《Lectures on Convex 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 2022 年 9 月至 2023 年 7 月期间优化讨论班上的读书笔记，内容围绕 Yurii Nesterov 所著《Lectures on Convex Optimization》的前三章展开。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。然而，书中大量的推导过程略去了中间步骤，尤其在一些关键定理与证明中跳步较多，这给我们的阅读和理解带来一定困难。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《Lectures on Convex Optimization》第 1.1.1 节—3.2.7 节的全部内容，具体包括：

- 非线性优化初步：梯度法、牛顿法、共轭梯度法的一般性理论
- 光滑凸优化理论：L-光滑凸函数的性质、一阶方法的复杂度界、Nesterov 加速梯度法、极小-极大问题
- 非光滑凸优化理论：闭凸函数的性质、次梯度理论、非光滑优化最优性条件、极小-极大定理、次梯度算法

如何阅读此笔记

本笔记按照《Lectures on Convex Optimization》第 1.1.1 节—3.2.7 节顺序整理，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《Lectures on Convex Optimization》前三章主讲的同学有：南开大学的王雨畑、张迪、贾晖、张震、贾茹茹、于梅灵、李浩然、葛涵、李庆生、谢君，和天津大学的王学友、许君霞、王一静；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

§1.1 The world of nonlinear Optimization

$$\mathcal{P} \equiv (\underbrace{\Sigma}_{\text{model}},\ \underbrace{\mathcal{O}}_{\text{oracles}},\ \underbrace{\mathcal{T}_\varepsilon}_{\text{accuracy}})$$

$\mathcal{M}(\Sigma, \mathcal{O})$: method 定义在 $(\Sigma, \mathcal{O})$ 上

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

$$B_n = \{ x \in \mathbb{R}^n \mid 0 \leq x^{(i)} \leq 1,\ i = 1, \cdots, n \}$$

设 $f(\cdot)$ 是 L-Lipschitz的, i.e.

$$|f(x) - f(y)| \leq L \|x - y\|_\infty \quad \forall x, y \in B_n$$

Method $\mathscr{G}(p)$

1. Form p^n points

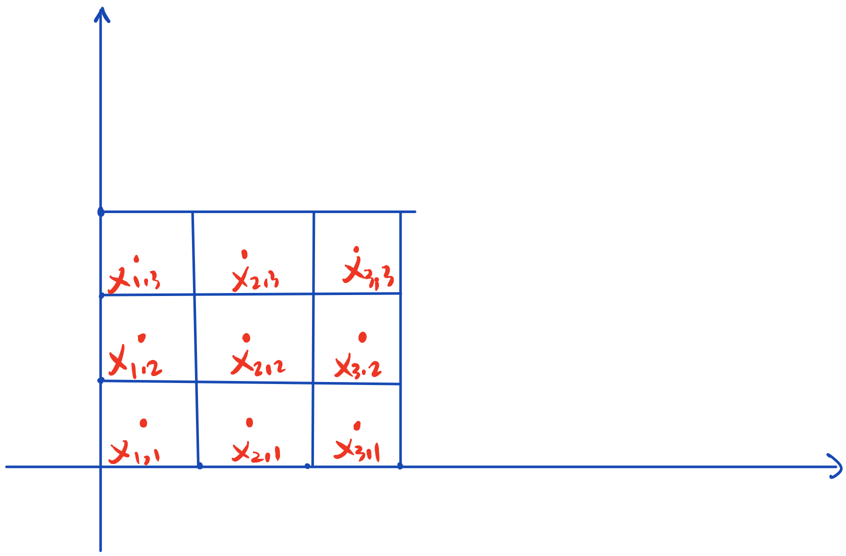
$$x_\alpha = \left(\frac{2i_1 - 1}{2p}, \frac{2i_2 - 1}{2p}, \ldots, \frac{2i_n - 1}{2p} \right)^T,$$

where $\alpha \equiv (i_1, \ldots, i_n) \in \{1, \ldots, p\}^n$.

2. Among all points x_α, find the point $\bar{x}$ with the minimal value of the objective function.

3. The pair $(\bar{x}, f(\bar{x}))$ is the output of the method.

(1.1.6)



Theorem 1.1.1

令 f^* 是 global optimal value，则 $f(\bar{x}) - f(x^*) \leq \frac{L}{2p}$

证明：对 $\alpha = (i_1, \cdots, i_n)$，令

$$X_\alpha = \left\{ x \in \mathbb{R}^n : \| x - x_\alpha \|_\infty \leq \frac{1}{2p} \right\}$$

则 $\bigsqcup_{\alpha \in \{1,\cdots,p\}^n} X_\alpha = B_n$，则设 x^* 是全局最优，则 $\exists$ 一个 multi-index α^*，

s.t. $x^* \in X_{\alpha^*}$，且 $\| x^* - x_{\alpha^*} \| \leq \frac{1}{2p}$，故

$$f(\bar{x}) - f(x^*) \leq f(x_{\alpha^*}) - f(x^*) \leq \frac{L}{2p}$$

$\boxtimes$

Corollary 1.1.1 对于method $\mathcal{G}$, analytical complexity 至多为：$\mathcal{A}(\mathcal{G}) = (\lfloor \frac{L}{2\varepsilon} \rfloor + 1)^n$

证明：令 $p = \lfloor \frac{L}{2\varepsilon} \rfloor + 1$，则 $p \geq \frac{L}{2\varepsilon}$，故

$$f(\bar{x}) - f^* \leq \frac{L}{2p} \leq \varepsilon \qquad \boxtimes$$

注：Local Black Box 的第二条：

对应了 Lipschiz 假设？

$$\mathcal{P}_\infty = (\Sigma, \mathcal{O}, \mathcal{T}_\varepsilon)$$

Σ：$f(x)$ 在 B_n 上是 ℓ_∞-Lipschiz 的

$\mathcal{O}$：任给 $x_0 \in B_n$，返回值是 $f(x_0)$

$\mathcal{T}_\varepsilon$：求 $\bar{x} \in B_n$，s.t. $f(\bar{x}) - f^* < \varepsilon$

Theorem 1.1.2

令 $\varepsilon < \frac{L}{2}$，则对于问题集合 $\mathcal{P}_\infty$ 来说，对于每个具体问题 $P \in \mathcal{P}_\infty$，它的解算法集合 $\mathcal{M} := \mathcal{M}(\mathcal{P}_\infty, \mathcal{O})$ 中任一算法 $\mathcal{S} \in \mathcal{M}$，分析复杂度有 $\mathcal{A}(P) \geq (\lfloor \frac{L}{2\varepsilon} \rfloor)^n$

证明：

考虑 $\mathcal{M}:=\mathcal{M}(\mathcal{P}_\infty, \mathcal{O})$ 集合的任何一个解算法 $\mathcal{S}$

$\mathcal{S}$ 通过在 B_n 上进行采样得到 $\{x_k\}$，再对 $\{x_k\}$ 调用 $\mathcal{O}$，对

$\{\mathcal{O}(x_k)\}$ 进行排序，得最优 ε-近似解

则 $\mathcal{M}$ 中解算法的差异仅在于 $\{x_k\}$ 的选取，上面的例子是均匀选取

下用反证法：

令 $p=\lfloor \frac{L}{2\varepsilon} \rfloor > 1$，若 $\exists\, \mathcal{S}\in\mathcal{M}$, s.t. $\exists\, P\in \mathcal{P}_\infty$，对于求解 P，

调用 $\mathcal{O}$ 的次数 $N<p^n$

则只需证：$\exists\, f_p \in \ell_\infty\text{-Lip}(B_n)$，s.t. $\mathcal{S}$ 在解 P 时，迭代

$N<p^n$ 次后的精度大于等于 ε

由 $N<p^n$，故存在非测试点 $\hat{x}$，集合

$$B=\left\{x \,\middle|\, \hat{x}\le x\le \hat{x}+\frac{1}{p}\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}\right\}\subseteq B_n \quad \text{（抽屉原理）}$$

（"$\le$"是按分量）

中不含任何测试点.

令 $x^*=\hat{x}+\frac{1}{2p}\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}\in B$，则令

$$f_p(x)=\min\{0,\ L\|x-x^*\|_\infty-\varepsilon\},\ x\in B$$

$$f_p(x)=0, \quad x\in B_n\setminus B$$

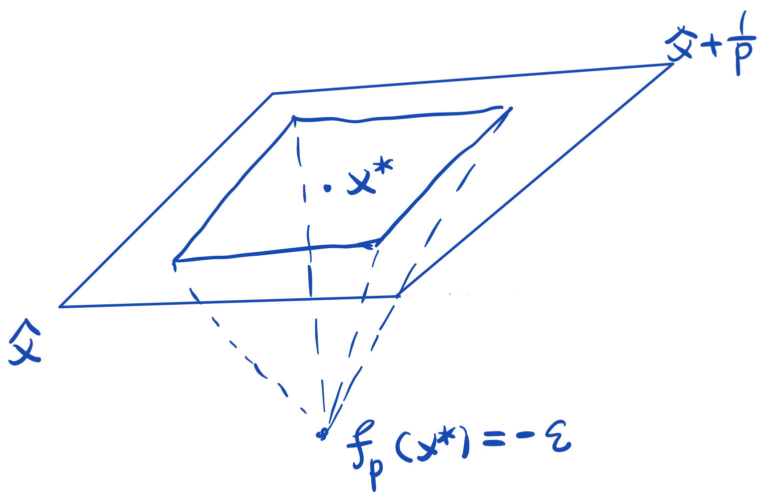
注意到：

$$\|\hat{x}-x^*\|_\infty = \frac{1}{2p} \geq \frac{1}{2\frac{L}{2\varepsilon}} = \frac{\varepsilon}{L}$$

则 $L\|\hat{x}-x^*\|-\varepsilon \geq \frac{\varepsilon}{L}\cdot L-\varepsilon \geq 0$

$$f_p(\hat{x})=\min\{0, L\|\hat{x}-x^*\|_\infty-\varepsilon\}=0$$

故



故 $f_p(x)$ 在 B_n 上 ℓ_∞-Lip，且 $f_p(x^*)=-\varepsilon$

而用 S 解 $f_p(x)$ 时，在 $\{x_k\}_{k=1}^N$ 调用 $\mathcal{O}$ 时，均返回

$$\mathcal{O}(x_k)=f(x_k)=0 \quad \forall k=1,\cdots,N$$

故 $f_p(\bar{x})=0$，有 $f_p(\bar{x})-f_p(x^*)\geq\varepsilon$，矛盾！

$\boxtimes$

ψ : $\left(\left\lfloor \frac{L}{2\varepsilon} \right\rfloor + 1\right)^n \iff \left\lfloor \frac{L}{2\varepsilon} \right\rfloor^n$

当 $\frac{L}{2\varepsilon} = \square.999\cdots9$ 时

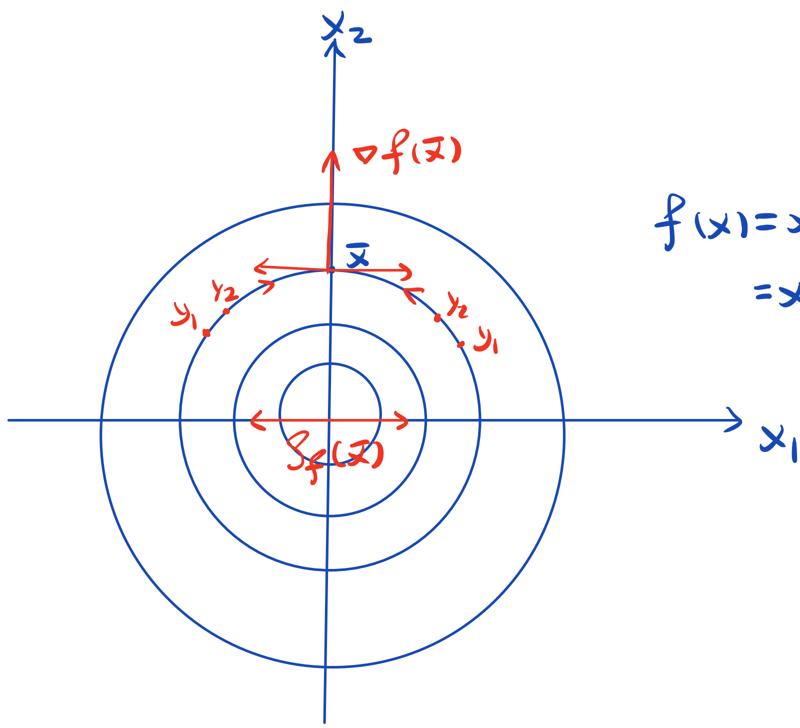
当 $\frac{L}{2\varepsilon} =$ 整数时

§ 1.2 Local Methods in Unconstrained Minimization

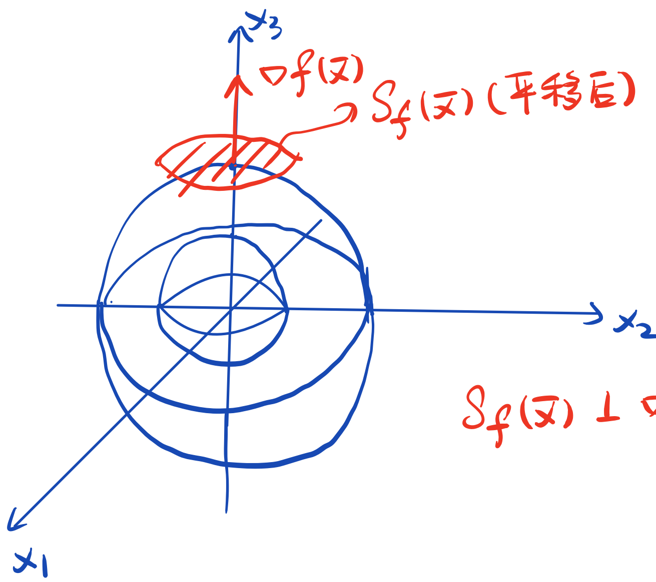
§ 1.2.1 Relaxation and Approximation

$$\mathcal{L}_f(\alpha) = \{ x \in \mathbb{R}^n \mid f(x) \le \alpha \}$$

$$S_f(\bar{x}) = \left\{ s \in \mathbb{R}^n \;\middle|\; s = \lim_{k \to \infty} \frac{y_k - \bar{x}}{\| y_k - \bar{x} \|},\ \exists\, y_k \to \bar{x},\ f(y_k) = f(\bar{x}),\ \forall k \right\}$$



$$f(x) = x^T x$$
$$= x_1^2 + x_2^2$$



$f(x) = x^T x$

$S_f(\bar{x}) \perp \nabla f(\bar{x})$

Lemma1.2.1 若 $s \in S_f(\bar{x})$，则 $\langle \nabla f(\bar{x}), s \rangle = 0$

证明:

$$f(y_k) = f(\bar{x}) + \langle \nabla f(\bar{x}), y_k - \bar{x} \rangle + o(\|y_k - \bar{x}\|)$$

$$\Rightarrow \quad \left\langle \nabla f(\bar{x}), \frac{y_k - \bar{x}}{\|y_k - \bar{x}\|} \right\rangle + o(1) = 0 \qquad \boxtimes$$

令 $\|s\| = 1$，令

$$\Delta(s) = \lim_{\alpha \to 0} \frac{1}{\alpha} \left(f(\bar{x} + \alpha s) - f(\bar{x}) \right)$$

由 $f(\bar{x} + \alpha s) - f(\bar{x}) = \alpha \langle \nabla f(\bar{x}), s \rangle + o(\alpha)$

故 $\Delta(s) = \langle \nabla f(\bar{x}), s \rangle \geqslant -\|\nabla f(\bar{x})\|$

取 $\xi = \frac{-\nabla f(\bar{x})}{\|\nabla f(\bar{x})\|}$，则 $\Delta(\xi) = -\|\nabla f(\bar{x})\|$

则 $-\nabla f(\bar{x})$ 是 $\bar{x}$ 处下降最快方向

<u>Theorm 1.2.1</u>

x^* 是 local minimum of 可微 func $f(\cdot)$，则 $\nabla f(x^*) = 0$

证明：

由 x^* local minimum，则 $\exists r > 0$, s.t. 对 $\forall y \in \mathbb{R}^n$, $\|y - x^*\| \le r$,

有 $f(y) \geq f(x^*)$，由 f 可微

$f(y) = f(x^*) + \langle \nabla f(x^*), y - x^* \rangle + o(\|y - x^*\|) \geq f(x^*)$

则 $\langle \nabla f(x^*), s \rangle \geq 0$，故取 $s = -\nabla f(x^*)$ 即证 $\boxtimes$

Corollary 1.2.1

优化问题 $\min f(x)$ 的局部极小是 x^*,
$\text{s.t. } Ax=b$

其中 $x\in\mathbb{R}^n$, $A\in\mathbb{R}^{m\times n}$ 是行满秩的，设 $\{x \mid Ax=b\}\neq\emptyset$,

则存在 $\lambda^*\in\mathbb{R}^m$, s.t. $\nabla f(x^*)=A^T\lambda^*$

证明：

当 $\nabla f(x^*)=0$ 时，取 $\lambda^*=0\in\mathbb{R}^m$ 即可，故不妨设 $\nabla f(x^*)\neq 0$

考虑：$g^*=\min\limits_{\lambda\in\mathbb{R}^m}\left\{g(\lambda)=\frac{1}{2}\|\nabla f(x^*)-A^T\lambda\|^2\right\}$

用反证法：设 $g^*>0$，则：

$$g(\lambda)=\frac{1}{2}\|\nabla f(x^*)\|^2-\langle\nabla f(x^*),A^T\lambda\rangle+\frac{1}{2}\langle B\lambda,\lambda\rangle$$

$B=AA^T\succeq\lambda_{\min}(B)I_m$，且 $\lambda_{\min}(B)>0$

由 $\text{rank}(B)=\text{rank}(AA^T)$
$=\text{rank}(A)=m$,

且 B 是半正定的，故 B 正定

由 $g(\lambda)$ 的下水平集是有界闭集，故是紧集，再由 $g(\lambda)$ 的连续性知：$\exists \lambda^* \in \mathbb{R}^m$, s.t. $g(\lambda^*) = \inf\limits_{\lambda \in \mathbb{R}^m} g(\lambda)$

（紧集上的连续func可以取到最大、最小值）

故 $0 = g(\lambda^*) = B\lambda^* - A\nabla f(x^*)$

即 $\lambda^* = B^{-1}A\nabla f(x^*)$，令 $s^* = (I_n - A^T B^{-1} A)\nabla f(x^*)$

则 $As^* = 0$（$s^*, -s^*$ 均是可行方向），由

$$\langle \nabla f(x^*), s^* \rangle = \langle \nabla f(x^*), \nabla f(x^*) - A^T B^{-1} A \nabla f(x^*) \rangle$$

$$= \|\nabla f(x^*)\|^2 - \langle B^{-1} A \nabla f(x^*), A\nabla f(x^*) \rangle$$

$$= 2g^* > 0$$

故设 $\alpha > 0$：

$$f(x^* - \alpha s^*) = f(x^*) - \alpha \nabla f(x^*)^T s^* + o(\alpha^2)$$

$$= f(x^*) - 2\alpha g^* + o(\alpha^2)$$

由 x^* 是局部极小，i.e. $\exists \delta > 0$, s.t. $\forall x \in \delta(x^*, \delta)$，$f(x) \geq f(x^*)$

故可取 α 充分小，s.t. $x^* - \alpha s^* \in B(x^*, \delta)$

故与 x^* 局部极小矛盾，从而假设不成立，$g^* = 0$ $\square$

注：A 行满秩是不失一般的，由 $Ax=b$ 解集非空，故可以将约束化简 $\tilde{A}x=\tilde{b}$，$(\tilde{A},\tilde{b})$ 是 (A,b) 的任一组极大线性无关组

1.2.2 Classes of Differentiable Functions

定义：$f\in C_L^{k,p}(Q)$，则 f 在 Q 上连续可微，且

$$\|\nabla^p f(x)-\nabla^p f(y)\|\leq L\|x-y\| \quad \forall x,y\in Q$$

设 $A:\mathbb{R}^n\longrightarrow\mathbb{R}^n$，则 $\|A\|=\sup\limits_{\|x\|=1}\|Ax\|$，即是算子范数

则设 $A=U\Sigma V^T$，则

$$\|A\|_2=\sup_{\|x\|=1}\|Ax\|_2=\sup_{\|x\|=1}\sqrt{x^TA^TAx}$$

$$=\sup_{\|x\|=1}\sqrt{x^TV\Sigma^2V^Tx}$$

$$=\sup_{\|y\|=1}\sqrt{y^T\Sigma^2y}=\sqrt{\lambda_{max}(A^TA)}$$

Lemma 1.2.2 $f(\cdot) \in C_L^{2,1}(\mathbb{R}^n) \subset C_L^{1,1}(\mathbb{R}^n) \Leftrightarrow$

$\forall x \in \mathbb{R}^n$, $\|\nabla^2 f(x)\| \leq L$

证明:

$(\Leftarrow)$

$$\nabla f(y) = \nabla f(x) + \int_0^1 \nabla^2 f(x+\tau(y-x))(y-x)d\tau$$

$$= \nabla f(x) + \left(\int_0^1 \nabla^2 f(x+\tau(y-x))d\tau\right)(y-x)$$

故

$$\|\nabla f(y) - \nabla f(x)\| = \left\|\left(\int_0^1 \nabla^2 f(x+\tau(y-x))d\tau\right)(y-x)\right\|$$

$$\leq \left\|\int_0^1 \nabla^2 f(x+\tau(y-x))d\tau\right\| \|y-x\|$$

$$\leq \int_0^1 \|\nabla^2 f(x+\tau(y-x))\| d\tau \, \|y-x\|$$

$$\leq L\|y-x\|$$

$(\Rightarrow)$ 对 $\forall s \in \mathbb{R}^n$, $\alpha > 0$

$$\left\|\left(\int_0^\alpha \nabla^2 f(x+\tau s)d\tau\right)s\right\| = \|\nabla f(x+\alpha s) - \nabla f(x)\|$$

$$\leq \alpha L \|s\|$$

则对 $\forall \alpha > 0$

$$\frac{1}{\alpha} \|\nabla f(x+\alpha s) - \nabla f(x)\| \leq L \|s\|$$

令 $\alpha \to 0^+$，

$$\lim_{\alpha \to 0^+} \frac{1}{\alpha} \|\nabla f(x+\alpha s) - \nabla f(x)\|$$

$$= \left\| \lim_{\alpha \to 0^+} \frac{\nabla f(x+\alpha s) - \nabla f(x)}{\alpha} \right\|$$

$$= \|\nabla^2 f(x) s\| \leq L \|s\| \quad (*)$$

由于有限维空间算子是可达范的，故 $\exists s \in \mathbb{R}^n$, $\|s\| = 1$,

s.t. $\|\nabla^2 f(x) s\| = \|\nabla^2 f(x)\|$，故不妨设

$\exists \tilde{x} \in \mathbb{R}^n$, s.t. $\|\nabla^2 f(\tilde{x})\| > L$，则 $\exists s$, s.t.

$$\left\|\nabla^2 f(x) \frac{s}{\|s\|}\right\| \|s\| = \|\nabla^2 f(x)\| \|s\| > L \|s\|$$

与 $(*)$ 矛盾，故 $\|\nabla^2 f(x)\| \leq L$，$\forall x \in \mathbb{R}^n$ $\square$

Lemma 1.2.3

$f\in C_L^{1,1}(\mathbb{R}^n)$, 则 $\forall x,y\in\mathbb{R}^n$, 有

$$|f(y)-f(x)-\langle\nabla f(x),y-x\rangle|\leq\frac{L}{2}\|y-x\|^2$$

证明: 对 $\forall x,y\in\mathbb{R}^n$

$$f(y)=f(x)+\int_0^1\langle\nabla f(x+\tau(y-x)),y-x\rangle d\tau$$

$$=f(x)+\langle\nabla f(x),y-x\rangle+\int_0^1\langle\nabla f(x+\tau(y-x))-\nabla f(x),y-x\rangle d\tau$$

则 $|f(y)-f(x)-\langle\nabla f(x),y-x\rangle|$

$$=\left|\int_0^1\langle\nabla f(x+\tau(y-x))-\nabla f(x),y-x\rangle d\tau\right|$$

$$\leq\int_0^1\|\nabla f(x+\tau(y-x))-\nabla f(x)\|\,\|y-x\|\,d\tau$$

$$\leq\int_0^1\tau L\|y-x\|^2d\tau=\frac{L}{2}\|y-x\|^2$$

$\boxtimes$

Lemma 1.2.4

$f\in C_M^{2,2}(\mathbb{R}^n)$, 则对 $\forall x,y\in\mathbb{R}^n$, 有

$$\|\nabla f(y)-\nabla f(x)-\nabla^2f(x)(y-x)\|\leq\frac{M}{2}\|y-x\|^2 \quad (1)$$

$$|f(y)-f(x)-\langle\nabla f(x),y-x\rangle-\frac{1}{2}\langle\nabla^2f(x)(y-x),y-x\rangle|\leq\frac{M}{6}\|y-x\|^3 \quad (2)$$

证明：(1) 与 Lemma 1.2.3 同理，只证 (2)

$$f(y) = f(x) + \langle \nabla f(x), y-x \rangle + \int_0^1 \langle \nabla f(x+\tau(y-x)) - \nabla f(x), y-x \rangle d\tau$$

$$\nabla f(x+\tau(y-x)) - \nabla f(x) = \int_0^1 \nabla^2 f(x+t\tau(y-x))\tau(y-x)\, dt$$

$$= \nabla^2 f(x)\tau(y-x) + \int_0^1 \left(\nabla^2 f(x+t\tau(y-x)) - \nabla^2 f(x)\right)\tau(y-x)\, dt$$

$$\int_0^1 \langle \nabla f(x+\tau(y-x)) - \nabla f(x), y-x \rangle d\tau$$

$$= \int_0^1 \tau(y-x)^T \nabla^2 f(x)(y-x) +$$

$$\int_0^1 \tau(y-x)^T \int_0^1 (\nabla^2 f(x+t\tau(y-x)) - \nabla^2 f(x))dt\ (y-x)\, d\tau$$

$$= \frac{1}{2}\langle \nabla^2 f(x)(y-x), y-x \rangle +$$

$$\int_0^1 \tau(y-x)^T \int_0^1 (\nabla^2 f(x+t\tau(y-x)) - \nabla^2 f(x))dt\ (y-x)\, d\tau$$

故 $$\left| f(y) - f(x) - \nabla f(x)^T(y-x) - \langle \nabla^2 f(x)(y-x), y-x \rangle \right|$$

$$= \left| \int_0^1 \tau(y-x)^T \int_0^1 (\nabla^2 f(x+t\tau(y-x)) - \nabla^2 f(x))dt\ (y-x)\, d\tau \right|$$

$$\leq \int_0^1 \tau(y-x)^T \int_0^1 \|\nabla^2 f(x+t\tau(y-x)) - \nabla^2 f(x)\|\, dt\ (y-x)\, d\tau$$

$$\leq \int_0^1 \tau (y-x)^T \int_0^1 M t\tau \|y-x\| dt \, (y-x) d\tau$$

$$= \int_0^1 \frac{1}{2} M \tau^2 \|y-x\|^3 d\tau = \frac{1}{6} M \|y-x\|^3$$

$\boxtimes$

1.2.3 The gradient Method

Choose $x_0 \in \mathbb{R}^n$

Interate $x_{k+1} = x_k - h_k \nabla f(x_k), \; k = 0, 1, \cdots$

Armijo rule: $x_{k+1} = x_k - h \nabla f(x_k), h > 0$, s.t.

$$\alpha \langle \nabla f(x_k), x_k - x_{k+1} \rangle \leq f(x_k) - f(x_{k+1}) \quad \cdots\cdots (1)$$

$$\beta \langle \nabla f(x_k), x_k - x_{k+1} \rangle \geq f(x_k) - f(x_{k+1}) \quad \cdots\cdots (2)$$

解释：

条件(1)：

$$f(x_k) - f(x_{k+1}) \geq \alpha h_k \|\nabla f(x_k)\|^2 \geq 0$$

即算法是下降的

条件(2)：

$$h_k \geq \frac{f(x_k) - f(x_{k+1})}{\beta \|\nabla f(x_k)\|^2} \geq \frac{h_k (1 - \frac{1}{2} h_k L)}{\beta}$$

则 $h_k \geq \frac{2(1-\beta)}{L}$

步长又不能太小

§1.2.3 The Gradient Method

考虑 $\min_{x\in\mathbb{R}^n} f(x)$，$f\in C_L^{1,1}(\mathbb{R}^n)$，且下方有界

令 $y=x-h\nabla f(x)$，则

$$f(y)\le f(x)+\langle\nabla f(x),y-x\rangle+\frac{L}{2}\|y-x\|^2$$

$$= f(x)\underbrace{-h(1-\frac{h}{2}L)\|\nabla f(x)\|^2}_{\text{下降量}}$$

故作 $\Delta(h)=-h(1-\frac{h}{2}L)$，找 $\min_{h>0}\Delta(h)$，从而有尽可能大的下降，由 $\Delta(h)$ 是关于 h 的二次 func，则

$$\boxed{h^*=\frac{1}{L}}$$

则一步 gradient descent 下降量有：

$$f(y)\le f(x)-\frac{1}{2L}\|\nabla f(x)\|^2$$

下面考虑另一种步长选取：

令 $x_{k+1}=x_k-h_k\nabla f(x_k)$ （对固定步长时，$h_k=h,\ \forall k$）

则同理:

$$f(x_k)-f(x_{k+1}) \geqslant h_k(1-\frac{1}{2}Lh_k)\|\nabla f(x_k)\|^2$$

若取 $h_k=\frac{2\alpha}{L}$, $\alpha\in(0,1)$,则

$$f(x_k)-f(x_{k+1}) \geqslant \frac{2}{L}\alpha(1-\alpha)\|\nabla f(x_k)\|^2$$

显然 $\alpha=\frac{1}{2}$ 时,取max,故 $h_k=\frac{1}{L}$,与刚才是统一的.

故对于 Full relaxation

$$f(x_k)-f(x_{k+1}) \geqslant \frac{1}{2L}\|\nabla f(x_k)\|^2$$

最后:考虑 Armijo rule,由 $x_{k+1}=x_k-h_k\nabla f(x_k)$

$$f(x_k)-f(x_{k+1}) \leq \beta\langle\nabla f(x_k), x_k-x_{k+1}\rangle=\beta h_k\|\nabla f(x_k)\|^2$$

$$f(x_k)-f(x_{k+1}) \geqslant h_k(1-\frac{h_k}{2}L)\|\nabla f(x_k)\|^2$$

故 $\beta \geqslant 1-\frac{h_k}{2}L$,即 $h_k \geqslant \frac{2}{L}(1-\beta)$

又由 Armijo rule:

$$f(x_k)-f(x_{k+1}) \geqslant \alpha\langle\nabla f(x_k), x_k-x_{k+1}\rangle=\alpha h_k\|\nabla f(x_k)\|^2$$

$$\geqslant \frac{2}{L}\alpha(1-\beta)\|\nabla f(x_k)\|^2$$

综上：无论3种步长选取方式，均有

$$f(x_k)-f(x_{k+1}) \geqslant \frac{w}{L}\|\nabla f(x_k)\|^2, \quad w\in \mathbb{R}^+$$

故

$$\frac{w}{L}\sum_{k=0}^{N}\|\nabla f(x_k)\|^2 \leqslant f(x_0)-f(x_{N+1}) \leqslant f(x_0)-f^*$$

取 $N\to+\infty$，则 $\|\nabla f(x_k)\| \longrightarrow 0 \ (k\to\infty)$

令 $g_N^* = \min\limits_{0\leq k\leq N} \|\nabla f(x_k)\|$

$$(N+1)\frac{w}{L} {g_N^*}^2 \leqslant \frac{w}{L}\sum_{k=0}^{N}\|\nabla f(x_k)\|^2 \leqslant f(x_0)-f^*$$

则 $g_N^* \leqslant \frac{1}{\sqrt{N+1}}\left[\frac{1}{w}L(f(x_0)-f^*)\right]^{\frac{1}{2}}$

可以说 $\|\nabla f(x_k)\|$ 的收敛，但没有 x_k 与 $f(x_k)$ 的收敛

例：

$$f(x)\equiv f(x^{(1)},x^{(2)}) = \frac{1}{2}(x^{(1)})^2+\frac{1}{4}(x^{(2)})^4-\frac{1}{2}(x^{(2)})^2$$

则

$$\nabla^2 f(x) = \begin{pmatrix} 1 & 0 \\ 0 & 3(x^{(2)})^2-1 \end{pmatrix}$$

则 $\|\nabla^2 f(x)\| = \max\{1, 3(x^{(2)})^2-1\}$，则 $\boxed{f \notin C_L^{1,1}(\mathbb{R}^2)}$

$\nabla f(x) = (x^{(1)}, (x^{(2)})^3 - x^{(2)})^T$

则 $x_1^* = (0,0)$，$x_2^* = (0,1)$，$x_3^* = (0,-1)$

则 x_1^* 不是局部极小 $\left(\nabla^2 f(x_1^*) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\right)$

但当 $x_0 = (1,0)$，则 $\nabla f(x_0)$ 的第二分量也是 0，

则 $\forall k$，$x_k^{(2)} = 0$，又由 $\|\nabla f(x_k)\| \to 0$，故 $x_k \to x_1^*$ $\boxtimes$

令 $g_N^* \le \frac{1}{\sqrt{N+1}}\left[\frac{1}{w} L(f(x_0) - f^*)\right]^{\frac{1}{2}} \le \varepsilon$

$N+1 \ge \frac{L}{w\varepsilon^2}(f(x_0) - f^*)$ 复杂度上界

下面考虑更特殊一类问题，从而保证梯度下降收敛到局部最优：

$$\min_{x\in\mathbb{R}^n} f(x)$$

假设

1) $f\in C_M^{2,2}(\mathbb{R}^n)$

2) $\exists$局部极小$x^*\in\mathbb{R}^n$，s.t. $\nabla^2 f(x^*)\succ 0$

3) 我们知道某个bound $0<\mu\le L<\infty$，s.t.

$$\mu I_n \preceq \nabla^2 f(x^*)\preceq L I_n$$

4) 初始点x_0与x^*足够近

考虑 $x_{k+1}=x_k-h_k\nabla f(x_k)$

则

$$\nabla f(x_k)=\nabla f(x_k)-\nabla f(x^*)$$

$$=\int_0^1\nabla^2 f(x^*+\tau(x_k-x^*))(x_k-x^*)d\tau$$

$$=G_k(x_k-x^*)$$

$$G_k=\int_0^1\nabla^2 f(x^*+\tau(x_k-x^*))d\tau$$

故

$$x_{k+1}-x^*=x_k-h_k\nabla f(x_k)-x^*=(I_n-h_kG_k)(x_k-x^*)$$

下证：$I_n - h_k G_k : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ 是一个压缩映射

令 $r_k = \|x_k - x^*\|$，由推论1.2.2

$$\underbrace{\nabla^2 f(x^*)}_{\substack{\succeq \\ \mu I_n}} - \tau M r_k I_n \preceq \nabla^2 f(x^* + \tau(x_k - x^*)) \preceq \underbrace{\nabla^2 f(x^*)}_{\substack{\preceq \\ L I_n}} + \tau M r_k I_n$$

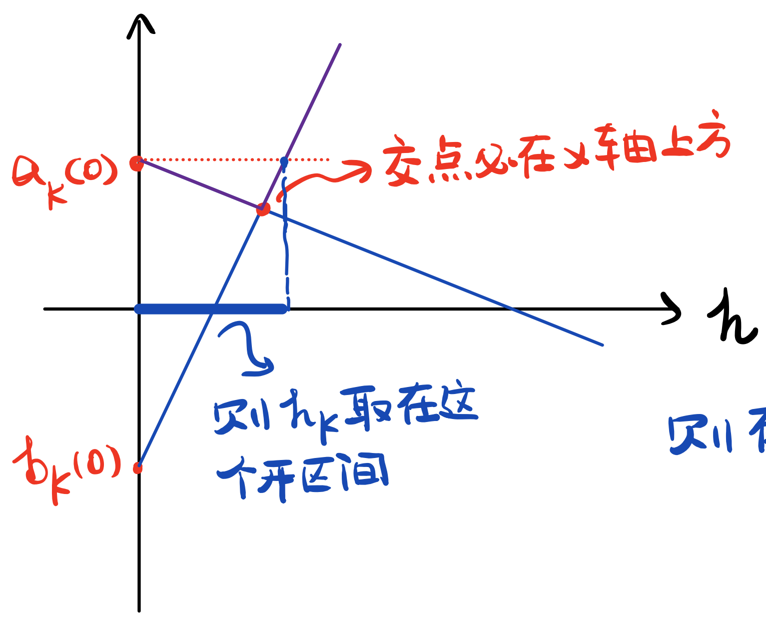
故 $(\mu - \frac{r_k}{2} M) I_n \preceq G_k \preceq (L + \frac{r_k}{2} M) I_n$

故 $(1 - h_k(L + \frac{r_k}{2} M)) I_n \preceq I_n - h_k G_k \preceq (1 - h_k(\mu - \frac{r_k}{2} M)) I_n$

$\Rightarrow \|I_n - h_k G_k\| \leq \max\{a_k(h_k), b_k(h_k)\}$

$a_k(h) = 1 - h(\mu - \frac{r_k}{2} M)$ $\quad$ $b_k = -1 + h_k(L + \frac{r_k}{2} M)$

若 $0 < r_k < \bar{r} \equiv \frac{2\mu}{M}$（令 $a_k(\cdot)$ 是↓的）



故 ∃ h_k 充分小，

s.t. $\|I_n - h_k G_k\| < 1$

则有 $r_{k+1} \leq \|I_n - h_k G_k\| r_k < r_k$

在 $r_0 < \bar{r}$ 的大前提下

$$找\ \min_{h} \max\{a_k(h), b_k(h)\}$$

则 $a_k(h) = b_k(h) \iff 1 - h(\mu - \frac{r_k}{2}M) = -1 + h(L + \frac{r_k}{2}M)$

故 $h_k^* = \frac{2}{L+\mu}$ (与Lipschiz常数M是无关的)

故取 h_k^* 作步长的话

$$r_{k+1} \leq \|I_n - h_k G_k\| r_k \leq a_k(h_k^*) r_k$$

$$= \frac{(L-\mu) r_k}{L+\mu} + \frac{M r_k^2}{L+\mu}$$

不妨令 $q = \frac{2\mu}{L+\mu}$, $a_k = \frac{M}{L+\mu} r_k$

则

$$\frac{M}{L+\mu} r_{k+1} \leq (1 - \frac{2\mu}{L+\mu}) \frac{M}{L+\mu} r_k + (\frac{M}{L+\mu} r_k)^2$$

$a_k - q = \frac{1}{L+\mu}(M r_k - 2\mu) < 0$

$$\iff a_{k+1} \leq (1-q) a_k + a_k^2 = a_k(1 + (a_k - q))$$

$$= \frac{a_k(1 - (a_k - q)^2)}{1 - (a_k - q)} \leq \frac{a_k}{1 + q - a_k}$$

$a_k - q > -1$，由 $a_k - q = \dfrac{M r_k - 2\mu}{L+\mu} \le -1$

$$\underbrace{M r_k}_{>0} \le \underbrace{-L+\mu}_{<0} \text{. 矛盾！}$$

故 $\dfrac{1}{a_{k+1}} \geq \dfrac{1+q}{a_k} - 1$，这等价于：

$$\frac{q}{a_{k+1}} - 1 \geq \frac{q(1+q)}{a_k} - q - 1$$

$$= (1+q)\left(\frac{q}{a_k} - 1\right)$$

$$\geq (1+q)^n \left(\frac{q}{a_0} - 1\right)$$

$$= (1+q)^n \left(\frac{2\mu}{L+\mu} \cdot \frac{L+\mu}{r_0 M} - 1\right)$$

$$= (1+q)^n \left(\frac{\bar{r}}{r_0} - 1\right)$$

故 $a_k \le \dfrac{q r_0}{r_0 + (1+q)^k (\bar{r} - r_0)} \le \dfrac{q r_0}{\bar{r} - r_0} \left(\dfrac{1}{1+q}\right)^k$

又由 $q = \dfrac{2\mu}{L+\mu}$，$a_k = \dfrac{M}{L+\mu} r_k$

$$\|x_k - x^*\| \le \frac{\bar{r} r_0}{\bar{r} - r_0} \left(1 - \frac{2\mu}{L+3\mu}\right)^k$$

1.2.4 得证 $\square$

1.2.4 Newton's Method

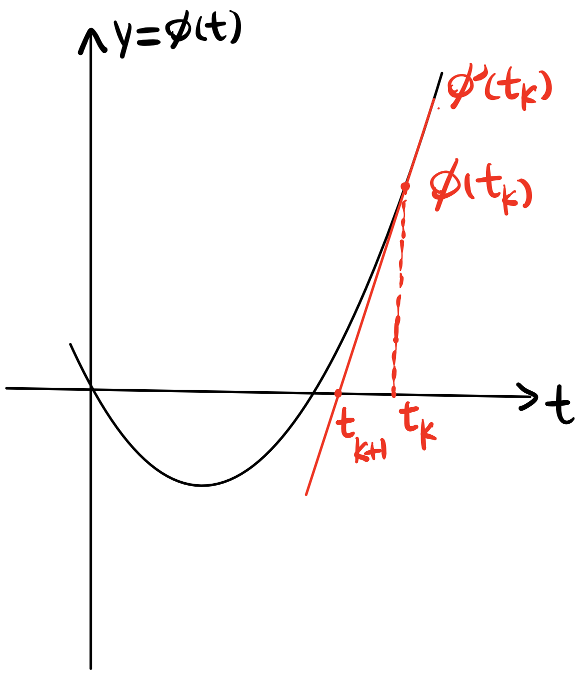
令 $\phi(\cdot):\mathbb{R}\longrightarrow\mathbb{R}$，考虑 $\phi(t^*)=0$（方程）

由
$$\phi(t+\Delta t)=\phi(t)+\phi'(t)\Delta t+o(|\Delta t|)$$

则 $\phi(t+\Delta t)=0$ 可以用 $\phi(t)+\phi'(t)\Delta t=0$ 估计

即 $\Delta t=-\dfrac{\phi(t)}{\phi'(t)}$

则有迭代序列：$t_{k+1}=t_k-\dfrac{\phi(t_k)}{\phi'(t_k)}$



如图：

$$\frac{\phi(t_k)}{t_k-t_{k+1}}=\phi'(t_k)$$

同理，对于方程组：$F(x)=0$

$$F(\cdot): \mathbb{R}^n \longrightarrow \mathbb{R}^n$$

那么设 $F'(x)$ 可逆，则同理有：

$$x_{k+1} = x_k - [F'(x_k)]^{-1} F(x_k)$$

将上述方法用在优化问题，可以将无约束优化问题

转化成求解方程：（必要条件）

$$\nabla f(x) = 0$$

则有：$x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$ ...... ①

注意到，迭代公式①可以用二次估计的角度获得：

令

$$\phi(x) = f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \frac{1}{2} \langle \nabla^2 f(x_k)(x - x_k), x - x_k \rangle$$

假设 $\nabla^2 f(x_k) \succ 0$，则

$$\nabla \phi(x) = \nabla f(x_k) + \nabla^2 f(x_k)(x - x_k)$$

令 x_{k+1} 满足：$\nabla \phi(x_{k+1}) = 0$，则

$$x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k) \quad \cdots\cdots \quad ①$$

Newton法的drawbacks:

① $\nabla^2 f(x_k)$ 奇异 ② 发散

例1.2.3

$$\phi(t) = \frac{t}{\sqrt{1+t^2}} \qquad (t^* = 0)$$

$$\phi'(t) = (1+t^2)^{-\frac{3}{2}}$$

则

$$t_{k+1} = t_k - \frac{\phi(t_k)}{\phi'(t_k)} = t_k - t_k(1+t_k^2) = -t_k^3$$

故若 $|t_0| < 1$，则 $t_k \longrightarrow 0 \ (k \to +\infty)$

$|t_0| = 1$ 时，振荡

$|t_0| > 1$ 时，发散

$\square$

所以，为了避免发散的情形，有如下 damped Newton 法

$$x_{k+1} = x_k - h_k [\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$$

h_k 在初始时取 gradient 法的步长，在最后，取 $h_k = 1$

为了建立 Newton 法收敛性，假设对优化问题

$$\min_{x \in \mathbb{R}^n} f(x)$$

① $f \in C_M^{2,2}(\mathbb{R}^n)$

② $\nabla^2 f(x^*) \succeq \mu I_n$ ，$\mu > 0$

③ x_0 离 x^* 足够近

则不妨设 $\nabla^2 f(x_k)$ 可逆（$\nabla^2 f(x_0)$ 可逆，由 x_0 与 x^* 足够近）

则：

$$x_{k+1} - x^* = x_k - x^* - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$$

$$= x_k - x^* - [\nabla^2 f(x_k)]^{-1} \int_0^1 \nabla^2 f(x^* + \tau(x_k - x^*))(x_k - x^*) d\tau$$

$$= [\nabla^2 f(x_k)]^{-1} G_k (x_k - x^*)$$

其中 $G_k = \int_0^1 [\nabla^2 f(x_k) - \nabla^2 f(x^* + \tau(x_k - x^*))] d\tau$

令 $r_k = \| x_k - x^* \|$，则

$$\| G_k \| = \left\| \int_0^1 [\nabla^2 f(x_k) - \nabla^2 f(x^* + \tau(x_k - x^*))] d\tau \right\|$$

$$\le \int_0^1 \| \nabla^2 f(x_k) - \nabla^2 f(x^* + \tau(x_k - x^*)) \| d\tau$$

$$\le \int_0^1 M \| x_k - \tau(x_k - x^*) - x^* \| d\tau$$

$$= \int_0^1 M(1-\tau) r_k d\tau = \frac{M}{2} r_k$$

由推论1.2.2

$$\nabla^2 f(x_k) \succeq \nabla^2 f(x^*) - M r_k I_n \succeq (\mu - M r_k) I_n$$

故若 $\mu - M r_k > 0$，$r_k < \frac{\mu}{M}$（即 $\exists c > 0, s.t. \nabla^2 f(x_k) \succeq cI$）

故 $\lambda_{\min}(\nabla^2 f(x_k)) \geq c$，则

$\lambda_{\max}(\nabla^2 f(x_k)^{-1}) \le c^{-1}$，即

$$\| [\nabla^2 f(x_k)]^{-1} \| \le (\mu - M r_k)^{-1}$$

$\| \cdot \|$只能是导出的算子范数

故

$$r_{k+1} \le \|[\nabla^2 f(x_k)]^{-1}\| \, \|G_k\| \, r_k$$

$$= \frac{M r_k}{2(\mu - M r_k)} \cdot r_k$$

则 $\frac{M r_k}{2(\mu - M r_k)} < 1$，即 $r_k < \frac{2\mu}{3M}$ 时，

有 $r_{k+1} \le \frac{M r_k^2}{2(\mu - M r_k)} \le r_k$

（自然地，$\nabla^2 f(x_{k+1})$ 也可逆，故迭代是有意义的，设 $\exists \delta > 0$.

s.t. $\forall x \in B(x^*, \delta)$，均有 $\nabla^2 f(x) > 0$，故只需 $\|x_0 - x^*\| \le \delta$

，则 $\forall k$，$x_k \in B(x^*, \delta)$）

定理 1.2.5 得证！

注：Gradient Method 要求 $r_0 < \frac{2\mu}{M}$，而

Newton Method 要求 $r_0 < \frac{2\mu}{3M}$ 是一个量级的，故

可以先用梯度法迭代有限步进入 $B(x^*, \frac{2\mu}{3M})$，再用 Newton 法

收敛率：

- 次线性 $r_k \le \frac{c}{\sqrt{k}}$ ....
- 线性 $r_k \le c(1-q)^k \le ce^{-qk}$
- 二次 $r_{k+1} \le cr_k^2$

1.3.1

对于优化问题

$$\min_{x\in\mathbb{R}^n} f(x)$$

$f\in C_M^{2,2}(\mathbb{R}^n)$，牛顿法与梯度法都为了找局部极小

牛顿法有二次收敛而梯度法只有线性收敛，为什么

会出现这样的不同呢？

对于固定点 $x\in\mathbb{R}^n$，作 $f(\cdot)$ 的估计 $\phi(\cdot)$

$$\phi_1(x) = f(x_k) + \langle \nabla f(x_k), x - x_k\rangle + \frac{1}{2h}\|x - x_k\|^2 \quad (h>0)$$

则 $x_{k+1} = \underset{x\in\mathbb{R}^n}{\operatorname{argmin}}\ \phi_1(x) \iff$

$$\nabla\phi_1(x_{k+1}) = \nabla f(x_k) + \frac{1}{h}(x_{k+1} - x_k) = 0$$

$$x_{k+1} = x_k - h\nabla f(x_k) \quad \cdots\cdots \text{梯度法}$$

注意到，由 Lemma 1.2.3，若 $h\in(0,\frac{1}{L}]$，则

$$f(x) \le \phi_1(x),\ \forall x\in\mathbb{R}^n$$

我们还可以对 $f(\cdot)$ 作二次估计

$$\phi_2(x) = f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \frac{1}{2} \langle \nabla^2 f(x_k)(x - x_k), x - x_k \rangle$$

则令 $x_{k+1} = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \ \phi_2(x)$ 知：

$$x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$$

容易看到 ϕ_2 的估计精度是高于 ϕ_1 的，因此易知牛顿法的收敛速度比梯度法快，但 ϕ_2 要算 $[\nabla^2 f(x_k)]^{-1}$，计算量更大

令 $G_k \in S_+^n$，定义

$$\phi_{G_k}(x) = f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \frac{1}{2} \langle G_k(x - x_k), x - x_k \rangle$$

则令 $x_{k+1} = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \ \phi_{G_k}(x)$，则

$$x_{k+1} = x_k - G_k^{-1} \nabla f(x_k)$$

生成矩阵序列 $H_k \equiv G_k^{-1} \longrightarrow [\nabla^2 f(x_k)]^{-1}$

这是一个一阶算法，生成 G_k 或 H_k 时仅用到了梯度信息，叫做 variable metric 法或 Quasi-Newton 法

定义内积空间 $(\mathbb{R}^n, \langle\cdot,\cdot\rangle_A)$

$$\langle x, y\rangle_A = \langle Ax, y\rangle, \ \forall x, y \in \mathbb{R}^n$$

则内积诱导的范数是 $\|x\|_A = \langle Ax, x\rangle^{\frac{1}{2}}$

$(\mathbb{R}^n, \langle\cdot,\cdot\rangle_A)$ 与欧式空间线性同胚

$$f(x+h) = f(x) + \langle \nabla f(x), h\rangle + \frac{1}{2}\langle \nabla^2 f(x)h, h\rangle + o(\|h\|)$$

$$= f(x) + \langle A^{-1}\nabla f(x), h\rangle_A + \frac{1}{2}\langle A^{-1}\nabla^2 f(x)h, h\rangle_A + o(\|h\|_A)$$

故 $\nabla f_A(x) = A^{-1}\nabla f(x)$

$$\nabla^2 f_A(x) = A^{-1}\nabla^2 f(x)$$

关于不同坐标系下求梯度

在有限维空间下，1-form df一定是不变量
(不是向量场)

(由于有限维空间各种拓扑等价，而

$$df(v_0)=\lim_{t\to 0}\frac{f(x+tv_0)-f(x)}{t}$$

只依赖拓扑，不依赖度量，故在坐标变换下是不变的)

在二维下考虑(不失一般地)

变换前坐标系是 (x,y)，变换后是 (u,v)

则在 (x,y) 下：$v=(a,b)$

$$df(v)=\frac{\partial f}{\partial x}dx+\frac{\partial f}{\partial y}dy$$

$$=af_x+bf_y$$

梯度最原始的定义是

$$\langle\nabla f,\vec{v}\rangle_g=df(\vec{v}),\text{对}\ \forall\vec{v}\in\mathbb{R}^n$$

由Riesz表示定理，∇f 一定是存在的，但依赖度量

$\langle\cdot,\cdot\rangle_g$ 表示任一内积

在 (x,y) 下，内积是通常的内积：

$$\langle (a,b),(c,d)\rangle = ac+bd$$

但在 (u,v) 下，内积有如下形式：

$$a\,dx+b\,dy=(a x_u+b y_u)\,du+(a x_v+b y_v)\,dv$$

故 $\langle (a,b),(c,d)\rangle_g=(x_u^2+y_u^2)ac+$

$(y_u x_v+y_u y_v)(bc+ad)+(x_v^2+y_v^2)bd$

所以，若变换是 $\begin{pmatrix} x' \\ y' \end{pmatrix}=\begin{pmatrix} \alpha_1 x \\ \alpha_2 y \end{pmatrix}$

则

$$\langle (a,b),(c,d)\rangle_g=\frac{ac}{\alpha_1^2}+\frac{bd}{\alpha_2^2}$$

则 $\nabla f=(p,q)$（在 (x,y) 下表示）满足

$$df(v)=\langle \nabla f, v\rangle \qquad \forall v$$

$$a f_x+b f_y=\frac{ap}{\alpha_1^2}+\frac{bq}{\alpha_2^2}$$

故令 $v=(1,0),(0,1)$ 即得

$$\nabla f = (\alpha_1^2 f_x, \ \alpha_2^2 f_y)$$

$$= \begin{pmatrix} \alpha_1^2 & \\ & \alpha_2^2 \end{pmatrix} \begin{pmatrix} f_x \\ f_y \end{pmatrix}$$

看回本例： $x^T A x = \langle x, x \rangle_A$

若此时有 $A = \begin{pmatrix} \frac{1}{\alpha_1^2} & \\ & \frac{1}{\alpha_2^2} \end{pmatrix}$

那么相当于作变换：

$$x' = A^{-\frac{1}{2}} x$$

原空间单位向量 v

$v \longmapsto A^{-\frac{1}{2}} v$

则 $\| A^{-\frac{1}{2}} v \|_A = v^T A^{-\frac{1}{2}} A A^{-\frac{1}{2}} v = 1$

即 $\begin{pmatrix} x_1' \\ x_2' \end{pmatrix} = \begin{pmatrix} \alpha_1 x_1 \\ \alpha_2 x_2 \end{pmatrix}$

则由刚刚的推导

$$\nabla f_A = \begin{pmatrix} \alpha_1^2 & \\ & \alpha_2^2 \end{pmatrix} \begin{pmatrix} f_x \\ f_y \end{pmatrix}$$

$$= A^{-1} \begin{pmatrix} f_x \\ f_y \end{pmatrix}$$

当 A不是对角阵时，不妨设

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

此时 x_v, y_u 项就有了 ⊛

同理可以计算，这时 $A = R^T \Lambda R$

$$= \underbrace{(\Lambda^{\frac{1}{2}} R)^T}(\Lambda^{\frac{1}{2}} R)$$

变换阵是 $R^T \Lambda^{-\frac{1}{2}}$，可代入

⊛ 时的计算结果

df 是拓扑不变量，但 ∇f 不是（依赖度量）

Variable metric method
0. Choose $x_0 \in \mathbb{R}^n$. Set $H_0 = I_n$. Compute $f(x_0)$ and $\nabla f(x_0)$.
1. kth iteration ($k \geq 0$). (a) Set $p_k = H_k \nabla f(x_k)$. (b) Find $x_{k+1} = x_k - h_k p_k$ (see Section 1.2.3 for step size rules). (c) Compute $f(x_{k+1})$ and $\nabla f(x_{k+1})$. (d) Update the matrix H_k to H_{k+1}.

Quasi-Newton rule

$$H_{k+1}(\nabla f(x_{k+1}) - \nabla f(x_k)) = x_{k+1} - x_k$$

1.3.2 Conjugate Gradient

$$\min_{x\in\mathbb{R}^n} f(x) = \alpha + \langle a, x\rangle + \frac{1}{2}\langle Ax, x\rangle, \quad A>0$$

则显然 $x^* = -A^{-1}a$，那么目标func可以变形

$$f(x) = \alpha + \langle a, x\rangle + \frac{1}{2}\langle Ax, x\rangle$$

$$= \alpha - \langle Ax^*, x\rangle + \frac{1}{2}\langle Ax, x\rangle$$

$$= \alpha - \frac{1}{2}\langle Ax^*, x^*\rangle + \frac{1}{2}\langle A(x-x^*), (x-x^*)\rangle$$

则 $f^* = \alpha - \frac{1}{2}\langle Ax^*, x^*\rangle$，$\nabla f(x) = A(x-x^*)$

给定初始点，$x_0\in\mathbb{R}^n$，则定义Krylov子空间：

$$\mathcal{L}_k = \mathrm{span}\{A(x_0-x^*), \cdots, A^k(x_0-x^*)\}$$

则 $$x_k = \arg\min\{f(x) \mid x\in x_0+\mathcal{L}_k\}, \quad k\geq 1$$

如此生成 $\{x_k\}$ 叫共轭梯度法

注：如此做必能找到 $x^*=-A^{-1}a$

$$\chi(\lambda)=\det(\lambda I-A)=\lambda^n+\alpha_1\lambda^{n-1}+\cdots+\alpha_n$$

则由 Hamilton-Caley 定理

$$\chi(A)=A^n+\alpha_1 A^{n-1}+\cdots+\alpha_n I$$

由 $A>0$，$\lambda_n\neq 0$，故 $\alpha_n\neq 0$

$$x^*-x_0=\frac{1}{\alpha_n}A^n(x_0-x^*)+\cdots+\frac{\alpha_{n-1}}{\alpha_n}A(x_0-x^*)\in\mathcal{L}_n$$

故 $x^*\in x_0+\mathcal{L}_n$ $\boxtimes$

（n 步之内有解）

注 2：

$$\mathcal{L}_1\subsetneqq\mathcal{L}_2\subsetneqq\cdots\subsetneqq\mathcal{L}_k=\mathcal{L}_{k+1}=\cdots=\mathcal{L}_n=\cdots$$

引理 1.3.1

若 $A^{k+1}(x_0-x^*) \notin \mathcal{K}_k$，则 $x_k \neq x_{k+1}$，反之 $x_k \neq x_{k+1}$，

则 $A^{k+1}(x_0-x^*) \notin \mathcal{K}_k$

证明：反过来是显然的，下证($\Rightarrow$)

由 $A^{k+1}(x_0-x^*) \notin \mathcal{K}_k$，不妨令

$x_{k+1} = x_0 + y + \lambda A^{k+1}(x_0-x^*)$，其中 y 满足

$x_k = x_0 + y$ (y 是存在唯一的)

则 $f(x_{k+1}) = \frac{1}{2}\langle A(x_{k+1}-x^*), x_{k+1}-x^*\rangle + \text{const}$

$= \frac{1}{2}\langle A(x_0+y+\lambda A^{k+1}(x_0-x^*)-x^*), x_0+y+\lambda A^{k+1}(x_0-x^*)-x^*\rangle + \text{const}$

$= \frac{1}{2}\langle A(x_0+y-x^*), x_0+y-x^*\rangle +$

$\frac{1}{2}\langle A(x_0+y-x^*), \lambda A^{k+1}(x_0-x^*)\rangle +$

$\frac{1}{2}\langle \lambda A^{k+2}(x_0-x^*), x_0+y-x^*\rangle + o(\lambda) + \text{const}$

$= f(x_k) + \lambda\langle A^{k+2}(x_0-x^*), x_0-x^*\rangle + \lambda\langle A^{k+2}(x_0-x^*), y\rangle$

$+ o(\lambda) +$

由 $A \succ 0$, $y=\sum_{j=1}^{k}\lambda^{(j)}A^{j}(x_0-x^*)$，故取 $\lambda<0$，且 $|\lambda|$ 足够小，

可以使 $f(x_{k+1})<f(x_k)$

$\boxed{//}$

Lemma 1.3.1

$\forall k\geqslant 1$, $\mathcal{L}_k=\mathrm{span}\{\nabla f(x_0),\cdots,\nabla f(x_{k-1})\}$

证明：用数学归纳法：

$n=1$ 时，$\nabla f(x_0)=A(x_0-x^*)$，成立

设对 $n=k$ 时成立，下证对 $n=k+1$ 成立

$$x_k=x_0+\sum_{i=1}^{k}\lambda^{(i)}A^{i}(x_0-x^*)\in x_0+\mathcal{L}_k \quad \exists\lambda\in\mathbb{R}^k$$

当 $\lambda^{(k)}=0$ 时：

有 $x_k\in\mathcal{L}_{k-1}$，故 $x_k=x_{k-1}$，故 $\mathcal{L}_k=\mathcal{L}_{k-1}$，从而

$$\begin{aligned}\mathcal{L}_k&=\mathrm{span}\{\nabla f(x_0),\cdots,\nabla f(x_{k-2})\}\\&=\mathrm{span}\{\nabla f(x_0),\cdots,\nabla f(x_{k-2}),\nabla f(x_{k-1})\}\end{aligned}$$

当 $\lambda^{(k)}\neq 0$ 时

$$\nabla f(x_k) = A(x_0 - x^*) + \sum_{i=1}^{k} \lambda^{(i)} A^{i+1}(x_0 - x^*)$$

$$= y + \lambda^{(k)} A^{i+1}(x_0 - x^*)$$

其中 $y \in \mathcal{L}_k$

$$\mathcal{L}_{k+1} \equiv \mathrm{span}\{\mathcal{L}_k \sqcup A^{k+1}(x_0 - x^*)\}$$

$$= \mathrm{span}\{\mathcal{L}_k \sqcup (\nabla f(x_k) - y)\}$$

$$= \mathrm{span}\{\mathcal{L}_k \sqcup \nabla f(x_k)\}$$

$$= \mathrm{span}\{\nabla f(x_0), \cdots, \nabla f(x_k)\} \qquad \boxed{/}$$

引理 1.3.2

$\forall\, k, i \geq 0,\ k \neq i$，有 $\langle \nabla f(x_k), \nabla f(x_i) \rangle = 0$

证明:

不妨令 $k > i$，考虑.

$$\phi(\lambda) = f\left(x_0 + \sum_{j=1}^{k} \lambda^{(j)} \nabla f(x_{j-1})\right), \quad \lambda \in \mathbb{R}^k$$

由引理 1.3.1，$\exists\, \lambda_* \in \mathbb{R}^k$, s.t.

$$x_k = x_0 + \sum_{j=1}^{k} \lambda_*^{(j)} \nabla f(x_{j-1})$$

由 x_k 是 $f(x)$ 在 $x_0 + \mathcal{L}_k$ 上的极小值，故

$$\nabla \phi(\lambda_*) = 0$$

$$0 = \left.\frac{\partial \phi(\lambda)}{\partial \lambda^{(j)}}\right|_{\lambda=\lambda_*} = \langle \nabla f(x_k), \nabla f(x_{j-1}) \rangle \quad \forall j = 1, \cdots, k$$

$\boxed{}$

令 $\delta_i = x_{i+1} - x_i$，则 $\mathcal{L}_k = \mathrm{span}\{\delta_0, \cdots, \delta_{k-1}\}$

证明：同样用数学归纳法

$n=0$ 时，设 $x_1 = x_0 + \lambda \nabla f(x_0)$

$\delta_0 = x_1 - x_0 = \lambda \nabla f(x_0)$，成立

设 $n=k-1$ 时成立，证明 $n=k$ 时成立

$$\begin{aligned}
\delta_k &= x_{k+1} - x_k \\
&= x_0 + \sum_{j=0}^{k} \lambda_1^{(j)} \nabla f(x_j) - x_0 + \sum_{j=0}^{k-1} \lambda_2^{(j)} \nabla f(x_j) \\
&= \sum_{j=0}^{k-1} (\lambda_1^{(j)} - \lambda_2^{(j)}) \nabla f(x_j) + \lambda_1^{(k)} \nabla f(x_k)
\end{aligned}$$

$\lambda_1^{(k)}=0$时，即 $x_{k+1}\in \mathcal{L}_k$，故 $\mathcal{L}_{k+1}=\mathcal{L}_k$ ✓

$\lambda_1^{(k)}\neq 0$时

$$\mathcal{L}_{k+1}=\operatorname{span}\{\delta_0,\dots,\delta_{k-1},\nabla f(x_k)\}$$
$$=\operatorname{span}\{\delta_0,\dots,\delta_k\}$$ ✓

$\square$

引理 1.3.3

$\forall k,i\geq 0,k\neq i$，有 $\langle A\delta_k,\delta_i\rangle=0$

证：WLOG，设 $k>i$，则

$$\langle A\delta_k,\delta_i\rangle=\langle A(x_{k+1}-x_k),\delta_i\rangle$$
$$=\langle \nabla f(x_{k+1})-\nabla f(x_k),\delta_i\rangle$$

$\delta_i=x_{i+1}-x_i\in\mathcal{L}_{i+1}\subseteq\mathcal{L}_k\subseteq\mathcal{L}_{k+1}$，从而

$$\langle A\delta_k,\delta_i\rangle=0$$

$\square$

由 $\mathcal{L}_k = span\{\delta_0, \cdots, \delta_{k-1}\} = span\{\delta_0, \cdots, \delta_{k-2}, \nabla f(x_{k-1})\}$

则 $x_{k+1} \in \mathcal{L}_{k+1}$ 总可以表示成如下形式：

$$x_{k+1} = x_k - h_k \nabla f(x_k) + \sum_{j=0}^{k-1} \lambda^{(j)} \delta_j$$

$$\delta_k = -h_k \nabla f(x_k) + \sum_{j=0}^{k-1} \lambda^{(j)} \delta_j$$

则对 $0 \leq i \leq k-1$，有

$$\begin{aligned}
0 &= \langle A\delta_k, \delta_i \rangle \\
&= -h_k \langle A \nabla f(x_k), \delta_i \rangle + \sum_{j=0}^{k-1} \lambda^{(j)} \langle A\delta_j, \delta_i \rangle \\
&= -h_k \langle \nabla f(x_k), A\delta_i \rangle + \lambda^{(i)} \langle A\delta_i, \delta_i \rangle \\
&= -h_k \langle \nabla f(x_k), \nabla f(x_{i+1}) - \nabla f(x_i) \rangle + \lambda^{(i)} \langle A\delta_i, \delta_i \rangle
\end{aligned}$$

故当 $i < k-1$ 时，$\lambda_i = 0$

$i = k-1$ 时

$$0 = -h_k \|\nabla f(x_k)\|^2 + \lambda^{(k-1)} \langle A\delta_i, \delta_i \rangle$$

$$\lambda^{(k-1)} = \frac{h_k \|\nabla f(x_k)\|^2}{\langle A\delta_{k-1}, \delta_{k-1} \rangle} = \frac{h_k \|\nabla f(x_k)\|^2}{\langle \nabla f(x_k) - \nabla f(x_{k-1}), \delta_{k-1} \rangle}$$

故

$$x_{k+1} = x_k - h_k \nabla f(x_k) + \lambda^{(k-1)} \delta_{k-1}$$

$$\doteq x_k - h_k p_k$$

$$p_k = \nabla f(x_k) - \frac{\|\nabla f(x_k)\|^2 \delta_{k-1}}{\langle \nabla f(x_k) - \nabla f(x_{k-1}), \delta_{k-1} \rangle}$$

$$= \nabla f(x_k) - \frac{\|\nabla f(x_k)\|^2 p_{k-1}}{\langle \nabla f(x_k) - \nabla f(x_{k-1}), p_{k-1} \rangle}$$

(由 $\delta_{k-1} = -h_{k-1} p_{k-1}$)

则我们定义了方向 $\{p_k\}$

关于Schmidt正交化过程，在内积空间 $\langle \cdot,\cdot \rangle_A$ 中

给定 $(\nabla f(x_0), \cdots, \nabla f(x_{n-1}))$，想找

$(p_0, \cdots, p_{n-1})$ 便得 $\langle p_i, p_j \rangle_A$ 正交

最原始的做法（不用二次func的性质）

$$p_{i+1} = \nabla f(x_{i+1}) + \sum_{j=0}^{i} \lambda_j^{(i+1)} p_j$$

则由，$k = 0, \cdots, i$ 时

$$0 = \langle p_{i+1}, p_k \rangle_A = \langle \nabla f(x_{i+1}), p_k \rangle_A + \sum_{j=0}^{i} \lambda_j^{(i+1)} \langle p_j, p_k \rangle_A$$

$$= \langle \nabla f(x_{i+1}), p_k \rangle_A + \lambda_k^{(i+1)} \langle p_k, p_k \rangle_A$$

故 $$\lambda_k^{(i+1)} = -\frac{\langle \nabla f(x_{i+1}), p_k \rangle_A}{\langle p_k, p_k \rangle_A}$$

故 $$p_{i+1} = \nabla f(x_{i+1}) - \sum_{j=0}^{i} \frac{\langle \nabla f(x_{i+1}), p_j \rangle_A}{\langle p_j, p_j \rangle_A} p_j$$

之后再化简就要用到二次func性质：

由 $\langle \nabla f(x_i), \nabla f(x_j) \rangle = 0,\ i \neq j$

$$\nabla f(x_{i+1}) - \nabla f(x_i) = A(x^{i+1} - x^i) = -h_i A p_i$$

故

$$\langle \nabla f(x_{i+1}), p_j \rangle_A = \langle \nabla f(x_{i+1}), \nabla f(x_{j+1}) - \nabla f(x_j) \rangle \cdot -\frac{1}{h_j}$$

$$= \begin{cases} 0, & j = 0, 1, \cdots, i-1 \\ -\frac{1}{h_i} \|\nabla f(x_{i+1})\|^2 \end{cases}$$

而 $\langle p_j, p_j \rangle_A = -\frac{1}{h_j} p_j^T (\nabla f(x_{j+1}) - \nabla f(x_j))$

故

$$p_{i+1} = \nabla f(x_{i+1}) - \frac{\|\nabla f(x_{i+1})\|^2}{\langle \nabla f(x_{i+1}) - \nabla f(x_i), p_i \rangle} \cdot p_i$$

与书上对起来了，实际上与书上等价

与 Schmidt 正交化不同的是：初始的线性无关组给的特殊，它们在 $\langle \cdot, \cdot \rangle$ 正交，所以很多系数是 0

1.3.3 Constrained Minimization

$$\min_{x\in Q} f_0(x)$$

$$f_j(x)\leqslant 0,\ j=1,\cdots,m$$

$Q\subset\mathbb{R}^n$是闭的，$f_0,\cdots,f_m$连续

1.3.3.1 Lagrangian Relaxation

定理 1.3.1

令$F(x,\lambda)$定义在$x\in Q_1\subseteq\mathbb{R}^n$，$\lambda\in Q_2\in\mathbb{R}^m$，$Q_1, Q_2\neq\emptyset$，则

$$\sup_{\lambda\in Q_2}\inf_{x\in Q_1} F(x,\lambda)\leqslant\inf_{x\in Q_1}\sup_{\lambda\in Q_2} F(x,\lambda)$$

证明：对$\forall x\in Q_1$，$\lambda\in Q_2$

$$F(x,\lambda)\leqslant\sup_{\xi\in Q_2} F(x,\xi)$$

则 $\inf_{x\in Q_1} F(x,\lambda)\leqslant\inf_{x\in Q_1}\sup_{\xi\in Q_2} F(x,\xi)$ 对$\forall\lambda\in Q_2$

故 $\sup_{\lambda\in Q_2}\inf_{x\in Q_1} F(x,\lambda)\leqslant\inf_{x\in Q_1}\sup_{\lambda\in Q_2} F(x,\lambda)$ $\square$

注意到

$$f^* = \inf_{x\in Q}\{ f_0(x) : f_j(x)\leq 0,\ j=1,\cdots,m\}$$

$$= \inf_{x\in Q}\sup_{\lambda\in\mathbb{R}^m_+} \mathcal{L}(x,\lambda)$$

其中 $\mathbb{R}^m_+ = \{\lambda\in\mathbb{R}^m, \lambda^{(j)}\geq 0,\ j=1,\cdots,m\}$

$$\mathcal{L}(x,\lambda) = f_0(x) + \langle \lambda, f(x)\rangle$$

则显然

$$\sup_{\lambda\in\mathbb{R}^m_+} \mathcal{L}(x,\lambda) = \begin{cases} +\infty, & x\notin\Omega \\ f_0(x), & x\in\Omega \end{cases}$$ ，Ω 是可行集

