非光滑优化算法

袁功林　盛　洲　著

科学出版社

非光滑优化算法

袁功林 盛 洲 著

科 学 出 版 社
北 京

内 容 简 介

本书旨在系统介绍基于 Moreau–Yosida 正则化的非光滑优化理论与方法，主要内容包括凸集和凸函数的概念、次梯度和 Moreau–Yosida 正则化有关性质；求解非光滑优化问题的束方法，以及牛顿束方法和有限记忆束方法；提出非光滑优化的共轭梯度算法，包括改进的 PRP 算法和改进的 HS 算法以及 Barzilai 和 Borwein（BB）算法，并给出了求解大规模非光滑问题的数值案例，供读者参考；提出非光滑优化的信赖域算法，包括调和信赖域算法和投影梯度信赖域算法在非光滑问题中的应用.

本书可作为应用数学、运筹学与控制论及经济管理有关专业的高年级本科生和研究生教材，同时也可供相关专业的科研工作者进行学术研究参考.

图书在版编目(CIP)数据

非光滑优化算法/袁功林，盛洲著. —北京：科学出版社，2017. 8

ISBN 978-7-03-054088-1

Ⅰ. ①非…　Ⅱ. ①袁…　②盛…　Ⅲ.①光滑化(数学)　Ⅳ. ①O189

中国版本图书馆 CIP 数据核字（2017）第 186954 号

责任编辑：李　欣／责任校对：彭　涛
责任印制：张　伟／封面设计：陈　敬

科 学 出 版 社 出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com

北京东华虎彩印刷有限公司 印刷
科学出版社发行　各地新华书店经销
*
2017 年 8 月第　一　版　开本：720 × 1000　B5
2018 年 1 月第三次印刷　印张：8 1/2
字数：135 000

定价：58.00 元

（如有印装质量问题，我社负责调换）

序　言

本书包含作者及合作者近几年关于非光滑优化方面的一些最新科研成果，特别是大规模非光滑优化问题，其中利用共轭梯度算法成功求解万维以上的问题，尚属首次，成果也得到国内外同行的关注，2014 年发表在 ESI 期刊 *JCAM* (Top Journal) 上的论文于 2016 年入选了 ESI 全球 Top 0.1%“热点论文”. 本书共分 4 章，前两章分别是非光滑基础和束方法，属于非光滑问题基础，第 3 章是共轭梯度法应用于大规模非光滑问题，第 4 章是非光滑问题的信赖域算法.

本书的总体策划、资金筹措和全书总撰工作由袁功林负责，参与编写的人员有：2015 级硕士生李春念和 2014 级硕士生盛洲参与编写第 1 章，第 2 章由 2016 级硕士生胡午杰和 2014 级硕士生盛洲参与完成，第 3 章由 2015 级硕士生王博朋和 2014 级硕士生盛洲参与完成，2014 级硕士生盛洲参与完成第 4 章. 在此对他们的辛勤工作表示感谢. 本书的顺利出版，感谢作者袁功林的博士生导师华东理工大学鲁习文教授和香港理工大学祁力群教授，他们给予的教诲使我终身受益；广西大学的韦增欣教授是作者袁功林的硕士导师，同样给予了很多的教导和支持；感谢广西大学给予的良好科研环境和广西大学数学与信息科学学院提供的帮助. 本书的撰写参阅了许多优秀成果，在此对他们的工作表示感谢；本书撰写不能做到面面俱到，个别基础成果或结论可能会与其他文献有相似之处，也非常感谢相关学者的理解和包容；特别要感谢作者袁功林的爱人李向荣女士对作者多年来工作的理解、支持和鼓励，儿子袁子轩和女儿袁子茜是作者的精神动力和快乐源泉，还要感谢作者的母亲、姐姐、妹妹和哥哥，他们一直默默地支持作者的工作，无怨无悔. 感谢国家自然科学基金 (编号：11261006 和 11161003) 和广西自然科学杰出青年基金项目 (编号：

2015GXNSFGA139001) 的资助.

非光滑问题是相对较为困难的问题, 因为其梯度不存在, 所以光滑问题的梯度算法不能直接应用, 往往需要求解次梯度, 而次梯度的计算很烦琐, 因此关于非光滑问题的成果在 20 世纪 90 年代之后的二三十年的时间里, 不是很多. 本书尽量采用较为易懂的方式撰写, 力求让读者比较容易接受. 本书可作为高等院校数学类专业、工科相关专业的研究生教材, 并可供从事科学计算的工程技术人员参考使用. 近年来, 关于非光滑问题的求解又相对火热起来, 也有一些比较好的成果, 本书并没有进行收集, 只是收集与自己相关的资料.

由于作者水平有限, 本书的不妥之处在所难免, 欢迎读者批评和指正.

