不确定性支持向量机原理及应用

●杨志民　刘广利　著

科学出版社
www.sciencep.com

不确定性支持向量机原理及应用

杨志民　刘广利　著

本书获浙江工业大学专著与研究生教材出版基金资助
（资助基金编号 20060105）

科学出版社
北京

内 容 简 介

不确定性支持向量机是商业智能和数据挖掘的一个新的研究领域，它能有效地处理不确定性信息条件下的模式分类、回归预测、聚类分析和有序回归等诸多问题，并可应用于预测预警、综合评价等领域. 本书从不确定性规划出发，结合模糊、粗糙和未确知等不确定性理论，详细阐述了适用于各类问题的不确定性支持向量机模型和算法. 目前国内外不确定性优化理论和支持向量机相结合的研究正处于快速发展阶段，希望本书的出版能促进不确定性支持向量机在我国各个应用领域的普及与提升.

本书适合于高等院校高年级本科生、研究生、教师和相关领域的实际工作者阅读. 以期能给相关领域的理论研究者和应用工作者提供一些思路和帮助.

图书在版编目(CIP)数据

不确定性支持向量机原理及应用/杨志民，刘广利著. —北京：科学出版社，2007

ISBN 978-7-03-018775-8

Ⅰ. 不… Ⅱ. ①杨… ②刘… Ⅲ. 向量计算机-算法理论 Ⅳ. TP301.6

中国版本图书馆 CIP 数据核字(2007)第 041018 号

责任编辑：范庆奎 贾瑞娜 / 责任校对：宋玲玲

责任印制：赵德静 / 封面设计：王 浩

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

源海印刷有限责任公司印刷

科学出版社发行 各地新华书店经销

*

2007 年 4 月第 一 版 开本：B5(720×1000)

2007 年 4 月第一次印刷 印张：15 1/2

印数：1—3 500 字数：290 000

定价：38.00 元

(如有印装质量问题，我社负责调换〈环伟〉)

前　言

支持向量机是 Vapnik 等人提出的一类新型机器学习方法，它能够非常成功地处理分类和回归问题. 由于支持向量机出色的学习性能，该技术已成为机器学习界的研究热点，并在很多领域得到了成功的应用. 但是，作为一种尚未成熟的新技术，支持向量机模型目前还存在许多局限. 客观世界存在大量的不确定性信息，如果支持向量机的训练集中含有不确定性信息，那么标准的支持向量机模型将无能为力. 基于含有不确定性信息的数据挖掘问题，本书尝试构建了一套完整的不确定性支持向量机框架体系.

本书首先介绍了最优化以及不确定性数学的基础理论，然后就统计学习理论和支持向量机进行了讨论. 之后，第 4、5 章提出了基于可能性理论以及基于模糊系数规划的模糊支持向量分类机，就模糊线性可分、近似模糊线性可分和模糊非线性问题三种情形分别构建了模糊支持向量分类模型及算法. 当最优超平面中的参数为模糊量 (模糊向量或模糊数) 时，第 6 章建立了模糊线性支持向量分类机和模糊线性支持向量回归机，给出了带有模糊决策的模糊机会约束规划模型，并讨论了带有模糊决策的模糊机会约束规划解法 —— 基于模糊模拟的遗传算法，以及模糊支持向量集 (模糊集合) 的概念. 本书的第 7 章提出了带有不确定信息的支持向量分类方法，该方法将专家意见和不确定信息融入分类系统. 针对多类分类问题的一种特殊问题 —— 有序回归，第 8 章建立了不确定有序支持向量回归分类模型. 第 9 章介绍了不确定理论、核理论与聚类问题的结合. 在第 10 章，作者尝试提出了未确知支持向量机模型建立的设想. 最后第 11 章则就相关的应用案例进行了介绍.

