Foundation of Optimization
最优化原理

李晓鹏
Li Xiao-Peng

电子与信息工程学院
School of Electronic and Information Engineering

Semester–I 24/25

目录

目录

引言

- 最优化问题无处不在
- 如何求解最优化问题？
- 基本又常用的最优化方法有哪些？
- 如何学习和掌握这些优化方法？

最优化问题无处不在

在日常生活中，大家经常遇到下面一类问题：

- 完成一项工作或任务，怎样合理安排时间才能做到既完成任务又耗时最少（比如：同学周末一天安排，厨师在规定的时间内炒多个菜等）；

最优化问题无处不在

在日常生活中，大家经常遇到下面一类问题：

- 完成一项工作或任务，怎样合理安排时间才能做到既完成任务又耗时最少（比如：同学周末一天安排，厨师在规定的时间内炒多个菜等）；
- 理财投资过程中，如何合理分配持有资金的的比例，才能既达到回报最高且风险较低的目标；

最优化问题无处不在

在日常生活中，大家经常遇到下面一类问题：

- 完成一项工作或任务，怎样合理安排时间才能做到既完成任务又耗时最少（比如：同学周末一天安排，厨师在规定的时间内炒多个菜等）；
- 理财投资过程中，如何合理分配持有资金的的比例，才能既达到回报最高且风险较低的目标；
- 城建规划中，怎样安排工厂、机关、学校、超市、医院等其它单位的合理布局，才能既方便群众又利于城市各行业的发展；

最优化问题无处不在

在日常生活中，大家经常遇到下面一类问题：

- 完成一项工作或任务，怎样合理安排时间才能做到既完成任务又耗时最少（比如：同学周末一天安排，厨师在规定的时间内炒多个菜等）；
- 理财投资过程中，如何合理分配持有资金的的比例，才能既达到回报最高且风险较低的目标；
- 城建规划中，怎样安排工厂、机关、学校、超市、医院等其它单位的合理布局，才能既方便群众又利于城市各行业的发展；
- 生产中，制定怎样的生产计划方案，使得能耗和成本降低、利润提高（比如：码头卸货、汽车制造等）

最优化问题无处不在

在日常生活中，大家经常遇到下面一类问题：

- 完成一项工作或任务，怎样合理安排时间才能做到既完成任务又耗时最少（比如：同学周末一天安排，厨师在规定的时间内炒多个菜等）；
- 理财投资过程中，如何合理分配持有资金的的比例，才能既达到回报最高且风险较低的目标；
- 城建规划中，怎样安排工厂、机关、学校、超市、医院等其它单位的合理布局，才能既方便群众又利于城市各行业的发展；
- 生产中，制定怎样的生产计划方案，使得能耗和成本降低、利润提高（比如：码头卸货、汽车制造等）
- ……

上述问题的共同点

- **最优化问题**：都属于寻求最优方案(决策)的问题，即在所有可能的方案中，寻找最合理的、能够使事先规定的目标达到最优的方案这样一类问题。
- **最优方案**：在所有可行的方案中，最合理的、能够达到事先规定最优目标的方案。
- **求解最优化问题的关键**：寻求最优方案.

最优化问题无处不在

如何寻求最优方案？

- 需要借助一些方法，这些方法统称最优化方法.
- 最优化方法: 寻求最优方案的方法.

本学期重点内容

1. 基本又常用的最优化方法有哪些?
2. 每种算法的产生背景和适用范围是什么?
3. 各自有什么优缺点?
4. 如何正确应用他们求解一些优化问题?

对以上问题的解答构成了我们这学期的重点内容: 最优化理论与方法

课程性质和授课特点

课程性质：这门课是建立在高等数学和线性代数的基础上，主要解决最优计划、最优分配、最优决策、最佳设计、最佳管理等最优化问题的一门非常实用的**数学**课程。

授课特点：先介绍一种方法 (包括该方法由来、适用范围、算法内容、优缺点等)，然后举例演示该方法求解具体的数学模型(问题)的详细步骤。

教材及主要参考资料

1. 《实用最优化方法》 (第三版), 唐焕文等, 大连理工大学出版社,2004.
2. 《最优化方法及其应用》, 郭科等, 高等教育出版社, 2006.
3. 《最优化方法及其MATLAB程序设计》, 马昌凤编著, 科学出版社,2009.
4. 《最优化方法》 (第二版), 施光燕等, 高等教育出版社, 2007.
5. ......

基本概念

- **最优方案:** 在所有可行的方案中，最合理的、达到事先规定的最优目标的方案.
- **最优化方法:** 寻找最优方案的方法.

问题: 在实际应用中，遇到最优化问题，如何寻找最优化方法求解他们呢?

求解最优化问题的步骤

1. **建立数学模型:**
   对所要解决的实际问题进行分析和研究，加以抽象、简化，形成数学上的最优化问题(数学模型)。
2. **进行数学加工和求解:**
   对已得到的数学模型进行整理和变换，变成易求解的形式，然后根据模型特点**选择或设计**合适的最优化算法，最后通过编程或利用数学软件求解、分析结果。

求解最优化问题的步骤

例1.1：生产问题 某企业用甲、乙两种原料生产A，B，C三种产品，已知生产每种产品一吨所需原料(吨)和每天原料的最大供应量(吨)及每吨不同产品可获利润(千元/吨)情况如表1所示。

产品 单位消耗 原料	A	B	C	原料总供应量（吨）
甲	1	2	2	100
乙	3	1	3	100
单位利润（万元）	4	3	7	

试问，该企业怎样安排生产，才会使每天的利润最大?