袁功林

2017 年 5 月

目　　录

第 1 章　非光滑优化基础 ······ 1
　1.1　向量和矩阵范数 ······ 1
　1.2　凸集和凸函数 ······ 3
　1.3　次梯度 ······ 6
　1.4　Moreau-Yosida 正则化 ······ 9
　1.5　非光滑优化问题 ······ 11
第 2 章　束方法 ······ 19
　2.1　Newton 束方法 ······ 19
　2.2　有限记忆束方法 ······ 38
第 3 章　共轭梯度法 ······ 54
　3.1　共轭梯度方法基本框架 ······ 54
　3.2　改进的 HS 算法 ······ 57
　3.3　改进的 PRP 共轭梯度法 ······ 66
　3.4　改进的 BB 共轭梯度法 ······ 73
　3.5　改进的 HZ 共轭梯度法 ······ 79
第 4 章　信赖域方法 ······ 87
　4.1　信赖域方法基本框架 ······ 87
　4.2　有限记忆法 ······ 89
　4.3　梯度信赖域算法 ······ 91
　4.4　带有限记忆 BFGS 更新的积极集投影梯度信赖域算法 ······ 99
参考文献 ······ 123
索引 ······ 128

第 1 章　非光滑优化基础

1.1　向量和矩阵范数

向量和矩阵范数的概念及其有关理论在最优化算法的收敛性分析中是较为重要的基础内容, 本节将扼要地介绍这些概念与理论.

设 $\Re^n$ 表示 n 维向量空间. 下面给出向量范数的定义:

向量 $\boldsymbol{x} \in \Re^n$ 的范数 $\|\cdot\|$ 是一个非负数, 它必须满足下列条件:

(1) $\|\boldsymbol{x}\| \geqslant 0, \|\boldsymbol{x}\| = 0 \Longleftrightarrow \boldsymbol{x} = \boldsymbol{0}$;

(2) $\|\alpha \boldsymbol{x}\| = |\alpha| \|\boldsymbol{x}\|, \forall \alpha \in \Re$;

(3) $\|\boldsymbol{x} + \boldsymbol{y}\| \leqslant \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$.

这里给出一种特殊的范数, 定义向量范数实 p-范数为

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}, \quad p \geqslant 1. \tag{1.1}$$

由 p-范数的定义 (1.1), 可以得到下列常用的向量范数:

1-范数: $\|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} |x_i|$;

2-范数: $\|\boldsymbol{x}\|_2 = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{\frac{1}{2}}$;

∞-范数: $\|\boldsymbol{x}\|_\infty = \max_{1 \leqslant i \leqslant n} |x_i|$.

设 $\Re^{m \times n}$ 表示 $m \times n$ 实矩阵全体所组成的线性空间. 在满足向量范数定义中条件 (1)-(3) 的基础上, 矩阵 $\boldsymbol{A} \in \Re^{m \times n}$ 的范数还需要满足下列乘法性质:

$$\|\boldsymbol{AB}\| \leqslant \|\boldsymbol{A}\| \|\boldsymbol{B}\|, \quad \forall \boldsymbol{A} \in \Re^{m \times n}, \boldsymbol{B} \in \Re^{n \times q}.$$

定义 1.1　称矩阵范数 $\|\cdot\|_v$ 和向量范数 $\|\cdot\|$ 是相容的, 若矩阵范

数 $\|\cdot\|_v$ 相对于向量范数 $\|\cdot\|$ 满足

$$\|\boldsymbol{A}\boldsymbol{x}\| \leqslant \|\boldsymbol{A}\|_v \|\boldsymbol{x}\|, \quad \forall \boldsymbol{A} \in \Re^{m\times n}, \boldsymbol{x} \in \Re^n.$$

定义 1.2 称矩阵范数 $\|\cdot\|_v$ 为从属于向量范数 $\|\cdot\|$ 的矩阵范数, 若 $\exists \boldsymbol{x} \neq \boldsymbol{0}$ 使得

$$\|\boldsymbol{A}\|_v = \max_{\boldsymbol{x}\neq \boldsymbol{0}} \frac{\|\boldsymbol{A}\boldsymbol{x}\|}{\|\boldsymbol{x}\|} = \max_{\|\boldsymbol{x}\|=1} \|\boldsymbol{A}\boldsymbol{x}\|.$$

从而, 也可以得到下列常用的矩阵范数:

1-范数 (列和范数): $\|\boldsymbol{A}\|_1 = \max_{1\leqslant j\leqslant n} \sum_{i=1}^m |a_{ij}|$;

∞-范数 (行和范数): $\|\boldsymbol{A}\|_\infty = \max_{1\leqslant i\leqslant m} \sum_{j=1}^n |a_{ij}|$;

2-范数 (谱范数): $\|\boldsymbol{A}\|_2 = \max\{\sqrt{\lambda} | \lambda \in \lambda(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})\}$.

本书在各种迭代算法的收敛性分析中, 经常会用到谱范数和 F-范数, 下面给出 F-范数的定义:

$$\|\boldsymbol{A}\|_F = \sqrt{\operatorname{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})}.$$

向量序列和矩阵序列同样具有收敛性, 下面给出基本定义.

定义 1.3 若 $\{\boldsymbol{x}_k\}_{k=1}^\infty \subset \Re^n$, 则

$$\lim_{k\to\infty} \boldsymbol{x}_k = \boldsymbol{x},$$

若 $\{\boldsymbol{A}_k\}_{k=1}^\infty \subset \Re^{m\times n}$, 则

$$\lim_{k\to\infty} \boldsymbol{A}_k = \boldsymbol{A}.$$

向量范数和矩阵范数分别具有下列等价定理.

