x_S$；$y_{S^c}=0$，故 Lemma 成立 $\square$

§7.1 Symmetric Func

§7.1.1 Def and Examples

Def 7.1 设 $\mathcal{A} \subseteq \mathcal{O}^n$ 是正交阵的集合，proper func f

关于 $\mathcal{A}$ 对称，若 $f(Ax) = f(x) \quad \forall x \in \mathbb{R}^n, A \in \mathcal{A}$

Example 7.2. $\mathcal{A} = \{-I\}$，则偶 func 关于 $\mathcal{A}$ 对称

Example 7.3 $\mathcal{A} = \{D_i\}_{i=1}^n \subseteq \mathbb{R}^{n \times n}$，$D_i = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & -1 & & \\ & & & & 1 & \\ & & & & & \ddots \\ & & & & & & 1 \end{bmatrix}$

i-index

则 f 关于 $\mathcal{A}$ 对称 $\Leftrightarrow f(x) = f(|x|) \quad \forall x \in \mathbb{R}^n$

$(\Rightarrow)$ 对 $\forall x \in \mathbb{R}^n$，$I = \{i : x_i < 0\}$，则由 f 关于 $\mathcal{A}$ 对称

$$f(x) = f\left(\prod_{i \in I} D_i x\right) = f(|x|)$$

$(\Leftarrow)$ 对 $\forall x$，$\forall D_i \in \mathcal{A}$

$f(D_i x) = f(|D_i x|) = f(|x|) = f(x)$

称这样的func 绝对对称func.

<u>claim</u>: f 是绝对对称的 $\Leftrightarrow \exists g: \mathbb{R}^n_+ \to (-\infty, \infty]$, s.t. $f(x) = g(|x|)$ $\forall x \in \mathbb{R}^n$

<u>Example 7.4</u> f 关于 $A = O^n$ 对称 $\Leftrightarrow$

$\exists g: \mathbb{R} \to (-\infty, \infty]$, s.t. $f(x) = g(\|x\|_2)$

证明:

$(\Rightarrow)$ 取 $g(u) = f(e_1 u)$,则 $\exists$ 正交阵 V, s.t. $x = V e_1 \|x\|_2$

$g(\|x\|_2) = f(e_1 \|x\|_2) = f(x)$

$(\Leftarrow)$ $f(Ux) = g(\|Ux\|_2) = g(\|x\|_2) = f(x)$ $\square$

<u>Example 7.7</u> f 关于 Λ_n 对称 $\Leftrightarrow$ Def

$f(x) = f(Px)$, $\forall x \in \mathbb{R}^n$, $P \in \Lambda_n$ $\Leftrightarrow$ 下证

$f(x) = f(x^{\downarrow})$, $\forall x \in \mathbb{R}^n$

证明：

$(\Rightarrow)$ $\forall x$, $\exists \tilde{P} \in \Lambda_n$, s.t. $x^{\downarrow} = \tilde{P}x$

$$f(x) \overset{\text{Def}}{=} f(\tilde{P}x) = f(x^{\downarrow})$$

$(\Leftarrow)$ $\forall x$, $\forall P \in \Lambda_n$

$$f(Px) = f((Px)^{\downarrow}) = f(x^{\downarrow}) \overset{\text{已知}}{=} f(x) \qquad \boxed{\diagup}$$

<u>Example 7.8</u> f 关于 Λ_n^G 对称 $\Leftrightarrow$ $f(x) = f(|x|^{\downarrow})$

§7.1.2 The Symmetric Conjugate Theorem

<u>Thm 7.9</u> f 是 proper，关于 $\mathcal{A} \subseteq O^n$ 对称，则 f^* 关于 $\mathcal{A}$ 对称

证明：令 $A \in \mathcal{A}$，令 $h = f$，故 $h(x) = f(Ax)$

由 $f^*(y) = h^*(y)$，$\forall y \in \mathbb{R}^n$，则

$h^*(y) = f^*((A^T)^{-1}y) = f^*(Ay)$，从而 $f^*(y) = f^*(Ay)$ $\forall y$

由 A 的任意性得证 $\boxed{\diagup}$

§ 7.2 Symmetric Spectral Funcs over $\mathbb{S}^n$

约定对 $X \in \mathbb{S}^n$, $\lambda(X) \equiv (\lambda_1(X), \dots, \lambda_n(X))^T$

$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$

Def 7.11 g 是 $\mathbb{S}^n$ 上的 spectral func, 若 $\exists$ proper f . s.t.

$g = f \circ \lambda$, 称 f 是 associated func

Def 7.12 $f: \mathbb{S}^n \to (-\infty, \infty]$ 是对称 spectral func, 若

$\exists$ proper $f: \mathbb{R}^n \to (-\infty, \infty]$ permutation 对称; s.t. $g = f \circ \lambda$.

Thm 7.14 $\forall X, Y \in \mathbb{S}^n$, $\mathrm{Tr}(XY) \leq \langle \lambda(X), \lambda(Y) \rangle$

等式成立 $\Leftrightarrow \exists V \in \mathcal{O}^n$. s.t. $\begin{cases} X = V \mathrm{diag}(\lambda(X)) V^T \\ Y = V \mathrm{diag}(\lambda(Y)) V^T \end{cases}$

Thm 7.15 $f: \mathbb{E} \to (-\infty, \infty]$ 是 permutation symmetric func. 则

$$(f\circ\lambda)^* = f^*\circ\lambda.$$

证明：设 $Y \in S^n$，则

默认用这种矩阵内积

$$
\begin{aligned}
(f\circ\lambda)^*(Y) &= \max_{X\in S^n}\left\{\mathrm{Tr}(XY) - f(\lambda(X))\right\} \\
&\le \max_{X\in S^n}\left\{\langle\lambda(X),\lambda(Y)\rangle - f(\lambda(X))\right\} \quad \text{(Fan's 不等式)} \\
&\le \max_{x\in\mathbb{R}^n}\left\{\langle x,\lambda(Y)\rangle - f(x)\right\} \quad (\lambda(X)\text{有序关系}, x\text{没}) \\
&= (f^*\circ\lambda)(Y)
\end{aligned}
$$

另一边，对 Y 做谱分解：$Y = U\mathrm{diag}(\lambda(Y))U^T$，$U \in O^n$

$$
\begin{aligned}
(f^*\circ\lambda)(Y) &= \max_{x\in\mathbb{R}^n}\left\{\langle x,\lambda(Y)\rangle - f(x)\right\} \\
&= \max_{x\in\mathbb{R}^n}\left\{\mathrm{Tr}\left(\mathrm{diag}(x)\mathrm{diag}(\lambda(Y))\right) - f(x)\right\} \\
&= \max_{x\in\mathbb{R}^n}\left\{\mathrm{Tr}\left(\mathrm{diag}(x)U^TYU\right) - f(x^{\downarrow})\right\} \\
&= \max_{x\in\mathbb{R}^n}\left\{\mathrm{Tr}\left(U\mathrm{diag}(x)U^TY\right) - f(\lambda(U\mathrm{diag}(x)U^T))\right\}
\end{aligned}
$$

$$\leq \max_{Z \in \mathbb{S}^n} \{ \mathrm{Tr}(ZY) - f(\lambda(Z)) \}$$

$$= (f \circ \lambda)^*(Y) \qquad \boxed{/\!/}$$

Thm7.17 设 $F: \mathbb{S}^n \to (-\infty, \infty]$，$F = f \circ \lambda$，其中 $f: \mathbb{R}^n \to (-\infty, \infty]$ 是 permutation symmetric proper func，则 F 闭凸 $\Leftrightarrow$ f 闭凸

证明：由 Spectrual Conjugate formula：

$$F^* = (f \circ \lambda)^* = f^* \circ \lambda$$

由 Thm7.9，f^* 是 permutation symmetric func，由用 Thm7.15

$$F^{**} = (f^* \circ \lambda)^* = f^{**} \circ \lambda \qquad (7.2)$$

- 若 f 闭凸，由 Thm4.8，结合 f proper 知 $f^{**} = f$

  故 $F^{**} = f \circ \lambda = F$

  故由 F 是 F^* 的 conjugate，由 Thm 4.3 知 F 闭凸

- 若 F 闭凸，由 Thm4.8，$F^{**} = F$

  (F 显然 proper)

由(7.2) $f\circ\lambda = F = F^{**} = f^{**}\circ\lambda$

故对 $\forall x\in\mathbb{R}^n$

$$f(x^{\downarrow}) = f(\lambda(\mathrm{diag}(x))) = f^{**}(\lambda(\mathrm{diag}(x))) = f^{**}(x^{\downarrow})$$

由 f, f^{**} 均是 permutation symmetric func, 有 $f(x) = f^{**}(x)$

$\boxed{/}$

故 f 闭凸

§7.2.2 The Proximal Operator of Symmetric Spectrual Functions over $\mathbb{S}^n$

<u>Thm 7.18</u> $F: \mathbb{S}^n \to (-\infty, \infty]$ $F = f\circ\lambda$, f 是闭凸的 permutation Symmetric Proper closed convex func, 设 $X\in\mathbb{S}^n$, $X = U\mathrm{diag}(\lambda(X))U^T$, $U\in\mathbb{O}^n$, 则 $\mathrm{prox}_F(X) = U\mathrm{diag}(\mathrm{prox}_f(\lambda(X)))U^T$

证明: $\mathrm{prox}_F(X) = \underset{Z\in\mathbb{S}^n}{\mathrm{argmin}}\left\{F(Z) + \frac{1}{2}\|Z-F\|_F^2\right\}$ (7.3)

记 $D = \mathrm{diag}(\lambda(X))$, 则对 $\forall Z\in\mathbb{S}^n$.

$$F(Z) + \frac{1}{2}\|Z-X\|_F^2 = F(Z) + \frac{1}{2}\|Z-UDU^T\|_F^2$$

$$= F(U^T Z U) + \frac{1}{2}\|U^T Z U - D\|_F^2$$

$$F(Z) = f(\lambda(Z)) = f(\underbrace{\lambda(U^T Z U)}_{\text{相似变换不改变特征值.}}) = F(U^T Z U)$$

令 $W = U^T Z U$，知 (17.3) 的 optimal value $Z = U W^* U^T$，W^* 是

$$\min_{W \in S^n} \left\{ G(W) \equiv F(W) + \frac{1}{2}\|W - D\|_F^2 \right\}$$

的最优解，下证 W^* 是 diagonal：令 $i \in \{1, 2, \cdots, n\}$，取

$$V_i = \begin{array}{c} \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & -1 & & \\ & & & & 1 & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \\ \uparrow \\ i\text{列} \end{array}$$

，令 $\widetilde{W}_i = V_i W^* V_i^T$

由 $V_i \in O^n$，$F(V_i W^* V_i^T) = f(\lambda(V_i W^* V_i^T)) = f(\lambda(W^*)) = F(W^*)$

故
$$\begin{aligned} G(\widetilde{W}_i) &= F(\widetilde{W}_i) + \frac{1}{2}\|\widetilde{W}_i - D\|_F^2 \\ &= F(V_i W^* V_i^T) + \frac{1}{2}\|V_i W^* V_i^T - D\|_F^2 \\ &= F(W^*) + \frac{1}{2}\|W^* - V_i^T D V_i\|_F^2 \\ &= F(W^*) + \frac{1}{2}\|W^* - D\|_F^2 \\ &= G(W^*) \end{aligned}$$

故 $\widetilde{W}_i$ 也是 optimal solution，但由 Thm 6.3，(17.3) 的解唯一

故 $W^* = V_i W^* V_i^T \Rightarrow W^*_{ij} = 0\ (j \neq i)$

由选取的任意性，W^* 是对角阵，设 $W^* = \mathrm{diag}(w^*)$，w^* 是

$$\min_{w} \left\{ F(\mathrm{diag}(w)) + \frac{1}{2}\|\mathrm{diag}(w) - D\|_F^2 \right\}$$ 最优解

由 $F(\mathrm{diag}(w)) = f(w^\downarrow) = f(w)$，$\|\mathrm{diag}(w) - D\|_F^2 = \|w - \lambda(X)\|_2^2$

$$w^* = \operatorname*{argmin}_{w} \left\{ f(w) + \frac{1}{2}\|w - \lambda(X)\|_2^2 \right\} = \mathrm{prox}_f(\lambda(X))$$

故 $W^* = \mathrm{diag}(\mathrm{prox}_f(\lambda(X)))$，从而

$$\mathrm{prox}_F(X) = U \mathrm{diag}(\mathrm{prox}_f(\lambda(X)))\, U^T \qquad \boxed{}$$

§7.3 完全平行于 §7.2

§ 8.1 From GD to Subgradient Descent

§ 8.1.1 Descent Directions?