定义:

$$\psi(\lambda) = \inf_{x\in Q} \mathcal{L}(x,\lambda)$$

$$\mathrm{dom}\,\psi = \{\lambda\in\mathbb{R}^m : \psi(\lambda) > -\infty\}$$

$$X^*(\lambda) = \mathrm{Arg}\inf_{x\in Q} \mathcal{L}(x,\lambda)$$

假设：对 $\forall \lambda \in \mathrm{dom}\psi \cap \mathbb{R}_+^m \neq \emptyset$ 时，有 $X^*(\lambda) \neq \emptyset$

有如下 Lagrangian dual problem：

$$f_* = \sup_{\lambda} \{ \psi(\lambda) : \lambda \in \mathrm{dom}\psi \cap \mathbb{R}_+^m \} \leq f^*$$

注意到：对 $\forall \lambda_1, \lambda_2 \in \mathrm{dom}\psi$，$\forall x_1 \in X^*(\lambda_1)$，$\forall x_2 \in X^*(\lambda_2)$

$$\psi(\lambda_2) = f_0(x_2) + \sum_{j=1}^{m} \lambda_2^{(j)} f_j(x_2)$$

$$\leq f_0(x_1) + \sum_{j=1}^{m} \lambda_2^{(j)} f_j(x_1)$$

$$= \psi(\lambda_1) + \langle f(x_1), \lambda_2 - \lambda_1 \rangle$$

问题在于 $\nabla \psi(\lambda) \Big|_{\lambda=\lambda_1} = f(x_1)$？

想证明 $\psi(\lambda)$ 是凹的是简单的，只需用定义

$$\psi(t\lambda_1 + (1-t)\lambda_2) = \inf \{ f_0(x) + \langle t\lambda_1 + (1-t)\lambda_2, f(x) \rangle \}$$

$$= \inf \{ t f_0(x) + t\langle \lambda_1, f(x) \rangle + (1-t) f_0(x) + (1-t)\langle \lambda_2, f(x) \rangle \}$$

$$\geq t \inf \{ f_0(x) + \langle \lambda_1, f(x) \rangle \} + (1-t) \inf \{ f_0(x) + \langle \lambda_2, f(x) \rangle \}$$

$$= t\psi(\lambda_1) + (1-t)\psi(\lambda_2)$$

$\psi(\lambda)$一定可导吗？不一定

引理：设 Q 紧，令 $\bar{\lambda} \in \mathbb{R}^m$，设 $X^*(\bar{\lambda}) = \{\bar{x}\}$ 是单点集

则若 $\lambda_k \to \bar{\lambda}$，则设 $x_k \in X^*(\lambda_k)$，有 $x_k \to \bar{x}$

证明：用反证法，设 $\lambda_k \to \bar{\lambda}$，$x_k \in X^*(\lambda_k)$，$\exists$ 指标集 N_0，

s.t. $\|x_k - \bar{x}\| > \varepsilon > 0$，对 $\forall k \in N_0$

由 Q 紧，故 $\exists N_1 \subset N_0$，s.t. $\{x_k\}_{N_1} \longrightarrow y \in Q$，且 $\|y - \bar{x}\| \geq \varepsilon$

对 $\forall \lambda_k$，$k \in N_1$，有

$f_0(x_k) + \langle \lambda_k, f(x_k) \rangle \leq f_0(\bar{x}) + \langle \lambda_k, f(\bar{x}) \rangle$

两边对 $k \to +\infty$，$k \in N_1$，有

$$f_0(y) + \langle \bar{\lambda}, f(y) \rangle \leq f_0(\bar{x}) + \langle \bar{\lambda}, f(\bar{x}) \rangle$$

故 $y \in X^*(\bar{x})$，与 $X^*(\bar{x})$ 是单点集矛盾

$\boxtimes$

定理：上引理条件下，ψ 在 $\bar{\lambda}$ 处可微，且 $\nabla\psi(\lambda)\big|_{\lambda=\bar{\lambda}} = f(\bar{x})$

证明：由 f_0，f 连续，则 $L(x,\lambda)$ 连续，且 Ω 紧，故对

$\forall \lambda \in \mathbb{R}^m$，$\exists x_\lambda \in X^*(\lambda)$，显然

$$\psi(\lambda)-\psi(\bar{\lambda}) = \inf_{x\in\Omega}\{f_0(x)+\langle\lambda, f(x)\rangle\} - \inf_{x\in\Omega}\{f_0(x)+\langle\bar{\lambda}, f(x)\rangle\}$$

$$\leq \inf_{x\in\Omega}\{\langle\lambda-\bar{\lambda}, f(x)\rangle\} \leq \langle\lambda-\bar{\lambda}, f(\bar{x})\rangle$$

$$\psi(\bar{\lambda})-\psi(\lambda) \leq \inf_{x\in\Omega}\{\langle\bar{\lambda}-\lambda, f(x)\rangle\} \leq \langle\bar{\lambda}-\lambda, f(x_\lambda)\rangle$$

故
$$0 \geq \psi(\lambda)-\psi(\bar{\lambda})-\langle f(\bar{x}), \lambda-\bar{\lambda}\rangle \geq \langle\lambda-\bar{\lambda}, f(x_\lambda)-f(\bar{x})\rangle \geq -\|\lambda-\bar{\lambda}\|\,\|f(x_\lambda)-f(\bar{x})\|$$

故

$$0 \geqslant \frac{\psi(\lambda)-\psi(\bar{x})-\langle f(\bar{x}), \lambda-\bar{x}\rangle}{\|\lambda-\bar{x}\|} \geqslant -\|f(x_\lambda)-f(\bar{x})\|$$

令 $\lambda \to \bar{x}$，则 $x_\lambda \to \bar{x}$，故 $f(x_\lambda) \to f(\bar{x})$

故 $$\lim_{\lambda \to \bar{x}} \frac{\psi(\lambda)-\psi(\bar{x})-\langle f(\bar{x}), \lambda-\bar{x}\rangle}{\|\lambda-\bar{x}\|} = 0$$

故 $\nabla \psi(\lambda)\big|_{\lambda=\bar{x}} = f(\bar{x})$

证

定理 1.3.2 (Certificate of Global Optimality)

令 λ^* 是 (1.3.8) 的最优解，设 $\exists\, \varepsilon>0$，有

$$\Delta_\varepsilon^+(\lambda^*) = \{\lambda \in \mathbb{R}_+^m : \|\lambda-\lambda^*\| \leq \varepsilon\} \subseteq \operatorname{dom}\psi$$

令 $x(\lambda) \in X^*(\lambda)$，$\lambda \neq \lambda^*$ 是唯一确定的，且极限：

$$x^* = \lim_{\substack{\lambda\to\lambda^* \\ \lambda\in\Delta_\varepsilon^+(\lambda^*)}} x(\lambda)$$

是存在的，则若 $x^* \in X^*(\lambda^*)$，则 x^* 是 (1.3.5) 的一个全局 optimal

证明：

令 $g(\lambda)=f(x(\lambda))$，令 $I^*=\{j : \lambda_*^{(j)}>0\}$，选取 $j\in I^*$，$\varepsilon>0$，

s.t. $\lambda_* \pm \varepsilon e_j \in \operatorname{dom}\psi \cap \mathbb{R}_+^m$，则有：

$$\psi(\lambda_*) \leq \psi(\lambda_*+\varepsilon e_j) + \langle g(\lambda_*+\varepsilon e_j), -\varepsilon e_j\rangle$$

$$\leq \psi(\lambda_*) + \langle g(\lambda_*+\varepsilon e_j), -\varepsilon e_j\rangle$$

$$\psi(\lambda_*) \leq \psi(\lambda_*-\varepsilon e_j) + \langle g(\lambda_*-\varepsilon e_j), \varepsilon e_j\rangle$$

$$\leq \psi(\lambda_*) + \langle g(\lambda_*-\varepsilon e_j), \varepsilon e_j\rangle$$

故 $\langle g(\lambda_*+\varepsilon e_j), e_j\rangle \leq 0 \leq \langle g(\lambda_*-\varepsilon e_j), e_j\rangle$

则令 $\varepsilon\to 0$ 得

$$\langle f(x^*), e_j\rangle \leq 0 \leq \langle f(x^*), e_j\rangle$$

故 $f_j(x^*)=0$

同样地，若 $j\notin I^*$，可取 ε 充分小，s.t. $\lambda_*+\varepsilon e_j\in \mathrm{dom}\psi\cap\mathbb{R}_+^m$

则

$$\psi(\lambda_*)\leq\psi(\lambda_*+\varepsilon e_j)+\langle g(\lambda_*+\varepsilon e_j), -\varepsilon e_j\rangle$$

$$\leq \psi(\lambda_*)+\langle g(\lambda_*+\varepsilon e_j), -\varepsilon e_j\rangle$$

故 $\langle g(\lambda_*+\varepsilon e_j), e_j\rangle\leq 0$，取 $\varepsilon\to 0$ 得 $f_j(x^*)\leq 0$

故 x^* 属于 (1.3.5) 的可行集，且

$$\lambda_*^{(j)} f_j(x^*)=0 \quad , 1\leq j\leq m$$

故

$$f_0(x^*)=f_0(x^*)+\sum_{j=1}^m \lambda_*^{(j)} f_j(x^*)=\psi(\lambda_*)\leq f^*$$

故 $f_0(x^*)=f^*$

$\boxed{}$

例： $\Omega=\mathbb{R}^2$， $f_0(x)=\frac{1}{2}\|x-\bar{e}_2\|^2$， $f_1(x)=x^{(1)}-\frac{1}{2}x^{(2)2}$

$\bar{e}_2=(1,1)^T$，则：

$$\mathcal{L}(x,\lambda)=\frac{1}{2}\|x-\bar{e}_2\|^2+\lambda\left[x^{(1)}-\frac{1}{2}x^{(2)2}\right]$$

$$=\frac{1}{2}(x^{(1)}-1)^2+\frac{1}{2}(x^{(2)}-1)^2+\lambda x^{(1)}-\frac{\lambda}{2}x^{(2)2}$$

$$=\frac{1}{2}x^{(1)2}+(\lambda-1)x^{(1)}+\frac{1}{2}(1-\lambda)x^{(2)2}-x^{(2)}+1$$

故令 $\psi(\lambda)=\inf\limits_{x\in\mathbb{R}^2}\mathcal{L}(x,\lambda)$，则

$\mathrm{dom}\,\psi(\lambda)=(-\infty,1)$

对于固定 λ，$x(\lambda)$ 可以如下方法求得

$$\begin{cases} x^{(1)}(\lambda)-1+\lambda=0 \\ x^{(2)}(\lambda)-1-\lambda x^{(2)}(\lambda)=0 \end{cases}$$

$$\begin{pmatrix} x^{(1)}(\lambda) \\ x^{(2)}(\lambda) \end{pmatrix}=\begin{pmatrix} 1-\lambda \\ \frac{1}{1-\lambda} \end{pmatrix}$$ ，代入 $L(x,\lambda)$

$$\psi(\lambda)=\lambda-\frac{1}{2}\lambda^2-\frac{1}{2(1-\lambda)}+\frac{1}{2}$$

则 $\lambda_* = \underset{\lambda \in (0,1)}{\arg\max}\ \psi(\lambda) = 1 - (\frac{1}{2})^{\frac{1}{3}}$

由 $x(\lambda)$ 是唯一确定的，且在 $(-\infty, 1)$ 上连续.

故 $x^* = \lim_{\substack{\lambda \to \lambda^* \\ \lambda \in \Delta_\epsilon^+(\lambda_*)}} x(\lambda) = x(\lambda^*) \in X(\lambda^*)$

故由 1.3.2

$x(\lambda_*) = (2^{-\frac{1}{3}}, 2^{\frac{1}{3}})$ 是全局极小！

$\boxed{/}$

1.3.3.2 Penalty Functions

Penalty Function Method
0. Choose $x_0 \in Q$. Choose a sequence of penalty coefficients: $0 < t_k < t_{k+1}$ and $t_k \to \infty$.   **1.** k**th iteration** ($k \geq 0$).   Find $x_{k+1} = \arg\min_{x \in Q}\{f_0(x) + t_k \Phi(x)\}$ using x_k as starting point.

假设 x_{k+1} 是辅助 func 的全局极小，定义：

$\Psi_k(x) = f_0(x) + t_k \Phi(x)$

$$\Psi_k^* = \min_{x\in Q} \Psi_k(x) = \Psi_k(x_{k+1})$$

定理 1.3.3

设 $\exists \bar{t}>0$, s.t. $S=\{x\in\mathbb{R}^n \mid f_0(x)+\bar{t}\Phi(x)\leq f_0(x^*)\}$

是有界的，则：

$$\lim_{k\to+\infty} f_0(x_k)=f_0(x^*) \quad \lim_{k\to\infty}\Phi(x_k)=0$$

证明：

注意到：

$\Psi_k^* \leq \Psi_k(x^*)=f_0(x^*)$，且对 $\forall x\in Q$，有 $\Psi_{k+1}(x)\geq\Psi_k(x)$

故 $\Psi_{k+1}^*\geq\Psi_k^*$，故存在极限

$$\lim_{k\to\infty}\Psi_k^* \equiv \Psi^* \leq f_0(x^*)$$

若 $t_k>\bar{t}$，则

$$f_0(x_{k+1})+\bar{t}\Phi(x_{k+1}) \leq f_0(x_{k+1})+t_k\Phi(x_{k+1})=\Psi_k^*\leq f_0(x^*)$$

故对 k 充分大，有 $x_k \in S$，由 S 是列紧的，则 $\{x_k\}$ 有聚点集合 X^*，由 $t_k \to \infty$，故对 $\forall x^* \in X^*$，均有 $\Phi(x_*)=0$

故 $x_* \in \mathcal{F}$，又由 $f_0(x_*) \le f_0(x^*)$，故 $f_0(x_*) = f_0(x^*)$

$\square$

1.3.3.3 Barrier Function

Barrier Function Method

0. Choose $x_0 \in \mathcal{F}_0$ and a sequence of penalty coefficients:

$$0 < t_k < t_{k+1} \text{ and } t_k \to \infty.$$

1. ***k*th iteration** ($k \ge 0$).

Find $x_{k+1} = \arg\min_{x \in \mathcal{F}_0} \left\{ f_0(x) + \frac{1}{t_k} F(x) \right\}$ using x_k as the starting point.

定义：

$$\Psi_k(x) = f_0(x) + \frac{1}{t_k} F(x)$$

$$\Psi_k^* = \min_{x \in \mathcal{F}_0} \Psi_k(x)$$

定理 1.3.4

令 barrier $F(\cdot)$ 是在 $\mathcal{F}_0$ 上下方有界，则

$$\lim_{k \to \infty} \psi_k^* = f^*$$

证明：令 $F(x) \geqslant F^*$，对 $\forall x \in \mathcal{F}_0$，故对 $\forall \bar{x} \in \mathcal{F}_0$，有

$$\limsup_{k\to\infty} \psi_k^* \leqslant \lim_{k\to\infty} \left[f_0(\bar{x}) + \frac{1}{t_k} F(\bar{x}) \right] = f_0(\bar{x})$$

故 $\limsup_{k\to\infty} \psi_k^* \leqslant f^*$，另一方面：

$$\Psi_k^* = \min_{x\in\mathcal{F}_0} \left\{ f_0(x) + \frac{1}{t_k} F(x) \right\}$$

$$\geqslant \inf_{x\in\mathcal{F}_0} \left\{ f_0(x) + \frac{1}{t_k} F^* \right\}$$

$$= f^* + \frac{1}{t_k} F^*$$

故 $\liminf_{k\to\infty} \Psi_k^* \geqslant f^*$

故 $f^* \leqslant \liminf_{k\to\infty} \Psi_k^* \leqslant \limsup_{k\to+\infty} \Psi_k^* \leqslant f^*$

故 $\lim_{k\to\infty} \Psi_k^* = f^*$

$\square$

1.3.3 Constrained Minimization

$$\min_{x \in Q} f_0(x)$$

$$f_j(x) \leq 0 \ , j=1,\cdots,m$$

$Q \subset \mathbb{R}^n$是闭的，$f_0,\cdots,f_m$连续

1.3.3.1 Lagrangian Relaxation

定理 1.3.1

令$F(x,\lambda)$定义在$x \in Q_1 \subseteq \mathbb{R}^n$，$\lambda \in Q_2 \in \mathbb{R}^m$，$Q_1, Q_2 \neq \varnothing$，则

$$\sup_{\lambda \in Q_2} \inf_{x \in Q_1} F(x,\lambda) \leq \inf_{x \in Q_1} \sup_{\lambda \in Q_2} F(x,\lambda)$$

证明：对$\forall x \in Q_1$，$\lambda \in Q_2$

$$F(x,\lambda) \leq \sup_{\xi \in Q_2} F(x,\xi)$$

则 $$\inf_{x \in Q_1} F(x,\lambda) \leq \inf_{x \in Q_1} \sup_{\xi \in Q_2} F(x,\xi) \quad 对 \forall \lambda \in Q_2$$

故 $$\sup_{\lambda \in Q_2} \inf_{x \in Q_1} F(x,\lambda) \leq \inf_{x \in Q_1} \sup_{\lambda \in Q_2} F(x,\lambda)$$ $\boxtimes$

注意到

$$f^* = \inf_{x\in Q}\{f_0(x) : f_j(x)\leqslant 0,\ j=1,\cdots,m\}$$

$$= \inf_{x\in Q}\sup_{\lambda\in\mathbb{R}_+^m} \mathcal{L}(x,\lambda)$$

其中 $\mathbb{R}_+^m = \{\lambda\in\mathbb{R}^m, \lambda^{(j)}\geqslant 0,\ j=1,\cdots,m\}$

$\mathcal{L}(x,\lambda) = f_0(x) + \langle\lambda, f(x)\rangle$

则显然

$$\sup_{\lambda\in\mathbb{R}_+^m}\mathcal{L}(x,\lambda) = \begin{cases} +\infty & ,\ x\notin\Omega \\ f_0(x) & ,\ x\in\Omega \end{cases}$$ ，Ω 是可行集

定义:

$$\psi(\lambda) = \inf_{x\in Q}\mathcal{L}(x,\lambda)$$

$$\mathrm{dom}\,\psi = \{\lambda\in\mathbb{R}^m : \psi(\lambda) > -\infty\}$$

$$X^*(\lambda) = \mathrm{Arg}\inf_{x\in Q}\mathcal{L}(x,\lambda)$$

假设：对 $\forall \lambda \in \mathrm{dom}\psi \cap \mathbb{R}_+^m \neq \phi$ 时，有 $X^*(\lambda) \neq \phi$

有如下 Lagrangian dual problem：

$$f_* = \sup_{\lambda} \{ \psi(\lambda) : \lambda \in \mathrm{dom}\psi \cap \mathbb{R}_+^m \} \leq f^*$$

注意到：对 $\forall \lambda_1, \lambda_2 \in \mathrm{dom}\psi$，$\forall x_1 \in X^*(\lambda_1)$，$\forall x_2 \in X^*(\lambda_2)$

$$\begin{aligned}
\psi(\lambda_2) &= f_0(x_2) + \sum_{j=1}^m \lambda_2^{(j)} f_j(x_2) \\
&\leq f_0(x_1) + \sum_{j=1}^m \lambda_2^{(j)} f_j(x_1) \\
&= \psi(\lambda_1) + \langle f(x_1), \lambda_2 - \lambda_1 \rangle
\end{aligned}$$

问题在于 $\nabla \psi(\lambda) \Big|_{\lambda=\lambda_1} = f(x_1)$？

想证明 $\psi(\lambda)$ 是凹的是简单的，只需用定义

$$\psi(t\lambda_1 + (1-t)\lambda_2) = \inf \{ f_0(x) + \langle t\lambda_1 + (1-t)\lambda_2, f(x) \rangle \}$$

$$= \inf \{ t f_0(x) + t\langle \lambda_1, f(x) \rangle + (1-t) f_0(x) + (1-t)\langle \lambda_2, f(x) \rangle \}$$

$$\geq t \inf \{ f_0(x) + \langle \lambda_1, f(x) \rangle \} + (1-t) \inf \{ f_0(x) + \langle \lambda_2, f(x) \rangle \}$$

$$= t\psi(\lambda_1) + (1-t)\psi(\lambda_2)$$

$\psi(\lambda)$一定可导吗? 不一定

引理: 设 Q 紧, 令 $\bar{\lambda} \in \mathbb{R}^m$, 设 $X^*(\bar{\lambda}) = \{\bar{x}\}$ 是单点集

则若 $\lambda_k \to \bar{\lambda}$, 则设 $x_k \in X^*(\lambda_k)$, 有 $x_k \to \bar{x}$

证明: 用反证法, 设 $\lambda_k \to \bar{\lambda}$, $x_k \in X^*(\lambda_k)$, $\exists$ 指标集 N_0,

s.t. $\|x_k - \bar{x}\| > \varepsilon > 0$, 对 $\forall k \in N_0$

由 Q 紧, 故 $\exists N_1 \subset N_0$, s.t. $\{x_k\}_{N_1} \to y \in Q$, 且 $\|y - \bar{x}\| \geq \varepsilon$

对 $\forall \lambda_k$, $k \in N_1$, 有

$$f_0(x_k) + \langle \lambda_k, f(x_k) \rangle \leq f_0(\bar{x}) + \langle \lambda_k, f(\bar{x}) \rangle$$

两边对 $k \to +\infty$, $k \in N_1$, 有

$$f_0(y) + \langle \bar{\lambda}, f(y) \rangle \leq f_0(\bar{x}) + \langle \bar{\lambda}, f(\bar{x}) \rangle$$

故 $y \in X^*(\bar{\lambda})$, 与 $X^*(\bar{\lambda})$ 是单点集矛盾

$\square$

定理：上引理条件下，ψ 在 $\bar{\lambda}$ 处可微，且 $\nabla\psi(\lambda)\big|_{\lambda=\bar{\lambda}} = f(\bar{x})$

证明：由 f_0，f 连续，则 $L(x,\lambda)$ 连续，且 Ω 紧，故对

$\forall \lambda \in \mathbb{R}^m$，$\exists x_\lambda \in X^*(\lambda)$，显然

$$\psi(\lambda)-\psi(\bar{\lambda}) = \inf_{x\in\Omega}\{f_0(x)+\langle\lambda, f(x)\rangle\} - \inf_{x\in\Omega}\{f_0(x)+\langle\bar{\lambda}, f(x)\rangle\}$$

$$\leq \inf_{x\in\Omega}\{\langle\lambda-\bar{\lambda}, f(x)\rangle\} \leq \langle\lambda-\bar{\lambda}, f(\bar{x})\rangle$$

$$\psi(\bar{\lambda})-\psi(\lambda) \leq \inf_{x\in\Omega}\{\langle\bar{\lambda}-\lambda, f(x)\rangle\} \leq \langle\bar{\lambda}-\lambda, f(x_\lambda)\rangle$$

故 $$0 \geq \psi(\lambda)-\psi(\bar{\lambda})-\langle f(\bar{x}), \lambda-\bar{\lambda}\rangle$$

$$\geq \langle\lambda-\bar{\lambda}, f(x_\lambda)-f(\bar{x})\rangle$$

$$\geq -\|\lambda-\bar{\lambda}\|\,\|f(x_\lambda)-f(\bar{x})\|$$

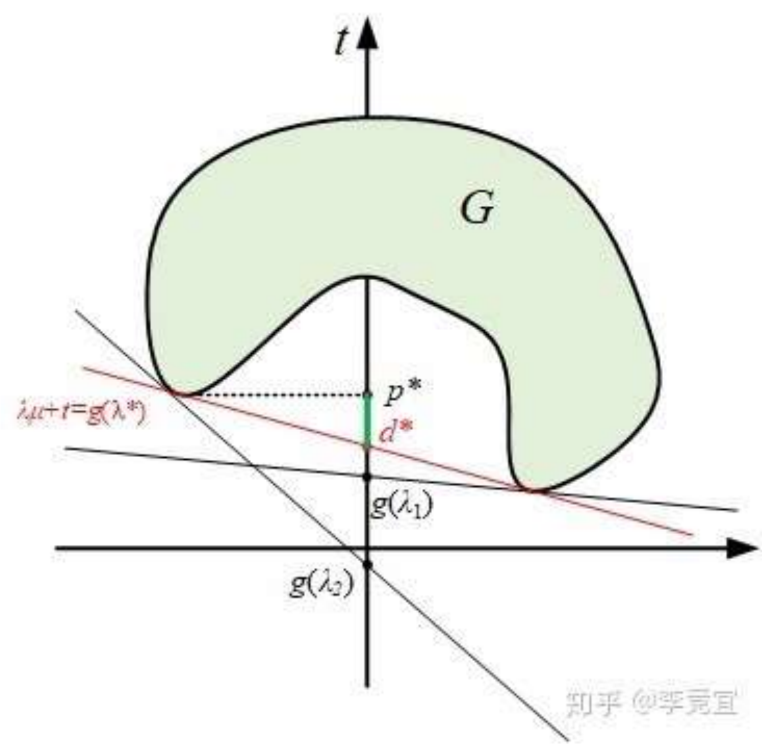
故

$$0 \geq \frac{\psi(\lambda)-\psi(\bar{x})-\langle f(\bar{x}),\lambda-\bar{x}\rangle}{\|\lambda-\bar{x}\|} \geq -\|f(x_\lambda)-f(\bar{x})\|$$

令 $\lambda\to\bar{x}$，则 $x_\lambda\to\bar{x}$，故 $f(x_\lambda)\to f(\bar{x})$

故 $$\lim_{\lambda\to\bar{x}} \frac{\psi(\lambda)-\psi(\bar{x})-\langle f(\bar{x}),\lambda-\bar{x}\rangle}{\|\lambda-\bar{x}\|} = 0$$

故 $\nabla\psi(\lambda)\big|_{\lambda=\bar{x}} = f(\bar{x})$ $\boxed{\text{证}}$



p^*-d^* 对偶间隙

定理1.3.2（Certificate of Global Optimality）

令λ^*是(1.3.8)的最优解，设$\exists\varepsilon>0$，有

$$\Delta_\varepsilon^+(\lambda^*)=\{\lambda\in\mathbb{R}_+^m:\|\lambda-\lambda^*\|\leq\varepsilon\}\subseteq \mathrm{dom}\,\psi$$

令$x(\lambda)\in X^*(\lambda)$，$\lambda\neq\lambda^*$是唯一确定的，且极限：

$$x^*=\lim_{\substack{\lambda\to\lambda^*\\ \lambda\in\Delta_\varepsilon^+(\lambda^*)}} x(\lambda)$$

是存在的，则若$x^*\in X^*(\lambda^*)$，则x^*是(1.3.5)的一个全局optimal

证明：

令$g(\lambda)=f(x(\lambda))$，令$I^*=\{j:\lambda_*^{(j)}>0\}$，选取$j\in I^*$，$\varepsilon>0$，

s.t. $\lambda_*\pm\varepsilon e_j\in\mathrm{dom}\,\psi\cap\mathbb{R}_+^m$，则有：

$$\psi(\lambda_*)\leq\psi(\lambda_*+\varepsilon e_j)+\langle g(\lambda_*+\varepsilon e_j),-\varepsilon e_j\rangle$$

$$\leq\psi(\lambda_*)+\langle g(\lambda_*+\varepsilon e_j),-\varepsilon e_j\rangle$$

$$\psi(\lambda_*)\leq\psi(\lambda_*-\varepsilon e_j)+\langle g(\lambda_*-\varepsilon e_j),\varepsilon e_j\rangle$$

$$\leq\psi(\lambda_*)+\langle g(\lambda_*-\varepsilon e_j),\varepsilon e_j\rangle$$

故 $\langle g(\lambda_*+\varepsilon e_j), e_j\rangle \le 0 \le \langle g(\lambda_*-\varepsilon e_j), e_j\rangle$

则令 $\varepsilon\to 0$ 得

$$\langle f(x^*), e_j\rangle \le 0 \le \langle f(x^*), e_j\rangle$$

故 $f_j(x^*)=0$

同样地，若 $j\notin I^*$，可取 ε 充分小，s.t. $\lambda_*+\varepsilon e_j\in \mathrm{dom}\,\psi\cap \mathbb{R}_+^m$

则

$$\psi(\lambda_*)\le \psi(\lambda_*+\varepsilon e_j)+\langle g(\lambda_*+\varepsilon e_j), -\varepsilon e_j\rangle$$

$$\le \psi(\lambda_*)+\langle g(\lambda_*+\varepsilon e_j), -\varepsilon e_j\rangle$$

故 $\langle g(\lambda_*+\varepsilon e_j), e_j\rangle \le 0$，取 $\varepsilon\to 0$ 得 $f_j(x^*)\le 0$

故 x^* 属于 (1.3.5) 的可行集，且

$$\lambda_*^{(j)} f_j(x^*)=0 \quad , 1\le j\le m$$

故

$$f_0(x^*)=f_0(x^*)+\sum_{j=1}^m \lambda_*^{(j)} f_j(x^*)=\psi(\lambda_*)\le f^*$$

故 $f_0(x^*)=f^*$ $\square$

例：$\Omega=\mathbb{R}^2$，$f_0(x)=\frac{1}{2}\|x-\bar{e}_2\|^2$，$f_1(x)=x^{(1)}-\frac{1}{2}x^{(2)2}$

$\bar{e}_2=(1,1)^T$，则：

$$\mathcal{L}(x,\lambda)=\frac{1}{2}\|x-\bar{e}_2\|^2+\lambda\left[x^{(1)}-\frac{1}{2}x^{(2)2}\right]$$

$$=\frac{1}{2}(x^{(1)}-1)^2+\frac{1}{2}(x^{(2)}-1)^2+\lambda x^{(1)}-\frac{\lambda}{2}x^{(2)2}$$

$$=\frac{1}{2}x^{(1)2}+(\lambda-1)x^{(1)}+\frac{1}{2}(1-\lambda)x^{(2)2}-x^{(2)}+1$$

故令 $\psi(\lambda)=\inf\limits_{x\in\mathbb{R}^2}\mathcal{L}(x,\lambda)$，则

$\mathrm{dom}\,\psi(\lambda)=(-\infty,1)$

对于固定 λ，$x(\lambda)$可以如下方法求得

$$\begin{cases} x^{(1)}(\lambda)-1+\lambda=0 \\ x^{(2)}(\lambda)-1-\lambda x^{(2)}(\lambda)=0 \end{cases}$$

$$\begin{pmatrix} x^{(1)}(\lambda) \\ x^{(2)}(\lambda) \end{pmatrix}=\begin{pmatrix} 1-\lambda \\ \frac{1}{1-\lambda} \end{pmatrix}$$，代入 $L(x,\lambda)$

$$\psi(\lambda)=\lambda-\frac{1}{2}\lambda^2-\frac{1}{2(1-\lambda)}+\frac{1}{2}$$

则
$$\lambda_* = \arg\max_{\lambda\in(0,1)} \psi(\lambda) = 1-\left(\frac{1}{2}\right)^{\frac{1}{3}}$$

由 $x(\lambda)$ 是唯一确定的，且在 $(-\infty, 1)$ 上连续.

故
$$x^* = \lim_{\substack{\lambda\to\lambda^* \\ \lambda\in\Delta_\epsilon^+(\lambda_*)}} x(\lambda) = x(\lambda^*) \in X(\lambda^*)$$

故由 1.3.2

$$x(\lambda_*) = (2^{-\frac{1}{3}}, 2^{\frac{1}{3}})$$ 是全局极小！

$\boxtimes$

1.3.3.2 Penalty Functions

Penalty Function Method
0. Choose $x_0 \in Q$. Choose a sequence of penalty coefficients: $0 < t_k < t_{k+1}$ and $t_k \to \infty$.
1. ***k*th iteration** ($k \geq 0$). Find $x_{k+1} = \arg\min_{x\in Q}\{f_0(x) + t_k\Phi(x)\}$ using x_k as starting point.

假设 x_{k+1} 是辅助 func 的全局极小，定义：

$$\Psi_k(x) = f_0(x) + t_k\Phi(x)$$

$$\Psi_k^* = \min_{x \in Q} \Psi_k(x) = \Psi_k(x_{k+1})$$

定理 1.3.3

设 $\exists \varepsilon > 0$, s.t. $S = \{x \in \mathbb{R}^n \mid f_0(x) + \varepsilon \Phi(x) \leq f_0(x^*)\}$

是有界的，则：

$$\lim_{k \to +\infty} f_0(x_k) = f_0(x^*) \quad \lim_{k \to \infty} \Phi(x_k) = 0$$

证明：

注意到：

$\Psi_k^* \leq \Psi_k(x^*) = f_0(x^*)$，且对 $\forall x \in Q$，有 $\Psi_{k+1}(x) \geq \Psi_k(x)$

故 $\Psi_{k+1}^* \geq \Psi_k^*$，故存在极限

$$\lim_{k \to \infty} \Psi_k^* \equiv \Psi^* \leq f_0(x^*)$$

若 $t_k > \varepsilon$，则

$$f_0(x_{k+1}) + \varepsilon \Phi(x_{k+1}) \leq f_0(x_{k+1}) + t_k \Phi(x_{k+1}) = \Psi_k^* \leq f_0(x^*)$$

故对 k 充分大，有 $x_k \in S$，由 S 是列紧的，则 $\{x_k\}$ 有

聚点集合 X^*，由 $t_k \to \infty$，故对 $\forall x^* \in X^*$，均有 $\Phi(x_*)=0$

故 $x_* \in \mathcal{F}$，又由 $f_0(x_*) \leq f_0(x^*)$，故 $f_0(x_*) = f_0(x^*)$

$\square$

1.3.3.3 Barrier Function

令$\mathcal{F}$是$\mathbb{R}^n$中闭集，且有非空内部。连续func $F(\cdot)$

称为 barrier func，若 $F(x)\to\infty \ (x\to\partial\mathcal{F})$

Slater条件：$\exists \bar{x}\in\mathbb{R}^n$, s.t. $f_j(\bar{x})<0$

$j=1,\cdots,m$

Barrier Function Method

0. Choose $x_0 \in \mathcal{F}_0$ and a sequence of penalty coefficients:

$$0 < t_k < t_{k+1} \text{ and } t_k \to \infty.$$

1. ***k*th iteration** ($k \geq 0$).

Find $x_{k+1} = \arg\min\limits_{x\in\mathcal{F}_0} \left\{ f_0(x) + \frac{1}{t_k}F(x) \right\}$ using x_k as the starting point.

定义：

$$\Psi_k(x) = f_0(x) + \frac{1}{t_k}F(x)$$

$$\Psi_k^* = \min_{x\in\mathcal{F}_0} \Psi_k(x)$$

定理 1.3.4

令 barrier $F(\cdot)$是在$\mathcal{F}_0$上下方有界，则

$$\lim_{k\to\infty} \psi_k^* = f^*$$

证明：令 $F(x) \geq F^*$，对 $\forall x \in \mathcal{F}_0$，故对 $\forall \bar{x} \in \mathcal{F}_0$，有

$$\limsup_{k\to\infty} \Psi_k^* \leq \lim_{k\to\infty} \left[f_0(\bar{x}) + \frac{1}{t_k} F(\bar{x}) \right] = f_0(\bar{x})$$

故 $\limsup\limits_{k\to\infty} \Psi_k^* \leq f^*$，另一方面：

$$\Psi_k^* = \min_{x\in\mathcal{F}_0} \left\{ f_0(x) + \frac{1}{t_k} F(x) \right\}$$

$$\geq \inf_{x\in\mathcal{F}_0} \left\{ f_0(x) + \frac{1}{t_k} F^* \right\}$$

$$= f^* + \frac{1}{t_k} F^*$$

故 $\liminf\limits_{k\to\infty} \Psi_k^* \geq f^*$

故 $f^* \leq \liminf\limits_{k\to\infty} \Psi_k^* \leq \limsup\limits_{k\to+\infty} \Psi_k^* \leq f^*$

故 $\lim\limits_{k\to\infty} \Psi_k^* = f^*$

$\square$

常见的 Barrier func 见例 1.3.4，如：

Log-Barrier：$F(x) = -\sum\limits_{j=1}^{m} \ln(-f_j(x))$

故 barrier func 方法实际上找 $\mathcal{F}_0 = Q \cap \tilde{\mathcal{F}}$ 中的极小

$\tilde{\mathcal{F}} = \{x : f_j(x) < 0, j = 1, 2, \cdots, m\}$

但 $\tilde{\mathcal{F}}$ 与 $\text{int}\,\mathcal{F}$ 不一定相等，例：

$$f_1(x) = \begin{cases} -e^{-\frac{1}{x}}, & x > 0 \\ 0 & x \le 0 \end{cases}$$

$f_1 \in C^\infty$

$$\boxed{\begin{aligned} &\min_{x \in \mathbb{R}} \ (x+2)^2 \\ &\text{s.t.} \ \ f_1(x) \le 0 \end{aligned}}$$

实际：$x=-2$ 时，$f_{min}=0$

barrier 法：下确界是 4

§ 2.1.1 Smooth Convex function

(可微的)

想建立一个函数类 $\mathcal{F}$，满足：

① 一阶最优性条件 ⇒ 全局最优

② 若 $f_1, f_2 \in \mathcal{F}$，则 $\alpha f_1 + \beta f_2 \in \mathcal{F}$，$\alpha, \beta \ge 0$

③ 线性函数类 $\subset \mathcal{F}$

设 $f\in\mathcal{F}$，取 $x_0\in\mathbb{R}^n$，考虑：

$$\phi(y)=f(y)-\underbrace{\langle\nabla f(x_0),y\rangle}_{\text{线性func}}\in\mathcal{F}$$

故 $\nabla\phi(y)\big|_{y=x_0}=\nabla f(x_0)-\nabla f(x_0)=0$

故 $\phi(y)\geqslant\phi(x_0)=f(x_0)-\langle\nabla f(x_0),x_0\rangle$

$\Leftrightarrow f(y)\geqslant f(x_0)+\langle\nabla f(x_0),y-x_0\rangle$

（凸func的一阶条件）

故 $f\in$ 函数类 $\mathcal{F}$ $\Rightarrow f\in$ 凸函数，下证反过来成立

定理2.1.1：若 $f\in\mathcal{F}^1(\mathbb{R}^n)$，且 $\nabla f(x^*)=0$，则 x^* 是 $f(\cdot)$ 在 $\mathbb{R}^n$ 上全局极小

证：$f(x)\geqslant f(x^*)+\langle\nabla f(x^*),x-x^*\rangle=f(x^*)$ $\boxed{\diagup}$

引理2.1.1：若 $f_1,f_2\in\mathcal{F}^1(Q)$，$\alpha,\beta\geqslant0$，则 $f=\alpha f_1+\beta f_2\in\mathcal{F}^1(Q)$

证明显然

以下建立凸func几个等价定义(光滑性足够)，f定义在凸集Q上：

1) $\forall x, y \in Q$, $\forall \alpha \in [0,1]$，有

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

最原始，f不需可微

2) $\forall x, y \in Q$, $f(y) \geq f(x)+\langle \nabla f(x), y-x\rangle$

3) $\forall x, y \in Q$ $\langle \nabla f(x)-\nabla f(y), x-y\rangle \geq 0$

4) 若f二次可微，$\forall x \in Q$ $\nabla^2 f(x) \succeq 0$

5) $\forall x \in Q, \forall v \in \mathbb{R}^n$, $g(t)=f(x+tv)$ 凸

证明：

2) $\Rightarrow$ 1) 令 $x_\alpha=\alpha x+(1-\alpha)y$

则 $f(x_\alpha) \leq f(y)-\langle \nabla f(x_\alpha), y-x_\alpha\rangle$

$$=f(y)-\alpha\langle \nabla f(x_\alpha), y-x\rangle \quad \cdots\cdots ①$$

$f(x_\alpha) \leq f(x)-\langle \nabla f(x_\alpha), x-x_\alpha\rangle$

$$=f(x)+(1-\alpha)\langle \nabla f(x_\alpha), y-x\rangle \quad \cdots\cdots ②$$

$(1-\alpha)\times$ ① $+\ \alpha\times$ ② 即证 1)

1) $\Rightarrow$ 2) 令 $x_\alpha=\alpha x+(1-\alpha)y$，取 $\alpha\in[0,1]$

$$f(y)\geq\frac{1}{1-\alpha}[f(x_\alpha)-\alpha f(x)]$$
$$=f(x)+\frac{1}{1-\alpha}[f(x_\alpha)-f(x)]$$
$$=f(x)+\frac{1}{1-\alpha}[f(x+(1-\alpha)(y-x))-f(x)]$$

令 $\alpha\to1$，则 $f(y)\geq f(x)+\langle\nabla f(x),y-x\rangle$

2) $\Rightarrow$ 3) $f(x)\geq f(y)+\langle\nabla f(y),x-y\rangle$ …… ③

$f(y)\geq f(x)+\langle\nabla f(x),y-x\rangle$ …… ④

③ + ④ 即证 3)

3) $\Rightarrow$ 2) 令 $x_\tau=x+\tau(y-x)\in Q$，则

$$f(y)=f(x)+\int_0^1\langle\nabla f(x+\tau(y-x)),y-x\rangle d\tau$$
$$=f(x)+\langle\nabla f(x),y-x\rangle+\int_0^1\langle\nabla f(x_\tau)-\nabla f(x),y-x\rangle d\tau$$
$$=f(x)+\langle\nabla f(x),y-x\rangle+\int_0^1\frac{1}{\tau}\langle\nabla f(x_\tau)-\nabla f(x),x_\tau-x\rangle d\tau$$
$$\geq f(x)+\langle\nabla f(x),y-x\rangle$$

3) $\Rightarrow$ 4) 令 $s \in \mathbb{R}^n$, $x_\tau = x + \tau s \in Q$, $\tau > 0$, 则

$$0 \le \frac{1}{\tau^2} \langle \nabla f(x_\tau) - \nabla f(x), x_\tau - x \rangle$$

$$= \frac{1}{\tau} \langle \nabla f(x_\tau) - \nabla f(x), s \rangle$$

$$= \frac{1}{\tau} \int_0^\tau \langle \nabla^2 f(x + \lambda s) s, s \rangle \, d\lambda$$

令 $\tau \to 0^+$, 由 $s \in \mathbb{R}^n$ 的任意性知: $\nabla^2 f(x) \succeq 0$

4) $\Rightarrow$ 2) 对 $\forall y \in Q$

$$f(y) = f(x) + \langle \nabla f(x), y - x \rangle + \int_0^1 \int_0^\tau \langle \nabla^2 f(x + \lambda(y-x))(y-x), y-x \rangle \, d\lambda \, d\tau$$

$$\geqslant f(x) + \langle \nabla f(x), y - x \rangle$$

1) $\Rightarrow$ 5) 取 $\forall \alpha \in [0,1]$

$$g(\alpha t_1 + (1-\alpha) t_2) = f(x + (\alpha t_1 + (1-\alpha) t_2) v)$$

$$= f(x + \alpha t_1 v + (1-\alpha) t_2 v)$$

$$= f(\alpha(x + t_1 v) + (1-\alpha)(x + t_2 v))$$

$$\leqslant \alpha g(t_1) + (1-\alpha) g(t_2)$$

5) $\Rightarrow$ 1) 取 $t \in [0,1]$

$$f(x+tv)=g(t)\le t\,g(1)+(1-t)g(0)$$
$$=t\,f(x+v)+(1-t)f(x)$$

取 $v=y-x$，$y\in\mathbb{R}^n$ 即证

$\boxed{\checkmark}$

例：$f(x)=|x|^p$ 凸

证：只需证明：$|w_1x+w_2y|\le(w_1|x|^p+w_2|y|^p)^{\frac{1}{p}}$，$w_1+w_2=1$

$$|w_1x+w_2y|\le w_1|x|+w_2|y| \quad \cdots\cdots (*)$$
$$=\left(w_1^{\frac{1}{p}}|x|\right)w_1^{\frac{1}{q}}+\left(w_2^{\frac{1}{p}}|y|\right)w_2^{\frac{1}{q}}$$
(Hölder) $$\le(w_1|x|^p+w_2|y|^p)^{\frac{1}{p}}(w_1+w_2)^{\frac{1}{q}}$$
$$=(w_1|x|^p+w_2|y|^p)^{\frac{1}{p}}$$

其中 $\frac{1}{p}+\frac{1}{q}=1$，事实上，$|x|^p\ (p>1)$ 严格 convex

只须验 Hölder 不等式等号成立条件：

$$\begin{cases} w_1^{\frac{1}{q}}=c\,w_1^{\frac{p-1}{p}}|x|^{p-1} \\ w_2^{\frac{1}{q}}=c\,w_2^{\frac{p-1}{p}}|y|^{p-1} \end{cases} \quad \cdots\cdots ①$$

由 $\frac{1}{q}=\frac{p-1}{p}$，则 ① $\Leftrightarrow$ $\begin{cases} 1=c|x|^{p-1} \\ 1=c|y|^{p-1} \end{cases}$

若 x,y 是异号的，则 ⊛ 是严格的 $\Rightarrow$ 严格 convex

若 x,y 是同号的，则 $|x|=|y|=\frac{1}{c} \Leftrightarrow x=y \Rightarrow$ 严格 convex

$\boxed{/}$

定义算子范数

有限维空间均可达范

$$\|g\|_* = \max_{x\in\mathbb{R}^n}\{\langle g,x\rangle : \|x\|\leq 1\}$$

即 $x\in(\mathbb{R}^n,\|\cdot\|)\doteq X$，则 $g\in X^*$

定义 $f\in\mathcal{F}_L^{1,1}(Q,\|\cdot\|)$，则 $Q\subseteq \mathrm{dom}f$，且：

$$\|\nabla f(x)-\nabla f(y)\|_*\leq L\|x-y\|,\ \forall x,y\in Q$$

注：$\left(t^p\right)^* = t^q \quad (1\leq p<\infty)$，特别地 $(t^2)^*=t^2$

等价意义

引理2.1.4：对 $\forall x,y\in\mathbb{R}^n$，$\alpha\in[0,1]$，有：

$$\alpha\|x\|^2+(1-\alpha)\|y\|^2\geqslant \alpha(1-\alpha)(\|x\|+\|y\|)^2\geqslant \alpha(1-\alpha)\|x-y\|^2$$

证明显然

Theorem 2.1.5 *All conditions below, holding for all $x, y \in \mathbb{R}^n$ and α from $[0, 1]$, are equivalent to the inclusion $f \in \mathscr{F}_L^{1,1}(\mathbb{R}^n, \| \cdot \|)$:*

$$0 \leq f(y) - f(x) - \langle \nabla f(x), y - x \rangle \leq \frac{L}{2} \| x - y \|^2, \tag{2.1.9}$$

$$f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2L} \| \nabla f(x) - \nabla f(y) \|_*^2 \leq f(y), \tag{2.1.10}$$

$$\frac{1}{L} \| \nabla f(x) - \nabla f(y) \|_*^2 \leq \langle \nabla f(x) - \nabla f(y), x - y \rangle, \tag{2.1.11}$$

$$0 \leq \langle \nabla f(x) - \nabla f(y), x - y \rangle \leq L \| x - y \|^2, \tag{2.1.12}$$

$$\begin{aligned} \alpha f(x) + (1-\alpha) f(y) &\geq f(\alpha x + (1-\alpha) y) \\ &\quad + \frac{\alpha(1-\alpha)}{2L} \| \nabla f(x) - \nabla f(y) \|_*^2, \end{aligned} \tag{2.1.13}$$

$$\begin{aligned} 0 &\leq \alpha f(x) + (1-\alpha) f(y) - f(\alpha x + (1-\alpha) y) \\ &\leq \alpha(1-\alpha) \frac{L}{2} \| x - y \|^2 . \end{aligned} \tag{2.1.14}$$

Moreover, if $f \in \mathscr{F}_L^{1,1}(Q)$, then inequalities (2.1.9), (2.1.12), and (2.1.14) are valid for all $x, y \in Q$.

证明:

$f \in \mathcal{F}_L^{1,1}(\mathbb{R}^n, \|\cdot\|) \Rightarrow (2.1.9)$

$0 \leq f(y) - f(x) - \langle \nabla f(x), y-x \rangle$ 显然

$f(y) - f(x) - \langle \nabla f(x), y-x \rangle$

$= \int_0^1 \langle \nabla f(x+\tau(y-x)) - \nabla f(x), y-x \rangle \, d\tau$

$\leq \int_0^1 \| \nabla f(x+\tau(y-x)) - \nabla f(x) \|_* \| y-x \| \, d\tau$

$\leq \int_0^1 L\tau \| y-x \|^2 d\tau = \frac{L}{2} \| y-x \|^2$

(2.1.9) $\Rightarrow$ (2.1.10)

固定 $x_0 \in \mathbb{R}^n$，考虑 $\phi(y) = f(y) - \langle \nabla f(x_0), y \rangle$

则 $\phi \in \mathcal{F}_L^{1,1}(\mathbb{R}^n, \|\cdot\|)$，且 ϕ 的 optimal 是 x_0

故：

$$\phi(x_0) = \min_{x \in \mathbb{R}^n} \phi(x)$$

$$\overset{(2.1.9)}{\leq} \min_{x \in \mathbb{R}^n} \left\{ \phi(y) + \langle \nabla \phi(y), x - y \rangle + \frac{L}{2} \|x - y\|^2 \right\}$$

$$\overset{r = \|x-y\|}{=} \min_{r \geq 0} \left\{ \phi(y) - r \|\nabla \phi(y)\|_* + \frac{L}{2} r^2 \right\}$$

注：$\langle \nabla \phi(y), x - y \rangle \geq -\|\nabla \phi(y)\|_* \|x - y\|$

且由 $x \in \mathbb{R}^n$，故 $\exists \tilde{x}$，s.t. 上式取等号

$$= \phi(y) - \frac{1}{2L} \|\nabla \phi(y)\|_*^2$$

$$= \phi(y) - \frac{1}{2L} \|\nabla f(y) - \nabla f(x_0)\|_*^2$$

(2.1.10) $\Rightarrow$ (2.1.11)

$$\begin{cases} f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2L} \|\nabla f(x) - \nabla f(y)\|_*^2 \leq f(y) & \cdots\cdots ① \\ f(y) + \langle \nabla f(y), x - y \rangle + \frac{1}{2L} \|\nabla f(y) - \nabla f(x)\|_*^2 \leq f(x) & \cdots\cdots ② \end{cases}$$

① + ② $\Rightarrow$ (2.1.11)

(2.1.11) $\Rightarrow$ 原始定义：由 Cauchy 不等式

(2.1.9) $\Rightarrow$ (2.1.12)

$$\begin{cases} 0 \le f(y) - f(x) - \langle \nabla f(x), y-x \rangle \le \frac{L}{2}\|x-y\|^2 & \cdots ③ \\ 0 \le f(x) - f(y) - \langle \nabla f(y), x-y \rangle \le \frac{L}{2}\|x-y\|^2 & \cdots ④ \end{cases}$$

③ + ④ $\Rightarrow$ (2.1.12)

(2.1.12) $\Rightarrow$ (2.1.9)

$$f(y) - f(x) - \langle \nabla f(x), y-x \rangle = \int_0^1 \langle \nabla f(x+\tau(y-x)) - \nabla f(y), y-x \rangle d\tau$$

$$\le \frac{1}{2} L \|y-x\|^2$$

(2.1.10) $\Rightarrow$ (2.1.13)

令 $x_\alpha = \alpha x + (1-\alpha) y$，则由 (2.1.10)

$$f(x) \ge f(x_\alpha) + \langle \nabla f(x_\alpha), (1-\alpha)(x-y) \rangle + \frac{1}{2L}\|\nabla f(x) - \nabla f(x_\alpha)\|_*^2 \quad \cdots ③$$

$$f(y) \ge f(x_\alpha) + \langle \nabla f(x_\alpha), \alpha(y-x) \rangle + \frac{1}{2L}\|\nabla f(y) - \nabla f(x_\alpha)\|_*^2 \quad \cdots ④$$

$\alpha \times ③ + (1-\alpha) \times ④ \Rightarrow$

$$\alpha f(x)+(1-\alpha)f(y) \geqslant f(x_\alpha)+\frac{\alpha}{2L}\|\nabla f(x)-\nabla f(x_\alpha)\|_*^2$$
$$+\frac{1-\alpha}{2L}\|\nabla f(y)-\nabla f(x_\alpha)\|_*^2$$

$$\geqslant f(x_\alpha)+\frac{\alpha(1-\alpha)}{2L}\|\nabla f(x)-\nabla f(y)\|_*^2$$

$(2.1.13) \Rightarrow (2.1.10)$

$$f(y) \geqslant \frac{f(\alpha x+(1-\alpha)y)-\alpha f(x)}{1-\alpha}+\frac{1}{2L}\|\nabla f(x)-\nabla f(y)\|_*^2$$

$$=\frac{f(\alpha x+(1-\alpha)y)-f(x)}{1-\alpha}+f(x)+\frac{1}{2L}\|\nabla f(x)-\nabla f(y)\|_*^2$$

取$\alpha\to 1$ $= \langle \nabla f(x), y-x\rangle + f(x) + \frac{1}{2L}\|\nabla f(x)-\nabla f(y)\|_*^2$

$(2.1.9) \Leftrightarrow (2.1.14)$ 同理 $\square$

$$\begin{tikzcd}
\text{原始定义} \arrow[r, Leftrightarrow] & 2.1.9 \arrow[r, Rightarrow] \arrow[dl, Leftrightarrow] \arrow[d, Leftrightarrow] & 2.1.10 \arrow[r, Rightarrow] \arrow[d, Leftrightarrow] & 2.1.11 \arrow[lll, bend right, Rightarrow] \\
2.1.14 & 2.1.12 & 2.1.13 &
\end{tikzcd}$$

$2.1.9 \Rightarrow$ 原始定义显然

不需要 $x\in\mathbb{R}^n$ 条件

定理 2.1.6 $f \in \mathcal{F}_L^{2,1}(\mathbb{R}^n, \|\cdot\|) \Leftrightarrow \forall x, h \in \mathbb{R}^n$,

$$0 \le \langle \nabla^2 f(x) h, h \rangle \le L \|h\|^2$$

证：左边显然，右边是 (2.1.12) 极限形式

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \le L \|x - y\|^2$$

设 $x - y = \alpha h$，则

$$\langle \nabla f(y + \alpha h) - \nabla f(y), \alpha h \rangle \le L \alpha^2 \|h\|^2$$

$\alpha \to 0^+$，则 $\langle \nabla^2 f(y) h, h \rangle \le L \|h\|^2$

$\square$

§ 2.1.2 lower complexity bounds for $\mathcal{F}_L^{\infty,1}(\mathbb{R}^n)$

假设 2.1.4：迭代方法 $\mathcal{M}$ 生成了序列 $\{x_k\}_{k=1}^{\infty}$ 满足

$$x_k \in x_0 + \mathrm{Lin}\{\nabla f(x_0), \cdots, \nabla f(x_{k-1})\}, \quad k \geq 1$$

$\boxed{L>0}$ 定义：

$$f_k(x) = \frac{L}{4}\left\{\frac{1}{2}\left[(x^{(1)})^2 + \sum_{i=1}^{k-1}(x^{(i)} - x^{(i+1)})^2 + (x^{(k)})^2\right] - x^{(1)}\right\}$$

故
$$\frac{\partial f_k(x)}{\partial x_i} = \frac{L}{4}\begin{cases} 2x^{(1)} - x^{(2)} - 1, & i=1 \\ -x^{(i-1)} + 2x^{(i)} - x^{(i+1)}, & 1<i<k \\ -x^{(k-1)} + 2x^{(k)}, & i=k \\ 0, & i \geq k \end{cases}$$

$$\nabla f_k(x) = A_k x - e_1$$

故
$$\nabla^2 f_k(x) = \left(\frac{\partial^2 f_k(x)}{\partial x_i \partial x_j}\right)_{n\times n} = \frac{L}{4}\cdot \left[\begin{array}{c|c} \begin{array}{c} \\ k行 \end{array}\left\{\begin{matrix} 2 & -1 & 0 & & \\ -1 & 2 & -1 & & \\ 0 & -1 & 2 & -1 & \\ & \ddots & \ddots & \ddots & -1 \\ & & 0 & -1 & 2 \end{matrix}\right. & 0 \\ \hline 0 & 0 \end{array}\right]_{n\times n}$$

$$\triangleq A_k$$

故

$$\langle \nabla^2 f_k(x) h, h\rangle = \frac{L}{4}\Big[(h^{(1)})^2 + \sum_{i=1}^{k-1}(h^{(i)} - h^{(i+1)})^2 + (h^{(k)})^2\Big] \geqslant 0$$

且

$$\langle \nabla^2 f_k(x) h, h\rangle \leqslant \frac{L}{4}\Big[(h^{(1)})^2 + \sum_{i=1}^{k-1} 2\big((h^{(i)})^2 + (h^{(i+1)})^2\big) + (h^{(k)})^2$$

$$\leqslant L\,\|h\|_2^2$$

从而，由定理2.1.6，$f_k \in \mathcal{F}_L^{\infty,1}(\mathbb{R}^n)$

方程 $\nabla f_k(x) = A_k x - e_1 = 0$ 的解是

$$\bar{x}_k^{(i)} = \begin{cases} 1 - \dfrac{i}{k+1}, & i = 1, \cdots, k \\ \text{任意}, & k+1 \leqslant i \leqslant n \end{cases}$$

为0，若 $\bar{x}_k \in \mathbb{R}^{k,n}$

由 f_k 的，故：

$$f_k^* = f_k(\bar{x}_k) = \frac{L}{4}\Big[\frac{1}{2}\langle A_k \bar{x}_k, \bar{x}_k\rangle - \langle e_1, \bar{x}_k\rangle\Big]$$

$$= \frac{L}{8}\Big(-1 + \frac{1}{k+1}\Big)$$

注意到：$\displaystyle\sum_{i=1}^{k} i^2 = \frac{k(k+1)(2k+1)}{6} \leqslant \frac{(k+1)^3}{3}$

故

$$\sum_{i=1}^{k}(\bar{x}_k^{(i)})^2=\sum_{i=1}^{k}\left(1-\frac{i}{k+1}\right)^2$$

$$=k-\frac{2}{k+1}\sum_{i=1}^{k}i+\frac{1}{(k+1)^2}\sum_{i=1}^{k}i^2$$

$$\leq k-\frac{2}{k+1}\cdot\frac{k(k+1)}{2}+\frac{1}{(k+1)^2}\frac{(k+1)^3}{3}$$

$$=\frac{1}{3}(k+1)$$

令 $\mathbb{R}^{k,n}=\{x\in\mathbb{R}^n \mid x^{(i)}=0,\ k+1\leq i\leq n\}$

则 $\forall x\in\mathbb{R}^{k,n}$，$f_p(x)\equiv f_k(x)$，$p=k,\cdots,n$

固定 p，$1\leq p\leq n$，且 $p\geq k$

引理 2.1.5：令 $x_0=0$，则对 $\forall\{x_k\}_{k=0}^{p}$，满足

$$x_k\in \mathcal{L}_k\triangleq \mathrm{Lin}\{\nabla f_p(x_0),\cdots,\nabla f_p(x_{k-1})\}$$

则 $\mathcal{L}_k\subseteq\mathbb{R}^{k,n}$

证明：由 $x_0=0$，故 $\nabla f_p(x_0)=-\frac{L}{4}e_1\in\mathbb{R}^{1,n}$

故 $\mathcal{L}_1\subseteq\mathbb{R}^{1,n}$，用数归：

设 $\mathcal{L}_k\subseteq\mathbb{R}^{k,n}$，则由 A_p 是三对角阵．$\forall x\in\mathbb{R}^{k,n}$，

有 $\nabla f_p(x) = \frac{L}{4}(A_p x - e_1) \in \mathbb{R}^{k+1,n}$

故 $\mathcal{J}_{k+1} \subseteq \mathbb{R}^{k+1,n}$，即证 $\square$

定理2.1.7：$\forall k,\ 1 \le k \le \frac{1}{2}(n-1),\ \forall x_0 \in \mathbb{R}^n,\ \exists f \in \mathcal{F}_L^{\infty,1}(\mathbb{R}^n)$，

使得对任意一阶方法 $\mathcal{M}$ 满足假设2.1.4，有：

$$f(x_k) - f^* \geq \frac{3L\|x_0 - x^*\|^2}{32(k+1)^2}$$

$$\|x_k - x^*\|^2 \geq \frac{1}{8}\|x_0 - x^*\|^2$$

x^* 是 f 全局极小

证明：

W.L.O.G.，设 $x_0 = 0$，否则 $\bar{f}(x) = f(x + x_0)$

固定 k，令 $f(x) = f_{2k+1}(x)$，则 $x^* = \bar{x}_{2k+1}$，$f^* = f^*_{2k+1}$

由推论2.1.1，有：

$$f(x_k) \equiv f_{2k+1}(x_k) = f_k(x_k) \geq f^*$$

由 $x_0 = 0$，故由(2.1.17)，(2.1.19)知

f^* f^*

$$\frac{\overset{J_k}{f(x_k)} - \overset{J_{2k+1}}{f^*}}{\|x_0 - x^*\|^2} \geqslant \frac{\frac{L}{8}\left(-1 + \frac{1}{k+1} + 1 - \frac{1}{2k+2}\right)}{\frac{1}{3}(2k+2)}$$

$$= \frac{3}{8}L \cdot \frac{1}{4(k+1)^2}$$

$$\|x_k - x^*\|^2 \geqslant \sum_{i=k+1}^{2k+1} \left(\bar{x}_{2k+1}^{(i)}\right)^2$$

$$= \sum_{i=k+1}^{2k+1} \left(1 - \frac{i}{2k+2}\right)^2$$

$$= k+1 - \frac{1}{k+1}\sum_{i=k+1}^{2k+1} i + \frac{1}{4(k+1)^2}\sum_{i=k+1}^{2k+1} i^2$$

$$= k+1 - \frac{1}{k+1}\,\frac{(3k+2)(k+1)}{2} + \frac{(2k+1)(7k+6)}{24(k+1)}$$

$$= \frac{2k^2+7k+6}{24(k+1)}$$

$$\overset{(2.1.19)}{\geqslant} \frac{2k^2+7k+6}{16(k+1)^2}\, \|x_0 - \bar{x}_{2k+1}\|^2$$

$$\geqslant \frac{1}{8}\|x_0 - x^*\|^2$$

$\boxtimes$

注：定理 2.1.7 证的是存在性，因为本书找的是于以上的最低复杂度，所以找一个最差的 func，它的复余度就是

$\mathcal{F}_L^{\infty,1}$ 的 Lower complexity bound

§ 2.1.3 Strongly convex functions

定义 2.1.3：连续可微 func $f(\cdot)$ 是强凸的，若 $\exists \mu > 0$. $\mathcal{S}_\mu^1(Q, \|\cdot\|)$

s.t. $\forall x, y \in Q$，有：

$$f(y) \geq f(x) + \langle \nabla f(x), y-x \rangle + \frac{1}{2}\mu \|y-x\|^2$$

定理 2.1.8：若 $f \in \mathcal{S}_\mu^1(\mathbb{R}^n)$，$\nabla f(x^*) = 0$，则

$$f(x) \geq f(x^*) + \frac{1}{2}\mu \|x - x^*\|^2 \quad \forall x \in \mathbb{R}^n$$

证：显然

引理 2.1.6 $f_1 \in \mathcal{S}_{\mu_1}^1(Q_1, \|\cdot\|)$，$f_2 \in \mathcal{S}_{\mu_2}^2(Q_2, \|\cdot\|)$，

$\alpha, \beta \geq 0$，则：

$$f = \alpha f_1 + \beta f_2 \in \mathcal{S}_{\alpha\mu_1+\beta\mu_2}^1(Q_1 \cap Q_2, \|\cdot\|)$$

证：显然

定理 2.1.9：以下等价于 $f \in \mathcal{F}_\mu^1(Q, \|\cdot\|)$

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \mu \|x - y\|^2$$

$$\alpha f(x) + (1-\alpha) f(y) \geq f(\alpha x + (1-\alpha) y) + \alpha(1-\alpha)\frac{\mu}{2}\|x - y\|^2$$

证明：$g(x) = f(x) - \frac{\mu}{2}\|x\|^2$ 是 convex func

证：$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|y - x\|^2$

故 $f(y) - \frac{\mu}{2}\|y\|^2 \geq f(x) - \frac{\mu}{2}\|x\|^2$

$$+ \langle \nabla f(x) - \mu x, y - x \rangle$$

下面用定理 2.1.2 和 2.1.3 对 $g(x)$ 即可 ☒

定理 2.1.10：若 $f \in \mathcal{F}_\mu^1(\mathbb{R}^n, \|\cdot\|)$，则 $\forall x, y$，有

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2\mu}\|\nabla f(x) - \nabla f(y)\|_*^2 \quad (2.1.24)$$

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \leq \frac{1}{\mu}\|\nabla f(x) - \nabla f(y)\|_*^2 \quad (2.1.25)$$

$$\mu\|x - y\|^2 \leq \|\nabla f(x) - \nabla f(y)\|_*^2 \quad (2.1.26)$$

$$\frac{1}{2}\|\nabla f(x)\|_*^2 \geq \mu(f(x) - f^*) \quad (P\text{-}L)$$

证明：固定 $x\in\mathbb{R}^n$，考虑

$$\phi(y)=f(y)-\langle\nabla f(x),y\rangle\in\mathcal{S}_{\mu}^{1}(\mathbb{R}^n,\|\cdot\|)$$

由 $\nabla\phi(y)=0$，故对 $\forall y\in\mathbb{R}^n$，有

$$\phi(x)=\min_{v\in\mathbb{R}^n}\phi(v)$$

$$\geqslant\min_{v\in\mathbb{R}^n}\left[\phi(y)+\langle\nabla\phi(y),v-y\rangle+\frac{1}{2}\mu\|y-v\|^2\right]$$

$$=\min_{r\geq0}\left[\phi(y)-\|\nabla\phi(y)\|_*r+\frac{1}{2}\mu r^2\right]$$

$$=\phi(y)-\frac{1}{2\mu}\|\nabla\phi(y)\|_*^2$$

成立是显然的，由有限维赋范空间均可达范，且 $v\in\mathbb{R}^n$，则 $v-y\in\mathbb{R}^n$

(2.1.24)即证，(2.1.25) 只需将 (2.1.24) x,y 互换，再相加.

(2.1.26)由 (2.1.25)，(2.1.22)可证

强凸 $\Rightarrow$ PL条件：

$$f(x^*)=\min_{y\in\mathbb{R}^n}f(y)\geqslant\min_{y\in\mathbb{R}^n}\left\{f(x)+\langle\nabla f(x),y-x\rangle+\frac{\mu}{2}\|y-x\|^2\right\}$$

$$=\min_{r\geqslant0}\left\{f(x)-r\|\nabla f(x)\|_*+\frac{\mu}{2}r^2\right\}$$

$$= f(x) - \frac{1}{2\mu}\|\nabla f(x)\|_*^2$$

定理2.1.11：f二次可微，则$f \in \mathcal{S}_\mu^2(Q, \|\cdot\|) \Leftrightarrow$
$\forall x \in \mathrm{int}\, Q, h \in \mathbb{R}^n$，有：

$$\langle \nabla^2 f(x) h, h\rangle \geq \mu \|h\|^2$$

证明：由(2.1.22)，取$y = x + \alpha h$，令$\alpha \to 0^+$即得
（同定理2.1.6）

例：
$$\eta(x) = \sum_{i=1}^n x^{(i)} \ln x^{(i)}$$
$$x \in \Delta_n^+ \doteq \{x \in \mathbb{R}_+^n : \langle \bar{e}_n, x\rangle \leq 1\} \quad (\text{凸集})$$

则对$\forall h \in \mathbb{R}^n$，有 $\langle \nabla^2 \eta(x) h, h\rangle = \sum_{i=1}^n \frac{(h^{(i)})^2}{x^{(i)}}$

故$\langle \nabla^2 \eta(x) h, h\rangle$关于$x^{(i)}$ $\forall i \downarrow$，则考察：

$$\min_{\langle \bar{e}_n, x\rangle = 1} \sum_{i=1}^n \frac{(h^{(i)})^2}{x^{(i)}}$$

故Lagrangian：

$$\mathcal{L}(x, \lambda) = \sum_{i=1}^n \frac{(h^{(i)})^2}{x^{(i)}} + \lambda(\langle \bar{e}_n, x\rangle - 1)$$

$$\frac{\partial \mathcal{L}}{\partial x_i} = -\frac{(h^{(i)})^2}{(x^{(i)})^2} + \lambda$$

$$\frac{\partial L}{\partial \lambda} = \langle \bar{e}_n, x \rangle - 1$$

故有：x^* 满足 $\lambda_* = \dfrac{(h^{(i)})^2}{(x_*^{(i)})^2} \rightarrow \sqrt{\lambda_*} = \dfrac{|h^{(i)}|}{x_*^{(i)}}$

故 $\sum_{i=1}^{n} x_*^{(i)} = \lambda_*^{-\frac{1}{2}} \sum_{i=1}^{n} |h^{(i)}| = 1$

$$\text{故 } \langle \nabla^2 \eta(x) h, h \rangle \geq \sum_{i=1}^{n} \frac{(h^{(i)})^2}{x_*^{(i)}}$$

$$= \lambda^* \sum_{i=1}^{n} x_*^{(i)}$$

$$= \lambda^*$$

$$= \left(\sum_{i=1}^{n} |h^{(i)}|\right)^2$$

$$= \|h\|_1^2$$

$\square$

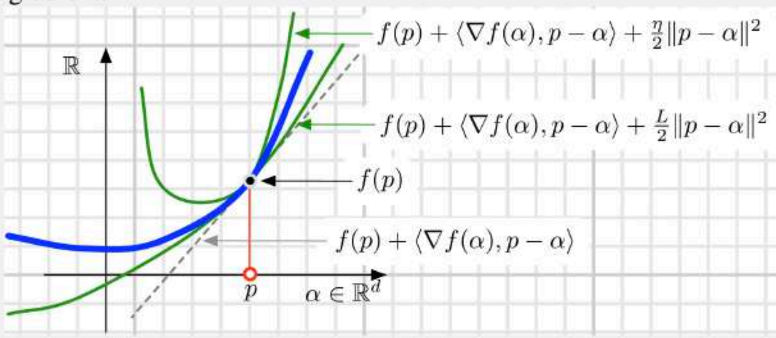
考虑函数类 $\mathscr{S}_{\mu,L}^{1,1}(\mathbb{R}^n)$ 2范数：

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \mu \|x - y\|^2$$

$$\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\|$$

Example: Strongly Convex Function

We show an example η-strongly convex function f, in blue. At any point p, it is sandwiched between two convex quadratic functions in green. The convex quadratic function which lower bounds f has an L-Lipschitz gradient.



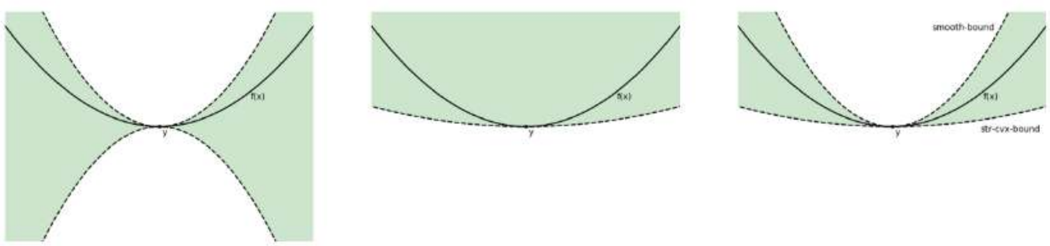


Figure 3: (Left) Upper and lower bounds from equation 4 for smoothness constraint (Middle) Lower bound from equation 5 for strong convexity constraint (Right) Combination of upper bound from smoothness and lower bound from strong convexity

定理 2.1.12：若 $f\in \mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$，对 $\forall x,y\in\mathbb{R}^n$，有：

$$\langle \nabla f(x)-\nabla f(y), x-y\rangle \geq \frac{\mu L}{\mu+L}\|x-y\|^2+\frac{1}{\mu+L}\|\nabla f(x)-\nabla f(y)\|^2$$

证明：定义 $\phi(x)=f(x)-\frac{1}{2}\mu\|x\|^2$

则 $\nabla\phi(x)=\nabla f(x)-\mu x$

由 $0\leq \phi(y)-\phi(x)-\langle\nabla\phi(x),y-x\rangle$

$$= f(y)-f(x)-\langle\nabla f(x),y-x\rangle-\frac{1}{2}\mu\|x-y\|^2$$

$$\leq \frac{L-\mu}{2}\|x-y\|^2$$

故 $\phi(x)\in\mathcal{F}^{1,1}_{L-\mu}(\mathbb{R}^n)$

$\mu=L$时结论已经证明了.