定理 1.1 (1) 设 $\|\cdot\|$ 和 $\|\cdot\|^*$ 是定义在 $\Re^n$ 上的两个向量范数, 则 $\forall \boldsymbol{x} \in \Re^n$, $\exists a_1 > 0, a_2 > 0$ 满足下列关系式

$$a_1\|\boldsymbol{x}\| \leqslant \|\boldsymbol{x}\|^* \leqslant a_2\|\boldsymbol{x}\|.$$

(2) 设 $\|\cdot\|$ 和 $\|\cdot\|^*$ 是定义在 $\Re^{m\times n}$ 上的两个矩阵范数, 则 $\forall \boldsymbol{A} \in \Re^{m\times n}$, $\exists b_1 > 0, b_2 > 0$ 满足下列关系式

$$b_1\|\boldsymbol{A}\| \leqslant \|\boldsymbol{A}\|^* \leqslant b_2\|\boldsymbol{A}\|.$$

类似定理 1.1, 可以给出向量序列和矩阵序列的收敛性.

定理 1.2 (1) 设 $\{\boldsymbol{x}_k\}$ 为 n 维向量序列, $\|\cdot\|$ 为定义在 $\Re^n$ 上的向量范数, 则

$$\lim_{k\to\infty}\boldsymbol{x}_k=\boldsymbol{x}\Longleftrightarrow\lim_{k\to\infty}\|\boldsymbol{x}_k-\boldsymbol{x}\|=0.$$

(2) 设 $\{\boldsymbol{A}_k\}$ 为 $m\times n$ 维矩阵序列, $\|\cdot\|$ 为定义在 $\Re^{m\times n}$ 上的矩阵范数, 则

$$\lim_{k\to\infty}\boldsymbol{A}_k=\boldsymbol{A}\Longleftrightarrow\lim_{k\to\infty}\|\boldsymbol{A}_k-\boldsymbol{A}\|=0.$$

1.2 凸集和凸函数

凸集和凸函数的概念在最优化理论分析中经常用到, 本节将介绍这些概念, 对于其中的性质和定理的证明过程感兴趣的读者可以参阅 Rockafellar 的专著 [51].

定义 1.4 设集合 $U\subset\Re^n$. 若对任意的 $\boldsymbol{x},\boldsymbol{y}\in U$ 及任意的实数 $\lambda\in[0,1]$, 都有 $\lambda\boldsymbol{x}+(1-\lambda)\boldsymbol{y}\in U$, 则称集合 U 为凸集.

凸集的几何意义是: 对非空集合 $U\in\Re^n$, 若连接其中任意两点的线段仍属于该集合, 则称该集合 U 为凸集. 凸集具有下列基本性质:

性质 1.1 设 U,U_1,U_2 是凸集, $a\in\Re$, 那么

(1) $aU=\{\boldsymbol{x}|\boldsymbol{x}=a\boldsymbol{y},\boldsymbol{y}\in U\}$ 是凸集;

(2) $U_1\cap U_2$ 是凸集;

(3) $U_1+U_2=\{\boldsymbol{x}=\boldsymbol{y}+\boldsymbol{z},\boldsymbol{y}\in U_1,\boldsymbol{z}\in U_2\}$ 是凸集.

n 维欧几里得空间中的 m 个点的凸组合是一个凸集, 即集合

$$\left\{\boldsymbol{x}=\sum_{i=1}^{m}a_i\boldsymbol{x}_i|\boldsymbol{x}_i\in\Re^n,a_i\geqslant 0,\sum_{i=1}^{m}a_i=1\right\}$$

是凸集.

以 $\boldsymbol{x}_0\in\Re^n$ 为起点, $\boldsymbol{d}\in\Re^n\setminus\{\boldsymbol{0}\}$ 为方向的射线

$$a(\boldsymbol{x}_0;\boldsymbol{d})=\{\boldsymbol{x}\in\Re^n|\boldsymbol{x}=\boldsymbol{x}_0+a\boldsymbol{d},a\geqslant 0\}$$

是凸集.

定义 1.5　集合 $U \in \Re^n$ 的凸包是指所有包含 U 的凸集的交集, U_1 为凸集. 记为

$$\operatorname{conv}(U) = \cap_{U \subseteq U_1} U_1.$$

定义 1.6　设非空集合 $U_1 \subset \Re^n$. 若对任意的 $\boldsymbol{x} \in U_1$ 和任意的实数 $\lambda > 0$, 有 $\lambda \boldsymbol{x} \in U_1$, 则称 U_1 为一个锥. 若 U_1 同时也是凸集, 则称 U_1 为一个凸锥. 对于锥 U_1, 若 $\mathbf{0} \in U_1$, 则称 U_1 为一个凸锥. 对于 U_1, 若 $\mathbf{0} \in U_1$, 则称 U_1 为一个尖锥. 相应地, 包含 $\mathbf{0}$ 的凸锥称为尖凸锥.

多面体 $\{\boldsymbol{x} \in \Re^n | \boldsymbol{A}\boldsymbol{x} \geqslant \mathbf{0}\}$ 是一个尖凸锥, 称为多面锥.