Example 8.3 $f(x_1, x_2) = |x_1| + 2|x_2|$

$$\partial f(1,0) = \{(1,x) : |x| \leq 2\}$$

故 $(1,2) \in \partial f(1,0)$，但 $-(1,2)$ 不是 descent dir：对 $\forall t > 0$

$$g(t) \equiv f((1,0) - t(1,2))$$

$$= |1-t| + 4t = \begin{cases} 1+3t, & t \in (0,1] \\ 5t-1, & t \geq 1 \end{cases}$$

故 $f'((1,0), -(1,2)) = g'_+(0) = 3 > 0$，即 $-(1,2)$ 不是 descent dir

且没有 $\{(1,0) - t(1,2) : t > 0\}$ 上的点，值小于 $(1,0)$

注 $g^* = \underset{g \in \partial f(x)}{\operatorname{argmin}} \|g\|$ 时，g^* 是 descent dir

§ 8.1.2 Wolfe's Example

取 $\gamma>1$，$f(x_1,x_2)=\begin{cases}\sqrt{x_1^2+\gamma x_2^2}, & |x_2|\le x_1\\ \dfrac{x_1+\gamma|x_2|}{\sqrt{1+\gamma}}, & \text{else}\end{cases}$

<u>Lemma 8.4</u> (P) $\max\{g(y): f_1(y)\le 0,\ f_2(y)\le 0\}$

g concave，f_1, f_2 convex，设 $\max\{g(y): f_1(y)\le 0\}$ 有唯一解 $\tilde{y}$，则设 Y^* 是(P)的最优解集，则有：

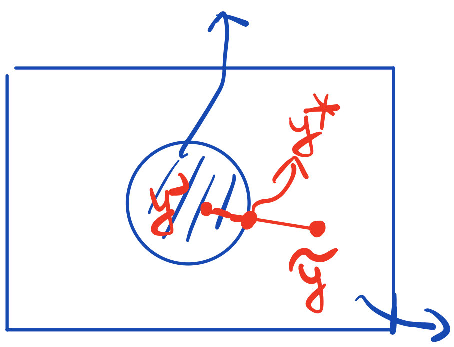
(i) $f_2(\tilde{y})\le 0$，此时有 $Y^*=\{\tilde{y}\}$

(ii) $f_2(\tilde{y})>0$，此时 $Y^*=\arg\max\{g(y): f_1(y)\le 0,\ f_2(y)=0\}$

证明 (i) 显然。

(ii)

(P)的可行集



$\max\{g(y): f_1(y)\le 0\}$ 可行集

$$g(y^*) \doteq g(\theta\tilde{y}+(1-\theta)y')$$

$$\geq \theta g(\tilde{y})+(1-\theta)g(y')$$

用了$\tilde{y}$的唯一性 $$> \theta g(y')+(1-\theta)g(y') = g(y')$$

故若 $y' \notin \partial F_{(P)}$；$F_{(P)}$是(P)的可行集，一定找到 $y^* \in F_{(P)}$

s.t. $g(y^*) > g(y')$，即证 $\boxed{\diagup}$

Lemma 8.5

(a) $f = \sigma_C$，其中

$$C = \left\{ (y_1, y_2) \in \mathbb{R}\times\mathbb{R} : y_1^2 + \frac{y_2^2}{\gamma} \leq 1,\ y_1 \geq \frac{1}{\sqrt{1+\gamma}} \right\}$$

(b) f 闭凸

(c)

$$\partial f(x_1, x_2) = \begin{cases} C & x_1 = x_2 = 0 \\ \dfrac{(x_1, \gamma x_2)}{\sqrt{x_1^2+\gamma x_2^2}} & |x_2| \leq x_1,\ x_1 \neq 0 \\ \left(\dfrac{1}{\sqrt{1+\gamma}}, \dfrac{\gamma\,\mathrm{sgn}(x_2)}{\sqrt{1+\gamma}}\right), & |x_2| > x_1,\ x_2 \neq 0 \\ \left\{\dfrac{1}{\sqrt{\gamma+1}}\right\} \times \left[-\dfrac{1}{\sqrt{1+\gamma}}, \dfrac{1}{\sqrt{1+\gamma}}\right], & x_2 = 0,\ x_1 < 0 \end{cases}$$

证明：

$$\sigma_C(x_1,x_2)=\max_{y_1,y_2}\left\{x_1y_1+x_2y_2:\ y_1^2+\frac{y_2^2}{\gamma}\le 1,\ y_1\ge\frac{1}{\sqrt{1+\gamma}}\right\}$$

- 当 $(x_1,x_2)=0$ 时，$\sigma_C(x_1,x_2)=0$，且

$$\operatorname*{Argmax}_{y_1,y_2}\left\{x_1y_1+x_2y_2:\ y_1^2+\frac{y_2^2}{\gamma}\le 1,\ y_1\ge\frac{1}{\sqrt{1+\gamma}}\right\}=C$$

- 设 $(x_1,x_2)\ne(0,0)$，记 $g(y_1,y_2)=x_1y_1+x_2y_2$

$$f_1(y_1,y_2)=y_1^2+\frac{y_2^2}{\gamma}-1,\quad f_2(y_1,y_2)=-y_1+\frac{1}{\sqrt{1+\gamma}}$$

则 (8.6) $\Longleftrightarrow$ $\max\limits_{y_1,y_2}\{g(y_1,y_2):f_1(y_1,y_2)\le 0,\ f_2(y_1,y_2)\le 0\}$

优化问题 $\max\limits_{y_1,y_2}\{g(y_1,y_2):f_1(y_1,y_2)\le 0\}$

的最优解∃!，$(\tilde{y}_1,\tilde{y}_2)=\dfrac{(x_1,\gamma x_2)}{\sqrt{x_1^2+\gamma x_2^2}}$ （用 Lagrangian）

<u>Case I</u> 若 $f_2(\tilde{y}_1,\tilde{y}_2)\le 0 \Longleftrightarrow \dfrac{x_1}{\sqrt{x_1^2+\gamma x_2^2}}\ge\dfrac{1}{\sqrt{1+\gamma}}$

$\Longleftrightarrow |x_2|\le x_1$，此时 $\sigma_C(x_1,x_2)=\sqrt{x_1^2+\gamma x_2^2}$

Case II $f_2(\tilde{y}_1,\tilde{y}_2)>0 \iff |x_2|>x_1$，由 lemma 8.4

(8.6)的最优解有 $y_1=\frac{1}{\sqrt{1+\gamma}}$，则 (8.6) 变成：

$$\max_{y_2}\left\{\frac{1}{\sqrt{1+\gamma}}x_1+x_2y_2 : y_2^2\le\frac{\gamma^2}{1+\gamma}\right\}$$

当 $x_2\neq 0$ 时，最优解是 $\left\{\frac{\gamma\,\mathrm{sgn}(x_2)}{\sqrt{1+\gamma}}\right\}$

当 $x_2=0$ 时，最优解是 $\left[-\frac{\gamma}{\sqrt{1+\gamma}},\frac{\gamma}{\sqrt{1+\gamma}}\right]$

Both options，$\sigma_C(x_1,x_2)=\frac{x_1+\gamma|x_2|}{\sqrt{1+\gamma}}$

即证(a)，故(b)自然得证，下证(c)：已经证明了：

$$\operatorname*{argmax}_{y_1,y_2}\left\{x_1y_1+x_2y_2 : (y_1,y_2)\in C\right\}$$

$$=\begin{cases} C, & x_1=x_2=0 \\ \dfrac{(x_1,x_2)}{\sqrt{x_1^2+\gamma x_2^2}}, & |x_2|\le x_1,\ x_1\neq 0 \\ \left(\dfrac{1}{\sqrt{1+\gamma}},\dfrac{\gamma\,\mathrm{sgn}(x_2)}{\sqrt{1+\gamma}}\right), & |x_2|>x_1,\ x_2\neq 0 \\ \left\{\dfrac{1}{\sqrt{1+\gamma}}\right\}\times\left[-\dfrac{\gamma}{\sqrt{1+\gamma}},\dfrac{\gamma}{\sqrt{1+\gamma}}\right], & x_2=0,\ x_1<0 \end{cases}$$

又由

$$\partial f(x_1,y_2)=\partial \sigma_C(x_1,x_2)$$

$$=\underset{y_1,y_2}{\operatorname{argmax}}\left\{x_1y_1+x_2y_2-\sigma_C^*(y_1,y_2)\right\}$$

$$=\underset{y_1,y_2}{\operatorname{argmax}}\left\{x_1y_1+x_2y_2-\delta_C(y_1,y_2)\right\}$$

$$=\underset{y_1,y_2}{\operatorname{argmax}}\left\{x_1y_1+x_2y_2,\ (y_1,y_2)\in C\right\} \qquad \square$$

Lemma 8.6 设 $\left\{(x_1^{(k)},x_2^{(k)})\right\}_{k\geq 0}$ 是 GD with exact line search 生成的序列，initial point $(x_1^0,x_2^0)=(\gamma,1)$，$\gamma>1$

则对 $\forall k\geq 0$

(a) f 在 $(x_1^{(k)},x_2^{(k)})$ 上可微

(b) $|x_2^{(k)}|\leq x_1^{(k)}$, $x_1^{(k)}\neq 0$

(c) $(x_1^{(k)},x_2^{(k)})=\left(\gamma\left(\frac{\gamma-1}{\gamma+1}\right)^k,\ \left(-\frac{\gamma-1}{\gamma+1}\right)^k\right)$

证明：只须证 (c)，用数归

- $k=0$时显然成立
- 设k时成立，i.e.，$(x_1^k, x_2^k) = \left(\gamma\left(\frac{\gamma-1}{\gamma+1}\right)^k, \left(-\frac{\gamma-1}{\gamma+1}\right)^k\right)$

证$k+1$时成立，i.e.，$(x_1^{k+1}, x_2^{k+1}) = (\beta_k, \gamma_k)$

$\beta_k = \gamma\left(\frac{\gamma-1}{\gamma+1}\right)^{k+1}$，$\gamma_k = \left(-\frac{\gamma-1}{\gamma+1}\right)^{k+1}$

由$|x_2^k| < x_1^k$，$x_1^k \neq 0$，知$f(x_1^k, x_2^k) = \sqrt{x_1^{k2} + \gamma x_2^{k2}}$

故f于(x_1^k, x_2^k)处可微。

$$\nabla f(x_1^k, x_2^k) = \frac{1}{\sqrt{x_1^{k2}+\gamma x_2^{k2}}}(x_1^k, \gamma x_2^k) \doteq \alpha_k(x_1^k, \gamma x_2^k)$$

令$g(t) \equiv f((x_1^k, x_2^k) - t(x_1^k, x_2^k))$，下证

(A) $(\beta_k, \gamma_k) = (x_1^k, x_2^k) - \frac{2}{\gamma+1}(x_1^k, \gamma x_2^k)$

(B) $g'\left(\frac{2}{\gamma+1}\right) = 0$

综合(A)(B)，由g的严格凸性即证，(A)(B)只须代入计算即可 $\square$

注：核心问题是 Fixed Stepsize，如 $y=|x|$, $t=1$

则 $|x_{k+1}-x_k|\equiv 1$，若 $x_0\notin\mathbb{Z}$，则永不收敛

§8.2 The Projected Subgradient Method

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

(A) f proper 闭凸

(B) C 非空闭凸

(C) $C\subseteq \mathrm{int}(\mathrm{dom} f)$

(D) (8.10) optimal set 非空，记为 X^*

注：$X^* = C\cap \mathrm{Lev}(f, f_{opt})$ 闭

取 $f'(x)\in\partial f(x)$，f' 是确定性的，记

$$f_{best}^k \equiv \min_{n=0,1,\cdots,k} f(x^n)$$

Lemma 8.11 $\forall x^*\in X^*$，$k\geq 0$

$$\|x^{k+1}-x^*\|^2 \le \|x^k-x^*\|^2 - 2t_k(f(x^k)-f_{opt}) + t_k^2\|f'(x^k)\|^2$$

证明：

$$\|x^{k+1}-x^*\|^2 = \|P_C(x^k-t_kf'(x^k)) - P_C(x^*)\|^2$$

$$\le \|x^k-t_kf'(x^k)-x^*\|^2$$

$$= \|x^k-x^*\|^2 - 2t_k\langle f'(x^k), x^k-x^*\rangle + t_k^2\|f'(x^k)\|^2$$

$$\le \|x^k-x^*\|^2 - 2t_k(f(x^k)-f_{opt}) + t_k^2\|f'(x^k)\|^2$$

$\boxed{/\!/}$

§ 8.2.2 Converge under Polayk's StepSize Rule.