本书由杨志民和刘广利共同撰稿完成. 其内容是作者多年在不确定性支持向量机方面的研究成果，同时也是许多同行专家指导和帮助的结果. 在本书的撰写过程中，许多人给予了无私的帮助. 首先要感谢中国农业大学邓乃扬教授的指导. 其次要感谢中国农业大学王库教授、北京科技大学尹怡欣教授、北京理工大学刘宝光教授、中国科学院田英杰博士、西南交通大学徐扬教授、安阳师范学院王爱民教授的指导和帮助. 此外，还要感谢浙江工业大学计建炳教授、方志民教授、张一青教授、池仁勇教授、周明华教授、邸继征教授、隋成华教授、乐孜纯教授和蒋天颖同志的帮助. 本书的撰写还得到了齐志泉、李春艳同志的帮助，在此一并表示感谢.

本书的出版还得到了浙江省自然科学基金 (编号 Y606082) 和中国农业大学信息与电气工程学院科学研究基金的资助，在此表示感谢.

由于时间仓促，加上作者水平有限，书中难免有不当之处，敬请读者批评指正.

作　者

2006 年 11 月 14 日

目　　录

前言
第 1 章　最优化理论基础 …… 1
1.1　最优化问题 …… 1
1.1.1　最优化问题 …… 1
1.1.2　线性规划 …… 1
1.1.3　凸最优化 …… 3
1.2　最优性条件 …… 5
1.2.1　几何最优性条件 …… 6
1.2.2　Fritz John 条件 …… 6
1.2.3　KKT 条件 …… 7
1.2.4　鞍点 …… 9
1.2.5　对偶理论 …… 9
1.2.6　二次规划 …… 11
1.3　最优化算法 …… 13
1.3.1　线性逼近法 …… 13
1.3.2　线性约束条件下的线性逼近法 …… 13
1.3.3　非线性约束条件下的线性逼近法 …… 15
1.3.4　可行方向法 …… 16
1.3.5　投影梯度法 …… 16
1.3.6　罚函数法 …… 18
第 2 章　不确定性数学基础 …… 19
2.1　模糊数学 …… 19
2.1.1　模糊子集及其运算 …… 19
2.1.2　模糊集的基本定理 …… 22
2.1.3　模糊矩阵 …… 24
2.1.4　模糊关系 …… 28

2.1.5 模糊等价矩阵 30
2.2 粗糙集 32
2.2.1 粗糙集理论的基本思想 32
2.2.2 粗糙集理论的产生和发展 32
2.2.3 粗糙集理论的一些基本概念 33
2.2.4 粗糙集的应用 37
2.3 未确知理论 38
2.3.1 未确知数的概念 38
2.3.2 未确知数的加减运算 40
2.3.3 未确知数的乘除运算 44
2.3.4 未确知数的大小关系 44
2.3.5 未确知数的数学期望与方差 47
2.3.6 高阶未确知数降阶方法 50
第 3 章 统计学习理论与支持向量机 55
3.1 统计学习理论 55
3.2 支持向量分类 57
3.2.1 基本概念 57
3.2.2 线性支持向量机 58
3.2.3 非线性支持向量机 61
3.2.4 支持向量分类算法 63
3.2.5 模型参数选择 65
3.2.6 基他分类模型 68
3.3 支持向量回归 69
3.3.1 ε - 支持向量回归 69
3.3.2 ν- 支持向量回归 71
3.3.3 其他回归模型 72
3.3.4 时间序列分析 74
3.4 核函数及其应用 75
3.4.1 核理论基础 75
3.4.2 核主成分分析 79
3.4.3 预警指标选择 81