集合

$$\Re_+^n = \{\boldsymbol{x} \in \Re^n | x_i \geqslant 0, i = 1, 2, ..., n\}$$

是一个尖凸锥, 称为非负锥. 相应地, 凸锥

$$\Re_{++}^n = \{\boldsymbol{x} \in \Re^n | x_i > 0, i = 1, 2, ..., n\},$$

则称为正锥.

下面定义凸集上的凸函数.

定义 1.7　设函数 $f : U \subset \Re^n \to \Re$, 其中 U 为凸集.

(1) 若对任意的 $\boldsymbol{x}, \boldsymbol{y} \in U$ 及任意的实数 $\lambda \in [0, 1]$, 都有

$$f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leqslant \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y}).$$

则称 f 是 U 上的凸函数.

(2) 若对任意的 $\boldsymbol{x}, \boldsymbol{y} \in U$ 及任意的实数 $\lambda \in (0, 1)$, 其中 $\boldsymbol{x} \neq \boldsymbol{y}$, 都有

$$f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) < \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y}).$$

则称 f 是 U 上的严格凸函数.

(3) 若对任意的 $\boldsymbol{x}, \boldsymbol{y} \in U$ 及任意的实数 $\lambda \in [0, 1]$, $\exists \alpha > 0$, 都有

$$f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) + \frac{1}{2}\lambda(1-\lambda)\alpha\|\boldsymbol{x} - \boldsymbol{y}\|^2 \leqslant \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y}).$$

则称 f 是 U 上的一致凸函数.

凸函数具有如下性质:

性质 1.2 设 f_1, f_2, f_3 都是凸集 U 上的凸函数, $a_1 > 0, a_2 > 0, a \in \Re$, 则

(1) $a_1 f_1(\boldsymbol{x}) + a_2 f_2(\boldsymbol{x})$ 也是 U 上的凸函数;

(2) 水平集

$$L_f(a) = \{\boldsymbol{x} | \boldsymbol{x} \in U, f(\boldsymbol{x}) \leqslant a\}$$

是凸集.

下面给出几个利用函数的梯度或 Hessian 阵来判别函数凸性的定理.

定理 1.3 设 f 在凸集 $U \subset \Re^n$ 上一阶连续可微, 则

(1) f 在 U 上为凸函数的充要条件是

$$f(\boldsymbol{x}) \geqslant f(\boldsymbol{y}) + \nabla f(\boldsymbol{y})^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{y}), \quad \forall \boldsymbol{x}, \boldsymbol{y} \in U;$$

(2) f 在 U 上为严格凸函数的充要条件是当 $\boldsymbol{x} \neq \boldsymbol{y}$ 时,

$$f(\boldsymbol{x}) \geqslant f(\boldsymbol{y}) + \nabla f(\boldsymbol{y})^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{y}), \quad \forall \boldsymbol{x}, \boldsymbol{y} \in U;$$

(3) f 在 U 上为一致凸函数的充要条件是存在 $a > 0$,

$$f(\boldsymbol{x}) \geqslant f(\boldsymbol{y}) + \nabla f(\boldsymbol{y})^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{y}) + a\|\boldsymbol{x} - \boldsymbol{y}\|^2, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in U.$$

在一元函数中, 若 $f(x)$ 在区间 (a, b) 上二阶可微且 $f''(x) \geqslant 0 (> 0)$, 则 $f(x)$ 在 (a, b) 内凸 (严格凸). 同样, 也有下列定义.

定义 1.8 设 n 元实函数 f 在凸集 U 上是二阶连续可微的, 若对一切 $\boldsymbol{h} \in \Re^n$, 有 $\boldsymbol{h}^{\mathrm{T}} \nabla^2 f(\boldsymbol{x}) \boldsymbol{h} \geqslant 0$, 则称 $\nabla^2 f$ 在点 $\boldsymbol{x}$ 处是半正定的. 若对一切 $\boldsymbol{0} \neq \boldsymbol{h} \in \Re^n$, 有 $\boldsymbol{h}^{\mathrm{T}} \nabla^2 f(\boldsymbol{x}) \boldsymbol{h} > 0$, 则称 $\nabla^2 f$ 在点 $\boldsymbol{x}$ 处是正定的. 进一步, 若存在 $c > 0$, 使得对任意的 $\boldsymbol{h} \in \Re^n$, $\boldsymbol{x} \in U$, 有 $\boldsymbol{h} \nabla^2 f(\boldsymbol{x}) \boldsymbol{h} \geqslant c\|\boldsymbol{h}\|^2$, 则称 $\nabla^2 f$ 在 U 上是一致正定的.

从而, 可以得到多元函数的关于二阶导数表示凸性的定理.

定理 1.4 设 n 元实函数 f 在凸集 $U \subset \Re^n$ 上二阶连续可微的, 则

(1) f 在 U 上为凸函数的充要条件是 $\nabla^2 f(\boldsymbol{x})$ 对一切 $\boldsymbol{x} \in U$ 半正定;

(2) f 在 U 上为严格凸函数的充分条件是 $\nabla^2 f(\boldsymbol{x})$ 对一切 $\boldsymbol{x} \in U$ 正定;

(3) f 在 U 上为一致凸函数的充要条件是 $\nabla^2 f(\boldsymbol{x})$ 对一切 $\boldsymbol{x} \in U$ 一致正定.