Assumption 8.12 $\exists L_f>0$, s.t. $\|g\|<L_f$ 对 $\forall g\in\partial f(x)$, $x\in C$

注：由 $C\subseteq \text{int dom} f$, 由 Thm 3.61，$f$ 在 C 上 L_f-Lip 连续：

$$|f(x)-f(y)|\le L_f\|x-y\|, \forall x,y\in C$$

取 Polyak 步长：

$$t_k = \begin{cases} \dfrac{f(x^k) - f_{opt}}{\|f'(x^k)\|^2}, & f'(x^k) \neq 0 \\ 1, & f'(x^k) = 0 \end{cases}$$

Thm 8.13 设 Assumption 8.7, 8.12 成立，$\{x^k\}_{k\geq 0}$ 是 projection subgradient method with Polyak 步长，有

(a) $\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2, \forall k \geq 0, x^* \in X^*$;

(b) $f(x^k) \to f_{opt}$ as $k \to \infty$

(c) $f_{best}^k - f_{opt} \leq \dfrac{L_f d_{X^*}(x^0)}{\sqrt{k+1}} \quad \forall k \geq 0$

证明：由 Lemma 8.11，

$$\|x^{n+1} - x^*\|^2 \leq \|x^n - x^*\|^2 - 2t_n(f(x^n) - f_{opt}) + t_n^2 \|f'(x^n)\|^2$$

若 $f'(x^n) \neq 0$，则将 $t_n = \dfrac{f(x^n) - f_{opt}}{\|f'(x^n)\|^2}$ 代入上式

$$\|x^{n+1} - x^*\|^2 \leq \|x^n - x^*\|^2 - \frac{(f(x^n) - f_{opt})^2}{\|f'(x^n)\|^2}$$

由 $\|f'(x^n)\| \leq L_f$，知

$$\|x^{n+1}-x^*\|^2 \leq \|x^n-x^*\|^2 - \frac{(f(x^n)-f_{opt})^2}{L_f^2} \quad (8.15)$$

若 $f'(x^n)=0$ 时，显然 (a) 成立，故 (a) 得证

将 (8.15) 对 $n=0,1,\cdots,k$ 求和：

$$\frac{1}{L_f^2}\sum_{n=0}^{k}(f(x^n)-f_{opt})^2 \leq \|x^0-x^*\|^2 - \|x^{k+1}-x^*\|^2 \leq \|x^0-x^*\|^2$$

$$\Rightarrow \sum_{n=0}^{k}(f(x^n)-f_{opt})^2 \leq L_f^2\|x^0-x^*\|^2$$

($x^* \in X^*$ 的任意性)

$$\Rightarrow \sum_{n=0}^{k}(f(x^n)-f_{opt})^2 \leq L_f^2 d_{X^*}^2(x^0)$$

故 $f(x^n)-f_{opt} \to 0$，(b) 得证

下证 (c)：由 $f(x^n) \geq f_{best}^k$ 对 $\forall n=0,1,\cdots,k$

$$\sum_{n=0}^{k}(f(x^n)-f_{opt})^2 \geq (k+1)(f_{best}^k-f_{opt})^2$$

$$\Rightarrow (k+1)(f_{best}^k-f_{opt})^2 \leq L_f^2 d_{X^*}^2(x^0)$$

$$\Rightarrow \quad f_{best}^{k} - f_{opt} \leq \frac{L_f d_{X^*}(x^0)}{\sqrt{k+1}} \qquad \square$$

<u>Def</u> $\{x^k\}_{k\geq 0} \subseteq \mathbb{E}$ 是 Fejér monotone w.r.t. $S \subseteq \mathbb{E}$，if

$$\|x^{k+1} - y\| \leq \|x^k - y\|，对 \forall k\geq 0，y \in S$$

<u>Thm 8.16</u> (Convergence under Fejér 单调性).

$\{x^k\}_{k\geq 0} \subseteq \mathbb{E}$，$D \subseteq S$ 且 S 包含 $\{x^k\}_{k\geq 0}$ 的所有聚点，若

$\{x^k\}_{k\geq 0}$ 是 Fejér monotone w.r.t. S，则 $\{x^k\}$ 收敛

证明：由 $\{x^k\}_{k\geq 0}$ 是 Fejér monotone，故有界且有 limit points

设 $\tilde{x}$ 是 $\{x^k\}_{k\geq 0}$ 任一聚点，i.e.，$\exists \{x^{k_j}\}_{j\geq 0}$，$x^{k_j} \to \tilde{x}$

由 $\tilde{x} \in D \subseteq S$，故由 Fejér 单调 w.r.t. S：对 $\forall k\geq 0$

$$\|x^{k+1} - \tilde{x}\| \leq \|x^k - \tilde{x}\|$$

故 $\{\|x^k - \tilde{x}\|\}_{k\geq 0}$ 是 nonincreasing 序列，且 lower bounded by 0

故该列收敛；由 $x^{k_j}\to\bar{x}$，故 $x^k\to\bar{x}$ $\boxtimes$

Thm 8.17 设 Assumption 8.7, 8.12 成立，则 $\{x^k\}_{k\geq 0}$ 收敛到 X^* 是一个 point

证明：由 Thm 8.13 (a)，$\{x^k\}_{k\geq 0}$ 是 Fejér monotone w.r.t. X^*.

故由 Thm 8.16，只需证 $\{x^k\}$ 任一聚点在 X^* 中。设 $\bar{x}$ 是 $\{x^k\}$ 的任一聚点，则 $\exists\{x^{k_j}\}_{j\geq 0}\to\bar{x}$，由 C 的闭性，$\bar{x}\in C$

由 Thm 8.13 (b) $\quad f(x^{k_j})\to f_{opt},\ j\to\infty$

由 $\bar{x}\in C\subseteq \mathrm{int}(\mathrm{dom} f)$，故 f 在 $\bar{x}$ 处连续. 故 $f(\bar{x})=f_{opt}$.

$\Rightarrow \bar{x}\in X^*$ $\boxtimes$

Thm 8.18 $\quad k\geq \dfrac{L_f^2 d_{X^*}^2(x^0)}{\varepsilon^2}-1$，则 $f_{best}^k-f_{opt}\leq\varepsilon$

证明：Thm 8.13 (c) 的直接推论

Example 8.19 $\quad \min\limits_{x_1,x_2}\{f(x_1,x_2)=|x_1+2x_2|+|3x_1+4x_2|\}$

易知 $(x_1,x_2)=(0,0)$ 时，$f_{opt}=0$，由 $f(x)=\|Ax\|_1$，

$A=\begin{pmatrix}1&2\\3&4\end{pmatrix}$，则对 $\forall x\in\mathbb{R}^2$：

$$\partial f(x)=A^T\partial h(Ax)$$

其中 $h(x)=\|x\|_1$，由例3.41，知对 $\forall w\in\mathbb{R}^2$

$$\partial h(w)=\left\{z\in\mathbb{R}^2:\ z_i=\operatorname{sgn}(w_i),\ i\in I_{\neq}(w),\ |z_j|\le 1,\ j\in I_0(w)\right\}$$

其中 $I_0(w)=\{i: w_i=0\}$，$I_{\neq}(w)=\{i: w_i\neq 0\}$

故若 $\eta\in\partial h(Ax)$，则 $\eta\in[-1,1]\times[-1,1]$，故 $\|\eta\|_2\le\sqrt{2}$

故 $\forall g\in\partial f(x)$，则 $g=A^T\eta$，$\eta\in\partial h(Ax)$，有

$$\|g\|_2=\|A^T\eta\|_2\le\|A^T\|_{2,2}\|\eta\|_2\le\|A^T\|_{2,2}\sqrt{2}=7.7287$$

故可以取 $L_f=7.7287$

$$\begin{pmatrix}x_1^{k+1}\\ x_2^{k+1}\end{pmatrix}=\begin{pmatrix}x_1^{k}\\ x_2^{k}\end{pmatrix}-\frac{|x_1^k+2x_2^k|+|3x_1^k+4x_2^k|}{\|v(x_1^k,x_2^k)\|_2^2}\,v(x_1^k,x_2^k)$$

其中：

$$v(x_1,x_2)=\begin{pmatrix} \operatorname{sgn}(x_1+2x_2)+3\operatorname{sgn}(3x_1+4x_2). \\ 2\operatorname{sgn}(x_1+2x_2)+4\operatorname{sgn}(3x_1+4x_2) \end{pmatrix}\in\partial f(x_1,x_2)$$

$\boxtimes$

§8.2.3 The Convex Feasibility Problem

设 $S_1,\cdots,S_m\subseteq\mathbb{E}$ 闭凸，$S\equiv\bigcap_{i=1}^{m}S_i\neq\phi$

想找 $x\in S\Leftrightarrow\min\limits_{x}\{f(x)\equiv\max\limits_{i=1,\cdots,m}d_{S_i}(x)\}$

设 $S\neq\phi$，则 $f_{opt}=0$，$X^*=S$，且 f 1-Lipschitz 连续

<u>Lemma 8.20</u> $S_1,\cdots,S_m$ 非空闭凸，则 f 1-Lip 连续

证明：设 $i\in\{1,2,\cdots,m\}$，$x,y\in\mathbb{E}$，则

$$\begin{aligned} d_{S_i}(x) &= \|x-P_{S_i}(x)\| \\ &\leq \|x-P_{S_i}(y)\| \\ &\leq \|x-y\|+\|y-P_{S_i}(y)\| \\ &= \|x-y\|+d_{S_i}(y) \end{aligned}$$

故 $d_{S_i}(x)-d_{S_i}(y)\leq\|x-y\|$

Replacing x, y，知：$d_{S_i}(y) - d_{S_i}(x) \le \|x-y\|$

$$\Rightarrow \quad |d_{S_i}(x) - d_{S_i}(y)| \le \|x-y\|$$

故 $\forall x, y \in \mathbb{E}$

$$|f(x)-f(y)| = \left|\max_{i=1,\cdots,m} d_{S_i}(x) - \max_{i=1,\cdots,m} d_{S_i}(y)\right|$$

$$= \left|\|V_x\|_\infty - \|V_y\|_\infty\right|$$

其中 $V_x = (d_{S_i}(x))_{i=1}^m$，$V_y = (d_{S_i}(y))_{i=1}^m$，故

$$|f(x)-f(y)| \le \left|\|V_x\|_\infty - \|V_y\|_\infty\right|$$

$$\le \|V_x - V_y\|_\infty \quad \text{(三角不等式)}$$

$$= \max_{i=1,\cdots,m} |d_{S_i}(x) - d_{S_i}(y)|$$

$$\le \|x-y\| \qquad \boxed{\angle}$$

用 Polyak 步长：取 $x^0 \in \mathbb{E}$，若 $x^k \in S$，取 $f'(x^k)=0$

故 $x^{k+1} = x^k$，否则由 Thm 3.50

(i) compute $i_k \in \operatorname{argmax}_{i=1,\cdots,m} d_{S_i}(x^k)$;

