大学教材

优化选讲

董云达著

内容简介

本书是作者在多年教学讲义的基础上编写而成的。它包含了优化的基本理论和方法。

首先，介绍了无约束优化方法：最陡下降法，共轭方向法，Newton-Raphson方法，Davidon 变尺度法（也称拟 Newton 法）等基本的迭代方法。然后，又详细地介绍了约束优化的 Karush-Kuhn-Tucker 定理和相应的迭代方法：增广 Lagrange 乘子法，原始-对偶内点法，线性规划的原始对偶内点法，大规模半定规划的 Douglas-Rachford 方法等。书中的带*部分，可以跳过不讲。

本书的特色不求杂而全，只求取材精要，再进行透彻地论述。绝大部分是经典理论，也有少部分的最新结果。

作为大学教材，本书适合于数学专业的高年级本科生、研究生，也是广大教师和科研人员了解优化、研究优化的好帮手。

目录

引理 4.4.2 设 f 在 X 上是自协调的, $x \in intX$, $h^T\nabla^2(x)h < 1$。则有

(i) $x + h \in intX$.

(ii) $f(x+h)$ 的上、下界可以表示为

$$f(x) + \nabla f(x)^T h - \sqrt{h^T\nabla^2 f(x)h} - \ln\left(1 - \sqrt{h^T\nabla^2 f(x)h}\right), \tag{4.9}$$

$$f(x) + \nabla f(x)^T h + \sqrt{h^T\nabla^2 f(x)h} - \ln\left(1 + \sqrt{h^T\nabla^2 f(x)h}\right). \tag{4.10}$$

(iii) 并且对于所有的 $d \in R^n$，总有

$$1 - \sqrt{h^T\nabla^2 f(x)h} \le \frac{\sqrt{d^T\nabla^2 f(x+h)d}}{\sqrt{d^T\nabla^2 f(x)d}} \le \frac{1}{1 - \sqrt{h^T\nabla^2 f(x)h}}.$$

证 分三部分。首先，我们先来证明结论 (ii) 与结论 (iii) 成立，然后，再来证明结论 (i) 也必然成立。

为了证明 (ii), 我们定义 $\phi(t) = h^T\nabla^2 f(x+th)h$。则 $\phi(t)$ 是连续可微的，并且

$$|\phi'(t)| = |\nabla^3 f(x+th)[h,h,h]| \le 2\phi(t)^{3/2}. \tag{4.11}$$

于是，将 $\phi(t)^{-1/2}$ 在 $t = 0$ 处展开，便可以得到

$$\begin{aligned}\phi(t)^{-1/2} &= \phi(0)^{-1/2} + t\Big[\frac{d}{ds}\phi(s)^{-1/2}\Big]_{s=\xi} \qquad 0 < \xi < t \\ &= \phi(0)^{-1/2} - t\frac{1}{2}\phi(\xi)^{-3/2}\phi'(\xi).\end{aligned}$$

结合 (4.11), 进一步可以推导出

$$\phi(0)^{-1/2} - t \le \phi(t)^{-1/2} \le \phi(0)^{-1/2} + t, \ \ \forall t \in [0,1],$$

即

$$\frac{\phi(0)}{\left(1 + t\sqrt{\phi(0)}\right)^2} \le \phi(t) \le \frac{\phi(0)}{\left(1 - t\sqrt{\phi(0)}\right)^2}, \qquad \forall t \in [0,1].$$

记 $r = \sqrt{\phi(0)} = \sqrt{h^T\nabla^2 f(x)h}$。由题设知 $r < 1$, 从而有

$$\frac{r}{(1+rt)^2} \le h^T\nabla^2 f(x+th)h \le \frac{r}{(1-rt)^2}. \tag{4.12}$$

将 (4.12) 中右边不等式积分两次

$$\begin{aligned}f(x+h) &\le f(x) + \nabla f(x)^T h + \int_0^1\int_0^\tau \frac{r^2}{(1-rt)^2}dt\,d\tau \\ &= f(x) + \nabla f(x)^T h - (r + \ln(1-r)) \\ &= f(x) + \nabla f(x)^T h - \sqrt{h^T\nabla^2 f(x)h} - \ln\left(1 - \sqrt{h^T\nabla^2 f(x)h}\right).\end{aligned}$$