$\mu<L$时，有

$$\langle\nabla\phi(x)-\nabla\phi(y), x-y\rangle \geq \frac{1}{L-\mu}\|\nabla\phi(x)-\nabla\phi(y)\|^2$$

这等价于 (2.1.32)

□

§2.1.4 Lower Complexity bounds for $\mathcal{S}^{\infty,1}_{\mu,L}(\mathbb{R}^n)$

定义条件数 $Q_f=\frac{L}{\mu}$，定义 $f_{\mu,Q_f}:\ell^2\longrightarrow\mathbb{R}$

$$f_{\mu,Q_f}(x)=\frac{\mu(Q_f-1)}{8}\left\{(x^{(1)})^2+\sum_{i=1}^{\infty}(x^{(i)}-x^{(i+1)})^2-2x^{(1)}\right\}+\frac{\mu}{2}\|x\|^2$$

令 $L=\mu Q_f$, $A=\begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & \ddots \\ 0 & 0 & \ddots & \ddots \end{pmatrix}$

则 $\nabla^2 f_{\mu,Q_f}(x)=\frac{\mu(Q_f-1)}{4}A+\mu I$

易知 $0\preceq A\preceq 4I$ (A 和 $4I-A$ 都严格主对角占优)

故 $\mu I\preceq\nabla^2 f_{\mu,Q_f}(x)\preceq(\mu(Q_f-1)+\mu)I=\mu Q_f I=LI$

故 $f_{\mu,Q_f}\in\mathcal{S}_{\mu,L}^{\infty,1}(\ell^2)$

则由一阶最优性条件

$$\nabla f_{\mu,Q_f}(x)=\left(\frac{\mu(Q_f-1)}{4}A+\mu I\right)x-\frac{\mu(Q_f-1)}{4}e_1=0$$

$$\Leftrightarrow\quad \left(A+\frac{4}{Q_f-1}I\right)x=e_1$$

$$\Leftrightarrow\begin{cases} 2\dfrac{Q_f+1}{Q_f-1}x^{(1)}-x^{(2)}=1 \\ x^{(k+1)}-2\dfrac{Q_f+1}{Q_f-1}x^{(k)}+x^{(k-1)}=0,\ k=2,\cdots \end{cases}$$

特征方程：

$$q^2 - 2\frac{Q_f+1}{Q_f-1}q+1=0$$

则 $q=\dfrac{\sqrt{Q_f}-1}{\sqrt{Q_f}+1}$ （为了令 $x^*\in\ell^2$，故舍去大于1解）

则 $(x^*)^{(k)}=q^k$，$k=1,2,\cdots$

定理2.1.13：对 $\forall x_0\in\ell^2$，$\forall \mu>0$，$Q_f>1$，$\exists f\in\mathcal{F}_{\mu,L}^{\infty,1}(\ell^2)$，

s.t. 任意一阶算法 $\mathcal{M}$ 满足假设2.1.4，有：

$$\|x_k-x^*\|^2 \geqslant \left(\frac{\sqrt{Q_f}-1}{\sqrt{Q_f}+1}\right)^{2k}\|x_0-x^*\|^2 \quad ①$$

$$f(x_k)-f(x^*) \geqslant \frac{\mu}{2}\left(\frac{\sqrt{Q_f}-1}{\sqrt{Q_f}+1}\right)^{2k}\|x_0-x^*\|^2 \quad ②$$

证明：W.L.O.G. 设 $x_0=0$，取 $f(x)=f_{\mu,Q_f}(x)$，则

$$\|x_0-x^*\|^2=\sum_{i=1}^{\infty}[(x^*)^{(i)}]^2$$

$$=\sum_{i=1}^{\infty}q^{2i}=\frac{q^2}{1-q^2}$$

由 $\nabla^2 f_{\mu,Q_f}(x)$ 是三对角算子，且 $\nabla f_{\mu,Q_f}(0)=-\frac{L-\mu}{4}e_1$

同样由数归，$x_k \in \mathbb{R}^{k,\infty}$，故

$$\|x_k - x^*\|^2 \geq \sum_{i=k+1}^{\infty} [(x^*)^{(i)}]^2 = \frac{q^{2(k+1)}}{1-q^2} = q^{2k}\|x_0 - x^*\|^2$$

又由定理 2.1.8，$f(x) \geq f(x^*) + \frac{1}{2}\mu\|x - x^*\|^2$

则②得证 $\square$

§ 2.1.5 The gradient method

优化的问题：$\min_{x\in\mathbb{R}^n} f(x)$，$f\in \mathcal{F}_L^{1,1}(\mathbb{R}^n)$

定理 2.1.14：令 $f\in \mathcal{F}_L^{1,1}(\mathbb{R}^n)$，$0<h<\frac{2}{L}$，则梯度法生成的序列 $\{x_k\}$ 满足：

$$f(x_k)-f^*\leq \frac{2(f(x_0)-f^*)\|x_0-x^*\|^2}{2\|x_0-x^*\|^2+kh(2-Lh)(f(x_0)-f^*)}, \quad k\geq 0$$

证明：令 $r_k=\|x_k-x^*\|$，则

$$r_{k+1}^2=\|x_k-x^*-h\nabla f(x_k)\|^2$$

$$=r_k^2-2h\langle\nabla f(x_k),x_k-x^*\rangle+h^2\|\nabla f(x_k)\|^2$$

$$\overset{(2.1.11)}{\leq} r_k^2-h\left(\frac{2}{L}-h\right)\|\nabla f(x_k)\|^2$$

由 (2.1.9)：

$$f(x_{k+1})\leq f(x_k)+\langle\nabla f(x_k),x_{k+1}-x_k\rangle+\frac{L}{2}\|x_{k+1}-x_k\|^2$$

$$=f(x_k)-h\|\nabla f(x_k)\|^2+\frac{L}{2}h^2\|\nabla f(x_k)\|^2$$

$$\doteq f(x_k)-\omega\|\nabla f(x_k)\|^2 \quad \cdots\cdots \quad (*)$$

其中 $\omega=h(1-\frac{L}{2}h)$，记 $\Delta_k=f(x_k)-f^*$

故由 f 的凸性：

$$\Delta_k \leqslant \langle \nabla f(x_k), x_k - x^* \rangle \leqslant \|\nabla f(x_k)\| r_k \leqslant r_0 \|\nabla f(x_k)\|$$

由㊉知：

$$\Delta_{k+1} \leqslant \Delta_k - \omega \|\nabla f(x_k)\|^2 \leqslant \Delta_k - \frac{\omega}{r_0^2} \Delta_k^2$$
$$\leqslant \Delta_k - \frac{\omega}{r_0^2} \Delta_{k+1} \Delta_k$$

故 $\frac{1}{\Delta_{k+1}} \geqslant \frac{1}{\Delta_k} + \frac{\omega}{r_0^2} \geqslant \frac{1}{\Delta_0} + (k+1)\frac{\omega}{r_0^2}$ $\square$

由：

$$f(x_k) - f^* \leqslant \frac{2(f(x_0)-f^*)\|x_0 - x^*\|^2}{2\|x_0 - x^*\|^2 + kh(2-Lh)(f(x_0)-f^*)}$$

$h = \frac{1}{L}$ 时，取极大值

故 $h^* = \frac{1}{L}$ 时

$$f(x_k) - f^* \leqslant \frac{2L(f(x_0)-f^*)\|x_0 - x^*\|^2}{2L\|x_0 - x^*\|^2 + k(f(x_0)-f^*)}$$

又由 (2.1.9)：

$$f(x_0) - f^* \leqslant \langle \nabla f(x^*), x_0 - x^* \rangle + \frac{L}{2}\|x_0 - x^*\|^2$$
$$= \frac{L}{2}\|x_0 - x^*\|^2$$

故 $f(x_k)-f^* \le \dfrac{2L\|y_0-x^*\|^2}{2L\dfrac{\|x_0-x^*\|^2}{f(x_0)-f^*}+k}$

$\le \dfrac{2L\|x_0-x^*\|^2}{k+4}$ …… (推论 2.1.2)

定理 2.1.15：若 $f\in \mathcal{F}_{\mu,L}^{1,1}(\mathbb{R}^n)$，$0<h\le\frac{2}{\mu+L}$，则梯度法生成的序列 $\{x_k\}$ 满足：

$$\|x_k-x^*\|^2 \le \left(1-\frac{2h\mu L}{\mu+L}\right)^k \|x_0-x^*\|^2$$

若 $h=\frac{2}{\mu+L}$，则

$$\|x_k-x^*\| \le \left(\frac{Q_f-1}{Q_f+1}\right)^k \|x_0-x^*\|$$

$$f(x_k)-f^* \le \frac{L}{2}\left(\frac{Q_f-1}{Q_f+1}\right)^{2k} \|x_0-x^*\|^2$$

证明：$r_k=\|x_k-x^*\|$，则

$$r_{k+1}^2 = \|x_k-x^*-h\nabla f(x_k)\|^2$$

$$= r_k^2-2h\langle\nabla f(x_k), x_k-x^*\rangle+h^2\|\nabla f(x_k)\|^2$$

$$\overset{(2.1.32)}{\le} \left(1-\frac{2h\mu L}{\mu+L}\right)r_k^2+\underbrace{h\left(h-\frac{2}{\mu+L}\right)}_{\le 0}\|\nabla f(x_k)\|^2$$

$$\leq \left(1-\frac{2h\mu L}{\mu+L}\right) r_k^2 \leq \left(1-\frac{2h\mu L}{\mu+L}\right)^{k+1} r_0^2$$

当 $h=\frac{2}{\mu+L}$ 时，结论是显然的 $\square$

注：$f \in \mathcal{F}^{1,1}_{\mu,L}(\mathbb{R}^n)$ 时 $0<\mu\leq L$

$$\frac{1}{2L}\|\nabla f(x)\|^2 \leq f(x)-f^* \leq \frac{1}{2\mu}\|\nabla f(x)\|^2$$

$$\frac{\mu}{2}\|x-x^*\|^2 \leq f(x)-f^* \leq \frac{L}{2}\|x-x^*\|^2$$

§ 2.2.1 Estimating Sequence

定义 2.2.1 序列 $\{\phi_k(x)\}_{k=0}^{\infty}$ 和 $\{\lambda_k\}_{k=0}^{\infty}$，$\lambda_k \geq 0$ 叫做 $f(\cdot)$ 的 estimating 序列，若：对 $\forall x \in \mathbb{R}^n$，$\forall k \geq 0$，有

$$\phi_k(x) \leq (1-\lambda_k)f(x) + \lambda_k \phi_0(x), \text{ 且 } \lambda_k \to 0$$

引理 2.2.1 若对序列 $\{x_k\}$ 有：$f(x_k) \leq \phi_k^* \doteq \min_{x \in \mathbb{R}^n} \phi_k(x)$

则 $f(x_k) - f^* \leq \lambda_k [\phi_0(x^*) - f^*] \to 0$

证明：

$$\begin{aligned} f(x_k) \leq \phi_k^* &= \min_{x \in \mathbb{R}^n} \phi_k(x) \\ &= \min_{x \in \mathbb{R}^n} [(1-\lambda_k)f(x) + \lambda_k \phi_0(x)] \\ &\leq (1-\lambda_k) f(x^*) + \lambda_k \phi_0(x^*) \end{aligned}$$

$\square$

引理 2.2.2 假设

1. $f \in \mathcal{F}_{\mu,L}^{1,1}(\mathbb{R}^n)$

2. $\phi_0(\cdot)$ 是 $\mathbb{R}^n$ 中任意的 func

3. $\{y_k\}_{k=0}^{\infty}$ 是 $\mathbb{R}^n$ 中任意序列

4. $\{\alpha_k\}_{k=0}^{\infty}$ 满足：$\alpha_k \in (0,1)$ ，$\sum_{k=0}^{\infty} \alpha_k = \infty$

5. $\lambda_0 = 1$

则定义序列 $\{\phi_k\}_{k=0}^{\infty}$ 和 $\{\lambda_k\}_{k=0}^{\infty}$ ：

$$\lambda_{k+1} = (1-\alpha_k)\lambda_k$$

$$\phi_{k+1}(x) = (1-\alpha_k)\phi_k(x) + \alpha_k[f(y_k) + \langle \nabla f(y_k), x-y_k\rangle + \frac{\mu}{2}\|x-y_k\|^2] \quad (2.2.4)$$

是 estimating 序列：

证明：$\phi_0(x) \le (1-\lambda_0)f(x) + \lambda_0\phi_0(x) \equiv \phi_0(x)$

设 k 时成立，则：

$$\phi_{k+1}(x) \le (1-\alpha_k)\phi_k(x) + \alpha_k f(x)$$

$$= (1-(1-\alpha_k)\lambda_k)f(x) + (1-\alpha_k)(\phi_k(x) - (1-\lambda_k)f(x))$$

$$\le (1-(1-\alpha_k)\lambda_k)f(x) + (1-\alpha_k)\lambda_k\phi_0(x)$$

$$\le (1-\lambda_{k+1})f(x) + \lambda_{k+1}\phi_0(x)$$

即 k+1 时成立 $\boxtimes$

引理2.2.3：令 $\phi_0(x) = \phi_0^* + \frac{\nu_0}{2}\|x-\nu_0\|^2$，则 (2.2.4) 有：

$$\phi_k(x) \equiv \phi_k^* + \frac{\nu_k}{2}\|x-\nu_k\|^2 \text{，其中：} \quad (2.2.5)$$

$$\gamma_{k+1}=(1-\alpha_k)\gamma_k+\alpha_k\mu$$

$$v_{k+1}=\frac{1}{\gamma_{k+1}}[(1-\alpha_k)\gamma_k v_k+\alpha_k\mu y_k-\alpha_k\nabla f(y_k)]$$

$$\phi_{k+1}^*=(1-\alpha_k)\phi_k^*+\alpha_k f(y_k)-\frac{\alpha_k^2}{2\gamma_{k+1}}\|\nabla f(y_k)\|^2$$

$$+\frac{\alpha_k(1-\alpha_k)\gamma_k}{\gamma_{k+1}}\left(\frac{\mu}{2}\|y_k-v_k\|^2+\langle\nabla f(y_k),v_k-y_k\rangle\right)$$

证明：注意到 $\nabla^2\phi_0(x)=\gamma_0 I_n$，下证 $\nabla^2\phi_k(x)=\gamma_k I_n, k\geq 0$

设 k 时成立，则

$$\nabla^2\phi_{k+1}(x)=(1-\alpha_k)\nabla^2\phi_k(x)+\alpha_k\mu I_n$$

$$=((1-\alpha_k)\gamma_k+\alpha_k\mu)I_n\equiv\gamma_{k+1}I_n$$

故 ϕ_k 应有 (2.2.5) 的形式，进一步地

$$\phi_{k+1}(x)=(1-\alpha_k)(\phi_k^*+\frac{\gamma_k}{2}\|x-v_k\|^2)$$

$$+\alpha_k[f(y_k)+\langle\nabla f(y_k),x-y_k\rangle+\frac{\mu}{2}\|x-y_k\|^2]$$

则令 $\nabla\phi_{k+1}(x)=0$，有：

$$(1-\alpha_k)\gamma_k(x-v_k)+\alpha_k\nabla f(y_k)+\alpha_k\mu(x-y_k)=0$$

则 $x=v_{k+1}=\frac{1}{\gamma_{k+1}}[(1-\alpha_k)\gamma_k v_k+\alpha_k\mu y_k-\alpha_k\nabla f(y_k)]$

最后计算 ϕ_{k+1}^*：

$$\phi_{k+1}^* + \frac{\gamma_{k+1}}{2}\|y_k - v_{k+1}\|^2 = \phi_{k+1}(y_k)$$

$$= (1-\alpha_k)(\phi_k^* + \frac{\gamma_k}{2}\|y_k - v_k\|^2) + \alpha_k f(y_k)$$

由 v_{k+1} 的递归定义：

$$v_{k+1} - y_k = \frac{1}{\gamma_{k+1}}[(1-\alpha_k)\gamma_k(v_k - y_k) - \alpha_k \nabla f(y_k)]$$

故

$$\frac{\gamma_{k+1}}{2}\|v_{k+1} - y_k\|^2 = \frac{1}{2\gamma_{k+1}}[(1-\alpha_k)^2\gamma_k^2\|v_k - y_k\|^2 -$$

$$2\alpha_k(1-\alpha_k)\gamma_k\langle\nabla f(y_k), v_k - y_k\rangle + \alpha_k^2\|\nabla f(y_k)\|^2]$$

故 $$\phi_{k+1}^* = -\frac{\gamma_{k+1}}{2}\|v_{k+1} - y_k\|^2 + (1-\alpha_k)(\phi_k^* + \frac{\gamma_k}{2}\|y_k - v_k\|^2) + \alpha_k f(y_k)$$

$$= \left(-\frac{1}{2\gamma_{k+1}}(1-\alpha_k)^2\gamma_k^2 + (1-\alpha_k)\frac{\gamma_k}{2}\right)\|y_k - v_k\|^2$$

$$- \frac{\alpha_k^2}{2\gamma_{k+1}}\|\nabla f(y_k)\|^2 + \frac{\alpha_k}{\gamma_{k+1}}(1-\alpha_k)\gamma_k\langle\nabla f(y_k), v_k - y_k\rangle$$

$$+ (1-\alpha_k)\phi_k^* + \alpha_k f(y_k)$$

$$= (1-\alpha_k)\frac{\gamma_k}{2}\left(1-\frac{(1-\alpha_k)\gamma_k}{\gamma_{k+1}}\right)$$

$$= (1-\alpha_k)\frac{\gamma_k}{2}\frac{\alpha_k\mu}{\gamma_{k+1}}$$

$\boxed{/}$

假设我们已经有 x_k，且有 $\phi_k^* \geqslant f(x_k)$，则由引理 2.2.3

$$\phi_{k+1}^* \geqslant (1-\alpha_k)f(x_k)+\alpha_k f(y_k)-\frac{\alpha_k^2}{2\gamma_{k+1}}\|\nabla f(y_k)\|^2$$
$$+\frac{\alpha_k(1-\alpha_k)\gamma_k}{\gamma_{k+1}}\langle\nabla f(y_k), v_k-y_k\rangle$$

$$\geqslant f(y_k)-\frac{\alpha_k^2}{2\gamma_{k+1}}\|\nabla f(y_k)\|^2$$
$$+(1-\alpha_k)\left\langle\nabla f(y_k), \frac{\alpha_k\gamma_k}{\gamma_{k+1}}(v_k-y_k)+x_k-y_k\right\rangle$$

我们想让 $\phi_{k+1}^* \geqslant f(x_{k+1})$，由：

$$f(y_k)-\frac{1}{2L}\|\nabla f(y_k)\|^2 \geqslant f(x_{k+1})$$

不妨令 $x_{k+1}=y_k-h_k\nabla f(y_k)$，$h_k=\frac{1}{L}$

且 α_k 定义为下二次方程的根：

$$L\alpha_k^2=(1-\alpha_k)\gamma_k+\alpha_k\mu \ (=\gamma_{k+1})$$

故 $$\begin{aligned}\phi_{k+1}^* &\geq f(y_k) - \frac{1}{2L}\|\nabla f(y_k)\|^2 \\ &\quad + (1-\alpha_k)\left\langle \nabla f(y_k), \frac{\alpha_k\gamma_k}{\gamma_{k+1}}(v_k - y_k) + x_k - y_k \right\rangle \\ &\geq f(x_{k+1}) + (1-\alpha_k)\left\langle \nabla f(y_k), \frac{\alpha_k\gamma_k}{\gamma_{k+1}}(v_k - y_k) + x_k - y_k \right\rangle\end{aligned}$$

由于$\{y_k\}$是任意选的,故可以令:

$$\frac{\alpha_k\gamma_k}{\gamma_{k+1}}(v_k - y_k) + x_k - y_k = 0$$

General Scheme of Optimal Method

0. Choose the point $x_0 \in \mathbb{R}^n$, some $\gamma_0 > 0$, and set $v_0 = x_0$.

1. ***k*th iteration** ($k \geq 0$).

(a) Compute $\alpha_k \in (0, 1)$ from the equation

$$L\alpha_k^2 = (1-\alpha_k)\gamma_k + \alpha_k\mu. \tag{2.2.7}$$

Set $\gamma_{k+1} = (1-\alpha_k)\gamma_k + \alpha_k\mu$.

(b) Choose $y_k = \frac{1}{\gamma_k + \alpha_k\mu}[\alpha_k\gamma_k v_k + \gamma_{k+1}x_k]$. Compute $f(y_k)$ and $\nabla f(y_k)$.

(c) Find x_{k+1} such that

$$f(x_{k+1}) \leq f(y_k) - \frac{1}{2L}\| \nabla f(y_k) \|^2$$

(see Sect. 1.2.3 for the step-size rules).

(d) Set $v_{k+1} = \frac{1}{\gamma_{k+1}}[(1-\alpha_k)\gamma_k v_k + \alpha_k\mu y_k - \alpha_k\nabla f(y_k)]$.

定理2.2.1：算法2.2.7生成的序列$\{x_k\}_{k=0}^{\infty}$满足：

$$f(x_k) - f^* \leq \lambda_k\left[f(x_0) - f^* + \frac{\gamma_0}{2}\|x_0 - x^*\|^2\right]$$

其中 $\lambda_0 = 1$, $\lambda_k = \prod_{i=0}^{k-1}(1-\alpha_i)$

证明：取 $\phi_0(x)=f(x_0)+\frac{\gamma_0}{2}\|x-v_0\|^2$，则 $f(x_0)=\phi_0^*$，

且 $f(x_k)\leq\phi_k^*$，故由引理2.2.1立即得 $\boxtimes$

引理 2.2.4 定义 $q_f = \frac{1}{Q_f} = \frac{\mu}{L}$

若 (2.2.7) 取 $\gamma_0 \in (\mu, 3L+\mu]$，则对 $\forall k \geq 0$. 有:

$$\lambda_k \leq \frac{4\mu}{(\gamma_0-\mu)\left[\exp\left(\frac{k+1}{2} q_f^{\frac{1}{2}}\right) - \exp\left(-\frac{k+1}{2} q_f^{\frac{1}{2}}\right)\right]^2} \leq \frac{4L}{(\gamma_0-\mu)(k+1)^2}$$

对 $\gamma_0 = \mu$，有 $\lambda_k = (1-\sqrt{q_f})^k$，$k \geq 0$

证明：设 $\gamma_0 > \mu$，则由

$$\gamma_{k+1} = (1-\alpha_k)\gamma_k + \alpha_k \mu$$

则 $\gamma_{k+1} - \mu = (1-\alpha_k)(\gamma_k - \mu) = \cdots = \lambda_{k+1}(\gamma_0 - \mu)$

由 $\alpha_k = 1 - \frac{\lambda_{k+1}}{\lambda_k}$，故由 $L\alpha_k^2 = (1-\alpha_k)\gamma_k + \alpha_k\mu = \gamma_{k+1}$

$$1 - \frac{\lambda_{k+1}}{\lambda_k} = \left[\frac{\gamma_{k+1}}{L}\right]^{\frac{1}{2}} = \left[\frac{\mu}{L} + \lambda_{k+1}\frac{\gamma_0-\mu}{L}\right]^{\frac{1}{2}}$$

故 $$\frac{1}{\lambda_{k+1}} - \frac{1}{\lambda_k} = \frac{1}{\lambda_{k+1}^{\frac{1}{2}}}\left[\frac{q_f}{\lambda_{k+1}} + \frac{\gamma_0-\mu}{L}\right]^{\frac{1}{2}}$$

$$= \left(\frac{1}{\lambda_{k+1}^{\frac{1}{2}}} + \frac{1}{\lambda_k^{\frac{1}{2}}}\right)\left(\frac{1}{\lambda_{k+1}^{\frac{1}{2}}} - \frac{1}{\lambda_k^{\frac{1}{2}}}\right)$$

$$\leq \frac{2}{\lambda_{k+1}^{\frac{1}{2}}}\left(\frac{1}{\lambda_{k+1}^{\frac{1}{2}}} - \frac{1}{\lambda_k^{\frac{1}{2}}}\right) \qquad \cdots\cdots (*)$$

定义 $\xi_k = \left[\frac{L}{(\gamma_0-\mu)\lambda_k}\right]^{\frac{1}{2}}$，则由 $(*)$

$$\left[\frac{q_f}{\lambda_{k+1}} + \frac{r_0-\mu}{L}\right]^{\frac{1}{2}} \le 2\left(\frac{1}{\lambda_{k+1}^{\frac{1}{2}}} - \frac{1}{\lambda_k^{\frac{1}{2}}}\right)$$

两边同除 $\sqrt{\frac{r_0-\mu}{L}}$，则有

$$\sqrt{\frac{L}{\lambda_{k+1}(r_0-\mu)}} - \sqrt{\frac{L}{\lambda_k(r_0-\mu)}} \geq \frac{1}{2}\left[q_f \cdot \frac{L}{\lambda_{k+1}(r_0-\mu)} + 1\right]^{\frac{1}{2}}$$

即 $\xi_{k+1} - \xi_k \geq \frac{1}{2}\left(q_f \xi_{k+1}^2 + 1\right)^{\frac{1}{2}}$

令 $\delta = \frac{1}{2}\sqrt{q_f}$，下证：（用数归）

$$\xi_k \geq \frac{1}{4\delta}\left[e^{(k+1)\delta} - e^{-(k+1)\delta}\right], \quad k \geq 0$$

对 $k=0$，有：

$$\xi_0 = \left[\frac{L}{r_0-\mu}\right]^{\frac{1}{2}} \geq \frac{1}{\sqrt{3}} > \frac{1}{2}\left(e^{\frac{1}{2}} - e^{-\frac{1}{2}}\right) \geq \frac{1}{4\delta}\left[e^{\delta} - e^{-\delta}\right] \checkmark$$

设 k 时成立，下证 $k+1$ 时成立，考虑 $\psi(t) = \frac{1}{4\delta}\left[e^{(t+1)\delta} - e^{-(t+1)\delta}\right]$

则 $\psi'(t) = \frac{1}{4}\left[e^{(t+1)\delta} + e^{-(t+1)\delta}\right] \uparrow$，故 $\psi(t)$ 凸

故 $\psi(k) \leq \xi_k \leq \xi_{k+1} - \frac{1}{2}\left[q_f \xi_{k+1}^2 + 1\right]^{\frac{1}{2}} \doteq \gamma(\xi_{k+1})$

由假设

注意到：
$$\gamma'(\xi) = 1 - \frac{1}{2}\frac{q_f \xi}{[q_f \xi^2 + 1]} \geq 1 - \frac{1}{2}\frac{\sqrt{q_f \xi^2 + 1}}{q_f \xi^2 + 1} = 1 - \frac{1}{2\sqrt{q_f \xi^2 + 1}} > 0$$

用反证法，设 $\xi_{k+1} < \psi(k+1)$

$$\psi(t) < \psi(t+1) - \frac{1}{2}\left[4\delta^2\left(\frac{1}{4\delta}\left[e^{(t+2)\delta} - e^{-(t+2)\delta}\right]\right)^2 + 1\right]^{\frac{1}{2}}$$

$$= \psi(t+1) - \frac{1}{4}\left[e^{(t+2)\delta} + e^{-(t+2)\delta}\right]$$

$$= \psi(t+1) + \psi'(t+1)(t-(t+1)) \leqslant \psi(t)$$

矛盾！故：

$$\xi_k \geqslant \frac{1}{4\delta}\left[e^{(k+1)\delta} - e^{-(k+1)\delta}\right]$$

即：
$$\lambda_k \leqslant \frac{4\mu}{(\gamma_0-\mu)\left[\exp\left(\frac{k+1}{2}q_f^{\frac{1}{2}}\right) - \exp\left(-\frac{k+1}{2}q_f^{\frac{1}{2}}\right)\right]^2}$$

当 $\gamma_0=\mu$ 时，有 $\gamma_k=\mu \ (\forall k \geqslant 0)$，则 $\alpha_k=\sqrt{q_f}$，$\forall k$ $\square$

<u>定理2.2.2</u> 在(2.2.7)中取 $\gamma_0=3L+\mu$，则生成的 $\{x_k\}$

满足：
$$f(x_k)-f^* \leqslant \frac{2(4+q_f)\mu\|x_0-x^*\|^2}{3\left[\exp\left(\frac{k+1}{2}q_f^{\frac{1}{2}}\right) - \exp\left(-\frac{k+1}{2}q_f^{\frac{1}{2}}\right)\right]^2}$$

$$\leqslant \frac{2(4+q_f)L\|x_0-x^*\|^2}{3(k+1)^2}$$

这意味着 (2.2.7) 对解无约束优化问题是optimal的

$f \in \mathcal{F}^{1,1}_{\mu,L}(\mathbb{R}^n)$，$\mu \geq 0$，当精度 $\varepsilon > 0$ 足够小：

$$\varepsilon \leq \frac{\mu}{2} \|x_0 - x^*\|^2$$

若 $\mu = 0$，则该方法对 $\varepsilon \leq \frac{3L}{32} \|x_0 - x^*\|^2$ 是 optimal 的

证明：由 $f(x_0) - f^* \leq \frac{L}{2} \|x_0 - x^*\|^2$，故由定理 2.2.1

$$f(x_k) - f^* \leq \lambda_k \left[f(x_0) - f^* + \frac{\gamma_0}{2} \|x_0 - x^*\|^2 \right]$$

$$\leq \frac{\lambda_k}{2} (L + \gamma_0) \|x_0 - x^*\|^2$$

由引理 2.2.4

$$f(x_k) - f^* \leq \frac{2\mu (L+\gamma_0) \|x_0 - x^*\|^2}{(\gamma_0 - \mu)\left[\exp\left(\frac{k+1}{2} q_f^{\frac{1}{2}}\right) - \exp\left(-\frac{k+1}{2} q_f^{\frac{1}{2}}\right)\right]^2}$$

$$\leq \frac{2L(L+\gamma_0) \|x_0 - x^*\|}{(\gamma_0 - \mu)(k+1)^2}$$

将 $\gamma_0 = 3L + \mu$ 代入上式即证

令 $\mu > 0$，有：

$$f(x_k) - f^* \geq \frac{\mu}{2} \left(\frac{\sqrt{Q_f} - 1}{\sqrt{Q_f} + 1} \right)^{2k} R^2 \geq \frac{\mu}{2} \exp\left(-\frac{4k}{\sqrt{Q_f} - 1} \right) R^2$$

其中 $R = \|x_0 - x^*\|$

注：下证 $\left(\frac{\sqrt{Q_f}-1}{\sqrt{Q_f}+1}\right)^{2k} \geqslant \exp\left(-\frac{4k}{\sqrt{Q_f}-1}\right)$

即证：$2k\ln\left(1-\frac{2}{\sqrt{Q_f}+1}\right) \geqslant -\frac{4k}{\sqrt{Q_f}-1}$ ⊛

$\Leftrightarrow \quad \frac{2}{\sqrt{Q_f}-1}+\ln\left(1-\frac{2}{\sqrt{Q_f}+1}\right) \geqslant 0$ 对 $\forall Q_f>1$ 成立

令 $g(x)=\frac{2}{x-1}+\ln\left(1-\frac{2}{x+1}\right)$, $x>1$

$g'(x)=-\frac{2}{(x-1)^2}+\frac{2}{(x-1)(x+1)}=\frac{2}{x-1}\left(\frac{1}{x+1}-\frac{1}{x-1}\right)<0$

而 $\lim\limits_{x\to+\infty} g(x)=0$，故⊛成立

故 $f(x_k)-f^* \leqslant \varepsilon$ 不可能好于：$\underbrace{k \geqslant \frac{\sqrt{Q_f}-1}{4}\ln\frac{\mu R^2}{2\varepsilon}}_{⊛}$

对于我们的算法：

$$f(x_k)-f^* \leqslant \frac{2(4+q_f)\mu\|x_0-x^*\|^2}{3\left[\exp\left(\frac{k+1}{2}q_f^{\frac{1}{2}}\right)-\exp\left(-\frac{k+1}{2}q_f^{\frac{1}{2}}\right)\right]^2}$$

$$\leqslant \frac{10}{3}\mu R^2\left[e^{(k+1)q_f^{\frac{1}{2}}}-2\right]^{-1}$$

则 $k>\sqrt{Q_f}\ln\left(2+\frac{10\mu R^2}{3\varepsilon}\right)$

而 $\ln\left(2+\frac{10\mu R^2}{3\varepsilon}\right) \leq \ln\left(\frac{\mu R^2}{\varepsilon}+\frac{10\mu R^2}{3\varepsilon}\right)$

$$= \ln\frac{13}{3}+\ln\frac{\mu R^2}{\varepsilon}$$

则由 $\sqrt{Q_f}\left(\ln\frac{13}{3}+\ln\frac{\mu R^2}{\varepsilon}\right)$ 与 ⊛ 成比例，故 (2.2.7) 是最优的

$\square$

注：对于强凸情形 $(\mu>0)$

$$f(x_k)-f^* \leq \frac{2(4+q_f)\mu\|x_0-x^*\|^2}{3\left[\exp\left(\frac{k+1}{2}q_f^{\frac{1}{2}}\right)-\exp\left(-\frac{k+1}{2}q_f^{\frac{1}{2}}\right)\right]^2}$$

$$\leq \frac{2(4+q_f)\mu\|x_0-x^*\|^2}{3\left[\exp\left(\frac{k+1}{2}\sqrt{q_f}\right)-1\right]^2}$$

$$= O\left(e^{-(k+1)\sqrt{q_f}}\right)$$

Constant Step Scheme I

0. Choose the point $x_0 \in \mathbb{R}^n$, some $\gamma_0 > 0$, and set $v_0 = x_0$.

1. **kth iteration ($k \geq 0$).**

(a) Compute $\alpha_k \in (0, 1)$ from the equation

$$L\alpha_k^2 = (1 - \alpha_k)\gamma_k + \alpha_k \mu. \tag{2.2.19}$$

Set $\gamma_{k+1} = (1 - \alpha_k)\gamma_k + \alpha_k \mu$.

(b) Choose $y_k = \frac{1}{\gamma_k + \alpha_k \mu}[\alpha_k \gamma_k v_k + \gamma_{k+1} x_k]$. Compute $f(y_k)$ and $\nabla f(y_k)$.

(c) Set $x_{k+1} = y_k - \frac{1}{L}\nabla f(y_k)$ and

$$v_{k+1} = \frac{1}{\gamma_{k+1}}[(1 - \alpha_k)\gamma_k v_k + \alpha_k \mu y_k - \alpha_k \nabla f(y_k)].$$

注意到：这里把(2.2.7)中的 Find x_{k+1}, s.t.

$$f(x_{k+1}) \leq f(y_k) - \frac{1}{2L}\|\nabla f(y_k)\|^2$$

换成了 $x_{k+1} = y_k - \frac{1}{L}\nabla f(y_k)$，则由1.2.3节的分析：

$$\begin{aligned} f(x_{k+1}) &\leq f(y_k) + \langle \nabla f(y_k), x_{k+1} - y_k \rangle + \frac{L}{2}\|x_{k+1} - y_k\|^2 \\ &= f(y_k) - \frac{1}{L}\|\nabla f(y_k)\|^2 + \frac{L}{2}\cdot\frac{1}{L^2}\|\nabla f(y_k)\|^2 \\ &= f(y_k) - \frac{1}{2L}\|\nabla f(y_k)\|^2 \end{aligned}$$

故取固定步长 $\frac{1}{L}$ 满足(2.2.7)的要求

下将(2.2.19)化简成更简单的形式，注意到：

$$y_k = \frac{1}{\gamma_k+\alpha_k\mu}(\alpha_k\gamma_k v_k+\gamma_{k+1}x_k)$$

$$x_{k+1} = y_k - \frac{1}{L}\nabla f(y_k)$$

$$v_{k+1} = \frac{1}{\gamma_{k+1}}[(1-\alpha_k)\gamma_k v_k+\alpha_k\mu y_k-\alpha_k\nabla f(y_k)]$$

故

$$v_{k+1} = \frac{1}{\gamma_{k+1}}\left\{\frac{(1-\alpha_k)}{\alpha_k}[(\gamma_k+\alpha_k\mu)y_k-\gamma_{k+1}x_k]+\alpha_k\mu y_k-\alpha_k\nabla f(y_k)\right\}$$

$$= \frac{1}{\gamma_{k+1}}\left\{\frac{(1-\alpha_k)\gamma_k}{\alpha_k}y_k+\mu y_k\right\}-\frac{1-\alpha_k}{\alpha_k}x_k-\frac{\alpha_k}{\gamma_{k+1}}\nabla f(y_k)$$

$$= x_k+\frac{1}{\alpha_k}(y_k-x_k)-\frac{1}{\alpha_k L}\nabla f(y_k)$$

$$= x_k+\frac{1}{\alpha_k}(x_{k+1}-x_k)$$

故：

$$y_{k+1} = \frac{1}{\gamma_{k+1}+\alpha_{k+1}\mu}(\alpha_{k+1}\gamma_{k+1}v_{k+1}+\gamma_{k+2}x_{k+1})$$

$$= x_{k+1}+\frac{\alpha_{k+1}\gamma_{k+1}(v_{k+1}-x_{k+1})}{\gamma_{k+1}+\alpha_{k+1}\mu}$$

$$= x_{k+1}+\beta_k(x_{k+1}-x_k)$$

(由 $v_{k+1} = x_{k+1}-\frac{\alpha_k-1}{\alpha_k}(x_{k+1}-x_k)$)

其中 $\beta_k = \dfrac{\alpha_{k+1}\gamma_{k+1}(1-\alpha_k)}{\alpha_k(\gamma_{k+1}+\alpha_{k+1}\mu)}$

$\{\gamma_k\}$ 的更新准则不变，仍是

$$\alpha_k^2 L = (1-\alpha_k)\gamma_k + \mu\alpha_k \equiv \gamma_{k+1}$$

故

$$\beta_k = \frac{\alpha_{k+1}\gamma_{k+1}(1-\alpha_k)}{\alpha_k(\gamma_{k+1}+\alpha_{k+1}\mu)}$$

$$= \frac{\alpha_{k+1}\gamma_{k+1}(1-\alpha_k)}{\alpha_k(\gamma_{k+1}+\alpha_{k+1}^2 L-(1-\alpha_{k+1})\gamma_{k+1})}$$

$$= \frac{\gamma_{k+1}(1-\alpha_k)}{\alpha_k(\gamma_{k+1}+\alpha_{k+1}L)}$$

$$= \frac{\alpha_k^2 L(1-\alpha_k)}{\alpha_k(\alpha_k^2 L+\alpha_{k+1}L)} = \frac{\alpha_k(1-\alpha_k)}{\alpha_k^2+\alpha_{k+1}}$$

又注意到 $\left(\frac{\gamma_{k+1}}{L} = \alpha_k^2\right)$

$$\alpha_{k+1}^2 = (1-\alpha_{k+1})\alpha_k^2 + q_f\,\alpha_{k+1}$$

$$\alpha_0^2 L = (1-\alpha_0)\gamma_0 + \mu\alpha_0$$

故 γ_0 可以视为 α_0 的函数！

Constant Step Scheme II

0. Choose the point $x_0 \in \mathbb{R}^n$, some $\alpha_0 \in (0,1)$, and set $y_0 = x_0$.

1. k**th iteration** ($k \geq 0$).

(a) Compute $f(y_k)$ and $\nabla f(y_k)$. Set $x_{k+1} = y_k - \frac{1}{L}\nabla f(y_k)$.

(b) Compute $\alpha_{k+1} \in (0,1)$ from the equation

$$\alpha_{k+1}^2 = (1-\alpha_{k+1})\alpha_k^2 + q_f \alpha_{k+1}.$$

Set $\beta_k = \frac{\alpha_k(1-\alpha_k)}{\alpha_k^2+\alpha_{k+1}}$ and $y_{k+1} = x_{k+1} + \beta_k(x_{k+1} - x_k)$.

(2.2.20)

定理2.2.3：(2.2.20)取α_0满足：

$$\sqrt{q_f} \leq \alpha_0 \leq \frac{2(3+q_f)}{3+\sqrt{21+4q_f}} \qquad (2.2.21)$$

则 $$f(x_k) - f^* \leq \frac{4\mu\left[f(x_0) - f^* + \frac{\gamma_0}{2}\|x_0 - x^*\|^2\right]}{(\gamma_0-\mu)\left[\exp\left(\frac{k+1}{2}\sqrt{q_f}\right) - \exp\left(-\frac{k+1}{2}\sqrt{q_f}\right)\right]^2}$$

$$\leq \frac{4L}{(\gamma_0-\mu)(k+1)^2}\left[f(x_0) - f^* + \frac{\gamma_0}{2}\|x_0 - x^*\|^2\right]$$

其中 $\gamma_0 = \frac{\alpha_0(\alpha_0 L - \mu)}{1-\alpha_0}$

证明：(2.2.21)与lemma 2.2.4中的$\gamma_0 \in [\mu, 3L+\mu)$等价

故该定理不用证。

若 $\alpha_0=\sqrt{q_f}$，则：$\alpha_k=\sqrt{q_f}$，$\beta_k=\dfrac{1-\sqrt{q_f}}{1+\sqrt{q_f}}$ $\quad\forall k\geqslant 0$

Constant Step scheme III

0. Choose $y_0=x_0\in\mathbb{R}^n$.
1. k**th iteration** ($k\geq 0$).

$$x_{k+1}=y_k-\tfrac{1}{L}\nabla f(y_k),$$

$$y_{k+1}=x_{k+1}+\tfrac{1-\sqrt{q_f}}{1+\sqrt{q_f}}(x_{k+1}-x_k). \tag{2.2.22}$$

故：

$$f(x_k)-f^*\leqslant(1-\sqrt{q_f})^k\Big[f(x_0)-f^*+\frac{\overset{=\mu}{\gamma_0}}{2}\|x_0-x^*\|^2\Big]$$

$$\leqslant\frac{L+\mu}{2}\|x_0-x^*\|^2(1-\sqrt{q_f})^k$$

$$\leqslant\frac{L+\mu}{2}\|x_0-x^*\|^2e^{-k\sqrt{q_f}},\quad k\geqslant 0$$

注：对 $\mu=0$（$q_f=0$）时，此法不适用

<u>定理 2.2.4</u>：(2.2.7)应用于func $f\in\mathcal{F}_L^{1,1}(\mathbb{R}^n)$，

则对 $\forall k\geqslant 0$，有：

$$\|v_k-x^*\|\leqslant\Big[1+\frac{1}{\gamma_0}L\Big]^{\frac12}\tilde{r}_0$$

$$\|x_k-x^*\|\leqslant\Big[1+\frac{1}{\gamma_0}L\Big]^{\frac12}\tilde{r}_0$$

其中 $\tilde{r}_0=\|x^*-x_0\|$，更多地，对 $g_k=\dfrac{\lambda_k}{1-\lambda_k}\displaystyle\sum_{i=0}^{k-1}\frac{\alpha_i}{\lambda_{i+1}}\nabla f(y_k)$

它的系数满足：$\sum_{i=0}^{k-1} \frac{\alpha_i}{\lambda_{i+1}} = \frac{1-\lambda_k}{\lambda_k}$，$k \geq 1$. 则

$$\|g_k\| \leq \frac{\lambda_k \gamma_0}{1-\lambda_k}\left(1+[1+\frac{1}{\gamma_0}L]^{\frac{1}{2}}\right)\tilde{r}_0$$

证明：

estimating func 列可以如下形式表示：

$$\phi_k(x) = l_k(x) + \lambda_k\left(f(x_0) + \frac{1}{2}\gamma_0\|x-x_0\|^2\right), \quad k \geq 0$$

其中 $l_k(x)$ 满足：$l_0(x) \equiv 0$，

$$l_{k+1}(x) = (1-\alpha_k) l_k(x) + \alpha_k[f(y_k) + \langle \nabla f(y_k), x-y_k \rangle], \quad k \geq 0$$

Why? 用数学归纳法：

$$\phi_0(x) = l_0(x) + \lambda_0\left(f(x_0) + \frac{1}{2}\gamma_0\|x-x_0\|^2\right)$$

$$= f(x_0) + \frac{1}{2}\gamma_0\|x-x_0\|^2 \quad \text{满足条件 ✓}$$

设 $\phi_k(x)$ 满足，下证 $\phi_{k+1}(x)$ 也满足：

$$\phi_{k+1}(x) = l_{k+1}(x) + \lambda_{k+1}\left(f(x_0) + \frac{1}{2}\gamma_0\|x-x_0\|^2\right)$$

$$= (1-\alpha_k) l_k(x) + \alpha_k[f(y_k) + \langle \nabla f(y_k), x-y_k \rangle]$$

$$+ (1-\alpha_k)\lambda_k\left(f(x_0) + \frac{1}{2}\gamma_0\|x-x_0\|^2\right)$$

$$= (1-\alpha_k)\phi_k(x) + \alpha_k[f(y_k) + \langle \nabla f(y_k), x-y_k \rangle]$$

由 (2.2.4)知，$\phi_{k+1}(x)$仍是估计func ☑

定义 $\nabla f_k \equiv \nabla f_k(x)$，$x\in\mathbb{R}^n$（$f_k$是仿射，故良定义）

由 $\phi_k(x)$的强凸性，且凸系数是$\lambda_k\gamma_0$，则对$\forall x\in\mathbb{R}^n$，有：

$$f(x_k)+\frac{1}{2}\lambda_k\gamma_0\|x-v_k\|^2 \leq \phi_k^* + \frac{1}{2}\lambda_k\gamma_0\|x-v_k\|^2$$

$$\leq \phi_k(x) \quad (\text{由}\phi_k\text{的强凸性})$$

$$\overset{(2.2.2)}{\leq} f(x)+\lambda_k(f(x_0)+\frac{1}{2}\gamma_0\|x-x_0\|^2-f(x))$$

故令 $x=x^*$，则

$$\frac{1}{2}\lambda_k\gamma_0\|x^*-v_k\|^2 \leq f(x^*)-f(x_k)+\lambda_k(f(x_0)+\frac{1}{2}\gamma_0\|x^*-x_0\|^2-f(x^*))$$

$$\leq \lambda_k(f(x_0)+\frac{1}{2}\gamma_0\|x^*-x_0\|^2-f(x^*))$$

即证(2.2.24)，下证 (2.2.25)，用数学归纳法：

当$k=0$时，有 $v_0=x_0$，(2.2.25)成立

设k时成立，下证$k+1$时也成立

$$\|y_k-x^*\| = \|\frac{1}{\gamma_k+\alpha_k\mu}[\alpha_k\gamma_k v_k+\gamma_{k+1}x_k]-x^*\|$$

$$\leq \frac{\alpha_k\gamma_k}{\gamma_k+\alpha_k\mu}\|v_k-x^*\|+\frac{\gamma_{k+1}}{\gamma_k+\alpha_k\mu}\|x_k-x^*\|$$

$$\leq [1+\frac{1}{\gamma_0}L]^{\frac{1}{2}}\widetilde{r_0} \quad (k+1 时成立)$$

由梯度法：

$$\|x_{k+1}-x^*\| \leq \|y_k-x^*\| \leq [1+\frac{1}{\gamma_0}L]^{\frac{1}{2}}\widetilde{r_0}$$

设 $S_k \triangleq \frac{1}{\lambda_k}\nabla t_k$，注意到 $S_0=0$，且

$$\nabla t_{k+1} = (1-\alpha_k)\nabla t_k + \alpha_k \nabla f(y_k)$$

$$= \frac{\lambda_{k+1}}{\lambda_k}\nabla t_k + \alpha_k \nabla f(y_k), \quad k \geq 0$$

故 $S_k = \sum_{i=0}^{k-1} \frac{\alpha_i}{\lambda_{i+1}}\nabla f(y_i)$，令 $\tau_i = \frac{\alpha_i}{\lambda_{i+1}}$

则 $\tau_i = \frac{\alpha_i}{(1-\alpha_i)\lambda_i} = \frac{1}{\lambda_{i+1}} - \frac{1}{\lambda_i}$

故 $\sum_{i=0}^{k-1}\tau_i = \frac{1}{\lambda_k}-1$，

$$g_k = \frac{\lambda_k S_k}{1-\lambda_k} \equiv \frac{\lambda_k}{1-\lambda_k}\nabla t_k(x), \quad x \in \mathbb{R}^n$$

注意到：

$$V_k = x_0 - \frac{1}{\lambda_k \gamma_0}\nabla t_k = x_0 - \frac{1-\lambda_k}{\lambda_k \gamma_0} g_k$$

同样用数学归纳法：

$v_0 = x_0$ ✓

设 k时成立，下验 k+1时成立

$$v_{k+1} = x_0 - \frac{1}{\lambda_{k+1}\gamma_0} \nabla \ell_{k+1}$$

$$= x_0 - \frac{1}{\lambda_{k+1}\gamma_0} \left(\frac{\lambda_{k+1}}{\lambda_k} \nabla \ell_k + \alpha_k \nabla f(y_k) \right)$$

$$= x_0 - \frac{1}{\lambda_k \gamma_0} \nabla \ell_k - \frac{\alpha_k}{\lambda_{k+1}\gamma_0} \nabla f(y_k)$$

$$= v_k - \frac{\alpha_k}{\prod_{n=0}^{k} (1-\alpha_n)\gamma_0} \nabla f(y_k)$$

$$= \frac{(1-\alpha_k)\gamma_k}{\gamma_{k+1}} - \frac{\alpha_k}{\gamma_{k+1}} \nabla f(y_k) \qquad \checkmark$$

故 $$\left[1 + \frac{1}{\gamma_0} L\right]^{\frac{1}{2}} \tilde{r}_0 \geq \left\| x_0 - \frac{1-\lambda_k}{\lambda_k \gamma_0} g_k - x^* \right\|$$

$$\geq \frac{1-\lambda_k}{\lambda_k \gamma_0} \| g_k \| - \tilde{r}_0$$

$\boxed{2}$

§ 2.2.2 Decreasing the norm of gradient

先考虑一般的梯度法 (2.1.37)，$h_k = \frac{1}{L}$

记 $R_0 = \|x_0 - x^*\|$，固定 $T \geqslant 3$，当 $0 \leqslant k < T$ 时

$$f(x_k) - f^* \overset{(2.1.39)}{\leqslant} \frac{2LR_0^2}{k+4}$$

设 $i \geqslant k$，则

$$\begin{aligned} f(x_i) - f(x_{i+1}) &\geqslant \langle \nabla f(x_i), x_i - x_{i+1} \rangle - \frac{L}{2} \|x_i - x_{i+1}\|^2 \\ &= \frac{1}{L} \|\nabla f(x_i)\|^2 - \frac{L}{2} \cdot \frac{1}{L^2} \|\nabla f(x_i)\|^2 \\ &= \frac{1}{2L} \|\nabla f(x_i)\|^2 \end{aligned}$$

记 $g_{k,T} = \min\limits_{k \leq i \leq T} \|\nabla f(x_i)\|$，则

$$\begin{aligned} (T-k+1) g_{k,T}^2 &\leqslant \sum_{i=k}^{T} \|\nabla f(x_i)\|^2 \\ &\leqslant 2L \sum_{i=k}^{T} (f(x_i) - f(x_{i+1})) \\ &= 2L (f(x_k) - f(x_{T+1})) \\ &\leqslant 2L (f(x_k) - f^*) \leqslant \frac{4L^2R_0^2}{k+4} \end{aligned}$$

故 $$g_{0,T}^2 \le \frac{4L^2R_0^2}{(k+4)(T-k+1)}$$

令 $q(k)=(k+4)(T-k+1)$，注意到：

$$q^* = \max_{k\in\mathbb{Z}} q(k) \geq q(\tau^*+\tfrac{1}{2}),\ \tau^* = \arg\max_{\tau\in\mathbb{R}} q(\tau)$$

$[\tau^*, \tau^*+\frac{1}{2}]$ 内总有一个整数

由 $\tau^* = \frac{T-3}{2}$，则 $q^* \geq q(\frac{T-2}{2}) = \frac{1}{4}(T+4)(T+6)$

从而得到下面的定理：

定理 2.2.5 令 $f\in\mathcal{F}_L^{1,1}(\mathbb{R}^n)$，取 (2.1.37) 中 $h_k=\frac{1}{L}$，则当 $T\geq 3$，有：

$$g_{0,T} \le \frac{4LR_0}{[(T+4)(T+6)]^{\frac{1}{2}}}$$

下面介绍单调版本的 Optimal Method.

Monotone Constant Step Scheme $\mathrm{I_A}$

0. Choose the point $x_0 \in \mathbb{R}^n$. Set $\lambda_0 = 1$ and $v_0 = x_0$.

1. k**th iteration** ($k \geq 0$).

(a) Compute $\alpha_k \in (0,1)$ from equation $\alpha_k^2 = 3(1-\alpha_k)\lambda_k$.

(b) Set $y_k = \alpha_k v_k + (1-\alpha_k)x_k$ and $\lambda_{k+1} = (1-\alpha_k)\lambda_k$.

(c) Compute $\nabla f(y_k)$ and set $\hat{x}_{k+1} = y_k - \frac{1}{L}\nabla f(y_k)$.

(d) Define $v_{k+1} = v_k - \frac{1}{L\alpha_k}\nabla f(y_k)$.

(e) Set $\hat{y}_k = \arg\min\left\{f(y):\ y \in \{x_k, \hat{x}_{k+1}\}\right\}$.

(f) Compute $\nabla f(\hat{y}_k)$ and set $x_{k+1} = \hat{y}_k - \frac{1}{L}\nabla f(\hat{y}_k)$.

(2.2.32)

上面的方法对应于(2.2.7)中取 $\gamma_0=3L$, $\mu=0$，则由

$$\begin{cases} \gamma_{k+1} = (1-\alpha_k)\gamma_k, & \gamma_0 = 3L \\ \lambda_{k+1} = (1-\alpha_k)\lambda_0, & \lambda_0 = 1 \end{cases}$$

$\Rightarrow \gamma_k \equiv 3L\lambda_k$，故

$$f(x_k) \overset{(2.2.32)e}{\geq} f(\hat{y}_k) \overset{(2.2.32)f}{\geq} f(x_{k+1}) + \frac{1}{2L}\|\nabla f(\hat{y}_k)\|^2$$

General Scheme of Optimal Method

0. Choose the point $x_0 \in \mathbb{R}^n$, some $\gamma_0 > 0$, and set $v_0 = x_0$.
1. k**th iteration** ($k \geq 0$).

(a) Compute $\alpha_k \in (0,1)$ from the equation

$$L\alpha_k^2 = (1-\alpha_k)\gamma_k + \alpha_k\mu. \tag{2.2.7}$$

Set $\gamma_{k+1} = (1-\alpha_k)\gamma_k + \alpha_k\mu$.

(b) Choose $y_k = \frac{1}{\gamma_k+\alpha_k\mu}[\alpha_k\gamma_k v_k + \gamma_{k+1}x_k]$. Compute $f(y_k)$ and $\nabla f(y_k)$.

(c) Find x_{k+1} such that

$$f(x_{k+1}) \leq f(y_k) - \frac{1}{2L}\|\nabla f(y_k)\|^2$$

(see Sect. 1.2.3 for the step-size rules).

(d) Set $v_{k+1} = \frac{1}{\gamma_{k+1}}[(1-\alpha_k)\gamma_k v_k + \alpha_k\mu y_k - \alpha_k\nabla f(y_k)]$.

(2.2.7)

$$x_k \longrightarrow y_k \longrightarrow x_{k+1}$$

(2.2.32)

$$x_k \longrightarrow y_k \longrightarrow \hat{x}_{k+1} \longrightarrow \hat{y}_k \longrightarrow x_{k+1}$$

$f(\hat{x}_{k+1}) \leq f(y_k) - \frac{1}{2L}\|\nabla f(y_k)\|^2 \Leftarrow$ 就是(2.2.7)中 x_{k+1} 的一种实现(2.2.19)

想证：(2.2.32)中：$f(x_{k+1}) \le f(y_k) - \frac{1}{2L}\|\nabla f(y_k)\|^2$

证：
$$
\begin{aligned}
f(x_{k+1}) &\le f(\hat{y}_k) - \frac{1}{2L}\|\nabla f(\hat{y}_k)\|^2 \\
&\le f(\hat{x}_{k+1}) - \frac{1}{2L}\|\nabla f(\hat{y}_k)\|^2 \\
&\le f(y_k) - \frac{1}{2L}\|\nabla f(y_k)\|^2 - \frac{1}{2L}\|\nabla f(\hat{y}_k)\|^2 \\
&\le f(y_k) - \frac{1}{2L}\|\nabla f(y_k)\|^2
\end{aligned}
$$

□

故 $T \geq 3$ 时，当 $0 \le k < T$：

$$f(x_k) - f^* \overset{(2.2.18)}{\le} \frac{8LR_0^2}{3(k+1)^2}$$

若 $i \geq k$，有：

$$f(x_i) - f(x_{i+1}) \overset{(2.2.23)}{\geq} \frac{1}{2L}\|\nabla f(\hat{y}_i)\|^2$$

令 $g_{k,T} = \min\limits_{k \le i \le T} \|\nabla f(\hat{y}_i)\|$，则

$$
\begin{aligned}
(T-k+1)\, g_{k,T}^2 &\le \sum_{i=k}^{T} \|\nabla f(\hat{y}_i)\|^2 \\
&\le 2L \sum_{i=k}^{T} (f(x_i) - f(x_{i+1})) \\
&= 2L\,(f(x_k) - f(x_{T+1})) \\
&\le 2L\,(f(x_k) - f^*)
\end{aligned}
$$

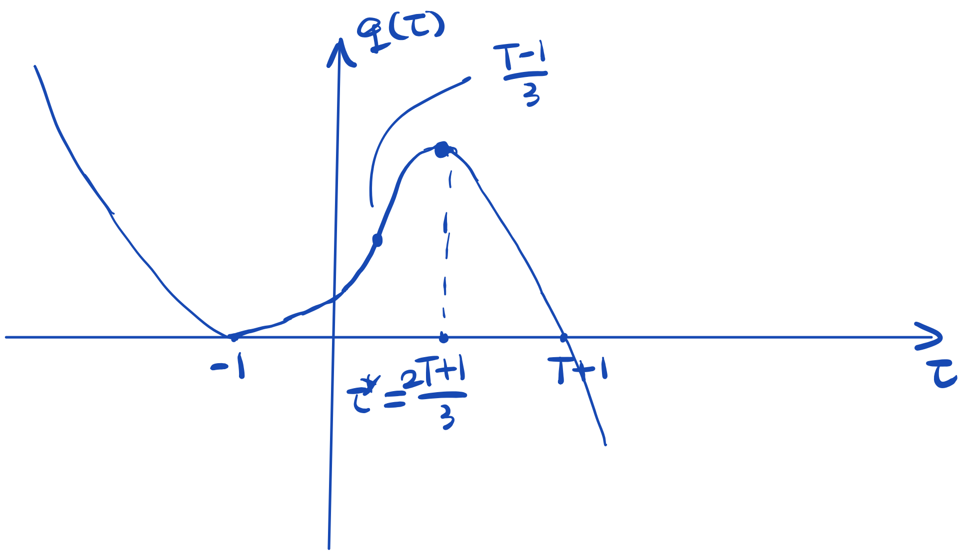
$$\le \frac{16L^2R_0^2}{3(k+1)^2}$$

故 $g_{0,T}^2 \le \dfrac{16L^2R_0^2}{3(k+1)^2(T-k+1)}$

令 $q(k) = (k+1)^2(T-k+1)$

$q'(k) = (k+1)(2T+1-3k)$

$q''(k) = 2T-2-6k$



故 $q^* \ge \min\{q(\tau^*-\frac{1}{2}),\ q(\tau^*+\frac{1}{2})\}$

$$= \min_{\delta=\pm\frac{1}{2}} \{q(\tau^*) + \tfrac{1}{2}q''(\tau^*)(\tfrac{1}{2})^2 + \tfrac{1}{6}q'''(\tau^*)\delta^3\}$$

$$= q(\tau^*) + \tfrac{1}{8}q''(\tau^*) - \tfrac{1}{8}$$

$$= \tfrac{4}{27}(T+2)^3 + \tfrac{1}{8}(-2T-4) - \tfrac{1}{8}$$

$$= \frac{4}{27}(T+2)^3 - \frac{1}{4}(T+2) - \frac{1}{8}$$

定理 2.2.6

若 $f \in \mathcal{F}_L^{1,1}(\mathbb{R}^n)$，则 (2.2.32) 有：

$$g_{0,T} \le \frac{4LR_0^2}{\left[\frac{4}{3}(T+2)^3 - \frac{9}{4}(T+2) - \frac{9}{8}\right]^{\frac{1}{2}}} \quad , T \ge 1$$

Regularization 技术

取 $f_\delta(x) = f(x) + \frac{1}{2}\delta \|x - x_0\|^2$

取 $f_\delta \in \mathcal{S}_{\delta, L+\delta}^{1,1}(\mathbb{R}^n)$，令 x_δ^* 是 f_δ 的全局最优，则

$$\nabla f(x_\delta^*) + \delta(x_\delta^* - x) = 0$$

注意到

$$f_\delta(x_\delta^*) + \frac{1}{2}\delta \|x_\delta^* - x^*\| \overset{(2.1.21)}{\le} f_\delta(x^*)$$

$$= f(x^*) + \frac{1}{2}\delta \|x^* - x_0\|^2$$

由 $f(x^*) \le f(x_\delta^*)$，有：

$$\|x_\delta^* - x_0\|^2 + \|x_\delta^* - x^*\|^2 \leqslant \|x_0 - x^*\|^2$$

又由

$$\|\nabla f(x_\delta^*)\| = \delta \|x_\delta^* - x_0\| \leqslant \delta R_0$$

故对 f_δ 用 (12.2.22),

$$\|\nabla f(x_T)\| \leqslant \|\nabla f(x_\delta^*)\| + \|\nabla f(x_T) - \nabla f(x_\delta^*)\|$$

$$\leqslant \delta R_0 + L\|x_T - x_\delta^*\|$$

$$\leqslant \delta R_0 + L\left[\frac{2}{\delta}\left(f_\delta(x_T) - f_\delta(x_\delta^*)\right)\right]^{\frac{1}{2}}$$

$$\overset{(12.2.23)}{\leqslant} \delta R_0 + L\left[\frac{L+2\delta}{\delta} R_0^2 e^{-T\sqrt{q_f}}\right]^{\frac{1}{2}}$$

取 $\delta R_0 = \frac{1}{2}\varepsilon$,则 $\frac{1}{q_f} = 1 + \frac{L}{\delta} = 1 + \frac{2LR_0}{\varepsilon}$

故要保证 $\|\nabla f(x_T)\| < \varepsilon$,只需

$$LR_0\left[\frac{L+2\delta}{\delta}\right]^{\frac{1}{2}} \leqslant \frac{\varepsilon}{2} e^{T\frac{\sqrt{q_f}}{2}}$$

即 $\quad T \geqslant \frac{2}{\sqrt{q_f}} \ln\left(\left(\frac{1}{q_f} - 1\right)\left(1 + \frac{1}{q_f}\right)^{\frac{1}{2}}\right)$

而 $\frac{2}{\sqrt{q_f}} \ln\left(\frac{1}{q_f}\right) > \frac{2}{\sqrt{q_f}} \ln\left(\left(\frac{1}{q_f} - 1\right)\left(1 + \frac{1}{q_f}\right)^{\frac{1}{2}}\right)$

定理 2.2.7:

令 $f\in\mathcal{F}_L^{1,1}(\mathbb{R}^n)$, $\delta=\frac{\varepsilon}{2R_0}$, 则迭代步数T使得

(2.2.22)作用在 f_δ 上, s.t. $\|\nabla f(x_T)\|\le\varepsilon$, 有

$$T\le 3\sqrt{1+\frac{2LR_0}{\varepsilon}}\,\ln\left(1+\frac{2LR_0}{\varepsilon}\right)$$

注: 对于 $\min\limits_{0\le k\le T}\|\nabla f(x_k)\|^2 = O\left(\frac{L^2}{k^3}\right)$

对一般的Nesterov算法也成立, 但系数不同

Theorem 6 *Let $f\in\mathcal{F}_L^1(\mathbb{R}^n)$. For any step size $0<s\le 1/(3L)$, the iterates $\{x_k\}_{k=0}^{\infty}$ generated by NAG-C obey*

$$\min_{0\le i\le k}\|\nabla f(x_i)\|^2\le\frac{8568\,\|x_0-x^\star\|^2}{s^2(k+1)^3},$$

for all $k\ge 0$. In additional, we have $f(x_k)-f(x^\star)\le\frac{119\|x_0-x^\star\|^2}{s(k+1)^2}$ for all $k\ge 0$.