(ii) take any $g^k \in \partial d_{S_{i_k}}(x^k)$

由例13.49 $\quad g^k = \dfrac{x^k - P_{S_{i_k}}(x^k)}{d_{S_{i_k}}(x^k)}$

$$x^{k+1} = x^k - \frac{d_{S_{i_k}}(x^k) - f_{opt}}{\|g^k\|^2} \cdot \frac{x^k - P_{S_{i_k}}(x^k)}{d_{S_{i_k}}(x^k)}$$

$$= x^k - d_{S_{i_k}}(x^k) \frac{x^k - P_{S_{i_k}}(x^k)}{d_{S_{i_k}}(x^k)}$$

$$= P_{S_{i_k}}(x^k)$$

$\square$

§ 8.2.4 Projected Subgradient with Dynamic Stepsize

Lemma 8.24 设 Assumption 8.7 holds, positive stepsize $\{t_k\}_{k\geq 0}$

则对 $\forall x^* \in X^*$ 和 $\forall k$，有：

$$\sum_{n=0}^{k} t_n (f(x^n) - f_{opt}) \leq \frac{1}{2}\|x^0 - x^*\|^2 + \frac{1}{2}\sum_{n=0}^{k} t_n^2 \|f'(x^n)\|^2$$

证明：由 lemma 8.11，对 $\forall n \geq 0$，$x^* \in X^*$

$$\frac{1}{2}\|x^{n+1} - x^*\|^2 \leq \frac{1}{2}\|x^n - x^*\|^2 - t_n (f(x^n) - f_{opt}) + \frac{t_n^2}{2}\|f'(x^n)\|^2$$

对 $n = 0, 1, \cdots, k$ 求和：

$$\sum_{n=0}^{k} t_n (f(x^n) - f_{opt}) \leq \frac{1}{2}\|x^0 - x^*\|^2 - \frac{1}{2}\|x^{k+1} - x^*\|^2 + \sum_{n=0}^{k} \frac{t_n^2}{2}\|f'(x^n)\|^2$$

$$\leq \frac{1}{2}\|x^0 - x^*\|^2 + \frac{1}{2}\sum_{n=0}^{k} t_n^2 \|f'(x^n)\|^2$$

Thm 8.25 设假设 8.7, 8.12，若 $\sum_{n=0}^{k} t_n^2 \Big/ \sum_{n=0}^{k} t_n \to 0 \quad (k \to \infty)$

则 $f_{best}^k - f_{opt} \to 0 \quad (k \to \infty)$.

证明：令 $L_f > 0$, s.t. $\|g\| \le L_f$ 对 $\forall g \in \partial f(x)$, $x \in C$

由 $\|f'(x^n)\| \le L_f$, $f(x^n) \ge f_{best}^k$ $\forall n \le k$

$$\Rightarrow \left(\sum_{n=0}^{k} t_n\right)\left(f_{best}^k - f_{opt}\right) \le \frac{1}{2}\|x^0 - x^*\|^2 + \frac{L_f^2}{2}\sum_{n=0}^{k} t_n^2$$

$$\Rightarrow \quad f_{best}^k - f_{opt} \le \frac{1}{2}\frac{\|x^0 - x^*\|^2}{\sum_{n=0}^{k} t_n} + \frac{L_f^2}{2}\frac{\sum_{n=0}^{k} t_n^2}{\sum_{n=0}^{k} t_n}$$

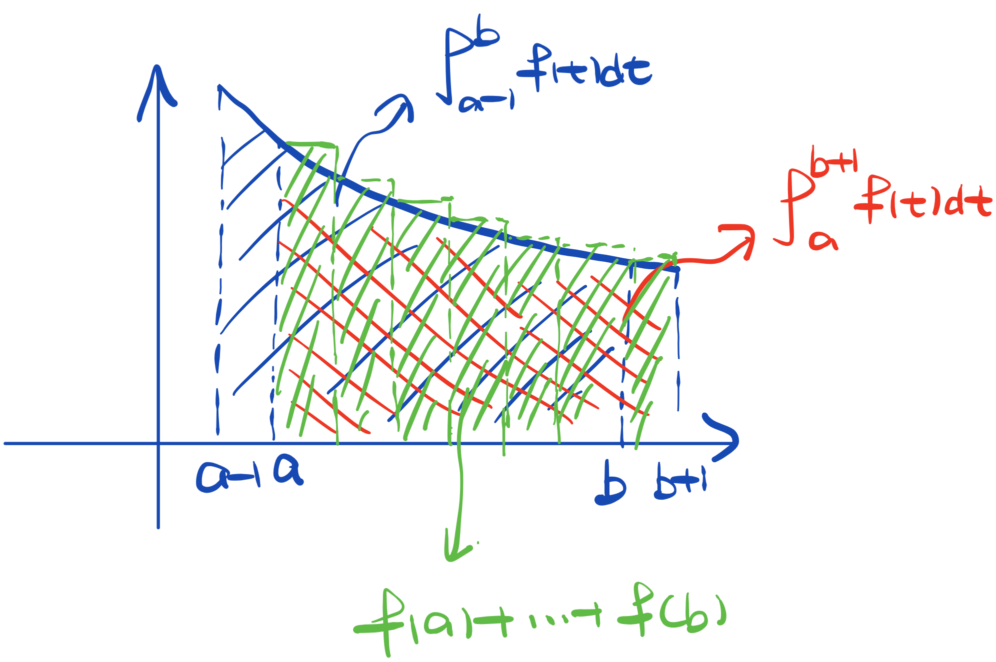
$$\longrightarrow 0$$

$\boxed{//}$

lemma 8.26 $f: [a-1, b+1] \to \mathbb{R}$ 连续 nonincreasing

$$\int_a^{b+1} f(t)dt \le f(a) + f(a+1) + \cdots + f(b) \le \int_{a-1}^{b} f(t)dt.$$

证明



$\boxed{//}$

Lemma 8.27 设 $D\in\mathbb{R}$

(a) $\forall k\geq 1$ $$\frac{D+\sum_{n=0}^{k}\frac{1}{n+1}}{\sum_{n=0}^{k}\frac{1}{\sqrt{n+1}}}\leq\frac{D+1+\log(k+1)}{\sqrt{k+1}};$$

(b) $\forall k\geq 2$ $$\frac{D+\sum_{n=\lceil k/2\rceil}^{k}\frac{1}{n+1}}{\sum_{n=\lceil k/2\rceil}^{k}\frac{1}{\sqrt{n+1}}}\leq\frac{4(D+\log 3)}{\sqrt{k+2}}$$

证明：(a) 由 lemma 8.26

$$\sum_{n=0}^{k}\frac{1}{n+1}=1+\sum_{n=1}^{k}\frac{1}{n+1}\leq 1+\int_0^k\frac{1}{x+1}dx=1+\log(k+1)$$

$$\sum_{n=0}^{k}\frac{1}{\sqrt{n+1}}\geq\int_0^{k+1}\frac{1}{\sqrt{x+1}}dx=2\sqrt{k+2}-2\geq\sqrt{k+1}$$

(b) $$\sum_{n=\lceil k/2\rceil}^{k}\frac{1}{n+1}\leq\int_{\lceil k/2\rceil-1}^{k}\frac{dt}{t+1}=\log(k+1)-\log(\lceil k/2\rceil)$$

$$=\log\left(\frac{k+1}{\lceil 0.5k\rceil}\right)\leq\log\left(\frac{k+1}{0.5k}\right)=\log\left(2+\frac{2}{k}\right)$$

$$\leq\log 3\quad(\forall k\geq 2)$$

$$\sum_{n=\lceil k/2\rceil}^{k}\frac{1}{\sqrt{n+1}} \geq \int_{\lceil k/2\rceil}^{k+1}\frac{dt}{\sqrt{t+1}} = 2\sqrt{k+2}-2\sqrt{\lceil k/2\rceil+1}$$

$$\geq 2\sqrt{k+2}-2\sqrt{k/2+2} = \frac{k}{\sqrt{k+2}+\sqrt{0.5k+2}}$$

$$\geq \frac{k}{2\sqrt{k+2}} \overset{\forall k\geq 2}{\geq} \frac{1}{4}\sqrt{k+2} \qquad \square$$

Thm 8.28 $t_k = \dfrac{1}{\|f'(x^k)\|\sqrt{k+1}}$ if $f'(x^k)\neq 0$, $t_k = \dfrac{1}{L_f}$ otherwise

(a) $$f^k_{best} - f_{opt} \leq \frac{L_f}{2}\,\frac{\|x^0-x^*\|^2+1+\log(k+1)}{\sqrt{k+1}}$$

(b) $$f(x^{(k)}) - f_{opt} \leq \frac{L_f}{2}\,\frac{\|x^0-x^*\|^2+1+\log(k+1)}{\sqrt{k+1}}$$

$$x^{(k)} = \frac{1}{\sum_{n=0}^{k} t_n}\sum_{n=0}^{k} t_n x^n$$

证明：由 (8.27)，且 $f(x^n)\geq f^k_{best}$ 对 $\forall n=0,\cdots,k$

$$f^k_{best}-f_{opt} \leq \frac{1}{2}\,\frac{\|x^0-x^*\|^2+\sum_{n=0}^{k}t_n^2\|f'(x^n)\|^2}{\sum_{n=0}^{k}t_n} \qquad (8.36)$$

由 Jensen's Inequality:

$$f(x^{(k)}) \le \frac{1}{\sum_{n=0}^{k} t_n} \sum_{n=0}^{k} t_n f(x^n)$$

由(8.27)知:

$$f(x^{(k)}) - f_{opt} \le \frac{1}{2} \frac{\|x^0 - x^*\|^2 + \sum_{n=0}^{k} t_n^2 \|f'(x^n)\|^2}{\sum_{n=0}^{k} t_n} \quad (8.37)$$

故结合(8.36)、(8.37)

$$\max\left\{ f_{best}^k - f_{opt},\ f(x^{(k)}) - f_{opt} \right\} \le \frac{1}{2} \frac{\|x^0 - x^*\|^2 + \sum_{n=0}^{k} t_n^2 \|f'(x^n)\|^2}{\sum_{n=0}^{k} t_n}$$

由t_n的定义: $t_n^2 \|f'(x^n)\|^2 \le \frac{1}{n+1}$, 且 $t_n \ge \frac{1}{L_f \sqrt{n+1}}$

$$\max\left\{ f_{best}^k - f_{opt},\ f(x^{(k)}) - f_{opt} \right\} \le \frac{L_f}{2} \frac{\|x^0 - x^*\|^2 + \sum_{n=0}^{k} \frac{1}{n+1}}{\sum_{n=0}^{k} \frac{1}{\sqrt{n+1}}}$$

由lemma 8.27(a), $D = \|x^0 - x^*\|^2$, 则:

$O(\frac{\log k}{\sqrt{k}})$.

$$\max\left\{ f_{best}^k - f_{opt},\ f(x^{(k)}) - f_{opt} \right\} \le \frac{L_f}{2} \frac{\|x^0 - x^*\|^2 + 1 + \log(k+1)}{\sqrt{k+1}} \quad \square$$

Thm 8.30 设假设8.7、8.12成立, 且C compact, 令

$$\Theta \ge \max_{x, y \in C} \frac{1}{2} \|x - y\|^2$$

取 $t_k = \frac{\sqrt{2\Theta}}{L_f \sqrt{k+1}}$ or

$$t_k = \begin{cases} \dfrac{\sqrt{2\Theta}}{\|f'(x^k)\|\sqrt{k+1}}, & f'(x^k) \neq 0 \\ \dfrac{\sqrt{2\Theta}}{L_f\sqrt{k+1}}, & f'(x^k) = 0 \end{cases}$$

则对 $\forall k \geqslant 2$ $\quad f_{best}^{k} - f_{opt} \leq \dfrac{\delta L_f \sqrt{2\Theta}}{\sqrt{k+1}}$, $\delta = 2(1+\log 3)$

证明：由 lemma 8.11，对 $\forall n \geqslant 0$

$$\frac{1}{2}\|x^{n+1} - x^*\|^2 \leq \frac{1}{2}\|x^n - x^*\|^2 - t_n(f(x^n) - f_{opt}) + \frac{t_n^2}{2}\|f'(x^n)\|^2$$