求解最优化问题的步骤

解：设该企业生产产品 A， B， C 分别为 x_1，x_2，x_3 吨，
则总利润的表达式为：

$$f := 4x_1 + 3x_2 + 7x_3$$

我们希望在现有资源条件下总利润最大。现有资源的限制为

$$x_1 + 2x_2 + 2x_3 \leq 100 \text{ (原料甲的限制)}$$

$$3x_1 + x_2 + 3x_3 \leq 100 \text{ (原料乙的限制)}$$

此外，由于未知数（我们称为决策变量）x_1，x_2，x_3 是计划产量，故 x_1，x_2，x_3 均非负，即

$$x_j \geq 0, j = 1, 2, 3$$

求解最优化问题的步骤

此问题的数学模型可以表述为:

$$\begin{aligned} \max\ f &:= 4x_1 + 3x_2 + 7x_3 \\ s.t.\ x_1 + 2x_2 + 2x_3 &\le 100 \\ 3x_1 + x_2 + 3x_3 &\le 100 \\ x_j \ge 0, j &= 1,2,3 \end{aligned}$$

其中未知数 x_1，x_2，x_3 称为决策变量；

函数 $f := 4x_1 + 3x_2 + 7x_3$称为目标函数；

s.t. 为英文subject to的缩写，表示决策变量 $x_j \ge 0, j = 1,2,3$ 受它后面的条件的约束，后面的条件称为约束条件。

例1.2：运输问题 一个企业有若干个生产基地与销售站点，如何制定调运方案，使某种一定量的产品从若干个产地运到若干个销售地的总的运费最小?

设某建材公司有三个水泥厂A1，A2， A3，四个销售点B1，B2，B3，B4。其产量、销量、运费（元/吨）见下表

销售点 单位运费 生产地	B1	B2	B3	B4	产量（吨）
A1	8	7	3	2	2000
A2	4	7	5	1	10000
A3	2	4	9	6	4000
销售量（吨）	3000	2000	4000	5000	

如何制定调运方案，使总的运费最小?

求解最优化问题的步骤

解：设由生产基地 $A_i(i=1,2,3)$ 运到销售地 $B_j(j=1,2,3,4)$ 的货运量为 x_{ij}，则相应的总运费为：

$$\begin{aligned} f :=& 8x_{11}+7x_{12}+3x_{13}+2x_{14}+4x_{21}+7x_{22}+5x_{23}+x_{24} \\ &+2x_{31}+4x_{32}+9x_{33}+6x_{34} \end{aligned}$$

我们的目标是在现有条件使总运费最小，即

$$\begin{aligned} \min f :=& 8x_{11}+7x_{12}+3x_{13}+2x_{14}+4x_{21}+7x_{22}+5x_{23}+x_{24} \\ &+2x_{31}+4x_{32}+9x_{33}+6x_{34} \end{aligned}$$

目前的限制条件有哪些呢？

求解最优化问题的步骤

现有的限制条件有:

$$x_{11}+x_{12}+x_{13}+x_{14}\le 2000\ \text{(A1的生产能力限制)}$$
$$x_{21}+x_{22}+x_{23}+x_{24}\le 10000\ \text{(A2的生产能力限制)}$$
$$x_{31}+x_{32}+x_{33}+x_{34}\le 4000\ \text{(A3的生产能力限制)}$$

以及

$$x_{11}+x_{21}+x_{31}\ge 3000\ \text{(B1的销售额限制)}$$
$$x_{12}+x_{22}+x_{32}\ge 2000\ \text{(B2的销售额限制)}$$
$$x_{13}+x_{23}+x_{33}\ge 4000\ \text{(B3的销售额限制)}$$
$$x_{14}+x_{24}+x_{34}\ge 5000\ \text{(B4的销售额限制)}$$

另外，货运量应为非负整数，故

$$x_{ij}\ge 0, i=1,2,3; j=1,2,3,4$$

且为整数。

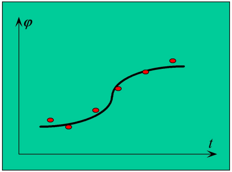
综上可得原问题对应的数学模型为：

$$
\begin{aligned}
\min\ f := {} & 8x_{11} + 7x_{12} + 3x_{13} + 2x_{14} + 4x_{21} + 7x_{22} \\
& + 5x_{23} + x_{24} + 2x_{31} + 4x_{32} + 9x_{33} + 6x_{34} \\
s.t.\ & x_{11} + x_{12} + x_{13} + x_{14} \leq 2000 \\
& x_{21} + x_{22} + x_{23} + x_{24} \leq 10000 \\
& x_{31} + x_{32} + x_{33} + x_{34} \leq 4000 \\
& x_{11} + x_{21} + x_{31} \geq 3000 \\
& x_{12} + x_{22} + x_{32} \geq 2000 \\
& x_{13} + x_{23} + x_{33} \geq 4000 \\
& x_{14} + x_{24} + x_{34} \geq 5000 \\
& x_{ij} \geq 0, i = 1, 2, 3; j = 1, 2, 3, 4 \text{且为整数}
\end{aligned}
$$

例1.3：曲线拟合问题 已知某物体的温度 φ 与时间 t 之间有如下形式的经验函数关系：

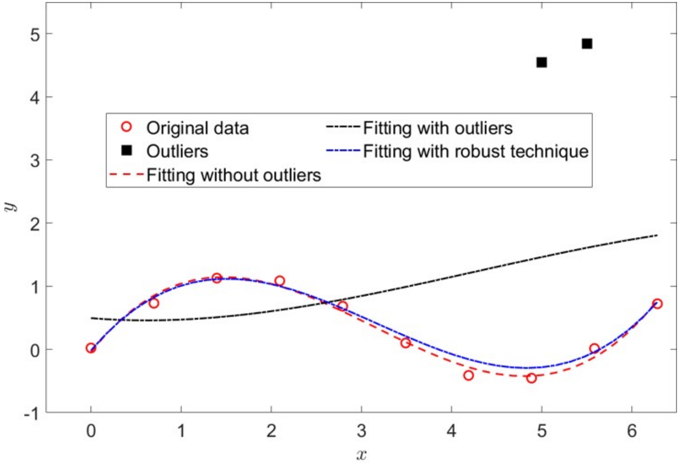
$$\varphi = c_1 + c_2 t + e^{c_3 t}$$

其中，c_1，c_2，c_3是待定参数。现通过测试获得 n 组 φ 与 t 之间的实验数据 $(t_i, \varphi_i), i = 1, ..., n$。试确定参数 c_1，c_2，c_3使得理论曲线尽可能地与n个测试点拟合。



$$\min \sum_{i=1}^{n} [\varphi_i - (c_1 + c_2 t_i + e^{c_3 t_i})]^2$$

示例：$y = \sin(x) + 0.1x$


Original data
Outliers
Fitting without outliers
Fitting with outliers
Fitting with robust technique
5
4
3
2
1
0
-1
0
1
2
3
4
5
6
x
y