1.3 次 梯 度

由于研究问题是非光滑的, 在算法研究中需要用次梯度的概念. 本节主要介绍这些概念. 在处理非光滑问题时, 得到一个沿着函数递增或递减的方向是比较重要的, 关于次梯度更多的性质和证明过程读者可以参阅文献 [8, 19]. 本节首先介绍方向导数.

定义 1.9 设 $D \subset \Re^n$, $f : D \to \Re$, $\bar{\boldsymbol{x}} \in D$ 和 $\boldsymbol{d}$ 是非零向量, 对 $\lambda > 0$ 且充分小满足 $\bar{\boldsymbol{x}} + \lambda \boldsymbol{d} \in D$, 若下列极限存在

$$f^o(\bar{\boldsymbol{x}}; \boldsymbol{d}) = \lim_{\lambda \to 0+} \frac{f(\bar{\boldsymbol{x}} + \lambda \boldsymbol{d}) - f(\bar{\boldsymbol{x}})}{\lambda}, \tag{1.2}$$

则称 $f^o(\bar{\boldsymbol{x}}; \boldsymbol{d})$ 为 f 在 $\bar{\boldsymbol{x}}$ 处沿着向量 $\boldsymbol{d}$ 的方向导数.

下面给出凸函数方向导数的存在性定理.

定理 1.5 设 $f : \Re^n \to \Re$ 是凸函数, 对任意的 $\bar{\boldsymbol{x}} \in \Re^n$ 和非零方向 $\boldsymbol{d}$, 方向导数 $f^o(\bar{\boldsymbol{x}}; \boldsymbol{d})$ 在 $\bar{\boldsymbol{x}}$ 处沿着向量 $\boldsymbol{d}$ 存在.

证明 设 $\lambda_1 > \lambda_2 > 0$, 由函数 f 的凸性, 可以得到

$$\begin{aligned} f(\bar{\boldsymbol{x}} + \lambda_2 \boldsymbol{d}) &= f\left(\frac{\lambda_2}{\lambda_1}(\bar{\boldsymbol{x}} + \lambda_1 \boldsymbol{d}) + \left(1 - \frac{\lambda_2}{\lambda_1}\right)\bar{\boldsymbol{x}}\right) \\ &\leqslant \frac{\lambda_2}{\lambda_1} f(\bar{\boldsymbol{x}} + \lambda_1 \boldsymbol{d}) + \left(1 - \frac{\lambda_2}{\lambda_1}\right) f(\bar{\boldsymbol{x}}), \end{aligned}$$

从而,

$$\frac{f(\bar{\boldsymbol{x}} + \lambda_2 \boldsymbol{d}) - f(\bar{\boldsymbol{x}})}{\lambda_2} \leqslant \frac{f(\bar{\boldsymbol{x}} + \lambda_1 \boldsymbol{d}) - f(\bar{\boldsymbol{x}})}{\lambda_1}.$$

因此, $\dfrac{f(\bar{\boldsymbol{x}}+\lambda\boldsymbol{d})-f(\bar{\boldsymbol{x}})}{\lambda}$ 是关于 λ 的单调递增函数. 对任意的 $\lambda \geqslant 0$, 有

$$\begin{aligned}f(\bar{\boldsymbol{x}}) &= f\left(\frac{\lambda}{1+\lambda}(\bar{\boldsymbol{x}}-\boldsymbol{d})+\frac{1}{1+\lambda}(\bar{\boldsymbol{x}}+\lambda\boldsymbol{d})\right)\\ &\leqslant \frac{\lambda}{1+\lambda}f(\bar{\boldsymbol{x}}-\boldsymbol{d})+\frac{1}{1+\lambda}f(\bar{\boldsymbol{x}}+\lambda\boldsymbol{d}),\end{aligned}$$

进而,

$$f(\bar{\boldsymbol{x}})-f(\bar{\boldsymbol{x}}-\boldsymbol{d}) \leqslant \frac{f(\bar{\boldsymbol{x}}+\lambda\boldsymbol{d})-f(\bar{\boldsymbol{x}})}{\lambda},$$

再根据当 $\lambda \to 0^+$ 时, $\dfrac{f(\bar{\boldsymbol{x}}+\lambda\boldsymbol{d})-f(\bar{\boldsymbol{x}})}{\lambda}$ 是关于 λ 的递增函数, 且有 $f(\bar{\boldsymbol{x}})-f(\bar{\boldsymbol{x}}-\boldsymbol{d})$, 因此

$$\lim_{\lambda\to 0^+}\frac{f(\bar{\boldsymbol{x}}+\lambda\boldsymbol{d})-f(\bar{\boldsymbol{x}})}{\lambda} = \inf_{\lambda>0}\frac{f(\bar{\boldsymbol{x}}+\lambda\boldsymbol{d})-f(\bar{\boldsymbol{x}})}{\lambda}.$$

从而, 定理得证. □

定义 1.10 设 $D \subset \Re^n$ 是非空凸集, $f: D \to \Re$ 是凸函数, 若对任意的 $\boldsymbol{x} \in D$ 满足

$$f(\boldsymbol{x})-f(\bar{\boldsymbol{x}}) \geqslant \boldsymbol{\xi}^{\mathrm{T}}(\boldsymbol{x}-\bar{\boldsymbol{x}}). \tag{1.3}$$

则称 $\boldsymbol{\xi}$ 为 f 在 $\bar{\boldsymbol{x}}$ 处的次梯度.