对 $n = \lceil k/2 \rceil, \lceil k/2 \rceil + 1, \cdots, k$，有

$$\sum_{n=\lceil \frac{k}{2} \rceil}^{k} t_n(f(x^n) - f_{opt}) \leq \frac{1}{2}\|x^{\lceil k/2 \rceil} - x^*\|^2 - \frac{1}{2}\|x^{k+1} - x^*\|^2 + \sum_{n=\lceil k/2 \rceil}^{k} \frac{t_n^2}{2}\|f'(x^n)\|^2$$

$$\leq \Theta + \sum_{n=\lceil k/2 \rceil}^{k} \frac{t_n^2}{2}\|f'(x^n)\|^2$$

$$\leq \Theta + \Theta \sum_{n=\lceil k/2 \rceil}^{k} \frac{1}{n+1}$$

由 $t_n \geqslant \dfrac{\sqrt{2\Theta}}{L_f\sqrt{n+1}}$, $f(x^n) \geqslant f_{best}^{k}$ $\forall n \leq k$

$$\sum_{n=\lceil k/2 \rceil}^{k} t_n(f(x^n) - f_{opt}) \geqslant \left(\sum_{n=\lceil k/2 \rceil}^{k} \frac{\sqrt{2\Theta}}{L_f\sqrt{n+1}} \right)(f_{best}^{k} - f_{opt})$$

$$f_{best}^{k} - f_{opt} \le \frac{L_f\sqrt{\Theta}}{\sqrt{2}} \frac{1+\sum\limits_{n=\lceil k/2\rceil}^{k}\frac{1}{n+1}}{\sum\limits_{n=\lceil k/2\rceil}^{k}\frac{1}{\sqrt{n+1}}} \le \frac{\delta L_f\sqrt{2\Theta}}{\sqrt{k+2}} \qquad \square$$

§8.2.5 The Strongly Convex Case

<u>Thm 8.31</u> 设假设8.7, 8.12成立, f σ-强凸, x^*是唯一minimizer

$$t_k = \frac{2}{\sigma(k+1)}$$

(a) 对 $\forall k \geq 0$, $f_{best}^{k} - f_{opt} \le \frac{2L_f^2}{\sigma(k+1)}$

$$\|x^{i_k} - x^*\| \le \frac{2L_f}{\sigma\sqrt{k+1}}, \quad i_k \in \operatorname*{argmin}_{j=0,\dots,k} f(x^j)$$

(b) $x^{(k)} = \sum\limits_{n=0}^{k} \alpha_n^k x^n$, 其中 $\alpha_n^k \equiv \frac{2n}{k(k+1)}$, 则 $\forall k \geq 0$

$$f(x^{(k)}) - f_{opt} \le \frac{2L_f^2}{\sigma(k+1)}$$

$$\|x^{(k)} - x^*\| \le \frac{2L_f}{\sigma\sqrt{k+1}}$$

证明: (a) $\forall n \geq 0$

$$\|x^{n+1} - x^*\|^2 = \|P_C(x^n - t_n f'(x^n)) - P_C(x^*)\|^2$$

$$\le \|x^n - t_n f'(x^n) - x^*\|^2$$

$$= \|x^n - x^*\|^2 - 2t_n\langle f'(x^n), x^n - x^*\rangle + t_n^2\|f'(x^n)\|^2$$

代入

由 f 是 σ-强凸，有

$$\langle f'(x^n), x^n - x^*\rangle \geq f(x^n) - f_{opt} + \frac{\sigma_n}{2}\|x^n - x^*\|^2$$

$$\Rightarrow \|x^{n+1} - x^*\|^2 \leq (1-\sigma t_n)\|x^n - x^*\|^2 - 2t_n(f(x^n) - f_{opt}) + t_n^2\|f'(x^n)\|^2$$

由 $\|f'(x^n)\| \leq L_f$，同除 $2t_n$ 知：

$$f(x^n) - f_{opt} \leq \frac{1}{2}(t_n^{-1} - \sigma)\|x^n - x^*\|^2 - \frac{1}{2}t_n^{-1}\|x^{n+1} - x^*\|^2 + \frac{t_n}{2}L_f^2$$

将 $t_n = \frac{2}{\sigma(n+1)}$ 代入上式

$$f(x^n) - f_{opt} \leq \frac{\sigma(n-1)}{4}\|x^n - x^*\|^2 - \frac{\sigma(n+1)}{4}\|x^{n+1} - x^*\|^2 + \frac{1}{\sigma(n+1)}L_f^2$$

$$\Rightarrow n(f(x^n) - f_{opt}) \leq \frac{\sigma(n-1)n}{4}\|x^n - x^*\|^2 - \frac{\sigma(n+1)n}{4}\|x^{n+1} - x^*\|^2 + \frac{n}{\sigma(n+1)}L_f^2$$

对 $n=0,\cdots,k$ 相加：

$$\sum_{n=0}^{k} n(f(x^n) - f_{opt}) \leq 0 - \frac{\sigma}{4}k(k+1)\|x^{k+1} - x^*\|^2 + \frac{L_f^2}{\sigma}\sum_{n=0}^{k}\frac{n}{n+1} \leq \frac{L_f^2 k}{\sigma}. \quad (8.49)$$

$$\Rightarrow \left(\sum_{n=0}^{k} n\right)(f_{best}^k - f_{opt}) \leq \frac{L_f^2 k}{\sigma}$$

$$\Rightarrow f_{best}^k - f_{opt} \leq \frac{2L_f^2}{\sigma(k+1)}$$

，故(8.44)得证

注意到 $f_{best}^k = f(x^{i_k})$，由Thm5.25(b)，对 σ-强凸func：

$f+\delta_C$ $\quad \frac{\sigma}{2}\|x^{i_k}-x^*\|^2 \le f_{best}^k - f_{opt} \le \frac{2L_f^2}{\sigma(k+1)}$

$$\Rightarrow \|x^{i_k}-x^*\| \le \frac{2L_f}{\sigma\sqrt{k+1}}$$

(b) 对(8.49)除以 $\frac{k(k+1)}{2}$

$$\sum_{n=0}^{k}\alpha_n^k(f(x^n)-f_{opt}) \le \frac{2L_f^2}{\sigma(k+1)}$$

由Jensen's Inequality（由 $(\alpha_n^k)_{n=0}^k \in \Delta_{k+1}$）

$$f(x^{(k)})-f_{opt} = f\left(\sum_{n=0}^{k}\alpha_n^k x^n\right)-f_{opt}$$

$$\le \sum_{n=0}^{k}\alpha_n^k(f(x^n)-f_{opt}) \le \frac{2L_f^2}{\sigma(k+1)}$$

$\square$

同理(8.47)成立

Thm 8.33：在Thm8.31假设下，对 $\forall k$ 满足 $k \ge \frac{2L_f^2}{\sigma\varepsilon}-1$

则 $f_{best}^k - f_{opt} \le \varepsilon$ 且 $f(x^{(k)})-f_{opt} \le \varepsilon$

§ 8.3 The Stochastic Projected Subgradient Method

§ 8.3.1 Settings and Method.

The Stochastic Projected Subgradient Method:

for $k=0,1,2,\cdots$

(A) $t_k>0$, random vector $g^k\in\mathbb{E}$

(B) $x^{k+1}=P_C(x^k-t_kg^k)$

<u>Assumption 8.34</u>

(A) $\forall k\geqslant 0$, $\mathbb{E}(g^k|x^k)\in\partial f(x^k)$

(B) $\exists\tilde{L}_f>0$, s.t. $\forall k\geqslant 0$, $\mathbb{E}(\|g^k\|^2|x^k)\leq\tilde{L}_f^2$

§ 8.3.2 Analysis

<u>Thm 8.35</u> 设假设8.7, 8.34成立，则

(a) $\sum\limits_{n=0}^{k}t_n^2\Big/\sum\limits_{n=0}^{k}t_n\to 0\ (k\to\infty)$，则 $\mathbb{E}(f_{best}^k)\to f_{opt}$

(b) 设C紧，$\Theta \geqslant \max\limits_{x,y\in C} \frac{1}{2}\|x-y\|^2$

若 $t_k = \dfrac{\sqrt{2\Theta}}{\tilde{L}_f\sqrt{k+1}}$，则对 $\forall k \geqslant 2$

$$\mathbb{E}(f_{best}^k) - f_{opt} \leqslant \frac{\delta \tilde{L}_f\sqrt{2\Theta}}{\sqrt{k+2}}, \quad \delta = 2(1+\log 3)$$

证明：对 $\forall n \geqslant 0$

$$\mathbb{E}(\|x^{n+1}-x^*\|^2 | x^n) = \mathbb{E}(\|P_C(x^n - t_n g^n) - P_C(x^*)\|^2 | x^n)$$

$$\leqslant \mathbb{E}(\|x^n - t_n g^n - x^*\|^2 | x^n)$$

$$= \|x^n - x^*\|^2 - 2t_n \mathbb{E}(\langle g^n, x^n - x^*\rangle | x^n) + t_n^2 \mathbb{E}(\|g^n\|^2 | x^n)$$

$$= \|x^n - x^*\|^2 - 2t_n \langle \mathbb{E}(g^n | x^n), x^n - x^*\rangle + t_n^2 \mathbb{E}(\|g^n\|^2 | x^n)$$

$$\leqslant \|x^n - x^*\|^2 - 2t_n(f(x^n) - f_{opt}) + t_n^2 \tilde{L}_f^2$$

两边对 x_n 取期望：

$$\mathbb{E}(\|x^{n+1}-x^*\|^2) \leqslant \mathbb{E}(\|x^n - x^*\|^2) - 2t_n(\mathbb{E}(f(x^n)) - f_{opt}) + t_n^2 \tilde{L}_f^2$$

对 $n = m, m+1, \cdots, k$ $(m \in \mathbb{Z}_+, m \leqslant k)$

$$\sum_{n=m}^{k} t_n (\mathbb{E}(f(x^n)) - f_{opt}) \leqslant \frac{1}{2}\left[\mathbb{E}(\|x^m - x^*\|^2) + \tilde{L}_f^2 \sum_{n=m}^{k} t_n^2\right]$$

$$\Rightarrow \left(\sum_{n=m}^{k} t_n\right)\left(\min_{n=m,\cdots,k} \mathbb{E}(f(x^n)) - f_{opt}\right) \leq \frac{1}{2}\left[\mathbb{E}(\|x^m - x^*\|^2) + \tilde{L}_f^2 \sum_{n=m}^{k} t_n^2\right]$$

由 $\mathbb{E}(f_{best}^k) \leq \mathbb{E}\left(\min_{n=m,\cdots,k} f(x^n)\right) \leq \min_{n=m,\cdots,k} \mathbb{E}(f(x^n))$

对 p 个 r.v. X_i, $i=1,\cdots,p$，有对 $\forall i \in \{1,\cdots,p\}$

$$\min\{X_1,\cdots,X_p\} \leq X_i$$

故 $\mathbb{E}(\min\{X_1,\cdots,X_p\}) \leq \min_{i=1,\cdots,p} \mathbb{E}(X_i)$

故：

$$\mathbb{E}(f_{best}^k) - f_{opt} \leq \frac{\mathbb{E}(\|x^m - x^*\|^2) + \tilde{L}_f^2 \sum_{n=m}^{k} t_n^2}{2\sum_{n=m}^{k} t_n^2} \quad (8.52)$$

取 $m=0$，有：

$$\mathbb{E}(f_{best}^k) - f_{opt} \leq \frac{\mathbb{E}(\|x^m - x^*\|^2) + \tilde{L}_f^2 \sum_{n=0}^{k} t_n^2}{2\sum_{n=0}^{k} t_n^2}$$

由假设(a)，知 $\mathbb{E}(f_{best}^k) \to f_{opt}$。(a)得证

(b) 取 $m=\lceil k/2 \rceil$，有：

$$\mathbb{E}(f_{best}^k) - f_{opt} \leq \frac{\Theta + \frac{\tilde{L}_f^2}{2}\sum_{n=\lceil k/2\rceil}^{k} t_n^2}{\sum_{n=\lceil k/2\rceil}^{k} t_n}$$