(该公式的推导要求 $1+v^Tu$ 和 $1+z^Tw$ 不同时为 0，这儿的假设条件正好使之满足)，则有

$$\det(\tilde{M}) = \det(M)\frac{s^Ty}{s^TMs} = \det(M)\frac{s^Ty}{\|s\|^2}\frac{1}{q}.$$

两边取对数

$$\ln\det(\tilde{M}) = \ln\det(M) - \ln q + \ln\frac{s^Ty}{\|s\|^2}. \tag{5.23}$$

将 (5.22) 减去 (5.23)，并且利用 $\varphi(\cdot) = \mathrm{trace}(\cdot) - \ln\det(\cdot)$，整理可以得到

$$\varphi(\tilde{M}) = \varphi(M) + 1 - \frac{q}{\cos^2\theta} + \ln\frac{q}{\cos^2\theta} + \ln\cos^2\theta + \frac{\|y\|^2}{s^Ty} - \ln\frac{s^Ty}{\|s\|^2} - 1.$$

既然对于所有的 $t>0$ 总有 $1-t+\ln t \le 0$ 以及根据假设条件 (5.18)

$$\frac{\|y\|^2}{s^Ty} - \ln\frac{s^Ty}{\|s\|^2} - 1 \le \bar{\mu} - \ln\mu - 1 \tag{5.24}$$

有一个正的上界 (记作 c)，那么我们就有

$$\varphi(\tilde{M}) \le \varphi(M) + \ln\cos^2\theta + c.$$

重新写出上、下标，则有

$$\varphi(M_{k+1}) \le \varphi(M_k) + \ln\cos^2\theta_k + c \le \varphi(M_0) + \sum_{j=0}^{k}\ln\cos^2\theta_j + (k+1)c.$$

由于当 M 正定时，

$$\varphi(M) = \mathrm{trace}(M) - \ln\det(M) = \sum(\lambda_i - \ln\lambda_i) > 0,$$

其中 λ_i 为 M 的特征值，所以，我们可以推导出

$$k\min_{j=0,\ldots,k-1}\{-\ln\cos^2\theta_j\} \le \sum_{j=0}^{k-1}(-\ln\cos^2\theta_j) < \varphi(M_0) + kc.$$

这个不等式表明，当 $k > \varphi(M_0)$ 时，我们有

$$\min_{j=0,\ldots,k-1}\{-\ln\cos^2\theta_j\} < \varphi(M_0)/k + c < 1 + c.$$

也就是说，当 $k > \varphi(M_0)$ 时，序列 $\{\cos^2\theta_k\}$ 存在一个子列使得 $1/e^{c+1}$ 为其一个正的下界。

第四步，我们证明：整个序列 $\{x^k\}$ 必然收敛到问题的唯一解点。 在第三步中，我们已经说明了 $\{\cos^2\theta_k\}$ 必然存在一个有正下界的子列。结合 (5.19) 可以知道：序列 $\{\|\nabla f(x^k)\|\}$ 相应地存在一个收敛于 0 的子列。记这个收敛于 0 的子列

7.7 Motzkin 定理的补充证明

本节主要介绍如何利用 Tucker 定理证明 Motzkin 定理。

定义 7.7.1 一个集合 $X \subseteq R^n$ 称为锥，当且仅当对于任给的 $x \in X$, $\lambda > 0$，λx 总在 X 中。

下面，我们介绍点与闭凸锥分离定理，它本身具有简洁优美的优点。

定理 7.7.1 *(点与闭凸锥分离定理)* 假设 n 维向量 c 不属于非空闭凸锥 $C \subset R^n$，则存在一个非零 n 维向量 p 使得

$$p^T c < 0 \leq p^T z, \ \ \forall z \in C. \tag{7.35}$$

证 既然集合 C 是一个非空闭凸集，那么必然存在唯一的点 $\bar{c} \in C$，使得

$$\|c - \bar{c}\| = \min_{u \in C} \|c - u\| \ \ \Rightarrow \ \ (c - \bar{c})^T (z - \bar{c}) \leq 0, \ \ \forall z \in C.$$