相应地, 若 $f: D \to \Re$ 是凹函数, 若对任意的 $\boldsymbol{x} \in D$ 满足

$$f(\boldsymbol{x})-f(\bar{\boldsymbol{x}}) \leqslant \boldsymbol{\xi}^{\mathrm{T}}(\boldsymbol{x}-\bar{\boldsymbol{x}}). \tag{1.4}$$

则称 $\boldsymbol{\xi}$ 为 f 在 $\bar{\boldsymbol{x}}$ 处的次梯度.

一般来说, 在 $\bar{\boldsymbol{x}}$ 处 f 的次梯度不止一个, 因此, 将满足式 (1.3) 或 (1.4) 的向量 $\boldsymbol{\xi}$ 的全体构成的集合记为 $\partial f(\bar{\boldsymbol{x}})$, 称之为 f 在点 $\bar{\boldsymbol{x}}$ 处的次微分.

定理 1.6 $f: \Re^n \to \Re$ 在点 $\boldsymbol{x} \in D$ 处的次微分 $\partial f(\boldsymbol{x})$ 为闭凸集.

凸函数的次梯度和方向导数之间存在如下关系.

定理 1.7 设 $f: \Re^n \to \Re$, $\boldsymbol{x} \in D$, 则 $\boldsymbol{\xi} \in \partial f(\boldsymbol{x})$ 的充要条件是对任意的 $\boldsymbol{d} \in \Re^n$ 满足

$$f^o(\boldsymbol{x};\boldsymbol{d}) \geqslant \langle\boldsymbol{\xi},\boldsymbol{d}\rangle. \tag{1.5}$$

下面次梯度的存在性定理可以由定理 1.7 得到.

定理 1.8　设 $f:\Re^n\to\Re$, $\boldsymbol{x}\in D$, 则 $\partial f(\boldsymbol{x})\neq\varnothing$ 的充要条件是对任意的 $\boldsymbol{y}\in\Re^n$, $\exists a>0$, 使得

$$f(\boldsymbol{y})-f(\boldsymbol{x})\geqslant -a\|\boldsymbol{y}-\boldsymbol{x}\|. \tag{1.6}$$

证明　假设 $\exists\boldsymbol{\xi}\in\partial f(\boldsymbol{x})$. 由式 (1.3) 和 Cauchy-Schwarz 不等式, 对任意的 $\boldsymbol{x},\boldsymbol{y}\in\Re^n$ 有

$$f(\boldsymbol{y})-f(\boldsymbol{x})\geqslant -\|\boldsymbol{\xi}\|\|\boldsymbol{y}-\boldsymbol{x}\|. \tag{1.7}$$

当 $\boldsymbol{\xi}=\mathbf{0}$ 时, 式 (1.6) 对任意的 $a>0$ 均成立. 当 $\boldsymbol{\xi}\neq\mathbf{0}$ 时, 记 $a=\|\boldsymbol{\xi}\|$, 则得到式 (1.6). 假设 $\partial f(\boldsymbol{x})=\varnothing$, 则 $\exists\boldsymbol{d}\in\Re^n$ 使得 $f^o(\boldsymbol{x};\boldsymbol{d})=-\infty$ 且 $\|\boldsymbol{d}\|=1$. 取 $\boldsymbol{y}=\boldsymbol{x}+t\boldsymbol{d}$, 并令 t 单调递减趋近于 0, 那么

$$\frac{f(\boldsymbol{y})-f(\boldsymbol{x})}{\|\boldsymbol{y}-\boldsymbol{x}\|}=\frac{f(\boldsymbol{x}+t\boldsymbol{d})-f(\boldsymbol{x})}{t}\to-\infty.$$

从而, 不存在 $a>0$ 满足式 (1.6), 与定理条件矛盾, 假设不成立. 因此, 定理得证. □

当所有 $f_i, i=1,2,\ldots$ 均在 $\boldsymbol{x}$ 处可微时, 则

$$\partial f(\boldsymbol{x})=\operatorname{conv}\{\nabla f_i(\boldsymbol{x})|i=1,2,\ldots\}. \tag{1.8}$$

例 1.1　定义如下函数 $f:\Re^3\to\Re$: $f(\boldsymbol{x})=\min\{-x_1-x_2+x_3, x_1-x_3, x_2-x_3\}$, 那么利用式 (1.8), f 在点 $\boldsymbol{x}_0=(0,0,0)^{\mathrm{T}}$ 处的次微分为

$$\partial f(\boldsymbol{x}_0)=\operatorname{conv}\{(-1,-1,1)^{\mathrm{T}},(1,0,-1)^{\mathrm{T}},(0,1,-1)^{\mathrm{T}}\}.$$

下面定理说明了 $\partial f(\boldsymbol{x})$ 作为映射具有连续性.

定理 1.9　设 $f:\Re^n\to\Re$ 为凸函数, 给定满足 $\boldsymbol{x}_k\to\boldsymbol{x}$ 的点列 $\{\boldsymbol{x}_k\}\subseteq\Re^n$, 若 $\boldsymbol{\xi}_k\in\partial f(\boldsymbol{x}_k)$ 且 $\boldsymbol{\xi}_k\to\boldsymbol{\xi}$, 则有 $\boldsymbol{\xi}\in\partial f(\boldsymbol{x})$.