取 $t_n = \frac{\sqrt{2\Theta}}{\tilde{L}_f\sqrt{n+1}}$，有

$$\mathbb{E}(f_{best}^k) - f_{opt} \leq \frac{\tilde{L}_f\sqrt{2\Theta}}{2}\frac{1+\sum_{n=\lceil k/2\rceil}^{k}\frac{1}{n+1}}{\sum_{n=\lceil k/2\rceil}^{k}\frac{1}{\sqrt{n+1}}}$$

由 Lemma 8.27(b)，得证 $\square$

Example 8.36

$$(P)\quad \min\left\{f(x) \equiv \sum_{i=1}^{m} f_i(x) : x \in C\right\}$$

其中 $f_1, \cdots, f_m$ proper, closed, convex；设 Assumption 8.7 成立

C compact，L_f 是 f 在 C 上的 Lipschitz constant，令

$$\frac{1}{2}\max_{x,y\in C}\|x-y\|^2 \leq \Theta$$

设 $\forall i = 1, \cdots, m$，$\exists L_{f_i}$，s.t.

$$\|g\| \leq L_{f_i}，\text{对 } \forall g \in \partial f_i(x), x \in C$$

考虑 §8.2 的确定性 Projected Subgradient Method：

取 $f_i'(x^k) \in \partial f_i(x^k), i=1,\cdots,m$

$$x^{k+1} = P_C\left(x^k - \frac{\sqrt{2\Theta}}{\left\|\sum_{i=1}^m f_i'(x^k)\right\| \sqrt{k+1}} \left(\sum_{i=1}^m f_i'(x^k)\right)\right)$$

对 $k \geq 2$，由 Thm 8.30：

$$f_{best}^k - f_{opt} \leq \frac{\delta L_f \sqrt{2\Theta}}{\sqrt{k+2}}, \quad \delta = 2(1+\log 3)$$

则 ε-optimal 要 $N_1 = \max\left\{\frac{2\delta^2 L_f^2 \Theta}{\varepsilon^2} - 2, 2\right\}$

m large 时，$\sum_{i=1}^m f_i'(x^k)$ 算起来很耗时，下考虑用随机次梯度：

令 $g^k = m f_{i_k}'(x^k)$

$i_k \in \text{Uniform}\{1,2,\cdots,m\}$，则无偏性：

$$\mathbb{E}(g^k | x^k) = \sum_{i=1}^m \frac{1}{m} m f_i'(x^k) = \sum_{i=1}^m f_i'(x^k) \in \partial f(x^k)$$

有界性：

$$\mathbb{E}(\|g^k\|^2 | x^k) = \frac{1}{m}\sum_{i=1}^m m^2 \|f_i'(x^k)\|^2 \leq m \sum_{i=1}^m L_{f_i}^2 \equiv \tilde{L}_f^2$$

- Pick $i_k \sim \text{Uniform}\{1,\cdots,m\}$, $f_{i_k}'(x^k) \in \partial f_{i_k}(x^k)$

- $x^{k+1} = P_C\left(x^k - \frac{\sqrt{2\Theta}\, m}{\tilde{L}_f \sqrt{k+1}} f_{i_k}'(x^k)\right)$, $\tilde{L}_f = \sqrt{m}\sqrt{\sum_{i=1}^m L_{f_i}^2}$

由 Thm 8.35：

$$\mathbb{E}(f_{best}^k) - f_{opt} \le \frac{\delta\sqrt{m}\sqrt{\sum_{i=1}^m L_{f_i}^2}\sqrt{2\Theta}}{\sqrt{k+2}}$$

$$N_2 = \max\left\{\frac{2\delta^2 m\Theta\sum_{i=1}^m L_{f_i}^2}{\varepsilon^2} - 2, 2\right\}$$

Which is better?

$$\frac{N_2}{N_1} \approx \frac{m\sum_{i=1}^m L_{f_i}^2}{L_f^2} \equiv \beta$$

$f_i(x) = |a_i^T x + b_i|$, $i=1,\cdots,m$, $a_i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$, $C = B_{\|\cdot\|_2}[0,1]$

$$f(x) = \|Ax+b\|_1$$

$A = \begin{pmatrix} a_1^T \\ \vdots \\ a_m^T \end{pmatrix} \in \mathbb{R}^{m\times n}$, $b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \in \mathbb{R}^m$，由

$$\partial f_i(x) = \begin{cases} a_i, & a_i^T x + b_i > 0 \\ -a_i, & a_i^T x + b_i < 0 \\ \{\xi a_i : \xi \in [-1,1]\}, & a_i^T x + b_i = 0 \end{cases}$$

故 $L_{f_i} = \|a_i\|_2$，下估计 L_f，由例 3.44，$g \in \partial f(x)$ 有 $g = A^T\eta$

$\eta \in [-1,1]^m$，故 $\|\eta\|_2 \le \sqrt{m}$，故：

$$\|g\|_2 = \|A^T\eta\|_2 \le \|A^T\|_{2,2}\|\eta\|_2 \le \sqrt{m}\|A^T\|_{2,2} \doteq L_f$$

$$\beta=\frac{m\sum_{i=1}^{m}\|a_i\|_2^2}{m\|A^T\|_{2,2}^2}=\frac{\|A^T\|_F^2}{\|A^T\|_{2,2}^2}=\frac{\sum_{i=1}^{n}\lambda_i(AA^T)}{\max_{i=1,2,\dots,n}\lambda_i(AA^T)}$$

故 $\beta\in[1,m]$ $\boxed{/}$

§8.3.3 The strongly convex case.

<u>Thm 8.37</u> $t_k=\frac{2}{6(k+1)}$

(a) $\mathbb{E}(f_{best}^k)-f_{opt}\leq\frac{2\tilde{L}_f^2}{6(k+1)}$

(b) 令 $x^{(k)}=\sum_{n=0}^{k}\alpha_n^k x^n$, $\alpha_n^k\equiv\frac{2n}{k(k+1)}$, 则

$$\mathbb{E}(f(x^k))-f_{opt}\leq\frac{2\tilde{L}_f^2}{6(k+1)}$$

证明: (a) $\forall x^*\in X^*$, $n\geq0$

$$\mathbb{E}(\|x^{n+1}-x^*\|\,|\,x^n)=\|x^n-x^*\|^2-2t_n\langle\mathbb{E}(g^n|x^n),x^n-x^*\rangle$$
$$+t_n^2\,\mathbb{E}(\|g^n\|^2|x^n)$$

由 f 6-强凸且 $\mathbb{E}(g^n|x^n)\in\partial f(x^n)$, 故

$$f(x^*)\geq f(x^n)+\langle\mathbb{E}(g^n|x^n),x^*-x^n\rangle+\frac{6}{2}\|x^n-x^*\|^2$$

$$\Rightarrow \quad \langle \mathbb{E}(g^n|x^n), x^n - x^* \rangle \geq f(x^n) - f_{opt} + \frac{\sigma}{2}\|x^n - x^*\|^2$$

$$\mathbb{E}\left(\|x^{n+1} - x^*\|^2 \middle| x^n\right) \leq (1 - \sigma t_n)\|x^n - x^*\|^2 - 2t_n(f(x^n) - f_{opt})$$
$$+ t_n^2 \mathbb{E}(\|g^n\|^2 | x^n)$$

$$\Rightarrow f(x^n) - f_{opt} \leq \frac{1}{2}(t_n^{-1} - \sigma)\|x^n - x^*\|^2 - \frac{1}{2}t_n^{-1}\mathbb{E}\left(\|x^{n+1} - x^*\|^2 \middle| x^n\right)$$
$$+ \frac{1}{2}t_n \tilde{L}_f^2$$

取 $t_n = \frac{2}{\sigma(n+1)}$，有

$$f(x^n) - f_{opt} \leq \frac{\sigma(n-1)}{4}\|x^n - x^*\|^2 - \frac{\sigma(n+1)}{4}\mathbb{E}\left(\|x^{n+1} - x^*\|^2 \middle| x^n\right)$$
$$+ \frac{1}{\sigma(n+1)}\tilde{L}_f^2$$

对 x^n 取期望：

$$n\left(\mathbb{E}(f(x^n)) - f_{opt}\right) \leq \frac{\sigma n(n-1)}{4}\mathbb{E}\left(\|x^n - x^*\|^2\right) - \frac{\sigma(n+1)n}{4}\mathbb{E}\left(\|x^{n+1} - x^*\|^2\right)$$
$$+ \frac{n}{\sigma(n+1)}\tilde{L}_f^2$$

对 $n = 0, 1, \cdots, k$ 求和：

$$\sum_{n=0}^{k} n\left(\mathbb{E}(f(x^n)) - f_{opt}\right) \leq 0 - \frac{\sigma}{4}k(k+1)\mathbb{E}\left(\|x^{k+1} - x^*\|^2\right)$$
$$+ \frac{\tilde{L}_f^2}{\sigma}\sum_{n=0}^{k}\frac{n}{n+1} \leq \frac{\tilde{L}_f k}{\sigma} \quad (8.56)$$

$$\Rightarrow \left(\sum_{n=0}^{k} n\right)\left(\mathbb{E}(f_{best}^{k}) - f_{opt}\right) \leq \frac{\tilde{L}_f^2 k}{6}$$

$$\Rightarrow \qquad \mathbb{E}(f_{best}^{k}) - f_{opt} \leq \frac{2\tilde{L}_f^2}{6(k+1)}$$

(b) 用 (8.56) 除 $\frac{k(k+1)}{2}$

$$\sum_{n=0}^{k} \alpha_n^k \left(\mathbb{E}(f(x^n)) - f_{opt}\right) \leq \frac{2\tilde{L}_f^2}{6(k+1)}$$

由 Jensen 不等式：

$$\begin{aligned}
\mathbb{E}(f(x^{(k)})) - f_{opt} &= \mathbb{E}\left(f\left(\sum_{n=0}^{k} \alpha_n^k x^n\right)\right) - f_{opt} \\
&\leq \sum_{n=0}^{k} \alpha_n^k \left(\mathbb{E}(f(x^n)) - f_{opt}\right) \\
&\leq \frac{2\tilde{L}_f^2}{6(k+1)}
\end{aligned}$$

§8.4 The Incremental Projected Subgradient Method

$$\min\left\{ f(x) = \sum_{i=1}^{m} f_i(x) : x \in C \right\}$$

Assumption 8.38

(a) f_i is proper closed convex for $i=1,\cdots,m$

(b) $\exists L>0$ for which $\|g\| \leq L$ for $\forall g \in \partial f_i(x)$, $i=1,\cdots,m$, $x \in C$

Init: pick $x^0 \in C$

General Step: $\forall k=0,1,2,\cdots,$ execute:

(a) Set $x^{k,0}=x^k$, pick $t_k>0$

(b) $\forall i=0,1,\cdots,m-1$, compute

$$x^{k,i+1} = P_C(x^{k,i} - t_k g^{k,i})$$

$g^{k,i} \in \partial f_{i+1}(x^{k,i})$

(c) $x^{k+1} = x^{k,m}$

lemma 8.39 设假设 8.7，8.38 成立

$$\|x^{k+1}-x^*\|^2 \leq \|x^k-x^*\|^2 - 2t_k(f(x^k)-f_{opt}) + t_k^2 m^2 L^2$$

证明：$\forall x^* \in X^*$，$k \geq 0$，$i \in \{0,1,\cdots,m-1\}$

$$\|x^{k,i+1}-x^*\|^2 = \|P_c(x^{k,i}-t_k g^{k,i}) - P_c(x^*)\|^2$$

$$\leq \|x^{k,i}-t_k g^{k,i}-x^*\|^2$$

$$\leq \|x^{k,i}-x^*\|^2 - 2t_k\langle g^{k,i}, x^{k,i}-x^*\rangle + t_k^2 L^2$$

$$\leq \|x^{k,i}-x^*\|^2 - 2t_k(f_{i+1}(x^{k,i})-f_{i+1}(x^*)) + t_k^2 L$$

Summing inequality for $i=0,1,\cdots,m-1$，且 $x^{k,0}=x^k$，$x^{k,m}=x^{k+1}$

则对 $\forall x^* \in X^*$：

$$\|x^{k+1}-x^*\|^2 \leq \|x^k-x^*\|^2 - 2t_k\sum_{i=0}^{m-1}(f_{i+1}(x^{k,i})-f_{i+1}(x^*)) + t_k^2 mL$$

$$= \|x^k-x^*\|^2 - 2t_k\left(f(x^k)-f_{opt} + \sum_{i=0}^{m-1}(f_{i+1}(x^{k,i})-f_{i+1}(x^k))\right) + t_k^2 mL$$