所以方法的单调性不是 $\min\limits_{0\le k\le T}\|\nabla f(x_k)\|^2$

收敛速度的关键

§ 2.2.3 Convex set

引理 2.2.5

若 $f(\cdot)$ 是 $\mathbb{R}^n$ 上凸 func，则对 $\forall \beta \in \mathbb{R}$

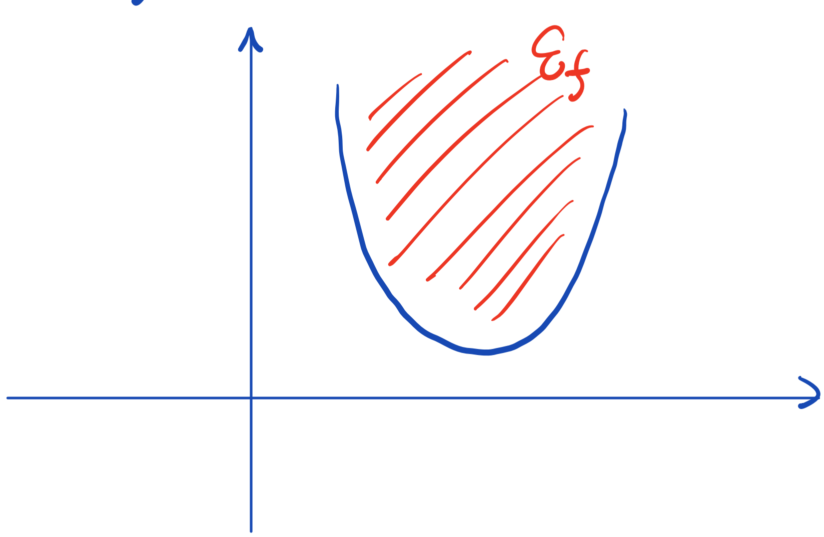
$$\mathcal{L}_f(\beta) = \{x \in \mathbb{R}^n \mid f(x) \le \beta\} \quad \text{凸/空}$$

证明简单

引理 2.2.6

若 $f(\cdot)$ 是 $\mathbb{R}^n$ 上凸 func，则上图

$$\mathcal{E}_f = \{(x,\tau) \in \mathbb{R}^{n+1} \mid f(x) \le \tau\} \quad \text{凸}$$



Theorem 2.2.8 *Let $Q_1 \subseteq \mathbb{R}^n$ and $Q_2 \subseteq \mathbb{R}^m$ be closed convex sets, and $\mathscr{A}(\cdot)$ be a linear operator:*

$$\mathscr{A}(x) = Ax + b : \mathbb{R}^n \to \mathbb{R}^m.$$

1. *The intersection of two sets ($m = n$), $Q_1 \bigcap Q_2 = \{x \in \mathbb{R}^n \mid x \in Q_1, x \in Q_2\}$, is convex and closed.*
2. *The sum of two sets ($m = n$), $Q_1 + Q_2 = \{z = x + y \mid x \in Q_1, y \in Q_2\}$, is convex. It is closed provided that one of the sets is bounded.*
3. *The direct product of two sets, $Q_1 \times Q_2 = \{(x, y) \in \mathbb{R}^{n+m} \mid x \in Q_1, y \in Q_2\}$ is convex and closed.*
4. *The conic hull of a set, $\mathscr{K}(Q_1) = \{z \in \mathbb{R}^n \mid z = \beta x, x \in Q_1, \beta \geq 0\}$, is convex. It is closed if the set Q_1 is bounded and does not contain the origin.*
5. *The convex hull of two sets,*

$$\mathrm{Conv}(Q_1, Q_2) = \{z \in \mathbb{R}^n \mid z = \alpha x + (1-\alpha)y, x \in Q_1, y \in Q_2, \alpha \in [0,1]\},$$

   is convex. It is closed if both sets are bounded.
6. *The affine image of a set, $\mathscr{A}(Q_1) = \{y \in \mathbb{R}^m \mid y = \mathscr{A}(x), x \in Q_1\}$, is convex and closed.* (若Q_1有界)
7. *The inverse affine image: $\mathscr{A}^{-1}(Q_2) = \{x \in \mathbb{R}^n \mid \mathscr{A}(x) \in Q_2\}$ is convex. It is closed ~~if Q_2 is bounded.~~*

证明:

① 显然

② 设 $z_1 = x_1 + y_1$, $z_2 = x_2 + y_2$, $x_i \in Q_1, y_i \in Q_2$

则 $\alpha z_1 + (1-\alpha) z_2 = \alpha x_1 + (1-\alpha) y_2 +$

$\alpha y_1 + (1-\alpha) y_2 \in Q_1 + Q_2$ 凸 ✓

不妨设 Q_2 有界，考虑序列 $z_k = x_k + y_k \to \bar{z}$

由 Q_2 有界，故 $\exists \{n_k\}_{k=1}^{\infty} \subset \{k\}_{k=1}^{\infty}$, $y_{n_k} \to \bar{y}$，故 $x_{n_k} \to \bar{z} - \bar{y}$

由 Q_1 与 Q_2 闭，故 $Q_1 + Q_2$ 闭

③ 显然

④ 设 $z_1=\beta_1 x_1$，$z_2=\beta_2 x_2$，则

$$\alpha z_1+(1-\alpha)z_2=\alpha\beta_1 x_1+(1-\alpha)\beta_2 x_2$$

$$=\gamma(\bar{\alpha}x_1+(1-\bar{\alpha})x_2)$$

$\gamma=\alpha\beta_1+(1-\alpha)\beta_2$，$\bar{\alpha}=\frac{\alpha\beta_1}{\gamma}\in[0,1]$，凸 ✓

考虑序列 $\{z_k=\beta_k x_k\to z\}$

若 Q_1 有界，则 $\{x_k\}$ 有界，若 $0\notin Q_1$，则 $\{\beta_k\}$ 有界

故 W.L.O.G.，设 $\{x_k\}$，$\{\beta_k\}$ 收敛，故 $z\in K(Q_1)$

⑤ 设 $z_1=\beta_1 x_1+(1-\beta_1)y_1$

$z_2=\beta_2 x_2+(1-\beta_2)y_2$

则 $\alpha z_1+(1-\alpha)z_2=\bar{\alpha}(\bar{\beta}_1 x_1+(1-\bar{\beta}_1)x_2)$

$$+(1-\bar{\alpha})(\bar{\beta}_2 y_1+(1-\bar{\beta}_2)y_2)$$

凸 ✓

取一列 $z_k=\beta_k x_k+(1-\beta_k)y_k\to z$

由 Q_1，Q_2 的有界性，证明完全类似④

⑥ 凸性显然，下证 $A(Q_1)$ 的闭性

任取一列 $\{y_k = Ax_k \to \bar{y}\}$，则由 Q_1 的有界性

W.L.O.G，$\{x_k\} \to \bar{x} \in Q_1$，故 $\bar{y} \in A(Q_1)$

⑦ 凸性显然，闭性由于连续 $\Leftrightarrow$ 闭集的原像是闭集

下面要研究带约束的优化问题

$$\min_{x \in Q} f(x),\quad f \in \mathcal{F}^1(Q, \|\cdot\|) \qquad (2.2.38)$$

的最优性条件：

<u>定理 2.2.9</u> 令 $f \in \mathcal{F}^1(Q)$，Q 凸 + 闭，x^* 是 (2.2.38) 的解 $\Leftrightarrow$ $\langle \nabla f(x^*), x - x^* \rangle \geq 0 \quad \forall x \in Q$

证明：

"$\Leftarrow$" $f(x) \geq f(x^*) + \langle \nabla f(x^*), x - x^* \rangle$

$$\geq f(x^*) \quad \forall x \in Q$$

"$\Rightarrow$" 设 $\exists x \in \Omega$, s.t. $\langle \nabla f(x^*), x - x^* \rangle < 0$

则令 $\phi(\alpha) = f(x^* + \alpha(x - x^*))$

$$\phi(0) = f(x^*)$$

$$\phi'(0) = \langle \nabla f(x^*), x - x^* \rangle < 0$$

则对充分小 α，有

$$x^* + \alpha(x - x^*) \in \Omega$$

且 $f(x^* + \alpha(x - x^*)) = \phi(\alpha)$

$$< \phi(0)$$

$$= f(x^*)$$

$\square$

Ω 的闭性是为了不让问题退化成一般的最优性条件 $(\nabla f(x^*) = 0)$

推论 2.2.2：令 $f \in C_L^{1,1}(\mathbb{R}^n, \|\cdot\|)$，则对 $x_1^*, x_2^* \in X^*$，

有：$\nabla f(x_1^*) = \nabla f(x_2^*)$，

$$\langle \nabla f(x_1^*), x_1^* \rangle = \langle \nabla f(x_2^*), x_2^* \rangle$$

证明：显然

$$\langle \nabla f(x_1^*), x_2^* - x_1^* \rangle \overset{(2.2.39)}{\geq} 0$$

$$\langle \nabla f(x_2^*), x_1^* - x_2^* \rangle \overset{(2.2.39)}{\geq} 0$$

$$0 \geq \langle \nabla f(x_1^*) - \nabla f(x_2^*), x_1^* - x_2^* \rangle$$

$$\overset{(2.1.11)}{\geq} \frac{1}{L} \| \nabla f(x_1^*) - \nabla f(x_2^*) \|_*^2 \geq 0$$

良定义

故 $\nabla f(x_1^*) = \nabla f(x_2^*)$，从而对 $x^* \in X$，令 $g^* = \nabla f(x^*)$

$$0 \overset{(2.2.39)}{\geq} \langle \nabla f(x_2^*), x_2^* - x_1^* \rangle$$

$$= \langle g^*, x_2^* - x_1^* \rangle$$

$$= \langle \nabla f(x_1^*), x_2^* - x_1^* \rangle \geq 0$$

□

定理 2.2.10：令 $f \in \mathcal{S}_\mu^1(Q, \|\cdot\|)$ $(\mu > 0)$，Q 闭凸

则存在问题 (2.2.38) 的唯一解 x^*

证明：令 $x_0 \in Q$，定义 $\bar{Q} = \{x \in Q \mid f(x) \le f(x_0)\}$

则 (2.2.38) $\Leftrightarrow$ $\min\{f(x) \mid x \in \bar{Q}\}$

集合 $\bar{Q}$ 是有界的：对 $\forall x \in \bar{Q}$，有：

$$f(x_0) \geq f(x) \geq f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \frac{\mu}{2}\|x - x_0\|^2$$

故 $\frac{\mu}{2}\|x - x_0\|^2 \leq \langle \nabla f(x_0), x_0 - x \rangle$

$$\leq \|\nabla f(x_0)\|_* \|x_0 - x\|$$

故 $\|x - x_0\| \leq \frac{2}{\mu}\|\nabla f(x_0)\|_*$ (有界)

紧集上的连续 func 有最小值，故 x^* 存在，下证

唯一性，设 x_1^* 也是 (2.2.38) 的最优点，则：

$$f^* = f(x_1^*) \geq f^* + \frac{\mu}{2}\|x_1^* - x^*\|^2$$

$\Rightarrow$ $x_1^* = x^*$ $\square$

注：Q一定是闭集吗？

答：是。要用到 紧集上的连续 func 有最小值

就要证 $\bar{Q}$ 紧. 若 Q 闭，则 $\forall$ Cauchy $\{x_k\} \subset \bar{Q}$

$x_k \xrightarrow{\|\cdot\|} \bar{x}$，则 $\{x_k\} \subset Q$，则 $\bar{x} \in Q$，且由 $f(x_k) \le f(x_0)$

及 f 连续，故 $\lim_{k\to\infty} f(x_k) = f(\bar{x}) \le f(x_0)$，则 $\bar{x} \in \bar{Q}$

即证 $\bar{Q}$ 闭。但 Q 不闭，不能证明 $\bar{x} \in Q$

例：$f(x) = x^2$，$Q = (-\infty, 0)$，不存在最小值

<u>例 2.2.4</u> 令 $f \in \mathcal{S}^1_\mu(Q, \|\cdot\|_p)$，考虑原问题：

$$f^* = \min_{x\in Q} \{ f(x) : Ax = b \}$$

$A \in \mathbb{R}^{m\times n}$，$b \in \mathbb{R}^m$，那么 Lagrangian：

$$\mathcal{L}(x, u) = f(x) + \langle u, b - Ax \rangle, \quad x \in Q, u \in \mathbb{R}^m$$

对偶 func：$\phi(u) = \min_{x\in Q} \mathcal{L}(x, u)$，由定理 2.2.10，

对 $\forall u \in \mathbb{R}^m$，$\phi(u)$ 是良定义的，令 $x(u) = \arg\min_{x \in Q} L(x,u)$

$\in Q$

引理：设 Q 紧，令 $\bar{\lambda} \in \mathbb{R}^m$，设 $X^*(\bar{\lambda}) = \{\bar{x}\}$ 是单点集

则若 $\lambda_k \to \bar{\lambda}$，则设 $x_k \in X^*(\lambda_k)$，有 $x_k \to \bar{x}$

定理：上引理条件下，ϕ 在 $\bar{\lambda}$ 处可微，且 $\nabla\phi(\lambda)\big|_{\lambda=\bar{\lambda}} = f(\bar{x})$

由定理 2.2.10 中的证明

$$\min_{x \in Q} \{ f(x) : Ax = b \}$$

$$\Leftrightarrow \min_{x \in \bar{Q}} \{ f(x) : Ax = b \}，且 \bar{Q} 紧，$$

再用定理 2.2.10，$X^*(\bar{\lambda}) = \{\bar{x}\}$ 是单点集成立 $\forall \bar{\lambda}$，故

$\nabla \phi(u) = b - Ax(u)$，$\forall u$，故对 $\forall u_1, u_2 \in \mathbb{R}^m$

令 $g(u) = b - Ax(u)$

$$\phi(u_1) = f(x(u_1)) + \langle u_1, b - Ax(u_1)\rangle$$
$$\le f(x(u_2)) + \langle u_1, b - Ax(u_2)\rangle$$
$$= \phi(u_2) + \langle u_1 - u_2, g(u_2)\rangle$$

在 $\mathbb{R}^m$ 上引入范数 $\|\cdot\|_d$，定义：

$$\|A\|_{p,d} = \max_{x,u}\{\langle Ax, u\rangle : \|x\|_p \le 1, \|u\|_d \le 1\}$$
$$= \max_{x,u}\{\langle x, A^T u\rangle : \|x\|_p \le 1, \|u\|_d \le 1\}$$
$$\overset{(2.1.6)}{=} \max_{u}\{\|A^T u\|_{p*} : \|u\|_d \le 1\} \quad (\text{由可达范})$$

注：$\min\limits_{\substack{x\in X\\ y\in Y}} f(x,y) = \min\limits_{x\in X}\min\limits_{y\in Y} f(x,y)$ 始终成立

$$= \min_{y\in Y}\min_{x\in X} f(x,y)$$

证：取 $x_0 \in X, y_0 \in Y$，则

$$f(x_0, y_0) \ge \min_{y\in Y} f(x_0, y) \ge \min_{x\in X}\min_{y\in Y} f(x,y)$$

取 $x_0, y_0 = \operatorname*{argmin}\limits_{x\in X, y\in Y} f(x,y)$

则 $\min\limits_{\substack{x\in X\\ y\in Y}} f(x,y) \geq \min\limits_{x\in X}\min\limits_{y\in Y} f(x,y)$

反之，由 $\min\limits_{\substack{x\in X\\ y\in Y}} f(x,y) \leq f(x_0,y_0)$，取 $y_0=\arg\min\limits_{y\in Y} f(x_0,y)$

$\Rightarrow \quad \min\limits_{\substack{x\in X\\ y\in Y}} f(x,y) \leq \min\limits_{y\in Y} f(x_0,y)$

再取 $x_0=\arg\min\limits_{x\in X}\min\limits_{y\in Y} f(x,y)$，故

$$\min\limits_{\substack{x\in X\\ y\in Y}} f(x,y) \leq \min\limits_{x\in X}\min\limits_{y\in Y} f(x,y)$$

故 $\min\limits_{\substack{x\in X\\ y\in Y}} f(x,y) = \min\limits_{x\in X}\min\limits_{y\in Y} f(x,y)$ $\boxed{\checkmark}$

max 同理

则对 $\forall u_1,u_2\in\mathbb{R}^n$，由 $\nabla_x L(x,u)=\nabla f(x)-A^Tu$

故 $\langle \nabla f(x(u_2)), x(u_1)-x(u_2)\rangle \overset{(2.2.39)}{\geq} \langle A^Tu_2, x(u_1)-x(u_2)\rangle$

故

$$\phi(u_1) = f(x(u_1)) + \langle u_1, b - Ax(u_1)\rangle$$

$$\overset{(2.1.20)}{\geqslant} f(x(u_2)) + \langle \nabla f(x(u_2)), x(u_1) - x(u_2)\rangle + \frac{\mu}{2}\|x(u_1) - x(u_2)\|_p^2 + \langle u_1, b - Ax(u_1)\rangle$$

$$\overset{(2.2.44)}{\geqslant} f(x(u_2)) + \langle u_2, A(x(u_1) - x(u_2))\rangle + \frac{\mu}{2}\|x(u_1) - x(u_2)\|_p^2 + \langle u_1, b - Ax(u_1)\rangle$$

$$= \phi(u_2) + \langle g(u_2), u_1 - u_2\rangle - \langle u_1 - u_2, A(x(u_1) - x(u_2))\rangle + \frac{\mu}{2}\|x(u_1) - x(u_2)\|_p^2$$

$$\geqslant \phi(u_2) + \langle g(u_2), u_1 - u_2\rangle - \frac{1}{2\mu}\|A^T(u_1 - u_2)\|_{p*}^2$$

由 $g(u_2) = \nabla\phi(u_2)$，且

$$\|A^T(u_1 - u_2)\|_{p*} \leqslant \|A\|_{p,d}\|u_1 - u_2\|_d$$

证：只需证 $\|A\|_{p,d} \geqslant \left\|A^T \frac{u_1 - u_2}{\|u_1 - u_2\|_d}\right\|_{p*}$

由 $\|A\|_{p,d} = \max\limits_{u}\{\|A^T u\|_{p*} : \|u\|_d \leq 1\}$，即证

故 $-\phi \in \mathcal{F}_L^{1,1}(\mathbb{R}^m, \|\cdot\|_d)$，$L = \frac{1}{\mu}\|A\|_{p,d}^2$

现在用任意求解光滑凸func的方法解：

$$\min_{u\in\mathbb{R}^m}\{-\phi(u)\} \quad \cdots\cdots \quad (2.2.45)$$

不妨设 u^* 存在，则

$0 = \nabla\phi(u^*) = b - Ax(u^*)$，故 $x(u^*)$ 是 (2.2.43)

的可行解，下证最优性：

$$f^* \overset{(1.3.6)}{\geq} f_* \triangleq \max_{u\in\mathbb{R}^m}\phi(u)$$

$$= f(x(u^*)) + \langle u^*, \nabla\phi(u^*)\rangle$$

$$= f(x(u^*)) \geq f^*$$

故 $f^* = f_*$，$x(u^*)$ 是 (2.2.43) 的最优解

现在设 $\bar{u}\in\mathbb{R}^m$ 是 (2.2.45) 的近似解，则

$$f(x(\bar{u})) - f^* = \phi(\bar{u}) - \langle \bar{u}, \nabla\phi(\bar{u})\rangle - \phi(u^*)$$

$$\leq \|\bar{u}\|_d \|\nabla\phi(\bar{u})\|_{d*}$$

又显然

$$\|\nabla\phi(\bar{u})\|_{d*} = \|b - Ax(\bar{u})\|_{d*}$$

故

$$\underbrace{\|b-Ax(\bar{u})\|_{d*}}_{\text{逼近可行域的速度}} = \|\nabla\phi(\bar{u})\|_{d*} \geq \frac{1}{\|\bar{u}\|_d}\underbrace{(f(x(\bar{u}))-f^*)}_{\text{收敛速度}}$$

注：能否将约束推广到非线性？

答：不可以，此时可行集 Q 一定不是凸集，故不再是一个凸优化问题，因为设问题

$$\min_{x\in\mathbb{R}^n} f(x) \qquad f\in\mathcal{F}^1(\mathbb{R}^n)$$

$$\text{s.t.} \quad g(x)=0 \quad , g\text{非线性}$$

故可行集在 $\mathbb{R}^n$ 空间上是一个$\underbrace{\text{超曲面}}_{\text{(流形)}}$，一定不凸，严格证明：

$$g(x)=0 \iff \begin{cases} g(x)\geqslant 0 \\ -g(x)\geqslant 0 \end{cases}$$

为了保证 Q 凸，则 $g(x),-g(x)$ 均凸，故 $g(x)$ 只能线性

注：例如 $(ax-b)^2=0$ 算做线性约束

定义：Q 是闭集，$x_0\in\mathbb{R}^n$，定义

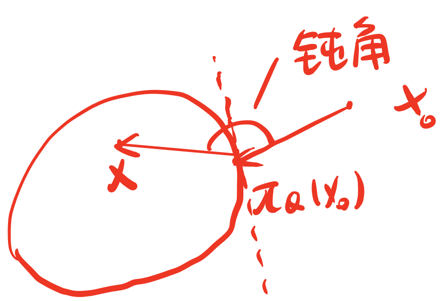
$$\pi_Q(x_0)=\arg\min_{x\in Q}\|x-x_0\|$$

定理 2.2.11 Q 凸，则存在唯一 $\pi_Q(x_0)$

易证

引理 2.2.7 Q 是闭凸集，$x_0\notin Q$，则对 $\forall x\in Q$，

有：$\langle \pi_Q(x_0)-x_0,\ x-\pi_Q(x_0)\rangle\geqslant 0$



证明：$\pi_Q(x_0)$ 是 $\min\limits_{x\in Q} f(x)$，$f(x)=\frac{1}{2}\|x-x_0\|^2$

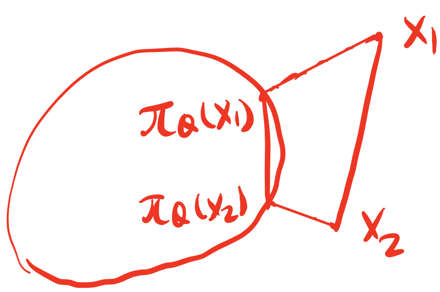
的解，由定理 2.2.9

$$\langle \underbrace{\nabla f(\pi_Q(x_0))}_{=\pi_Q(x_0)-x_0},\ x-\pi_Q(x_0)\rangle \geqslant 0$$

$\square$

推论 2.2.3：$x_1, x_2 \in \mathbb{R}^n$，

$$\|\pi_Q(x_1)-\pi_Q(x_2)\| \leqslant \|x_1-x_2\|$$



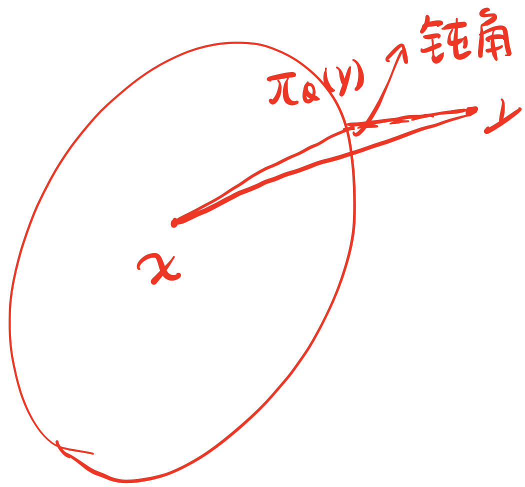
证明：$\langle \pi_Q(x_1)-x_1,\ \pi_Q(x_2)-\pi_Q(x_1)\rangle \geqslant 0$

$$\langle \pi_Q(x_2)-x_2,\ \pi_Q(x_1)-\pi_Q(x_2)\rangle \geqslant 0$$

相加即证 $\square$

引理 2.2.8 $x\in Q$, $y\in\mathbb{R}^n$

$$\|x-\pi_Q(y)\|^2+\|\pi_Q(y)-y\|^2\le\|x-y\|^2$$



由上图是显然的 $\boxtimes$

定理 2.2.12 令 x^* 是 (2.2.38) 的 optimal，则对 $\forall \gamma>0$，有 $\pi_Q(x^*-\frac{1}{\gamma}\nabla f(x^*))=x^*$

证明：考虑 $\min\limits_{x\in Q}\frac{1}{2}\|x-x^*+\frac{1}{\gamma}\nabla f(x^*)\|^2$，则存在唯一 optimal x_*，又由定理 2.2.9 $\Leftrightarrow$

$$\langle x_*-x^*+\frac{1}{\gamma}\nabla f(x^*), x-x_*\rangle\ge 0\quad\forall x\in Q$$

故 $x_*=x^*$ $\boxtimes$

§2.2.3

定义 $\rho_Q(x) \doteq \frac{1}{2}\|x-\pi_Q(x)\|^2, \ x \in \mathbb{R}^n$

引理 2.2.9　ρ_Q 在 $\mathbb{R}^n$ 上是凸的、可微的，且

$$\nabla \rho_Q(x) = x - \pi_Q(x), \ x \in \mathbb{R}^n$$

是 Lipschiz 连续的（关于标准欧式 norm），且 Lipschiz 常数是 1

证明：固定任意两点 $x_1, x_2 \in \mathbb{R}^n$，令：

$$\pi_1 = \pi_Q(x_1) \in Q, \ \pi_2 = \pi_Q(x_2) \in Q$$
$$g_1 = x_1 - \pi_1, \ g_2 = x_2 - \pi_2$$

$$\frac{1}{2}\|g_2\|^2 = \frac{1}{2}\|g_1\|^2 + \langle g_1, g_2 - g_1\rangle + \frac{1}{2}\|g_2 - g_1\|^2$$

故
$$\begin{aligned}\rho_Q(x_2) &\geq \rho_Q(x_1) + \langle x_1 - \pi_Q(x_1), x_2 - x_1\rangle \\ &\quad + \langle \pi_Q(x_1) - x_1, \pi_Q(x_2) - \pi_Q(x_1)\rangle \\ &\overset{(2.2.47)}{\geq} \rho_Q(x_1) + \langle g_1, x_2 - x_1\rangle\end{aligned}$$

另一方面：

$$\rho_Q(x_2)-\rho_Q(x_1)\overset{(2.2.53)}{=}\langle g_1, g_2-g_1\rangle+\frac{1}{2}\|g_2-g_1\|^2$$

$$=\langle g_1, x_2-x_1\rangle+\langle g_1, \pi_1-\pi_2-g_2\rangle+\frac{1}{2}\|g_1\|^2+\frac{1}{2}\|g_2\|^2$$

$$\overset{(2.2.46)}{\leq}\langle g_1, x_2-x_1\rangle+\langle g_1, \pi_1-x_2\rangle+\frac{1}{2}\|g_1\|^2+\frac{1}{2}\|x_2-\pi_1\|^2$$

$$=\langle g_1, x_2-x_1\rangle+\frac{1}{2}\|x_2-x_1\|^2$$

故对 $\forall x_1, x_2\in\mathbb{R}^n$，有

$$\langle g_1, x_2-x_1\rangle\leq\rho_Q(x_2)-\rho_Q(x_1)$$
$$\leq\langle g_1, x_2-x_1\rangle+\frac{1}{2}\|x_2-x_1\|^2$$

即证引理 2.2.9 $\square$

§ 2.2.4 The gradient Mapping

定义 2.2.3 固定 $\gamma>0$，定义

$$x_Q(\bar{x};\gamma)=\underset{x\in Q}{\operatorname{argmin}}\left[f(\bar{x})+\langle\nabla f(\bar{x}), x-\bar{x}\rangle+\frac{\gamma}{2}\|x-\bar{x}\|^2\right]$$

$$g_Q(\bar{x};\gamma)=\gamma(\bar{x}-x_Q(\bar{x};\gamma))$$

$x_Q(\bar{x};\gamma)$ 等价于极小化：

$$\underbrace{f(\bar{x})}_{} + \frac{\gamma}{2}\|x-\bar{x}+\frac{1}{\gamma}\nabla f(\bar{x})\|^2 - \underbrace{\frac{1}{2\gamma}\|\nabla f(\bar{x})\|^2}_{}$$

都是定值，故 $x_Q(\bar{x};\gamma)$ 可以看作 $\bar{x}-\frac{1}{\gamma}\nabla f(\bar{x})$ 到可行集 Q 的投影

故 $Q\equiv\mathbb{R}^n$ 时：

$$x_Q(\bar{x};\gamma)=\bar{x}-\frac{1}{\gamma}\nabla f(\bar{x}),\quad g_Q(\bar{x};\gamma)=\nabla f(\bar{x})$$

注：gradient mapping $x_Q(\bar{x};\gamma)$ 中，$\bar{x}\in\mathbb{R}^n$，而不是 $x\in Q$

<u>定理 2.2.13</u> 令 $f\in\mathscr{S}^{1,1}_{\mu,L}(Q)$，$\gamma\geq L$，$\bar{x}\in\mathbb{R}^n$，则对 $\forall x\in Q$，有：

$$f(x)\geq f(x_Q(\bar{x};\gamma))+\langle g_Q(\bar{x};\gamma),x-\bar{x}\rangle$$
$$+\frac{1}{2\gamma}\|g_Q(\bar{x};\gamma)\|^2+\frac{\mu}{2}\|x-\bar{x}\|^2$$

证明：令 $x_Q=x_Q(\bar{x};\gamma)$，$g_Q=g_Q(\bar{x};\gamma)$

$$\phi(x)=f(\bar{x})+\langle\nabla f(\bar{x}),x-\bar{x}\rangle+\frac{\gamma}{2}\|x-\bar{x}\|^2$$

则 $\nabla\phi(x)=\nabla f(\bar{x})+\gamma(x-\bar{x})$，对 $\forall x\in Q$，有：

$$\langle \nabla f(\bar{x}) - g_Q, x - x_Q \rangle = \langle \nabla \phi(x_Q), x - x_Q \rangle \overset{(2.2.39)}{\geq} 0 \quad ①$$

故 $f(x) - \frac{\mu}{2}\|x - \bar{x}\|^2 \geq f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle$

$$= f(\bar{x}) + \langle \nabla f(\bar{x}), x_Q - \bar{x} \rangle + \langle \nabla f(\bar{x}), x - x_Q \rangle$$

$$\overset{①}{\geq} f(\bar{x}) + \langle \nabla f(\bar{x}), x_Q - \bar{x} \rangle + \langle g_Q, x - x_Q \rangle$$

$$= \phi(x_Q) - \frac{\gamma}{2}\|x_Q - \bar{x}\|^2 + \langle g_Q, x - x_Q \rangle$$

$$= \phi(x_Q) + \frac{1}{2\gamma}\|g_Q\|^2 + \langle g_Q, x - \bar{x} \rangle$$

且 $\phi(x_Q) \geq f(\bar{x}) + \langle \nabla f(\bar{x}), x_Q - \bar{x} \rangle + \frac{L}{2}\|x_Q - \bar{x}\|^2$

$$\geq f(x_Q)$$

☑

<u>推论 2.2.4</u> 令 $f \in \mathcal{S}_{\mu,L}^{1,1}(Q)$, $\gamma \geq L$, $\bar{x} \in Q$, 则

$$f(x_Q(\bar{x};\gamma)) \leq f(\bar{x}) - \frac{1}{2\gamma}\|g_Q(\bar{x};\gamma)\|^2 \qquad (2.2.58)$$

$$\langle g_Q(\bar{x};\gamma), \bar{x} - x^* \rangle \geq \frac{1}{2\gamma}\|g_Q(\bar{x};\gamma)\|^2 + \frac{\mu}{2}\|\bar{x} - x^*\|^2$$

$$+ \frac{\mu}{2}\|x_Q(\bar{x};\gamma) - x^*\|^2 \qquad (2.2.59)$$

证明:

(2.2.57) 中令 $x=\bar{x}$，则 $\Rightarrow$ (2.2.58)

(2.2.57) 中令 $x=x^*$，则 $\Rightarrow$

$$f(x^*) \geq f(x_Q(\bar{x};\gamma)) + \langle g_Q(\bar{x};\gamma), x^*-\bar{x}\rangle + \frac{1}{2\gamma}\|g_Q(\bar{x};\gamma)\|^2 + \frac{\mu}{2}\|x^*-\bar{x}\|^2$$

又由 $f(x_Q(\bar{x};\gamma)) \overset{(2.2.40)}{\geq} f(x^*) + \frac{\mu}{2}\|x_Q(\bar{x};\gamma)-x^*\|^2$

$\square$

§2.2.5 Minimization over Simple Sets

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

设: ① $f \in \mathcal{S}^{1,1}_{\mu,L}(Q)$

② Q是闭凸集，且足够简单（使得 gradient mapping 可以用一个 closed form 表示出来）

Gradient Method for Simple Set
0. 取 $x_0 \in Q$，$\gamma > 0$ 1. kth 迭代 $(k \geq 0)$ $x_{k+1} = x_k - \frac{1}{\gamma} g_Q(x_k;\gamma)$

(2.2.60)

注：$x_{k+1} = x_Q(x_k;\gamma) = \pi_Q\left(x_k - \frac{1}{\gamma}\nabla f(x_k)\right)$

定理 2.2.14 令 $f \in \mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$，若 (2.2.60) 中 $\gamma \geq \frac{L+\mu}{2}$，

$$\|x_k - x^*\| \leq \left(1 - \frac{\mu}{\gamma}\right)^k \|x_0 - x^*\|$$

证明：令 $r_k = \|x_k - x^*\|$，则由定理 2.2.12

$$r_{k+1}^2 = \left\|\pi_Q\left(x_k - \frac{1}{\gamma}\nabla f(x_k)\right) - \pi_Q\left(x^* - \frac{1}{\gamma}\nabla f(x^*)\right)\right\|^2$$

$$\overset{(2.2.48)}{\leq} \left\|x_k - x^* - \frac{1}{\gamma}\left(\nabla f(x_k) - \nabla f(x^*)\right)\right\|^2$$

$$= r_k^2 - \frac{2}{\gamma}\langle \nabla f(x_k) - \nabla f(x^*), x_k - x^*\rangle + \frac{1}{\gamma^2}\|\nabla f(x_k) - \nabla f(x^*)\|^2$$

$$\overset{(2.1.32)}{\leq} \left(1 - \frac{2}{\gamma}\frac{\mu L}{\mu+L}\right) r_k^2 + \left(\frac{1}{\gamma^2} - \frac{2}{\gamma}\frac{1}{\mu+L}\right)\|\nabla f(x_k) - \nabla f(x^*)\|^2$$

$$\overset{(2.1.26)}{\leq} \left(1-\frac{2}{\gamma}\frac{\mu L}{\mu+L}+\mu^2\left(\frac{1}{\gamma^2}-\frac{2}{\gamma}\frac{1}{\mu+L}\right)\right) r_k^2$$

$$= \left(1-\frac{\mu}{\gamma}\right)^2 r_k^2 \qquad \square$$

取 $\gamma=\frac{L+\mu}{2}$，则：

$$\|x_k - x^*\| \leq \left(\frac{L-\mu}{L+\mu}\right)^k \|x_0 - x^*\|$$

注：与无约束问题收敛速率一致

设 $x_0 \in Q$，定义

$$\phi_0(x) = f(x_0) + \frac{\gamma_0}{2}\|x-x_0\|^2$$

$$\phi_{k+1}(x) = (1-\alpha_k)\phi_k(x) + \alpha_k\Big[f(x_Q(y_k;L)) + \frac{1}{2L}\|g_Q(y_k;L)\|^2$$

$$+\langle g_Q(y_k;L), x-y_k\rangle + \frac{\mu}{2}\|x-y_k\|^2\Big], \; k \geq 0$$

这一项在无约束问题的 ϕ_{k+1} 中没有

$\{\phi_k(x)\}$ 可以表示为：$\phi_k(x) = \phi_k^* + \frac{\gamma_k}{2}\|x-v_k\|^2$

其中：
$$\gamma_{k+1} = (1-\alpha_k)\gamma_k + \alpha_k\mu$$

$$v_{k+1} = \frac{1}{\gamma_{k+1}}\left[(1-\alpha_k)\gamma_k v_k + \alpha_k\mu y_k - \alpha_k g_Q(y_k;L)\right]$$

$$\phi_{k+1}^* = (1-\alpha_k)\phi_k^* + \alpha_k f(x_Q(y_k;L)) + \left(\frac{\alpha_k}{2L} - \frac{\alpha_k^2}{2\gamma_{k+1}}\right)\|g_Q(y_k;L)\|^2$$

$$+ \frac{\alpha_k(1-\alpha_k)\gamma_k}{\gamma_{k+1}}\left(\frac{\mu}{2}\|y_k - v_k\|^2 + \langle g_Q(y_k;L), v_k - y_k\rangle\right)$$

证明：注意到 $\phi_0(x) = \gamma_0 I_n$，下证 $\nabla^2\phi_k(x) = \gamma_k I_n$，用数归，设 k 成立

则 $\nabla^2\phi_{k+1}(x) = (1-\alpha_k)\nabla^2\phi_k(x) + \alpha_k \mu I_n$

$$= ((1-\alpha_k)\gamma_k + \alpha_k\mu) I_n$$

$$\equiv \gamma_{k+1} I_n$$

✓

故 $\phi_{k+1}(x) = (1-\alpha_k)\left(\phi_k^* + \frac{\gamma_k}{2}\|x - v_k\|^2\right)$

$$+ \alpha_k\left[f(x_Q(y_k;L)) + \frac{1}{2L}\|g_Q(y_k;L)\|^2 + \langle g_Q(y_k;L), x - y_k\rangle + \frac{\mu}{2}\|x - y_k\|^2\right]$$

故 $\nabla\phi_{k+1}(x) = (1-\alpha_k)\gamma_k(x - v_k) + \alpha_k g_Q(y_k;L)$

$$+ \alpha_k\mu(x - y_k) = 0$$

即可解得 $v_{k+1} = \frac{1}{\gamma_{k+1}}\left[(1-\alpha_k)\gamma_k v_k + \alpha_k\mu y_k - \alpha_k g_Q(y_k;L)\right]$

最后证明 ϕ_{k+1}^* 有上面的形式

$$\phi_{k+1}^* + \frac{\gamma_{k+1}}{2}\|y_k - v_{k+1}\|^2 = \phi_{k+1}(y_k)$$

①

$$= (1-\alpha_k)\left(\phi_k^* + \frac{\gamma_k}{2}\|y_k - v_k\|^2\right) + \alpha_k f(x_Q(y_k;L)) + \frac{\alpha_k}{2L}\|g_Q(y_k;L)\|^2$$

$$v_{k+1} - y_k = \frac{1}{\gamma_{k+1}}\left[(1-\alpha_k)\gamma_k(v_k - y_k) - \alpha_k g_Q(y_k;L)\right]$$

故 $\frac{\gamma_{k+1}}{2}\|y_k - v_{k+1}\|^2 = \frac{1}{2\gamma_{k+1}}\left[(1-\alpha_k)^2\gamma_k^2\|v_k - y_k\|^2\right.$

$\left. - 2\alpha_k(1-\alpha_k)\gamma_k\langle g_Q(y_k;L), v_k - y_k\rangle + \alpha_k^2\|g_Q(y_k;L)\|^2\right]$ ②

整理 ①②即证 $\boxed{\checkmark}$

进一步假设 $\phi_k^* \geq f(x_k)$，由

$$f(x_k) \overset{(2.2.57)}{\geq} f(x_Q(y_k;L)) + \langle g_Q(y_k;L), x_k - y_k\rangle + \frac{1}{2L}\|g_Q(y_k;L)\|^2 + \frac{\mu}{2}\|x_k - y_k\|^2$$

故 $\phi_{k+1}^* \geq (1-\alpha_k)f(x_k) + \alpha_k f(x_Q(y_k;L))$

$$+\left(\frac{\alpha_k}{2L} - \frac{\alpha_k^2}{2\gamma_{k+1}}\right)\|g_Q(y_k;L)\|^2 + \frac{\alpha_k(1-\alpha_k)\gamma_k}{\gamma_{k+1}}\langle g_Q(y_k;L), v_k - y_k\rangle$$

$$\geq f(x_Q(y_k;L)) + \left(\frac{1}{2L} - \frac{\alpha_k^2}{2\gamma_{k+1}}\right)\|g_Q(y_k;L)\|^2$$

$$+ (1-\alpha_k)\underbrace{\left\langle g_Q(y_k;L), \frac{\alpha_k\gamma_k}{\gamma_{k+1}}(v_k-y_k)+x_k-y_k\right\rangle}_{=0}$$

使这一项=0

故取
$$x_{k+1} = x_Q(y_k;L)$$

$$L\alpha_k^2 = (1-\alpha_k)\gamma_k + \alpha_k\mu \equiv \gamma_{k+1}$$

$$y_k = \frac{1}{\gamma_k+\alpha_k\mu}(\alpha_k\gamma_k v_k + \gamma_{k+1}x_k)$$

$$\Rightarrow \phi_{k+1}^* \geq f(x_{k+1})$$

故得到下面(2.2.20)的变体

Constant Step Scheme II for Simple Set

0. Choose $x_0 \in \mathbb{R}^n$ and $\alpha_0 \in \left[\sqrt{q_f}, \frac{2(3+q_f)}{3+\sqrt{21+4q_f}}\right]$. Set $y_0 = x_0$.

1. k**th iteration** ($k \geq 0$).

(a) Compute $f(y_k)$ and $\nabla f(y_k)$. Set $x_{k+1} = x_Q(y_k; L)$.

(b) Compute $\alpha_{k+1} \in (0,1)$ from the equation

$$\alpha_{k+1}^2 = (1-\alpha_{k+1})\alpha_k^2 + q_f\alpha_{k+1}.$$

Set $\beta_k = \frac{\alpha_k(1-\alpha_k)}{\alpha_k^2+\alpha_{k+1}}$ and $y_{k+1} = x_{k+1} + \beta_k(x_{k+1} - x_k)$. (2.2.63)

§ 2.3.1 The Minmax Problem

$$\min_{x\in Q}\left[f(x)=\max_{1\le i\le m} f_i(x)\right]$$

$f_i\in \mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n,\|\cdot\|)$, $i=1,\cdots,m$ ，Q是闭凸集

记 $f\in \mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n,\|\cdot\|)$若 $f_i\in \mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n,\|\cdot\|)$, $i=1,\cdots,m$

定义 2.3.1 令 f是 max-type function:

$$f(x)=\max_{1\le i\le m} f_i(x)$$

则 $$f(\bar{x};x)=\max_{1\le i\le m}\left[f_i(\bar{x})+\langle\nabla f_i(\bar{x}),x-\bar{x}\rangle\right]$$

叫做 f在$\bar{x}$的线性化

引理 2.3.1 对 $\forall x,\bar{x}\in\mathbb{R}^n$, 有:

$$f(x)\geqslant f(\bar{x};x)+\frac{\mu}{2}\|x-\bar{x}\|^2 \qquad (2.3.2)$$

$$f(x)\leqslant f(\bar{x};x)+\frac{L}{2}\|x-\bar{x}\|^2 \qquad (2.3.3)$$

证明: 对 $\forall i=1,\cdots,m$, 有:

$$f_i(x) \geqslant f_i(\bar{x}) + \langle \nabla f_i(\bar{x}), x-\bar{x} \rangle + \frac{\mu}{2}\|x-\bar{x}\|^2$$

两边同时取最大，即得 (2.3.2)，又由 $\forall i=1,\cdots,m$

$$f_i(x) \leqslant f_i(\bar{x}) + \langle \nabla f_i(\bar{x}), x-\bar{x} \rangle + \frac{L}{2}\|x-\bar{x}\|^2$$

两边同时取最大，即得 (2.3.3) $\boxed{/}$

定理 2.3.1 $x^* \in Q$ 是 (2.3.1) 的 optimal $\Longleftrightarrow$

对 $\forall x \in Q$，$f(x^*; x) \geqslant f(x^*, x^*) = f(x^*)$

证明：

$(\Leftarrow)$ $f(x) \geqslant f(x^*; x) \geqslant f(x^*, x^*) = f(x^*) \quad \forall x \in Q$

$(\Rightarrow)$ 用反证法，设 $\exists x \in Q$, s.t. $f(x^*; x) < f(x^*)$

考虑 $\phi_i(\alpha) = f_i(x^* + \alpha(x - x^*)) \quad i=1,\cdots,m$

注意到 $f_i(x^*) + \langle \nabla f_i(x^*), x - x^* \rangle$

$$\leqslant f(x^*; x) \overset{(2.3.2)}{<} f(x^*) = \max_{1\leqslant i\leqslant m} f_i(x^*)$$

故对 $\forall i=1,\cdots,m$

$$\begin{cases} \phi_i(0) \equiv f_i(x^*) < f(x^*) \\ \\ \phi_i(0) \equiv f_i(x^*) = f(x^*) \\ \phi_i'(0) = \langle \nabla f_i(x^*), x - x^* \rangle < f(x^*) - f_i(x^*) = 0 \end{cases}$$

二者择一发生，故对充分小的 α，有

$$f_i(x^* + \alpha(x - x^*)) = \phi_i(\alpha) < f(x^*) \quad \forall i = 1, \cdots, m$$

与 x^* 是 (2.3.1) 的 optimal 矛盾 $\boxed{/}$

推论 2.3.1 令 x^* 是 max-type func $f(\cdot)$ 在 Q 上的 minimum

若 $f \in \mathcal{S}_\mu^1(\mathbb{R}^n, \|\cdot\|)$，则

$$f(x) \geq f(x^*) + \frac{\mu}{2}\|x - x^*\|^2, \text{对 } \forall x \in Q$$

证明：对 $\forall x \in Q$，有

$$\begin{aligned} f(x) &\geq f(x^*; x) + \frac{\mu}{2}\|x - x^*\|^2 \\ &\geq f(x^*; x^*) + \frac{\mu}{2}\|x - x^*\|^2 \\ &= f(x^*) + \frac{\mu}{2}\|x - x^*\|^2 \end{aligned}$$

$\boxed{/}$

定理2.3.2 设 max-type func $f \in \mathcal{S}_\mu^1(\mathbb{R}^n, \|\cdot\|)$, $\mu > 0$

Q是闭凸集，则(2.3.1)的解 x^* 存在唯一

证明：令 $\bar{x} \in Q$，考虑 $\bar{Q} = \{x \in Q \mid f(x) \le f(\bar{x})\}$

$\bar{Q}$是闭的，且 (2.3.1) $\Leftrightarrow$ $\max\{f(x) \mid x \in \bar{Q}\}$

下证 $\bar{Q}$是有界的：对 $\forall x \in \bar{Q}$，有 $\forall i = 1, \cdots, m$

$$f(\bar{x}) \geqslant f_i(\bar{x}) \geqslant f_i(x) \geqslant f_i(\bar{x}) + \langle \nabla f_i(\bar{x}), x - \bar{x} \rangle + \frac{\mu}{2}\|x - \bar{x}\|^2$$

故
$$\begin{aligned}\frac{\mu}{2}\|x - \bar{x}\|^2 &\le \langle \nabla f_i(\bar{x}), \bar{x} - x \rangle + f(\bar{x}) - f_i(\bar{x}) \\ &\leqslant \|\nabla f_i(\bar{x})\|_* \|\bar{x} - x\| + f(\bar{x}) - f_i(\bar{x}) \quad (\forall i = 1, \cdots, m)\end{aligned}$$

令 $R = \|x - \bar{x}\|$，则

$$\frac{\mu}{2}R^2 - \|\nabla f_i(\bar{x})\|_* R + f_i(\bar{x}) - f(\bar{x}) \leqslant 0 \quad (\forall i = 1, \cdots, m)$$

开口向上的抛物线，且 $\Delta = \|\nabla f_i(\bar{x})\|_*^2 + 2\mu(f(\bar{x}) - f_i(\bar{x})) \geqslant 0$

故 R有界，即 $\bar{Q}$有界，从而 $\bar{Q}$紧，x^*存在

下证唯一性：若 x_1^* 也是 (2.3.1) 的 optimal，则

$$f(x^*) = f(x_1^*) \geq f(x^*; x_1^*) + \frac{\mu}{2} \| x_1^* - x^* \|$$

$$\geq f(x^*) + \frac{\mu}{2} \| x_1^* - x^* \|$$

$\Rightarrow x^* = x_1^*$

$\square$

§2.3.2 Gradient Mapping

固定$\gamma>0$，$\bar{x}\in\mathbb{R}^n$，对于max-type function f，定义

$$f_\gamma(\bar{x};x)=f(\bar{x};x)+\frac{\gamma}{2}\|x-\bar{x}\|^2$$

定义 2.3.2

$$f^*(\bar{x};\gamma)=\min_{x\in Q} f_\gamma(\bar{x};x)$$

$$x_f(\bar{x};\gamma)=\arg\min_{x\in Q} f_\gamma(\bar{x};x)$$

$$g_f(\bar{x};\gamma)=\gamma(\bar{x}-x_f(\bar{x};\gamma))$$

注：$f_\gamma(\bar{x},\cdot)$是一个max-type func，分支是：

$$f_i(\bar{x})+\underbrace{\langle\nabla f_i(\bar{x}),x-\bar{x}\rangle}_{\text{线性}}+\underbrace{\frac{\gamma}{2}\|\bar{x}-x\|^2}_{\text{二次}}\in \mathcal{S}^{1,1}_{\gamma,\gamma}(\mathbb{R}^n),\ i=1,2,\cdots,m$$

故由定理2.3.2：$x_f(\bar{x};\gamma)$是良定义的

定理 2.3.3：对$\forall x\in Q$，$\gamma\geqslant L$，$\bar{x}\in\mathbb{R}^n$，有：

$$f(\bar{x};x)\geqslant f^*(\bar{x};\gamma)+\langle g_f(\bar{x};\gamma),x-\bar{x}\rangle+\frac{1}{2\gamma}\|g_f(\bar{x};\gamma)\|^2$$

证明：令 $x_f = x_f(\bar{x};\gamma)$，$g_f = g_f(\bar{x};\gamma)$

显然 $f_\gamma(\bar{x};\cdot) \in \mathcal{F}^{1,1}_{\gamma,\gamma}(\mathbb{R}^n)$，且是 max-type function

$$f(\bar{x};x) = f_\gamma(\bar{x};x) - \frac{\gamma}{2}\|\bar{x}-x\|^2$$

(推论2.3.1) $\geq f_\gamma(\bar{x};x_f) + \frac{\gamma}{2}\|x-x_f\|^2 - \frac{\gamma}{2}\|\bar{x}-x\|^2$

$\|x\|^2-\|y\|^2 = \langle x+y, x-y\rangle$

$$= f^*(\bar{x};\gamma) + \frac{\gamma}{2}\langle \bar{x}-x_f, 2x-x_f-\bar{x}\rangle$$

$$= f^*(\bar{x};\gamma) + \frac{\gamma}{2}\langle \bar{x}-x_f, 2(x-\bar{x})+\bar{x}-x_f\rangle$$

$$= f^*(\bar{x};\gamma) + \langle g_f, x-\bar{x}\rangle + \frac{1}{2\gamma}\|g_f\|^2$$

$\boxtimes$

推论2.3.2：令 $f \in \mathcal{F}^{1,1}_{\mu,L}(\mathbb{R}^n)$，$\gamma \geq L$，则：

1) 对 $\forall x \in Q$，$\bar{x} \in \mathbb{R}^n$，有

$$f(x) \geq f(x_f(\bar{x};\gamma)) + \langle g_f(\bar{x};\gamma), x-\bar{x}\rangle + \frac{1}{2\gamma}\|g_f(\bar{x};\gamma)\|^2 + \frac{\mu}{2}\|x-\bar{x}\|^2$$

2) 若 $\bar{x} \in Q$，则

$$f(x_f(\bar{x};\gamma)) \leq f(\bar{x}) - \frac{1}{2\gamma}\|g_f(\bar{x};\gamma)\|^2$$

3) 对 $\forall \bar{x} \in \mathbb{R}^n$，有

$$\langle g_f(\bar{x};\gamma), \bar{x}-x^*\rangle \geqslant \frac{1}{2\gamma}\|g_f(\bar{x};\gamma)\|^2+\frac{\mu}{2}\|\bar{x}-x^*\|^2$$

证明：设 $\gamma \geqslant L$，则

$$\begin{aligned} f^*(\bar{x};\gamma) &= f_\gamma(\bar{x};x_f) \\ &= f(\bar{x};x_f)+\frac{\gamma}{2}\|\bar{x}-x_f\|^2 \\ \text{(引理 2.3.1)}\quad &\geqslant f(x_f)-\frac{L}{2}\|\bar{x}-x_f\|^2+\frac{\gamma}{2}\|\bar{x}-x_f\|^2 \\ &\geqslant f(x_f) \end{aligned}$$

故由 $f(x)\geqslant f(\bar{x};x)+\frac{\mu}{2}\|x-\bar{x}\|^2$，即证 1)

令 $x=\bar{x}$，即得 2)

令 $x=x^*$，由 $f(x_f(\bar{x};\gamma))\geqslant f(x^*)$，即证 3) $\square$

<u>引理 2.3.2</u>：对 $\forall \gamma_1,\gamma_2>0$，$\bar{x}\in\mathbb{R}^n$，有

$$f^*(\bar{x};\gamma_2)\geqslant f^*(\bar{x};\gamma_1)+\frac{\gamma_2-\gamma_1}{2\gamma_1\gamma_2}\|g_f(\bar{x};\gamma_1)\|^2$$

证明：令 $x_i=x_f(\bar{x};\gamma_i)$，$g_i=g_f(\bar{x};\gamma_i)$，$i=1,2$

由 (2.3.6) $$f(\bar{x};x)+\frac{\gamma_2}{2}\|x-\bar{x}\|^2 \geqslant f^*(\bar{x};\gamma_1)+\langle g_1, x-\bar{x}\rangle + \frac{1}{2\gamma_1}\|g_1\|^2+\frac{\gamma_2}{2}\|x-\bar{x}\|^2 \quad \forall x\in Q$$

令 $x=x_2$，得到：

$$
\begin{aligned}
f^*(\bar{x};\gamma_2) &= f(\bar{x};x_2)+\frac{\gamma_2}{2}\|x_2-\bar{x}\|^2 \\
&\geq f^*(\bar{x};\gamma_1)+\langle g_1, x_2-\bar{x}\rangle+\frac{1}{2\gamma_1}\|g_1\|^2+\frac{\gamma_2}{2}\|x_2-\bar{x}\|^2 \\
&= f^*(\bar{x};\gamma_1)-\underline{\frac{1}{\gamma_2}\langle g_1, g_2\rangle}+\frac{1}{2\gamma_1}\|g_1\|^2+\underline{\frac{1}{2\gamma_2}\|g_2\|^2} \\
&\geq f^*(\bar{x};\gamma_1)+\frac{1}{2\gamma_1}\|g_1\|^2-\frac{1}{2\gamma_2}\|g_1\|^2 \\
&= f^*(\bar{x};\gamma_1)+\frac{\gamma_2-\gamma_1}{2\gamma_1\gamma_2}\|g_1\|^2
\end{aligned}
$$

$\boxtimes$

§ 2.3.3 Minimization Methods for the Minmax Problem

Gradient Method for Minmax Problem
0. 取 $x_0 \in Q$, $h>0$
1. kth 迭代 $(k \geq 0)$ (2.3.11)
$x_{k+1} = x_k - h g_f(x_k;L)$

<u>定理2.3.4</u> 令 $f \in \mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$，若 (2.3.11) 取 $h \leq \frac{1}{L}$，则：

$$\|x_k - x^*\|^2 \leq (1-\mu h)^k \|x_0 - x^*\|^2, \quad k \geq 0$$

证明：令 $r_k = \|x_k - x^*\|$, $g_k = g_f(x_k;L)$，则

$$r_{k+1}^2 = \| x_k - h g_k - x^* \|^2 = r_k^2 - 2h\langle g_k, x_k - x^* \rangle + h^2 \| g_k \|^2$$

$$\overset{(2.3.9)}{\leq} (1-h\mu) r_k^2 + h(h - \tfrac{1}{L}) \| g_k \|^2$$

$$\leq (1-h\mu) r_k^2$$

令 $\alpha = hL \leq 1$，则

$$x_{k+1} = x_k - h g_f(x_k; L)$$

$$= x_k - hL(x_k - x_f(x_k; L))$$

$$= (1-\alpha) x_k + \alpha x_f(x_k; L) \in Q$$

$\blacksquare$

注：取 $h = \frac{1}{L}$，则 $x_{k+1} = x_f(x_k; L)$，有收敛速率

$$\| x_k - x^* \|^2 \leq (1 - \tfrac{\mu}{L})^k \| x_0 - x^* \|^2$$

考虑加速，定义估计序列

$$\phi_0(x) = f(x_0) + \frac{\gamma_0}{2} \| x - x_0 \|^2$$

$$\phi_{k+1}(x) = (1-\alpha_k)\phi_k(x) + \alpha_k \Big[f(x_f(y_k; L)) + \frac{1}{2L} \| g_f(y_k; L) \|^2$$

$$+ \langle g_f(y_k; L), x - y_k \rangle + \frac{\mu}{2} \| x - y_k \|^2 \Big]$$

注：这样定义的原因是为了证明 $\phi_k(x)$ 是估计序列，即

$$\phi_k(x) \leq (1-\lambda_k)f(x) + \lambda_k \phi_0(x) \quad (\text{定义})$$

这时引理 2.2.2 中的证明要利用 (2.3.7) 式，不再用 (2.1.20)
强凸带来的不等式

引理 2.3.3 对 $\forall k \geq 0$，有

$$\phi_k(x) \equiv \phi_k^* + \frac{\gamma_k}{2}\|x - v_k\|^2$$

其中 $\{\gamma_k\}$、$\{v_k\}$、$\{\phi_k^*\}$ 定义如下：$v_0 = x_0$，$\phi_0^* = f(x_0)$

$$\gamma_{k+1} = (1-\alpha_k)\gamma_k + \alpha_k \mu$$

$$v_{k+1} = \frac{1}{\gamma_{k+1}}\left[(1-\alpha_k)\gamma_k v_k + \alpha_k \mu y_k - \alpha_k g_f(y_k;L)\right]$$

$$\phi_{k+1}^* = (1-\alpha_k)\phi_k^* + \alpha_k\left(f(x_f(y_k;L)) + \frac{1}{2L}\|g_f(y_k;L)\|^2\right) - \frac{\alpha_k^2}{2\gamma_{k+1}}\|g_f(y_k;L)\|^2$$

$$+ \frac{\alpha_k(1-\alpha_k)\gamma_k}{\gamma_{k+1}}\left(\frac{\mu}{2}\|y_k - v_k\|^2 + \langle g_f(y_k;L), v_k - y_k\rangle\right)$$

证明：注意到$\nabla^2\phi_0(x)=\gamma_0 I_n$，下证$\nabla^2\phi_k(x)=\gamma_k I_n$，用数归，设$k$成立

则 $\nabla^2\phi_{k+1}(x)=(1-\alpha_k)\nabla^2\phi_k(x)+\alpha_k\mu I_n$

$$=((1-\alpha_k)\gamma_k+\alpha_k\mu)I_n$$

$$\equiv \gamma_{k+1}I_n \qquad \checkmark$$

故 $\phi_{k+1}(x)=(1-\alpha_k)(\phi_k^*+\frac{\gamma_k}{2}\|x-v_k\|^2)$

$$+\alpha_k[f(x_f(y_k;L))+\frac{1}{2L}\|g_f(y_k;L)\|^2 + \langle g_f(y_k;L),x-y_k\rangle+\frac{\mu}{2}\|x-y_k\|^2]$$

故 $\nabla\phi_{k+1}(x)=(1-\alpha_k)\gamma_k(x-v_k)+\alpha_k g_f(y_k;L)$

$$+\alpha_k\mu(x-y_k)=0$$

即可解得 $v_{k+1}=\frac{1}{\gamma_{k+1}}[(1-\alpha_k)\gamma_k v_k+\alpha_k\mu y_k-\alpha_k g_f(y_k;L)]$

最后证明 ϕ_{k+1}^* 有上面的形式

$$\phi_{k+1}^*+\frac{\gamma_{k+1}}{2}\|y_k-v_{k+1}\|^2=\phi_{k+1}(y_k)$$

①

$$=(1-\alpha_k)(\phi_k^*+\frac{\gamma_k}{2}\|y_k-v_k\|^2)+\alpha_k f(x_f(y_k;L))+\frac{\alpha_k}{2L}\|g_f(y_k;L)\|^2$$

$$v_{k+1}-y_k=\frac{1}{\gamma_{k+1}}\left[(1-\alpha_k)\gamma_k(v_k-y_k)-\alpha_k g_f(y_k;L)\right]$$

故 $\frac{\gamma_{k+1}}{2}\|y_k-v_{k+1}\|^2=\frac{1}{2\gamma_{k+1}}\left[(1-\alpha_k)^2\gamma_k^2\|v_k-y_k\|^2\right.$

$$\left.-2\alpha_k(1-\alpha_k)\gamma_k\langle g_f(y_k;L),v_k-y_k\rangle+\alpha_k^2\|g_f(y_k;L)\|^2\right] \quad ②$$

整理 ①②即证 $\boxed{/}$

假设 $\phi_k^*\geqslant f(x_k)$，由(2.3.7) $x=x_k$，$\bar{x}=y_k$

$$\Rightarrow\quad f(x_k)\geqslant f(x_f(y_k;L))+\langle g_f(y_k;L),x_k-y_k\rangle+\frac{1}{2L}\|g_f(y_k;L)\|^2$$

$$+\frac{\mu}{2}\|x_k-y_k\|^2$$

$$\phi_{k+1}^*=(1-\alpha_k)\overset{\geqslant f(x_k)}{\phi_k^*}+\alpha_k\left(f(x_f(y_k;L))+\frac{1}{2L}\|g_f(y_k;L)\|^2\right)-\frac{\alpha_k^2}{2\gamma_{k+1}}\|g_f(y_k;L)\|^2$$

$$+\frac{\alpha_k(1-\alpha_k)\gamma_k}{\gamma_{k+1}}\left(\frac{\mu}{2}\|y_k-v_k\|^2+\langle g_f(y_k;L),v_k-y_k\rangle\right)$$

$$\geqslant(1-\alpha_k)f(x_k)+\alpha_k f(x_f(y_k;L))+\left(\frac{\alpha_k}{2L}-\frac{\alpha_k^2}{2\gamma_{k+1}}\right)\|g_f(y_k;L)\|^2$$

$$+\frac{\alpha_k(1-\alpha_k)\gamma_k}{\gamma_{k+1}}\langle g_f(y_k;L),v_k-y_k\rangle$$

$$\geq f(x_f(y_k;L)) + \left(\frac{1}{2L} - \frac{\alpha_k^2}{2\gamma_{k+1}}\right)\|g_f(y_k;L)\|^2$$

$$+ (1-\alpha_k)\left\langle g_f(y_k;L), \frac{\alpha_k \gamma_k}{\gamma_{k+1}}(v_k - y_k) + x_k - y_k \right\rangle$$

故取
$$\begin{cases} x_{k+1} = x_f(y_k;L) \\ L\alpha_k^2 = (1-\alpha_k)\gamma_k + \alpha_k\mu \equiv \gamma_{k+1} \\ y_k = \dfrac{1}{\gamma_k + \alpha_k\mu}(\alpha_k\gamma_k v_k + \gamma_{k+1}x_k) \end{cases}$$

则有 $\phi_{k+1}^* \geq f(x_f(y_k;L)) = f(x_{k+1})$

Constant Step Scheme II for Minimax Problem
0. Choose $x_0 \in \mathbb{R}^n$ and $\alpha_0 \in \left[\sqrt{q_f}, \frac{2(3+q_f)}{3+\sqrt{21+4q_f}}\right]$. Set $y_0 = x_0$. **1.** k**th iteration** ($k \geq 0$). (a) Compute $\{f_i(y_k)\}_{i=1}^m$ and $\{\nabla f_i(y_k)\}_{i=1}^m$. Set $x_{k+1} = x_f(y_k; L)$. (b) Compute $\alpha_{k+1} \in (0,1)$ from the equation $$\alpha_{k+1}^2 = (1-\alpha_{k+1})\alpha_k^2 + q_f\alpha_{k+1}.$$ Set $\beta_k = \frac{\alpha_k(1-\alpha_k)}{\alpha_k^2+\alpha_{k+1}}$ and $y_{k+1} = x_{k+1} + \beta_k(x_{k+1} - x_k)$.

(2.3.12)

定理2.3.5 令 max-type function $f\in \mathcal{S}_{\mu,L}^{1,1}(\mathbb{R}^n)$，若

(2.3.12)取 $\alpha_0\in\left(\sqrt{q_f},\ \frac{2(3+q_f)}{3+\sqrt{21+4q_f}}\right]$，则

$$f(x_k)-f^*\le\frac{4\mu\left[f(x_0)-f^*+\frac{\gamma_0}{2}\|x_0-x^*\|^2\right]}{(\gamma_0-\mu)\left[\exp\left(\frac{k+1}{2}\sqrt{q_f}\right)-\exp\left(-\frac{k+1}{2}\sqrt{q_f}\right)\right]^2}$$

$$\le\frac{4L}{(\gamma_0-\mu)(k+1)^2}\left[f(x_0)-f^*+\frac{\gamma_0}{2}\|x_0-x^*\|^2\right]$$

证明：重新回顾引理2.2.4，修正一下 $\mu=0$ 时的证明：

$$\gamma_{k+1}=(1-\alpha_k)\gamma_k=\cdots=\lambda_{k+1}\gamma_0$$

由 $\alpha_k=1-\frac{\lambda_{k+1}}{\lambda_k}$，由 $L\alpha_k^2=\gamma_{k+1}=\lambda_{k+1}\gamma_0$

$$\Rightarrow\quad 1-\frac{\lambda_{k+1}}{\lambda_k}=\left[\frac{\gamma_{k+1}}{L}\right]^{\frac12}=\left[\frac{\lambda_{k+1}\gamma_0}{L}\right]^{\frac12}$$

$$\Rightarrow\quad \frac{1}{\lambda_{k+1}}-\frac{1}{\lambda_k}=\sqrt{\frac{\gamma_0}{\lambda_{k+1}L}}$$ ，故

$$\sqrt{\frac{\gamma_0}{\lambda_{k+1}L}}=\left(\frac{1}{\sqrt{\lambda_{k+1}}}+\frac{1}{\sqrt{\lambda_k}}\right)\left(\frac{1}{\sqrt{\lambda_{k+1}}}-\frac{1}{\sqrt{\lambda_k}}\right)\le\frac{2}{\sqrt{\lambda_{k+1}}}\left(\frac{1}{\sqrt{\lambda_{k+1}}}-\frac{1}{\sqrt{\lambda_k}}\right)\quad\cdots\circledast$$

故令 $\xi_k=\sqrt{\frac{L}{\gamma_0\lambda_k}}$，则由 $\circledast$ 知 $\xi_{k+1}-\xi_k\ge\frac12$

下面证明 $\xi_k \geq \frac{k+1}{2}$，用数学归纳法

$k=0$ 时，$\xi_0 = \sqrt{\frac{L}{\gamma_0 \lambda_0}} = \sqrt{\frac{L}{\gamma_0}} \geq \frac{1}{\sqrt{3}} \geq \frac{1}{2}$

设 k 时成立，下证 $k+1$ 时成立

由 $\xi_{k+1} - \xi_k \geq \frac{1}{2}$，$\Rightarrow \sqrt{\frac{\gamma_0}{L}} \leq 2\left(\frac{1}{\sqrt{\lambda_k}\sqrt{1-\alpha_k}} - \frac{1}{\sqrt{\lambda_k}}\right)$

则 $\frac{1}{\sqrt{1-\alpha_k}} \geq 1 + \frac{\sqrt{\lambda_k}}{2}\sqrt{\frac{\gamma_0}{L}}$，则有

$$
\begin{aligned}
\xi_{k+1} = \sqrt{\frac{L}{\gamma_0 \lambda_{k+1}}} &= \frac{1}{\sqrt{1-\alpha_k}} \sqrt{\frac{L}{\gamma_0 \lambda_k}} \\
&\geq \left(1 + \frac{1}{2}\sqrt{\frac{\lambda_k \gamma_0}{L}}\right) \sqrt{\frac{L}{\gamma_0 \lambda_k}} \\
&\geq \frac{k+1}{2} + \frac{1}{2} = \frac{k+2}{2}
\end{aligned}
$$

证明了 $\xi_k = \sqrt{\frac{L}{\gamma_0 \lambda_k}} \geq \frac{k+1}{2} \Leftrightarrow \lambda_k \leq \frac{4L}{\gamma_0 (k+1)^2}$

为什么能直接用引理2.2.4的结论：

利用(2.3.7)，可以证明引理2.2.2的结论，结合 $\phi_k^* \geq f(x_k)$，

可以利用定理2.2.1，而 $\{\alpha_k\}$ 的更新准则

$$
L\alpha_k^2 = (1-\alpha_k)\gamma_k + \alpha_k \mu = \gamma_{k+1}
$$

$$
\alpha_0 \to \gamma_1 \to \alpha_1 \to \gamma_2 \to \cdots \to \alpha_k \to \gamma_{k+1} \to \cdots
$$

与2.2节完全一致，自然 $\lambda_k=\prod_{i=0}^{k}(1-\alpha_i)$ 收敛分析完全一致

针对 $\mu>0$ 时，得到算法的变体：

Optimal Method for Minimax Problem with $f\in\mathscr{S}_{\mu,L}^{1,1}(\mathbb{R}^n)$
0. Choose $x_0\in Q$. Set $y_0=x_0$, $\beta=\frac{1-\sqrt{q_f}}{1+\sqrt{q_f}}$. **1.** k**th iteration** ($k\geq 0$). Compute $\{f_i(y_k)\}$ and $\{\nabla f_i(y_k)\}$. Set $x_{k+1}=x_f(y_k;L)$ and $y_{k+1}=x_{k+1}+\beta(x_{k+1}-x_k)$.

(2.3.13)

定理2.3.6 对于算法(2.3.13)，有

$$f(x_k)-f^*\leq 2\left(1-\sqrt{\tfrac{\mu}{L}}\right)^k(f(x_0)-f^*)$$

证明：相当于 $\alpha_0=\sqrt{\frac{\mu}{L}}$，$\gamma_0=\mu$，则由引理2.2.4

$$f(x_k)-f^*\leq(1-\sqrt{q_f})^k\left[f(x_0)-f^*+\frac{\mu}{2}\|x_0-x^*\|^2\right]$$

$$\text{(推论2.3.1)}\quad\leq 2(1-\sqrt{q_f})^k[f(x_0)-f^*]$$

$\square$

我们要解：

$$\min_{x\in Q}\left\{\max_{1\leq i\leq m}\left[f_i(x_0)+\langle\nabla f_i(x_0),x-x_0\rangle\right]+\frac{\gamma}{2}\|x-x_0\|^2\right\}$$

$$\Leftrightarrow \quad \min_{x,t} \left\{ t + \frac{\gamma}{2}\|x - x_0\|^2 \right\}$$

$$\text{s.t.} \quad f_i(x_0) + \langle \nabla f_i(x_0), x - x_0 \rangle \leq t, \; i=1,\cdots,m$$

$$x \in Q, \; t \in \mathbb{R}$$

可以用单纯形法/内点法解

§ 2.3.4

$$\min_{x\in Q} f(x) \qquad \text{s.t. } f_i(x)\le 0,\ i=1,\cdots,m \tag{2.3.16}$$

设 $f_i \in \mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$, $\mu>0$

定义参数化的 max-type func:

$$f(t;x)=\max\{f_0(x)-t,\ f_i(x),\ i=1,\cdots,m\},\ t\in\mathbb{R},\ x\in Q$$

$$f^*(t)=\min_{x\in Q} f(t;x) \tag{2.3.17}$$

$f(t;x)$ 的每一个分支是关于 x 的强凸 func, 由 Thm 2.3.2

对 $\forall t\in\mathbb{R}$, (2.3.7) 的 solution 存在唯一

引理 2.3.4　令 t^* 是 (2.3.4) 的 optimal value, 则

$$f^*(t)\le 0, \text{ 对 } \forall t\ge t^*$$

$$f^*(t)>0, \text{ 对 } \forall t<t^*$$

证明: 令 x^* 是 (2.3.16) 的 solution, 若 $t\ge t^*$, 则

$$f^*(t) \le f^*(t; x^*)$$
$$= \max\{f_0(x^*) - t;\ f_i(x^*)\}$$
$$= \max\{t^* - t \quad ;\ f_i(x^*)\} \le 0$$

另一方面，当 $t < t^*$ 时，用反证法，设 $f^*(t) \le 0$，即

$$f^*(t) = \min_{x \in \otimes} \max\{f_0(x) - t;\ f_i(x)\} \le 0$$

则 $\exists y \in \otimes$, s.t.

$$f_0(y) \le t < t^*,\ f_i(y) \le 0,\ i = 1, \cdots, m$$

与 x^* 是 (2.3.16) 的 solution 矛盾 $\boxed{/\!/}$

故 $f^*(\cdot)$ 的最小根是 (2.3.16) 的 optimal value!

<u>引理 2.3.5</u>：对 $\forall \Delta \ge 0$，有

$$f^*(t) - \Delta \le f^*(t+\Delta) \le f^*(t)$$

证明：
$$f^*(t+\Delta) = \min_{x \in \otimes} \max_{1 \le i \le m} \{f_0(x) - t - \Delta;\ f_i(x)\}$$
$$\le \min_{x \in \otimes} \max_{1 \le i \le m} \{f_0(x) - t;\ f_i(x)\} = f^*(t)$$

$$f^*(t+\Delta)=\min_{x\in\otimes}\max_{1\le i\le m}\{f_0(x)-t;f_i(x)+\Delta\}-\Delta$$

$$\geq \min_{x\in\otimes}\max_{1\le i\le m}\{f_0(x)-t;f_i(x)\}-\Delta=f^*(t)-\Delta$$

$\boxtimes$

注：$f^*(t)$是递减的，1- Lipschiz的

<u>引理 2.3.6</u> 对$\forall t_1<t_2,\Delta\geq 0$，有

$$f^*(t_1-\Delta)\geq f^*(t_1)+\Delta\frac{f^*(t_1)-f^*(t_2)}{t_2-t_1} \qquad (2.3.18)$$

证明：令$t_0=t_1-\Delta$，$\alpha=\dfrac{\Delta}{t_2-t_0}=\dfrac{\Delta}{t_2-t_1+\Delta}\in[0,1]$

则 $\alpha=\dfrac{t_1-t_0}{t_2-t_0}\Rightarrow t_1=(1-\alpha)t_0+\alpha t_2$

(2.3.18) $\Leftrightarrow$ $f^*(t_1)\leq(1-\alpha)f^*(t_0)+\alpha f^*(t_2)$

令$x_\alpha=(1-\alpha)x^*(t_0)+\alpha x^*(t_2)$，则

$$f^*(t_1)\leq\max_{1\le i\le m}\{f_0(x_\alpha)-t_1;f_i(x_\alpha)\}$$

$$\leq\max_{1\le i\le m}\{(1-\alpha)(f_0(x^*(t_0))-t_0)+\alpha(f_0(x^*(t_2))-t_2);\ (1-\alpha)f_i(x^*(t_0))+\alpha f_i(x^*(t_2))\}$$

$$\leq(1-\alpha)\max_{1\le i\le m}\{f_0(x^*(t_0))-t_0;f_i(x^*(t_0))\}$$

$$+\alpha \max_{1\le i\le m}\{f_0(x^*(t_2))-t_2;\ f_i(x^*(t_2))\}$$

$$=(1-\alpha)f^*(t_0)+\alpha f^*(t_2) \qquad \boxed{/\!/}$$

注：引理2.3.5，2.3.6对任意参数化的max-type均成立！

定义 linearization of parametric max-type func $f(t;\cdot)$

$$f(t;\bar{x};x)=\max_{1\le i\le m}\{f_0(\bar{x})+\langle\nabla f_0(\bar{x}),x-\bar{x}\rangle-t;\ f_i(\bar{x})+\langle\nabla f_i(\bar{x}),x-\bar{x}\rangle\}$$

固定 $\gamma>0$，定义：

$$f_\gamma(t;\bar{x};x)=f(t;\bar{x};x)+\frac{\gamma}{2}\|x-\bar{x}\|^2$$

$$f^*(t;\bar{x};\gamma)=\min_{x\in Q}f_\gamma(t;\bar{x};x)$$

$$x_f(t;\bar{x};\gamma)=\arg\min_{x\in Q}f_\gamma(t;\bar{x};x)$$

$$g_f(t;\bar{x};\gamma)=\gamma(\bar{x}-x_f(t;\bar{x};\gamma))$$

由于 max-type func $f_\gamma(t;\bar{x};\cdot)$ 的各个分支是：

$$f_0(\bar{x})+\langle\nabla f_0(\bar{x}),x-\bar{x}\rangle-t+\frac{\gamma}{2}\|x-\bar{x}\|^2$$

$$f_i(\bar{x})+\langle\nabla f_i(\bar{x}),x-\bar{x}\rangle+\frac{\gamma}{2}\|x-\bar{x}\|^2,\quad i=1,\cdots,m$$

故 $f_\gamma(t;\bar{x};\cdot)\in \mathcal{F}^{1,1}_{\gamma,\gamma}(\mathbb{R}^n)$，则由 Thm 2.3.2：

对 $\forall t\in\mathbb{R}$，f^*，x_f，g_f 均良定义

由 $f(t;\cdot)\in \mathcal{F}^{1,1}_{\mu,L}(\mathbb{R}^n)$，故由引理 2.3.1

$$f_\mu(t;\bar{x};x)\leqslant f(t;x)\leqslant f_L(t;\bar{x};x)\quad \forall x\in\mathbb{R}^n$$

$$\Rightarrow\ f^*(t;\bar{x};\mu)\leqslant f^*(t)\leqslant f^*(t;\bar{x};L)$$

再由引理 2.3.6：对 $\forall \bar{x}\in\mathbb{R}^n$，$\gamma>0$，$\Delta\geqslant 0$，$t_1<t_2$

$$f^*(t_1-\Delta;\bar{x};\gamma)\geqslant f^*(t_1;\bar{x};\gamma)+\frac{\Delta}{t_2-t_1}\left(f^*(t_1;\bar{x};\gamma)-f^*(t_2;\bar{x};\gamma)\right)$$

应用引理 2.3.2，$\gamma_1=L$，$\gamma_2=\mu$：

$$f^*(t;\bar{x};\mu)\geqslant f^*(t;\bar{x};L)-\frac{L-\mu}{2\mu L}\|g_f(t;\bar{x};L)\|^2$$

定义 $t^*(\bar{x},t)=\mathrm{root}_t(f^*(t;\bar{x};\mu))$

最小根

引理 2.3.7：令 $\bar{x}\in\mathbb{R}^n$, $\bar{t}<t^*$, s.t. $\exists\ \kappa\in(0,1)$，有

$$f^*(\bar{t};\bar{x};\mu)\geqslant(1-\kappa)f^*(\bar{t};\bar{x};L)$$

则 $\bar{t}<t^*(\bar{x},\bar{t})\leqslant t^*$，且对 $\forall\ t<\bar{t}$, $x\in\mathbb{R}^n$，有

$$f^*(t;x;L)\geqslant 2(1-\kappa)f^*(\bar{t};\bar{x};L)\sqrt{\frac{\bar{t}-t}{t^*(\bar{x},\bar{t})-\bar{t}}}$$

证明：由 $\bar{t}<t^*$，则

$$0<f^*(\bar{t})\leqslant f^*(\bar{t};x;L)\leqslant\frac{1}{1-\kappa}f^*(\bar{t};\bar{x};\mu)$$

$\Rightarrow$ $f^*(\bar{t};\bar{x};\mu)>0$，由 $f^*(\cdot;\bar{x};\mu)\downarrow$，故

$$t^*(\bar{x};\bar{t})>\bar{t}$$

又由 $f^*(t;\bar{x};\mu)\leqslant f^*(t)$，故 $t^*(\bar{x};\bar{t})\leqslant t^*$

令 $\Delta=\bar{t}-t$，则由 (2.3.20)，有

$$\begin{aligned}
f^*(t;\bar{x};L)&\geqslant f^*(t)\geqslant f^*(t;\bar{x};\mu)\\
&\geqslant f^*(\bar{t};\bar{x};\mu)+\frac{\Delta}{t^*(\bar{x};\bar{t})-\bar{t}}f^*(\bar{t};\bar{x};\mu)\\
&\geqslant(1-\kappa)\left(1+\frac{\Delta}{t^*(\bar{x};\bar{t})-\bar{t}}\right)f^*(\bar{t};\bar{x};L)\\
&\geqslant 2(1-\kappa)f^*(\bar{t};\bar{x};L)\sqrt{\frac{\Delta}{t^*(\bar{x};\bar{t})-\bar{t}}}
\end{aligned}$$

$\square$

§ 2.3.5 The Method for Constrained Minimization

Constrained Minimization Scheme
0. Choose $x_0 \in Q$, $\varkappa \in (0, \frac{1}{2})$, $t_0 < t^*$, and accuracy $\epsilon > 0$. **1.** k**th iteration** ($k \geq 0$). (a) Generate the sequence $\{x_{k,j}\}$ by method (2.3.13) as applied to $f(t_k; \cdot)$ with starting point $x_{k,0} = x_k$. If $$f^*(t_k; x_{k,j}; \mu) \geq (1 - \varkappa) f^*(t_k; x_{k,j}; L),$$ then stop the internal process and set $j(k) = j$, $$j^*(k) = \arg \min_{0 \leq j \leq j(k)} f^*(t_k; x_{k,j}; L),$$ $$x_{k+1} = x_f(t_k; x_{k,j^*(k)}; L).$$ **Global Stop:** $f^*(t_k; x_{k,j}; L) \leq \epsilon$ at some iteration of the internal scheme. (b) Set $t_{k+1} = t^*(x_{k,j(k)}, t_k)$.