考虑到 C 又是一个锥，于是对于所有的 $\gamma > 0$，总有 $\gamma z \in C$，从而

$$(c - \bar{c})^T (\gamma z - \bar{c}) \leq 0 \ \ \Rightarrow \ \ \gamma (c - \bar{c})^T z \leq (c - \bar{c})^T c - \|c - \bar{c}\|^2, \ \forall z \in C.$$

令 $p = \bar{c} - c$，则 $p \neq 0$。并且，

$$\gamma p^T z \geq p^T c + \|p\|^2 \quad \Leftrightarrow \quad p^T z \geq (p^T c + \|p\|^2)/\gamma.$$

对于左边的不等式，令 $\gamma \to 0^+$ 便可以推导出 $p^T c \leq -\|p\|^2 < 0$；而对于右边的不等式，我们再令 $\gamma \to +\infty$ 就得到了 (7.35) 右边的关系式。 □

定理 7.7.2 *(Tucker 定理)* 设 $D \in R^{m \times n}$，则不等式组

$$Dx \geq 0, \ D^T y = 0, \ y \geq 0$$

总存在一组解 $(\bar{x}, \bar{y})$ 使得 $D\bar{x} + \bar{y} > 0$。

证 记 $D^T = (d^1, d^2, \cdots, d^m)$。构造相应的集合

$$X_1 = \left\{ \sum_{i \neq 1} - y_i d^i : y_i \geq 0, i \neq 1 \right\}. \tag{7.36}$$

分两种情形进行讨论。

情形一 $d^1 \in X_1$。此时，必然存在 $y_i \geq 0$, $i = 2, ..., m$ 使得

$$d^1 = \sum_{i \neq 1} - y_i d^i.$$

设 $\gamma \in (0, 0.01)$，记 $t = x^T z/n$，则中心路径的无穷模邻域定义如下

$$\mathcal{N}_{\infty}^{-}(\gamma) := \{(x, \lambda, z) \in \mathcal{F}^0 : x_i z_i \geq \gamma t,\ i = 1, \cdots, n\}.$$

显然，当 γ 充分小时，中心路径的无穷模邻域可以看作 $\mathcal{F}^0$ 的一个近似。而且，它还是中心路径的一个大邻域。

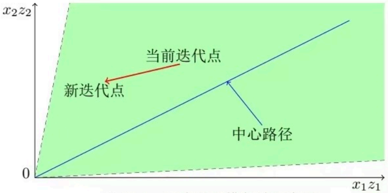

图 10.3：二维无穷模邻域示意图

当 $n = 2$ 时，分别以 x_1z_1, x_2z_2 为横轴与纵轴，图 10.3 显示了该方法是如何顺着（xz 平面上的）中心路径的无穷模邻域进行迭代的。

Salahi 预估校正算法，每次迭代主要分为：预估步和校正步。但它要求迭代序列在中心路径的无穷模邻域内，其中的参数 γ 取为 0.001 或更小。

预估步 已知 $(x,\ \lambda,\ z) \in \mathcal{N}_{\infty}^{-}(\gamma)$ 解下面的线性方程组

$$A^T \Delta\lambda^a + \Delta z^a = 0,\ \ A\Delta x^a = 0, \tag{10.15}$$

$$z_i \Delta x_i^a + x_i \Delta z_i^a = -x_i z_i,\ i = 1, ..., n. \tag{10.16}$$

得到 Δx^a, $\Delta\lambda^a$, Δz^a。

校正步 取 $\mu_{\min}$ 为下边 (10.19) 的较小的正根。解下面的线性方程组

$$A^T \Delta\lambda + \Delta z = 0,\ \ A\Delta x = 0, \tag{10.17}$$

$$z_i \Delta x_i + x_i \Delta z_i = \mu_{\min} - x_i z_i - \Delta x_i^a \Delta z_i^a,\ i = 1, ..., n. \tag{10.18}$$