求解最优化问题的步骤

例1.5：投资决策问题 设一段时间内，有B万元的资金可用于投资，有m个项目 $A_1, A_2, ..., A_m$ 可供挑选。若对项目 A_i 进行投资，花费的资金为 a_i 万元，可获得收益 c_i 万元，试确定最佳投资方案。
解：引入变量

$$x_i = \begin{cases} 1, \text{若对}A_i\text{投资，} \\ 0, \text{若对}A_i\text{不投资,} \end{cases} \quad i = 1, 2, \cdots, m$$

最佳投资方案应当满足：投资少，收益大。

求解最优化问题的步骤

由此得到问题的数学模型为:

$$\begin{aligned}
\min\ f_1(x_1, x_2, \cdots, x_m) &= \sum_{i=1}^{m} a_i x_i \\
\max\ f_2(x_1, x_2, \cdots, x_m) &= \sum_{i=1}^{m} c_i x_i \\
s.t.\ \sum_{i=1}^{m} a_i x_i &\le B \\
x_i \in \boldsymbol{I} = \{0, 1\}, i &= 1, 2, \cdots, m
\end{aligned}$$

模型分类

根据求解模型所用方法不同，数学模型分为:

初等模型、简单优化模型、数学规划模型、组合优化模型、统计回归模型、动态规划模型、图论与网络流模型等。

- 这学期主要学习求解简单优化模型和数学规划模型的一些最优化方法(数值解法).

数学规划模型（约束优化模型）

定义：
在变量受一些等式或不等式的约束下，求目标函数的极大（或极小）的优化模型。

三要素: 决策变量、约束条件和目标函数

【注】 上述定义是对目前应用较多的带约束条件的数学规划模型的一个描述，很多教材上把它称为约束数学规划模型; 对于不含约束条件的则称为无约束数学规划模型。

最优化问题分类

1. **无约束优化模型**

$$\max f\left(\boldsymbol{x}\right) \quad \left(\text{或}\min f\left(\boldsymbol{x}\right)\right) \quad \boldsymbol{x} \in \mathbb{R}^n$$

2. **约束优化模型（约束数学规划）**

$$\begin{aligned} \max\ f\left(\boldsymbol{x}\right) \quad & \left(\text{或}\min f\left(\boldsymbol{x}\right)\right) \quad \boldsymbol{x} \in \mathbb{R}^n \\ s.t.\ h_i(\boldsymbol{x}) &= 0, i = 1, 2, \cdots, l \\ g_j(\boldsymbol{x}) &\geq 0, j = 1, 2, \cdots, m \end{aligned}$$

约束优化模型的分类

1. 根据目标函数和约束函数是否全是线性函数，分为线性规划和非线性性规划.
   - **线性规划：** $f(\boldsymbol{x})$, $h_i(\boldsymbol{x})$, $g_j(\boldsymbol{x})$ 都是线性函数;
   - **非线性性规划：** $f(\boldsymbol{x})$, $h_i(\boldsymbol{x})$, $g_j(\boldsymbol{x})$ 不全是线性函数;

   例如：上一节例1.1: 线性规划，例1.3: 非线性性规划.
2. 根据决策变量的是否要求全取整数，分为整数规划和混合规划
   - **整数规划：** 要求 $x_i \in Z, i = 1, 2, \cdots, n$的数学规划模型.
     特殊地，若要求决策变量的分量只取0或1，这样的整数规划又称0–1规划.
   - **混合规划：** x_i 可以取整数也可以取分数的数学规划.

   例如：上一节例1.2: 整数规划；例1.5: 0–1规划。

与约束优化模型相关的几个概念

$$\begin{aligned} &\min\ f(\boldsymbol{x}), \boldsymbol{x} \in \mathbb{R}^n \\ &s.t.\ h_i(\boldsymbol{x}) = 0, i = 1, 2, \cdots, l \\ &\qquad g_j(\boldsymbol{x}) \geq 0, j = 1, 2, \cdots, m \end{aligned}$$

1. **可行解（点）：** 满足所有约束条件的点.
2. **可行集（域）：** 所有可行解（点）组成的集合，一般记为 $\boldsymbol{S}$，即

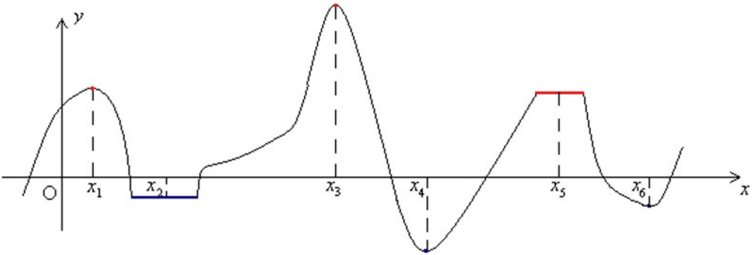
$$\boldsymbol{S} = \{\boldsymbol{x} \in \mathbb{R}^n \,|h_i(\boldsymbol{x}) = 0, g_j(\boldsymbol{x}) \geq 0, i = 1, 2, \cdots, l; j = 1, 2, \cdots, m\}$$