3.4.4 核聚类 84
第 4 章 基于可能性理论的模糊支持向量分类机 86
4.1 可能性测度与模糊机会约束规划 86
4.2 模糊特征及其表示 89
4.3 模糊支持向量分类机 90
4.3.1 模糊线性可分模糊支持向量分类机 93
4.3.2 近似模糊线性可分模糊支持向量分类机 97
4.3.3 模糊非线性模糊支持向量分类机 102
4.4 数据试验 106
4.5 最佳置信水平 109
第 5 章 基于模糊系数规划的模糊支持向量分类机 111
5.1 模糊系数规划 111
5.2 模糊支持向量分类机 114
5.2.1 含有模糊信息的线性可分问题模糊支持向量分类机 114
5.2.2 含有模糊信息的近似线性可分问题模糊支持向量分类机 119
5.2.3 含有模糊信息的非线性问题模糊支持向量分类机 125
5.2.4 数据试验 129
5.3 最佳阈值 130
第 6 章 模糊线性支持向量机 132
6.1 带有模糊决策的模糊机会约束规划 132
6.2 模糊线性支持向量分类机 132
6.3 模糊线性支持向量回归机 134
6.4 基于模糊模拟的遗传算法 137
6.4.1 模糊模拟 137
6.4.2 基于模糊模拟的遗传算法 139
6.5 模糊支持向量集 141
第 7 章 不确定支持向量机 143
7.1 粗糙集支持向量机 143
7.1.1 知识约简方法 143
7.1.2 基于粗糙集预处理的支持向量分类 144
7.1.3 基于粗糙集预处理的支持向量回归 146

7.1.4 财务困境预警应用实例 …… 146
7.2 加权支持向量机 …… 148
7.2.1 样本不平衡问题 …… 148
7.2.2 加权支持向量机模型 …… 149
7.2.3 参数选择 …… 150
7.2.4 数据试验 …… 151
7.3 模糊模式识别与不完全支持向量机 …… 152
7.3.1 引言 …… 152
7.3.2 不完全支持向量机 …… 152
7.3.3 模糊隶属度的确定 …… 154
7.3.4 模糊模式识别 …… 156
7.3.5 模糊模式识别与模糊支持向量分类机的比较 …… 161
7.4 不确定支持向量机 …… 162
7.4.1 问题提出 …… 162
7.4.2 不确定支持向量分类 …… 163
7.4.3 USVC 与 FSVM 的关系 …… 167
7.4.4 数据试验 …… 168
第 8 章　不确定有序支持向量回归 …… 170
8.1 多类问题与有序回归 …… 170
8.2 模糊多类支持向量机 …… 172
8.2.1 多类支持向量分类方法 …… 172
8.2.2 模糊多类 SVM …… 173
8.3 基于间隔最大化的有序回归模型 …… 173
8.3.1 最小间隔最大化 …… 174
8.3.2 总间隔最大化 …… 176
8.4 模糊有序回归模型 …… 179
8.4.1 模糊 OSVR 原理 …… 179
8.4.2 隶属度的确定 …… 180
8.4.3 算法描述 …… 181
8.4.4 数据试验 …… 181
8.5 不确定有序回归模型 …… 182

第 9 章 不确定聚类方法 187
9.1 模糊核 k–均值算法 187
9.2 可能性核聚类算法 189
9.3 加权有序支持向量聚类算法 191
9.3.1 有序支持向量聚类 191
9.3.2 加权聚类算法 192
第 10 章 建立未确知支持向量机的设想 194
10.1 未确知事件的可靠度 194
10.2 未确知机会约束规划模型及其算法 195
10.3 建立未确知支持向量机的设想 200
第 11 章 应用 203
11.1 冠心病诊断 203
11.2 城市空气质量评价 209
11.3 粮食预警 213
11.3.1 中国粮食产量预警 213
11.3.2 中国粮食安全预警 214
11.3.3 粮食产量增长率回归预测 216
11.3.4 粮食产量增长率时间序列分析 217
11.3.5 粮食安全综合评价 218
11.3.6 粮食预警指标选择 219
11.3.7 粮食安全区划 221
11.4 棉花预警 223
11.4.1 棉花产量预警 (两类) 223
11.4.2 棉花产量预警 (多类) 224
11.4.3 有序回归棉花预警 225
11.5 财务困境识别 226
11.6 股票预测 229
11.7 遥感影像分类 232
参考文献 234

第 1 章　最优化理论基础

1.1　最优化问题

1.1.1　最优化问题

所谓最优化就是找出一个多变量函数的极小值点. 最优化问题数学模型的一般形式为

$$\begin{cases} \min & f(x) \\ \text{s.t.} & x \in S, \end{cases} \tag{1.1}$$

其中, $x = (x_1, x_2, \cdots, x_n)^{\mathrm{T}} \in \mathbf{R}^n$ 称为决策变量; $f(x)$ 是 x 的函数, 称为目标函数; s.t. 是 “subject to” 的缩写, S 是 $\mathbf{R}^n$ 的子集.