(2.3.22)

引理 2.3.8

$$f^*(t_k; x_{k+1}; L) \leq \frac{t^* - t_0}{1-\varkappa} \left[\frac{1}{2(1-\varkappa)}\right]^k$$

证明：令 $\beta = \frac{1}{2(1-\varkappa)} < 1$，$\delta_k = \frac{f^*(t_k; x_{k,j(k)}; L)}{\sqrt{t_{k+1} - t_k}}$

由 $t_{k+1} = t^*(x_{k,j(k)}, t_k)$，由引理 2.3.7：对 $k \geq 1$

$$2(1-\varkappa) \frac{f^*(t_k; x_{k,j(k)}; L)}{\sqrt{t_{k+1} - t_k}} \leq \frac{f^*(t_{k-1}; x_{k-1,j(k-1)}; L)}{\sqrt{t_k - t_{k-1}}}$$

$$\Rightarrow \quad \delta_k \leq \beta \delta_{k-1}$$

$$\Rightarrow \quad f^*(t_k; x_{k,j(k)}; L) = \delta_k \sqrt{t_{k+1}-t_k}$$

$$\leq \beta^k \delta_0 \sqrt{t_{k+1}-t_k}$$

$$= \beta^k f^*(t_0; x_{0,j(0)}; L)\sqrt{\frac{t_{k+1}-t_k}{t_1-t_0}}$$

由引理2.3.5 : $t_1 - t_0 \geq f^*(t_0; x_{0,j(0)}; \mu)$

故 $f^*(t_k; x_{k,j(k)}; L) \leq \beta^k f^*(t_0; x_{0,j(0)}; L)\sqrt{\frac{t_{k+1}-t_k}{f^*(t_0; x_{0,j(0)}; \mu)}}$

$$\leq \frac{\beta^k}{1-\kappa}\sqrt{f^*(t_0; x_{0,j(0)}; \mu)(t_{k+1}-t_k)}$$

$$\leq \frac{\beta^k}{1-\kappa}\sqrt{f^*(t_0)(t^*-t_0)}$$

$$\leq \frac{\beta^k}{1-\kappa}(t^*-t_0) \quad \text{(Lemma 2.3.5)}$$

又由 $f^*(t_k; x_{k+1}; L) \equiv f^*(t_k; x_{k,j^*(k)}; L) \leq f^*(t_k; x_{k,j(k)}; L)$

$\boxed{\angle}$

§ 2.3.5 The Method for constrained Minimization

0. 取 $x_0 \in Q$，$\kappa \in (0,\frac{1}{2})$，$t_0 < t^*$，$\varepsilon > 0$

1. 应用 (2.3.13) 于 $f(t_k;\cdot)$，$x_{k,0} = x_k$ 生成序列 $\{x_{k,j}\}$，若

$$f^*(t_k; x_{k,j}; \mu) \geqslant (1-\kappa) f^*(t_k; x_{k,j}; L)$$

则停止内层，令 $j(k) = j$，

$$j^*(k) = \underset{0 \leq j \leq j(k)}{\operatorname{argmin}} f^*(t_k; x_{k,j}; L),$$

$$x_{k+1} = x_f(t_k; x_{k,j^*(k)}; L)$$

注：$t_0 < t_1 < \cdots < t_{k+1} < \cdots \leq t^*$

全局停止条件：$f^*(t_k; x_{k,j}; L) \leq \varepsilon$

(b) 令 $t_{k+1} = t^*(x_{k,j(k)}, t_k)$

引理 2.3.8

$$f^*(t_k; x_{k+1}; L) \leqslant \frac{t^* - t_0}{1-\kappa}\left[\frac{1}{2(1-\kappa)}\right]^k$$

证明：令 $\beta = \dfrac{1}{2(1-\kappa)}$，$\delta_k = \dfrac{f^*(t_k; x_{k,j(k)}; L)}{\sqrt{t_{k+1} - t_k}}$

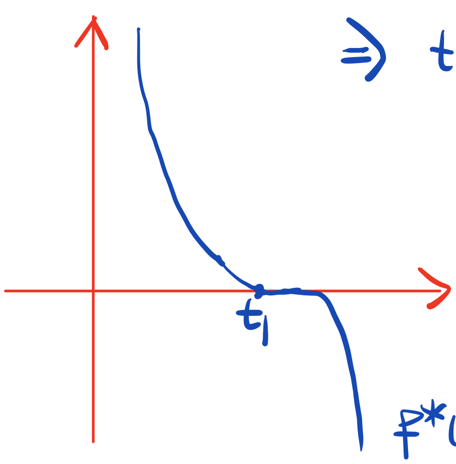
由 $t_{k+1} = t^*(x_{k,j(k)}, t_k)$. 由 lemma 2.3.7，对 $k \geq 1$

$$2(1-\kappa)\frac{f^*(t_k; x_{k,j(k)}; L)}{\sqrt{t_{k+1}-t_k}} \leq \frac{f^*(t_{k-1}; x_{k-1,j(k-1)}; L)}{\sqrt{t_k - t_{k-1}}}$$

$\Rightarrow$ $\delta_k \leq \beta \delta_{k-1}$，且

$$\begin{aligned} f^*(t_k; x_{k,j(k)}; L) &= \delta_k \sqrt{t_{k+1}-t_k} \\ &\leq \beta^k \delta_0 \sqrt{t_{k+1}-t_k} \\ &= \beta^k f^*(t_0; x_{0,j(0)}; L)\sqrt{\frac{t_{k+1}-t_k}{t_1-t_0}} \end{aligned}$$

<u>由引理2.3.5</u>： $f^*(t) - \Delta \leq f^*(t+\Delta)$



$\Rightarrow$ $t_1 - t_0 \geq f^*(t_0; x_{0,j(0)}; \mu)$

故 $f^*(t_k; x_{k,j(k)}; L) \leq \beta^k f^*(t_0; x_{0,j(0)}; L)\sqrt{\dfrac{t_{k+1}-t_k}{f^*(t_0; x_{0,j(0)}; \mu)}}$

$$\leq \frac{\beta^k}{1-\kappa}\sqrt{f^*(t_0; x_{0,j(0)}; L)(t_{k+1}-t_k)}$$

$$\leq \frac{\beta^k}{1-\kappa}\sqrt{f^*(t_0)(t^*-t_0)}$$

$$\leq \frac{\beta^k}{1-\kappa}(t^*-t_0) \quad (\text{lemma } 2.3.5)$$

怎么来的?

又由

$$f^*(t_k; x_{k+1}; L) \equiv f^*(t_k; x_{k,j^*(k)}; L) \leq f^*(t_k; x_{k,j(k)}; L)$$

上面的结论给了我们估计(2.3.16)的 ε-solution的外层迭代上界的方法:

令 $f^*(t_k; x_{k,j}; L) \leq \varepsilon$, 则对 $x_* = x_f(t_k; x_{k,j}; L)$ 有:

$$f(t_k; x_*) = \max_{1\leq i\leq m}\{f_0(x_*)-t_k; f_i(x_*)\}$$

$$\leq f_L(t_k; x_{k,j}; x_*) = f^*(t_k; x_{k,j}; L) \leq \varepsilon$$

由 $t_k \leq t^*$, 故

$$\begin{cases} f_0(x_*) \leq t^* + \varepsilon \\ f_i(x_*) \leq \varepsilon, \quad i = 1, \cdots, m \end{cases} \quad (2.3.23)$$

是原问题的 ε-近似解

由lemma 2.3.8，得到(2.3.23)至多运行外层循环

$$N(\varepsilon)=\frac{1}{\ln[2(1-\kappa)]}\ln\frac{t^*-t_0}{(1-\kappa)\varepsilon}\text{ 次}$$

下面分析内层循环次数，由定理2.3.6，有

$$f(t_k;x_{k,j})-f^*(t_k)\leqslant 2(1-\sqrt{q_f})^j(f(t_k;x_k)-f^*(t_k))$$

$$\leqslant 2e^{-\sigma j}(f(t_k;x_k)-f^*(t_k))$$

$$\leqslant 2e^{-\sigma j}f(t_k;x_k)\quad(\sigma=\sqrt{q_f})$$

令N表示算法(2.3.22)的完整迭代次数 $(N<N(\varepsilon))$

故 $j(k)$ 对 $0\leqslant k\leqslant N$ 良定义，注意到

$$t_k=t^*(x_{k-1,j(k-1)},t_{k-1})>t_{k-1}$$

故 $f(t_k;x_k)\leqslant f(t_{k-1};x_k)\leqslant f_L(t_{k-1};x_{k-1,j^*(k-1)};x_k)$

$$=f^*(t_{k-1};x_{k-1,j^*(k-1)};L)$$

定义 $\Delta_k=f^*(t_{k-1};x_{k-1;j^*(k-1)};L)$，$k\geqslant1$，$\Delta_0=f(t_0;x_0)$

则对 $\forall k\geqslant0$，有

$$f(t_k;x_k)-f^*(t_k)\leqslant\Delta_k$$

引理 2.3.9 对 $\forall k, 0 \le k \le N$，若满足

$$f(t_k; x_{k,j}) - f^*(t_k) \le \frac{\kappa}{Q_f - 1} f^*(t_k; x_{k,j}; L)$$

则内层循环停止

证明：设(2.3.25)满足，由(2.3.8)

$$\frac{1}{2L}\|g_f(t_k; x_{k,j}; L)\|^2 \le f(t_k; x_{k,j}) - f(t_k; x_f(t_k; x_{k,j}; L))$$
$$\le f(t_k; x_{k,j}) - f^*(t_k)$$

由(2.3.21)

$$f^*(t_k; x_{k,j}; \mu) \ge f^*(t_k; x_{k,j}; L) - \frac{L-\mu}{2\mu L}\|g_f(t_k; x_{k,j}; L)\|^2$$
$$\ge f^*(t_k; x_{k,j}; L) - (Q_f - 1)(f(t_k; x_{k,j}) - f^*(t_k))$$
$$\overset{(2.3.25)}{\ge} (1-\kappa) f^*(t_k; x_{k,j}; L)$$

$\square$

Lemma 2.3.10 对 $\forall k, 0 \le k \le N$，有

$$j(k) \le 1 + \sqrt{Q_f} \ln \frac{2(Q_f - 1)\Delta_k}{\kappa \Delta_{k+1}}$$

证明：设 $j(k)-1 > \frac{1}{\sigma}\ln\frac{2(Q_f-1)\Delta_k}{\kappa\Delta_{k+1}}$，由 Δ_{k+1} 的定义,

注意到内层循环的停止准则在 $j(k)-1$ 步不成立，故由 lemma 2.3.9

$$f^*(t_k;x_{k,j};L) \le \frac{Q_f-1}{\kappa}\left(f(t_k;x_{k,j})-f^*(t_k)\right)$$

$$\le 2\frac{Q_f-1}{\kappa}e^{-\sigma j}\Delta_k$$

$$< \Delta_{k+1}$$

与 Δ_{k+1} 的定义矛盾 $\square$

推论 2.3.3

$$\sum_{k=0}^{N} j(k) \le (N+1)\left[1+\sqrt{Q_f}\ln\frac{2(L-\mu)}{\kappa\mu}\right]+\sqrt{Q_f}\ln\frac{\Delta_0}{\Delta_{N+1}}$$

考虑最后一步迭代次数 j^*

lemma 2.3.11

$$j^* \le 1+\sqrt{Q_f}\ln\frac{2(Q_f-1)\Delta_{N+1}}{\kappa\varepsilon}$$

证明：设 $j^*-1 > \sqrt{Q_f}\ln\frac{2(Q_f-1)\Delta_{N+1}}{\kappa\varepsilon}$

当 $j=j^*-1$ 时，有

$$\varepsilon \le f^*(t_{N+1}; x_{N+1}; L)$$

$$\le \frac{Q_f-1}{\kappa}\left(f(t_{N+1}; x_{N+1,j}) - f^*(t_{N+1})\right)$$

$$\le 2\frac{Q_f-1}{\kappa} e^{-\sigma j}\Delta_{N+1} < \varepsilon \qquad \text{矛盾！} \quad \boxed{\checkmark}$$

推论2.3.4

$$j^* + \sum_{j=0}^{N} j(k) \le (N+2)\left[1+\sqrt{Q_f}\ln\frac{2(Q_f-1)}{\kappa}\right] + \sqrt{Q_f}\ln\frac{\Delta_0}{\varepsilon}$$

$$\sim \ln\frac{t_0-t^*}{\varepsilon}\sqrt{Q_f}\ln Q_f$$

suboptimal的，但是

与optimal的收敛速率差的不大

两个技术性问题

1° 在(2.3.22)中，假设能找到$t_0 < t^*$，这是不难的，

可以取t_0为：$\min\limits_{x\in Q}\left[f_0(x_0) + \langle \nabla f_0(x_0), x-x_0\rangle + \frac{\mu}{2}\|x-x_0\|^2\right]$

$$= f_0(x_0) - \frac{1}{2\mu}\|\nabla f(x_0)\|^2 \underbrace{\le}_{\text{PL condition}} f_0^* < t^*$$

2° 需要计算 $t^*(\bar{x},t)$，即 $f^*(t;\bar{x};\mu)=\min\limits_{x\in\otimes} f_\mu(t;\bar{x};x)$ 的根

可以用simplex-type方法和内点法解决

§3.1.1 Motivation and definitions

$$\operatorname{dom} f = \{x \in \mathbb{R}^n \mid |f(x)| < \infty\}$$

始终假设 $\operatorname{dom} f \neq \emptyset$

Def 3.1.1 $f(\cdot)$是凸的，若 $\operatorname{dom} f$是凸的,且对$\forall x, y \in \operatorname{dom} f$, $\alpha \in [0,1]$，有： $f(\alpha x + (1-\alpha) y) \leqslant \alpha f(x) + (1-\alpha) f(y)$

若不等式严格成立叫严格凸

Lemma 3.1.1 (Jensen's 不等式)

对 $\forall x_1, \cdots, x_m \in \operatorname{dom} f$，正系数 $\alpha_1, \cdots, \alpha_m$, s.t. $\sum_{i=1}^{m} \alpha_i = 1$. 则

$$f\left(\sum_{i=1}^{m} \alpha_i x_i\right) \leqslant \sum_{i=1}^{m} \alpha_i f(x_i)$$

证明：用数归：

$m=2$时，由凸func定义，成立

设$m \geqslant 2$时成立，下证 $m+1$时成立：

$$\sum_{i=1}^{m} \alpha_i x_i = \alpha_1 x_1 + (1-\alpha_1) \sum_{i=1}^{m} \beta_i x_{i+1} \quad , \quad \beta_i = \frac{\alpha_{i+1}}{1-\alpha_1}, i=1,\cdots,m$$

故 $\sum_{i=1}^{m}\beta_i=1$, $\beta_i>0$, $i=1,\cdots,m$

$$
\begin{aligned}
f\left(\sum_{i=1}^{m}\alpha_i x_i\right) &= f\left(\alpha_1 x_1+(1-\alpha_1)\sum_{i=1}^{m}\beta_i x_{i+1}\right)\\
&\leqslant \alpha_1 f(x_1)+(1-\alpha_1)f\left(\sum_{i=1}^{m}\beta_i x_{i+1}\right)\\
&\leqslant \sum_{i=1}^{m+1}\alpha_i f(x_i)
\end{aligned}
$$

$\boxed{/\!/}$

注: $x=\sum_{i=1}^{m}\alpha_i x_i$ 称为 $\{x_i\}_{i=1}^{m}$ 的凸组合

<u>推论3.1.1</u> x 是 $x_1,\cdots,x_m$ 的凸组合,则

$$f(x)\leqslant \max_{1\leqslant j\leqslant m} f(x_i)$$

证明:

$$f(x)=f\left(\sum_{i=1}^{m}\alpha_i x_i\right)\leqslant \sum_{i=1}^{m}\alpha_i f(x_i)\leqslant \max_{1\leqslant j\leqslant m} f(x_i)$$

$\boxed{/\!/}$

<u>推论3.1.2</u> 令 $\Delta=\mathrm{Conv}\{x_1,\cdots,x_m\}$

$$\equiv\left\{x=\sum_{i=1}^{m}\alpha_i x_i \,\middle|\, \alpha_i\geqslant 0,\ \sum_{i=1}^{m}\alpha_i=1\right\}$$

则 $\max_{x\in\Delta} f(x)=\max_{1\leqslant j\leqslant m} f(x_i)$

证明：由推论3.1.1，$f(x)\le \max\limits_{1\le j\le m} f(x_j)$，$\forall x\in\Delta$

显然，$\exists \tilde{x}\in\Delta$, s.t. $f(\tilde{x})=\max\limits_{1\le j\le m} f(x_j)$，故

$$\max_{x\in\Delta} f(x)=\max_{1\le j\le m} f(x_j) \qquad \boxed{/\!/}$$

<u>定理3.1.1</u> f是convex的 $\Leftrightarrow$ 对$\forall x,y\in \mathrm{dom} f$，$\beta\geqslant 0$，s.t.

$y+\beta(y-x)\in \mathrm{dom} f$，则

$$f(y+\beta(y-x))\geqslant f(y)+\beta(f(y)-f(x))$$

证明：

$(\Rightarrow)$ 定义 $\alpha=\dfrac{\beta}{1+\beta}$，$u=y+\beta(y-x)$，则

$$y=\frac{1}{1+\beta}(u+\beta x)=(1-\alpha)u+\alpha x$$

故 $f(y)\leqslant(1-\alpha)f(u)+\alpha f(x)$

$$=\frac{1}{1+\beta}f(u)+\frac{\beta}{1+\beta}f(x)$$

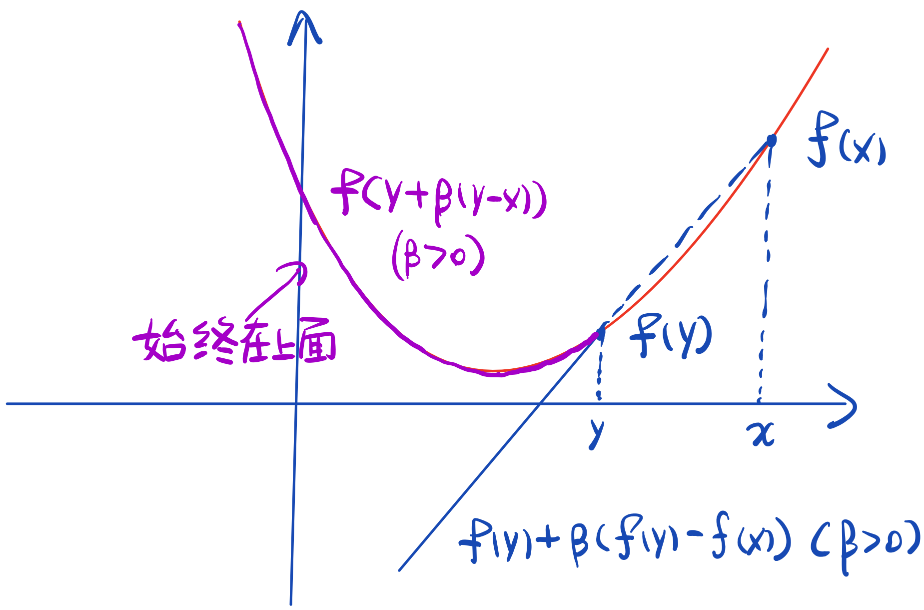
$(\Leftarrow)$ 固定$x,y\in\mathrm{dom} f$，$\alpha\in[0,1]$，定义 $\beta=\dfrac{1-\alpha}{\alpha}$，$u=\alpha x+(1-\alpha)y$

则 $x=\dfrac{1}{\alpha}(u-(1-\alpha)y)=u+\beta(u-y)$

故 $f(x) \geqslant f(u) + \beta(f(u) - f(y)) = \frac{1}{\alpha} f(u) - \frac{1-\alpha}{\alpha} f(y)$

$\boxed{/}$

注 1°: 几何解释



注 2°: 与第二章中 func 定义的关系

$$\frac{f(y+\beta(y-x)) - f(y)}{\beta} \geqslant f(y) - f(x)$$

而 $\lim\limits_{\beta \to 0} \dfrac{f(y+\beta(y-x)) - f(y)}{\beta}$ 是在 y 处关于 $(y-x)$ 的方向导数（设 f 可微）

$= \langle \nabla f(y), y-x \rangle$，此时恰好是第二章的定义

定理3.1.2 f 是 convex $\Longleftrightarrow$ $\text{epi}(f)=\{(x,t)\in \text{dom}f\times\mathbb{R} \mid t\geqslant f(x)\}$ 是凸集

证明:

$(\Rightarrow)$ 若 $(x_1,t_1)\in\text{epi}(f)$, $(x_2,t_2)\in\text{epi}(f)$, 则对 $\forall\alpha\in[0,1]$, 有

$$\alpha t_1+(1-\alpha)t_2\geqslant\alpha f(x_1)+(1-\alpha)f(x_2)\geqslant f(\alpha x_1+(1-\alpha)x_2)$$

故 $(\alpha x_1+(1-\alpha)x_2,\ \alpha t_1+(1-\alpha)t_2)\in\text{epi}(f)$

$(\Leftarrow)$ 由 $x_1,x_2\in\text{dom}f$, 则

$$(x_1,f(x_1))\in\text{epi}(f)\ ,\ (x_2,f(x_2))\in\text{epi}(f)$$

$(\alpha x_1+(1-\alpha)x_2,\ \alpha f(x_1)+(1-\alpha)f(x_2))\in\text{epi}(f)$, 意味着

$$f(\alpha x_1+(1-\alpha)x_2)\leqslant\alpha f(x_1)+(1-\alpha)f(x_2)$$ $\square$

定理3.1.3 f 是 convex 的, 则所有的水平集

$$\mathcal{L}_f(\beta)=\{x\in\text{dom}f \mid f(x)\leqslant\beta\}\ ,\ \beta\in\mathbb{R}$$

是 convex 的或空集

证明: 若 $x_1\in\mathcal{L}_f(\beta)$, $x_2\in\mathcal{L}_f(\beta)$, 则对 $\forall\alpha\in[0,1]$, 有

$$f(\alpha x_1+(1-\alpha)x_2) \le \alpha f(x_1)+(1-\alpha)f(x_2)$$
$$\le \alpha\beta+(1-\alpha)\beta=\beta$$

$\boxed{/}$

定义 3.1.2 f是凸集 Q上的闭凸func,若

$$\mathrm{epi}_Q(f)=\{(x,t)\in Q\times\mathbb{R} : t > f(x)\}$$

是闭凸集. 若$Q=\mathrm{dom} f$, 则 f是闭凸func

Lemma 3.1.2 令 f在 Q上闭凸, 则对 $\forall$闭凸集 $Q_1\subseteq Q$,

f在 Q_1上闭凸

证明: $\{(x,t): x\in Q_1\}$是闭集 ,故由定理 2.2.8 (1)

即证

$\boxed{/}$

定理3.1.4 设f是闭凸的

1) 对$\forall$序列$\{x_k\}\subset \mathrm{dom} f$收敛到$\bar{x}\in \mathrm{dom} f$，有

$$\liminf_{k\to\infty} f(x_k)\geqslant f(\bar{x}) \qquad (3.1.6)$$

证明：注意到$\{(x_k, f(x_k))\}\sqsubseteq \mathrm{epi}(f)$，而$\mathrm{epi}(f)$闭.

若存在子序列收敛到$(\bar{x}, \bar{f})\in \mathrm{epi}(f)$，则$\bar{x}\in \mathrm{dom} f$且

$\bar{f}\geqslant f(\bar{x})$，即(3.1.6)成立

若$\{f(x_k)\}$不存在收敛子序列，则需要考虑以下两种情况：

① $\liminf\limits_{k\to\infty} f(x_k)=-\infty$，不妨设$\lim\limits_{n\to+\infty} f(x_{k_n})=-\infty$

由于$\bar{x}\in \mathrm{dom} f$，则序列$\{(x_{k_n}, f(\bar{x})-1)\}\subset \mathrm{epi}(f)$

对充分大的n均成立

这是因为 $(x_{k_n}, f(\bar{x})-1)\in \mathrm{epi}(f) \Leftrightarrow f(x_{k_n})\leqslant f(\bar{x})-1$

而$\lim\limits_{n\to+\infty} f(x_{k_n})=-\infty$，故对充分大的$n$成立

但$\{(x_{k_n}, f(\bar{x})-1)\}\xrightarrow{n\to+\infty}(\bar{x}, f(\bar{x})-1)\notin \mathrm{epi}(f)$

这与$\mathrm{epi}(f)$的闭性矛盾，故这种情况不成立

② $\lim\limits_{k\to+\infty} f(x_k)=+\infty$，(3.1.6)自然成立 $\square$

2) 对 $\forall$ 序列 $\{x_k\} \subset \mathrm{dom} f$，$x_k \longrightarrow \bar{x} \notin \mathrm{dom} f$，则

$$\lim_{k\to\infty} f(x_k) = +\infty$$

证明：设 $\{f(x_k)\}$ 包含一个有界的子序列，不妨设为 $\{f(x_{k_n})\}_{n=1}^{+\infty}$

那么 $\exists$ 充分大的 τ，s.t. $(x_{k_n}, \tau) \in \mathrm{epi}(f)$，但 $(x_{k_n}, \tau) \to (\bar{x}, \tau)$

$\notin \mathrm{epi}(f)$（由 $\bar{x} \notin \mathrm{dom} f$），与 $\mathrm{epi}(f)$ 的闭性矛盾

下证不存在 $\{x_k\}$ 的子列 $\{x_{k_j}\}_{j=1}^{\infty}$，s.t. $f(x_{k_j}) \to -\infty$.

用反证法，设 $\exists \{x_{k_j}\}$，s.t. $f(x_{k_j}) \to -\infty$. 对充分大的 j'，s.t. $\forall j > j'$

$\{(x_{k_j}, 0)\}_{j=j'}^{+\infty} \subseteq \mathrm{epi}(f)$，但 $(x_{k_j}, 0) \to (\bar{x}, 0) \notin \mathrm{epi}(f)$，与 $\mathrm{epi}(f)$

的闭性矛盾

综上：$\lim_{k\to\infty} f(x_k) = +\infty$ $\boxed{/\!/}$

3) f 的所有水平集或是闭凸的，或是空的

证明：由定义 $(\mathcal{L}_f(\beta), \beta) = \mathrm{epi}(f) \cap \{(x,t) \mid t = \beta\}$

由 $\mathrm{epi}(f)$ 和 $\{(x,t) \mid t=\beta\}$ 均是凸集，故 $(\mathcal{L}_f(\beta), \beta)$ 是凸集或空集

故 $\mathcal{L}_f(\beta)$ 是凸集或空集

$\boxed{/\!/}$

4) 令 f 是 Q 上的闭凸 func，约束水平集均有界，则

$\min\limits_{x\in Q} f(x)$ 是可解的

证明：考虑序列 $\{x_k\}\subset Q$，s.t. $\lim\limits_{k\to\infty} f(x_k)=f_*\triangleq\inf\limits_{x\in Q} f(x)$

故 $\exists k'$, s.t. $\forall k\geq k'$，$f(x_k)\leq f_*+1$，由于 $\mathcal{L}_f(f_*+1)$ 是有界的，故 $\{x_k\}_{k=k'}^{+\infty}$ 是有界的，因此不妨设 $\lim\limits_{k\to\infty} x_k=x^*$

设 $f_*=-\infty$，考虑点列 $y_k=(1-\alpha_k)x_0+\alpha_k x_k\in Q$，$k\geq 0$

其中 $\{\alpha_k\}\downarrow 0$，s.t. $f(y_k)\leq f(x_0)+\alpha_k(f(x_k)-f(x_0))\to-\infty$

注：这样的 $\{\alpha_k\}$ 是 $\exists$ 的，不妨取 $\alpha_k=\dfrac{1}{\sqrt{|f(x_k)-f(x_0)|}}\to 0$

则此时 $y_k\to x_0\in Q$，故由 1) 知道：

$$\liminf_{k\to\infty} f(y_k)\geq f(x_0) \quad \cdots \quad (*)$$

而 $\lim\limits_{k\to\infty} f(y_k)=-\infty$，故与 $(*)$ 矛盾

故 $f_*>-\infty$，因此不妨设 $\{(x_k, f(x_k))\}$ 收敛到 $(x^*, f_*)\in \mathrm{epi}_Q(f)$，

故 $f(x^*)\leq f^*=\inf\limits_{x\in Q} f(x)$，且 $x^*\in Q$，故 $\min\limits_{x\in Q} f(x)$ 可解 $\square$

5) 令 f 是 Ω 上的闭凸func，若 $X^*=\arg\min_{x\in\Omega} f(x)$ 是非空有界的

则 f 在 Ω 上任意水平集是有界集或空集

证明：设 $\exists\beta>f^*=\min_{x\in\Omega} f(x)$，s.t. $\mathcal{L}_f(\beta)$ 是无界的

固定点 $x^*\in X^*$，取 $R>\max_{y\in X^*}\|y-x^*\|$

考虑序列 $\{x_k\}\subseteq\mathcal{L}_f(\beta)$，s.t. $\rho_k\triangleq\|x_k-x^*\|\to\infty$，不失一般地，

设 $\rho_k\geqslant R$ 对 $\forall k\geqslant 0$ 均成立，定义 $y_k=x^*+\frac{1}{\rho_k}R(x_k-x^*)$

故 $y_k\in\Omega$ 且 $\|y_k-x^*\|=R$，但

$$f(y_k)\leqslant f^*+\frac{1}{\rho_k}R(f(x_k)-f^*)\longrightarrow f^*,\ k\to\infty$$

由 $\|y_k-x^*\|=R$，故 $\{y_k\}_{k\geqslant 0}$ 有界，且 $\mathcal{L}_f(\beta)$ 是闭的，故不妨

设 $\underbrace{\lim_{k\to\infty} y_k=\bar{y}\in\mathcal{L}_f(\beta)}$

这是因为 $f(y_k)\to f^*$，故 $\exists\tilde{k}$，s.t. $\forall k>\tilde{k}$，有 $f(y_k)\leqslant\beta$，故

$\{y_k\}_{k=\tilde{k}}^{\infty}\subseteq\mathcal{L}_f(\beta)$，又由 $\mathcal{L}_f(\beta)$ 的闭性知 $\bar{y}\in\mathcal{L}_f(\beta)$

由1)知 $\underbrace{\liminf_{k\to\infty} f(y_k)}_{= f^*}\geqslant f(\bar{y})$，故 $\bar{y}\in X^*$，与 $\|y_k-x\|>R$

矛盾

$\boxtimes$

Example 3.1.1

1) 线性func是闭凸的

2) $f(x)=|x|$是闭凸的，由

$$\mathrm{epi}(f)=\{(x,t)\mid t\geq x, t\geq -x\}$$

是两个凸集的交

3) $\mathbb{R}^n$上的连续凸func均是闭凸func

证明：设$\{x_k,t_k\}$是$\mathrm{epi}(f)$中的Cauchy列，且$x_k\to\bar{x}$，$t_k\to\bar{t}$

则 $f(x_k)\leq t_k$，两边取极限，由f的连续性知：$f(\bar{x})\leq\bar{t}$

4) $f(x)=\frac{1}{x}$，$x>0$是闭凸func，但$\mathrm{dom} f=\mathrm{int}\,\mathbb{R}_+$是开集

5) $f(x)=\|x\|$，$\|\cdot\|$是任一范数是闭凸func：

$$f(\alpha x_1+(1-\alpha)x_2)=\|\alpha x_1+(1-\alpha)x_2\|$$

$$\leq\alpha\|x_1\|+(1-\alpha)\|x_2\|$$

连续＋凸$\Rightarrow$闭

6)

$$f(x,y)=\begin{cases}0, & x^2+y^2<1\\ \phi(x,y), & x^2+y^2=1\end{cases}$$

其中 $\phi(x,y)\geq 0,\ \forall (x,y)\in D=\{(x,y)\mid x^2+y^2=1\}$

f 是凸 func：

$$f((\alpha x_1+(1-\alpha)x_2,\ \alpha y_1+(1-\alpha)y_2))\leq \alpha f(x_1,y_1)+(1-\alpha)f(x_2,y_2) \quad (*)$$

case1：若 $(x_1,y_1),(x_2,y_2)\notin D$，则 $(*)\Leftrightarrow\ 0\leq 0$ ✓

case2：若 $(x_1,y_1)\in D,(x_2,y_2)\notin D$，$(*)\Leftrightarrow\ 0\leq \alpha\phi(x_1,y_1)$ ✓

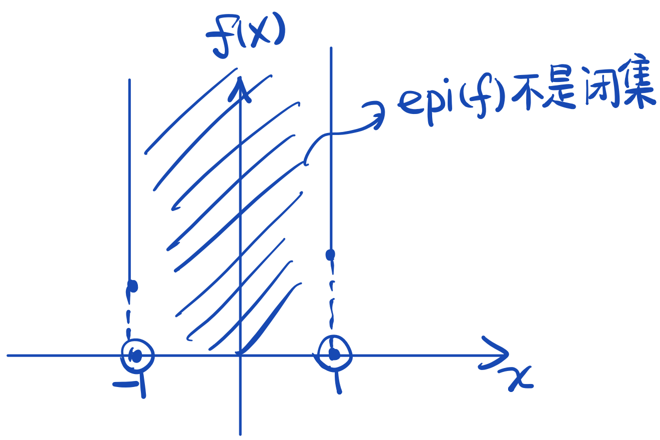
case3：若 $(x_1,y_1)\notin D,(x_2,y_2)\in D$，$(*)\Leftrightarrow\ 0\leq (1-\alpha)\phi(x_2,y_2)$ ✓

case4：若 $(x_1,y_1),(x_2,y_2)\in D$，$(*)\Leftrightarrow\ 0\leq \alpha\phi(x_1,y_1)+(1-\alpha)\phi(x_2,y_2)$ ✓

设 $\phi(x,y)$ 不恒为0，不妨设 $\phi(1,0)=t>0$

则 $\{(1-\frac{1}{n},0)\}_{n=1}^{+\infty}\subseteq \mathrm{epi}(f)$，但 $(1,0)\notin \mathrm{epi}(f)$，故 f 不闭

$\boxed{\text{一维case}}$



§3.1.2 Operations with convex functions

Thm 3.1.5 f_1, f_2分别是凸集$\mathcal{Q}_1, \mathcal{Q}_2$上的闭凸func, $\beta \geq 0$

则下面所有的func是$\mathcal{Q}$上的闭凸func

1. $f(x) = \beta f_1(x)$, $\mathcal{Q} = \mathcal{Q}_1$
2. $f(x) = f_1(x) + f_2(x)$, $\mathcal{Q} = \mathcal{Q}_1 \cap \mathcal{Q}_2$
3. $f(x) = \max\{f_1(x), f_2(x)\}$, $\mathcal{Q} = \mathcal{Q}_1 \cap \mathcal{Q}_2$

证明:

1) 凸性: $f(\alpha x_1 + (1-\alpha)x_2) \leq \beta\left(\alpha f_1(x_1) + (1-\alpha) f_1(x_2)\right)$

$$= \alpha f(x_1) + (1-\alpha) f(x_2)$$

闭性: 由f_1的闭性: $\forall$ Cauchy列 $\{(x_k, t_k)\} \subseteq \mathrm{epi}_{\mathcal{Q}}(f)$,

且$x_k \to \bar{x}$, $t_k \to \bar{t}$, 即 $f(x_k) \leq t_k \Leftrightarrow f_1(x_k) \leq \frac{t_k}{\beta}$

由f_1闭$\Rightarrow (\bar{x}, \frac{\bar{t}}{\beta}) \in \mathrm{epi}_{\mathcal{Q}}(f_1) \Leftrightarrow f_1(\bar{x}) \leq \frac{\bar{t}}{\beta} \Leftrightarrow f(\bar{x}) \leq \bar{t}$

$\Leftrightarrow (\bar{x}, \bar{t}) \in \mathrm{epi}_{\mathcal{Q}}(f)$, 故$f$闭

2) 对$\forall x_1, x_2 \in \mathcal{Q} = \mathcal{Q}_1 \cap \mathcal{Q}_2$, $\alpha \in [0,1]$, 有

$$f_1(\alpha x_1 + (1-\alpha)x_2) + f_2(\alpha x_1 + (1-\alpha)x_2)$$

$$\leq \alpha f_1(x_1) + (1-\alpha) f_1(x_2) + \alpha f_2(x_1) + (1-\alpha) f_2(x_2)$$

$= \alpha(f_1(x_1)+f_2(x_1))+(1-\alpha)(f_1(x_2)+f_2(x_2))$

故 f 在 $\otimes$ 上凸，下证在 $\otimes$ 上是闭的，考虑Cauchy列 $\{(x_k,t_k)\}\subset \mathrm{epi}_{\otimes}(f)$：

即 $t_k \geqslant f_1(x_k)+f_2(x_k)$，$x_k\in\otimes$，$\lim\limits_{k\to\infty} x_k=\bar{x}$，$\lim\limits_{k\to\infty} t_k=\bar{t}$

$\boxed{\text{下证 } \bar{x}\in\otimes}$

先证 $\bar{x}\in\otimes_1$，若不然，由 f_1 的闭凸性知 $\lim\limits_{k\to\infty} f(x_k)\to+\infty$

与 $t_k\geqslant f_1(x_k)+f_2(x_k)$ 矛盾，同理 $\bar{x}\in\otimes_2$

故由 f_1,f_2 分别是 $\otimes_1,\otimes_2$ 上的闭凸func，故

$\liminf\limits_{k\to\infty} f_1(x_k)\geqslant f_1(\bar{x})$；$\liminf\limits_{k\to\infty} f_2(x_k)\geqslant f_2(\bar{x})$

故 $\bar{t}=\lim\limits_{k\to\infty} t_k \geqslant \liminf\limits_{k\to\infty} f_1(x_k)+\liminf\limits_{k\to\infty} f_2(x_k)$

$$\geqslant f_1(\bar{x})+f_2(\bar{x})=f(\bar{x})$$

故 $(\bar{x},\bar{t})\in \mathrm{epi}_{\otimes}(f)$，即证 f 闭

3) 凸：$\forall x_1,x_2\in\otimes$，$\forall\alpha\in[0,1]$

$$\begin{aligned}
f(\alpha x_1+(1-\alpha)x_2) &= \max\{f_1(\alpha x_1+(1-\alpha)x_2),\ f_2(\alpha x_1+(1-\alpha)x_2)\}\\
&\leqslant \max\{\alpha f_1(x_1)+(1-\alpha)f_1(x_2),\ \alpha f_2(x_1)+(1-\alpha)f_2(x_2)\}\\
&\leqslant \alpha\max\{f_1(x_1)+f_1(x_2)\}+(1-\alpha)\max\{f_2(x_1),f_2(x_2)\}\\
&= \alpha f(x_1)+(1-\alpha)f(x_2)
\end{aligned}$$

闭性：$\mathrm{epi}_{Q}(f)=\{(x,t)\mid t\geq f_1(x),t\geq f_2(x),x\in Q_1\cap Q_2\}$

$$=\mathrm{epi}_{Q_1}(f_1)\cap \mathrm{epi}_{Q_2}(f_2)$$

是闭的 $\square$

Thm3.1.6 令ϕ是$S\subseteq\mathbb{R}^n$上的闭凸func，考虑线性算子$\mathcal{A}(x)=Ax+b$

$\mathbb{R}^n\to\mathbb{R}^m$，则函数$f(x)=\phi(\mathcal{A}(x))$是$Q$上的闭凸func，

$$Q=\{x\in\mathbb{R}^n\mid \mathcal{A}(x)\in S\}$$

证明：$\forall x_1,x_2\in Q$，定义$y_1=\mathcal{A}(x_1)$，$y_2=\mathcal{A}(x_2)$，则对$\forall\alpha\in[0,1]$有

$$\begin{aligned}f(\alpha x_1+(1-\alpha)x_2)&=\phi(\mathcal{A}(\alpha x_1+(1-\alpha)x_2))\\&=\phi(\alpha y_1+(1-\alpha)y_2)\\&\leq\alpha\phi(y_1)+(1-\alpha)\phi(y_2)\\&=\alpha f(x_1)+(1-\alpha)f(x_2)\end{aligned}$$

下证闭性：取任意Cauchy列$\{(x_k,t_k)\}\subseteq Q$，且$x_k\to\bar{x}$，$t_k\to\bar{t}$

则 $f(x_k)\leq t_k$，即$\phi(\mathcal{A}(x_k))\leq t_k\Rightarrow(\mathcal{A}(x_k),t_k)\in\mathrm{epi}(\phi)$

由ϕ的闭凸性，$(\mathcal{A}(\bar{x}),\bar{t})\in\mathrm{epi}(\phi)$，故$\phi(\mathcal{A}(\bar{x}))\leq\bar{t}$

$\Rightarrow f(\bar{x})\leq\bar{t}\Rightarrow(\bar{x},\bar{t})\in\mathrm{epi}_{Q}(f)\Rightarrow f$闭 $\square$

Thm 3.1.7 令 Ω 是凸集，ϕ 在 $\mathrm{dom}\,\phi \supseteq \Omega$ 上凸，则

$$f(x)=\inf_{y}\{\phi(x,y):(x,y)\in\Omega\}$$

在 $\hat{\Omega}=\{x:\exists y, s.t. (x,y)\in\Omega\}$ 上凸

证明：取 $\forall x_1,x_2\in\hat{\Omega}$，考虑序列 $\{y_{1,k}\}$，$\{y_{2,k}\}$，s.t.

$\{(x_1,y_{1,k})\}\subset\Omega$，$\{(x_2,y_{2,k})\}\subset\Omega$，且

$\lim\limits_{k\to\infty}\phi(x_1,y_{1,k})=f(x_1)$，$\lim\limits_{k\to\infty}\phi(x_2,y_{2,k})=f(x_2)$

由 ϕ 关于 (x,y) 联合凸，故对 $\forall\alpha\in[0,1]$ 有

$$f(\alpha x_1+(1-\alpha)x_2)\leqslant\phi(\alpha y_1+(1-\alpha)x_2,\ \alpha y_{1,k}+(1-\alpha)y_{2,k})$$

$$\leqslant\alpha\phi(x_1,y_{1,k})+(1-\alpha)\phi(x_2,y_{2,k})$$

对不等号右边取极限即证 $\boxed{/\!/}$

Thm 3.1.8 令 Δ 是任意集合，$f(x)=\sup\limits_{y}\{\phi(x,y)\mid y\in\Delta\}$

设对 $\forall y\in\Delta$，$\phi(\cdot,y)$ 在某个集合 Ω 上是闭凸的，则 f 在 $\hat{\Omega}$ 上闭凸

$$\hat{\Omega}=\left\{x\in\Omega\,\middle|\,\sup_{y\in\Delta}\phi(x,y)<+\infty\right\}$$

证明：若 $x\in\hat{\Omega}$，则 $f(x)<+\infty$，故 $\hat{\Omega}\subseteq\mathrm{dom} f$，显然

$(x,t)\in\mathrm{epi}_{\Omega}(f)\iff$ 对 $\forall y\in\Delta$，有 $x\in\Omega$，$t\geqslant\phi(x,y)$

则 $\mathrm{epi}_{\otimes}(f) = \bigcap_{y\in\Delta} \mathrm{epi}_{\otimes}(\phi(\cdot,y))$

由 $\mathrm{epi}_{\otimes}(\phi(\cdot,y))$是闭凸集，故 $\mathrm{epi}_{\otimes}(f)$是闭凸集

而 $\mathrm{epi}_{\hat{\otimes}}(\phi(\cdot,y)) \equiv \mathrm{epi}_{\otimes}(\phi(\cdot,y))$，即证 ☑

Thm 3.1.9 令$\psi(\cdot)$凸，φ是单变量非减凸func于集合

$$\mathrm{Im}\psi = \{\tau = \psi(x), x\in \mathrm{dom}\psi\}$$

则 $f(x) = \varphi(\psi(x))$，$x\in\mathrm{dom}\psi$是凸func

证明：对$\forall x,y\in \mathrm{dom} f$，$\alpha\in[0,1]$，有

$$\begin{aligned} f(\alpha x+(1-\alpha)y) &= \varphi(\psi(\alpha x+(1-\alpha)y)) \\ &\leq \varphi(\alpha\psi(x)+(1-\alpha)\psi(y)) \\ &\leq \alpha\varphi(\psi(x))+(1-\alpha)\varphi(\psi(y)) \\ &= \alpha f(x)+(1-\alpha)f(y) \end{aligned}$$

☑

Example 3.1.2

1. $f(x) = \max\limits_{1\le i\le n}\{x^{(i)}\}$ 是闭凸func（Thm3.1.5 3.）

$\phi_*(s) = \sup\limits_{x\in \mathrm{dom}\phi} [\langle s, x\rangle - \phi(x)]$，其中 ϕ 是 $\mathbb{R}^n$ 上任一func

是闭凸func，因为对 $\forall x \in \mathrm{dom}\phi$，$\langle s, x\rangle - \phi(x)$ 是关于 s 的

线性func，故由Thm3.1.8 知 $\phi_*(s)$ 的闭凸性

注：ϕ_* 称为 ϕ 的 Fenchel 对偶

2. 令 $\lambda = (\lambda^{(1)}, \cdots, \lambda^{(m)})$，令 Δ 是 $\mathbb{R}^m_+$ 的任一集合，考虑

$$f(x) = \sup_{\lambda\in\Delta}\left\{\sum_{i=1}^m \lambda^{(i)} f_i(x)\right\}$$

其中所有 f_i 是闭凸的，由定理3.1.5 知 $\sum\limits_{i=1}^m \lambda^{(i)} f_i(x)$

是闭凸的，由 3.1.8 知. $f(x)$ 是闭凸的

$\boxtimes$

3. Θ是任意集合，考虑

$$\xi_\Theta(x) = \sup_{g\in\Theta} \langle g, x\rangle$$

称为Θ上的支撑func，由3.1.8 $\xi_\Theta(\cdot)$是闭凸的，

且是正齐次的(degree=1)，即

$$\xi_\Theta(\tau x) = \sup_{g\in\Theta}\langle g, \tau x\rangle = \tau\,\xi_\Theta(x) \qquad x\in \mathrm{dom}\,\xi_\Theta,\ \tau\geq 0$$

若Θ是有界的，则 $\mathrm{dom}\,\xi_\Theta = \mathbb{R}^n$

Lemma 3.1.3 对两个集合Θ_1, Θ_2，定义$\Theta = \mathrm{Conv}\{\Theta_1, \Theta_2\}$

$$\xi_\Theta(x) = \max\{\xi_{\Theta_1}(x), \xi_{\Theta_2}(x)\}$$

证明：由$\Theta_1 \subseteq \Theta$，$\Theta_2\subseteq\Theta$，故显然

$$\xi_\Theta(x) \geq \max\{\xi_{\Theta_1}(x), \xi_{\Theta_2}(x)\}$$

另一方面

$$\xi_\Theta(x) = \sup_{\alpha, g_1, g_2}\{\langle \alpha g_1 + (1-\alpha) g_2, x\rangle : g_1\in\Theta_1, g_2\in\Theta_2, \alpha\in[0,1]\}$$

$$\le \sup_{\alpha\in[0,1]} \left\{\alpha\, \xi_{\otimes_1}(x) + (1-\alpha)\, \xi_{\otimes_2}(x)\right\}$$

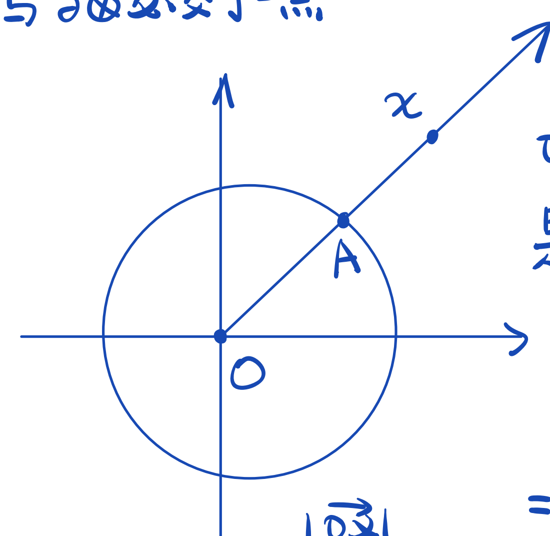
$$= \max\left\{\xi_{\otimes_1}(x),\ \xi_{\otimes_2}(x)\right\} \qquad \square$$

4. 另一个重要的凸齐次func是关于凸集的Minkowski func:

令$\otimes$是有界闭凸集, $0\in \operatorname{int}\otimes$, 定义

$$\psi_{\otimes}(x) = \min_{\tau\geq 0}\left\{\tau : x\in \tau\otimes\right\}$$

优化问题解的存在性是显然的, 过原点与x作射线l, l与$\partial\otimes$必交于一点



记唯一解为$\tau(x)$, 则 $\tau(x)\neq 0$时, $\frac{x}{\tau(x)}\in\partial\otimes$, $\psi_{\otimes}(x)$ 是正齐次的, 由: $\forall\lambda>0, \forall x\in\mathbb{R}^n$

$$\psi_{\otimes}(\lambda x) = \min_{\tau\geq 0}\left\{\tau : \lambda x\in\tau\otimes\right\}$$

$$= \lambda\min_{\tau\geq 0}\left\{\frac{\tau}{\lambda} : x\in\frac{\tau}{\lambda}\otimes\right\}$$

$$= \lambda\min_{\tau'\geq 0}\left\{\tau' : x\in\tau'\otimes\right\}$$

$$= \lambda\,\psi_{\otimes}(x)$$

此时 $\tau(x) = \dfrac{|\overrightarrow{OX}|}{|\overrightarrow{OA}|}$

且 $\psi_{\Theta}(x)$ 是凸 func, $\mathrm{dom}\,\psi_{\Theta}=\mathbb{R}^n$:

对 $\forall x_1,x_2\in\mathbb{R}^n\setminus\{0\}$, $\alpha\in[0,1]$, 有

$$\frac{\alpha x_1+(1-\alpha)x_2}{\alpha\tau(x_1)+(1-\alpha)\tau(x_2)}=\frac{\alpha\tau(x_1)\underbrace{\frac{x_1}{\tau(x_1)}}_{\in\partial\Theta}+(1-\alpha)\tau(x_2)\underbrace{\frac{x_2}{\tau(x_2)}}_{\in\partial\Theta}}{\alpha\tau(x_1)+(1-\alpha)\tau(x_2)}$$

$$\in\Theta$$

故 $\psi_{\Theta}(\alpha x_1+(1-\alpha)x_2)\leqslant\alpha\tau(x_1)+(1-\alpha)\tau(x_2)$

对 $x_1=0$ 或 $x_2=0$ 时, 由正齐性立即知凸性

$\square$

5. 令 Θ 是 $\mathbb{R}^n$ 上的集合, 考虑 $\psi(g,\gamma)=\sup\limits_{y\in\Theta}\phi(y,g,\gamma)$,

$$\phi(y,g,\gamma)=\langle g,y\rangle-\frac{\gamma}{2}\|y\|^2$$

对固定的 y, $\phi(y,g,\gamma)$ 关于 (g,γ) 是闭凸的, 因为

令 $x=\begin{pmatrix}g\\\gamma\end{pmatrix}$, 则 $\phi(x,\gamma)=(g^T,\gamma)\begin{pmatrix}y\\-\frac{1}{2}\|y\|^2\end{pmatrix}$

$$=x^T\begin{pmatrix}y\\-\frac{1}{2}\|y\|^2\end{pmatrix}$$

是线性func, 故是闭凸的

因此由<u>Thm3.1.8</u>，$\psi(g,\gamma)$是闭凸func

当Q是有界的,则$\mathrm{dom}\,\psi=\mathbb{R}^{n+1}$，下面讨论$Q=\mathbb{R}^n$

若$\gamma<0$，则$y\to+\infty$时，$\phi(y,g,\gamma)\to+\infty$，该情况一定不在$\mathrm{dom}\,\psi$中

若$\gamma=0$，$g\neq0$时，取$y_k=\alpha_k g$，其中$\alpha_k\to+\infty$，则$\phi(y_k,g,0)\to+\infty$也不在$\mathrm{dom}\,\psi$中

若$\gamma=0$，$g=0$时，$\psi(0,0)=0$

若$\gamma>0$时，$y^*(g,\gamma)=\frac{1}{\gamma}g$，此时$\psi(g,\gamma)=\frac{\|g\|^2}{2\gamma}$

故

$$\psi(g,\gamma)=\begin{cases}0, & g=0,\gamma=0\\ \frac{\|g\|^2}{2\gamma}, & \gamma>0\end{cases}$$

$\mathrm{dom}\,\psi=(\mathbb{R}^n\times\{\gamma>0\})\cup(0,0)$，该集合是凸集：

$\forall x_1,x_2\in\mathrm{dom}\,\psi\setminus(0,0)$，$\alpha x_1+(1-\alpha)x_2\in\mathrm{dom}\,\psi$

$x_1=(0,0)$，$x_2\in\mathrm{dom}\,\psi\setminus(0,0)$，$\alpha x_1+(1-\alpha)x_2=(1-\alpha)x_2\in\mathrm{dom}\,\psi$

$x_2=(0,0)$时，同理可证

此时ψ是凸集$\mathrm{dom}\,\psi$下的闭凸func，但在原点不连续：

$$\psi(\sqrt{r}g,r)\equiv\frac{1}{2}\|g\|^2,\ r\neq 0$$

考虑闭凸集$Q=\{(g,r):r\geq\|g\|^2\}$，则ψ是Q上的有界闭凸func，但在原点处不连续

为了说明下面的例子，先提出引理：

Lemma：$f:K\to\mathbb{R}$是convex function，$K\subseteq\mathbb{R}^n$是开集.

则f是连续func：

证明：用数学归纳法，$n=0$时成立是显然的

设$n-1$时成立，下证n时成立：

对$\forall x_0\in K$，$\exists r>0$，s.t. $Q=\{x\in\mathbb{R}^n \mid \|x-x_0\|_\infty\leq r\}\subseteq K$

定义 $H_{i,e}=\{x\in\mathbb{R}^n \mid x^{(i)}=x_0^{(i)}+er\}$，$e=\{\pm1\}$

那么$f\big|_{K\cap H_{i,e}}$是连续func，$\forall i=1,\cdots,n$，$e=\pm1$

（相当于开集$K\cap H_{i,e}$上$\mathbb{R}^{n-1}$维的凸func）

由 $\Theta \cap H_{i,e}$ 是紧集，故 $f\big|_{\Theta \cap H_{i,e}}$ 是有界的，且

$\bigcup_{\substack{i=1,\cdots,n\\ e=\pm 1}} \Theta \cap H_{i,e} = \partial\Theta$，故 $f\big|_{\partial\Theta}$ 是有界的，不妨设 $|f(x)|<M$，

$\forall x \in \partial\Theta$，考虑 $\forall x \in \Theta \setminus \{x_0\}$，令

$$g(t) = x_0 + (x - x_0)t$$

令 $t_1 = \frac{r}{\|x-x_0\|_\infty} > 1$，则 $t = \pm t_1$ 时，$g(t) \in \partial\Theta$

声明：$h = f\circ g$ 是 $[-t_1, t_1]$ 上的凸 func

对 $\forall \tilde{t}_1, \tilde{t}_2 \in [-t_1, t_1]$，$\forall \alpha \in [0,1]$

$$h(\alpha\tilde{t}_1 + (1-\alpha)\tilde{t}_2) = f(x_0 + (x-x_0)(\alpha\tilde{t}_1 + (1-\alpha)\tilde{t}_2))$$

$$= f\Big(\alpha\underbrace{(x_0 + (x-x_0)\tilde{t}_1)}_{\in\,\Theta} + (1-\alpha)\underbrace{(x_0 + (x-x_0)\tilde{t}_2)}_{\in\,\Theta}\Big)$$

$$\leq \alpha f\circ g(\tilde{t}_1) + (1-\alpha) f\circ g(\tilde{t}_2)$$

$$= \alpha h(\tilde{t}_1) + (1-\alpha) h(\tilde{t}_2)$$

故 $\dfrac{h(0)-h(-t_1)}{0-(-t_1)} \leqslant \dfrac{h(1)-h(0)}{1-0} \leqslant \dfrac{h(t_1)-h(0)}{t_1-0}$

$$\Rightarrow \quad \frac{h(0)-h(-t_1)}{t_1} \leqslant h(1)-h(0) \leqslant \frac{h(t_1)-h(0)}{t_1}$$

其中 $h(0)=f(x_0)$，$h(1)=f(x)$，$|h(t_1)|<M$，$|h(-t_1)|<M$

$$0<\frac{1}{t_1}=\frac{\|x-x_0\|_\infty}{r} \leqslant \frac{\|x-x_0\|_2}{r}$$

故 $|f(x)-f(x_0)| \leqslant \dfrac{M+|f(x_0)|}{r}\|x-x_0\|_2$

故对 $\forall \varepsilon>0$，取 $\|x-x_0\|<\delta=\min\left\{r, \dfrac{r\varepsilon}{M+|f(x_0)|}\right\}$，

有 $|f(x_0)-f(x)|<\varepsilon$

$\boxed{/\!/}$

6. 令 f 是 $\mathbb{R}^n$ 上的凸 func，故 f 连续，定义

$$\hat{f}(\tau,x)=\tau f\left(\frac{x}{\tau}\right)$$

$\hat{f}$ 在 $\mathbb{R}^n\times\{\tau>0\}$ 上是 well-defined，补充定义：$\hat{f}(0,0)=0$

$\hat{f}$ 是凸 func：考虑 $z_1=(\tau_1,x_1)$，$z_2=(\tau_2,x_2)$，$\tau_1,\tau_2>0$

则对 $\forall \alpha\in[0,1]$ 时：

$$\hat{f}(\alpha z_1+(1-\alpha)z_2) = (\alpha\tau_1+(1-\alpha)\tau_2)\, f\left(\frac{\alpha x_1+(1-\alpha)x_2}{\alpha\tau_1+(1-\alpha)\tau_2}\right)$$

$$= (\alpha\tau_1+(1-\alpha)\tau_2)\, f\left(\frac{\alpha\tau_1\frac{x_1}{\tau_1}+(1-\alpha)\tau_2\frac{x_2}{\tau_2}}{\alpha\tau_1+(1-\alpha)\tau_2}\right)$$

$$\leq \alpha\tau_1 f\left(\frac{x_1}{\tau_1}\right)+(1-\alpha)\tau_2 f\left(\frac{x_2}{\tau_2}\right)$$

$$= \alpha\hat{f}(z_1)+(1-\alpha)\hat{f}(z_2)$$

当 $z_1=0$，$z_2\neq 0$ 时，用齐次性可证凸性

当 $z_2=0$ 时，同理可得凸性

为了说明闭性，设 $\lim\limits_{\tau\to 0}\tau f\left(\frac{x}{\tau}\right)=+\infty$，$\forall x\in\mathbb{R}^n$

则任取Cauchy列 $\{\tau_k,x_k,t_k\}\subseteq \mathrm{epi}(\hat{f})$，$\{\tau_k,x_k,t_k\}$

$\to\{\bar{\tau},\bar{x},\bar{t}\}$，s.t. $\hat{f}(\tau_k,x_k)=\tau_k f\left(\frac{x_k}{\tau_k}\right)\leq t_k$

则 $\bar{\tau}>0$，否则 $\{t_k\}$ 发散

由f的连续性，且 $\bar{\tau}>0$，两边取极限：

$$\bar{\tau} f\left(\frac{\bar{x}}{\bar{\tau}}\right)\leq \bar{t} \iff (\bar{\tau},\bar{x},\bar{t})\in \mathrm{epi}(\hat{f})$$

即证 $\hat{f}$ 是闭的

注：取 $f=\frac{1}{2}\|x\|^2$，则 $\tilde{f}$ 就是例5的 ψ

注：条件 $\lim\limits_{\tau\to 0}\tau f(\frac{x}{\tau})=+\infty$ 是必要的，否则不妨设

$\exists \bar{x}\in\mathbb{R}^n\setminus\{0\}$ s.t. $\lim\limits_{\tau\to 0}\tau f(\frac{\bar{x}}{\tau})=M$，则考虑序列 $\{(\frac{1}{n},\bar{x})\}_{n=1}^{+\infty}$，

不妨设 $\tilde{f}(\frac{1}{n},\bar{x})\le M$，则 $\{(\frac{1}{n},\bar{x},M)\}_{n=1}^{+\infty}\subseteq \mathrm{epi}(\tilde{f})$，

但 $(0,\bar{x},M)\notin \mathrm{epi}(\tilde{f})$，因为 $(0,\bar{x})\notin \mathrm{dom}\tilde{f}$

Lemma 3.1.4 $\forall$ 单变量闭凸 func 在它的定义域连续

证明：令 f 闭凸，$\bar{x}\in \mathrm{dom}f\subseteq\mathbb{R}$，由 Thm 3.1.4 f 在 $\bar{x}$

是下半连续的，下证 f 在 $\bar{x}$ 上半连续：

对 $\forall x_k\to\bar{x}$，在去除 $\{x_k\}$ 有限个点下，有3种情况

① $\forall k\in\mathbb{N}$，$x_k\le\bar{x}$

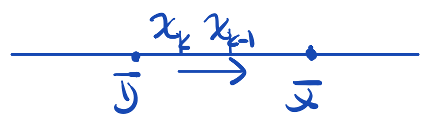
② $\forall k\in\mathbb{N}$，$x_k\ge\bar{x}$

③ $\{x_k\}$ 取子列 $\{x_{k_i}\}_{i=1}^{\infty}$ $\{x_{k_j}\}_{j=1}^{\infty}$，$\{k_i\}\cup\{k_j\}=\mathbb{N}$.

$x_{k_i}\ge\bar{x}\ \forall i$，$x_{k_j}\le\bar{x}$

对 case①：取 $\bar{y}<\bar{x}$，可以不失一般地设：

$$x_k=(1-\alpha_k)\bar{x}+\alpha_k\bar{y}$$



其中 $\alpha_k\to 0$，$\alpha_k\in[0,1]$，故由 f 的凸性

$$f(x_k)\leq(1-\alpha_k)f(\bar{x})+\alpha_k f(\bar{y})$$

$$\limsup_{k\to\infty} f(x_k)\leq f(\bar{x})$$，从而 f 上半连续

对 case②：只需取 $\bar{y}>\bar{x}$ 即可

对 case③：对子列 $\{x_{k_i}\}$ 取 $\bar{y}_1>\bar{x}$；对子列 $\{x_{k_j}\}$ 取 $\bar{y}_2<\bar{x}$，

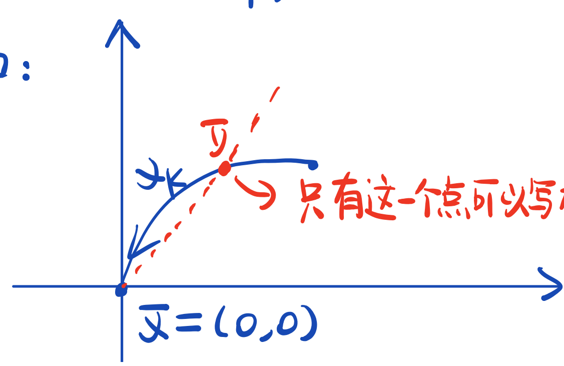
其余证明与①完全一致

综上，任意序列 $\{x_k\}\to\bar{x}$，有 $\limsup\limits_{k\to\infty} f(x_k)\leq f(\bar{x})$，

即证，f 是连续的 $\square$

注：在高维情形 $\{x_k\}\to\bar{x}$ 推不出来 $x_k=(1-\alpha_k)\bar{x}+\alpha_k\bar{y}$

的形式，如：



只有这一个点可以写成 $x_k=(1-\alpha_k)\bar{x}+\alpha_k\bar{y}$.

Thm3.1.8 令 Δ 是任意集合，$f(x)=\sup_y\{\phi(x,y)\mid y\in\Delta\}$

设对 $\forall y\in\Delta$，$\phi(\cdot,y)$ 在某个集合 $\otimes$ 上是闭凸的，则 f 在 $\hat{\otimes}$ 上闭凸

$$\hat{\otimes}=\left\{x\in\otimes \,\middle|\, \sup_{y\in\Delta}\phi(x,y)<+\infty\right\}$$

注：$\hat{\otimes}$ 一定是凸集：对 $\forall x_1,x_2\in\hat{\otimes}$，则 $x_1,x_2\in\otimes$，且

$$\begin{cases}\sup\limits_{y\in\Delta}\phi(x_1,y)<+\infty\\[2ex] \sup\limits_{y\in\Delta}\phi(x_2,y)<+\infty\end{cases}$$

由 $\alpha x_1+(1-\alpha)x_2\in\otimes$，则

$$\phi(\alpha x_1+(1-\alpha)x_2,y)\leq\alpha\phi(x_1,y)+(1-\alpha)\phi(x_2,y)$$

$$\Rightarrow\ \sup_{y\in\Delta}\phi(\alpha x_1+(1-\alpha)x_2,y)$$

$$\leq\ \sup_{y\in\Delta}\ \alpha\phi(x_1,y)+(1-\alpha)\phi(x_2,y)$$

$$\leq\alpha\sup_{y\in\Delta}\phi(x_1,y)+(1-\alpha)\sup_{y\in\Delta}\phi(x_2,y)<+\infty$$

注 2°：$\mathrm{epi}_{\otimes}(f)\equiv\mathrm{epi}_{\hat{\otimes}}(f)$

$\mathrm{epi}_{\hat{\otimes}}(f) \subseteq \mathrm{epi}_{\otimes}(f)$ 是显然的，下证 $\mathrm{epi}_{\otimes}(f) \subseteq \mathrm{epi}_{\hat{\otimes}}(f)$

$\forall (x,t) \in \mathrm{epi}_{\otimes}(f)$，则 $f(x) \leqslant t$ 且 $x \in \otimes$，设 $x \notin \hat{\otimes}$，则 $f(x) = +\infty$，与 $f(x) \leqslant t$ 矛盾，故 $x \in \hat{\otimes}$，故 $(x,t) \in \mathrm{epi}_{\hat{\otimes}}(f)$

所以 $\mathrm{epi}_{\hat{\otimes}}(f)$ 是闭凸集，即证 f 在 $\hat{\otimes}$ 上闭凸

定理3.1.10 令 f_1, f_2 是 Ω 上的闭凸func，约束下水平集是有界的，则 $\exists \lambda^* \in [0,1]$, s.t.

$$\min_{x\in\Omega} \left(f(x) \triangleq \max\{f_1(x), f_2(x)\} \right)$$
$$= \min_{x\in\Omega} \{\lambda^* f_1(x) + (1-\lambda^*) f_2(x)\}$$

证明：定义 $\phi(\lambda) = \min_{x\in\Omega} \{\lambda f_1(x) + (1-\lambda) f_2(x)\}$, $\lambda \in [0,1]$

由 f_1, f_2 是闭凸func，由 Thm3.1.5 知 $\lambda f_1(x) + (1-\lambda) f_2(x) \triangleq \tilde{f}$ 是闭凸func，设 $\tilde{f}$ 的约束下水平集 Ω 非空，则 Ω 是闭凸集，又由 f_1, f_2 的约束下水平集是有界的，则下证 $\tilde{f}(x)$ 的下水平集有界，不妨设

$$\hat{\Omega}_\lambda = \{x \mid \tilde{f}(x) = \lambda f_1(x) + (1-\lambda) f_2(x) \le \beta\} \quad \forall \beta \text{ s.t. } \hat{\Omega} \ne \phi$$

用反证法，设 $\exists \tilde{\lambda}$, s.t. $\hat{\Omega}_{\tilde{\lambda}}$ 无界，那么 $\tilde{\lambda} \in (0,1)$，则

$$\hat{\Omega}_{\lambda} = \left\{ x \mid f_1(x) \leq \frac{1}{\lambda}\left(\beta - (1-\lambda) f_2(x)\right) \right\}$$

$$\subseteq \left\{ x \mid f_1(x) \leq \frac{1}{\lambda}\left(\beta - (1-\lambda) f_2^*\right) \right\} \triangleq \mathcal{L}$$

其中 $f_2^* = \min\limits_{x \in \Omega} f_2(x)$，由引理知 f_2^* 是 well-defined

由 $f_1(x)$ 的约束水平集有界，故 $\mathcal{L}$ 有界，矛盾

故 $\tilde{f}$ 的约束下水平集有界，故 $\hat{\Omega}$ 是紧的

故 $\min\limits_{x \in \Omega} \tilde{f}(x) \Leftrightarrow \min\limits_{x \in \hat{\Omega}} \tilde{f}(x)$

由 3.1.4，$\phi(\lambda)$ 是 well-defined.