解之得 Δx, $\Delta\lambda$, Δz。计算最大可行步长 $\alpha_c \in [0, 1]$ 使得

$$x(\alpha_c) = x + \alpha_c \Delta x,\ \ \lambda(\alpha_c) = \lambda + \alpha_c \Delta\lambda,\ \ z(\alpha_c) = z + \alpha_c \Delta z$$

仍在 $\mathcal{N}_{\infty}^{-}(\gamma)$ 内。将 $x(\alpha_c)$, $\lambda(\alpha_c)$, $z(\alpha_c)$ 作为新的迭代点。

为了分析 Salahi 预估校正算法的收敛性，我们先做一些准备工作。

记

$$\phi(x,\ z,\ \mu) := \frac{1}{n} \sum_{i=1}^{n} \left(\frac{x_i z_i}{\mu} - \ln \frac{x_i z_i}{\mu} - 1 \right).$$

12.3 半正定锥上的投影

对于大规模的半定规划，内点法往往是无能为力的。因此，人们转而考虑如何将增广 Lagrange 乘子法和 Douglas-Rachford 分裂方法用于大规模的半定规划。为了介绍这方面的工作，我们先了解一下有关概念和半正定锥上的投影及其性质。

令 S^n 和 S_+^n 分别表示所有 n 阶实对称矩阵和正半定矩阵的集合。则 S_+^n 是一个非空闭凸锥，简称半正定锥。对于给定的 $Q \in S^n$, 极小化问题

$$\min \ \|Q - X\|_F, \quad \text{s.t. } X \succeq 0$$

的解，记作 Q_+，存在并且唯一。以后，我们称之为到半正定锥上的投影。

接下来，我们来讨论一个实对称矩阵到半正定锥上的投影的基本性质。

定理 12.3.1 若 $Q = P^T \Lambda P$ 是 $Q \in S^n$ 的谱分解，其中 P 满足 $P^T P = PP^T = I$，则 $Q_+ = P^T \Lambda_+ P$，其中 Λ_+ 是将 Λ 对角线上的负元变为零得到，且 $Q_+ = Q + [-Q]_+$.

证 对于任意给定的 $X \in S_+^n$, 令 $Y = PXP^T$，则

$$\begin{aligned}(Q-X)^T(Q-X) &= (P^T(\Lambda - Y)P)^T(P^T(\Lambda - Y)P)\\ &= P^T(\Lambda - Y)^T PP^T(\Lambda - Y)P\\ &= P^T(\Lambda - Y)^T(\Lambda - Y)P.\end{aligned}$$

从而

$$\begin{aligned}\|Q-X\|_F^2 &= \text{trace}\,(Q-X)^T(Q-X)\\ &= \text{trace}\, P^T(\Lambda - Y)^T(\Lambda - Y)P \qquad (12.5)\\ &= \text{trace}\,(\Lambda - Y)^T(\Lambda - Y). \qquad (12.6)\end{aligned}$$

最后等式源于(12.5)与(12.6)中的两个矩阵有相同的特征根。这样，我们就有

$$\|Q-X\|_F^2 = \|\Lambda - Y\|_F^2 = \textstyle\sum_{j\neq i} y_{ij}^2 + \sum_{i=1}^n (\lambda_i - y_{ii})^2.$$

显然，为了使得右边的和最小，每个 y_{ij} $(j \neq i)$ 都必须为 0。而且，根据 X 的正半定性容易推导出 y_{ii} 的非负性。于是，相应的结论成立。 □

定理 12.3.2 （*Fejer* 定理） 对于任意给定的实对称矩阵 Q，它是正半定的当且仅当对于任给的正半定矩阵 P 总有 $\langle Q,\ P\rangle \geq 0$。

显然，f, g 都是一元凸函数，并且满足 $\text{dom} f \cap \text{dom} g = \{0\} \neq \emptyset$。但

$$\partial(f+g)(0) = R \neq \emptyset = \partial f(0) + \partial g(0).$$

事实上，若 $s \in \partial f(0)$, 则

$$f(y) \geq f(0) + sy \quad \Rightarrow \quad 0 \geq 1 + sy, \, \forall\, y < 0 \quad \Rightarrow \quad s(-y) \geq 1, \, \forall\, y < 0.$$

这意味着: s 是某一个正数，并且对于任给的 $y < 0$, 总有 $s > 1/(-y)$. 但这是不可能的!