1.4 Moreau-Yosida 正则化

考虑非光滑最优化问题

$$\min_{\boldsymbol{x}\in\Re^n} f(\boldsymbol{x}), \tag{1.9}$$

其中 $f:\Re^n\to\Re$ 为非光滑凸函数. 利用 Moreau-Yosida 正则化函数,

$$F(\boldsymbol{x})=\min_{\boldsymbol{z}\in\Re^n}\left\{f(\boldsymbol{z})+\frac{1}{2\lambda}\|\boldsymbol{z}-\boldsymbol{x}\|^2\right\}, \tag{1.10}$$

其中 $F:\Re^n\to\Re$ 是 f 的 Moreau-Yosida 正则化函数, $\lambda>0$ 是一个参数, $\|\cdot\|$ 表示欧氏范数. $F(\boldsymbol{x})$ 具有一些很好的性质, 见文献 [3, 10, 29]. 则得到与 (1.9) 等价的优化问题

$$\min_{\boldsymbol{x}\in\Re^n} F(\boldsymbol{x}). \tag{1.11}$$

令

$$\theta(\boldsymbol{z})=f(\boldsymbol{z})+\frac{1}{2\lambda}\|\boldsymbol{z}-\boldsymbol{x}\|^2, \tag{1.12}$$

并定义 $\boldsymbol{p}(\boldsymbol{x})=\arg\min\limits_{\boldsymbol{z}\in\Re^n}\theta(\boldsymbol{z})$. 结合 (1.10), $F(\boldsymbol{x})$ 可以表示为

$$F(\boldsymbol{x})=f(\boldsymbol{p}(\boldsymbol{x}))+\frac{1}{2\lambda}\|\boldsymbol{p}(\boldsymbol{x})-\boldsymbol{x}\|^2, \tag{1.13}$$

记 $\boldsymbol{g}$ 是 F 的梯度, 它具有较好的性质. $F(\boldsymbol{x})$ 的广义 Jacobi 矩阵和 BD-正则性质可以分别参见文献 [6, 48].

性质 1.3 (1) 函数 F 是有界凸的且处处可微,

$$\boldsymbol{g}(\boldsymbol{x})=\nabla F(\boldsymbol{x})=\frac{\boldsymbol{x}-\boldsymbol{p}(\boldsymbol{x})}{\lambda}, \tag{1.14}$$

此处, 梯度 $\boldsymbol{g}:\Re^n\to\Re^n$ 是满足模长为 λ 的 Lipschitz 连续函数,

$$\|\boldsymbol{g}(\boldsymbol{x})-\boldsymbol{g}(\boldsymbol{y})\|\leqslant\frac{1}{\lambda}\|\boldsymbol{x}-\boldsymbol{y}\|,\quad \forall\boldsymbol{x},\boldsymbol{y}\in\Re^n. \tag{1.15}$$

(2) $\boldsymbol{x}$ 是问题 (1.10) 的最优解, 当且仅当 $\nabla F(\boldsymbol{x})=0$, 即 $\boldsymbol{p}(\boldsymbol{x})=\boldsymbol{x}$.

显然, 在求解 $f(\boldsymbol{x})$ 的最优解即 $\arg\min\limits_{\boldsymbol{z}\in\Re^n}\theta(\boldsymbol{z})$ 的过程中可以得到 $F(\boldsymbol{x})$ 和 $\boldsymbol{g}(\boldsymbol{x})$. 但解出 $\theta(z)$ 的最小值 $\boldsymbol{p}(\boldsymbol{x})$ 是非常困难的, 甚至是不可能的. 因此, 不能用 $\boldsymbol{p}(\boldsymbol{x})$ 的精确值来定义 $F(\boldsymbol{x})$ 和 $\boldsymbol{g}(\boldsymbol{x})$. 幸运的是, 对于每一个 $\boldsymbol{x}\in\Re^n$ 和任意的 $\varepsilon>0$, 存在一个向量 $\boldsymbol{p}^\alpha(\boldsymbol{x},\varepsilon)\in\Re^n$ 满足

$$f(\boldsymbol{p}^\alpha(\boldsymbol{x},\varepsilon))+\frac{1}{2\lambda}\|\boldsymbol{p}^\alpha(\boldsymbol{x},\varepsilon)-\boldsymbol{x}\|^2\leqslant F(\boldsymbol{x})+\varepsilon, \tag{1.16}$$

因此, 当 ε 很小时, 可以用 $\boldsymbol{p}^\alpha(\boldsymbol{x},\varepsilon)$ 分别定义 $F(\boldsymbol{x})$ 和 $\boldsymbol{g}(\boldsymbol{x})$, 具体如下:

$$F^\alpha(\boldsymbol{x},\varepsilon)=f(\boldsymbol{p}^\alpha(\boldsymbol{x},\varepsilon))+\frac{1}{2\lambda}\|\boldsymbol{p}^\alpha(\boldsymbol{x},\varepsilon)-\boldsymbol{x}\|^2, \tag{1.17}$$

$$\boldsymbol{g}^\alpha(\boldsymbol{x},\varepsilon)=\frac{\boldsymbol{x}-\boldsymbol{p}^\alpha(\boldsymbol{x},\varepsilon)}{\lambda}. \tag{1.18}$$

一些计算不可微凸函数 $\boldsymbol{p}^\alpha(\boldsymbol{x},\varepsilon)$ 的算法可参考见文献 [18]. 下面给出一些 $F^\alpha(\boldsymbol{x},\varepsilon)$ 和 $\boldsymbol{g}^\alpha(\boldsymbol{x},\varepsilon)$ 的性质.