由 $-\phi(\lambda)=\max\limits_{x\in\otimes}\{-\lambda f_1(x)-(1-\lambda)f_2(x)\}$

由Thm3.1.8，$-\phi$ 是闭凸func，故由lemma 3.1.4 $-\phi$ 是

$[0,1]$ 上的连续func，故 $\phi^*=\max\limits_{\lambda\in[0,1]}\phi(\lambda)$ 是well-defined

由 $\phi^*=\phi(\lambda^*)=\max\limits_{\lambda\in[0,1]}\phi(\lambda)$

$$=\max_{\lambda\in[0,1]}\min_{x\in\otimes}\{\lambda f_1(x)+(1-\lambda)f_2(x)\}$$

$$(\text{Thm1.3.1})\leq\min_{x\in\otimes}\max_{\lambda\in[0,1]}\{\lambda f_1(x)+(1-\lambda)f_2(x)\}$$

$$=\min_{x\in\otimes}\max\{f_1(x),f_2(x)\}$$

$$=\min_{x\in\otimes}f(x)=f^*$$

要证Thm3.1.10，只需证 $\phi^*=f^*$，故下证 $f^*\leq\phi^*$

对 $\forall\lambda\in[0,1]$，固定 λ，定义：

$$x(\lambda)\in\operatorname*{Argmin}_{x\in\otimes}\{\lambda f_1(x)+(1-\lambda)f_2(x)\}$$

$$g(\lambda)=f_1(x(\lambda))-f_2(x(\lambda))$$

对 $\forall\lambda_1,\lambda_2\in[0,1]$，有：

$$\phi(\lambda_1)=\min_{x\in\Theta}\{\lambda_1 f_1(x)+(1-\lambda_1)f_2(x)\}$$

$$\leqslant \lambda_1 f_1(x(\lambda_2))+(1-\lambda_1)f_2(x(\lambda_2))$$

$$=\phi(\lambda_2)+g(\lambda_2)(\lambda_1-\lambda_2) \quad \cdots\cdots ①$$

同理: $\phi(\lambda_2)\leqslant\phi(\lambda_1)+g(\lambda_1)(\lambda_2-\lambda_1) \quad \cdots\cdots ②$

$① + ② \Rightarrow (g(\lambda_2)-g(\lambda_1))(\lambda_1-\lambda_2)\geqslant 0$, $\lambda_1, \lambda_2\in[0,1]$

故 $g(\lambda)$是非增的于$[0,1]$，定义 $f_i^*=\min_{x\in\Theta} f_i(x)$，$i=1,2$

若$\lambda^*=1$，将 $\lambda_1=1$，$\lambda_2=\lambda\in[0,1)$代入①式

$$\phi(1)\leqslant\phi(\lambda)+g(\lambda)(1-\lambda)$$

$$\Rightarrow g(\lambda)(1-\lambda)\geqslant\phi(1)-\phi(\lambda)=\phi^*-\phi(\lambda)\geqslant 0$$

$$\Rightarrow g(\lambda)\geqslant 0$$

故由ϕ的连续性：

$\lim_{\lambda\to 1}\{f_2(x(\lambda))-g(\lambda)\cdot\lambda\}$ 极限∃

单调收敛定理

$$\phi^*=\lim_{\lambda\to 1}\phi(\lambda)$$

同理 $f_1(x(\lambda))$极限∃，故

$$=\lim_{\lambda\to 1}\min_{x\in\Theta}\{\lambda f_1(x)+(1-\lambda)f_2(x)\}$$

$\max\{\lim_{\lambda\to 1}f_1, \lim_{\lambda\to 1}f_2\}$

$$=\lim_{\lambda\to 1}\{\lambda f_1(x(\lambda))+(1-\lambda)f_2(x(\lambda))\}$$

$=\lim_{\lambda\to 1}\max(f_1, f_2)$

$=\lim_{\lambda\to 1}f(x(\lambda))$ ∃!

极限的存在性?

$$\geq \lim_{\lambda\to 1}\{\lambda f(x(\lambda)) + (1-\lambda) f_2^*\}$$

why? $\lambda f_1(x(\lambda)) \geq \lambda f(x(\lambda))$

$\Leftrightarrow \lambda f_1(x(\lambda)) \geq \lambda \max\{f_1(x(\lambda)), f_2(x(\lambda))\}$

· 若 $f_1(x(\lambda)) \geq f_2(x(\lambda))$ 时，成立

· 若 $f_1(x(\lambda)) \leq f_2(x(\lambda))$ 时，$\Leftrightarrow \lambda g(\lambda) \geq 0 \quad \lambda\in[0,1)$ ✓

$$= \lim_{\lambda\to 1} f(x(\lambda)) \geq f^* \quad (\text{极限的保号性})$$

若 $\lambda^*=0$，将 $\lambda_1=0$，$\lambda_2=\lambda\in(0,1]$ 代入①

$$\phi(0) \leq \phi(\lambda) + g(\lambda)(0-\lambda) \Rightarrow g(\lambda)\leq 0$$

故由 ϕ 的连续性：

$$\phi^* = \lim_{\lambda\to 0}\phi(\lambda)$$

$$= \lim_{\lambda\to 0}\min_{x\in\Omega}\{\lambda f_1(x) + (1-\lambda) f_2(x)\}$$

$$= \lim_{\lambda\to 0}\{\lambda f_1(x(\lambda)) + (1-\lambda) f_2(x(\lambda))\}$$

$$\geq \lim_{\lambda\to 0}\{\lambda f_1^* + (1-\lambda) f(x(\lambda))\}$$

$$= \lim_{\lambda \to 0} f(x(\lambda)) \geq f^*$$

下面考虑 $\lambda^* \in (0,1)$ 时

- 将 $\lambda_1 = \lambda^*$, $\lambda_2 = \lambda \in [0, \lambda^*)$ 代入 ① 时

$$\phi(\lambda^*) \leq \phi(\lambda) + g(\lambda)(\lambda^* - \lambda)$$

$$\Leftrightarrow g(\lambda) \geq 0, \quad \forall \lambda \in [0, \lambda^*)$$

- 将 $\lambda_1 = \lambda^*$, $\lambda_2 = \lambda \in (\lambda^*, 1]$ 代入 ① 时

$$g(\lambda) \leq 0, \quad \forall \lambda \in (\lambda^*, 1]$$

$\boxed{\text{case 1}}$ $\exists$ 序列 $\{\lambda_k\}_{k \geq 0} \subset [0,1]$, s.t.

$$\lambda_k \to \lambda^*, \quad g(\lambda_k) \longrightarrow 0 \quad k \to \infty$$

由 lemma 3.1.4

$$\phi^* = \lim_{k \to \infty} \min_{x \in Q} \{\lambda_k f_1(x) + (1-\lambda_k) f_2(x)\}$$

$$= \lim_{k \to \infty} \{\lambda_k f_1(x(\lambda_k)) + (1-\lambda_k) f_2(x(\lambda_k))\}$$

$$= \lim_{k \to \infty} \{f_2(x(\lambda_k)) + \lambda_k g(\lambda_k)\}$$

$$= \lim_{k\to\infty} f_2(x(\lambda_k))$$

同理:

$$\phi^* = \lim_{k\to\infty} \phi(\lambda_k)$$

$$= \lim_{k\to\infty} \min_{x\in \Omega} \{\lambda_k f_1(x) + (1-\lambda_k) f_2(x)\}$$

$$= \lim_{k\to\infty} \{\lambda_k f_1(x(\lambda_k)) + (1-\lambda_k) f_2(x(\lambda_k))\}$$

$$= \lim_{k\to\infty} \{f_1(x(\lambda_k)) - (1-\lambda_k) g(\lambda_k)\}$$

$$= \lim_{k\to\infty} f_1(x(\lambda_k))$$

$\max\{a,b\} = \frac{a+b+|b-a|}{2}$

由 $\max\{\cdot,\cdot\}$ 是连续 func，故

$$\phi^* = \max\{\lim_{k\to\infty} f_1(x(\lambda_k)), \lim_{k\to\infty} f_2(x(\lambda_k))\}$$

$$= \lim_{k\to\infty} \max\{f_1(x(\lambda_k)), f_2(x(\lambda_k))\}$$

$$= \lim_{k\to\infty} f(x(\lambda_k)) \geq f^*$$

$\boxed{\text{case 2}}$ 设不存在序列 $\{\lambda_k\}$ 满足 (3.1.14)，考虑:

$$\{\alpha_k\}_{k\geq 0} : \alpha_k \uparrow \lambda^*, \quad \{\beta_k\}_{k\geq 0} : \beta_k \downarrow \lambda^*$$

由(3.1.14)不满足，且 g 的单调性：$\exists\, a, b>0$, s.t.

$$\lim_{k\to\infty} g(\alpha_k)=a, \quad \lim_{k\to\infty} g(\beta_k)=-b$$

(由单调收敛定理，$g(\alpha_k)$，$g(\beta_k)$ 极限存在)

令 $\gamma=\frac{b}{a+b}$，则由lemma 3.1.4：

$$\begin{aligned}
\phi^* &= \lim_{k\to\infty}\left\{\gamma\phi(\alpha_k)+(1-\gamma)\phi(\beta_k)\right\} \\
&= \lim_{k\to\infty}\left\{\gamma\left[f_2(x(\alpha_k))+\alpha_k g(\alpha_k)\right]+(1-\gamma)\left[f_2(x(\beta_k))+\beta_k g(\beta_k)\right]\right\} \\
&= \lim_{k\to\infty}\left\{\gamma f_2(x(\alpha_k))+(1-\gamma) f_2(x(\beta_k))\right\} \\
&\geqslant \limsup_{k\to\infty} f_2(\gamma x(\alpha_k)+(1-\gamma)x(\beta_k))
\end{aligned}$$

同理：

$$\begin{aligned}
\phi^* = \lim_{k\to\infty}\Big\{&\gamma\left[f_1(x(\alpha_k))-(1-\alpha_k)g(\alpha_k)\right]+ \\
&(1-\gamma)\left[f_1(x(\beta_k))-(1-\beta_k)g(\beta_k)\right]\Big\}
\end{aligned}$$

$$= \lim_{k\to\infty}\left\{\gamma f_1(x(\alpha_k))+(1-\gamma) f_1(x(\beta_k))\right\}$$

$$\geq \limsup_{k\to\infty} f_1(\gamma x(\alpha_k)+(1-\gamma)x(\beta_k))$$

由序列 $\{f_i(\gamma x(\alpha_k)+(1-\gamma)x(\beta_k))\}_{k=0}^{+\infty} \subseteq [f_i^*, \phi^*]$ $i=1,2$

故可以取 $\{\alpha_{k_j}\}_{j=0}^{\infty}$ $\{\beta_{k_j}\}_{j=0}^{+\infty}$, s.t.

$$\{f_i(\gamma x(\alpha_{k_j})+(1-\gamma)x(\beta_{k_j})\}_{j=0}^{+\infty} \text{ 收敛}, i=1,2$$

故 $$\phi^* \geq \max\{\lim_{j\to+\infty} f_1(\gamma x(\alpha_{k_j})+(1-\gamma)x(\beta_{k_j})), \lim_{j\to+\infty} f_2(\gamma x(\alpha_{k_j})+(1-\gamma)x(\beta_{k_j}))\}$$

$$= \lim_{j\to+\infty} f(\gamma x(\alpha_{k_j})+(1-\gamma)x(\beta_{k_j})) \geq f^*$$

$\boxed{/}$

推论3.1.3 令 f_i, $i=1,\cdots,m$ 是 Ω 上的闭凸func, 约束水平集均有界, 则 $\exists \lambda_* \in \Delta_m$, s.t.

$$\min_{x\in\Omega}(F(x) \triangleq \max_{1\leq i\leq m} f_i(x)) = \min_{x\in\Omega}\left\{\sum_{i=1}^{m} \lambda_*^{(i)} f_i(x)\right\}$$

证明: 令 $F_k(x) = \max_{k\leq i\leq m} f_i(x)$

则 $F(x)=\max\{f_1(x), F_2(x)\}$

$$F_k(x)=\max\{f_k(x), F_{k+1}(x)\}, \quad k=2,\cdots,m-1$$

由 Thm 3.1.10，$\exists\ \lambda_*^{(1)}\in[0,1]$, s.t.

$$F^*\triangleq\min_{x\in\otimes}F(x)=\min_{x\in\otimes}\{\psi_1(x)=\lambda_*^{(1)}f_1(x)+(1-\lambda_*^{(1)})F_2(x)\}$$

$$=\min_{x\in\otimes}\max\{\lambda_*^{(1)}f_1(x)+(1-\lambda_*^{(1)})f_2(x),\ \lambda_*^{(1)}f_1(x)+(1-\lambda_*^{(1)})F_3(x)\}$$

又利用 Thm 3.1.10，$\exists\ \xi^*\in[0,1]$, s.t. $F^*=\min_{x\in\otimes}\psi_2(x)$

$$\psi_2(x)=\xi^*(\lambda_*^{(1)}f_1(x)+(1-\lambda_*^{(1)})f_2(x))+$$
$$(1-\xi^*)(\lambda_*^{(1)}f_1(x)+(1-\lambda_*^{(1)})F_3(x))$$

$$=\lambda_*^{(1)}f_1(x)+\xi^*(1-\lambda_*^{(1)})f_2(x)+(1-\xi^*)(1-\lambda_*^{(1)})F_3(x)$$

令 $\xi^*(1-\lambda_*^{(1)})=\lambda_*^{(2)}$，则

$$\psi_2(x)=\lambda_*^{(1)}f_1(x)+\lambda_*^{(2)}f_2(x)+(1-\lambda_*^{(1)}-\lambda_*^{(2)})F_3(x)$$

依此类推 ……

$\square$

§ 3.1.3 Continuity and Differentiability

Thm 3.1.11 令 f 凸，且 $x_0 \in \text{int}(\text{dom} f)$，则 f 是在 x_0 处局部有界且局部 Lipschiz 的

证明：先证 f 是 locally bounded 的，由 $x_0 \in \text{int}(\text{dom} f)$，

则 $\exists\, \varepsilon > 0$, s.t. $x_0 \pm \varepsilon e_i \in \text{int}(\text{dom} f)$, $i = 1, 2, \cdots, n$

定义：$\Delta = \text{conv}\{x_0 \pm \varepsilon e_i,\ i = 1, \cdots, n\} \overset{(3.1.8)}{=} B_1(x_0, \varepsilon)$

由 dom f 的凸性，$\Delta \subseteq \text{dom} f$，故由推论 3.1.2

$$\max_{x \in \Delta} f(x) = \max_{1 \le i \le n} f(x_0 \pm \varepsilon e_i) \triangleq M$$

即 x_0 处是 locally bounded 的，下证局部 Lipschiz：

考虑 $\forall\, y \in B_1(x_0, \varepsilon) \setminus \{x_0\}$，令

$$\alpha = \frac{1}{\varepsilon} \| y - x_0 \|_{(1)}, \quad z = x_0 + \frac{1}{\alpha}(y - x_0)$$

故 $\| z - x_0 \|_{(1)} = \frac{1}{\alpha} \| y - x_0 \|_{(1)} = \varepsilon$，$\alpha \le 1$ 且

$$y = \alpha z + (1-\alpha)x_0$$

故由 f 的凸性

$$\begin{aligned} f(y) &\le \alpha f(z) + (1-\alpha) f(x_0) \\ &\le f(x_0) + \alpha (M - f(x_0)) \\ &= f(x_0) + \frac{M - f(x_0)}{\varepsilon} \|y - x_0\|_{(1)} \end{aligned}$$

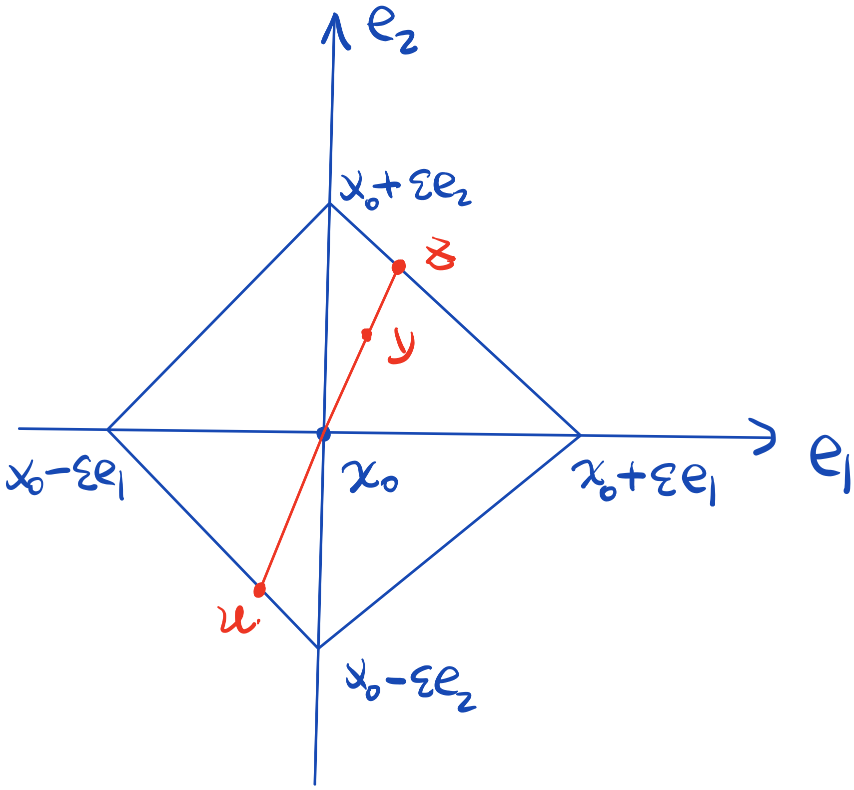
另一方面：令 $u = x_0 + \frac{1}{\alpha}(x_0 - y)$，则 $\|u - x_0\|_{(1)} = \varepsilon$

$$y = x_0 + \alpha (x_0 - u)$$

由定理 3.1.1

$$\begin{aligned} f(y) &\ge f(x_0) + \alpha (f(x_0) - f(u)) \\ &\ge f(x_0) - \alpha (M - f(x_0)) \\ &= f(x_0) - \frac{M - f(x_0)}{\varepsilon} \|y - x_0\|_{(1)} \end{aligned}$$

综上 $|f(y) - f(x_0)| \le \frac{M - f(x_0)}{\varepsilon} \|y - x_0\|_{(1)}$ $\square$



Def3.1.3 令 $x\in \mathrm{dom} f$，称 f 在 x_0 处关于方向 $p\neq 0$ 可微，

若下面极限存在：

$$f'(x;p)=\lim_{\alpha\downarrow 0}\frac{1}{\alpha}[f(x+\alpha p)-f(x)]$$

Thm3.1.12 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 $x\in \mathrm{int}(\mathrm{dom} f)$，考虑

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

取 ε 足够小, s.t. $x+\varepsilon p\in\mathrm{dom}f$, 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. $x-\gamma p\in\mathrm{dom}f$,

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 $\exists$

$\boxed{}$

lemma 3.1.5 令 f 是 convex 的且 $x\in \mathrm{int}(\mathrm{dom} f)$,

则 $f'(x;\cdot)$ 是 convex 的且正齐次的 (degree=1), 对 $\forall y\in$

$\mathrm{dom} f$, 有 $f(y)\geqslant f(x)+f'(x;y-x)$

证明: 首先证明方向导数是齐次的:

对 $\forall p\in\mathbb{R}^n\setminus\{0\}$ 时 $\tau>0$, 有

$$f'(x;\tau p)=\lim_{\alpha\downarrow 0}\frac{1}{\alpha}[f(x+\tau\alpha p)-f(x)]$$

$$=\tau\lim_{\beta\downarrow 0}\frac{1}{\beta}[f(x+\beta p)-f(x)]$$

$$=\tau f'(x_0;p)$$

当 $p=0$ 时, 补充定义:

$$f'(x;\tau p)=\tau f'(x;p)=0$$

故 $f'(x;\cdot)$ 是正齐次的. degree=1, 下证凸性:

对 $\forall p_1, p_2\in\mathbb{R}^n$, $\beta\in[0,1]$, 有:

$$f'(x;\beta p_1+(1-\beta)p_2)=\lim_{\alpha\downarrow 0}\frac{1}{\alpha}[f(x+\alpha(\beta p_1+(1-\beta)p_2))-f(x)]$$

$$\leq \lim_{\alpha \downarrow 0} \frac{1}{\alpha}\{\beta[f(x+\alpha p_1)-f(x)]+(1-\beta)[f(x+\alpha p_2)-f(x)]\}$$

$$= \beta f'(x;p_1)+(1-\beta)f'(x;p_2)$$

故 $f'(x;p)$ 是 p 的 convex func. 令 $\alpha \in (0,1]$, $y \in \mathrm{dom} f$,

$y_\alpha = x+\alpha(y-x)$, 故由定理 3.1.1

$$f(y) = f(y_\alpha + \frac{1}{\alpha}(1-\alpha)(y_\alpha - x))$$

$$\geq f(y_\alpha)+\frac{1}{\alpha}(1-\alpha)[f(y_\alpha)-f(x)]$$

对 $\alpha \downarrow 0$ 取极限:

$$f(y) \geq f(x)+f'(x;y-x)$$ $\boxed{/\!/}$

§3.1.4 分离定理

Def 3.1.4 令 Ω 是凸集, 超平面

$$\mathcal{H}(g,r) = \{x \in \mathbb{R}^n \mid \langle g, x\rangle = r\},\ g \neq 0$$

称 $\mathcal{H}(g,r)$ 是 Ω 的 supporting 若对 $\forall x \in \Omega$, $\langle g, x\rangle \leq r$

称 $H(g,r)$ 分离 x_0 和 Θ 若对 $\forall x\in\Theta$，

$$\langle g,x\rangle \le r \le \langle g,x_0\rangle \qquad (3.1.19)$$

若 (3.1.19) 的一个不等号是严格的，则称是强分离的 $\boxed{\angle}$

同理可以定义凸集之间的分离性，两个凸集 Θ_1，Θ_2 是可分的，若 $\exists g\in\mathbb{R}^n$，$g\ne 0$，$r\in\mathbb{R}^n$，s.t.

$$\langle g,x\rangle \le r \le \langle g,y\rangle,\ \forall x\in\Theta_1,\ \forall y\in\Theta_2$$

分离是严格的，若

$$\sup_{x\in\Theta_1}\langle g,x\rangle < r < \inf_{y\in\Theta_2}\langle g,y\rangle$$

<u>定理 3.1.13</u> 令 Θ_1，Θ_2 是 $\mathbb{R}^n$ 中的闭凸集，s.t. $\Theta_1\cap\Theta_2=\phi$，设 Θ_1，Θ_2 中有一个集合是有界的，则 Θ_1，Θ_2 是强分离的

证明：不妨设 Θ_1 是有界的，考虑极小化问题：

$$\rho^* = \min_{x\in\Theta_1} \rho_{\Theta_2}(x)$$

$P_{\Theta_2}(x)=\frac{1}{2}\|x-\pi_{\Theta_2}(x)\|^2$，而 $\pi_{\Theta_2}(x_0)=\underset{x\in\Theta_2}{\arg\min}\|x-x_0\|$

故由Thm 2.2.11 知，$\pi_{\Theta_2}(x_0)$ 是存在唯一的，故 $P_{\Theta_2}(x)$

是良定义的，由Lemma 2.2.9 知，$P_{\Theta_2}(x)$ 是 $\mathbb{R}^n$ 上的 $\mathcal{F}^1$ func，

由 Θ_1 的紧性知：P^* 是良定义的

对 $\forall x^*\in X^*$，有：

$$\nabla P_{\Theta_2}(x^*)\overset{(2.2.41)}{=}g^*，\langle g^*,x^*\rangle\overset{(2.2.4)}{=}\gamma^*$$

故对 $\forall x_1\in\Theta_1$，有：

$$\begin{aligned}\langle g^*,x_1\rangle-\gamma^*&=\langle g^*,x_1-x^*\rangle\\&=\langle\nabla P_{\Theta_2}(x^*),x_1-x^*\rangle\overset{(2.2.39)}{\geq}0\end{aligned}$$

对 $\forall x_2\in\Theta_2$，有

$$\begin{aligned}\langle g^*,x_2\rangle-\gamma^*&=\langle g^*,x_2-x^*\rangle\\&=\langle x^*-\pi_{\Theta_2}(x^*),x_2-x^*\rangle\quad(2.2.52)\\&=\langle\pi_{\Theta_2}(x^*)-x^*,(x^*-\pi_{\Theta_2}(x^*))-(x_2-\pi_{\Theta_2}(x^*))\rangle\end{aligned}$$

$$= -\|\pi_{\Theta_2}(x^*) - x^*\|^2 - \langle \pi_{\Theta_2}(x^*) - x^*, x_2 - \pi_{\Theta_2}(x^*) \rangle$$

$$\overset{(2.2.47)}{\leqslant} -\|\pi_{\Theta_2}(x^*) - x^*\|^2 = -2\rho^*$$

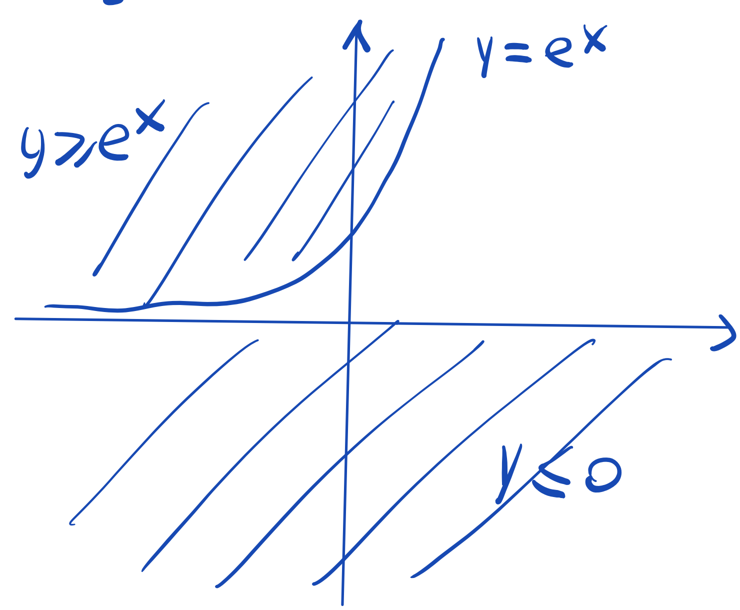
故 $\langle g^*, x_2 \rangle \leq r^* - 2\rho^* < r^* \leq \langle g^*, x_1 \rangle \quad \forall x_1 \in \Theta_1, \ \forall x_2 \in \Theta_2$

由 $\Theta_1 \cap \Theta_2 = \phi$，故 $\rho^* > 0$

故 $\sup\limits_{x_2 \in \Theta_2} \langle g^*, x_2 \rangle < r^* - \rho^* < \inf\limits_{x_1 \in \Theta_1} \langle g^*, x_1 \rangle$

$\square$

注：Θ_1，Θ_2 有一个有界是必要的



不是强分离的

<u>推论3.1.4</u> 令 Θ 是闭凸集，$x \notin \Theta$，则 x 与 Θ 强分离

推论3.1.5 令 $\otimes_1, \otimes_2$ 是两个闭凸集

1. 若 $\xi_{\otimes_1}(g) \le \xi_{\otimes_2}(g)$ 对 $\forall g \in \operatorname{dom} \xi_{\otimes_2}$ 成立，则 $\otimes_1 \subseteq \otimes_2$

2. 令 $\operatorname{dom} \xi_{\otimes_1} = \operatorname{dom} \xi_{\otimes_2}$，且对 $\forall g \in \operatorname{dom} \xi_{\otimes_1}$，有：

$\xi_{\otimes_1}(g) = \xi_{\otimes_2}(g)$，有 $\otimes_1 = \otimes_2$

证明 $\quad \xi_{\otimes}(g) = \sup\{\langle x, g\rangle \mid x \in \otimes\}$

1. 设 $\exists x_0 \in \otimes_1$, s.t. $x_0 \notin \otimes_2$，由推论3.1.4，$\exists g$, s.t.

$$\langle g, x_0\rangle > r \ge \langle g, x\rangle$$

对 $\forall x \in \otimes_2$ 成立，故 $g \in \operatorname{dom} \xi_{\otimes_2}$，且

$$\xi_{\otimes_1}(g) > \xi_{\otimes_2}(g) \qquad \text{矛盾}$$

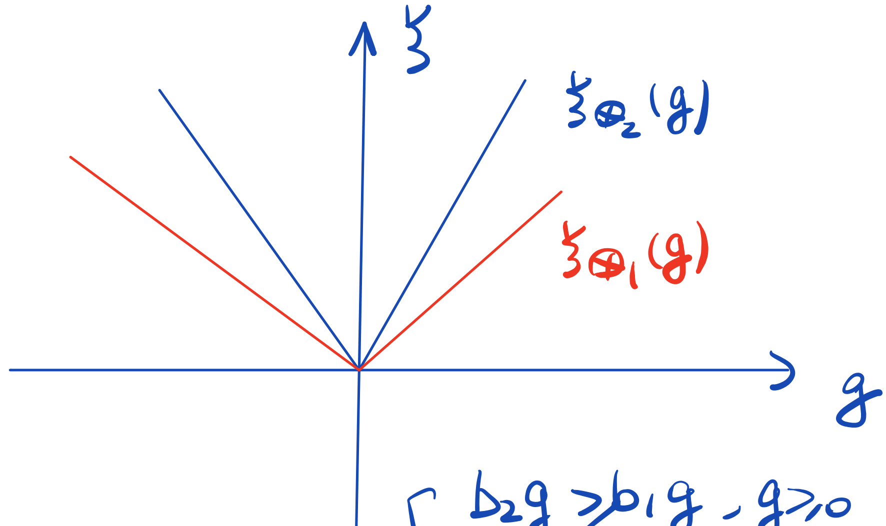
2. 用两次1即证 $\square$

一维下：设 $\otimes_1 = [a_1, b_1]$，$\otimes_2 = [a_2, b_2]$

① 设 $a_1, a_2 < 0$，$b_1, b_2 > 0$

则 $\xi_{\otimes_1}(g)=\sup\limits_{x\in[a_1,b_1]} xg=\begin{cases} b_1 g & g\geq 0 \\ a_1 g & g<0 \end{cases}$

$\xi_{\otimes_2}(g)=\sup\limits_{x\in[a_2,b_2]} xg=\begin{cases} b_2 g & g\geqslant 0 \\ a_2 g & g<0 \end{cases}$

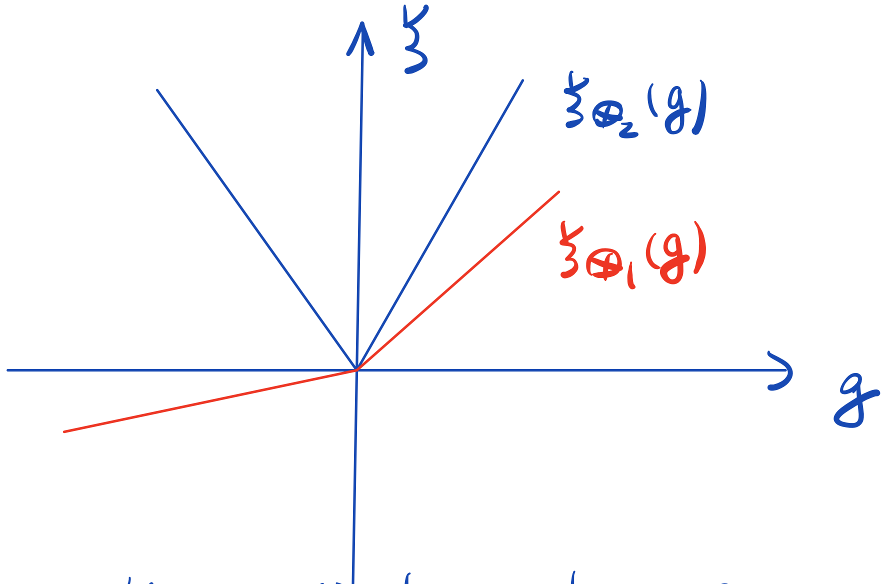


$\xi_{\otimes_2}(g)\geqslant\xi_{\otimes_1}(g) \iff \begin{cases} b_2 g\geqslant b_1 g,\ g\geqslant 0 \\ a_2 g\geqslant a_1 g,\ g<0 \end{cases}$

$\iff \begin{cases} b_2\geqslant b_1 \\ a_2\leqslant a_1 \end{cases} \iff \otimes_2\supseteq\otimes_1$

② 设 $a_2<0$、$a_1, b_1, b_2>0$，则

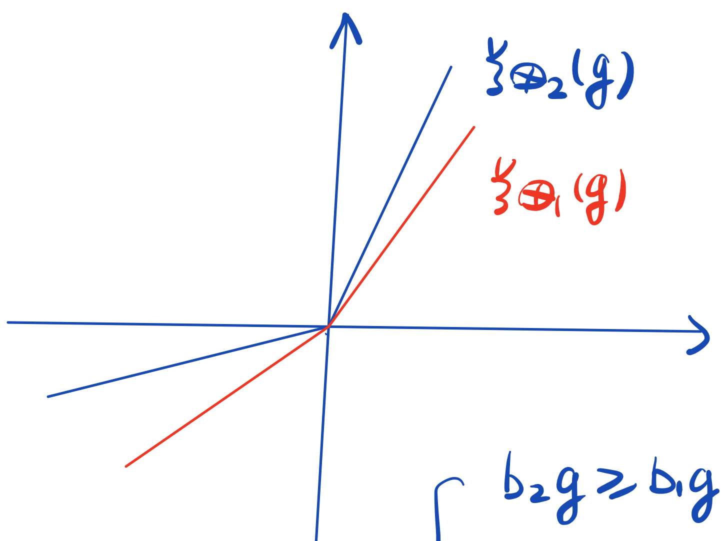
$\xi_{\otimes_1}(g)=\begin{cases} b_1 g,\ g\geqslant 0 \\ a_1 g,\ g\leqslant 0 \end{cases}$



$$\xi_{\Theta_2}(g) \geqslant \xi_{\Theta_1}(g) \iff b_2 g \geqslant b_1 g,\quad g \geqslant 0$$

$$\iff b_2 \geqslant b_1 \iff \Theta_2 \geqq \Theta_1$$

③ 设 $a_1, b_1, a_2, b_2 > 0$



$$\xi_{\Theta_2}(g) \geqslant \xi_{\Theta_1}(g) \iff \begin{cases} b_2 g \geqslant b_1 g & g \geqslant 0 \\ a_2 g \geqslant a_1 g & g < 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} b_2 \geqslant b_1 \\ a_2 \leqslant a_1 \end{cases} \Rightarrow \Theta_2 \geqq \Theta_1$$

定理 3.1.14　令 Ω 是闭凸集，若 $x_0 \in \partial\Omega$，则 $\exists\, \Omega$ 的支撑超平面 $H(g,r)$，且 $x_0 \in H(g,r)$

证明：考虑序列 $\{y_k\}$, s.t. $y_k \notin \Omega$，且 $y_k \to x_0$，令

$$g_k = \frac{y_k - \pi_\Omega(y_k)}{\|y_k - \pi_\Omega(y_k)\|}, \quad r_k = \langle g_k, \pi_\Omega(y_k) \rangle$$

由 Lemma 2.2.7，对 $\forall x \in \Omega$，有：

$$\langle \pi_\Omega(y_k) - y_k,\ x - \pi_\Omega(y_k) \rangle \geq 0$$

$$\Rightarrow \langle g_k, x \rangle \leq r_k \leq \langle g_k, y_k \rangle \quad \cdots\cdots \quad (*)$$

由 $\|g_k\| = 1$，由 Lemma 2.2.8，序列 $\{r_k\}$ 是有界的：

$$|r_k| = |\langle g_k, \pi_\Omega(y_k) - x_0 \rangle + \langle g_k, x_0 \rangle|$$

$$\leq \|\pi_\Omega(y_k) - x_0\| + \|x_0\|$$

$$\leq \|y_k - x_0\| + \|x_0\|$$

故 WLOG，设 $\exists\, g^* = \lim_{k\to\infty} g_k$，$r^* = \lim_{k\to\infty} r_k$

对⊗取极限：对 $\forall x \in \otimes$

$$\langle g^*, x\rangle \leq r^* \leq \langle g^*, x_0\rangle$$

故取 $H(g^*, r^*) = \{x \in \mathbb{R}^n \mid \langle g^*, x\rangle = r^*\}$ 即证

$\Box$

§3.1.5 Subgradient

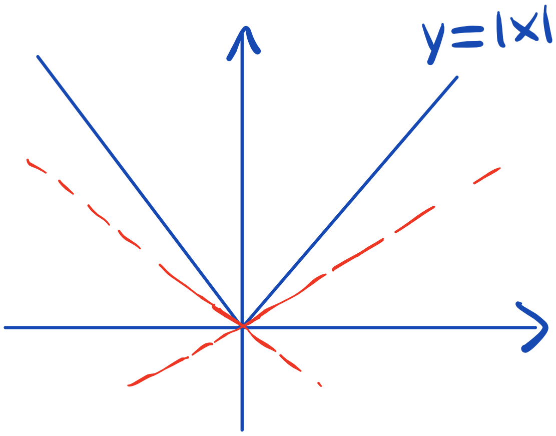
<u>Def 3.1.5</u> 向量 g 称为 f 在 $x_0 \in \mathrm{dom} f$ 的次梯度，若

对 $\forall y \in \mathrm{dom} f$，有 $f(y) \geq f(x_0) + \langle g, y - x_0\rangle$ (3.1.23)

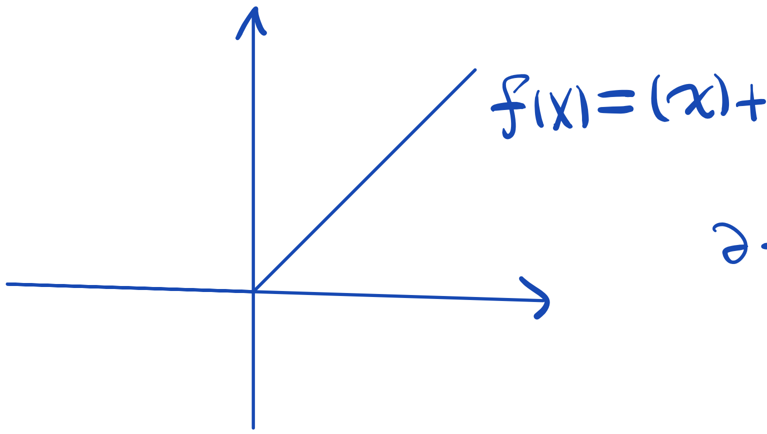
所有 f 在 x_0 的次梯度构成集合记为 $\partial f(x_0)$ 称为 f 在 x_0 处的次微分

若 (3.1.23) 只对 $y \in \otimes$ 成立，则用记号 $g \in \partial_{\otimes} f(x_0)$ 表示约束次微分，若 $\otimes \subseteq \mathrm{dom} f$，则 $\partial f(x_0) \subseteq \partial_{\otimes} f(x_0)$

$\partial f(x_0)$ 对于非凸 func 也可能非空



$\partial f(0)=[-1,1]$



$\partial f(0)=[0,1]$

$$
\begin{aligned}
\partial_{\otimes} f(x_0) &= \bigcap_{y\in\otimes} \partial_{\{y\}} f(x_0) \\
&= \bigcap_{y\in\otimes} \{ g\in\mathbb{R}^n \mid f(y) \geq f(x_0)+\langle g, y-x_0\rangle \} \\
&= \bigcap_{y\in\otimes} \mathcal{L}_{\phi}(f(y))
\end{aligned}
$$

其中 $\phi(g)=\langle g, y-x_0\rangle + f(x_0)$ ，$g\in\mathbb{R}^n$ 是闭凸func

$\Rightarrow \mathcal{L}_{\phi}(x)$ 是闭凸集或空集，又由 $\phi(g)$ 是线性func.

故 $\mathcal{L}_{\phi}(x)$ 对 $\forall x\in\mathbb{R}$ 是闭凸集

故 $\partial_{\otimes} f(x_0)$是闭凸集（空集也算进闭凸集）

<u>lemma 3.1.6</u> 令 $\otimes$是闭凸集，设对 $\forall x \in \otimes \subseteq \mathrm{dom} f$，

$\partial_{\otimes} f(x) \neq \phi$，则 f是$\otimes$上的闭凸集

证明：对 $\forall x \in \otimes$，定义

$$\hat{f}(x) = \sup_{y \in \otimes} \{ f(y) + \langle g(y), x-y \rangle \} \geq f(x)$$

其中 $g(y)$是 $\partial_{\otimes} f(y)$中任意次梯度

由<u>Thm 3.1.8</u>，$\hat{f}$是$\hat{\otimes}$上 闭凸func，且由次梯度定义：

$$f(x) \overset{(3.1.23)}{\geq} \hat{f}(x) \quad \forall x \in \otimes$$

故 $\hat{\otimes} = \otimes$，$\hat{f}$是$\otimes$上的闭凸func

故 $f(x) = \hat{f}(x)$ 是闭凸func

$\boxed{/}$

<u>Thm 3.1.15</u> 令 f是凸func，若$x_0 \in \mathrm{int}(\mathrm{dom} f)$，

则 $\partial f(x_0)$是非空有界的

证明：由 $(x_0, f(x_0)) \in \partial\, \mathrm{epi}(f)$，则

$(x_0, f(x_0)) \in \partial\, \overline{\mathrm{epi}(f)}$

注：$\partial A \subseteq \partial \overline{A}$ 并不一定正确，如 $\boxed{\partial A \triangleq \overline{A} \setminus A^\circ}$

$A = [0,1] \setminus \{\frac{1}{2}\}$，则 $\partial A = \{0, 1, \frac{1}{2}\}$

$\overline{A} = [0,1]$，则 $\partial \overline{A} = \{0, 1\}$

则 $\frac{1}{2} \in \partial A$，但 $\frac{1}{2} \notin \partial \overline{A}$，但对凸集，$\partial A \subseteq \partial \overline{A}$

证明：$\partial A \subseteq \partial \overline{A} \Leftrightarrow \overline{A} \setminus \mathrm{int}(A) \subseteq \overline{A} \setminus \mathrm{int}(\overline{A})$

$\Leftrightarrow \mathrm{int}(\overline{A}) \subseteq \mathrm{int}(A)$

要证 $\mathrm{int}(\overline{A}) \subseteq \mathrm{int}(A)$，只需证 $\mathrm{int}(\overline{A}) \subseteq A$

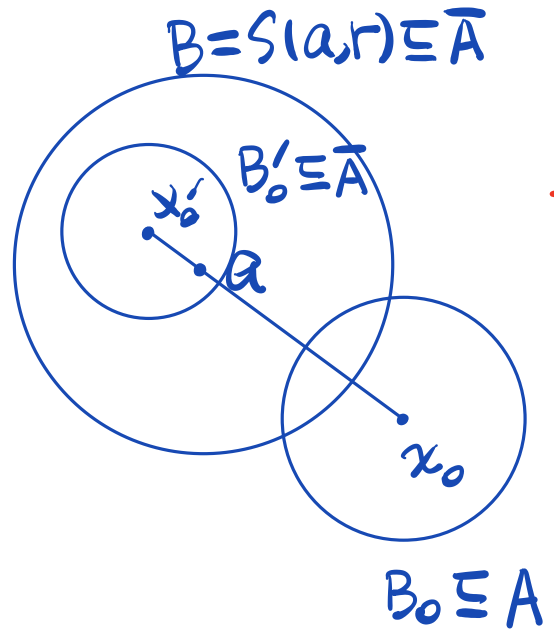
因为对 $\forall a \in \mathrm{int}(\overline{A})$，则 $\exists$ 开球 $S(a,r) \subseteq \mathrm{int}(\overline{A})$，

故 $S(a,r) \subseteq A$，故 $a \in \mathrm{int}(A)$，即证 $\mathrm{int}(\overline{A}) \subseteq \mathrm{int}(A)$

下证 $\mathrm{int}(\overline{A}) \subseteq A$，对 $\forall a \in \mathrm{int}(\overline{A})$，则 $\exists$ 开球 $B = S(a, r_1)$

$\subseteq \mathrm{int}(\overline{A})$，取 $x_0 \in \mathrm{int}(A)$，取开球 $B_0 = S(x_0, r_2) \subseteq A$

$t x_0 + (1-t) x_0'$，$t \in (0,1]$



取 x_0', s.t. $a\in[x_0,x_0')$,

且映射 $h(x)=a+k(x-a)$ 满足

$$\begin{cases} h(x_0)=x_0' \\ h: B_0 \longrightarrow B_0'\subseteq B \end{cases}$$

由 B_0' 开，故可以取 $x'\in B_0'$, s.t.

$x'\in A$，否则对 $\forall y\in B_0'$，有 $y\notin A$，但 $y\in\overline{A}$，故

$$B_0'\subseteq\overline{A}\setminus A\subseteq\overline{A}\setminus \mathrm{int}(A)=\partial A$$

与 ∂A 的定义矛盾，故 $\exists x'\in B_0'$, s.t. $x'\in A$

设 $x'=h(x)$，$x\in B_0\subseteq A$，则 $a\in[x,x')$，由 $x\in A$，$x'\in A$.

故由 A 的凸性，$a\in A$

故对 $\overline{\mathrm{epi}(f)}$ 用定理3.1.14，$\exists$ 超平面在 $(x_0,f(x_0))$ 处支撑

$\overline{\mathrm{epi}(f)}$：$-\alpha\tau+\langle d,x\rangle\leqslant-\alpha f(x_0)+\langle d,x_0\rangle$

对 $\forall(x,\tau)\in\overline{\mathrm{epi}(f)}$，归一化系数，s.t.

$$\|d\|^2+\alpha^2=1$$

由 $(x_0,\tau)\in \mathrm{epi}(f)$ 对 $\forall \tau\geqslant f(x_0)$ 成立，故 $\alpha\geqslant 0$

由定理 3.1.11，凸 func 是局部 Lipschiz 的，即

$\exists \varepsilon>0,\ M>0$, s.t. $B_2(x_0,\varepsilon)\subseteq \mathrm{dom} f$，且

$$f(x)-f(x_0)\leqslant M\|x-x_0\|$$

对 $\forall x\in B_2(x_0,\varepsilon)$ 成立，故对 $\forall x\in B_2(x_0,\varepsilon)$

$$\langle d, x-x_0\rangle \leqslant \alpha(f(x)-f(x_0))\leqslant \alpha M\|x-x_0\|$$

取 $x=x_0+\varepsilon d$，有 $\|d\|^2\leqslant M\alpha\|d\|$，则

$$\begin{cases} \|d\|=0 \text{时}，\ \alpha=1 \\ \|d\|>0 \text{时}\quad \sqrt{1-\alpha^2}\leqslant M\alpha \Rightarrow \alpha\geqslant [1+M^2]^{-\frac{1}{2}} \end{cases}$$

故 $\alpha>0$，取 $g=\frac{d}{\alpha}$，则

$$f(x)\overset{(3.1.25)}{\geqslant} f(x_0)+\langle g, x-x_0\rangle \quad 对\ \forall x\in \mathrm{dom} f$$

故 $\partial f(x_0)$ 是非空的，若 $g\in\partial f(x_0)$，$g\neq 0$，取 $x=x_0+\frac{\varepsilon g}{\|g\|}$

$$\Rightarrow \varepsilon\|g\|=\langle g, x-x_0\rangle\leqslant f(x)-f(x_0)\leqslant M\|x-x_0\|=M\varepsilon$$

$\Rightarrow \|g\| \leq M$，故 $\partial f(x_0)$ 是有界的

$\boxed{/}$

例 3.1.4　$x_0 \notin \mathrm{int}(\mathrm{dom} f)$ 时上面定理不成立.

$f(x) = -\sqrt{x}$，$x \in \mathbb{R}_+$，在 0 处次微分不存在

定理 3.1.16　对 func f，定义 Fenchel dual

$$f_*(s) = \sup_{y \in \mathrm{dom} f} [\langle s, y \rangle - f(y)]$$

和 f_* 的 Fenchel Dual

$$f_{**}(x) = \sup_{s \in \mathrm{dom} f_*} [\langle s, x \rangle - f_*(s)]$$

则 $f(x) \geq f_{**}(x)$ 对 $\forall x \in \mathrm{dom} f$ 成立，若 $\partial f(x) \neq \emptyset$ 对 $x \in \mathrm{dom} f$ 成立，则 $\partial f(x) \subseteq \mathrm{dom} f_*$，且 $f(x) = f_{**}(x)$

证明：对 $\forall x \in \mathrm{dom} f$，有

$$f_{**}(x) = \sup_{s \in \mathrm{dom} f_*} [\langle s, x \rangle - f_*(s)]$$

$$= \sup_{s\in \mathrm{dom} f_*} \inf_{y\in \mathrm{dom} f} [\langle s,x\rangle - \langle s,y\rangle + f(y)]$$

$$\leqslant \inf_{y\in \mathrm{dom} f} \sup_{s\in \mathrm{dom} f_*} [\langle s,x-y\rangle + f(y)]$$

$$\leqslant f(x)$$

对 $\forall g\in \partial f(x)$，则对 $\forall y\in \mathrm{dom} f$，

$$\langle g,y\rangle - f(y) \leqslant \langle g,y\rangle - f(x) - \langle g,y-x\rangle$$

$$= \langle g,x\rangle - f(x)$$

故 $g\in \mathrm{dom} f_*$，故

$$f_{**}(x) = \sup_{s\in \mathrm{dom} f_*} \inf_{y\in \mathrm{dom} f} [\langle s,x\rangle - \langle s,y\rangle + f(y)]$$

$$\geqslant \inf_{y\in \mathrm{dom} f} [\langle g,x\rangle - \langle g,y\rangle + f(y)]$$

$$\geqslant f(x)$$

$\square$

Thm3.1.7 设 f 是convex的，$x_0 \in \mathrm{int}(\mathrm{dom} f)$，则

$$\partial_2 f'(x_0;0) = \partial f(x)$$

对 $\forall p \in \mathbb{R}^n$，有

$$f'(x_0;p) = \max\{\langle g,p\rangle \mid g \in \partial f(x_0)\}$$

证明：注意到：

$$f'(x_0;p) = \lim_{\alpha \downarrow 0} \frac{1}{\alpha}[f(x_0+\alpha p) - f(x_0)] \geq \langle g,p\rangle$$

其中 $g \in \partial f(x_0)$ 是任意次梯度，故

$$f'(x_0;p) \geq f'(x_0;0) + \langle g,p\rangle$$

即 $\partial f(x_0) \subseteq \partial_2 f'(x_0;0)$，另一方面 $f'(x_0;p)$ 关于 p 是 convex 的，且由 lemma 3.1.5，对 $\forall y \in \mathrm{dom} f$，有

$$\begin{aligned} f(y) &\geq f(x_0) + f'(x_0;y-x_0) \\ &\geq f(x_0) + \langle g, y-x_0\rangle \end{aligned}$$

其中 $g \in \partial_2 f'(x_0;0)$，由 $f'(x_0;p)$ 的凸性，$\partial_2 f'(x_0;0)$ 非空，

故 $\partial_2 f'(x_0;0) \subseteq \partial f(x_0)$，综上

$$\partial_2 f'(x_0;0) = \partial f(x_0)$$

$\forall g \in \partial_2 f'(x_0;p)$，则由 (3.1.18)，对 $\forall v \in \mathbb{R}^n$，$\tau > 0$

$$\tau f'(x_0;v) = f'(x_0;\tau v) \geqslant f'(x_0;p) + \langle g, \tau v - p \rangle$$

令 $\tau \to \infty$： $f'(x_0;v) \geqslant \langle g, v \rangle$ ①

令 $\tau \to 0$： $f'(x_0;p) - \langle g, p \rangle \leqslant 0$ ②

由①知，$g \in \partial_2 f'(x_0;0)$，故 $\partial_2 f'(x_0;p) \subseteq \partial_2 f'(x_0;0)$

$$\max_{g \in \partial_2 f'(x_0;0)} \langle g, p \rangle \overset{\text{由 } \partial_2 f'(x_0;0) \supseteq \partial_2 f'(x_0;p)}{\geqslant} \max_{g \in \partial_2 f'(x_0;p)} \langle g, p \rangle$$

$$\geqslant \min_{g \in \partial_2 f'(x_0;p)} \langle g, p \rangle \overset{②}{\geqslant} f'(x_0;p) \overset{\text{次梯度定义}}{\geqslant} \max_{g \in \partial_2 f'(x_0;0)} \langle g, p \rangle$$

max，min 能取到，由 <u>Thm3.1.15</u> 知，$\partial_2 f'(x_0;0)$，

$\partial_2 f'(x_0;p)$ 均是紧的 $\square$

Thm 3.1.8 对 $\forall x_0\in \mathrm{dom} f$, $\forall g\in\partial f(x_0)$ 是下水平集 $\mathcal{L}_f(f(x_0))$ 的支撑：

$\langle g, x_0-x\rangle \geqslant 0$, $\forall x\in \mathcal{L}_f(f(x_0))=\{x\in \mathrm{dom} f: f(x)\leq f(x_0)\}$

证明：若 $f(x)\leqslant f(x_0)$, $g\in\partial f(x_0)$, 则

$$f(x_0)+\langle g, x-x_0\rangle \leqslant f(x)\leqslant f(x_0)$$

$\boxed{\diagup}$

推论 3.1.6 令 $Q\subseteq \mathrm{dom} f$ 是闭凸集, $x_0\in Q$, 且

$$x^*\in \operatorname*{Argmin}_{x\in Q} f(x)$$

则对 $\forall g\in\partial f(x_0)$, 有 $\langle g, x_0-x^*\rangle \geqslant 0$

$\boxed{\diagup}$

Def 3.1.6 令 $X\subseteq \mathrm{dom} f$ 是闭凸集, 集合

$$\widehat{\partial f}(X)=\bigcap_{x\in X}\partial f(x)$$

叫做集合的上图面

Thm 3.1.19 令集合 X 是闭凸集，且 $\widehat{\partial f}(X) \neq \emptyset$，则

$$f((1-\alpha)x_0+\alpha x_1) = (1-\alpha)f(x_0) + \alpha f(x_1), \quad \forall x_0, x_1 \in X, \ \alpha \in [0,1]$$

且对 $\forall g \in \widehat{\partial f}(X)$，$\forall x_0, x_1 \in X$，有：

$$f(x_1) = f(x_0) + \langle g, x_1 - x_0 \rangle$$

证明：令 $g \in \widehat{\partial f}(X) \subseteq \partial f(x_0) \cap \partial f(x_1)$，则

$$f(x_0) + \langle g, x_1 - x_0 \rangle \leqslant f(x_1) \leqslant f(x_0) + \langle g, x_1 - x_0 \rangle$$

$$\Rightarrow \quad f(x_1) = f(x_0) + \langle g, x_1 - x_0 \rangle$$

令 $x_\alpha = (1-\alpha)x_0 + \alpha x_1 \quad \alpha \in [0,1]$

$$(1-\alpha)f(x_0) + \alpha f(x_1) \geqslant f(x_\alpha) \geqslant f(x_0) + \langle g, x_\alpha - x_0 \rangle$$

$$= f(x_0) + \alpha \langle g, x_1 - x_0 \rangle = (1-\alpha)f(x_0) + \alpha f(x_1)$$

$\boxtimes$

<u>Thm 3.1.20</u> 令 $X^* = \operatorname{Argmin}\limits_{x\in \mathrm{dom} f} f(x)$，则闭凸集 X_* 是 X^* 的子集 $\iff 0\in \widehat{\partial f}(X_*)$

证明：若 $0\in \widehat{\partial f}(X_*)$，则对 $\forall x^*\in X_*$，有

对 $\forall x\in \mathrm{dom} f$，有 $f(x)\geqslant f(x^*)+\langle 0, x-x^*\rangle = f(x^*)$

故 $x^*\in X^*$

另一方面，若 $f(x)\geqslant f(x^*)$ 对 $\forall x\in \mathrm{dom} f$，$\forall x^*\in X_*$

成立，则 $0\in \bigcap\limits_{x^*\in X_*} \partial f(x^*) = \widehat{\partial f}(X_*)$

$\square$

§ 3.1.6 Computing Subgradient

<u>lemma 3.1.7</u> f是凸的，设在$x\in \mathrm{int}(\mathrm{dom} f)$是可微的，则$\partial f(x)=\{\nabla f(x)\}$

证明：对$\forall p\in\mathbb{R}^n$，有

$$
\begin{aligned}
f'(x;p) &= \langle \nabla f(x), p\rangle \\
&= \max\{\langle g, p\rangle \mid g\in\partial f(x)\} \\
&= \xi_{\partial f(x)}(p)
\end{aligned}
$$

故由推论3.1.5 (ii) $\partial f(x)=\{\nabla f(x)\}$ $\boxed{/}$

注：推论3.1.5 (ii)的条件 $\mathrm{dom}\,\xi_{Q_1}=\mathrm{dom}\,\xi_{Q_2}$ 是成立的，

由<u>Thm3.1.5</u>，$\partial f(x)$是有界的，故 $\mathrm{dom}\,\xi_{\partial f(x)}=\mathrm{dom}\,\xi_{\{\nabla f(x)\}}=\mathbb{R}^n$

<u>lemma 3.1.8</u> 令$\psi(\cdot)$是凸的，φ是单变量凸func，

在集合$\mathrm{Im}\psi=\{\tau=\psi(x),\ x\in\mathrm{dom} f\}$是非减的，则$f(\cdot)=$

$\varphi(\psi(\cdot))$是凸的，且对$\forall x\in \mathrm{int}(\mathrm{dom}\psi)$，有

$$\partial f(x)=\mathrm{Conv}\{\lambda\partial\psi(x),\ \lambda\in\partial\varphi(\psi(x))\}$$

证明：由Thm3.1.9，f是凸的，固定$\forall x\in\mathrm{int}(\mathrm{dom}\psi)$，$\forall p\in\mathbb{R}$

$$f'(x;p)=\lim_{\alpha\downarrow 0}\frac{1}{\alpha}[f(x+\alpha p)-f(x)]$$

$$=\lim_{\alpha\downarrow 0}\frac{1}{\alpha}[\varphi(\psi(x+\alpha p))-\varphi(\psi(x))]$$

$$=\lim_{\alpha\downarrow 0}\frac{1}{\alpha}\left[\varphi\left[\alpha\cdot\frac{1}{\alpha}[\psi(x+\alpha p)-\psi(x)]+\psi(x)\right]-\varphi(\psi(x))\right]$$

$$=\lim_{\alpha\downarrow 0}\frac{1}{\alpha}[\varphi(\psi(x)+\alpha\psi'(x;p)+o(\alpha))-\varphi(\psi(x))]$$

$$=\varphi'(\psi(x);\psi'(x;p))$$

$$=\max_{\lambda}\{\lambda\psi'(x;p):\lambda\in\partial\varphi(\psi(x))\}$$

$$=\max_{\lambda}\{\lambda\max_{\tilde{g}}\{\langle\tilde{g},p\rangle:\tilde{g}\in\partial\psi(x)\}:\lambda\in\partial\varphi(\psi(x))\}$$

$$=\max_{\lambda,g}\{\langle g,p\rangle:g\in\lambda\partial\psi(x),\ \lambda\in\partial\varphi(\psi(x))\}$$

$= \xi_{\mathcal{Q}}(p)$

其中 $\mathcal{Q} = \{\lambda \partial \psi(x) : \lambda \in \partial \varphi(\psi(x))\}$

下证 $\xi_{\mathcal{Q}}(p) \equiv \xi_{\hat{\mathcal{Q}}}(p)$，其中 $\hat{\mathcal{Q}} = \mathrm{Conv}\,\mathcal{Q}$

① $\xi_{\mathcal{Q}}(p) \leq \xi_{\hat{\mathcal{Q}}}(p)$ 是显然的

② 下证 $\xi_{\mathcal{Q}}(p) \geq \xi_{\hat{\mathcal{Q}}}(p)$，先证

$$\mathrm{Conv}(\mathcal{Q}) = \Big\{ \underbrace{\sum_{i=1}^{n} \lambda_i q_i \,\Big|\, n \in \mathbb{N},\ \sum_{i=1}^{n} \lambda_i = 1,\ \lambda_i \geq 0,\ q_i \in \mathcal{Q}}_{R \,//\, \Delta} \Big\}$$

"$\supseteq$" 考虑 $\forall$ 凸集 $C \supseteq \mathcal{Q}$，则由定义，$C \supseteq R$，由 C 的任意性知，$\mathrm{Conv}(\mathcal{Q}) \supseteq R$

"$\subseteq$" 只需证 R 是凸集，对 $\forall p_1, p_2 \in R$，不失一般地，

设 $p_1 = \sum_{i=1}^{n} \lambda_i q_i$，$p_2 = \sum_{i=1}^{n} \mu_i q_i$，则 $\forall \alpha \in [0,1]$

$$\alpha p_1 + (1-\alpha) p_2 = \sum_{i=1}^{n} \big(\alpha \lambda_i + (1-\alpha)\mu_i\big) q_i \in R$$

$$\sum_{i=1}^{n} \big(\alpha \lambda_i + (1-\alpha)\mu_i\big) = 1$$

故对 $\forall g\in\hat{\mathcal{Q}}$, $g=\sum_{i=1}^{n}\xi_i g_i$, 则

$$\langle g,p\rangle=\left\langle\sum_{i=1}^{n}\xi_i g_i,p\right\rangle=\sum_{i=1}^{n}\xi_i\langle g_i,p\rangle\leq\max_{1\leq i\leq n}\langle g_i,p\rangle$$

$\mathcal{Q}$
$\uparrow$

最后只需证明：$\hat{\mathcal{Q}}=\mathrm{Conv}(\mathcal{Q})$ 是闭集

① 先证 $\mathcal{Q}$ 是紧的，$\mathcal{Q}$ 的有界性（$\mathrm{dom}\,\varphi=\mathbb{R}$?）

闭性：$\{x_k\}_{k=1}^{\infty}\to\bar{x}$，则 $x_k=\lambda_k\xi_k$

$\lambda_k\in\partial\varphi(\psi(x))$，$\xi_k\in\partial\psi(x)$，则由 $\{\lambda_k\}$，$\{\xi_k\}$ 的有界性

$\exists$ 子列 $\lambda_{k_n}\to\lambda^*$，$\xi_{k_n}\to\xi^*$，s.t. $\bar{x}=\lambda^*\xi^*$，由 $\partial\varphi(\psi(x))$ 和 $\partial\psi(x)$

的闭性知，$\mathcal{Q}$ 是闭的

② 证 $\mathrm{Conv}(\mathcal{Q})$ 的闭性要用 Carathéodory's 定理

对 $\forall P\subset\mathbb{R}^n$，$g\in\mathrm{Conv}(P)$，则 $\exists\,k\leq n+1$ 个点 $p_1,\cdots,p_k\in P$，

s.t. $g\in\mathrm{Conv}(p_1,\cdots,p_k)$

用上面定理，可以将 $\mathrm{Conv}\,\mathcal{Q}=f(\mathcal{Q}^{n+1}\times\Sigma)$，其中

$$f:\underbrace{\mathbb{R}^n\times\cdots\times\mathbb{R}^n}_{n+1\text{个}}\times\Sigma\to\mathbb{R}^n \text{ 定义为}$$

$$f((x_1,\cdots,x_{n+1},\mu))=\sum_{k=1}^{n+1}\mu_k x_k$$

$\Sigma=\{\mu\in\mathbb{R}^{n+1}\mid \mu_i\geqslant 0,\sum_{i=1}^{n+1}\mu_i=1\}$ 是 $n+1$ 维单纯形,

由 f 是连续的，且 $\mathcal{Q}^{n+1}$，Σ 是紧的，故

$$\mathrm{Conv}\,\mathcal{Q}=f(\mathcal{Q}^{n+1}\times\Sigma)\text{ 是紧的}$$

故推论3.1.5：$\partial f(x)=\mathrm{Conv}\,\mathcal{Q}=\hat{\mathcal{Q}}$ $\square$

<u>Lemma 3.1.9</u> f 是凸的，且

$$\bar{z}=(\bar{x},\bar{y})\in\mathrm{int}(\mathrm{dom}f)\subseteq\mathbb{R}^n\times\mathbb{R}^m$$

设 f 关于第一个分量可微，对应的偏梯度是 $\nabla_1 f(\cdot,\cdot)\in\mathbb{R}^n$

在 $\bar{z}$ 处关于任意方向连续，则

$$\partial f(\bar{z})=\left(\nabla_1 f(\bar{x},\bar{y}),\ \partial_2 f(\bar{x},\bar{y})\right)$$

证明：固定 $\forall$ 方向 $h=(h_x,h_y)\in\mathbb{R}^n\times\mathbb{R}^m$，则对 $\alpha>0$ 充分小，

$$\frac{1}{\alpha}\left(f(\bar{x}+\alpha h_x,\bar{y}+\alpha h_y)-f(\bar{x},\bar{y})\right)$$

$$=\frac{1}{\alpha}\left(f(\bar{x}+\alpha h_x,\bar{y}+\alpha h_y)-f(\bar{x},\bar{y}+\alpha h_y)\right)$$

$$+\frac{1}{\alpha}\left(f(\bar{x},\bar{y}+\alpha h_y)-f(\bar{x},\bar{y})\right)$$

由 f 的凸性：

$$\alpha\langle\nabla_1 f(\bar{x},\bar{y}+\alpha h_y),h_x\rangle\leqslant f(\bar{x}+\alpha h_x,\bar{y}+\alpha h_y)-f(\bar{x},\bar{y}+\alpha h_y)$$

$$\leqslant\alpha\langle\nabla_1 f(\bar{x}+\alpha h_x,\bar{y}+\alpha h_y),h_x\rangle$$

令 $\alpha\to 0$，由 $\nabla_1 f(\cdot,\cdot)$ 的连续性知：

$$\frac{1}{\alpha}\left(f(\bar{x}+\alpha h_x,\bar{y}+\alpha h_y)-f(\bar{x},\bar{y}+\alpha h_y)\right)$$

$$=\langle\nabla_1 f(\bar{x},\bar{y}),h_x\rangle$$

故 $\quad f'(\bar{z},h)=\langle\nabla_1 f(\bar{x},\bar{y}),h_x\rangle+f'(\bar{z},(0,h_y))$

令 $f_x(y)=f(x,y)$，则

$$f'(\bar{z},(0,h_y))=\lim_{\alpha\downarrow 0}\frac{1}{\alpha}\left[f(\bar{x},\bar{y}+\alpha h_y)-f(\bar{x},\bar{y})\right]$$

$$= \lim_{\alpha\downarrow 0} \frac{1}{\alpha}[f_{\bar{x}}(\bar{y}+\alpha h_y) - f_{\bar{x}}(\bar{y})]$$

$$= f'_{\bar{x}}(\bar{y}; h_y)$$

$$= \max_{g} \{\langle g, h_y\rangle \mid g \in \partial f_{\bar{x}}(\bar{y})\}$$

$$= \max_{g} \{\langle g, h_y\rangle \mid g \in \partial_2 f(\bar{x},\bar{y})\}$$

故 $f'(\bar{z}; h) = \max\{\langle (\nabla_1 f(\bar{x},\bar{y}), g), (h_x, h_y)\rangle \mid g \in \partial_2 f(\bar{x},\bar{y})\}$

$$= \max_{\tilde{g}} \{\langle \tilde{g}, h\rangle \mid \tilde{g} \in \{\nabla_1 f(\bar{x},\bar{y})\} \times \partial_2 f(\bar{x},\bar{y})\}$$

$$= \max_{\tilde{g}} \{\langle \tilde{g}, h\rangle \mid \tilde{g} \in \partial f(\bar{z})\}$$

故由推论 3.1.5 (ii)

$$\partial f(\bar{z}) = (\nabla_1 f(\bar{x},\bar{y}), \partial_2 f(\bar{x},\bar{y}))$$

$\boxed{/\!/}$

lemma 3.1.10 令 f 是凸的，$x_0 \in \operatorname{int}(\operatorname{dom} f)$，设 $\exists g(x) \in \partial f(x)$，$g(x)$ 在 x_0 处连续，则 f 在 x_0 处可微，且 $\nabla f(x_0) = g(x_0)$

证明：对 $\forall$ 方向 $h \in \mathbb{R}^n$，$\alpha > 0$ 充分小，有

$$\langle g(x_0), h\rangle \leq \frac{1}{\alpha}[f(x_0+\alpha h)-f(x_0)]$$
$$\leq \langle g(x_0+\alpha h), h\rangle$$

故取 $\alpha \downarrow 0$，有 $f'(x_0;h)=\langle g(x_0), h\rangle$ 对 $\forall h\in\mathbb{R}^n$

取 $h=e_1, e_2, \cdots, e_n$ 知，$f(x)$ 在 x_0 处有连续的偏导数，

故 $f(x)$ 在 x_0 处可微，由lemma 3.1.7

$$g(x_0)=\nabla f(x_0)$$

$\boxed{//}$

<u>lemma 3.1.11</u> 设 f 在有界集 $S\subseteq \mathrm{dom} f\subseteq\mathbb{R}^m$ 上是闭凸的，

考虑线性算子 $\mathcal{A}(x)=Ax+b: \mathbb{R}^n\to\mathbb{R}^m$

则 $\phi(x)=f(\mathcal{A}(x))$ 在 $\mathcal{Q}=\{x \mid \mathcal{A}(x)\in S\}$ 上是闭凸的，

且对 $\forall x\in\mathcal{Q}$，$\partial f(\mathcal{A}(x))$ 非空，则

$$\partial\phi(x)=A^T\partial f(\mathcal{A}(x))$$

证明：<u>Thm3.1.6</u> 证明了第一部分，令 $y_0=\mathcal{A}(x_0)$，则

对 $\forall p\in\mathbb{R}^n$，有

$$\phi'(x_0, p) = f'(y_0; Ap)$$
$$= \max\{\langle g, Ap\rangle \mid g \in \partial f(y_0)\}$$
$$= \max\{\langle \bar{g}, p\rangle \mid \bar{g} \in A^T \partial f(y_0)\}$$

故 $\partial \phi(x_0) = A^T \partial f(A(x_0))$ $\boxed{\diagup}$

<u>Thm3.1.12</u> 令 f_1, f_2 是闭凸的，$\alpha_1, \alpha_2 \geq 0$，则

$f(x) = \alpha_1 f_1(x) + \alpha_2 f_2(x)$ 也是闭凸的，且

$$\partial f(x) = \alpha_1 \partial f_1(x) + \alpha_2 \partial f_2(x)$$

对 $\forall x \in \text{int}(\text{dom} f) = \text{int}(\text{dom} f_1) \cap \text{int}(\text{dom} f_2)$

证明：$f(x)$ 的闭凸性在 <u>Thm3.1.5</u>

考虑 $x_0 \in \text{int}(\text{dom} f)$，由 <u>Thm3.1.15</u> x_0 处次微分有界

对 $\forall p \in \mathbb{R}^n$，有：

$$f'(x_0; p) = \alpha_1 f_1'(x_0; p) + \alpha_2 f_2'(x_0; p)$$
$$= \max\{\langle g_1, \alpha_1 p\rangle \mid g_1 \in \partial f_1(x_0)\}$$
$$+ \max\{\langle g_2, \alpha_2 p\rangle \mid g_2 \in \partial f_2(x_0)\}$$

$$= \max\left\{\langle \alpha_1 g_1 + \alpha_2 g_2, p\rangle \mid g_1 \in \partial f_1(x_0), g_2 \in \partial f_2(x_0)\right\}$$

$$= \max\left\{\langle g, p\rangle \mid g \in \alpha_1 \partial f_1(x_0) + \alpha_2 \partial f_2(x_0)\right\}$$

故 $$\partial f(x) = \alpha_1 \partial f_1(x) + \alpha_2 \partial f_2(x)$$ $\boxed{\checkmark}$

lemma 3.1.13 令 $f_i, i=1,\cdots,m$ 是闭凸的，则 $f(x) = \max\limits_{1\le i\le m} f_i(x)$ 是闭凸的，且对 $\forall x \in \text{int}(\text{dom} f) = \bigcap\limits_{i=1}^{m} \text{int}(\text{dom} f_i)$，

$$\partial f(x) = \text{Conv}\left\{\partial f_i(x) \mid i \in I(x)\right\}$$

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

证明：f 的闭凸性在 Thm 3.1.15 已经证明

设 $I(x) = \{1,\cdots,k\}$，下证 $f'(x;p) = \max\limits_{1\le i\le k} f_i'(x;p)$

对 $i \notin I(x)$，则由 $x \in \text{int}(\text{dom} f)$，$f_i(x)$ 在 x 处连续，

$\forall i=1,\cdots,m$，故 $i \notin I(x)$ 的 f_i 对 $f'(x;p)$ 没有影响

$$\lim_{\alpha\downarrow 0}\frac{1}{\alpha}[f(x+\alpha p)-f(x)] \geqslant \liminf_{\alpha\downarrow 0}\frac{f_i(x+\alpha p)-f(x)}{\alpha}$$

$$=\liminf_{\alpha\downarrow 0}\frac{f_i(x+\alpha p)-f_i(x)}{\alpha}=f_i'(x;p),\ \forall i\in I(x)$$

故 $f'(x;p)\geqslant \max\limits_{1\leq i\leq k} f_i'(x;p)$

下证 $f'(x;p)\leq \max\limits_{1\leq i\leq k} f_i'(x;p)$，用反证法，设

$$\lim_{\alpha\downarrow 0}\frac{f(x+\alpha p)-f(x)}{\alpha}>\max_{1\leq i\leq k} f_i'(x;p)$$

则 $\exists\, \varepsilon>0$，$\exists \{\alpha_n\}\downarrow 0$，s.t.

$$\frac{f(x+\alpha_n p)-f(x)}{\alpha_n}\geqslant \max_{1\leq i\leq k} f_i'(x;p)+\varepsilon$$

设 $\dot{\imath}_n\in \operatorname*{Argmax}\limits_{1\leq i\leq k} f_i(x+\alpha_n p)$

由 $\dot{\imath}_n\in\{1,\cdots,k\}$ 是有限集，故 $\exists\{\alpha_n\}$ 的子列 $\{\alpha_{v_n}\}$，s.t

对 $\forall n\in\mathbb{N}$，$\dot{\imath}_{v_n}=\dot{\imath}^*$

$$\frac{f_{\partial^*}(x+\alpha_{v_n}p)-f_{\partial^*}(x)}{\alpha_{v_n}} \geqslant \max_{1\leq i\leq k} f_i'(x;p)+\varepsilon$$

取 $n\to\infty$，有 $f_{\partial^*}'(\bar{x};p) \geqslant \max\limits_{1\leq i\leq k} f_i'(x;p)+\varepsilon$，矛盾！

故 $f'(x;p)=\max\limits_{1\leq i\leq k} f_i'(x;p)$

$$f'(x;p)=\max_{1\leq i\leq k} f_i'(x;p)$$

$$=\max_{1\leq i\leq k}\ \max\{\langle g_i,p\rangle \mid g_i\in\partial f_i(x)\}$$

注意到

$$\max_{1\leq i\leq k} a_i=\max\left\{\sum_{i=1}^k \lambda_i a_i \,\middle|\, \{\lambda_i\}\in\Delta_k\right\}$$

$\Delta_k=\{\lambda_i\geqslant 0,\ \sum\limits_{i=1}^k\lambda_i=1\}$ 是 k 阶单纯形

则 $f'(x;p)=\max\limits_{\{\lambda_i\}\in\Delta_k}\left\{\sum\limits_{i=1}^k\lambda_i\max\{\langle g_i,p\rangle \mid g_i\in\partial f_i(x)\}\right\}$

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

$$= \max\left\{\langle g, p\rangle \,\middle|\, g=\sum_{i=1}^{k}\lambda_i g_i,\ g_i\in\partial f_i(x),\ \{\lambda_i\}\in\Delta_k\right\}$$

$$= \max\left\{\langle g, p\rangle \,\middle|\, g\in \mathrm{Conv}\{\partial f_i(x), i\in I(x)\}\right\}$$

$\boxed{/\!/}$

<u>lemma 3.1.14</u> 令 Δ 是任意集合，$f(x)=\sup\{\phi(x,y) \mid y\in\Delta\}$.