当 $S \subset \mathbf{R}^n$ 时, x 的取值受限制, x 必须属于 S, S 便称为约束条件. 公式 (1.1) 就是在集合 S 中求使 $f(x)$ 取极小值的 x, 称为约束最优化问题; 当 $S = \mathbf{R}^n$ 时, 问题 (1.1) 就转化为在 $\mathbf{R}^n$ 内求 $f(x)$ 的极小值, 称为无约束最小化问题.

当约束条件是线性的等式或不等式方程, 而目标函数是 x 的线性函数时, 称为线性规划. 当约束条件或目标函数中出现非线性时, 称为非线性规划. 非线性规划问题中最简单的一种是二次规划问题, 它的目标函数是 x 的二次函数, 而约束条件是 x 的线性等式或不等式方程, 这是非线性规划问题中重要的一种类型. 当规划问题与时间相关, 即前一时刻的行为将影响下时刻的行为时, 则称为动态规划, 又称为多级决策理论. 此外, 最优化问题还包括运输问题、最小二乘问题、最大最小问题等. 关于最优化问题更多的知识, 请参阅文献 [54]、[56]、[65]、[78]、[86].

1.1.2　线性规划

线性规划是数学规划的重要分支. 自从 1947 年 George Dantzig 建立了适合在计算机上进行计算的单纯形法, 线性规划得到了迅速的发展并日益扩大其应用领域. 其应用涵盖管理科学、系统工程、应用数学、数理经济学、信息科学等众多学科领域.

线性规划都具备线性函数约束、符号约束、线性目标函数三个特征. 线性规划中的所有常数和决策变量都是在实数范围中考虑的, 决策变量的个数和线性函数约束的数目都是有限的. 当要求决策变量取整数时, 这种线性规划称为整数规划.

满足线性规划的函数约束和符号约束的解称为可行解, 在一般情况下, 线性规划的可行解有无限多个.

标准的线性规划问题可以表示为

$$\begin{cases} \min & z = c^{\mathrm{T}} x \\ \text{s.t.} & Ax \geqslant b, \\ & x \geqslant 0, \end{cases} \tag{1.2}$$

其中, A 是 $m \times n$ 维结构矩阵; b 为 $m \times 1$ 维条件向量; c 是 $n \times 1$ 维的费用向量; x 是 $n \times 1$ 维的决策向量.

如果向量 x 满足式 (1.2) 的约束条件 $Ax \geqslant b, x \geqslant 0$, 则称 x 为可行解; 如果 x 同时满足式 (1.2) 的目标函数, 则称 x 为最优解. 并不是所有线性规划问题都有可行解和最优解, 即使线性规划的最优解存在, 也不一定唯一.

每个线性规划问题都存在另一个与它密切关联的线性规划问题, 将其中一个称为原问题, 简记 P, 另一个称为它的对偶问题, 简记为 D.

定义 1.1 将线性规划问题

$$\begin{cases} \max & w = y^{\mathrm{T}} b \\ \text{s.t.} & y^{\mathrm{T}} A \leqslant c^{\mathrm{T}}, \\ & y \geqslant 0, \end{cases} \tag{1.3}$$

称为线性规划问题 (1.2) 的对偶问题.

问题 (1.3) 和问题 (1.2) 互为对偶问题.