3. **全局最优解（点）：**
   若存在 $\boldsymbol{x}^* \in \boldsymbol{S}$，使得对一切 $\boldsymbol{x} \in \boldsymbol{S}$ 恒有 $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$ ，则称 $\boldsymbol{x}^*$ 为该最优化问题的全局（整体）最优解（点）。
   若 $\boldsymbol{x} \in \boldsymbol{S}, \boldsymbol{x} \neq \boldsymbol{x}^*$，恒有 $f(\boldsymbol{x}^*) < f(\boldsymbol{x})$，则称 $\boldsymbol{x}^*$ 为最优化问题的严格全局（整体）最优解。

与约束优化模型相关的几个概念

4. **局部最优解（点）：** 若存在$\boldsymbol{x}^* \in \boldsymbol{S}$，存在 $\boldsymbol{x}^*$的某邻域$N_\delta(\boldsymbol{x}^*)$，使得对于一切 $\boldsymbol{x} \in \boldsymbol{S} \bigcap N_\delta\left(\boldsymbol{x}^*\right)$恒有$f\left(\boldsymbol{x}^*\right) \leq f\left(\boldsymbol{x}\right)$，则称$\boldsymbol{x}^*$为最优化问题的局部优解.
   同样也有: 严格局部最优解.
5. **最优值：** 最优解（点）对应的目标函数值 $f(\boldsymbol{x}^*)$.

与约束优化模型相关的几个概念



注意:

- 显然，整体最优解一定是局部最优解，而局部最优解不一定是整体最优解。
- 求解最优化问题，理论上是求可行域S上的整体最优解。但在一般情况下，整体最优解是很难求出的，往往只能求出局部最优解。

补充内容

说明

- 线性函数：函数图象为直线的函数。
- 初等数学中，线性函数指只拥有一个变量的一阶多项式函数。
  形如：$f(x) = ax + b$；其中 a, b 为实数.
- 在高等数学里，线性函数是一个线性变换，是在两个向量空间之间，维持向量加法与数量运算的映射.
  形如：$f(\boldsymbol{x}) = \boldsymbol{M}\boldsymbol{x} + \boldsymbol{b}$，其中 $\boldsymbol{M}$ 是矩阵， $\boldsymbol{b}$ 是列向量.

目录

多元函数的梯度及性质

定义 2.1:

设 $f:\mathbb{R}^n\to\mathbb{R}, \bar{\boldsymbol{x}}\in\mathbb{R}^n$, 如果 $f(\boldsymbol{x})$ 在点 $\bar{\boldsymbol{x}}$ 处关于变量 $\boldsymbol{x}$ 的各分量的偏导数都存在，则称 $f(\boldsymbol{x})$ 在点 $\bar{\boldsymbol{x}}$ 处一阶可导，并称向量

$$\nabla f\left(\bar{\boldsymbol{x}}\right)=\left(\frac{\partial f\left(\bar{\boldsymbol{x}}\right)}{\partial x_1},\frac{\partial f\left(\bar{\boldsymbol{x}}\right)}{\partial x_2},\cdots,\frac{\partial f\left(\bar{\boldsymbol{x}}\right)}{\partial x_n}\right)^T$$

为 $f(\boldsymbol{x})$ 在点 $\bar{\boldsymbol{x}}$ 处的一阶导数或梯度.

多元函数的梯度及性质

如果 $f(\boldsymbol{x})$ 在任意点 $\bar{\boldsymbol{x}} \in \mathbb{R}^n$ 处一阶可导的，则称下面的向量函数：

$$\nabla f(\boldsymbol{x}) = \left(\frac{\partial f(\boldsymbol{x})}{\partial x_1}, \frac{\partial f(\boldsymbol{x})}{\partial x_2}, \cdots, \frac{\partial f(\boldsymbol{x})}{\partial x_n} \right)^T$$

为 $f(\boldsymbol{x})$ 的梯度函数，也称为一阶导函数.

可见：求梯度的关键是求偏导 $\frac{\partial f(\boldsymbol{x})}{\partial x_i}$。

多元函数的梯度及性质

定理 2.1:

若函数 $f(\boldsymbol{x})$ 定义在 $\boldsymbol{\Omega} \subseteq \mathbb{R}^n$，并具有一阶连续偏导数， $\boldsymbol{x}^*$ 是 $\boldsymbol{\Omega}$ 的一个内点，若 $\boldsymbol{x}^*$ 是 $f(\boldsymbol{x})$ 的极值点，则

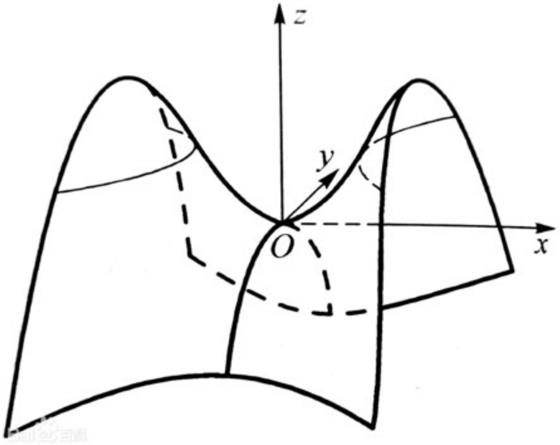
$$\nabla f\left(\boldsymbol{x}^*\right)=\mathbf{0}$$

注意: 此定理是极值点的一阶必要条件.