设对 $\forall y\in\Delta$，$\phi(\cdot,y)$ 是关于第一个变量在 Q 上闭凸的，

则 f 是 $\hat{Q}$ 上的闭凸 func，其中

$$\hat{Q}=\left\{x\in Q \,\middle|\, \sup_{y\in\Delta}\phi(x,y)<+\infty\right\}$$

且对 $\forall x\in\hat{Q}$，有

$$\partial_{\hat{Q}} f(x) \supseteq \mathrm{Conv}\left\{\partial_{Q,x}\phi(x,y) \mid y\in I(x)\right\}$$

$$I(x)=\{y\in\Delta \mid \phi(x,y)=f(x)\}$$

证明：f 的闭凸性在 <u>Thm 3.1.8</u> 上证明过了，

对 $\forall x\in\hat{\otimes}$，$y_0\in I(x_0)$，$g_0\in\partial_{\otimes,x}\phi(x_0,y_0)$，有

$$f(x)\geq\phi(x,y_0)\geq\phi(x_0,y_0)+\langle g_0,x-x_0\rangle$$
$$=f(x_0)+\langle g_0,x-x_0\rangle$$

跟据 Carathéodory's 定理即证

$\boxed{\quad}$

Example 3.1.5

① $f(x)=(x)_+=\max\{x,0\}=\max\limits_{g\in[0,1]} gx$

故 $f(y)=\max\{y,0\}\geq gy=f(0)+\langle g,y\rangle$, $g\in[0,1]$

故 $[0,1]\subset\partial f(0)$，且当 $g<0$ 时，取 $y<0$

则 $f(y)<f(0)+\langle g,y\rangle$

当 $g>1$ 时，取 $y>0$，则 $f(y)<f(0)+\langle g,y\rangle$，故

$$[0,1]=\partial f(0)$$

$\boxtimes$

② 对 $\forall i=1,\cdots,m$，$|\langle a_i,x\rangle|$ 是闭凸 func

case1：当 $\langle a_i,x\rangle>0$ 时，在 x 处可微

$$\partial|\langle a_i,x\rangle|=\nabla_x\langle a_i,x\rangle=a_i$$

case2：当 $\langle a_i,x\rangle<0$ 时，在 x 处可微

$$\partial|\langle a_i,x\rangle|=-\nabla_x\langle a_i,x\rangle=-a_i$$

case3: 当 $\langle a_i, x\rangle = 0$ 时, $g \in \partial|\langle a_i, x\rangle| \Leftrightarrow$

对 $\forall y \in \mathbb{R}^n$,

$$|\langle a_i, y\rangle| \geq \langle g, y-x\rangle \quad (*)$$

由 $\langle a_i, x\rangle = 0$, 可以扩展成 $\mathbb{R}^n$ 一组基 $\{a_i, x, e_3, \cdots, e_n\}$

$$\forall y \in \mathbb{R}^n,\ y = k_1 a_i + k_2 x + k_3 e_3 + \cdots + k_n e_n$$

$$(*) \Leftrightarrow |k_1| \|a_i\|_2^2 \geq \langle g, k_1 a_i + (\underbrace{k_2 - 1}_{\tilde{k}_2}) x + k_3 e_3 + \cdots + k_n e_n\rangle \quad ①$$

对 $\forall k_1, \tilde{k}_2, k_3, \cdots, k_n$ 恒成立, 故 $g = k a_i$, 若不然

不失一般地设 $g = k a_i + p e_3$, 则 ① 变为:

$$|k_1| \|a_i\|_2^2 \geq k_1 k \|a_i\|_2^2 + \underbrace{p k_3 \|e_3\|_2^2}_{\text{可以任意大}} \quad \text{对 } \forall k_1, k_3 \text{ 成立}$$

显然矛盾. 故 $g = k a_i$

$$(*) \Leftrightarrow |k_1| \|a_i\|_2^2 \geq k_1 k \|a_i\|_2^2 \quad \text{对 } \forall k_1 \text{ 成立}$$

故 $k \in [-1, 1]$, $g \in [-a_i, a_i]$

故 $\partial|\langle a_i, x\rangle| = [-a_i, a_i]$

由 lemma 3.1.12

$$\partial f(x) = \sum_{i \in I_+(x)} a_i - \sum_{i \in I_-(x)} a_i + \sum_{i \in I_0(x)} [-a_i, a_i]$$

$\boxed{/}$

③ $f(x) = \max_{1 \leq i \leq n} x^{(i)}$，定义 $I(x) = \{ i : x^{(i)} = f(x) \}$

由 lemma 3.1.13

$$\partial f(x) = \operatorname{Conv} \{ \partial f_i(x) \mid i \in I(x) \}$$

其中 $f_i(x) = x^{(i)}$ 是可微的，$f_i(x) = \langle e_i, x \rangle$

$$\partial f_i(x) = \nabla f_i(x) = e_i$$

故 $\partial f(x) = \operatorname{Conv} \{ e_i \mid i \in I(x) \}$

④ $f(x) = \|x\|$，$\forall y \in \mathbb{R}^n$，当 $x=0$ 时

$$\|y\| \geq \langle g, y \rangle$$

① 令 $y = g$ 时 $\|g\| \geq \|g\|^2 \Rightarrow \|g\| \leq 1$（必要性）

② $\|g\| \leq 1$ 时 $\langle g, y\rangle \leq \|g\|\|y\| \leq \|y\|$ (充分性)

故 $\partial f(0) = B_2(0,1) = \{x \in \mathbb{R}^n \mid \|x\| < 1\}$

当 $x \neq 0$ 时，是可微的 $\partial f(x) = \left\{\frac{x}{\|x\|}\right\}$

⑤ 由②

$x = 0$ 时

$$\partial f(0) = \sum_{i=1}^{n} [-e_i, e_i] = B_\infty(0,1)$$

$x \neq 0$ 时

$$\partial f(0) = \sum_{i \in I_+(x)} e_i - \sum_{i \in I_-(x)} e_i + \sum_{i \in I_0(x)} [-e_i, e_i]$$

⑥ Minkowski function

$$\psi_Q(x) = \min_{\tau \geq 0} \{\tau : x \in \tau Q\}$$

Q 是有界闭凸集，$0 \in \operatorname{int} Q$

polar 集 $\mathcal{P}_Q = \{g \in \mathbb{R}^n : \langle g, x\rangle \leq 1, \forall x \in Q\}$

当 $x=0$ 时，$g\in\partial\psi_{\otimes}(0) \Leftrightarrow \forall y\in\mathbb{R}^n$

$$\psi_{\otimes}(y)\geqslant\langle g,y\rangle$$

$$\Leftrightarrow \quad \left\langle g,\frac{y}{\psi_{\otimes}(y)}\right\rangle\leqslant 1 \quad \forall y\in\mathbb{R}^n$$

$$\Leftrightarrow \quad \langle g,z\rangle\leqslant 1 \quad \forall z\in\partial\otimes$$

$$\Leftrightarrow \quad \langle g,x\rangle\leqslant 1 \quad \forall x\in\otimes$$

令 $A=\{g\in\mathbb{R}^n:\langle g,z\rangle\leqslant 1,\forall z\in\partial\otimes\}$

$B=\{g\in\mathbb{R}^n:\langle g,x\rangle\leqslant 1,\forall x\in\otimes\}$

$A\supseteq B$ 是显然的，下证 $B\subseteq A$，若不然，设 $\exists g_0\in A$ 但 $g_0\notin B$

故 $\exists x_0\in\operatorname{int}\otimes$，s.t. $\langle g_0,x_0\rangle>1$

但 $\left\langle g_0,\frac{x_0}{\psi_{\otimes}(x_0)}\right\rangle\leqslant 1 \Rightarrow \psi_{\otimes}(x_0)>1$

与 $x_0\in\operatorname{int}\otimes$ 矛盾

故 $\partial\psi_{\otimes}(0)=P_{\otimes}$

下证 $\partial\psi_{\otimes}(x) = \underset{g\in\mathcal{P}_{\otimes}}{\mathrm{Argmax}}\langle g,x\rangle$

① 先证 $\psi_{\otimes}(x)$ 的次可列可加性 $\psi_{\otimes}(x+y)\leq\psi_{\otimes}(x)+\psi_{\otimes}(y)$

设 $\psi_{\otimes}(x)=a_1$, $\psi_{\otimes}(y)=a_2$ $\Rightarrow$ $a_1^{-1}x\in\otimes$, $a_2^{-1}y\in\otimes$

$$\frac{a_1}{a_1+a_2}\underbrace{a_1^{-1}x} + \frac{a_2}{a_1+a_2}\underbrace{a_2^{-1}y}\in\otimes$$

$x\in a_1\otimes\Rightarrow a_1^{-1}x\in\otimes$

$$\Rightarrow \frac{x+y}{a_1+a_2}\in\otimes \Rightarrow x+y\in(a_1+a_2)\otimes$$

$$\psi_{\otimes}(x+y)\leq a_1+a_2=\psi_{\otimes}(x)+\psi_{\otimes}(y)$$

② 证明 g 是次梯度 $\Rightarrow$ $g\in\mathcal{P}_{\otimes}$

$$\psi_{\otimes}(x)+\langle g,y-x\rangle\leq\psi_{\otimes}(y)=\psi_{\otimes}(y-x+x)$$
$$\leq\psi_{\otimes}(x)+\psi_{\otimes}(y-x)$$

$$\Rightarrow \langle g, y-x\rangle \leq \psi_{\otimes}(y-x) \quad \forall y$$

$$\Rightarrow \langle g, y\rangle \leq \psi_{\otimes}(y) \quad \forall y$$

$$\Rightarrow \langle g, \frac{y}{\psi_{\otimes}(y)}\rangle \leq 1 \Rightarrow \langle g, z\rangle \leq 1 \quad \forall z \in \partial Q$$

$$\Rightarrow g \in P_{\otimes} = \{ g \in \mathbb{R}^n : \langle g, x\rangle \leq 1, \forall x \in \otimes \}$$

③ 引理 (Schneider 1993)

记 $\xi_{\otimes}(x) = \sup\limits_{\eta \in \otimes} \langle \eta, x\rangle$ 是 $\otimes$ 的 support function (P151)

则 $\forall y \in \mathbb{R}^n$, $\boxed{\psi_{\otimes}(y) = \xi_{P_{\otimes}}(y)}$

④ 由引理,

$$\xi_{P_{\otimes}}(y) \geq \xi_{P_{\otimes}}(x) + \langle g, y-x\rangle \quad \forall y$$

$$\Leftrightarrow \max_{\xi \in P_{\otimes}} \langle y, \xi\rangle - \langle g, y\rangle \geq \max_{\xi \in P_{\otimes}} \langle x, \xi\rangle - \langle g, x\rangle \quad (*)$$

case1: $g = \operatorname{Argmax}\limits_{g \in P_{\otimes}} \langle g, x\rangle$, 则右边 $=0$

$\circledast \Leftrightarrow \max\limits_{\xi\in P_{\otimes}} \langle y,\xi\rangle \geq \langle g,y\rangle$ 由 $g\in P_{\otimes}$，显然成立

故 $\operatorname*{Argmax}\limits_{g\in P_{\otimes}} \langle g,x\rangle \in \partial\psi_{\otimes}(x)$

<u>case2</u> $g\neq \operatorname*{Argmax}\limits_{g\in P_{\otimes}} \langle g,x\rangle$，且 $g\in \partial P_{\otimes}$

则右边严格 >0，左边：

$$\max_{\xi\in P_{\otimes}} \langle \xi,y\rangle - \langle g,y\rangle$$

故 $\exists y\in\mathbb{R}^n$, s.t. $\max\limits_{\xi\in P_{\otimes}} \langle \xi,y\rangle = \langle g,y\rangle$，即左边 $=0$ 矛盾

故 $\partial P_{\otimes} \setminus \{g^*\} \notin \partial\psi_{\otimes}(x)$

$$g^* = \operatorname*{Argmax}_{g\in P_{\otimes}} \langle g,x\rangle$$

<u>case3</u>：$g\notin \partial P_{\otimes}$，设 $\tilde{g} = kg \in \partial P_{\otimes}$，$k>1$

$$\circledast \Leftrightarrow \max_{\xi\in P_{\otimes}} \langle y,\xi\rangle - \langle \tilde{g}/k, y\rangle \geq \max_{\xi\in P_{\otimes}} \langle x,\xi\rangle - \langle \tilde{g}/k, x\rangle$$

取 $y=\frac{1}{k}x$，则

$$\max_{\xi\in P_\oplus}\langle x/k,\xi\rangle-\langle x/k,\hat{g}/k\rangle\geq\max_{\xi\in P_\oplus}\langle x,\xi\rangle-\langle x,\hat{g}/k\rangle$$

令 $A=\max\limits_{\xi\in P_\oplus}\langle x,\xi\rangle-\langle x,\hat{g}/k\rangle$，则上式 $\Leftrightarrow$

$$A/k\geq A,\ A>0$$

$\Leftrightarrow$ $k\leq 1$ 与 $k>1$ 矛盾，故 $\forall g\in P_\oplus^\circ$，$g\notin\partial\psi_\oplus(x)$

综上：$\partial\psi_\oplus(x)=\operatorname*{Argmax}\limits_{g\in P_\oplus}\langle g,x\rangle$

$\square$

Def 3.1.7 f 称为齐次的 $(p>0)$，若 $\operatorname{dom} f$ 是锥且 $f(\tau x)=\tau^p f(x)$ $\forall x\in \operatorname{dom} f$，$\forall \tau\geq 0$

Thm 3.1.21 令 f 是凸的、次可微的，若齐次度 $p\geq 1$，则 $\langle g,x\rangle = p f(x)$，$\forall x\in \operatorname{dom} f$，$\forall g\in\partial f(x)$

证明：令 $x\in \operatorname{dom} f$，$g\in\partial f(x)$，则对 $\forall \tau\geq 0$，有

$$\tau^p f(x) = f(\tau x) \geq f(x) + (\tau-1)\langle g,x\rangle$$

当 $\tau>1$ 时，$\dfrac{\tau^p-1}{\tau-1} f(x) \geq \langle g,x\rangle$

取 $\tau\downarrow 1$，则 $p f(x)\geq \langle g,x\rangle$

当 $\tau<1$ 时，$\dfrac{1-\tau^p}{1-\tau} f(x)\leq \langle g,x\rangle$，故取 $\tau\uparrow 1$，有

$$p f(x)\leq \langle g,x\rangle$$

故 $p f(x) = \langle g,x\rangle$ $\boxed{\diagup}$

对 degree=1 的齐次 func:

$$\langle g, x\rangle = f(x), \forall x \in \mathrm{dom} f, \forall g \in \partial f(x)$$

则设 $\mathrm{dom} f = \mathbb{R}^n$, 则对 $\forall x \in \mathbb{R}^n$, 有

$$f(x) = f'(0, x) = \max_g \{\langle g, x\rangle : g \in \partial f(0)\}$$

对 $f(x) = \|x\|$, 有

$$\|x\| = \max_g \{\langle g, x\rangle : \|g\|_* \leq 1\}$$

故 $\partial \|x\| \Big|_{x=0} = \{g \in \mathbb{R}^n : \|g\|_* \leq 1\}$

<u>Lemma 3.1.5</u> 设 f 是凸的, 齐次的 (degree=1),

$\mathrm{dom} f = \mathbb{R}^n$, 则对 $\forall x \in \mathbb{R}^n$, 有

$$\partial f(x) = \{g \in \partial f(0) : \langle g, x\rangle = f(x)\}$$

证明:

令 $G(x) = \{g \in \partial f(0) : \langle g, x\rangle = f(x)\}$

若 $g\in\partial f(x)$，则对 $\forall y\in\mathbb{R}^n$，有：

$$f(y)\geqslant f(x)+\langle g,y-x\rangle \overset{(3.1.40)}{=\!=\!=} \langle g,y\rangle$$

故 $g\in\partial f(0)$，由假设 $g\in\partial f(x)\Rightarrow\langle g,x\rangle=f(x)$，故 $g\in G(x)$

若 $g\in G(x)$，则对 $\forall y\in\mathbb{R}^n$，有

$$f(y)\geqslant\langle g,y\rangle=f(x)+\langle g,y-x\rangle$$

故 $g\in\partial f(x)$ $\boxed{\diagup}$

注：用该引理可以立刻得到例3.1.5 ⑥

<u>Thm 3.1.22</u> 令 Q_1,Q_2 是有界闭凸集，$Q=Q_1\cap Q_2$ 有非空内部，则

$$\xi_Q(x)=\min_{y\in\mathbb{R}^n}\{\xi_{Q_1}(x+y)+\xi_{Q_2}(-y)\},\ x\in\mathbb{R}^n$$

证明：首先证明优化问题是良定义的，若 $g\in Q_1\cap Q_2$，则对 $\forall y\in\mathbb{R}^n$，有

$$\phi_x(y)\doteq\xi_{Q_1}(x+y)+\xi_{Q_2}(-y)\geqslant\langle g,x+y\rangle+\langle g,-y\rangle=\langle g,x\rangle$$

故 $\phi_x(y)$ 是下方有界的，设 $\phi_x^* = \inf\limits_{y\in\mathbb{R}^n} \phi_x(y)$，则 $\phi_*^* \geqslant \xi_{\otimes}(x)$

考虑序列 $\{y_k\}$, s.t. $\phi_x(y_k) \to \phi_x^*$

注：$\phi_x(y) = \xi_{\otimes}(x+y) + \xi_{\otimes}(-y)$ 是连续的，由 $\xi_{\otimes}(y)$ 是闭的 func，故 $\xi_{\otimes}(y)$ 是连续的 (Thm 3.1.11)，故 $\xi_{\otimes}(x+y)$，$\xi_{\otimes}(-y)$ 都是连续的

<u>case 1</u>：当 $\{y_k\}$ 有界，则显然 ϕ_x^* 是可达的

<u>case 2</u>：当 $\{y_k\}$ 无界，有 $t_k = \|y_k\| \to \infty$，令 $\bar{y}_k = \frac{1}{t_k} y_k$

则 $\lim\limits_{k\to\infty} [\xi_{\otimes_1}(\frac{1}{t_k}x + \bar{y}_k) + \xi_{\otimes_2}(-\bar{y}_k)]$

$= \lim\limits_{k\to\infty} [\frac{1}{t_k}\xi_{\otimes_1}(x+y_k) + \frac{1}{t_k}\xi_{\otimes_2}(-y_k)]$ (齐次性)

$= \lim\limits_{k\to\infty} \frac{1}{t_k}\phi_x(y_k) = 0$

由 $\{\bar{y}_k\}$ 有界，不妨设 $\bar{y}_k \to \bar{y}$，$\|\bar{y}\| = 1$

由 $\lim\limits_{k\to\infty} [\xi_{\otimes_1}(\frac{1}{t_k}x + \bar{y}_k) + \xi_{\otimes_2}(-\bar{y}_k)]$

$= \xi_{\Theta_1}(\bar{y}) + \xi_{\Theta_2}(-\bar{y}) = 0$ (连续性)

故 $\langle g_1, \bar{y}\rangle \leqslant \xi_{\Theta_1}(\bar{y}) = -\xi_{\Theta_2}(-\bar{y}) \leqslant \langle g_2, \bar{y}\rangle$

$\forall g_1 \in \Theta_1, \forall g_2 \in \Theta_2$

故 $\langle g, \bar{y}\rangle = \text{const}$，对 $\forall g \in \Theta_1 \cap \Theta_2 = \Theta$

即 Θ 是 $H = \{g : \langle g, \bar{y}\rangle = \text{const}\}$ 的子集，与 Θ 有

非空内部矛盾. 故 case 2 不成立，优化问题 well-defined

记 y^* 是 (3.1.44) 的最优值点，由 Thm 3.1.20

$$0 \in \partial\phi_x(y^*) = \partial\xi_{\Theta_1}(x+y^*) + \partial\xi_{-\Theta_2}(y^*)$$

求 $\partial\xi_\Theta(0)$：对 $\forall y \in \mathbb{R}^n$, $\xi_\Theta(y) \geqslant \langle g, y\rangle$

则 $\partial\xi_\Theta(0) = \Theta$

故存在 g, s.t.

$$g \in \Theta_1, \langle g, x+y^*\rangle = \xi_{\Theta_1}(x+y^*)$$

$$-g \in -\Theta_2, \langle -g, y^*\rangle = \xi_{-\Theta_2}(y^*)$$

故 $\phi_x^* = \xi_{\otimes_1}(x+y^*) + \xi_{\otimes_2}(-y^*)$

$$= \xi_{\otimes_1}(x+y^*) + \xi_{-\otimes_2}(y^*)$$

$$= \langle g, x\rangle$$

由 $g\in\otimes$，故 $\phi_x^* \le \xi_{\otimes}(x)$，综上 $\xi_{\otimes}(x) = \phi_x^*$ $\square$

<u>Lemma 3.1.16</u> 考虑 $\psi(g) = \max\limits_{\lambda\in\Lambda}\langle\lambda, g\rangle$，其中 $\Lambda\subset\mathbb{R}_+^m$

是有界闭凸集。$F(x) = (f_1(x), \cdots, f_m(x))$，$x\in\mathbb{R}^n$ 有可微凸分支.

则 $f(x) = \psi(F(x))$ 是凸的，且

$$\partial f(x) = \left\{\sum_{i=1}^{m}\lambda^{(i)}\nabla f_i(x) : \lambda\in \operatorname*{Argmax}_{\lambda\in\Lambda}\langle\lambda, F(x)\rangle\right\}$$

证明：$\psi(\cdot)$ 是单调的：设 $g_1\le g_2$（按分量 $\le$），则

$$\psi(g_1) = \max_{\lambda\in\Lambda}\langle\lambda, g_1\rangle = \langle\lambda^*, g_1\rangle \le \langle\lambda^*, g_2\rangle \le \max_{\lambda\in\Lambda}\langle\lambda, g_2\rangle = \psi(g_2)$$

故对 $\forall x, y\in\mathbb{R}^n$，$\alpha\in[0,1]$，有：

$$f(\alpha x+(1-\alpha)y)\leq\psi(\alpha F(x)+(1-\alpha)F(y))$$
$$\leq\alpha f(x)+(1-\alpha)f(y)$$

定义 $F'(x)=(\nabla f_1(x),\cdots,\nabla f_m(x))\in\mathbb{R}^{n\times m}$，则对 $\forall h\in\mathbb{R}^n$

$$f'(x;h)=\psi'(F(x);(F'(x))^T h)$$ 与§3.1.6节方法一致

$$=\max\{\langle\lambda,(F'(x))^T h\rangle:\lambda\in\partial\psi(F(x))\}$$

$$=\max\{\langle\lambda,(F'(x))^T h\rangle:\lambda\in\partial\psi(0),\langle\lambda,F(x)\rangle=\psi(F(x))\}$$

$\forall y\in\mathbb{R}^n,\psi(y)\geq\psi(0)+\langle g,y\rangle=\langle g,y\rangle$

故易证 $\partial\psi(0)=\Lambda$

$$=\max\{\langle\lambda,(F'(x))^T h\rangle:\lambda\in\operatorname*{Argmax}_{\lambda\in\Lambda}\langle\lambda,F(x)\rangle\}$$

$$=\max\{\langle F'(x)\lambda,h\rangle:\lambda\in\operatorname*{Argmax}_{\lambda\in\Lambda}\langle\lambda,F(x)\rangle\}$$

$$=\max\{\langle g,h\rangle:g=F'(x)\lambda,\lambda\in\operatorname*{Argmax}_{\lambda\in\Lambda}\langle\lambda,F(x)\rangle\}$$

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

故由推论3.1.5 (ii)

$$\partial f(x) = \left\{ \sum_{i=1}^{m} \lambda^{(i)} \nabla f_i(x) : \lambda \in \operatorname*{Argmax}_{\lambda \in \Lambda} \langle \lambda, F(x) \rangle \right\}$$

$\square$

<u>lemma 3.1.7</u> F 是 $\mathbb{R}^m$ 上可微凸的单调 func. f_i 在开的凸集 Q 上凸，则

$$\phi(x) = F(f_1(x), \cdots, f_m(x))$$

在 Q 上凸，且

$$\partial \phi(x) = \sum_{i=1}^{m} \nabla_i F(f(x)) \cdot \partial f_i(x) \quad , \quad x \in Q$$

其中 $f(x) = (f_1(x), \cdots, f_m(x))^T \in \mathbb{R}^m$

证明：对 $\forall x, y \in Q$, $\alpha \in [0,1]$，有

$$\begin{aligned} \phi(\alpha x + (1-\alpha) y) &\leqslant F(\alpha f(x) + (1-\alpha) f(y)) \\ &\leqslant \alpha \phi(x) + (1-\alpha) \phi(y) \end{aligned}$$

对 $\forall p \in \mathbb{R}^n$,

$$\phi'(x;p) = \sum_{i=1}^{m} \nabla_i F(f(x))\, f_i'(x;p)$$

$$= \sum_{i=1}^{m} \nabla_i F(f(x))\, \xi_{\partial_i f(x)}(p)$$

$$= \xi_{\sum_{i=1}^{m} \nabla_i F(f(x))\, \partial_i f(x)}(p)$$

故 $\partial \phi(x) = \sum_{i=1}^{m} \nabla_i F(f(x))\, \partial_i f(x)$ $\boxed{\mathscr{Z}}$

推论3.1.7 对 $\forall f_i$, $i=1,\cdots,m$, 是凸函数，则

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

是凸 func

证明：lemma3.1.17 的直接应用

§3.1.7 Optimality Conditions

$$\min_{x \in Q} \left\{ \tilde{f}(x) = f(x) + \psi(x) \right\} \qquad (3.1.48)$$

Q 是闭凸集，$f \in C^1(Q)$ 是连续可微凸 func，Ψ 是 Q 上的闭凸 func

<u>**Thm 3.1.23**</u> x^* 是 (3.1.48) 的解 $\Leftrightarrow$ $\forall x \in Q$

$$\langle \nabla f(x^*), x - x^* \rangle + \Psi(x) \geqslant \Psi(x^*) \qquad (3.1.49)$$

证明：

$(\Leftarrow)$

$$\begin{aligned} \tilde{f}(x) &= f(x) + \Psi(x) \\ &\geqslant f(x^*) + \langle \nabla f(x^*), x - x^* \rangle + \Psi(x) \\ &\geqslant f(x^*) + \Psi(x^*) \\ &= \tilde{f}(x^*) \end{aligned}$$

($\Rightarrow$) 设 $\exists x \in \Omega$, s.t.

$$\langle \nabla f(x^*), x - x^* \rangle + \Phi(x) < \Phi(x^*)$$

由 $\lim\limits_{\alpha \downarrow 0} \frac{1}{\alpha}[f(\alpha x + (1-\alpha)x^*) - f(x^*)] = \langle \nabla f(x^*), x - x^* \rangle$

故对充分小的 α, 有

$$\frac{1}{\alpha}[f(\alpha x + (1-\alpha)x^*) - f(x^*)] < \Phi(x^*) - \Phi(x)$$

$\Leftrightarrow$ $$f(\alpha x + (1-\alpha)x^*) < f(x^*) + \alpha[\Phi(x^*) - \Phi(x)]$$

$$= \tilde{f}(x^*) + \alpha[\Phi(x^*) - \Phi(x)] - \Phi(x^*)$$

$$\leq \tilde{f}(x^*) - \Phi(\alpha x + (1-\alpha)x^*)$$

故 $\tilde{f}(\alpha x + (1-\alpha)x^*) < \tilde{f}(x^*)$, 矛盾

$\boxed{/\!/}$

注: $\Phi \in C^1(\Omega)$ 时, (2.1.2) $\Rightarrow$ (3.1.49) 是显然的

(3.1.49) $\Rightarrow$ (2.1.2) 只需重复一遍 (2.1.2) 的证明

(3.1.49) $\iff$ $-\nabla f(x^*) \in \partial_{\Omega} \Phi(x^*)$

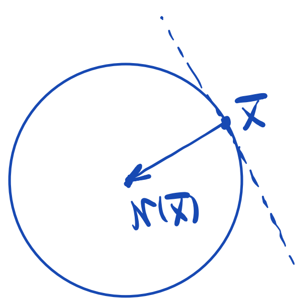
$$\min_{x\in\Omega} f(x)$$

$\Omega \subseteq \mathbb{R}^n$ 是闭凸集，f 是闭凸 func，$\mathrm{dom} f \supseteq \Omega$

对 $\bar{x}\in\Omega$，定义 normal cone：对 $\bar{x}\in\Omega$

$$\mathcal{N}(\bar{x}) = \left\{ g\in\mathbb{R}^n \,\middle|\, \langle g, x-\bar{x}\rangle \geqslant 0,\ \forall x\in\Omega \right\}$$

是闭凸集，由 $\mathcal{N}(\bar{x}) = \bigcap_{x\in\Omega} \underbrace{\{g : \langle g, x-\bar{x}\rangle \geqslant 0\}}_{\text{闭凸集}}$



$\mathcal{N}(\bar{x}) = \{0\} \iff \bar{x}\in \mathrm{int}\,\Omega$

$(\Leftarrow)$ $\bar{x}\in\mathrm{int}\,\Omega \iff \exists\delta$, s.t. $S(\bar{x},\delta)\subset\Omega$

故 $\mathcal{N}(\bar{x}) = 0$

$(\Rightarrow)$ 设 $\mathcal{N}(\bar{x}) = \{0\}$ 且 $\bar{x}\in\partial\Omega$，则

由 Thm3.1.14，$\exists$ Ω 的支撑 $H(g,\gamma)$，且 $\bar{x}\in H(g,\gamma)$

即 $\langle g, x\rangle \leqslant \gamma \leqslant \langle g, \bar{x}\rangle$，对 $\forall x\in\Omega$，且 $g\neq 0$

$\Leftrightarrow$ $\langle g, x-\bar{x}\rangle \leq 0$，对 $\forall x \in \Omega$，且 $g \neq 0$

$\Leftrightarrow$ $-g \in \mathcal{N}(\bar{x})$，且 $g \neq 0$ 与 $\mathcal{N}(\bar{x})$ 矛盾

tangent cone

$$\mathcal{T}(\bar{x}) = \{ p \in \mathbb{R}^n \mid \langle g, p\rangle \geq 0, \forall g \in \mathcal{N}(\bar{x}) \}$$

Lemma 3.1.18 令 $\bar{x} \in \partial\Omega$，则 $\Omega - \bar{x} \subset \mathcal{T}(\bar{x})$，且

$$\mathcal{T}(\bar{x}) = cl\,(\mathcal{K}(\Omega - \bar{x}))$$

证明：由定义：$\langle g, x-\bar{x}\rangle \geq 0 \quad \forall x \in \Omega$，$g \in \mathcal{N}(\bar{x})$

故 $\Omega - \bar{x} \subset \mathcal{T}(\bar{x})$，由 $\mathcal{T}(\bar{x})$ 的闭性，故

$$\bar{\mathcal{K}} = cl\,(\mathcal{K}(\Omega - \bar{x})) \subseteq \mathcal{T}(\bar{x})$$

设 $\exists \bar{p} \in \mathcal{T}(\bar{x})$，但 $\bar{p} \notin \bar{\mathcal{K}}$，则由推论 3.1.4，$\exists \bar{g}$ 强分离 $\bar{p}$ 和 $\bar{\mathcal{K}}$

i.e. $\langle \bar{g}, \bar{p}\rangle < \gamma \leq \langle \bar{g}, \alpha(x-\bar{x})\rangle$，$\forall x \in \Omega$，$\alpha \geq 0$

$\langle \bar{g}, x-\bar{x}\rangle \geq \frac{\gamma}{\alpha}$, $\alpha \to +\infty \Rightarrow \langle \bar{g}, x-\bar{x}\rangle \geq 0 \quad \forall x \in \Omega$

故 $\bar{g} \in N(\bar{x})$，取 $\alpha=0$，取 $\gamma \leq 0$，故 $\langle \bar{g}, \bar{p}\rangle < 0$，

故 $\bar{p} \notin T(\bar{x})$，矛盾 $\boxed{/}$

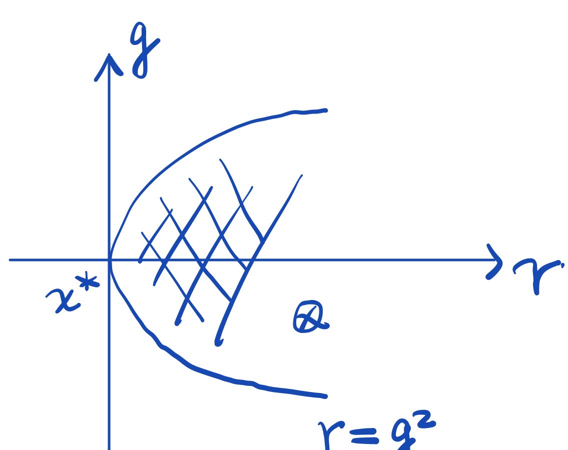
考虑 P152 页例 5：

$$\psi(g,r)=\begin{cases} 0 & , g=0, r=0 \\ \frac{|g|^2}{2r} & , r>0 \end{cases}$$

$\mathrm{dom}\,\psi = (\mathbb{R}\times\{r>0\}) \cup (0,0)$ 显然 $\mathrm{Argmin}\,\psi(g,r)=(0,0)$

$\Omega = \{(g,r): r \geq \|g\|^2\}$

故 $N(x^*)=\{kx: k\in\mathbb{R}^+, x=(1,0)\}$



$T(x^*)=\{r\geq 0\}\times\mathbb{R}$

故 $p=(0,1)^T \in T(x^*)$

但 $f(x^*+tp)$，$t\neq 0$ 没有定义，故 $f'(x^*;p)$ 显然不存在，故

Lemma 3.1.19 加条件 $x^* \in \mathrm{int}(\mathrm{dom} f)$

令 x^* 是 (3.1.50) 的 optimal solution，则

$$f'(x^*;p) \geq 0 \quad \forall p \in T(x^*)$$

证明：设 $\exists \bar{p} \in T(x^*)$, s.t. $f'(x^*;\bar{p}) < 0$

由 lemma 3.1.8，$\exists \{\alpha_k\} \subset \mathbb{R}_+$，$\{x_k\} \subset \Omega$, s.t.

$$\bar{p} = \lim_{k\to\infty} \alpha_k (x_k - x^*)$$

由于 $f'(x^*;\cdot)$ 是连续的（Lemma 3.1.5 在 $\mathbb{R}^n$ 上是凸的）

$$0 > f'(x^*;p) = \lim_{k\to\infty} \alpha_k f'(x^*; x_k - x^*)$$

$$= \lim_{k\to\infty} \lim_{\beta\downarrow 0} \frac{\alpha_k}{\beta} [f(x^*+\beta(x_k - x^*)) - f(x^*)] \geq 0$$

矛盾！

注：关于延拓的思路，还是P152页例15

设 $x^*=(0,0)$ 邻域 $B(x^*,\delta)$ 有延拓 $\tilde{\varphi}$，且 $\tilde{\varphi}$ 是凸的

s.t. $\tilde{\varphi}((0,\varepsilon_2))$ 对 $\varepsilon_2<\delta$ 充分小时有定义

设 $x_1=(0,\varepsilon_2)$，$x_2=(\varepsilon_1,\varepsilon_2)$，其中 $\varepsilon_1>0$ 任取

则 $\tilde{\varphi}(\frac{1}{2}x_1+\frac{1}{2}x_2)\leq\frac{1}{2}\tilde{\varphi}(x_1)+\frac{1}{2}\tilde{\varphi}(x_2)$

$\Leftrightarrow$ $\tilde{\varphi}((\frac{1}{2}\varepsilon_1,\varepsilon_2))\leq\frac{1}{2}\tilde{\varphi}((0,\varepsilon_2))+\frac{1}{2}\tilde{\varphi}((\varepsilon_1,\varepsilon_2))$

$$\Leftrightarrow\quad 2\frac{\varepsilon_2^2}{\varepsilon_1}\leq\frac{1}{2}\tilde{\varphi}((0,\varepsilon_2))+\frac{1}{2}\frac{\varepsilon_2^2}{\varepsilon_1}$$

$$\Leftrightarrow\quad 3\frac{\varepsilon_2^2}{\varepsilon_1}\leq\tilde{\varphi}((0,\varepsilon_2))$$

由 $\varepsilon_1>0$ 的任意性，$\tilde{\varphi}((0,\varepsilon_2))=+\infty$，故矛盾

$\boxed{/}$

Thm 3.1.24 $x^* \in \Omega$，则 $x^* \in X^* \Leftrightarrow$

$\exists g^* \in \partial f(x^*)$，s.t. $\langle g^*, x - x^* \rangle \geq 0 \quad \forall x \in \Omega$

证明：

$(\Leftarrow)$ $f(x) \geq f(x^*) + \langle g^*, x - x^* \rangle \geq f(x^*) \quad \forall x \in \Omega$

$(\Rightarrow)$ 设 $x^* \in X^*$ 是 (3.1.50) 的最优解，设不存在 $g \in \partial f(x^*)$，

s.t. $\langle g, x - x^* \rangle \geq 0$，$\forall x \in \Omega$

故 $\partial f(x^*) \cap \mathcal{N}(x^*) = \phi$，考虑优化问题：

$$\min_{g_1, g_2} \left\{ \phi(g_1, g_2) = \frac{1}{2} \|g_1 - g_2\|^2 : g_1 \in \partial f(x^*), g_2 \in \mathcal{N}(x^*) \right\}$$

由 $\boxed{x^* \in \operatorname{int}(\operatorname{dom} f)}$，故 $\partial f(x^*)$ 是有界的，且 $\phi(g_1, g_2)$

关于 g_2 是强制的，故最优解存在，记为 (g_1^*, g_2^*)

$$p^* = \phi(g_1^*, g_2^*)$$

由 Thm 2.2.9

$$\langle \nabla_{g_1}\phi(g_1^*, g_2^*), g_1 - g_1^* \rangle = \langle g_1^* - g_2^*, g_1 - g_1^* \rangle \geq 0 \quad \forall g_1 \in \partial f(x^*)$$

$$\langle \nabla_{g_2}\phi(g_1^*, g_2^*), g_2 - g_2^* \rangle = \langle g_2^* - g_1^*, g_2 - g_2^* \rangle \geq 0 \quad \forall g_2 \in \mathcal{N}(x^*)$$

显然，$0 \in \mathcal{N}(x^*)$，$\alpha g_2^* \in \mathcal{N}(x^*)$，$\forall \alpha \in \mathbb{R}_+$，故取

$g_2 = 0$ 与 $g_2 = \alpha g_2^*$，$\alpha \to +\infty$，有

$$\langle g_2^* - g_1^*, g_2^* \rangle \leq 0 \leq \langle g_2^* - g_1^*, g_2^* \rangle$$

故令 $p^* = g_2^* - g_1^*$，有 $\langle p^*, g_2^* \rangle = 0$，故由 (3.1.58)

$$\langle g_2, p^* \rangle \geq 0, \ \forall g_2 \in \mathcal{N}(x^*)$$

故 $p^* \in T(x^*)$；另一方面，对 $\forall g_1 \in \partial f(x^*)$，有

$$\langle g_1, p^* \rangle \overset{(3.1.57)}{\leq} \langle g_1^*, p^* \rangle = \langle g_1^* - g_2^*, p^* \rangle = -2p^*$$

则 $f'(x^*; p^*) = \max\{\langle g, p^* \rangle \mid g \in \partial f(x^*)\}$

$$= \langle g_1^*, p^* \rangle = -2p^* < 0$$

由 Lemma 3.1.19，推出矛盾，故 $\exists g^* \in \partial f(x^*)$，s.t.

$$\langle g^*, x - x^* \rangle \geq 0 \quad \forall x \in X$$

注意到对$\forall$其它点 $x_1^* \in X^*$，有

$$f(x^*) = f(x_1^*) \geq f(x^*) + \langle g^*, x_1^* - x^* \rangle \geq f(x^*)$$

故 $\langle g^*, x_1^* - x^* \rangle = 0 \Rightarrow g^* \in \partial f(x_1^*)$

故 $g^* \in \partial f(X_*)$，又由

$$\langle g^*, x - x_1^* \rangle = \langle g^*, (x - x^*) - (x_1^* - x^*) \rangle$$

$$= \langle g^*, x - x^* \rangle - \langle g^*, x_1^* - x^* \rangle$$

$$= \langle g^*, x - x^* \rangle \geq 0 \Rightarrow g^* \in N(x_1^*)$$

$\square$

<u>Thm 3.1.25</u> 令 ϕ 是闭凸 func，$Q_1 \subseteq \mathbb{R}^n$，$Q_2 \subseteq \mathbb{R}^m$

是两个闭凸集，s.t. $Q_1 \times Q_2 \subseteq \text{int dom}\,\phi$，定义

$$f(x) = \inf_{y \in Q_2} \phi(x,y)$$

则 f 在 Q_1 上凸，且若 $Y(x) = \operatorname*{Argmin}_{y \in Q_2} \phi(x,y) \neq \phi$，则

$$\partial_{Q_1} f(x) \supseteq \Big\{ g_x \in \mathbb{R}^n : \exists g_y ,\ \text{s.t. } (g_x, g_y) \in \bigcap_{y \in Y(x)} \partial\phi(x,y),$$

$$\text{且 } \langle g_y, y - y_x \rangle \geqslant 0 \ \ \forall y \in Q_2,\ \forall y_x \in Y(x) \Big\} \quad (3.1.60)$$

证明：f 的凸性是 <u>Thm 3.1.7</u> 的结论

先证 (3.1.60) 非空，为此先证明

$$\Big\{ g_x \in \mathbb{R}^n : \exists g_y ,\ \text{s.t. } (g_x, g_y) \in \partial\phi(x,y)，\text{且}$$

$$\langle g_y, y - y_x \rangle \geqslant 0 \ \ \forall y \in Q_2,\ \forall y_x \in Y(x) \Big\} \neq \phi$$

为此证明：$\partial_2\phi(x,y_x) \subseteq \mathrm{Proj}_{\mathcal{X}_2}\partial\phi(x,y_x)$ 对

$\forall x\in\mathcal{X}_1$ 成立：对 $\forall g\in\partial_2\phi(x,y_x)$

有 $\phi(x,y)\geq\phi(x,y_x)+\langle g, y-y_x\rangle \quad \forall y\in\mathcal{X}_2$

想证明：$\exists g_x\in\mathbb{R}^n$, s.t.

$$\phi(\tilde{x},y)\geq\phi(x,y_x)+\left\langle \begin{pmatrix} g_x \\ g\end{pmatrix}, \begin{pmatrix}\tilde{x}-x \\ y-y_x\end{pmatrix}\right\rangle$$

$$=\phi(x,y_x)+\langle g_x,\tilde{x}-x\rangle+\langle g,y-y_x\rangle \quad \begin{array}{l}\forall\tilde{x}\in\mathcal{X}_1\\ \forall y\in\mathcal{X}_2\end{array}$$

只需证明：

$$\phi(\tilde{x},y)-\phi(x,y)\geq\langle g_x,\tilde{x}-x\rangle \quad \forall\tilde{x}\in\mathcal{X}_1$$

故只需取 $g_x\in\partial_1\phi(x,y)$，$g_x$ 与 y 相关了！

故 $\partial_2\phi(x,y_x) \not\subseteq \mathrm{Proj}_{\mathcal{X}_2}\partial\phi(x,y_x)$

但 $\mathrm{Proj}_{\mathcal{X}_2}\partial\phi(x,y_x)\subseteq\partial_2\phi(x,y_x)$ 是显然的 非空性?

Corollary 3.1.8 若 $\gamma(x) \neq \phi$ 对 $\forall x \in \operatorname{dom} f$ 成立，

则 f 是闭凸 func 于 Q_1

证明：由 $\partial_{Q_1} f(x) \neq \phi$，故用 lemma 3.1.6 立刻证

$\boxed{\diagup}$

$$\min_{x \in Q} \left\{ f_0(x) \mid f_i(x) \leq 0, i=1, \cdots, m \right\}$$

Thm 3.1.6 (KKT) 设 f_i，$i=0, \ldots, m$ 是可微

凸 func，$\operatorname{int}(\operatorname{dom} f_i) \supset Q$，设存在 $\bar{x} \in Q$，s.t.

$f_i(\bar{x}) < 0, i=1, \cdots, m$ (Slater 条件)

则 x^* 是 (3.1.61) 的最优解 $\Leftrightarrow$ $\exists$ 非负 λ_i^*，$i=1, \cdots, m$，

s.t. $\left\langle \nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*), x - x^* \right\rangle \geq 0$，$\forall x \in Q$

$$\lambda_i^* f_i(x^*) = 0 \ , \ i=1,\cdots,m$$

证明：(3.1.61)的最优解 $\Leftrightarrow$

$$\operatorname*{Argmin}_{x\in\Theta} \phi(x) = \max\{f_0(x)-f^*;\ f_i(x),\ i=1,\cdots,m\}$$

由 <u>Thm3.1.24</u>，$\Leftrightarrow \exists\, g^* \in \partial\phi(x^*)$, s.t.

$$\langle g^*, x-x^* \rangle \geqslant 0 \quad \forall x \in \Theta$$

由 <u>lemma3.1.13</u>，$g^* \in \partial f(x^*) \Leftrightarrow$

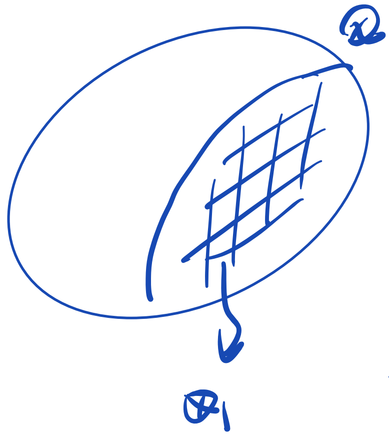
$\exists\, \bar\lambda_i$, $i=0,\cdots,m$. s.t.

$$\bar\lambda_0 \nabla f_0(x^*) + \sum_{i\in I^*} \bar\lambda_i \nabla f_i(x^*) = g^*$$

$$\bar\lambda_0 + \sum_{i\in I^*} \bar\lambda_i = 1$$

其中 $I^* = \{ i \in \{1,\cdots,m\} : f_i(x^*) = 0 \}$

注：$\min\limits_{x\in\Theta}\phi(x)=0$，设 $\Theta_1=\{x\in\Theta \mid f_i(x)\le 0,\ i=1,\cdots,m\}$



$\forall x\in\Theta\backslash\Theta_1$，$\exists \bar{i}$，s.t. $f_{\bar{i}}(x)>0$

故 $\phi(x)>0$

当 $x\in\Theta$ 时，$f_i(x)\le 0$，$\forall i=1,\cdots,m$

且 $f_0(x)-f^*\geqslant 0$（=可以取到）

故只需证 $\bar{\lambda}_0>0$，若不然，$\bar{\lambda}_0=0$，则

$$\sum_{i\in I^*}\bar{\lambda}_i f_i(\bar{x})\geqslant\sum_{i\in I^*}\bar{\lambda}_i[f_i(x^*)+\langle\nabla f_i(x^*),\bar{x}-x^*\rangle]$$

$$=\left\langle\sum_{i\in I^*}\bar{\lambda}_i\nabla f_i(x^*),\bar{x}-x^*\right\rangle$$

$$=\langle g^*,\bar{x}-x^*\rangle\geqslant 0$$

与 Slater 条件矛盾，故 $\bar{\lambda}_0>0$，取 $\lambda_i^*=\bar{\lambda}_i/\bar{\lambda}_0$ $\forall i\in I^*$

$\lambda_i^*=0$ $\forall i\notin I^*$，即证 ☑

lemma 3.1.20 令 $A \succ 0$

$$\max_{x} \{ \langle c, x\rangle : \langle Ax, x\rangle \le 1 \} = \langle c, A^{-1}c\rangle^{\frac{1}{2}}$$

证明:

$$(KKT) \Rightarrow \begin{cases} -c + \lambda^* A x^* = 0 \\ \lambda^* (\langle Ax^*, x^*\rangle - 1) = 0 \end{cases}$$

若 $c \neq 0$，则 $\lambda^* \neq 0$

$$\Rightarrow \begin{cases} \lambda^* A x^* = c \\ \langle Ax^*, x^*\rangle = 1 \end{cases} \Rightarrow x^* = \frac{A^{-1}c}{\langle c, A^{-1}c\rangle^{\frac{1}{2}}}$$

$\boxed{/\!/}$

LP情形:

$$\min \ \langle c, x\rangle$$
$$\text{s.t.} \ \langle a_i, x\rangle \le 0, \ i=1,\cdots,m$$

$$(KKT) \Rightarrow \begin{cases} c + \sum_{i=1}^{m} \lambda_i^* a_i = 0 \\ \lambda_i^* \langle a_i, x\rangle = 0 \end{cases}$$

(互补松弛条件)

$$[a_1, \cdots, a_m] \begin{bmatrix} \lambda_1^* \\ \vdots \\ \lambda_m^* \end{bmatrix} = -c.$$

考虑 $\min\limits_{x\in\Omega}\{f(x): Ax=b\}$ (3.1.68)

Ω是闭凸集，A是行满秩的

Thm 3.1.27 f在$\Omega\subset \mathrm{int}(\mathrm{dom} f)$上是凸的，且在$\Omega$上的水平集是有界的，设$\exists\bar{x}, \varepsilon>0$, s.t.

$A\bar{x}=b$, $B(\bar{x},\varepsilon)\subseteq\Omega$ (Slater condition for equalities)

则 x^*是(3.1.68)的最优解 $\Leftrightarrow$ $Ax^*=b$，且存在$y^*\in\mathbb{R}^m$,

$g^*\in\partial f(x^*)$, s.t. $\langle g^*-A^Ty^*, x-x^*\rangle\geq 0$, $\forall x\in\Omega$

且 $\|A^Ty^*\|\leq\frac{1}{\varepsilon}\left(\max\limits_{x\in B(\bar{x},\varepsilon)} f(x)-\min\limits_{x\in\Omega} f(x)\right)$

证明：

($\Leftarrow$) 对$\forall x\in\Omega$，且$Ax=b$，则

$f(x)-f(x^*)\geq\langle g^*, x-x^*\rangle$

$$\geqslant \langle y^*, A(x-x^*)\rangle = 0$$

($\Rightarrow$) 考虑 $\phi(x) = f(x) + K\|b-Ax\|$，$K>0$ 是常数

则 $\mathcal{L}_\phi(\beta) \subseteq \mathcal{L}_f(\beta)$，$\forall \beta \in \mathbb{R}$，其中 $\mathcal{L}_\phi$，$\mathcal{L}_f$ 表示 ϕ，f

的在 Ω 上的水平集；ϕ 的凸性是显然的，由 Thm 3.1.11

$f(x)$ 在 Ω 上是连续的，故 $\phi(x)$ 是连续的，由 Ω 是闭的，故

ϕ 是 Ω 上的闭凸 func，由 lemma 3.1.4，$\exists x_*$，s.t.

$$x_* \in \operatorname*{Argmin}_{x\in\Omega} \phi(x)$$

故由 Thm 3.1.24，$\exists g_\phi^* \in \partial\phi(x_*)$，s.t.

$$\langle g_\phi^*, x - x_*\rangle \geqslant 0 \quad \forall x \in \Omega$$

由 lemma 3.1.12，lemma 3.1.11，知

$$g_\phi^* = g^* - KA^T\partial\|b-Ax_*\|, \quad g^* \in \partial f(x_*)$$

由 (3.1.42) 知 $\partial\|x\|_0 = \{\bar{y} \in \mathbb{R}^m : \|\bar{y}\| \leqslant 1\}$

故由lemma 3.1.15，$\exists\, \bar{y} \in \mathbb{R}^m$，$\|\bar{y}\| \leq 1$，s.t.

$$g_\phi^* = g^* - K A^T \bar{y},$$

$$\langle \bar{y}, b - A x_* \rangle = \| b - A x_* \|$$

另一方面 $\forall\, \delta \in B(0, \varepsilon)$，有 $x_\delta = \bar{x} + \delta \in \otimes$，故

$$\begin{aligned} \langle g^*, x_\delta - x_* \rangle &\geq K \langle A^T \bar{y}, \bar{x} + \delta - x_* \rangle \\ &= K \langle \bar{y}, A\delta + b - A x_* \rangle \\ &= K \| b - A x_* \| + K \langle A^T \bar{y}, \delta \rangle \end{aligned}$$

由Thm 3.1.11，$M = \max\limits_{x} \{ f(x) : x \in B(\bar{x}, \varepsilon) \} < +\infty$

则 $\langle g^*, x_\delta - x_* \rangle \leq f(x_\delta) - f(x_*) \leq M - f_*$

其中 $f_* = \min\limits_{x \in \otimes} f(x)$，则

$$\begin{aligned} M - f^* &\geq K \| b - A x_* \| + \max_{\delta \in B(0,\varepsilon)} K \langle A^T \bar{y}, \delta \rangle \\ &\geq \max_{\delta \in B(0,\varepsilon)} K \langle A^T \bar{y}, \delta \rangle \end{aligned}$$

$$\geq K\varepsilon\|A^T\bar{y}\|$$

$$\geq K\varepsilon\mu\|\bar{y}\|$$

其中 $\mu=\lambda_{min}^{\frac{1}{2}}(AA^T)>0$，定义 $y^*=K\bar{y}$，则 (3.1.71) 成立.

另一方面，取 $K>\frac{1}{\varepsilon\mu}(M-f_*)$，则 $\|\bar{y}\|<1$，由 lemma 3.1.15，

$Ax_*=b$，故 x_* 是 (3.1.68) 的最优解

故当 K 足够大时，$\forall$ (3.1.68) 的最优解 x^* 是 ϕ 的全局极小.

故 (3.1.70) 成立

$\boxed{/\!/}$

Thm 3.1.25 令 ϕ 是闭凸 func, $\Omega_1 \subseteq \mathbb{R}^n$, $\Omega_2 \subseteq \mathbb{R}^m$

是两个闭凸集, s.t. $\Omega_1 \times \Omega_2 \subseteq \text{int dom}\,\phi$, 定义

$$f(x) = \inf_{y \in \Omega_2} \phi(x,y)$$

则 f 在 Ω_1 上凸, 且若 $Y(x) = \operatorname*{Argmin}_{y \in \Omega_2} \phi(x,y) \neq \phi$, 则

$$\partial_{\Omega_1} f(x) \supseteq \Big\{ g_x \in \mathbb{R}^n : \exists g_y, \text{ s.t. } (g_x, g_y) \in \bigcap_{y \in Y(x)} \partial\phi(x,y),$$

$$\text{且 } \langle g_y, y - y_x \rangle \geq 0 \quad \forall y \in \Omega_2, \forall y_x \in Y(x) \Big\} \quad (3.1.60)$$

证明: 定义 $\Omega_x = \{(x,y) \mid y \in \Omega_2\}$ 是闭凸集

则 $f(x) = \inf_{y \in \Omega_2} \phi(x,y)$

$$= \inf_{(x,y) \in \Omega_x} \phi(x,y)$$

故由 Thm 3.1.24, $\exists g^* = \begin{pmatrix} g_x \\ g_y \end{pmatrix} \in \widehat{\partial\phi}(x, Y(x))$, s.t.

$$\left\langle g^*, \begin{pmatrix} x - x \\ y - y_x \end{pmatrix} \right\rangle \geq 0 \quad \forall y \in \Omega_2, \forall y_x \in Y(x)$$

$\Leftarrow$ $\langle g_y, y-y_x\rangle \geqslant 0$，故右边集合非空

考虑 $\forall (g_x, g_y)$ 满足上面的条件，令 $x_1 \in \Theta_1$，$\varepsilon>0$

取 $y_1 \in \Theta_2$，s.t. $\phi(x_1, y_1) \leqslant f(x_1)+\varepsilon$，则

$$
\begin{aligned}
f(x_1)+\varepsilon &\geqslant \phi(x_1,y_1) \\
&\geqslant \phi(x,y_x)+\langle g_x, x_1-x\rangle+\langle g_y, y_1-y_x\rangle \\
&\geqslant \phi(x,y_x)+\langle g_x, x_1-x\rangle \\
&= f(x)+\langle g_x, x_1-x\rangle
\end{aligned}
$$

取 ε 任意小，则 $g_x \in \partial_{\Theta_1} f(x)$ $\boxed{}$

Lemma 3.1.21 $\forall \bar{x}$ 是 (3.1.61) 的可行点，则

$$
f_0(\bar{x})-f_0(x^*) \geqslant \sum_{i=1}^{m}(-f_i(\bar{x}))\lambda_i^*
$$

证明：

$$
\begin{aligned}
& f_0(\bar{x})+\sum_{i=1}^{m}\lambda_i^* f_i(\bar{x}) \\
\geqslant\ & f_0(x^*)+\langle \nabla f_0(x^*), \bar{x}-x^*\rangle+\sum_{i=1}^{m}\lambda_i^*[f_i(x^*)+\langle \nabla f_i(x^*), \bar{x}-x^*\rangle] \\
=\ & f_0(x^*)+\sum_{i=1}^{m}\lambda_i^* f_i(x^*)+\langle \nabla f_0(x^*)+\sum_{i=1}^{m}\nabla f_i(x^*), \bar{x}-x^*\rangle
\end{aligned}
$$

(KKT)

$\geqslant 0$

Exact Penalty function

令 $\bar{x} \in \otimes$ 满足 Slater 条件，设已知了 $f_0(\bar{x}) - f_0(x^*)$ 的上界 D

如：$D = \max_{x \in \otimes} \langle \nabla f_0(\bar{x}), \bar{x} - x \rangle$

考虑 $\Lambda = \{ \lambda \in \mathbb{R}^m_+ : \sum_{i=1}^m (-f_i(\bar{x}))\lambda_i \leq D \}$

由 lemma 3.1.21 ：$\lambda^* \in \Lambda$，定义：

$$\psi(g) = \max_{\lambda \in \Lambda} \langle \lambda, g \rangle = D \left(\max_{1 \leq i \leq m} \frac{g^{(i)}}{-f_i(\bar{x})} \right)_+ , \ g \in \mathbb{R}^m$$

证明：

$\forall i$，都非负

$$\langle \lambda, g \rangle = \sum_{i=1}^m \lambda_i g^{(i)} = \sum_{i=1}^m \frac{g^{(i)}}{-f_i(\bar{x})} (-f_i(\bar{x})\lambda_i)$$

令 $\delta_i = \frac{g^{(i)}}{-f_i(\bar{x})}$，$i = 1, \cdots, m$，若

① $\forall i = 1, \cdots, m$ $\delta_i \leq 0$，则 $\langle \lambda, g \rangle \leq 0 = D \left(\max_{1 \leq i \leq m} \delta_i \right)_+$

② $\exists i$, s.t. $\delta_i > 0$，则 $\langle \lambda, g \rangle \leq \max_{1 \leq i \leq m} (\delta_i) \left(\sum_{i=1}^m -f_i(\bar{x})\lambda_i \right)$

$$\leq D\left(\max_{1\leq i\leq m}\left(\frac{g^{(i)}}{-f_i(\bar{x})}\right)\right)_+$$

故 $\max_{\lambda\in\Lambda}\langle\lambda,g\rangle \leq D\left(\max_{1\leq i\leq m}\left(\frac{g^{(i)}}{-f_i(\bar{x})}\right)\right)_+$

另一方面：

① $\forall i=1,\cdots,m$, $g^{(i)}\leq 0$，则取 $\tilde{\lambda}=0\in\Lambda$，有 $\langle\tilde{\lambda},g\rangle=0$

② $\exists i$, $g^{(i)}>0$，令

$$\tilde{\lambda}_i=\begin{cases}\dfrac{D}{-f_j(\bar{x})}, & j=\min\left(\operatorname*{argmax}_{1\leq i\leq m}\dfrac{g^{(i)}}{-f_i(\bar{x})}\right)\\ 0, & \text{otherwise}\end{cases}$$

取 $\tilde{\lambda}=\begin{pmatrix}\tilde{\lambda}_1\\ \vdots\\ \tilde{\lambda}_m\end{pmatrix}\in\Lambda$ 是显然的

$$\langle\tilde{\lambda},g\rangle=g^{(j)}\cdot\frac{D}{-f_j(\bar{x})}$$

$$=\frac{g^{(j)}}{-f_j(\bar{x})}\cdot D=\left(\max_{1\leq i\leq m}\frac{g^{(i)}}{-f_i(\bar{x})}\right)_+ D$$

综上：$\psi(g)=\max_{\lambda\in\Lambda}\langle\lambda,g\rangle=D\left(\max_{1\leq i\leq m}\frac{g^{(i)}}{-f_i(\bar{x})}\right)_+$

$\square$

$$\min_{x \in \Theta} \left\{ \phi(x) = f_0(x) + \psi(f(x)) \right\}$$

其中 $f(x) = (f_1(x), \cdots, f_m(x))$，$x^*$是(3.1.61)的最优点、

注意到 $\max \langle \lambda, f(x^*) \rangle = 0$，定义：

$$\Lambda_+ = \{ \lambda \in \Lambda : \langle \lambda, f(x^*) \rangle = 0 \} = \{ \lambda \in \Lambda : \lambda_i = 0, i \notin I(x^*) \}$$

其中 $I(x^*) = \{ i : f_i(x^*) = 0 \}$，由 $\lambda^* \in \Lambda_+$ (KKT条件)

故 $\displaystyle\sum_{i \in I(x^*)} \lambda_i^* \nabla f_i(x^*) \in \partial \psi(f(x^*))$ (lemma 3.1.16)

故 $\displaystyle g^* = \nabla f_0(x^*) + \sum_{i \in I(x^*)} \lambda_i^* \nabla f_i(x^*) \in \partial \phi(x^*)$

注： $\displaystyle\sum_i \partial f_i \subseteq \partial \sum_i f_i$，$f_i$在$x$处次可微

设 $g_k \in \partial f_k(x)$，则 $f_k(z) \geq f_k(x) + \langle g_k, z-x \rangle \ \forall z$

$\displaystyle\Rightarrow \sum_k f_k(z) \geq \sum_k f_k(x) + \left\langle \sum_k g_k, z-x \right\rangle$，故 $\displaystyle\sum_k g_k \in \partial \sum_k f_k$

上面的证明只需 $f_k(x)$ 在 x 处次可微，若 f_k 有凸性，则有：

Thm 23.8 (Rockafellar)

令 f_i，$i=1,\cdots,m$ 是 proper convex function 于 $\mathbb{R}^n$，则令

$f=\sum_{i=1}^m f_i$，有 $\partial f(x) \supset \partial f_1(x)+\cdots+\partial f_m(x) \quad \forall x$

进一步，若 $\bigcap_{k=1}^m \mathrm{ri}(\mathrm{dom} f_i)$，$i=1,\cdots,m$ 非空，则

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

注：lemma 3.1.12 的确不要闭性，但是要求 $x\in \mathrm{int}(\mathrm{dom} f)$

故由 Thm 3.1.26

$$\langle g^*, x-x^*\rangle \geq 0 \quad \forall x\in \otimes$$

由 Thm 3.1.24，$x^* \in \operatorname{Argmin}_{x\in\otimes} \phi(x)$

下证 $\phi(x)$ 的最小值点也一定是 (3.1.61) 的最优解

令 $\hat{x}$ 是 (3.1.67) 的任意最优解，故 $\exists \hat{\lambda} \in \operatorname{Argmax}_{\lambda\in\Lambda} \langle \lambda, f(\hat{x})\rangle$，

s.t. $\left\langle \nabla f_0(\hat{x}) + \sum_{i=1}^m \hat{\lambda}_i \nabla f_i(\hat{x}), x-\hat{x}\right\rangle \geq 0 \quad \forall x\in\otimes$

不妨设 $\Phi(f(\hat{x}))>0$

则 $\langle\hat{\lambda},-f(\bar{x})\rangle=0$（即 Λ 中的不等式约束是积极的）

注：由 $\hat{\lambda}\in\underset{\lambda\in\Lambda}{\mathrm{Argmax}}\langle\lambda,f(\hat{x})\rangle$，由 $\Phi(f(\hat{x}))>0$

则 $\langle\hat{\lambda},f(\hat{x})\rangle>0$，用反证法，设 $\langle\hat{\lambda},-f(\bar{x})\rangle<0$

则 $\exists k>1$, s.t. $k\hat{\lambda}\in\Lambda$，且

$$\langle k\hat{\lambda},f(\hat{x})\rangle>\langle\hat{\lambda},f(\hat{x})\rangle\text{，与 }\hat{\lambda}\in\underset{\lambda\in\Lambda}{\mathrm{Argmax}}\langle\lambda,f(\hat{x})\rangle$$

矛盾，故

$$0\geqslant f_0(\bar{x})-f_0(\hat{x})\geqslant\langle\nabla f_0(\hat{x}),\bar{x}-\hat{x}\rangle$$

$$\geqslant\sum_{i=1}^{m}\hat{\lambda}_i\langle\nabla f_i(\hat{x}),\hat{x}-\bar{x}\rangle$$

$$\geqslant\langle\hat{\lambda},f(\hat{x})-f(\bar{x})\rangle$$

$$=\Phi(f(\hat{x}))+0$$ 矛盾！

故 $\Phi(f(\hat{x}))=0$ （说明 $\hat{x}$ 一定在 (3.1.61) 的可行域中）

且由 $\bar{x}$ 是 $\phi(x) = f_0(x) + \Phi(f(x))$ 的任一最小值

且 $\Phi(f(\bar{x})) = 0$，故 $\bar{x}$ 是 (3.1.6) 的解

$\square$

Thm 3.1.28 设 f 是凸的，$Q\subseteq \mathrm{int}(\mathrm{dom} f)$ 是闭凸集，设 f 在 Q 上的水平集有界，令 $A\in\mathbb{R}^{m\times n}$ $(n>m)$ 是行满秩，考虑 $\phi(u)=\min\limits_{x\in Q}\{f(x): Ax=u\}$

则 ϕ 是凸的，且对 $\forall u\in\mathbb{R}^m$，$\{x\in \mathrm{int}(Q): Ax=u\}\neq\phi$，有

$$\{y^*: \exists x^*\in Q,\ Ax^*=u,\ g^*\in\partial f(x^*),\ s.t.$$
$$\langle g^*-A^Ty^*, x-x^*\rangle\geq 0\ \ \forall x\in Q\}\subseteq\partial\phi(u)$$

证明：

令 $Q(u)=\{x\in Q: Ax=u\}$，则 $\mathrm{dom}\,\phi=\{u\in\mathbb{R}^m: Q(u)\neq\phi\}$

对 $\forall u\in \mathrm{dom}\,\phi$，$\exists x(u)\in \operatorname*{Argmin}\limits_{x\in Q(u)} f(x)$

由 $Q\subseteq \mathrm{int}(\mathrm{dom} f)$ 是闭凸的，故 f 在 Q 上是连续的

且 $Q(u)$ 是闭凸的，且 $Q(u)\subseteq Q$，故 f 在 $Q(u)$ 上连续，

且闭凸，且由假设下水平集有界，故 $\operatorname*{Argmin}\limits_{x\in Q(u)} f(x)$ 可解

令 $u_1, u_2 \in \mathrm{dom}\,\phi$，$\alpha \in [0,1]$，则

$$x_\alpha \doteq \alpha x(u_1) + (1-\alpha) x(u_2) \in \otimes\big(\overset{\in\,\mathrm{dom}\,\phi}{\underline{\alpha u_1 + (1-\alpha) u_2}}\big)$$

故 $\mathrm{dom}\,\phi$ 是凸集，且

$$\phi(\alpha u_1 + (1-\alpha) u_2) \leq f(x_\alpha) \leq \alpha f(x(u_1)) + (1-\alpha) f(x(u_2))$$

$$= \alpha \phi(u_1) + (1-\alpha) \phi(u_2)$$

由 Thm 3.1.27，(3.1.74) 左侧非空，取 (x^*, y^*, g^*) 对某个 $u = u_1 \in \mathrm{dom}\,\phi$，则对 $u_2 \in \mathrm{dom}\,\phi$，有

$$\phi(u_2) = f(x(u_2)) \geq f(x^*) + \langle g^*, x(u_2) - x^* \rangle$$

$$\geq f(x^*) + \langle A^T y^*, x(u_2) - x^* \rangle$$

$$\overset{\text{Thm3.1.27}}{=} \phi(u_1) + \langle y^*, u_2 - u_1 \rangle$$

故 $y^* \in \partial \phi(u_1)$

$\square$

§3.1.8 Minimax Theorems

$\Phi(\cdot,\cdot)$定义在两个凸集的直积，$P\subseteq\mathbb{R}^n$，$S\subseteq\mathbb{R}^m$

设 $\Phi(\cdot,u)$是闭凸的于 $P\subseteq \mathrm{dom}\,\Phi(\cdot,u)$，$\forall u\in S$

$\Phi(x,\cdot)$是闭凹的于 $S\subseteq \mathrm{dom}\,\Phi(x,\cdot)$，$\forall x\in P$

核心目的是：

$$\inf_{x\in P}\sup_{u\in S}\Phi(x,u)=\sup_{u\in S}\inf_{x\in P}\Phi(x,u)$$

何时成立？ 一般地，只有左边$\geq$右边

由
$$\Phi(x,u)\geq \boxed{\inf_{x\in P}\Phi(x,u)}\ \phi(u)$$

$$\Rightarrow \overset{f(x)}{\boxed{\sup_{u\in S}\Phi(x,u)}}\geq \sup_{u\in S}\inf_{x\in P}\Phi(x,u)$$

$$\Rightarrow \inf_{x\in P}\sup_{u\in S}\Phi(x,u)\geq \sup_{u\in S}\inf_{x\in P}\Phi(x,u)$$

lemma 3.1.22 设对 $\forall u \in S$，$\Phi(\cdot, u)$ 的下水平集在 P 上有界，且 ϕ 在 $u^* \in S$ 上达到最大，则对 $\forall u \in S$，有

$$\min_{x \in P} \max\{\Phi(x,u), \Phi(x,u^*)\} = \phi(u^*)$$

证明：

$$f_u(x) = \max\{\Phi(x,u), \Phi(x,u^*)\}$$
$$\geq \max\{\phi(u), \phi(u^*)\} = \phi(u^*)$$

首先说明 $\min_{x \in P} f_u(x)$ 可解

由 Thm3.1.5，$f_u(x)$ 是闭的 func，且对 $\forall \beta \in \mathbb{R}$

$$\{\underset{P}{x} : f_u(x) \leq \beta\} \subseteq \{\underset{P}{x} : \Phi(x,u) \leq \beta\} \text{ 有界}$$

故 $\min_{x \in P} f_u(x)$ 可解

由 Thm3.1.10，$\exists \lambda^* \in [0,1]$，s.t.

$$\min_{x \in P} f_u(x) = \min_{x \in P}\{\lambda^* \Phi(x,u) + (1-\lambda^*)\Phi(x,u^*)\}$$

$$\leq \min_{x\in P} \Phi(x, \lambda^* u+(1-\lambda^*)u^*)$$

$$= \phi(\lambda^* u+(1-\lambda^*)u^*)$$

故 $\phi(u^*) \leq \min_{x\in P} f_u(x) \leq \phi(\lambda^* u+(1-\lambda^*)u^*)$

$$\leq \phi(u^*)$$

故 $\min_{x\in P} f_u(x) = \phi(u^*)$ $\square$

Thm 3.1.29 设 $\Phi(\cdot,u)$ 在 P 上可达唯一最小，ϕ 在 S 上可达最大值，则

$$\min_{x\in P} f(x) = \max_{u\in S} \phi(u)$$

证明：由 $x(u) = \operatorname{argmin}_{x\in P} \Phi(x,u)$ 是唯一定义的，则

由 Thm 3.1.4 (5)，$\Phi(\cdot,u)$，$u\in S$ 的下水平集在 P 上有界，

故由 lemma 3.1.22：对 $\forall u\in S$

$$\min_{x\in P} \max\{\Phi(x,u), \Phi(x,u^*)\} = \phi(u^*)$$

由 $\phi(u^*)=\Phi(x(u^*),u^*)$，知 $x(u^*)$ 是 (3.1.76) 唯一最小值点：这是由于 x^* 是 (3.1.76) 的最小值点，

$$\Rightarrow \quad \Phi(x^*,u^*) \leq \phi(u^*) = \min_{x\in P} \Phi(x,u^*)$$

故 $x(u^*)$ 是唯一使 $\Phi(x(u^*),u^*)=\phi(u^*)$ 的点。

故 $\Phi(x(u^*),u) \leq \Phi(x(u^*),u^*) \leq \Phi(x,u^*)$，$x\in P$

故 $\sup\limits_{u\in S} \Phi(x(u^*),u) \leq \inf\limits_{x\in P} \Phi(x,u^*)$

$$f(x(u^*)) \leq \phi(u^*)$$

$$\Rightarrow \quad \min_{x\in P} f(x) = f(x(u^*)) \leq \phi(u^*) = \max_{u\in S}\phi(u)$$

$$\begin{aligned} f(x(u^*)) &= \sup_{u\in S} \Phi(x(u^*),u) \\ &\leq \phi(u^*) \\ &= \inf_{x\in P}\Phi(x,u^*) \\ &\leq \inf_{x\in P}\sup_{u\in S}\Phi(x,u) = \inf_{x\in P} f(x). \end{aligned}$$

$\square$

Thm 3.1.10 设 P, Ω 有界，则

$$\min_{x\in P} f(x) = \max_{u\in S} \phi(u)$$

证明：固定 $\varepsilon>0$，$\|\cdot\|$ 是欧式 norm，定义

$$\Phi_\varepsilon(x,u) = \Phi(x,u) + \frac{1}{2}\varepsilon\|x\|^2,\ x\in P,\ u\in S$$

则 $\Phi_\varepsilon(\cdot,u)$ 是强凸的 ($\forall u\in S$)，故在 P 上存在唯一最小值

由 $\{x\in P: \Phi_\varepsilon(x,u)\le\beta\} \subseteq \underbrace{\{x\in P: \Phi(x,u)\le\beta\}}_{\text{是紧集，设为}B}$

故 $\inf_{x\in P}\Phi_\varepsilon(x,u) \Leftrightarrow \inf_{x\in B}\Phi_\varepsilon(x,u)$，且 $\Phi_\varepsilon(\cdot,u)$ 在 B 上

是闭凸的，可以用 Thm 3.1.4 4) 说明 $\min_{x\in B}\Phi_\varepsilon(x,u)$ 的

well-define，且唯一性见 (P210，推论 3.2.3)

，故 $\phi_\varepsilon(u)=\min\limits_{x\in P}\Phi_\varepsilon(x,u)$ 是 well-defined.