值得指出的是，若 f 是若干个线性函数的逐点最大值函数，则有如下定理。

定理 13.8.5 设 $f\colon E \to (-\infty, +\infty]$ 和 $g(x) = \max\{\langle a^i, x\rangle - b_i\colon i = 1, ..., l\}$, 其中 $a^i \in E$, $b_i \in R$, 均为凸函数。若 $\text{ri dom} f \cap \text{dom} g \neq \emptyset$，则

$$\partial(f(x) + g(x)) = \partial f(x) + \partial g(x).$$

定理 13.8.6 设 $f\colon E \to R$ 为凸函数，A 是一个 $m \times n$ 矩阵。则

$$\partial[f(Ax)] = A^T \partial f(Ax) = \{A^T s\colon s \in \partial f(Ax)\}.$$

证 $\Leftarrow$ 任取 $s \in \partial f(Ax)$，则

$$f(Ay) \geq f(Ax) + \langle s, Ay - Ax\rangle = f(Ax) + \langle A^T s, y - x\rangle,$$

即 $A^T s \in \partial[f(Ax)]$。

$\Rightarrow$ 若 $A^T \tilde{s} \notin A^T \partial(f(Ax))$，往证 $A^T \tilde{s} \notin \partial[f(Ax)]$。实际上 $\tilde{s} \notin \partial(f(Ax))$，即存在某个 $y \in E$，使得

$$f(Ay) < f(Ax) + \langle \tilde{s}, Ay - Ax\rangle = f(Ax) + \langle A^T \tilde{s}, y - x\rangle,$$

从而 $A^T \tilde{s} \notin \partial[f(Ax)]$。 □

接下来，我们讨论逐点最大值函数的次微分。为此，先引入有限个集合的凸包的概念。则由一组集合 $S_1, ..., S_l$ 生成的凸包定义为

$$co\{S_1, ..., S_l\} := \lambda_1 S_1 + ... + \lambda_l S_l,$$

其中 $\lambda_1, ..., \lambda_l$ 是一组和为 1 的非负系数。

定理 13.8.7 设 $f_i\colon E \to (-\infty, +\infty]$, $i = 1, ..., l$, 均为闭的真凸函数。记 $f = \max\{f_i\colon i = 1, ..., l\}$. 若 $x \in \text{ri dom} f_i$, $i = 1, ..., l$, 则该逐点最大值函数在 x 处的次微分可以表示为

$$\partial f(x) = co\{\partial f_i(x)\colon i \in I(x)\},$$

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

14.7 一致有界性原则与共鸣定理

在 1876 年，P. du Bois Reymond 构造了一个周期为 2π 的连续函数，使得其 Fourier 级数在给定的点处发散。在总结前人五十年来大量工作的基础上，Banach 和 Steinhaus 通过借鉴 Osgood 定理的证明方法，抽象提取出了共鸣(resonance)定理。它刻画了有界线性算子族的基本性质，从而成为泛函分析中最为重要的研究成果之一。

引理 14.7.1 *设 T 是从 Banach 空间 $\mathcal{B}$ 映射到赋范线性空间 $\mathcal{Y}$ 上的有界线性算子。则对于任意的 $x \in \mathcal{B}$ 和 $r > 0$，有*

$$\sup\{\|Tx'\|\colon x' \in \mathcal{B}(x,r)\} \geq r\|T\|,$$

其中 $\mathcal{B}(x,r) = \{x' \in \mathcal{B}\colon \|x' - x\| \leq r\}$。

证 对于任意的 $\xi \in \mathcal{B}$，我们有

$$\begin{aligned} 2\max\{\|T(x+\xi)\|, \|T(x-\xi)\|\} &\geq \|T(x+\xi)\| + \|T(x-\xi)\| \\ &\geq \|T(x+\xi-(x-\xi))\| \\ &= 2\,\|T\xi\|. \end{aligned} \tag{14.13}$$