定义 2.2:

满足 $\nabla f\left(\boldsymbol{x}^*\right)=0$ 的点 $\boldsymbol{x}^*$ 称为 $f(\boldsymbol{x})$ 的稳定点或驻点。

多元函数的梯度及性质



例2.1 考虑求解如下问题

$$\min f(\boldsymbol{x}), \boldsymbol{x} \in \mathbb{R}^n$$

设 $\boldsymbol{x}^k$ 为第 k 步的迭代点。

问题： 判断该点是否是极值点？

例2.1 考虑求解如下问题

$$\min f(\boldsymbol{x}), \boldsymbol{x} \in \mathbb{R}^n$$

设 $\boldsymbol{x}^k$ 为第 k 步的迭代点。

问题： 判断该点是否是极值点？

方法： 根据定理，若 $f(\boldsymbol{x})$ 具有一阶连续偏导数，那么只需检验

$$\left\| \nabla f(\boldsymbol{x}^k) \right\|_2 = \mathbf{0}$$

是否成立即可。

下降方向

定义2.3：

设函数$f:\boldsymbol{\Omega}\subseteq\mathbb{R}^n\to\mathbb{R}$ 在点$\boldsymbol{x}^0$处可导，$\boldsymbol{p}\in\mathbb{R}^n$为$n$维非零向量，如果存在实数$\delta>0$，使得任意的$\alpha\in(0,\delta)$ 有：

$$f\left(\boldsymbol{x}^0+\alpha\boldsymbol{p}\right)<f\left(\boldsymbol{x}^0\right)$$

则称$\boldsymbol{p}$为$f(\boldsymbol{x})$在$\boldsymbol{x}^0$处的一个下降方向。

即$f(\boldsymbol{x})$从$\boldsymbol{x}^0$出发沿 $\boldsymbol{p}$方向函数值是下降的。

下降方向

定理 2.2:

设函数$f:\boldsymbol{\Omega}\subseteq\mathbb{R}^n\to\mathbb{R}$在点$\boldsymbol{x}^k$处一阶可导， $\boldsymbol{p}\in\mathbb{R}^n$为$n$维非零向量，如果满足:

$$\nabla f\left(\boldsymbol{x}^k\right)^T\boldsymbol{p}<0$$

则 $\boldsymbol{p}$ 必为 $f(\boldsymbol{x})$ 在 $\boldsymbol{x}^k$处的下降方向。

定理 2.3:

梯度的反方向(负梯度方向)$-\nabla f(\boldsymbol{x}^k)$ 是 $f(\boldsymbol{x})$从$\boldsymbol{x}^k$出发下降最快的方向。因此，也称$-\nabla f(\boldsymbol{x}^k)$为该点处的最速下降方向。

下降方向

例2.2 考虑求解如下问题

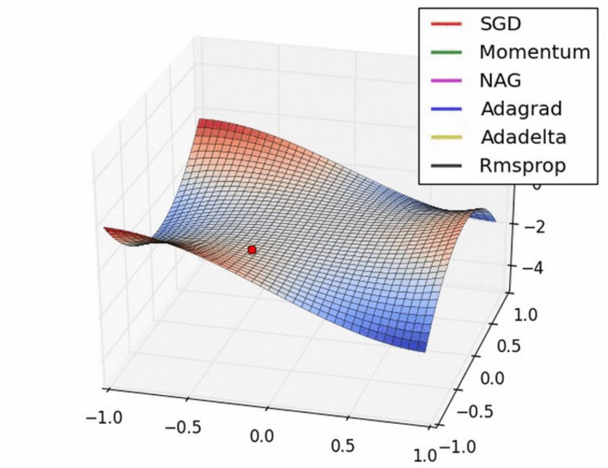
$$\min f(\boldsymbol{x}), \boldsymbol{x} \in \mathbb{R}^n$$

设$\boldsymbol{x}^k$为第k步的迭代点，它不是极小点。

对于该问题，为了尽快求得最优解，在进行下一步迭代前选取搜索方向时，总希望取或接近于负梯度$-\nabla f(\boldsymbol{x}^k)$ 的方向，使函数值尽快地下降到目标函数的最低点。

下降方向

最速真的“最速”吗？



下降方向

例2.3 求 $f(\boldsymbol{x}) = x_1^2 + x_2^2 + 1$ 在 $\boldsymbol{x}^0 = (0, 3)^T$ 处的最速下降方向并求沿该方向移动一个单位后得到的新迭代点以及对应的函数值.

下降方向

例2.3 求 $f(\boldsymbol{x}) = x_1^2 + x_2^2 + 1$ 在 $\boldsymbol{x}^0 = (0,3)^T$ 处的最速下降方向并求沿该方向移动一个单位后得到的新迭代点以及对应的函数值.

解:

$$\nabla f(\boldsymbol{x}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}\right)^T = (2x_1, 2x_2)^T$$

故所求最速下降方向为: $-\nabla f\left(\boldsymbol{x}^0\right) = (0,-6)^T$

将该向量单位化得: $\boldsymbol{p} = \frac{-\nabla f\left(\boldsymbol{x}^0\right)}{\|\nabla f(\boldsymbol{x}^0)\|} = (0,1)^T$

故所求新的迭代点为:

$$\boldsymbol{x}^1 = \boldsymbol{x}^0 + \boldsymbol{p} = (0,3)^T + (0,-1)^T = (0,2)^T$$

对应的函数值为:

$$f(\boldsymbol{x}^1) = 0^2 + 2^2 + 1 = 5$$

多元函数的Hessian矩阵及性质

定义2.4:

设 $f:\mathbb{R}^n \to \mathbb{R}, \bar{\boldsymbol{x}} \in \mathbb{R}^n$，如果$f(\boldsymbol{x})$在点$\bar{\boldsymbol{x}}$处关于变量 $\boldsymbol{x}$ 的各分量的二阶偏导数都存在，则称$f(\boldsymbol{x})$ 在点$\bar{\boldsymbol{x}}$处二阶可导，并称如下矩阵：

$$\nabla^2 f(\bar{\boldsymbol{x}}) = \left(\frac{\partial^2 f(\bar{\boldsymbol{x}})}{\partial x_i \partial x_j}\right)_{n\times n}$$

即

$$\nabla^2 f(\bar{\boldsymbol{x}}) = \begin{pmatrix} \frac{\partial^2 f(\bar{\boldsymbol{x}})}{\partial x_1^2} & \frac{\partial^2 f(\bar{\boldsymbol{x}})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\bar{\boldsymbol{x}})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\bar{\boldsymbol{x}})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\bar{\boldsymbol{x}})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\bar{\boldsymbol{x}})}{\partial x_2 \partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f(\bar{\boldsymbol{x}})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\bar{\boldsymbol{x}})}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(\bar{\boldsymbol{x}})}{\partial x_n^2} \end{pmatrix}$$