可以证明, 如果 x^* 是 (1.2) 的可行解, y^* 是对偶问题 (1.3) 的可行解, 而且 $c^{\mathrm{T}} x^* = y^{*\mathrm{T}} b$, 则 x^* 和 y^* 分别是问题 (1.2) 和问题 (1.3) 的最优解.

尽管 1977 年 George Dantzig 提出单纯形法以来, 出现了若干新的算法, 如椭圆形法以及 Karmarkar 方法等, 但单纯形法仍然是线性规划中最重要的实用算法, 这里我们只介绍单纯形法. 单纯形法的基本思想, 简单地可概括为如下几点:

(1) 求出一个基可行解, 或者判定其无可行解, 如无可行解则计算停止, 否则进入下一阶段;

(2) 对基可行解进行最优性判别, 如果已达到最优, 则计算停止; 或者无最优解, 计算也可停止;

(3) 替换基, 求出新的基可行解, 转回到第 (2) 点.

上述第 (1) 点是第一阶段, 第 (2)、(3) 点属于第二阶段, 而第一阶段的具体算法与第二阶段是完全相同的, 只不过考虑的问题略有不同.

1.1.3 凸最优化

定义 1.2 集合 $S \subset \mathbf{R}^N$, 若任两点 $x, y \in S$, 有

$$(1-\theta)x + \theta y \in S \qquad (0 \leqslant \theta \leqslant 1), \tag{1.4}$$

我们称 S 为 $\mathbf{R}^N$ 中的凸集. 根据凸集的定义, 可以证明有下述性质, 即在同一空间中:

(1) 有限个凸集的交也是凸集;

(2) 如果 S_1 和 S_2 都是凸集, 则集合

$$S = \{z : z = x - y, x \in S_1, y \in S_2\} \tag{1.5}$$

也是凸集;

(3) 如果 S_1 和 S_2 都是凸集, 则集合

$$S = \{z : z = x + y, x \in S_1, y \in S_2\} \tag{1.6}$$

也是凸集;

(4) 如果 $S_1, \cdots, S_n$ 都是凸集, 则集合

$$S = \left\{ z : z = \sum_{i=1}^{n} a_i x^{(i)}, x^{(i)} \in S_i, a_i \geqslant 0, \sum_{i=1}^{n} a_i = 1 \right\} \tag{1.7}$$

也是凸集.

定义 1.3 如果集合 S 包含它的一切极限点, 则称 S 为闭集. 换言之, 如果点列 $x^{(1)}, x^{(2)}, x^{(3)}, \cdots \in S$ 及 $x^{(k)} \to x^*(k \to \infty)$, 有 $x^* \in S$, 则 S 是闭的.

定义 1.4 若 $A = [a^{(1)}, \cdots, a^{(n)}]$ 是 $m \times n$ 矩阵, 集合

$$C = \{Ax : x \geqslant 0\} \tag{1.8}$$

或写为

$$C = \left\{ x_1 a^{(1)} + \cdots + x_n a^{(n)},\ \forall x_j \geqslant 0,\ j = 1, \cdots, n \right\} \tag{1.9}$$

称为向量 $a^{(1)}, \cdots, a^{(n)}$ 的有限集生成的有限锥.

定义 1.5 设 $f : C \subset \mathbf{R}^n \to \mathbf{R}^1$, 其中 C 是非空凸集, 若对于任给的 $x^{(1)}$, $x^{(2)} \in C$ 及任给的 $\lambda \in [0, 1]$ 都有

$$f(\lambda x^{(1)} + (1-\lambda)x^{(2)}) \leqslant \lambda f(x^{(1)}) + (1-\lambda) f(x^{(2)}), \tag{1.10}$$

则称 $f(x)$为C 上的凸函数, 如果上述定义式严格不等, 则称为严格凸函数.

把上述定义式中的不等号反向, 即可得到凹函数和严格凹函数的定义. 显然, 若函数 $f(x)$ 是凸函数 (严格凸函数), 则 $f(x)$ 一定是凹函数 (严格凹函数).

下面介绍凸函数的性质.