性质 1.4　$\boldsymbol{p}^\alpha(\boldsymbol{x},\varepsilon)$ 是一个向量且满足 (1.16), $F^\alpha(\boldsymbol{x},\varepsilon)$ 和 $\boldsymbol{g}^\alpha(\boldsymbol{x},\varepsilon)$ 分别由 (1.17) 和 (1.18) 定义给出, 可以得到

$$F(\boldsymbol{x})\leqslant F^\alpha(\boldsymbol{x},\varepsilon)\leqslant F(\boldsymbol{x})+\varepsilon, \tag{1.19}$$

$$\|\boldsymbol{p}^\alpha(\boldsymbol{x},\varepsilon)-\boldsymbol{p}(\boldsymbol{x})\|\leqslant\sqrt{2\lambda\varepsilon}, \tag{1.20}$$

$$\|\boldsymbol{g}^\alpha(\boldsymbol{x},\varepsilon)-\boldsymbol{g}(\boldsymbol{x})\|\leqslant\sqrt{\frac{2\varepsilon}{\lambda}}. \tag{1.21}$$

由性质 1.4 可以得到 $F^\alpha(\boldsymbol{x},\varepsilon)$ 和 $\boldsymbol{g}^\alpha(\boldsymbol{x},\varepsilon)$ 的近似值. 通过令参数 $\varepsilon\to0$, 使得 $F^\alpha(\boldsymbol{x},\varepsilon)$ 和 $\boldsymbol{g}^\alpha(\boldsymbol{x},\varepsilon)$ 会尽可能地接近 $F(\boldsymbol{x})$ 和 $\boldsymbol{g}(\boldsymbol{x})$.

性质 1.5　令 $\boldsymbol{x}_k\subset\mathrm{dom}f$ 且定义如下:

$$\boldsymbol{x}_{k+1}=\boldsymbol{x}_k-\alpha_k\boldsymbol{v}_k,\quad k=1,2,\ldots, \tag{1.22}$$

其中步长 $\alpha_k>0$, $\boldsymbol{v}_k$ 是在 $\boldsymbol{x}_k$ 处的近似次梯度, 即

$$\boldsymbol{v}_k\in\partial_{\varepsilon_k}f(\boldsymbol{x}_k),\quad k=1,2,\ldots, \tag{1.23}$$

(1) 如果 v_k 满足

$$v_k \in \partial f(\boldsymbol{x}_{k+1}), \quad k = 1, 2, \ldots, \tag{1.24}$$

则 (1.23) 成立, 其中

$$\varepsilon_k = f(\boldsymbol{x}_k) - f(\boldsymbol{x}_{k+1}) - \alpha_k \|\boldsymbol{v}_k\|^2 \geqslant 0. \tag{1.25}$$

(2) 反之, 如果 (1.23) 成立, ε_k 由 (1.25) 定义, 则 (1.24) 成立, 即 $\boldsymbol{x}_{k+1} = \boldsymbol{p}^{\alpha}(\boldsymbol{x}_k, \varepsilon_k)$.

1.5 非光滑优化问题

本节首先给出一些小规模非光滑优化问题, 其中 $\boldsymbol{x}_0$ 是测试问题的初始点, $f_{\text{ops}}(\boldsymbol{x})$ 是对应测试问题的最优解.

问题 1 Rosenbrock[41]

$$f(\boldsymbol{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2,$$

$$\boldsymbol{x}_0 = (-1.2, 1.0)^{\mathrm{T}}, \quad f_{\text{ops}}(\boldsymbol{x}) = 0.$$

问题 2 Crescent[41]

$$f(\boldsymbol{x}) = \max\{x_1^2 + (x_2 - 1)^2 + x_2 - 1,\ -x_1^2 - (x_2 - 1)^2 + x_2 + 1\},$$

$$\boldsymbol{x}_0 = (-1.5, 2.0)^{\mathrm{T}}, \quad f_{\text{ops}}(\boldsymbol{x}) = 0.$$

问题 3 CB2[7]

$$f(\boldsymbol{x}) = \max\{x_1^2 + x_2^4,\ (2 - x_1)^2 + (2 - x_2)^2, 2e^{-x_1 + x_2}\},$$

$$\boldsymbol{x}_0 = (1.0, -1.0)^{\mathrm{T}}, \quad f_{\text{ops}}(\boldsymbol{x}) = 1.9522245.$$

问题 4 CB3[7]

$$f(\boldsymbol{x}) = \max\{x_1^4 + x_2^2,\ (2 - x_1)^2 + (2 - x_2)^2, 2e^{-x_1 + x_2}\},$$