Thm 3.61

$$\leq \|x^k-x^*\|^2 - 2t_k(f(x^k)-f_{opt}) + 2t_k\sum_{i=0}^{m-1} L\|x^{k,i}-x^k\| + t_k^2 mL$$

$$\|x^{k,1}-x^k\|=\|P_C(x^{k,0}-t_k g^{k,0})-P_C(x^k)\|\le t_k\|g^{k,0}\|\le t_k L$$

$$\|x^{k,2}-x^k\|\le \|x^{k,1}-x^k\|+t_k\|g^{k,1}\|\le 2t_k L$$

In general, for $\forall i=0,1,\cdots,m-1$

$$\|x^{k,i}-x^k\|\le t_k i L \quad \text{，故}$$

$$\|x^{k+1}-x^*\|^2\le\|x^k-x^*\|^2-2t_k(f(x^k)-f_{opt})+2t_k^2\sum_{i=0}^{m-1} iL^2+t_k^2 mL^2$$

$$=\|x^k-x^*\|^2-2t_k(f(x^k)-f_{opt})+t_k^2 m^2L^2 \qquad \square$$

Thm 8.40

(a) If $\left.\sum_{n=0}^{k}t_n^2\middle/\sum_{n=0}^{k}t_n\right.\to 0\ (k\to\infty)$, then $f_{best}^k\to f_{opt}$

(b) 设 C 紧，$\Theta\ge\max\limits_{x,y\in C}\frac{1}{2}\|x-y\|^2$，if $t_k=\dfrac{\sqrt{\Theta}}{Lm\sqrt{k+1}}$, then $\forall k\ge 2$

$$f_{best}^k-f_{opt}\le\frac{\delta mL\sqrt{\Theta}}{\sqrt{k+2}}, \quad \delta=2(2+\log 3)$$

Proof: By lemma 8.39, $\forall n\ge 0$

$$\|x^{n+1}-x^*\|^2\le\|x^n-x^*\|^2-2t_n(f(x^n)-f_{opt})+L^2m^2t_n^2$$

Summing over $n=p, p+1, \ldots, k$,

$$\|x^{k+1}-x^*\|^2 \le \|x^p - x^*\|^2 - 2\sum_{n=p}^{k} t_n\left(f(x^n)-f_{opt}\right) + L^2 m^2 \sum_{n=p}^{k} t_n^2$$

$$\Rightarrow \quad 2\sum_{n=p}^{k} t_n\left(f(x^n)-f_{opt}\right) \le \|x^p - x^*\|^2 + L^2 m^2 \sum_{n=p}^{k} t_n^2$$

$$\Rightarrow \quad f_{best}^{k} - f_{opt} \le \frac{\|x^p - x^*\|^2 + L^2 m^2 \sum_{n=p}^{k} t_n^2}{2\sum_{n=p}^{k} t_n}$$

令 $p=0$，由 $\sum_{n=0}^{k} t_n^2 \Big/ \sum_{n=0}^{k} t_n \to 0$，then $f_{best}^{k} \to f_{opt}$

(b) 令 $p = \lceil k/2 \rceil$，then

$$f_{best}^{k} - f_{opt} \le \frac{2\Theta + L^2 m^2 \sum_{n=\lceil k/2 \rceil}^{k} t_n^2}{2\sum_{n=\lceil k/2 \rceil}^{k} t_n}$$

取 $t_n = \dfrac{\sqrt{\Theta}}{Lm\sqrt{n+1}}$

$$f_{best}^{k} - f_{opt} \le \frac{Lm\sqrt{\Theta}}{2} \cdot \frac{2 + \sum_{n=\lceil k/2 \rceil}^{k} \frac{1}{n+1}}{\sum_{n=\lceil k/2 \rceil}^{k} \frac{1}{\sqrt{n+1}}}$$，由 lemma 8.27 (b)

$\square$

§ 8.5 The Dual Projected Subgradient Method

§ 8.5.1 The Dual Problem

$$
\begin{aligned}
f_{opt} = \min\ & f(x) \\
\text{s.t. } & g(x) \le 0 \qquad\qquad (8.62) \\
& x \in X
\end{aligned}
$$

<u>Assumption 8.41</u>

(A) $X \subseteq \mathbb{E}$ is convex

(B) $f : \mathbb{E} \to \mathbb{R}$ convex

(C) $g(\cdot) = (g_1(\cdot), \cdots, g_m(\cdot))^T$, where g_i is convex for $i = 1, \cdots, m$

(D) (8.62) has a finite optimal value denoted f_{opt}, $X^* \neq \emptyset$

(E) $\exists \bar{x} \in X$, s.t. $g(\bar{x}) < 0$

(F) $\forall \lambda \in \mathbb{R}^m_+$, $\min\limits_{x \in X} \{ f(x) + \lambda^T g(x) \}$ has an optimal solution

Lagrangian dual problem:

$$q(\lambda) = \min_{x \in X} \{ L(x; \lambda) = f(x) + \lambda^T g(x) \} \qquad (8.63)$$

由 (F), (8.63) possesses a solution

$$q_{opt} = \max \{ q(\lambda) : \lambda \in \mathbb{R}^m_+ \}$$

由(E) (slater condition) $f_{opt} = q_{opt}$

Thm 8.42 : 假设8.41 holds, $\bar{x} \in X$ s.t. $g(\bar{x}) < 0$, then

$\forall \lambda \in S_\mu \equiv \{\lambda \in \mathbb{R}_+^m : q(\lambda) \geq \mu\}$,

$$\|\lambda\|_2 \leq \frac{f(\bar{x}) - \mu}{\min_{j=1,\cdots,m} \{-g_j(\bar{x})\}} .$$

证明: 由 $\lambda \in S_\mu$, 有

$$\mu \leq q(\lambda) \leq f(\bar{x}) + \lambda^T g(\bar{x}) = f(\bar{x}) + \sum_{j=1}^m \lambda_j g_j(\bar{x})$$

$$\Rightarrow \quad -\sum_{j=1}^m \lambda_j g_j(\bar{x}) \leq f(\bar{x}) - \mu$$

由 $\lambda_j \geq 0$, $g_j(\bar{x}) < 0 \ \forall j \Rightarrow$

$$\sum_{j=1}^m \lambda_j \leq \frac{f(\bar{x}) - \mu}{\min_{j=1,\cdots,m} \{-g_j(\bar{x})\}}$$

由 $\lambda \geq 0 \Rightarrow \|\lambda\|_2 \leq \sum_{j=1}^m \lambda_j$, 即证 $\square$

§ 8.5.2 The Dual Projected Subgradient Method.

By Example 3.7, For given $\lambda \in \mathbb{R}_+^m$, the minimum of problem $q(\lambda)$ is attained at x_λ, then $-g(x_\lambda) \in \partial(-q)(\lambda)$

The Dual Projected Subgradient Method.

Init: pick $\lambda^0 \in \mathbb{R}_+^m$ arbitrarily

General step: $\forall k = 0, 1, 2, \cdots$ execute:

(a) pick $\gamma_k > 0$;

(b) compute $x^k \in \underset{x \in X}{\operatorname{argmin}} \{ f(x) + (\lambda^k)^T g(x) \}$

(c) if $g(x^k) = 0$, then terminate with an output x^k; otherwise

$$\lambda^{k+1} = \left[\lambda^k + \gamma_k \frac{g(x^k)}{\|g(x^k)\|_2} \right]_+$$

Lemma 8.44 假设 8.41 holds. Let $\bar{\lambda} \in \mathbb{R}_+^m$, let $\bar{x} \in X$ s.t.

$$\bar{x} \in \underset{x \in X}{\operatorname{argmin}} \{ f(x) + \bar{\lambda}^T g(x) \}$$

且 $g(\bar{x}) = 0$. Then $\bar{x}$ is an optimal solution of (8.62)

Proof: let x be a feasible point of (8.62), i.e., $x \in X$ and $g(x) \le 0$

then:

$$
\begin{aligned}
f(x) &\geq f(x)+\bar{\lambda}^T g(x) && [g(x)\leq 0,\ \bar{\lambda}\geq 0] \\
&\geq f(\bar{x})+\bar{\lambda}^T g(\bar{x}) && [(8.65)] \\
&= f(\bar{x}) && [g(\bar{x})=0]
\end{aligned}
$$

$\boxed{/\!/}$

§ 8.5.3 Convergence Analysis

Primal Sequence $\{x^k\}$ converge?

- $$x^{(k)}=\sum_{n=0}^{k}\mu_n^k x^n,\quad \mu_n^k=\frac{\gamma_n/\|g(x^n)\|_2}{\sum_{j=0}^{k}\frac{\gamma_j}{\|g(x^j)\|_2}},\quad n=0,\cdots,k$$

- $$x^{\langle k\rangle}=\sum_{n=\lceil k/2\rceil}^{k}\eta_n^k x^n,\quad \eta_n^k=\frac{\gamma_n/\|g(x^n)\|_2}{\sum_{j=\lceil k/2\rceil}^{k}\frac{\gamma_j}{\|g(x^j)\|_2}},\quad n=\lceil k/2\rceil,\cdots,k$$

<u>lemma 8.45</u> 设假设 8.41 holds，且 $\exists L>0$, s.t. $\|g(x)\|_2\leq L\ \forall x\in X$.

let $\rho>0$，则对 $\forall k\geq 2$，

$$f(x^{(k)})-f_{opt}+\rho\|[g(x^{(k)})]_+\|_2\leq\frac{1}{2}\frac{(\|\lambda^0\|_2+\rho)^2+\sum_{n=0}^{k}\gamma_n^2}{\sum_{n=0}^{k}\gamma_n}$$

$$f(x^{(k)}) - f_{opt} + \rho \left\| [g(x^{(k)})]_+ \right\|_2 \le \frac{1}{2} \frac{(\|\lambda^{\lceil k/2 \rceil}\|_2 + \rho)^2 + \sum_{n=\lceil k/2 \rceil}^{k} \gamma_n^2}{\sum_{n=\lceil k/2 \rceil}^{k} \gamma_n}$$

证明: Let $\bar{\lambda} \in \mathbb{R}_+^m$, then for $\forall n \ge 0$

$$\|\lambda^{n+1} - \bar{\lambda}\|_2^2 = \left\| [\lambda^n + \gamma_n \frac{g(x_n)}{\|g(x_n)\|_2}]_+ - [\bar{\lambda}]_+ \right\|_2^2$$

$$\le \left\| \lambda^n + \gamma_n \frac{g(x_n)}{\|g(x_n)\|_2} - \bar{\lambda} \right\|_2^2$$

$$= \|\lambda^n - \bar{\lambda}\|_2^2 + \gamma_n^2 + \frac{2\gamma_n}{\|g(x_n)\|_2} g(x_n)^T (\lambda^n - \bar{\lambda})$$

Let $p \in \{0, 1, \cdots, k\}$, Summing for $n = p, \cdots, k$

$$\|\lambda^{k+1} - \bar{\lambda}\|_2^2 \le \|\lambda^p - \bar{\lambda}\|_2^2 + \sum_{n=p}^{k} \gamma_n^2 + 2\sum_{n=p}^{k} \frac{\gamma_n}{\|g(x^n)\|_2} g(x^n)^T (\lambda^n - \bar{\lambda})$$

$$\Rightarrow \quad 2\sum_{n=p}^{k} \frac{\gamma_n}{\|g(x^n)\|_2} g(x^n)^T (\bar{\lambda} - \lambda^n) \le \|\lambda^p - \bar{\lambda}\|_2^2 + \sum_{n=p}^{k} \gamma_n^2$$