为 $f(\boldsymbol{x})$ 在点$\bar{\boldsymbol{x}}$处的 Hessian 矩阵或二阶导数。

多元函数的Hessian矩阵及性质

说明：求Hessian矩阵的关键是求二阶偏导数 $\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j}$

若 $f(\boldsymbol{x})$ 的所有二阶偏导数在点， $\bar{\boldsymbol{x}}$ 处都连续, 则

$$\frac{\partial^2 f(\bar{\boldsymbol{x}})}{\partial x_i \partial x_j} = \frac{\partial^2 f(\bar{\boldsymbol{x}})}{\partial x_j \partial x_i}$$

此时，该点对应的Hessian矩阵是一个对称矩阵

例2.4 求 $f(\boldsymbol{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$ 在 $\boldsymbol{x}^0 = (0,0)^T$ 处的梯度和 Hessian 矩阵。

多元函数的Hessian矩阵及性质

例2.4 求$f(\boldsymbol{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$ 在 $\boldsymbol{x}^0 = (0, 0)^T$ 处的梯度和 Hessian 矩阵。

解:

$$
\begin{aligned}
\frac{\partial f}{\partial x_1} &= 200(x_2 - x_1^2)(-2x_1) + 2(1 - x_1)(-1) \\
&= -400x_1(x_2 - x_1^2) - 2(1 - x_1) \\
\frac{\partial f}{\partial x_2} &= 200(x_2 - x_1^2)
\end{aligned}
$$

故: $\nabla f(\boldsymbol{x}^0) = (-2, 0)^T$

例2.5 求 $f(\boldsymbol{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$ 在 $\boldsymbol{x}^0 = (0,0)^T$ 处的梯度和 Hessian 矩阵。

解:

$$\frac{\partial^2 f}{\partial x_1^2} = 400(x_1^2 - x_2) + 800x_1^2 + 2$$

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial^2 f}{\partial x_2 \partial x_1} = -400x_1$$

$$\frac{\partial^2 f}{\partial x_2^2} = 200$$

故: $\nabla^2 f(\boldsymbol{x}^0) = \begin{pmatrix} 2 & 0 \\ 0 & 200 \end{pmatrix}$

多元函数的Hessian矩阵及性质

例2.6 设$\boldsymbol{a} \in \mathbb{R}^n, \boldsymbol{x} \in \mathbb{R}^n, b \in \mathbb{R}$, 求$f(\boldsymbol{x}) = \boldsymbol{a}^T\boldsymbol{x} + b$在任意点$\boldsymbol{x}$处的梯度和Hessian矩阵.

例2.6 设$\boldsymbol{a} \in \mathbb{R}^n, \boldsymbol{x} \in \mathbb{R}^n, b \in \mathbb{R}$，求$f(\boldsymbol{x}) = \boldsymbol{a}^T\boldsymbol{x} + b$在任意点$\boldsymbol{x}$处的梯度和Hessian矩阵.

解： 设$\boldsymbol{a} = (a_1, a_2, \cdots, a_n)^T, \boldsymbol{x} = (x_1, x_2, \cdots, x_n)^T$

则： $f(\boldsymbol{x}) = a_1x_1 + a_2x_2 + \cdots + a_nx_n + b$

故： $\frac{\partial f}{\partial x_i} = a_i, i = 1, 2, \cdots, n$

即： $\nabla f(\boldsymbol{x}) = (a_1, a_2, \cdots, a_n)^T = \boldsymbol{a}$

例2.6 设$\boldsymbol{a} \in \mathbb{R}^n, \boldsymbol{x} \in \mathbb{R}^n, b \in \mathbb{R}$，求$f(\boldsymbol{x}) = \boldsymbol{a}^T\boldsymbol{x} + b$在任意点$\boldsymbol{x}$处的梯度和Hessian矩阵.

解： 又由$\frac{\partial f}{\partial x_i} = a_i, i = 1, 2, \cdots, n$可求得：

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = 0, \quad i, j = 1, 2, \cdots, n$$

故Hessian矩阵为：$\nabla^2 f(\boldsymbol{x}) = \mathbf{0}$为零矩阵

例2.7 设 $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$为实对称矩阵，$\boldsymbol{b} \in \mathbb{R}^n$，$c \in \mathbb{R}$ 求二次函数$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x} + c$在任意点$\boldsymbol{x}$处的梯度和Hessian矩阵。

例2.7 设 $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$为实对称矩阵，$\boldsymbol{b} \in \mathbb{R}^n$， $c \in \mathbb{R}$ 求二次函数$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x} + c$在任意点$\boldsymbol{x}$处的梯度和Hessian矩阵。

解: 设 $\boldsymbol{Q} = (q_{ij})_{n\times n}, \boldsymbol{b} = (b_1, \cdots, b_n)^T, \boldsymbol{x} = (x_1, \cdots, x_n)^T$

则

$$
\begin{aligned}
f(\boldsymbol{x}) =& \frac{1}{2}\sum_{l=1}^{n}\sum_{j=1}^{n} q_{lj}x_l x_j + \sum_{j=1}^{n} b_j x_j + c \\
=& \frac{1}{2}(\sum_{l\neq i}\sum_{j\neq i} q_{ij}x_l x_j + \sum_{j\neq i} q_{ij}x_i x_j + \sum_{l\neq i} q_{li}x_l x_i + q_{ii}x_i^2) \\
&+ \sum_{j=1}^{n} b_j x_j + c
\end{aligned}
$$

多元函数的Hessian矩阵及性质

则

$$
\begin{aligned}
\frac{\partial f}{\partial x_i} &= \frac{1}{2}(\sum_{j\neq i} q_{ij}x_j + \sum_{l\neq i} q_{li}x_l + 2q_{ii}x_i) + b_i \\
&= \frac{1}{2}(\sum_{j\neq i} q_{ij}x_j + \sum_{j\neq i} q_{ji}x_j + 2q_{ii}x_i) + b_i \\
&= \sum_{j\neq i} q_{ij}x_j + q_{ii}x_i + b_i \\
&= \sum_{j=1}^{n} q_{ij}x_j + b_i, i = 1, 2, \cdots, n
\end{aligned}
$$