由线性算子范数的定义得 $\sup\{\|T(r^{-1}\xi)\|\colon \|r^{-1}\xi\| \leq 1\} \geq \|T\|$ (以二维空间为例，前者在单位圆面上取而后者在单位圆周上取上确界。当然，由于 T 是有界线性的，所以实际上两者的上确界相等)。于是，我们就有

$$\forall\, \xi \in \mathcal{B}(0,r),\ \sup\|T\xi\| = r\sup\|T(r^{-1}\xi)\| \geq r\|T\|.$$

另一方面，因为当 $\xi \in \mathcal{B}(0,r)$ 时，$x \pm \xi \in \mathcal{B}(x,r)$，所以

$$\sup\{\|Tx'\|\colon x' \in \mathcal{B}(x,r)\} \geq \max\{\|T(x+\xi)\|, \|T(x-\xi)\|\}.$$

在（14.13）式两边关于 $\xi \in \mathcal{B}(0,r)$ 取上确界即可。 □

定理 14.7.1 *(一致有界性原则) 设 $\mathcal{F}$ 是一族从 Banach 空间 $\mathcal{B}$ 映射到赋范线性空间 $\mathcal{Y}$ 上的的有界线性算子。若任给的 $T \in \mathcal{F}$ 是逐点有界的，即对于所有的 $x \in \mathcal{B}$，都存在相应的 $\beta_x > 0$ 使得 $\|Tx\| \leq \beta_x$，则必然存在不再依赖于 x 的 $\beta > 0$ 使得*

$$\|T\| \leq \beta, \quad \forall T \in \mathcal{F}.$$

讨论 取 $\mathcal{B}$ 和 $\mathcal{Y}$ 为一维实空间，$T_n(x) = nx$, $n = 1, 2, ...$ 是一族线性算子。试问：满足上述定理的题设条件吗？举例说明：如果 $\mathcal{B}$ 换成不完备的赋范线性空间，那么一致有界性原则不成立。

一致有界性原则的逆否命题为下面的共鸣定理。

记 $w^k := \lambda_{k-1}^{-1}(x^{k-1} - x^k)$。则序列 $\{w^k\}$ 强收敛于 0。同时，由 (15.23) 可知序列 $\{\|x^k - x^*\|\}$ 极限存在，因此序列 $\{x^k\}$ 以范数有界，从而存在某个子列 $\{x^{k_j}\}$ 使得

$$x^{k_j} \rightharpoonup x^\infty, \quad k_j \to +\infty.$$

结合 $w^{k_j} \in A(x^{k_j})$，A 的极大单调性和定理 15.3.2，我们有

$$0 \in A(x^\infty).$$

即弱聚点 $\{x^\infty\}$ 是一个解点。

下证弱聚点的唯一性。若不然，则存在某个子列 $\{x^{k_l}\}$ 使得

$$x^{k_l} \rightharpoonup \hat{x}^\infty \neq x^\infty, \quad k_l \to +\infty.$$

根据刚才的讨论，$\{\|x^k - x^\infty\|\}$ 和 $\{\|x^k - \hat{x}^\infty\|\}$ 极限都存在。记

$$l = \lim_{k\to\infty} \|x^k - x^\infty\|^2, \quad \hat{l} = \lim_{k\to\infty} \|x^k - \hat{x}^\infty\|^2.$$

考虑

$$\begin{aligned}\|x^k - x^\infty\|^2 &= \|x^k - \hat{x}^\infty + \hat{x}^\infty - x^\infty\|^2 \\ &= \|x^k - \hat{x}^\infty\|^2 + 2\langle x^k - \hat{x}^\infty, \hat{x}^\infty - x^\infty\rangle + \|\hat{x}^\infty - x^\infty\|^2.\end{aligned}$$

沿 k_l 取极限，则

$$l = \hat{l} + \|\hat{x}^\infty - x^\infty\|^2.$$

类似地

$$\hat{l} = l + \|x^\infty - \hat{x}^\infty\|^2.$$

两式相加可以得到 $\|x^\infty - \hat{x}^\infty\| = 0$. 与假设 $x_1^\infty \neq x_2^\infty$ 矛盾! □

在邻近点方法中，如何证明弱聚点的唯一性是一个挑战。上述简洁的技巧归功于 Bregman.