性质 1.1　若 $f(x)$ 是凸集 $C \in \mathbf{R}^n$ 上的凸函数, 对任意正整数 $n \geqslant 2$ 及任意 $x(1), x(2), \cdots, x(n) \in C$, 以及任意 $\lambda_1, \lambda_2, \cdots, \lambda_n \geqslant 0$, 满足 $\sum\limits_{i=1}^{n} \lambda_i = 1$, 有

$$f\left(\sum_{i=1}^{n} \lambda_i x^{(i)}\right) \leqslant \sum_{i=1}^{n} \lambda_i f\left(x^{(i)}\right) \tag{1.11}$$

成立.

性质 1.2　若 $f_1(x)$ 和 $f_2(x)$ 是凸集 $C \in \mathbf{R}^n$ 上的凸函数, 则

$$f(x) = f_1(x) + f_2(x) \tag{1.12}$$

也是 C 上的凸函数.

性质 1.3　若 $f(x)$ 是凸集 $C \in \mathbf{R}^n$ 上的凸函数, 则对任意的 $\lambda \geqslant 0$, 函数 $\lambda f(x)$ 也是 C 上的凸函数.

可以推出：若 $f_1(x), f_2(x), \cdots, f_k(x)$ 是凸集 $C \in \mathbf{R}^n$ 上的凸函数, $\lambda_1, \lambda_2, \cdots, \lambda_k \geqslant 0$, 则非负线性组成

$$f(x) = \sum_{i=1}^{n} \lambda_i f_i(x) \tag{1.13}$$

也是 C 上的凸函数.

性质 1.4　设 $f(x)$ 是凸集 $C \in \mathbf{R}^n$ 上的凸函数, 则对每一实数 β, 集合

$$C_\beta \{x : f(x) \leqslant \beta, x \in C\} \tag{1.14}$$

是凸集.

定理 1.1　设 $C \in \mathbf{R}^n$ 为非空开凸集, $f : C \to \mathbf{R}^1$ 在 C 上可微, 则 f 是 C 上的凸函数的充要条件是对 $\forall x^{(1)}, x^{(2)} \in C$, 恒有

$$f(x^{(2)}) \geqslant f(x^{(1)} + \nabla f(x^{(1)})^{\mathrm{T}}(x^{(2)} - x^{(1)}) \tag{1.15}$$

成立, 其中 $\nabla f(x^{(1)})$ 为函数 f 在点 $x^{(1)}$ 处的梯度.

定理 1.2　设 $C \in \mathbf{R}^n$ 为非空开凸集, $f : C \to \mathbf{R}^1$ 在 C 上二阶可微, 则 f 是 C 上的凸函数的充要条件是每一 $x \in C$, f 在 x 处的 Hessen 矩阵

$$H(x) = \nabla^2 f(x) \tag{1.16}$$

为半正定.

定理 1.3 设 $C\in\mathbf{R}^n$ 为非空开凸集, $f:C\to\mathbf{R}^1$ 在 C 上二阶可微, 若对一切 $x\in C$, 由式 (1.16) 定义的矩阵 $H(x)$ 正定, 则 f 是 C 上严格凸函数.

定义 1.6 (凸约束问题) 称约束问题

$$\begin{cases} \min & f(x), \quad x\in\mathbf{R}^n \\ \text{s.t.} & c_i(x)\leqslant 0, \quad i=1,\cdots,p, \\ & c_i(x)=0, \quad i=p+1,\cdots,p+q \end{cases} \tag{1.17}$$

为凸约束最优化问题, 如果其中的目标函数 $f(x)$ 和约束函数 $c_i(x)(i=1,\cdots,p)$ 都是凸函数, 而 $c_i(x)=0(i=p+1,\cdots,p+q)$ 是线性函数.