多元函数的Hessian矩阵及性质

即

$$
\begin{aligned}
\nabla f(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix} &= \begin{pmatrix} \sum_{j=1}^n q_{1j} x_j + b_1 \\ \vdots \\ \sum_{j=1}^n q_{nj} x_j + b_n \end{pmatrix} \\
&= \begin{pmatrix} \sum_{j=1}^n q_{1j} x_j \\ \vdots \\ \sum_{j=1}^n q_{nj} x_j \end{pmatrix} + \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \\
&= \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b}
\end{aligned}
$$

多元函数的Hessian矩阵及性质

又由

$$\frac{\partial f}{\partial x_i} = \sum_{j=1}^{n} q_{ij} x_j + b_i, \quad i = 1, 2, \cdots, n$$

得:

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = q_{ij}, \quad i = 1, 2, \cdots, n$$

故Hessian矩阵为: $\nabla^2 f(\boldsymbol{x}) = \boldsymbol{Q}$.

结论:

设 $f(\boldsymbol{x}) = \boldsymbol{a}^T\boldsymbol{x} + b$ 则:

$$\begin{aligned} \nabla f(\boldsymbol{x}) &= \boldsymbol{a}, \\ \nabla^2 f(\boldsymbol{x}) &= \boldsymbol{0}. \end{aligned}$$

设$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x} + c$ 则:

$$\begin{aligned} \nabla f(\boldsymbol{x}) &= \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b}, \\ \nabla^2 f(\boldsymbol{x}) &= \boldsymbol{Q} \end{aligned}$$

结论:

$$\frac{\partial y}{\partial \boldsymbol{x}} = \begin{pmatrix} \frac{\partial y}{\partial x_1} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{pmatrix}$$

$$\frac{\partial \boldsymbol{y}}{\partial x} = \left(\frac{\partial y_1}{\partial x} \cdots \frac{\partial y_n}{\partial x} \right)$$

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_2} \\ \vdots & \vdots & \vdots \\ \frac{\partial y_1}{\partial x_n} & \cdots & \frac{\partial y_m}{\partial x_n} \end{pmatrix}$$

结论:

$$
\begin{aligned}
\frac{\partial \boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} &= ? \\
\frac{\partial \boldsymbol{x}^T\boldsymbol{A}}{\partial \boldsymbol{x}} &= ? \\
\frac{\partial \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} &= ? \\
\frac{\partial \|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}\|_2^2}{\partial \boldsymbol{x}} &= ?
\end{aligned}
$$

结论:

$$
\begin{aligned}
\frac{\partial \boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} &= \boldsymbol{A}^T \\
\frac{\partial \boldsymbol{x}^T\boldsymbol{A}}{\partial \boldsymbol{x}} &= \boldsymbol{A} \\
\frac{\partial \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} &= (\boldsymbol{A} + \boldsymbol{A}^T)\boldsymbol{x} \\
\frac{\partial \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2}{\partial \boldsymbol{x}} &= \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y})
\end{aligned}
$$

定理2.4：

设 $f:\mathbb{R}^n\to\mathbb{R}, \boldsymbol{x}_0\in\mathbb{R}^n$。如果 $f(\boldsymbol{x})$ 在 $\boldsymbol{x}_0$ 的某邻域 $N_\delta(\boldsymbol{x}_0)$具有一阶连续偏导数，则:

$$\begin{aligned} f(\boldsymbol{x}) =& f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0)^T(\boldsymbol{x}-\boldsymbol{x}_0) \\ & + o\left(\|\boldsymbol{x}-\boldsymbol{x}_0\|\right), \quad \boldsymbol{x}\in N_\delta(\boldsymbol{x}_0) \end{aligned}$$

说明： 在上式中略去高阶无穷小量后，有

$$f(\boldsymbol{x}) \approx f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0)^T(\boldsymbol{x}-\boldsymbol{x}_0), \quad \boldsymbol{x}\in N_\delta(\boldsymbol{x}_0)$$

通常，称上式右端的函数

$$q_1(\boldsymbol{x}) = f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0)^T(\boldsymbol{x}-\boldsymbol{x}_0)$$

为$f(\boldsymbol{x})$在$\boldsymbol{x}_0$处的线性(一阶)近似函数。

二阶Taylor展式

定理2.5:

设$f:R^n\to\mathbb{R}, \boldsymbol{x}_0\in\mathbb{R}^n$。如果$f(\boldsymbol{x})$ 在 $\boldsymbol{x}_0$的某邻域$N_\delta(\boldsymbol{x}_0)$具有二阶连续偏导数，则：

$$
\begin{aligned}
f(\boldsymbol{x}) =& f(\boldsymbol{x}_0)+\nabla f(\boldsymbol{x}_0)^T(\boldsymbol{x}-\boldsymbol{x}_0)\\
&+\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_0)^T\nabla^2 f(\boldsymbol{x}_0)(\boldsymbol{x}-\boldsymbol{x}_0)\\
&+o\left(\|\boldsymbol{x}-\boldsymbol{x}_0\|^2\right),\quad \boldsymbol{x}\in N_\delta(\boldsymbol{x}_0)
\end{aligned}
$$

说明： 在上式中略去高阶无穷小量后，有

$$f(\boldsymbol{x}) \approx f(\boldsymbol{x}_0) + \nabla f(x_0)^T(\boldsymbol{x} - \boldsymbol{x}_0) \\ + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_0)^T \nabla^2 f(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0), \boldsymbol{x} \in N_\delta(\boldsymbol{x}_0).$$

通常，称上式右端的函数

$$q_2(\boldsymbol{x}) = f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0)^T(\boldsymbol{x} - \boldsymbol{x}_0) \\ + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_0)^T \nabla^2 f(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0)$$