注：在 1976 年，Rockafellar 假设：λ_k 有一个正的下界。在 1978 年，Brézis, Lions 将其弱化为 (15.20) 并分析了邻近点方法的 $O(1/k)$ 的收敛率。在 2014-2015 年，Dong 将其收敛率改进为定理 15.4.3 中的 $1/k$ 的高阶无穷小。Lions 系 1994 年 Fields 奖得主。

15.5 附：Yosida 逼近

本附录主要介绍 Euclid 空间中极大单调算子的 Yosida 逼近。

结合 Cauchy-Schwarz 不等式

$$\frac{\|r(x,\alpha')\|^2}{\alpha'}+\frac{\|r(x,\alpha)\|^2}{\alpha}\leq\left(\frac{1}{\alpha'}+\frac{1}{\alpha}\right)\|r(x,\alpha')\|\|r(x,\alpha)\|.$$

若 $\|r(x,\alpha)\|=0$, 则 $\|r(x,\alpha')\|=0$, 结论成立。否则，$\|r(x,\alpha)\|\neq 0$. 从而，我们可以将上述不等式重新写作

$$\frac{\|r(x,\alpha')\|^2}{\|r(x,\alpha)\|^2}-\left(1+\frac{\alpha'}{\alpha}\right)\frac{\|r(x,\alpha')\|}{\|r(x,\alpha)\|}+\frac{\alpha'}{\alpha}\leq 0.$$

既然抛物线 $p(t):=t^2-(1+\alpha'/\alpha)t+\alpha'/\alpha$ 开口向上(这儿 α' 和 α 都是给定的正数) 并且方程 $p(t)=0$ 有两个根 $t=1$, $t=\alpha'/\alpha\geq 1$, 那么，$p(t)\leq 0$ 当且仅当

$$1\leq\frac{\|r(x,\alpha')\|}{\|r(x,\alpha)\|}\leq\frac{\alpha'}{\alpha}.$$

现在我们来证明 (ii). 对于符号 $x(\alpha)$，我们有

$$\frac{x-x(\alpha)}{\alpha}-F(x)\in A(x(\alpha)),\ \ \alpha>0,$$

根据 $w-z\in A(x)$ 和 A的单调性, 可以得到

$$\begin{aligned}0&\leq\langle x(\alpha)-x,\frac{x-x(\alpha)}{\alpha}-F(x)-w+F(x)\rangle\\&=\langle x(\alpha)-x,\frac{x-x(\alpha)}{\alpha}-w\rangle.\end{aligned}$$

通过利用 Cauchy-Schwarz 不等式，进一步得到

$$\frac{1}{\alpha}\|x-x(\alpha)\|^2\leq\langle w,x-x(\alpha)\rangle\leq\|w\|\|x-x(\alpha)\|.$$

因此，$x-x(\alpha)=0$ 或者 $\|x-x(\alpha)\|/\alpha\leq\|w\|$. 这意味着

$$\alpha\to 0\quad\Rightarrow\quad x-x(\alpha)\to 0.$$

接下来，只需证

$$\liminf_{\alpha\to 0}\alpha^{-1}\|x-x(\alpha)\|\geq\min\{\|w\|\colon w\in A(x)+F(x)\}.$$

引入记号

$$z(\alpha)=\alpha^{-1}(x-x(\alpha)).\tag{16.4}$$

则 $z(\alpha)$ 是依范数有界的。显然，存在 $\{\alpha_k\}\to 0$，使得

$$\|z(\alpha_k)\|\to\liminf_{\alpha\to 0}\|z(\alpha)\|,\quad z(\alpha_k)\rightharpoonup z.$$