定理 1.4 (凸函数的极小) 考虑凸约束问题 (1.17). 设问题的可行域

$$D=\{x|c_i(x)\leqslant 0, i=1,\cdots,p; c_i(x)=0, i=p+1,\cdots,p+q; x\in\mathbf{R}^N\}, \tag{1.18}$$

则

(1) 若问题有局部解 x^*, 则 x^* 是问题的全局解;

(2) 问题的全局解组成的集合是凸集;

(3) 问题有局部解 x^*, $f(x)$ 是 D 上的严格凸函数, 则 x^* 是问题的唯一解.

有下面的推论成立:

(1) 考虑不定式约束问题

$$\begin{cases} \min & f(x), \quad x\in\mathbf{R}^n \\ \text{s.t.} & c_i(x)\leqslant 0, \quad i=1,\cdots,p, \end{cases} \tag{1.19}$$

其目标函数 $f(x)$ 和约束条件 $c_i(x)(i=1,\cdots,p)$都是$\mathbf{R}^N$ 上的凸函数, 则定理 1.4 的结论成立.

(2) 考虑无约束问题

$$\min \quad f(x), \qquad x\in\mathbf{R}^n. \tag{1.20}$$

设其目标函数 $f(x)$ 是 $\mathbf{R}^N$ 上的凸函数, 则定理 1.4 的结论成立.

1.2 最优性条件

考虑约束最优化问题

$$\begin{cases} \min & f(x) \\ \text{s.t.} & g_i(x)\leqslant 0\ , \quad i=1,2,\cdots,m, \\ & h_i(x)=0, \quad j=1,2,\cdots,r, \end{cases} \tag{1.21}$$

其中, $f:\mathbf{R}^n\to\mathbf{R}^1$, $g_i:\mathbf{R}^n\to\mathbf{R}^1(i=1,2,\cdots,m)$, $h_j:\mathbf{R}^n\to\mathbf{R}^1(j=1,2,\cdots,r)$. 设

$$R=\{x:g_j(x)\leqslant 0,i=1,\cdots,m;h_j(x)=0,j=1,\cdots,r;x\in\mathbf{R}^n\}.$$

现在来推导问题 (1.21) 在一般假设条件下的最优性条件.

定义 1.7 　若问题 (1.21) 的一个可行点 $\bar{x}$ 使某个不等式约束 $g_i(x)\leqslant 0$ 变成等式, 即 $g_i(\bar{x})=0$, 则称该不等式约束为关于可行 $\bar{x}$ 起作用约束 (或紧约束); 否则, 若 $\bar{x}$ 使得某个 $g_i(\bar{x})<0$, 则该不等式称为关于可行点 $\bar{x}$ 的不起作用约束 (或松约束). 用集合

$$I=\{i:gi(\bar{x})=0,\bar{x}\in R\} \tag{1.22}$$

表示在可行点 $\bar{x}$ 的起作用约束的指标集.

显然, 等式约束 $h_j(x)=0$ $(j=1,\cdots,r)$ 可看作是关于任一可行点 $\bar{x}\in R$ 的起作用约束.

1.2.1　几何最优性条件

先讨论有约束最优化问题的几何最优性条件.

定理 1.5 　设 x^* 是问题 (1.21) 可行点, f 和 $g_i(i\in I)$ 在 x^* 可微, $g_i(i\notin I)$ 在 x^* 连续, $h_j(j=1,\cdots,r)$ 在 x^* 连续可微; $\nabla h_j(x^*)(j=1,\cdots,r)$ 线性无关. 如果 x^* 是问题的局部极小点, 则

$$F_0\cap G_0\cap H_0=\varnothing, \tag{1.23}$$

其中

$$F_0=\left\{d:\nabla f(x^*)^{\mathrm{T}}d<0\right\}, \tag{1.24}$$

$$G_0=\left\{d:\nabla g(x^*)^{\mathrm{T}}d<0,\forall i\in I\right\}, \tag{1.25}$$

$$H_0=\left\{d:\nabla h_j(x^*)^{\mathrm{T}}d=0,j=1,\cdots,r\right\}, \tag{1.26}$$

$\varnothing$ 表示空集.