Def: $\forall p \in \{0, 1, \cdots, k\}$

$$x^{k,p} \equiv \sum_{n=p}^{k} \alpha_n^{k,p} x^n, \quad \alpha_n^{k,p} = \frac{\frac{\gamma_n}{\|g(x^j)\|_2}}{\sum_{j=p}^{k} \frac{\gamma_j}{\|g(x^j)\|_2}}$$

$\{x^{k,0}\}_{k \ge 0}$ 与 $\{x^{k,\lceil k/2 \rceil}\}_{k \ge 0}$ 与 $\{x^{(k)}\}_{k \ge 0}$ 与 $\{x^{(k)}\}_{k \ge 0}$ 是相同列

由 $\|g(x^n)\|_2 \le L$，由 (8.72)

$$\sum_{n=p}^{k} \alpha_n^{k,p} g(x^n)^T(\bar{\lambda}-\lambda^n) \le \frac{L}{2} \frac{\|\lambda^p-\bar{\lambda}\|_2^2+\sum_{n=p}^{k}\gamma_n^2}{\sum_{n=p}^{k}\gamma_n}$$

对 $\forall x^* \in X^*$，有

$$f(x^*) \ge f(x^*) + (\lambda^n)^T g(x^*) \quad [\lambda^n \ge 0, g(x^*) \le 0]$$

$$\ge f(x^n) + (\lambda^n)^T g(x^n) \quad [x^n \in \underset{x\in X}{\operatorname{argmin}} f(x) + \lambda^{n T} g(x)]$$

$$\Rightarrow \quad -(\lambda^n)^T g(x^n) \ge f(x^n) - f_{opt}$$

$$\Rightarrow \sum_{n=p}^{k} \alpha_n^{k,p} g(x^n)^T(\bar{\lambda}-\lambda^n) \ge \sum_{n=p}^{k} \alpha_n^{k,p} g(x^n)^T\bar{\lambda} + \sum_{n=p}^{k} \alpha_n^{k,p} f(x^n) - \sum_{n=p}^{k} \alpha_n^{k,p} f_{opt}$$

$$\ge \bar{\lambda}^T g(x^{k,p}) + f(x^{k,p}) - f_{opt} \quad \text{(Jensen)}$$

$$\Rightarrow \bar{\lambda}^T g(x^{k,p}) + f(x^{k,p}) - f_{opt} \le \frac{L}{2} \frac{(\|\lambda^p\|_2+\|\bar{\lambda}\|_2)^2+\sum_{n=p}^{k}\gamma_n^2}{\sum_{n=p}^{k}\gamma_n}$$

plugging

$$\bar{\lambda} = \begin{cases} \rho \dfrac{[g(x^{k,p})]_+}{\|[g(x^{k,p})]_+\|_2}, & [g(x^{k,p})]_+ \ne 0 \\ 0, & \text{otherwise.} \end{cases}$$

into (8.76)

$$f(x^{k,p}) - f_{opt} + \rho \| [g(x^{k,p})]_+ \|_2 \le \frac{L}{2} \frac{(\|x^p\|_2 + \rho \quad)^2 + \sum_{n=p}^{k} \gamma_n^2}{\sum_{n=p}^{k} \gamma_n}$$

令 $p=0$，$p=\lceil k/2 \rceil$ 即证 $\square$

§9.1 From Projected Subgradient to Mirror Descent

$$(P) \quad \min\{f(x), x \in C\}$$

假设9.1:

(A) $f: \mathbb{E} \to (-\infty, \infty]$ proper closed convex

(B) $C \subseteq \mathbb{E}$ 非空闭凸

(C) $C \subseteq \operatorname{int}(\operatorname{dom} f)$

(D) X^* 非空

第八章的迭代格式:

$$x^{k+1} = \underset{x \in C}{\operatorname{argmin}} \left\{ f(x^k) + \langle f'(x^k), x - x^k \rangle + \underbrace{\frac{1}{2t_k} \|x - x^k\|^2}_{\text{用Bregman度量替换}} \right\}$$

Def 9.2 $\omega: \mathbb{E} \to (-\infty, \infty]$ proper闭凸,且在 $\operatorname{dom}\partial\omega$ 上可微

Bregman distance $B_\omega: \operatorname{dom}\partial\omega \times \operatorname{dom}\partial\omega \to \mathbb{R}$

$$B_\omega(x, y) = \omega(x) - \omega(y) - \langle \nabla\omega(y), x - y \rangle$$

假设9.3

- ω proper 闭凸
- ω 在 $\mathrm{dom}(\partial\omega)$ 上可微 → ω 在 $\mathrm{int}(\mathrm{dom} f)$ 上可微。
  由Thm3.18，$\mathrm{ri}(\mathrm{dom} f) \subseteq \mathrm{dom}\,\partial\omega$
- $C \subseteq \mathrm{dom}\,\omega$
- $\omega + \delta_C$ σ-强凸

lemma 9.4 C 非空闭凸，ω 满足假设9.3

- $B_\omega(x,y) \geq \frac{\sigma}{2}\|x-y\|^2 \quad \forall x \in C,\ y \in C \cap \mathrm{dom}\,\partial\omega$
- 令 $x \in C$ 且 $y \in C \cap \mathrm{dom}\,\partial\omega$，有
  - $B_\omega(x,y) \geq 0$
  - $B_\omega(x,y) = 0$ iff $x = y$.

证明：由 $\omega + \delta_C$ 上强凸，lemma 9.4 显然 ☑

$$x^{k+1} = \underset{x \in C}{\operatorname{argmin}} \left\{ f(x^k) + \langle f'(x^k), x - x^k \rangle + \frac{1}{t_k} B_\omega(x, x^k) \right\}$$

$$\Leftrightarrow \quad x^{k+1} = \underset{x \in C}{\operatorname{argmin}} \left\{ \langle f'(x^k), x \rangle + \frac{1}{t_k} B_\omega(x, x^k) \right\}$$

$$\Leftrightarrow \quad x^{k+1} = \underset{x \in C}{\operatorname{argmin}} \left\{ \langle t_k f'(x^k) - \nabla \omega(x^k), x \rangle + \omega(x) \right\}$$

Remark 9.6 令 $\tilde{\omega} = \omega + \delta_C$，上式 $\Leftrightarrow$

$$x^{k+1} = \underset{x \in \mathbb{E}}{\operatorname{argmin}} \left\{ \langle t_k f'(x^k) - \nabla \omega(x^k), x \rangle + \tilde{\omega}(x) \right\}$$

由 $\nabla \omega(x^k) \in \partial \tilde{\omega}(x^k)$，故上式 $\Rightarrow$

$$x^{k+1} = \underset{x \in \mathbb{E}}{\operatorname{argmin}} \left\{ \langle t_k f'(x^k) - \tilde{\omega}'(x^k), x \rangle + \tilde{\omega}(x) \right\}$$

由 Thm 5.26，$\tilde{\omega}^*$ 可微，又由 Corollary 4.21，上式 $\Leftrightarrow$

$$x^{k+1} = \nabla \tilde{\omega}^* (\tilde{\omega}'(x^k) - t_k f'(x^k)).$$

由 Thm 4.20 $\quad x = \nabla \tilde{\omega}^*(y) \Leftrightarrow y \in \partial \tilde{\omega}'(x)$

$$\nabla \tilde{\omega}^* = (\partial \tilde{\omega}')^{-1}$$

Subproblem of Mirror Descent:

$$\min_{x \in C} \left\{ \langle a, x \rangle + \omega(x) \right\}$$

Lemma 9.7 设

- ω proper 闭凸，在 $\mathrm{dom}\,\partial\omega$ 可微
- ψ proper 闭凸，且 $\mathrm{dom}\,\psi \subseteq \mathrm{dom}\,\omega$
- $\omega + \delta_{\mathrm{dom}\,\psi}$ σ-强凸

则
$$\min_{x\in\mathbb{E}} \{\psi(x) + \omega(x)\} \qquad (9.11)$$
在 $\mathrm{dom}\,\psi \cap \mathrm{dom}\,\partial\omega$ 有唯一解.

证明: $(9.11) \Leftrightarrow \min_{x\in\mathbb{E}} \varphi(x)$, $\varphi = \psi + \omega$

φ 闭, proper, 且 σ-强凸, 由 Thm 5.25 (a), (9.12) 在

$\mathrm{dom}\,\varphi = \mathrm{dom}\,\psi$ 上有唯一解 x^*, 下证 $x^* \in \mathrm{dom}\,\partial\omega$

由 Fermat's optimality 条件: $0 \in \partial\varphi(x^*)$, 故由 Thm 3.40:

$$\partial\varphi(x^*) = \partial\psi(x^*) + \partial\omega(x^*), \text{ 故 } x^* \in \mathrm{dom}(\partial\omega). \quad \boxed{/}$$

Thm 9.8 设假设 9.1, 9.3 holds, $a \in \mathbb{E}^*$, 则

$$\min_{x\in C} \{\langle a, x\rangle + \omega(x)\}$$

在 $C\cap \mathrm{dom}\,\partial\omega$ 上有唯一解

证明：在 lemma 9.7 取 $\psi(x)\equiv\langle a,x\rangle+\delta_C(x)$ 即证 $\square$

Example 9.10 $\quad C=\Delta_n$

$$\omega(x)=\begin{cases}\sum_{i=1}^{n} x_i\log x_i, & x\in\mathbb{R}_+^n \\ \infty, & \text{else}\end{cases} \qquad (0\log 0 \doteq 0).$$

由 Example 5.27，$\omega+\delta_{\Delta_n}$ 1-strongly convex w.r.t. l_1-norm.

$$\mathrm{dom}(\partial\omega)=\mathbb{R}_{++}^n$$

对 $\forall x\in\Delta_n$，$y\in\Delta_n^+$，有

$$B_\omega(x,y)=\sum_{i=1}^{n}x_i\log x_i-\sum_{i=1}^{n}y_i\log y_i-\sum_{i=1}^{n}(\log y_i+1)(x_i-y_i)$$

$$=\sum_{i=1}^{n}x_i\log\frac{x_i}{y_i}$$

则

$$x^{k+1}=\underset{x\in\Delta_n}{\operatorname{argmin}}\left\{\sum_{i=1}^{n}(t_k f_i'(x^k)-1-\log x_i^k)x_i+\sum_{i=1}^{n}x_i\log x_i\right\}$$

Oracle

$$x^{k+1} = \operatorname*{argmin}_{x \in \Delta_n} \left\{ \langle a, x \rangle + \sum_{i=1}^{n} x_i \log x_i \right\}.$$

$$\mathcal{L}(x; \lambda, \mu) = \langle a, x \rangle + \sum x_i \log x_i - \lambda^T x + \mu (e^T x - 1)$$

KKT condition:

$$\begin{cases} x^* \in \Delta_n \\ \lambda^* \geq 0 \\ \lambda_i^* x_i^* = 0 \quad \forall i \quad \Rightarrow \lambda_i^* = 0 \\ a_i + 1 + \log x_i^* - \lambda_i^* + \mu^* = 0 \quad \forall i \end{cases}$$

$$\Rightarrow \quad \log x_i^* = -a_i - 1 - \mu^* \Rightarrow x_i^* = e^{-a_i - 1} \cdot e^{-\mu^*}$$

$$\sum x_i^* = e^{-\mu^*} \sum e^{-a_i - 1} = 1 \quad , \mu^* \text{ closed-form}$$

但，对 Oracle

$$x^{k+1} = \operatorname*{argmin}_{x \in \Delta_n} \left\{ \langle a, x \rangle + \frac{1}{2} \|x\|^2 \right\}.$$

$$\mathcal{L}(x; \lambda, \mu) = \langle a, x \rangle + \frac{1}{2} \|x\|^2 - \lambda^T x + \mu (e^T x - 1)$$

KKT条件：

(i) $x^* \in \Delta_n$

(ii) $\lambda^* \geqslant 0$

无 closed-form solution!

(iii) $\lambda_i^* x_i^* = 0$

(iv) $x^* + a - \lambda^* + \mu^* e = 0$