由 Thm3.1.8　$\phi_\varepsilon(u)=-\max\limits_{x\in P}\left(-\Phi_\varepsilon(x,u)\right)$ 是闭凹的

下面证明 $\max\limits_{u\in S}\phi_\varepsilon(u)$ 是可解的，只需证 $\min\limits_{u\in S}-\phi_\varepsilon(u)$ 闭凸

是 well-defined：直接用 Thm3.1.4　4) 即可

对 $\Phi_\varepsilon(x,u)$ 用 Thm3.1.29：

$\exists\, u_\varepsilon^*\in\underset{u\in S}{\mathrm{Argmax}}\,\phi_\varepsilon(u)$，$x_\varepsilon^*=\underset{x\in P}{\mathrm{Argmin}}\,\Phi_\varepsilon(x,u_\varepsilon^*)$，s.t.

$$\Phi_\varepsilon(x_\varepsilon^*,u)\leq\Phi_\varepsilon(x_\varepsilon^*,u_\varepsilon^*)\leq\Phi_\varepsilon(x,u_\varepsilon^*)\ ,x\in P,u\in S$$

$$\Rightarrow\quad \Phi(x_\varepsilon^*,u)\leq\Phi(x_\varepsilon^*,u_\varepsilon^*)$$

$$\Rightarrow\quad f(x_\varepsilon^*)=\sup_{u\in S}\Phi(x_\varepsilon^*,u)\leq\Phi(x_\varepsilon^*,u_\varepsilon^*)$$

另一方面：对 $\forall x\in P$，有

$$\Phi(x_\varepsilon^*,u_\varepsilon^*)\leq\Phi_\varepsilon(x_\varepsilon^*,u_\varepsilon^*)\leq\Phi(x,u_\varepsilon^*)+\frac{1}{2}\varepsilon D^2\ ,\ D\geq\sup_{x\in P}\|x\|$$

故 $f(x_\varepsilon^*) \leq \min\limits_{x\in P} \Phi(x,u_\varepsilon^*) + \frac{1}{2}\varepsilon D^2$

$= \phi(u_\varepsilon^*) + \frac{1}{2}\varepsilon D^2$

Thm 3.1.4

$f(x) = \max\limits_{u\in S} \Phi(x,u)$

注：$f(x) = \sup\limits_{u\in S} \Phi(x,u)$ 是闭凸的

$\phi(u) = \min\limits_{x\in P} \Phi(x,u) = -\max\limits_{x\in P}(-\Phi(x,u))$ 是闭凹的

取 $\varepsilon\downarrow 0$，有

$$\liminf_{\varepsilon\downarrow 0} f(x_\varepsilon^*) \leq \limsup_{\varepsilon\downarrow 0} \phi(u_\varepsilon^*)$$

由 $\{x_\varepsilon^*\}_{\varepsilon\downarrow 0} \subseteq B$（是紧的），故不妨设 $x_\varepsilon^* \to \bar{x} \in B \subseteq P$

而 $\{u_\varepsilon^*\}$ 是 $\min\limits_{u\in S} -\phi_\varepsilon(u)$ 的某个解

$$\{u\in S: -\phi_\varepsilon(u) \leq \beta\} = \{u\in S: \max_{x\in P}(-\Phi_\varepsilon(x,u)) \leq \beta\}$$

$$\subseteq \{u\in S: -\Phi_\varepsilon(\tilde{x},u) \leq \beta\},\ \tilde{x}\in P$$

$$\subseteq \underbrace{\{u\in S: -\Phi_{\hat{\varepsilon}}(\tilde{x},u) \leq \beta\}}_{\text{是一个紧集}} \overset{\Delta}{=} C,\ \tilde{x}\in P,\ \hat{\varepsilon} = \sup\varepsilon$$

$$\min_{u\in S} -\phi_\varepsilon(u) \iff \min_{u\in C} -\phi_\varepsilon(u)$$

故 $\{u_\varepsilon^*\}\subseteq C\subseteq S$，不妨设 $u_\varepsilon^*\to\bar{u}\in C\subseteq S$

故 $f(\bar{x})\le \liminf\limits_{\varepsilon\downarrow 0} f(x_\varepsilon^*)\le \limsup\limits_{\varepsilon\downarrow 0}\phi(u_\varepsilon^*)\le\phi(\bar{u})$

故 $\min\limits_{x\in P} f(x)\le \max\limits_{u\in S}\phi(u)$ $\square$

<u>Thm 3.1.31</u> 设 f 在 $x^*\in P$ 处达到极小，设对某个 $g_*\in\partial_P f(x)$，

满足一阶最优性条件：

$$\langle g_*, x-x^*\rangle\geqslant 0,\quad x\in P$$

若 g_* 存在如下表示：

$$g_*=\sum_{i=1}^{k}\lambda^{(i)}g_i$$

对于确定的 $k\geqslant 1$，$\lambda\in\Delta_k$，$g_i\in\partial_{p,x}\Phi(x^*,u_i)$，其中

$u_i\in I(x^*)$，$i=1,\cdots k$，$I(x^*)=\{u\in S:\Phi(x^*,u)=f(x^*)\}$

则 (3.1.78) 成立

证明：令 $\bar{u}=\sum_{i=1}^{k}\lambda^{(i)}u_i$，则对 $\forall x\in P$，有：

$$
\begin{aligned}
f(x^*) &\leqslant f(x^*)+\langle g_*, x-x^*\rangle \\
&= f(x^*)+\sum_{i=1}^{k}\lambda^{(i)}\langle g_i, x-x^*\rangle \\
&\leqslant f(x^*)+\sum_{i=1}^{k}\lambda^{(i)}[\Phi(x,u_i)-\Phi(x^*,u_i)] \\
&= \sum_{i=1}^{k}\lambda^{(i)}\Phi(x,u_i) \quad (\text{由 } u_i \text{ 的定义}) \\
&\leqslant \Phi(x,\bar{u}) \quad (\text{由Jensen不等式})
\end{aligned}
$$

故 $f(x^*)\leqslant \inf_{x\in P}\Phi(x,\bar{u})=\phi(\bar{u})$

故 $\min_{x\in P} f(x)\leqslant \max_{u\in S}\phi(u)$ $\boxtimes$

Lemma 3.1.14 *Let Δ be an arbitrary set, and $f(x) = \sup\{\phi(x, y) \mid y \in \Delta\}$. Suppose that for any $y \in \Delta$ the function $\phi(\cdot, y)$ is closed and convex on some convex set Q. Then f is closed convex on the set*

$f(x)=\sup\limits_{u\in S} \Phi(x,u)$

$$\hat{Q} = \left\{ x \in Q \mid \sup_{y \in \Delta} \phi(x, y) < +\infty \right\}.$$

Moreover, for any $x \in \hat{Q}$ we have $\{\partial_{Q,x}\Phi(x,u) \mid u\in I(x)\}$

$$\partial_{\hat{Q}} f(x) \supseteq \text{Conv}\,\{\partial_{Q,x}\phi(x, y) \mid y \in I(x)\},$$

where $I(x) = \{y \in \Delta \mid \phi(x, y) = f(x)\}$.

$I(x)=\{u\in S : \Phi(x,u)=f(x)\}$

lemma 3.1.14

$$\partial_{\hat{P}} f(x^*) \supseteq \text{Conv}\{\partial_{P,x}\Phi(x^*,u) \mid u\in I(x^*)\}$$

$$I(x^*)=\{u\in S : \Phi(x^*,u)=f(x^*)\}$$

故 Thm 3.1.31 条件成立的充分条件是

$$\partial_{\hat{P}} f(x^*) = \text{Conv}\{\partial_{P,x}\Phi(x^*,u) : u\in I(x^*)\}$$

$\|$

$\partial_P f(x^*)$

对 $\forall g\in\partial_{\hat{P}} f(x^*)$

$f(y)\geq f(x^*)+\langle g, y-x^*\rangle,\ \forall y\in\hat{P}$

$\Rightarrow f(y)\geq f(x^*)+\langle g, y-x^*\rangle,\ \forall y\in P$

$\Rightarrow g\in\partial_P f(x^*)$

注1° 有错 Thm 3.1.29 的条件可以把 $\Phi(\cdot,u)$ 在 P 上有唯一最小放松成 $\Phi(\cdot,u)$ 在 P 上的最小值集合有界

证明：只需注意到对 $\forall\, x_* \in \underset{x\in P}{\operatorname{argmin}}\, \Phi(x,u^*)$，有

$\phi(u^*) = \Phi(x_*,u^*)$，不妨设 x^* 是 (3.1.76) 的最小值点。

与 u 相关

则 $$\Phi(x^*,u^*) \le \phi(u^*) = \min_{x\in P} \Phi(x,u^*)$$

故 $x^* \in \underset{x\in P}{\operatorname{argmin}}\, \Phi(x,u^*)$，故对 $\forall u\in S$：

$$\Phi(x^*,u) \le \Phi(x^*,u^*) \le \Phi(x,u^*),\ x\in P$$

$$\Rightarrow f(x^*) \le \phi(u^*)$$ $\boxtimes$

注2° $$\partial_p f(x^*) = \operatorname{Conv}\{\partial_{p,x}\Phi(x^*,u) : u\in I(x^*)\} \quad (*)$$

成立的充分条件：

THEOREM 3. *Let $x\in\mathbb{R}^n$ be such that, for some $\varepsilon_0>0$,*

(i) *the set $T_{\varepsilon_0}(x)$ is compact,*

(ii) *for each $z\in\operatorname{dom} f$ the function $t\mapsto f_t(z)$ is usc on $T_{\varepsilon_0}(x)$.*

Then

$$\partial f(x) = \operatorname{co}\left\{\bigcup_{t\in T(x)} \partial(f_t + \mathrm{I}_{\operatorname{dom} f})(x)\right\} \tag{20}$$

and, under condition (18),

$$\partial f(x) = \operatorname{co}\left\{\bigcup_{t\in T(x)} \partial f_t(x)\right\} + \mathrm{N}_{\operatorname{dom} f}(x). \tag{21}$$

(i) $T_{\varepsilon_0}(x) = \{t \in T \mid f_t(x) \geq f(x) - \varepsilon_0\}$

换成书中的记号：$\exists\, \varepsilon_0 > 0$, s.t.

$$T_{\varepsilon_0}(x) = \{u \in S \mid \Phi(x, u) \geq f(x) - \varepsilon_0\} \text{ 是紧的}$$

(ii) $\Phi(x, u)$ 对 $u \in T_{\varepsilon_0}(x)$ 是上半连续的，这是自然成立的。由于 $\Phi(x, \cdot)$ 在 S 上是闭凹的

(iii)

$$N_{\mathrm{dom}f}(x) = \begin{cases} 0, & x \in \mathrm{int}(\mathrm{dom}f) \\ \phi, & x \notin \mathrm{dom}f \end{cases}$$

故⊛成立只需 $x^* \in \mathrm{int}(\mathrm{dom}f)$ + 验证条件 (i)

注 3°：Thm 3.1.29 的对称命题是：令 $\Phi(x, \cdot)$ 在 S 上极大值集合存在且有界，f 在 P 上达到最小，则 (3.1.78) 成立

而 $\Phi(x, \cdot)$ 的极大值集有界 $\xRightarrow{\text{Thm 3.1.45}}$ 条件 (i) $\xRightarrow{\text{SIAM Opt}}$ Thm 3.1.31 条件

$\Rightarrow$ (3.1.78)

§ 3.1.9 Basic Elements of Primal-dual Methods

考虑问题 $f^* = \min_{x \in P} f(x)$

f 是 P 上的闭凸func，设 f 有下面的 max-表示：

$$f(x) = \max_{u \in S} \Psi(x, u)$$

Ψ 满足 §3.1.8 的所有假设，则对偶问题是

$$\phi^* = \max_{u \in S} \phi(u), \quad \phi(u) = \min_{x \in P} \Psi(x, u)$$

原问题和对偶问题在数值上并非是对称的，因为 $\phi(u)$ 难以拿到，但若假设 Oracle 能访问 (3.1.83) 的内部结构：

为了计算 $f(x)$，只需计算 $u(x) \in \operatorname{Arg}\max_{u \in S} \Psi(x, u)$

令 $g(x) \in \partial_{P,x} \Psi(x, u(x))$

设 Oracle 的返回值是：$f(x), g(x), u(x)$

设在 $\{y_k\}_{k=0}^{N} \subset P$ 上积累了信息，取系数：

$$\alpha_k > 0,\ k=0,\cdots,N,\ \sum_{k=0}^{N} \alpha_k = 1$$

考虑目标func的线性model：

$$\ell_N(x) = \sum_{k=0}^{N} \alpha_k [f(y_k) + \langle g(y_k), x-y_k \rangle] \leq f(x),\ x \in P$$

在某些算法中 $\boxed{f(x_N) \leq \min_{x \in P} \ell_N(x) + r_N}$ (3.1.85)

其中 $\{x_N\}$ 是最小化序列，$N \to \infty$ 时，$r_N \to 0$

$$\text{取 } \hat{u}_N = \sum_{k=0}^{N} \alpha_k u(y_k) \in S$$

lemma 3.1.23 设 x_N 满足 (3.1.85)，则

$$0 \leq (f(x_N) - f^*) + (\phi^* - \phi(\hat{u}_N)) \leq f(x_N) - \phi(\hat{u}_N) \leq r_N$$

证明： $g(y_k) \in \partial_{p,x} \Phi(y_k, u(y_k))$，则

$$\min_{x \in P} \ell_N(x) = \min_{x \in P} \sum_{k=0}^{N} \alpha_k [\Phi(y_k, u(y_k)) + \langle g(y_k), x-y_k \rangle]$$

$$\leq \min_{x\in P} \sum_{k=0}^{N} \alpha_k \Phi(x, u(y_k))$$

$$\leq \min_{x\in P} \Phi(x, \hat{u}_N)$$

$$= \phi(\hat{u}_N)$$

故 $f(x_N) \leq \min_{x\in P} t_N(x) + r_N \leq \phi(\hat{u}_N) + r_N$ $\boxtimes$

注：对偶问题的 good solution $\hat{u}_N = \sum_{k=1}^{N} \alpha_k u(y_k)$ 不需要计算 $\phi(u)$！

返过来，利用 Test points $\{y_k\}_{k=0}^{N}$ 生成原问题的 good solution

定义 $$\delta_N(x) = \sum_{k=0}^{N} \alpha_k \langle g(y_k), y_k - x \rangle$$

$$\hat{f}_N = \sum_{k=0}^{N} \alpha_k f(y_k) \geq f(\sum \alpha_k y_k)$$

<u>lemma 3.1.24</u> 设 $\max_{x\in P} \delta_N(x) \leq r_N \to 0$，则

$$0 \leq (\hat{f}_N - f^*) + (\phi^* - \phi(\hat{u}_N)) \leq \hat{f}_N - \phi(\hat{u}_N) \leq r_N \to 0$$

证明：

$$\max_{x\in P}\delta_N(x)=\max_{x\in P}\sum_{k=0}^{N}\alpha_k\langle g(y_k),y_k-x\rangle$$

$$\geqslant \max_{x\in P}\sum_{k=0}^{N}\alpha_k\Big(\Psi(y_k,u(y_k))-\Psi(x,u(y_k))\Big)$$

$$=\sum_{k=0}^{N}\alpha_k f(y_k)-\min_{x\in P}\sum_{k=0}^{N}\alpha_k\Psi(x,u(y_k)).$$

$$\geqslant \hat{f}_N-\min_{x\in P}\Psi(x,\hat{u}_N)$$

$$=\hat{f}_N-\phi(\hat{u}_N)$$

$\boxtimes$

§ 3.2.1 General Lower Complexity Bounds

$$\min_{x \in \mathbb{R}^n} f(x)$$

f 是凸 func，x^* 是最优解

Model	① Unconstrained minimization ② f 是凸函，在有界集上 Lipschiz
Oracle	First-order Black Box： 在任意点 $\hat{x}$，可以算 $f(\hat{x})$，$g(\hat{x}) \in \partial f(\hat{x})$， $g(\hat{x})$ 是任意一个次梯度
Approximate solution	Find $\bar{x} \in \mathbb{R}^n$: $f(\bar{x}) - f^* \leq \varepsilon$
Methods	$x_k \in x_0 + \mathrm{Lin}\{g(x_0), \cdots, g(x_{k-1})\}$

固定 $\mu>0, \gamma>0$，考虑

$$f_k(x)=\gamma \max_{1\le i\le k} x^{(i)}+\frac{\mu}{2}\|x\|^2 , k=1,\cdots,n$$

则 $\partial f_k(x)=\mu x+\gamma \operatorname{Conv}\{e_i \mid i\in I(x)\}$

$$I(x)=\{j \mid 1\le j\le k, x^{(j)}=\max_{1\le i\le k} x^{(i)}\}$$

令 x_k^* 是 f_k 的全局最小（存在性和唯一性根据 f_k 的强凸）

对 $\forall x, y\in B_2(x_k^*, \rho), \rho>0, g_k(y)\in\partial f_k(y)$，则

$$\begin{aligned}
f_k(y)-f_k(x) &\le \langle g_k(y), y-x\rangle \\
&\le \|g_k(y)\|\|y-x\| \\
&= \|\mu y+\gamma \sum_{i\in I(y)}\alpha_i e_i\| \quad \|y-x\| \\
&\le \left(\mu\|y\|+\gamma\sum_{i\in I(y)}\alpha_i\|e_i\|\right)\|y-x\| \\
&\le \left(\mu\|x_k^*\|+\mu\rho+\gamma\right)\|y-x\|
\end{aligned}$$

同理 $f_k(x)-f_k(y)\leqslant(\mu\|x_k^*\|+\mu\rho+\gamma)\|y-x\|$

故 f_k 在 $B_2(x_k^*,\rho)$ 上 Lipschiz 连续，$M=\mu\|x_k^*\|+\mu\rho+\gamma$

由 Thm 3.1.20，$0\in\partial f_k(x_k^*)$

$$\partial f_k(x_k^*)=\mu x_k^*+\gamma\,\mathrm{Conv}\{e_i\mid i\in I(x_k^*)\}$$

① $(x_k^*)^{(i)}=0$，$i=k+1,\cdots,n$，否则 $\forall g\in\partial f_k(x_k^*)$，

$g^{(i)}=\mu(x_k^*)^{(i)}\neq 0$，与 $0\in\partial f_k(x_k^*)$ 矛盾

② $(x_k^*)^{(1)}=\cdots=(x_k^*)^{(k)}$，否则 $\exists\,\xi_1,\xi_2\in\{1,\cdots,k\}$，s.t.

$(x_k^*)^{(\xi_1)}<(x_k^*)^{(\xi_2)}$，那么对 $\forall g\in\partial f_k(x_k^*)$

$g^{(\xi_1)}=\mu(x_k^*)^{(\xi_1)}$，除非 $(x_k^*)^{(\xi_1)}=0$，否则 $0\notin\partial f_k(x_k^*)$

但若 $(x_k^*)^{(\xi_1)}=0$，则 $(x_k^*)^{(\xi_2)}>0$，此时 x_k^* 一定不是最小值

故 $(x_k^*)^{(1)}=\cdots=(x_k^*)^{(k)}=\lambda$

$$\mu\lambda + \frac{\gamma}{k} = 0 \Rightarrow \lambda = -\frac{\gamma}{\mu k}$$

故

$$(x_k^*)^{(i)} = \begin{cases} -\frac{\gamma}{\mu k}, & 1 \le i \le k \\ 0, & k+1 \le i \le n \end{cases}$$

定义：$R_k \doteq \|x_k^*\| = \frac{\gamma}{\mu\sqrt{k}}$

$$f_k^* = -\frac{\gamma^2}{\mu k} + \frac{\mu}{2} R_k^2 = -\frac{\gamma^2}{2\mu k}$$

$$M = \mu\|x_k^*\| + \mu p + \gamma = \mu p + \gamma\frac{\sqrt{k}+1}{\sqrt{k}}$$

给定 Oracle

Input	$x \in \mathbb{R}^n$
MainLoop	$f := -\infty,\ i^* = 0$ for $j := 1$ to k do if $x^{(j)} > f$ then $\{ f := x^{(j)},\ i^* = j \}$ $f = \gamma f + \frac{\mu}{2}\|x\|^2$; $g = \gamma e_{i^*} + \mu x$
Output	$f_k(x) = f,\ g_k(x) = g$

定义 $\mathbb{R}^{p,n}=\{x\in\mathbb{R}^n \mid x^{(i)}=0,\ p+1\le i\le n\}$

则对 $\forall i,\ 1\le i\le k$，有 $x_i\in\mathbb{R}^{i,n}$，故对 $1\le i\le k-1$

$$f_k(x_i)\geq \gamma \max_{1\le j\le k} x_i^{(j)}=0$$

考虑问题类 $\mathcal{P}(x_0,R,M)$，在(3.2.2)下，再假设：

- x_0 离 x^* 足够近：$\|x_0-x^*\|\le R$
- f 在 $B_2(x^*,R)$ 上是 M-Lipschiz连续的

<u>Thm 3.2.1</u> 对任意类 $\mathcal{P}(x_0,R,M)$，$\forall k,\ 0\le k\le n-1$，存在

$f\in\mathcal{P}(x_0,R,M)$，s.t.

$$f(x_k)-f^*\geqslant\frac{MR}{2(2+\sqrt{k+1})}$$

对任意一阶方法成立，i.e.，

$$x_k\in x_0+\mathrm{Lin}\{g(x_0),\cdots,g(x_{k-1})\}$$

证明：WLOG，设 $x_0=0$，取 $f(x)=f_{k+1}(x)$，

$$\gamma = \frac{\sqrt{k+1}\,M}{2+\sqrt{k+1}}, \quad \mu = \frac{M}{(2+\sqrt{k+1})R}$$

则 $f^* = f^*_{k+1} = -\dfrac{\gamma^2}{2\mu(k+1)} = -\dfrac{MR}{2(2+\sqrt{k+1})}$

$\|x_0-x^*\| = R_{k+1} = \dfrac{\gamma}{\mu\sqrt{k+1}} = R$

且 f 在 $B_2(x^*,R)$ 上 M-Lipschiz连续，故由 $x_k\in\mathbb{R}^{k,n}$

$$f(x_k)-f^* \geq -f^* = \frac{MR}{2(2+\sqrt{k+1})}$$

$\square$

§ 3.2.2 Estimating Quality of Approximate Solutions

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

Ω是闭凸集，f是$\mathbb{R}^n$上的凸func

$$\boxed{对\ \forall x\in\Omega,\ \langle g(x), x-x^*\rangle \geq 0}$$

固定$\bar{x}\in\mathbb{R}^n$，对$x\in\mathbb{R}^n$，$g(x)\neq 0$，定义

$$v_f(\bar{x}, x)=\frac{1}{\|g(x)\|}\langle g(x), x-\bar{x}\rangle$$

若$g(x)=0$，则定义$v_f(\bar{x},x)=0$，故由C-S不等式

$$v_f(\bar{x},x)\leq \|x-\bar{x}\|$$

$v_f(\bar{x},x)$有自然的几何解释，设x满足$g(x)\neq 0$，且

$\langle g(x), x-\bar{x}\rangle \geq 0$，则定义

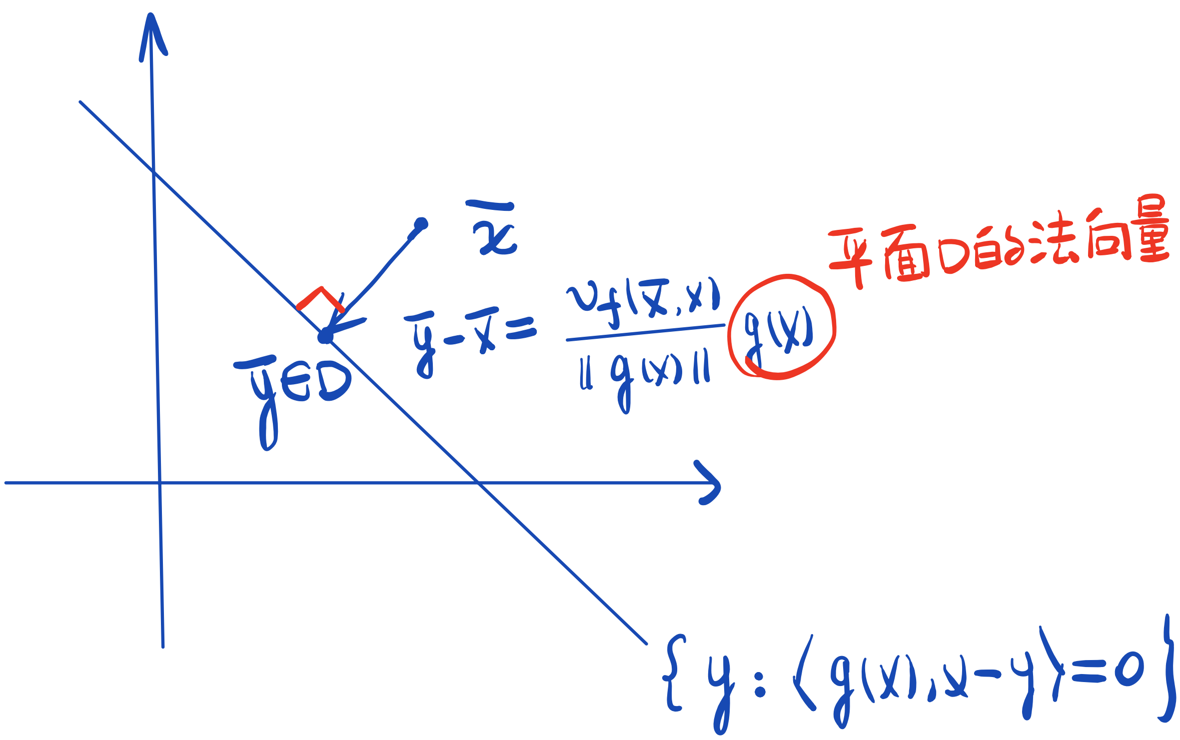
$$\bar{y} = \bar{x} + v_f(\bar{x}, x)\frac{g(x)}{\|g(x)\|}$$

则 $\langle g(x), x-\bar{y}\rangle = \langle g(x), x-\bar{x}\rangle - v_f(\bar{x}, x)\|g(x)\| = 0$

令 $D=\{y : \langle g(x), x-y\rangle = 0\}$，则 $\bar{y}\in D$

且 $\|\bar{y}-\bar{x}\| = v_f(\bar{x}, x)$

断言：此时 $v_f(\bar{x}, x)$ 是 $\bar{x}$ 到平面D的距离



故 $(\bar{y}-\bar{x})\perp D$，即 $\forall y\in D$，$\langle \bar{y}-\bar{x}, y-\bar{y}\rangle = 0$

对 $t\geqslant 0$，定义

$$\omega_f(\bar{x};t)=\max_x\{f(x)-f(\bar{x}):\|x-\bar{x}\|\leqslant t\}$$

若 $t<0$，令 $\omega_f(\bar{x};t)=0$

- $\omega_f(\bar{x};t)=0$，对 $\forall t\leqslant 0$
- $\omega_f(\bar{x};t)$ 关于 $t\in\mathbb{R}$ 是不减的
- $f(x)-f(\bar{x})\leqslant\omega_f(\bar{x};\|x-\bar{x}\|)$

Lemma 3.2.1 对 $\forall x\in\mathbb{R}^n$，有

$$f(x)-f(\bar{x})\leqslant\omega_f(\bar{x};v_f(\bar{x};x))\qquad(3.2.11)$$

若 $f(\cdot)$ 在 $B_2(\bar{x},R)$ 上是 M-Lipschitz 的，则

$$f(x)-f(\bar{x})\leqslant M(v_f(\bar{x};x))_+\qquad(3.2.12)$$

对 $\forall x\in\mathbb{R}^n$ 满足 $v_f(\bar{x};x)\leqslant R$ 成立

证明：若 $\langle g(x), x-\bar{x}\rangle < 0$，则

$$f(\bar{x}) \geq f(x) + \langle g(x), \bar{x}-x\rangle \geq f(x)$$

由 $v_f(\bar{x};x)<0$，故 $w_f(\bar{x}; v_f(\bar{x};x)) = 0$，故 (3.2.11) 成立

另一方面，若 $\langle g(x), x-\bar{x}\rangle \geq 0$，令

$$\bar{y} = \bar{x} + v_f(\bar{x};x)\frac{g(x)}{\|g(x)\|}$$

则有 $\langle g(x), \bar{y}-x\rangle = 0$，$\|\bar{y}-\bar{x}\| = v_f(\bar{x};x)$，故

$$f(\bar{y}) \geq f(x) + \langle g(x), \bar{y}-x\rangle = f(x)$$

因

$$f(x)-f(\bar{x}) \leq f(\bar{y})-f(\bar{x}) \leq w_f(\bar{x}; \|\bar{y}-\bar{x}\|)$$

$$= w_f(\bar{x}; v_f(\bar{x};x))$$

若 f 在 $B_2(\bar{x}, R)$ 上 Lipschiz 连续，且 $0 \leq v_f(\bar{x};x) \leq R$，

则 $\bar{y} \in B_2(\bar{x}, R)$，故

$$f(x)-f(\bar{x}) \leq f(\bar{y})-f(\bar{x}) \leq M\|y-x\| = M\, v_f(\bar{x};x)$$

若 $v_f(\bar{x};x)<0$，则 $f(\bar{x}) \geq f(x)+\langle g(x), \bar{x}-x\rangle \geq f(x)$，

综上，(3.2.12) 成立 $\square$

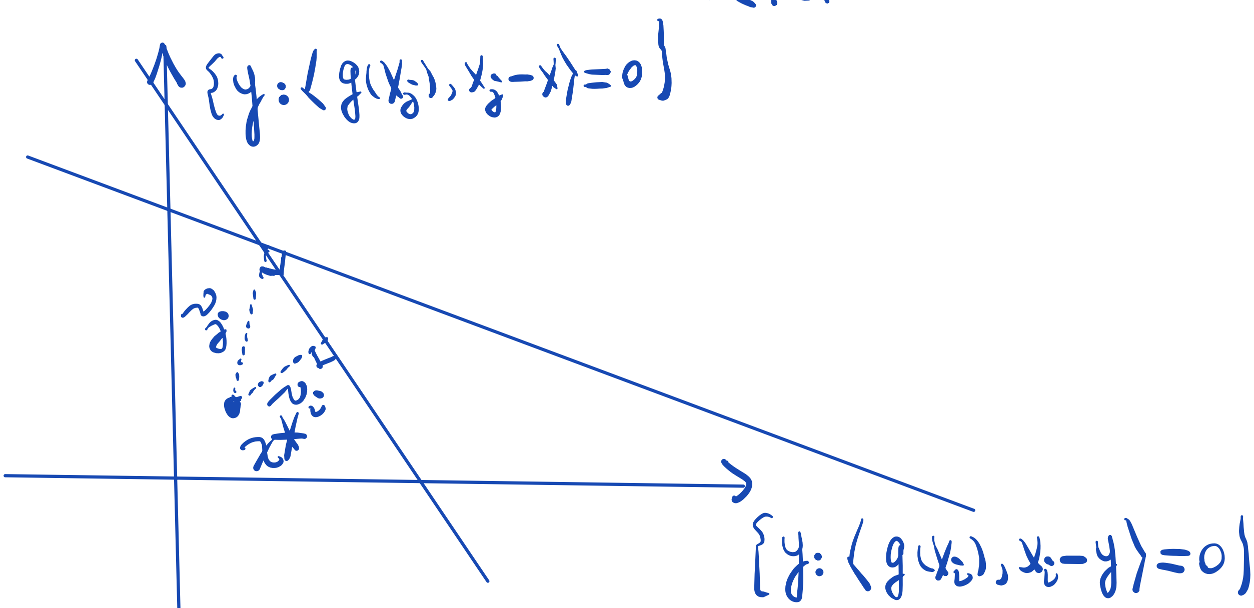
Def 3.2.1 令 $\{x_i\}_{i=0}^{\infty}$ 是 Q 上的序列，定义

$$S_k = \{x\in Q \mid \langle g(x_i), x_i - x\rangle \geq 0, i=0,\cdots,k\}$$

称 S_k 是由序列 $\{x_i\}_{i=0}^{\infty}$ 生成的问题 (3.2.8) 的 localization set

S_k 非空，由于 $x^* \in S_k$

令 $v_i = v_f(x^*;x_i) \geq 0$，$v_k^* = \min\limits_{0\leq i\leq k} v_i$



故

$$v_k^* = \max\{r : \langle g(x_i), x_i - x\rangle \geq 0, i=0,\cdots,k, \forall x \in B_2(x^*, r)\}$$

<u>lemma 3.2.2</u> 令 $f_k^* = \min\limits_{0\leq i\leq k} f(x_i)$，则

$$f_k^* - f^* \leq \omega_f(x^*; v_k^*)$$

证明：

$$\omega_f(x^*, v_k^*) = \min_{0\leq i\leq k} \omega_f(x^*, v_i) \geq \min_{0\leq i\leq k}[f(x_i) - f^*]$$

$$= f_k^* - f^*$$

$\square$

§ 3.2.3 The Subgradient Method.

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

f 是 $\mathbb{R}^n$ 上的凸 func，Q 是 simple 闭凸集

Subgradient Method for Simple Set

0. 取 $x_0 \in Q$，序列 $\{h_k\}_{k=0}^{\infty}$

$$h_k > 0 \,,\, h_k \to 0 \,,\, \sum_{k=0}^{\infty} h_k = +\infty$$

1. kth iteration $(k \geq 0)$

计算 $f(x_k)$，$g(x_k)$，令 $x_{k+1} = \pi_Q\left(x_k - h_k \frac{g(x_k)}{\|g(x_k)\|}\right)$

<u>Lemma 3.2.2</u> 设 f 是 $B_2(x^*, R)$ 上的 M-Lipschiz func，

其中 $R \geq \|x_0 - x^*\|$，则

$$f_k^* - f^* \leq M \frac{R^2 + \sum_{i=0}^{k} h_i^2}{\sum_{i=0}^{k} h_i}$$

证明：令 $r_i = \|x_i - x^*\|$，则由 <u>lemma 2.2.8</u>

$$r_{i+1}^2 = \left\|\pi_Q\left(x_i - h_i \frac{g(x_i)}{\|g(x_i)\|}\right) - x^*\right\|^2$$

$$\leq \left\|x_i - h_i \frac{g(x_i)}{\|g(x_i)\|} - x^*\right\|^2$$

$$= r_i^2 - 2h_i v_i + h_i^2$$

则 $$r_0^2 + \sum_{i=0}^{k} h_i^2 \geqslant 2\sum_{i=0}^{k} h_i v_i + r_{k+1}^2 \geqslant 2 v_k^* \sum_{i=0}^{k} h_i$$

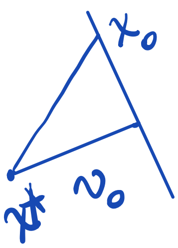
$$\Rightarrow \quad v_k^* \leqslant \frac{R^2 + \sum_{i=0}^{k} h_i^2}{2\sum_{i=0}^{k} h_i}$$

故由<u>lemma 3.2.2</u>

$$f_k^* - f^* \leqslant \omega_f(x^*; v_k^*)$$

$$= \max_{x}\{ f(x) - f^* : \|x - x^*\| \leqslant v_k^* \}$$

(由 $v_k^* \leqslant R$) $$\leqslant \max_{x}\{ M\|x - x^*\| : \|x - x^*\| \leqslant v_k^* \}$$



$$\leqslant M v_k^* \leqslant M \frac{R^2 + \sum_{i=0}^{k} h_i^2}{2\sum_{i=0}^{k} h_i}$$ $\square$

注: $$\lim_{k\to\infty} \frac{R^2 + \sum_{i=0}^{k} h_i^2}{2\sum_{i=0}^{k} h_i} = \lim_{k\to\infty} \frac{\sum_{i=0}^{k} h_i^2}{\sum_{i=0}^{k} h_i} \overset{\text{stolz}}{=\!=} \lim_{k\to\infty} \frac{h_k^2}{h_k} = 0$$

§ 3.2.3

设考虑一个 N 步的 Subgradient Method，则

$$\min_{\{h_k\}} \Delta_N = \frac{R^2+\sum_{i=0}^{N} h_i^2}{2\sum_{i=0}^{N} h_i}$$

由于 $\{h_k\}_{k=0}^{N}$ 是对称的，故 $h_0=\cdots=h_N=h$ 时取 min

易证，$h=\frac{R}{\sqrt{N+1}}$，此时 $\Delta_N=\frac{R}{\sqrt{N+1}}$，故：

$$f_N^*-f^* \leq \frac{MR}{\sqrt{N+1}}$$

另一种定义步长的方法是利用精度 $\varepsilon>0$，即

$$\frac{MR}{\sqrt{N+1}}=\varepsilon \Rightarrow N+1=\frac{M^2R^2}{\varepsilon^2}$$

故 $h_i=\frac{\varepsilon}{M}$，$i\geq 0$ （好处是 h_i 的定义不依赖 R, N）

此时 $f_N^*-f^* \leq \frac{MR^2}{2\varepsilon N}+\frac{1}{2}\varepsilon$

得到 ε-solution 需要 $N \geq \frac{M^2R^2}{\varepsilon^2}$ 步

故次梯度方法 (3.2.14), (3.2.16) 对 (3.2.13) 是 optimal 的.

(关于n是一致的)

若取 $h_i = \frac{r}{\sqrt{i+1}}$, $i=0,\cdots$, 则

$$\Delta_k \sim \frac{R^2 + r^2 \ln(k+1)}{4r\sqrt{k+1}} \quad \text{suboptimal}$$

§3.2.4 Minimization with Functional Constraints

考虑问题:

$$\min_{x \in Q} \{ f(x) : f_j(x) \leq 0, \ j=1,\cdots,m \} \quad (3.2.22)$$

f 和 f_j 是闭凸的, Q 是 simple 闭凸集

定义聚合约束 $\bar{f}(x) = \max_{1 \leq j \leq m} f_j(x)$, 则 (3.2.22) $\Leftrightarrow$

$$\min_{x \in Q} \{ f(x) : \bar{f}(x) \leq 0 \}$$

由 lemma 3.1.13，若 ∂f_j 可求，则 $\bar{g}(x)\in\partial\bar{f}(x)$ 易得

Subgradient Method with Functional Constraints

0. 取初始点 $x_0\in Q$

1. kth iteration $(k\geq 0)$

   (a) 计算 $f(x_k)$, $g(x_k)\in\partial f(x_k)$, $\bar{f}(x_k)$, $\bar{g}(x_k)\in\partial\bar{f}(x_k)$

   (b) 若 $\bar{f}(x_k)\leq\varepsilon$，则

$$x_{k+1}=\pi_Q\left(x_k-\frac{\varepsilon}{\|g(x_k)\|^2}g(x_k)\right) \quad \text{(case A)}$$

否则，令 $$x_{k+1}=\pi_Q\left(x_k-\frac{\bar{f}(x_k)}{\|\bar{g}(x_k)\|^2}\bar{g}(x_k)\right) \quad \text{(case B)}$$

定义 $\mathcal{T}_A(N)$ 是 type A 型迭代，$\mathcal{T}_B(N)$ 是 type B 型，故

$$\bar{f}(x_k)\leq\varepsilon,\ \forall k\in\mathcal{T}_A(N)$$

Thm 3.2.3 设 $f, f_j, j=1,\cdots,m$ 在 $B_2(x^*, \|x_0-x^*\|)$ 上 M-Lipschiz，若 N 充分大，s.t.

$$N \geqslant \frac{M^2}{\varepsilon^2}\|x_0-x^*\|^2 \qquad (3.2.26)$$

则 $\mathcal{F}_A(N)\neq\phi$，且

$$f_N^* \doteq \min_k\{f(x_k): k\in\mathcal{F}_A(N)\} \leqslant f(x^*)+\varepsilon$$

证明：定义 $r_k=\|x_k-x^*\|$，设 N 满足 (3.2.26)，但

$$f(x_k)-f^* \geqslant \varepsilon, \ \forall k\in\mathcal{F}_A(N)$$

若 $k\in\mathcal{F}_A(N)$，则

$$r_{k+1}^2 = \|x_{k+1}-x^*\|^2$$

$$\overset{(2.2.49)}{\leqslant} \left\|x_k-\frac{\varepsilon}{\|g(x_k)\|^2}g(x_k)-x^*\right\|^2$$

$$= r_k^2-\frac{2\varepsilon}{\|g(x_k)\|^2}\langle g(x_k), x_k-x^*\rangle+\frac{\varepsilon^2}{\|g(x_k)\|^2}$$

$$\overset{(3.1.23)}{\leq} r_k^2 - \frac{2\varepsilon}{\|g(x_k)\|^2}(f(x_k) - f^*) + \frac{\varepsilon^2}{\|g(x_k)\|^2}$$

$$\overset{(3.2.28)}{\leq} r_k^2 - \frac{\varepsilon^2}{\|g(x_k)\|^2}$$

Case B情形下：

$$r_{k+1}^2 = \| x_{k+1} - x^* \|^2$$

$$\overset{(2.2.49)}{\leq} \| x_k - \frac{\bar{f}(x_k)}{\|\bar{g}(x_k)\|^2} \bar{g}(x_k) - x^* \|^2$$

$$= r_k^2 - \frac{2\bar{f}(x_k)}{\|\bar{g}(x_k)\|} \langle \bar{g}(x_k), x_k - x^* \rangle + \frac{\bar{f}(x_k)^2}{\|\bar{g}(x_k)\|^2}$$

$$\leq r_k^2 - \frac{2\bar{f}(x_k)}{\|\bar{g}(x_k)\|^2}(\bar{f}(x_k) - \bar{f}(x^*)) + \frac{\bar{f}(x_k)^2}{\|\bar{g}(x_k)\|^2}$$

$$= r_k^2 - \frac{\bar{f}(x_k)^2}{\|\bar{g}(x_k)\|^2}$$

故case A，B都有：$r_{k+1} \leq r_k \leq \| x_0 - x^* \|$

由 f 在 $B_2(x^*, \| x_0 - x^* \|)$ 是 M-Lipschiz 的，由

Thm 3.1.15 $\|g(x_k)\| \le M,\ k\in \mathcal{F}_A(N)$

由 f_j 在 $B_2(x^*, \|x_0-x^*\|)$ 是 M-Lipschiz 的 故 于是 M-Lipschiz的, 故 $\|\bar{g}(x_k)\| \le M,\ k\in\mathcal{F}_B(N)$

故 $r_{k+1}^2 \le r_k^2 - \frac{\varepsilon^2}{M^2}$ 对 $\forall k=0,\cdots,N$

$$\Rightarrow \quad 0 \le r_{N+1}^2 \le r_0^2 - \frac{\varepsilon^2}{M^2}(N+1)$$

与 (3.2.26) 矛盾 $\square$

无约束优化问题并不比有约束问题简单

§3.2.5 Approximating the optimal Lagrange Multipliers

对 $\varepsilon>0$, 定义

$$\mathcal{F}(\varepsilon) = \{x\in Q : f_j(x)\le\varepsilon,\ j=1,\cdots,m\}$$

定义 Lagrangian:

$$\mathcal{L}(x,\lambda)=f(x)+\sum_{j=1}^{m}\lambda^{(j)}f_j(x)\ ,\ x\in\otimes\ ,\ \lambda=(\lambda^{(1)},\cdots,\lambda^{(m)})\in\mathbb{R}_+^m$$

定义 Lagrangian dual 问题：

$$\phi^*=\sup_{\lambda\in\mathbb{R}_+^m}\phi(\lambda)$$

其中 $\phi(\lambda)=\min\limits_{x\in\otimes}\mathcal{L}(x,\lambda)$，显然 $f^*\geqslant\phi^*$

Subgradient Method for Lagrange Multipliers

0. 取初始点 $x_0\in\otimes$

1. kth iteration ($k\geqslant 0$)

(a) 定义 $\mathcal{F}_k=\{j: f_j(x_k)>h\|g_j(x_k)\|\}$

(b) 若 $\mathcal{F}_k=\phi$，则 $x_{k+1}=\pi_\otimes\left(x_k-\dfrac{h\,g(x_k)}{\|g(x_k)\|}\right)$

(c) 若 $\mathcal{F}_k\neq\phi$，则取 $j_k\in\mathcal{F}_k$，定义

$$h_k=\frac{f_{j_k}(x_k)}{\|g_{j_k}(x_k)\|^2}\ ,\quad x_{k+1}=\pi_\otimes(x_k-h_k\,g_{j_k}(x_k))$$

t次迭代后，定义 $A_0(t)=\{k\in\{0,\cdots,t\}: \mathcal{F}_k=\phi\}$

令 $A_j(t)=\{k\in\{0,\cdots,t\}: j_k=j\}$，$1\le j\le m$

令 $N(t)=|A_0(t)|$，若 $N(t)>0$，可以定义估计对偶乘子：

$$\sigma_t=h\sum_{k\in A_0(t)}\frac{1}{\|g(x_k)\|}, \quad \lambda_t^{(j)}=\frac{1}{\sigma_t}\sum_{k\in A_j(t)}h_k, \quad j=1,\cdots,m$$

令 $S_t=\sum_{k\in A_0(t)}\frac{1}{\|g(x_k)\|}$，若 $N(t)=0$，定义 $S_t=0$，故

$$\sigma_t=hS_t$$

原理来自证明3

$\delta_t\le Mh\doteq\varepsilon$

定义 gap function

$$\delta_t=\frac{1}{S_t}\sum_{k\in A_0(t)}\frac{f(x_k)}{\|g(x_k)\|}-\phi(\lambda_t)$$

Thm3.2.4 设 $\mathbb{Q}$ 是有界的：$\|x-x_0\|\le R$，$\forall x\in\mathbb{Q}$

若迭代次数t充分大，s.t. $t>\frac{R^2}{h^2}$，

则 $N(t)>0$，且

$$\max_{1\le j\le m} f_j(x_k)\le Mh,\quad k\in A_0(t)$$

$$\delta_t\le Mh$$

其中 $M=\max_{0\le k\le t}\max_{0\le j\le m}\|g_j(x_k)\|$

证明：注意到

$$\sigma_t\cdot\delta_t=\max_{x\in\otimes}\left\{\sum_{k\in A_0(t)}\frac{hf(x_k)}{\|g(x_k)\|}-\sigma_t f(x)-\underbrace{\sum_{j=1}^{m}\sum_{k\in A_j(t)}h_k f_j(x)}\right\}$$

$$=\max_{x\in\otimes}\left\{\sum_{k\in A_0(t)}\frac{h(f(x_k)-f(x))}{\|g(x_k)\|}-\underbrace{\sum_{k\notin A_0(t)}h_k f_{j_k}(x)}\right\}$$

$$\sum_{k\in A_1(t)}h_k\underbrace{f_1}_{=f_{j_k}}(x)+\cdots+\sum_{k\in A_m(t)}h_k\underbrace{f_m}_{f_{j_k}}(x)=\sum_{k\notin A_0(t)}h_k f_{j_k}(x)$$

$$\le\max_{x\in\otimes}\left\{\sum_{k\in A_0(t)}\frac{h\langle g(x_k),x_k-x\rangle}{\|g(x_k)\|}+\sum_{k\notin A_0(t)}h_k\left[\langle g_{j_k}(x_k),x_k-x\rangle-f_{j_k}(x_k)\right]\right\}$$

对 $\forall x\in\mathcal{Q}$，令 $r_k(x)=\|x-x_k\|$，设 $k\in A_0(t)$，则

$$r_{k+1}^2(x)\le\left\|x_k-x-\frac{hg(x_k)}{\|g(x_k)\|}\right\|^2$$

$$=r_k^2(x)-\frac{2h}{\|g(x_k)\|}\langle g(x_k),x_k-x\rangle+h^2$$

若 $k\notin A_0(t)$，则

$$r_{k+1}^2(x)\le\|x_k-x-h_k g_{j_k}(x_k)\|^2$$

$$=r_k^2(x)-2h_k\langle g_{j_k}(x_k),x_k-x\rangle+h_k^2\|g_{j_k}(x_k)\|^2$$

故

$$2h_k[\langle g_{j_k}(x_k),x_k-x\rangle-f_{j_k}(x_k)]$$

$$\le r_k^2(x)-r_{k+1}^2(x)-\frac{f_{j_k}^2(x_k)}{\|g_{j_k}(x_k)\|^2}$$

$$\le r_k^2(x)-r_{k+1}^2(x)-h^2$$

故 $\sigma_t \delta_t \le \frac{1}{2} r_0^2(x) + \frac{1}{2} N(t) h^2 - \frac{1}{2}(t - N(t)) h^2$

$$= \frac{1}{2} r_0^2(x) - \frac{1}{2} t h^2 + N(t) h^2$$

$$\le \frac{1}{2} R^2 - \frac{1}{2} t h^2 + N(t) h^2$$

由于 $t > \frac{R^2}{h^2}$，则若 $N(t)=0$，则有 $\sigma_t \delta_t < 0$

而 $\sigma_t \ge 0$，$\delta_t \ge f(\hat{x}_k) - \phi(\lambda_t) \ge 0$ 矛盾，其中

$$\hat{x}_k = \frac{1}{\sigma_t} \sum_{k \in A_0(t)} \frac{x_k}{\|g(x_k)\|}$$

故 $N(t) > 0$，又由 $k \in A_0(t)$，则

$$f_j(x_k) \le h \|g_j(x_k)\| \le hM \quad 对\ \forall j = 1, \cdots, m$$

故 $\max\limits_{1 \le j \le m} f_j(x_k) \le hM$，$k \in A_0(t)$

由 $\sigma_t \ge h \frac{N(t)}{M}$，故由 $t > \frac{R^2}{h^2}$

$$\sigma_t \delta_t \le N(t) h^2$$

$\Rightarrow \quad \delta_t \leq Mh$ $\quad\quad \boxtimes$

§ 3.2.6 Strongly Convex Functions

Def 3.2.2 f在凸集$\otimes$上是强凸的，若$\exists \mu>0$, s.t. 对$\forall x, y\in\otimes$, $\alpha\in[0,1]$，有

$$f(\alpha x+(1-\alpha)y)\leqslant \alpha f(x)+(1-\alpha)f(y)-\frac{1}{2}\mu\alpha(1-\alpha)\|x-y\|^2$$

Lemma 3.2.3 令$f\in \mathcal{F}_{\mu}^{0}(\otimes)$，则对$\forall x\in \operatorname{int}\otimes$, $y\in\otimes$，有

$$f(y)\geqslant f(x)+f'(x;y-x)+\frac{1}{2}\mu\|y-x\|^2$$

证明：

$$f(y)\geqslant \frac{1}{\alpha}\left[f((1-\alpha)x+\alpha y)-(1-\alpha)f(x)+\frac{1}{2}\mu\alpha(1-\alpha)\|x-y\|^2\right]$$

$$=f(x)+\frac{1}{\alpha}\left[f(x+\alpha(y-x))-f(x)\right]+\frac{1}{2}\mu(1-\alpha)\|x-y\|^2$$

令$\alpha\downarrow 0$，得到(3.2.38) $\square$

Corollary 3.2.1 令 $f \in \mathcal{S}_\mu^0(\mathcal{Q})$，则对 $\forall g \in \partial f(x)$，有

$$f(y) \geq f(x) + \langle g, y-x \rangle + \frac{1}{2}\mu \|y-x\|^2$$

证明：由 **Thm 3.1.17**，对 $\forall g \in \partial f(x)$，有

$$f'(x; y-x) \geq \langle g, y-x \rangle$$

$\boxtimes$

Corollary 3.2.2 若问题 (3.2.13) 的目标 func

$f \in \mathcal{S}_\mu^0(\mathcal{Q})$，则下水平集有界，且最优解存在

证明：由 f 是 $\mathbb{R}^n$ 上的凸 func，故 f 在 $\mathbb{R}^n$ 上连续，

对 $\forall c \in \mathbb{R}$，$\forall x, x' \in \mathcal{Q}$

$$c \geq f(x) \geq f(x') + \langle g(x'), x-x' \rangle + \frac{\mu}{2}\|x-x'\|^2$$

$$= f(x') + \frac{\mu}{2}\|x-x' + \frac{1}{\mu}g(x')\|^2 - \frac{1}{2\mu}\|g(x')\|^2$$

$$\Rightarrow \frac{\mu}{2}\|x-x'+\frac{1}{\mu}g(x')\|^2 \le C+\frac{1}{2\mu}\|g(x')\|^2-f(x')$$

故下水平集有界，由 Thm3.1.4 4) 知最优解存在

$\square$

Corollary 3.2.3 令 $x^* \in \mathrm{int\,dom} f$ 是 (3.2.13) 的最优解，$f \in \mathcal{S}_\mu^0(\mathcal{Q})$，则对 $\forall x \in \mathcal{Q}$，有

$$f(x) \ge f^* + \frac{1}{2}\mu\|x-x^*\|^2$$

故，最优解存在唯一

证明：由 Thm3.1.24，$\exists\, g^* \in \partial f(x^*)$, s.t.

$$\langle g^*, x-x^* \rangle \ge 0$$

故由 (3.2.39) 知，结论成立

$\square$

强凸func的性质

1. Addition $f_1 \in \mathcal{F}^0_{\mu_1}(\otimes)$, $f_2 \in \mathcal{F}^0_{\mu_2}(\otimes)$, 则对 $\forall \alpha_1, \alpha_2 \geq 0$,

$$\alpha_1 f_1 + \alpha_2 f_2 \in \mathcal{F}^0_{\alpha_1\mu_1+\alpha_2\mu_2}(\otimes)$$

证明是显然的

2. Maximum $f_1 \in \mathcal{F}^0_{\mu_1}(\otimes)$, $f_2 \in \mathcal{F}^0_{\mu_2}(\otimes)$, 则

$$f(x) = \max\{f_1(x), f_2(x)\} \in \mathcal{F}^0_{\mu}(\otimes)$$

其中 $\mu = \min\{\mu_1, \mu_2\}$

证明：对 $\forall x_1, x_2 \in \otimes$, $\alpha \in [0,1]$, 有

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \max\{\alpha f_1(x_1) + (1-\alpha) f_1(x_2)$$

$$- \frac{1}{2}\mu_1\alpha(1-\alpha)\|x_1 - x_2\|^2, \alpha f_2(x_1) + (1-\alpha) f_2(x_2)$$

$$- \frac{1}{2}\mu_2\alpha(1-\alpha)\|x_1 - x_2\|^2\}$$

$$\leq \alpha f(x_1) + (1-\alpha) f(x_2) - \frac{1}{2}\mu\alpha(1-\alpha)\|x_1 - x_2\|^2$$

$\boxed{/}$

3. Subtraction $f \in \mathcal{F}^0_\mu(\mathcal{Q})$，则 $\hat{f}(x) = f(x) - \frac{1}{2}\mu\|x\|^2$

是凸的

证明：对 $\forall x, y \in \mathcal{Q}$，$\alpha \in [0,1]$

$$\hat{f}(\alpha x + (1-\alpha)y) = f(\alpha x + (1-\alpha)y) - \frac{1}{2}\mu\|\alpha x + (1-\alpha)y\|^2$$

$$\leq \alpha f(x) + (1-\alpha) f(y) - \frac{\mu}{2}\alpha(1-\alpha)\|x-y\|^2 - \frac{1}{2}\mu\|\alpha x + (1-\alpha)y\|^2$$

$$= \alpha f(x) + (1-\alpha) f(y) - \frac{1}{2}\alpha\mu\|x\|^2 - \frac{1}{2}(1-\alpha)\mu\|y\|^2$$

$$= \alpha\hat{f}(x) + (1-\alpha)\hat{f}(y)$$

$\boxed{/}$

仍然记 $f_k(x) = \gamma \max\limits_{1\leq i\leq k} x^{(i)} + \frac{\mu}{2}\|x\|^2$，$k = 1, \cdots, n$

在问题 (3.2.2) 的基础上加上下面的假设：

- f 在 $B_2(x^*, \|x_0 - x^*\|)$ 上是 M-Lipschiz 的
- $f \in \mathcal{S}_\mu^0(B_2(x^*, \|x_0 - x^*\|))$，$\mu > 0$

记该问题类是 $\mathcal{P}_s(x_0, \mu, M)$

Thm 3.2.5 对任意类 $\mathcal{P}_s(x_0, \mu, M)$，任意 k，$0 \le k \le n-1$，

$\exists f \in \mathcal{P}_s(x_0, \mu, M)$，s.t.

$$f(x_k) - f^* \geq \frac{M^2}{2\mu(2+\sqrt{k+1})^2}$$

对任意生成序列 $\{x_k\}$ 成立，其中

$$x_k \in x_0 + \mathrm{Lin}\{g(x_0), \cdots, g(x_{k-1})\}$$

证明：应用 func (3.2.2) 以及 Oracle (3.2.6)

W.L.O.G.，令 $x_0 = 0$，取 $f(x) = f_{k+1}(x)$，$\gamma = \dfrac{M\sqrt{k+1}}{2+\sqrt{k+1}}$

由 (3.2.41)，$f_{k+1} \in \mathcal{S}_\mu^0(\mathbb{R}^n)$，同时

$$R_{k+1} \doteq \| x_0 - x_{k+1}^* \| = \frac{r}{\mu\sqrt{k+1}} = \frac{M}{\mu(2+\sqrt{k+1})}$$

由(3.2.4)，f_{k+1} 在 $B_2(x_{k+1}^*, R_{k+1})$ 的 Lipschiz constant 是：

$$2\mu R_{k+1} + r = \frac{2M}{2+\sqrt{k+1}} + \frac{M\sqrt{k+1}}{2+\sqrt{k+1}} = M$$

故 $f = f_{k+1} \in \mathcal{P}_s(x_0, \mu, M)$，故

$$f(x_k) - f^* \geq -f_{k+1}^* = \frac{r^2}{2\mu(k+1)} = \frac{M^2}{2\mu(2+\sqrt{k+1})^2}$$

$\boxtimes$

<u>Thm 3.2.6</u> 设问题(3.2.13)中的 f 满足(3.2.42)，

令 $\varepsilon > 0$ 是优化问题的精度，考虑序列 $\{x_k\} \subset Q$ 由

$$x_{k+1} = \pi_Q\left(x_k - \frac{2\varepsilon g(x_k)}{\| g(x_k) \|^2} \right), \quad k \geq 0$$

生成，其中 $g(x_k) \in \partial f(x_k)$. 则若 N 足够大，s.t.

$$N \geq \frac{M^2}{\mu\varepsilon} \ln \frac{M\|x_0 - x^*\|}{\varepsilon}$$

有 $f_N^* = \min\limits_{0 \leq k \leq N} f(x_k) \leq f^* + \varepsilon$

证明：令 $r_k = \|x_k - x^*\|$，$h_k = \frac{2\varepsilon}{\|g(x_k)\|^2}$，设 N 满足

(3.2.46)，且 $f(x_k) - f^* > \varepsilon$，$k = 0, \cdots, N$

则
$$\begin{aligned}
r_{k+1}^2 &\leq \|x_k - h_k g(x_k)\|^2 \\
&= r_k^2 - 2h_k \langle g(x_k), x_k - x^* \rangle + \frac{4\varepsilon^2}{\|g(x_k)\|^2} \\
&\leq r_k^2 - \frac{4\varepsilon}{\|g(x_k)\|^2}\left[f(x_k) - f^* + \frac{1}{2}\mu r_k^2 \right] + \frac{4\varepsilon^2}{\|g(x_k)\|^2} \\
&\overset{(3.2.47)}{\leq} \left(1 - \frac{2\mu\varepsilon}{\|g(x_k)\|^2}\right) r_k^2
\end{aligned}$$

故 $x_k \in B(x^*, r_0)$，故 $\|g(x_k)\| \leq M$，从而

$$\varepsilon < f(x_N) - f^* \leq M r_N \leq M\left(1 - \frac{2\mu\varepsilon}{M^2}\right)^{\frac{N}{2}} r_0$$

$$\leq M\exp\left\{-\frac{\mu\varepsilon N}{M^2}\right\} r_0$$

与 (3.2.46)矛盾 $\boxed{2}$

§ 3.2.7 Complexity Bounds in Finite Dimension

本节研究 feasibility问题：

$$\text{Find } x^* \in S \qquad (3.2.49)$$

其中 S 是闭凸集，假设有下面的 seperation oracle：

或者报告 $\bar{x} \in S$

或者返回 $\bar{g}$，分离 $\bar{x}$ 与 S：$\langle \bar{g}, \bar{x} - x \rangle \geq 0 \quad \forall x \in S$

Assumption 3.2.1 $\exists x^* \in S$, s.t. $\exists \varepsilon > 0$, s.t. $B(x^*, \varepsilon) \subseteq S$

例如：已知 (3.2.8) 的最优值 f^*，则考虑 feasibility 问题，其中：

$$S = \{(t, x) \in \mathbb{R}^{n+1} \mid t \geq f(x), t \leq f^* + \varepsilon, x \in Q\}$$

为什么不能考虑 $S = \{x \in \mathbb{R}^n \mid f(x) \leq f^* + \varepsilon\}$ ？

下面考虑(3.2.49)的 resisting oracle：

序列 $\{B_k\}_{k=0}^{\infty}$，$B_{k+1} \subset B_k$，其中

$$B_k = \{x \in \mathbb{R}^n \mid a_k \leq x \leq b_k\}$$

对每个 box B_k，$k \geq 0$，记中心 $c_k = \frac{1}{2}(a_k + b_k)$

对每个box B_k，$k \geq 1$，oracle 生成各自的分离向量 g_k

g_k 除去符号的选取，始终是坐标向量

定义： m 是生成 box 的数量

i 是积极坐标

定义 $\bar{e}_n \in \mathbb{R}^n$ 是全1向量，oracle 从下面的初值开始：

$$a_0 = -R\bar{e}_n, \quad b_0 = R\bar{e}_n, \quad m = 0, \quad i = 1$$

输入任意 test point $x \in \mathbb{R}^n$

Resisting oracle for feasibity problem

if $x \notin B_0$, then return a separator of x from B_0

else:

1. 寻找最大的 $k \in [0, \cdots, m]$: $x \in B_k$

2. If $k < m$ Return g_k

   else : Create a new box :

   If $x^{(i)} \geq c_m^{(i)}$, then $a_{m+1} = a_m$

   $b_{m+1} = b_m + (c_m^{(i)} - b_m^{(i)}) e_i$, $g_m = e_i$

   else $a_{m+1} = a_m + (c_m^{(i)} - a_m^{(i)}) e_i$

   $b_{m+1} = b_m$, $g_m = -e_i$

   $m := m+1$, $i := i+1$; If $i > n$, then $i := 1$

   Return g_m

这样 box 的构造方法有如下两个重要性质

- $\mathrm{Vol}_n B_{k+1} = \frac{1}{2}\mathrm{Vol}_n B_k$
- 对 $\forall k \geq 0$，有 $b_{k+n} - a_{k+n} = \frac{1}{2}(b_k - a_k)$

<u>Lemma 3.2.4</u> 对 $\forall k \geq 0$，有：

$$B_2(c_k, r_k) \subset B_k, \quad r_k = \frac{R}{2}\left(\frac{1}{2}\right)^{\frac{k}{n}}$$

证明：对 $\forall k \in [0, \cdots, n-1]$，有：

$$B_k \supset B_n = \left\{ x \,\middle|\, c_n - \frac{1}{2}R\bar{e}_n \leq x \leq c_n + \frac{1}{2}R\bar{e}_n \right\} \supset B_2(c_n, \frac{1}{2}R)$$

故，对这样的 k，有：

$$B_k \supset B_2(c_k, \frac{1}{2}R) \supset B_2(c_k, r_k)$$

令 $k = nl + p$，其中 $p \in [0, \cdots, n-1]$：由

$$b_k - a_k = \left(\frac{1}{2}\right)^l (b_p - a_p)$$

故 $B_k \supset B_2(C_k, (\frac{1}{2})^k \cdot \frac{1}{2}R)$，又由 $r_k \le \frac{R}{2}(\frac{1}{2})^k$，即证 $\boxed{/\!/}$

Thm 3.2.7 考虑一类 feasibility 问题 (3.2.49)，满足

Assumption 3.2.1，S 是 $B_\infty(0,R)$ 的子集，则 lower analytical complexity 是 $n\ln\frac{R}{2\varepsilon}$ 次调用 separation oracle

证明：

$$x_0 \xrightarrow{\text{Separation}} g_0 \xrightarrow{f(x_0,g_0)} x_1 \longrightarrow g_1 \to x_2 \to \cdots$$

黑盒

在考虑的问题是：得到 $x_k \in S$ 的步数（即调用 Separation Oracle 的次数）的 lower analytical complexity：

即定义 $F(f, S, x_0)$ 是用黑盒 f，对集合 S 所需步数（x_0 为初始点。），求：

$$\inf_{f\in\mathcal{F}} \sup_{\substack{S\in\mathcal{S} \\ x_0\in B_\infty(0,R)}} F(f,S,x_0) \text{ 的下界}$$

而固定 $f\in\mathcal{F}$，$x_0\in B_\infty(0,R)$ 断言：

$$\sup_{S\in\mathcal{S}} F(f,S,x_0) \overset{①}{\geq} F(f,B_{k^*}(f,x_0),x_0) \overset{②}{\geq} n\log_2\frac{R}{2\varepsilon}$$

其中 $k^*=\lfloor n\log_2\frac{R}{2\varepsilon}\rfloor$，$B_{k^*}$ 是 resisting oracle 生成的第 k^* 个 box（以 x_0 为初值，$x_{k+1}=f(x_k,g_k)$）。

先证明①：由 $\left(\frac{R}{2}\right)\left(\frac{1}{2}\right)^{\frac{k}{n}}\geq\varepsilon \Rightarrow k\leq n\log_2\frac{R}{2\varepsilon}$

故 $k^*\leq k$，知 $B_{k^*}\in\mathcal{S}$（即 $\exists x^*\in B_{k^*}$，s.t. $B_2(x^*,\varepsilon)\subseteq S$）。

从而①成立

再证明②：调用 Resisting oracle 最好的情形

用 k^* 次调用得到 $\{x_k\}_{k=0}^{k^*}$，$\{g_k\}_{k=0}^{k^*}$，$\{B_k\}_{k=0}^{k^*}$

由于 $x_k\notin B_{k^*}$，$\forall k=0,\dots,k^*-1$

故②成立，从而得证 $\inf_{f\in\mathcal{F}}\sup_{S\in\mathcal{S}} F(f,S)\geq n\log_2\frac{R}{2\varepsilon}$

$\boxtimes$