这一定理给出的有约束极小问题 (1.21) 的最优性条件 (1.23) 是用点集形式表示的, 所以称为几何最优性条件. 这一定理给出的几何条件仅仅是必要的, 而不是充分的. 几何最优性条件虽然直观易懂, 但实际计算中并不好用. 以下讨论的 Fritz John 条件和 KKT 条件是较常用的代数最优条件.

1.2.2　Fritz John 条件

下面把约束最优化问题 (1.21) 的最优条件表示成代数形式.

定理 1.6 (Fritz John 必要条件)　设 x^* 是问题 (1.21) 的可行解, $f,g_i(i\in I)$ 在 x^* 可微, $g_i(i\notin I)$ 在 x^* 连续, $h_j(j=1,\cdots,r)$ 在 x^* 连续可微. 如果 x^* 是问题的局

部极小点, 则存在不全为零的 $u_0, u_i(i \in I)$ 和 $v_j(j=1,\cdots,r)$, 使得

$$\begin{cases} u_0\nabla f(x^*) + \sum_{i\in I} u_i \nabla g_i(x) + \sum_{j=1}^{r} v_j \nabla h_j(x^*) = 0, \\ u_0, u_i \geqslant 0, \qquad i \in I. \end{cases} \tag{1.27}$$

如果 $g_i(i \notin I)$ 在 x^* 也可微, 则上述条件可写成: 存在不全为零的 $u_0, u_i(i=1,\cdots,m)$ 和 $v_j(j=1,\cdots,r)$, 使得

$$\begin{cases} u_0\nabla f(x^*) + \sum_{i\in I} u_i \nabla g_i(x) + \sum_{j=1}^{r} v_j \nabla h_j(x^*) = 0, \\ u_i g_i(x^*) = 0, \qquad i = 1,\cdots,m, \\ u_0, u_i \geqslant 0, \qquad i = 1,\cdots,m. \end{cases} \tag{1.28}$$

式 (1.27) 或式 (1.28) 称为 Fritz John 条件, 满足 Fritz John 条件的点称为 Fritz John 点, 简称为 F-J 点. 定理 1.6 说明局部极小点必须是 Fritz John 点, 但反之却未必真.

1.2.3 KKT 条件

在 Fritz John 条件中, 当 $u_0 = 0$ 时, 目标函数梯度 $\nabla f(x^*)$ 在条件中消失, 而条件仅仅表明极小点处起作用约束的梯度与等式约束的梯度线性相关, 这对于寻找极小点没有什么意义. 因此, 当 $u_0 = 0$ 时 Fritz John 条件失效. 为了保证 $u_0 > 0$, 要对约束再增加一些限制条件, 这样的附加条件通常称为约束规格. 约束规格有多种形式, 下述定理所加的约束规格是起作用约束的梯度与等式约束的梯度线性无关. 这样便得到著名的 Karush-Kuhn-Tucker(KKT) 条件.

定理 1.7 (KKT 必要条件) 设 x^* 是问题 (1.21) 的可行解, $f, g_i(i \in I)$ 在 x^* 可微, $g_i(i \notin I)$ 在 x^* 连续, $h_j(j=1,\cdots,r)$ 在 x^* 连续可微. 再假设 $\nabla g_i(x^*)(i \in I)$ 和 $\nabla h_j(x^*)(j=1,\cdots,r)$ 线性无关, 如果 x^* 是问题 (1.21) 的局部极小点, 则存在不全为零的 $u_i(i \in I)$ 和 $v_j(j=1,\cdots,r)$, 使得

$$\begin{cases} u_0\nabla f(x^*) + \sum_{i\in I} u_i \nabla g_i(x) + \sum_{j=1}^{r} v_j \nabla h_j(x^*) = 0, \\ u_i \geqslant 0, \qquad i \in I. \end{cases} \tag{1.29}$$

如果 $g_i(i \notin I)$ 在 x^* 也可微, 则上述条件可写成: 存在 $u_i(i=1,\cdots,m)$ 和