为$f(\boldsymbol{x})$在$\boldsymbol{x}_0$处的二次近似函数。

重要结论

1. 记 $q_1(\boldsymbol{x}) = f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0)^T(\boldsymbol{x} - \boldsymbol{x}_0)$
则

$$\nabla q_1(\boldsymbol{x}) = \nabla f(\boldsymbol{x}_0), \nabla^2 q_1(\boldsymbol{x}) = 0$$

2. 记 $q_2(\boldsymbol{x}) = f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0)^T(\boldsymbol{x} - \boldsymbol{x}_0) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_0)^T \nabla^2 f(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0)$
则

$$\begin{aligned} \nabla q_2(\boldsymbol{x}) &= \nabla f(\boldsymbol{x}_0) + \nabla^2 f(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0) \\ \nabla^2 q_2(\boldsymbol{x}) &= \nabla^2 f(\boldsymbol{x}_0) \end{aligned}$$

重要结论–证明

$$
\begin{aligned}
q_2(\boldsymbol{x}) =& f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0)^T(\boldsymbol{x}-\boldsymbol{x}_0) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_0)^T\nabla^2 f(\boldsymbol{x}_0)(\boldsymbol{x}-\boldsymbol{x}_0) \\
=& f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0)^T\boldsymbol{x} - \nabla f(\boldsymbol{x}_0)^T\boldsymbol{x}_0 \\
& + \frac{1}{2}[(\boldsymbol{x}-\boldsymbol{x}_0)^T\nabla^2 f(\boldsymbol{x}_0)\boldsymbol{x} - (\boldsymbol{x}-\boldsymbol{x}_0)^T\nabla^2 f(\boldsymbol{x}_0)\boldsymbol{x}_0] \\
=& f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0)^T\boldsymbol{x} - \nabla f(\boldsymbol{x}_0)^T\boldsymbol{x}_0 + \frac{1}{2}[\boldsymbol{x}^T\nabla^2 f(\boldsymbol{x}_0)\boldsymbol{x} \\
& - \boldsymbol{x}_0^T\nabla^2 f(\boldsymbol{x}_0)\boldsymbol{x} - \boldsymbol{x}^T\nabla^2 f(\boldsymbol{x}_0)\boldsymbol{x}_0 + \boldsymbol{x}_0^T\nabla^2 f(\boldsymbol{x}_0)\boldsymbol{x}_0] \\
=& f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0)^T\boldsymbol{x} - \nabla f(\boldsymbol{x}_0)^T\boldsymbol{x}_0 + \frac{1}{2}[\boldsymbol{x}^T\nabla^2 f(\boldsymbol{x}_0)\boldsymbol{x} \\
& - (\nabla^2 f(\boldsymbol{x}_0)\boldsymbol{x}_0)^T\boldsymbol{x} - (\nabla^2 f(\boldsymbol{x}_0)\boldsymbol{x}_0)^T\boldsymbol{x} + \boldsymbol{x}_0^T\nabla^2 f(\boldsymbol{x}_0)\boldsymbol{x}_0]
\end{aligned}
$$

重要结论–证明

$$
\begin{aligned}
&= \frac{1}{2}\boldsymbol{x}^T\nabla^2 f(\boldsymbol{x}_0)\boldsymbol{x} + \nabla f(\boldsymbol{x}_0)^T\boldsymbol{x} - (\nabla^2 f(\boldsymbol{x}_0)\boldsymbol{x}_0)^T\boldsymbol{x} \\
&\quad + \frac{1}{2}\boldsymbol{x}_0^T\nabla^2 f(\boldsymbol{x}_0)\boldsymbol{x}_0 - \nabla f(\boldsymbol{x}_0)^T\boldsymbol{x}_0 + f(\boldsymbol{x}_0) \\
&= \frac{1}{2}\boldsymbol{x}^T\nabla^2 f(\boldsymbol{x}_0)\boldsymbol{x} + [\nabla f(\boldsymbol{x}_0) - \nabla^2 f(\boldsymbol{x}_0)\boldsymbol{x}_0]^T\boldsymbol{x} \\
&\quad + \frac{1}{2}\boldsymbol{x}_0^T\nabla^2 f(\boldsymbol{x}_0)\boldsymbol{x}_0 - \nabla f(\boldsymbol{x}_0)^T\boldsymbol{x}_0 + f(\boldsymbol{x}_0)
\end{aligned}
$$

故

$$
\begin{aligned}
\nabla q_2(\boldsymbol{x}) &= \nabla^2 f(\boldsymbol{x}_0)x + \nabla f(\boldsymbol{x}_0) - \nabla^2 f(\boldsymbol{x}_0)\boldsymbol{x}_0 \\
&= \nabla^2 f(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0)
\end{aligned}
$$

附录：定理2.2的证明

设$f(\boldsymbol{x})$具有连续的一阶偏导数，由 Taylor公式:

$$f\left(\boldsymbol{x}^k+\alpha \boldsymbol{p}\right)=f\left(\boldsymbol{x}^k\right)+\alpha \nabla f\left(\boldsymbol{x}^k\right)^T \boldsymbol{p}+o(\alpha)$$

当$\nabla f\left(\boldsymbol{x}^k\right)^T \boldsymbol{p}<0$ **时，有** $f\left(\boldsymbol{x}^k+\alpha \boldsymbol{p}\right)<f(\boldsymbol{x}^k)$**，**

所以(α充分小时)p 是 $f(x)$ 在x^k**处的一个下降方向。**

因此，满足：$\nabla f\left(\boldsymbol{x}^k\right)^T \boldsymbol{p}<0$的方向$\boldsymbol{p}$为$f(\boldsymbol{x})$在 $\boldsymbol{x}^k$处的下降方向。

目录

凸集

定义

设集合$\boldsymbol{\Omega} \subseteq \mathbb{R}^n$，若对于任意两点 $\boldsymbol{x}$，$\boldsymbol{y} \in \boldsymbol{\Omega}$及实数$\lambda\,(0 \leq \lambda \leq 1)$, 都有:

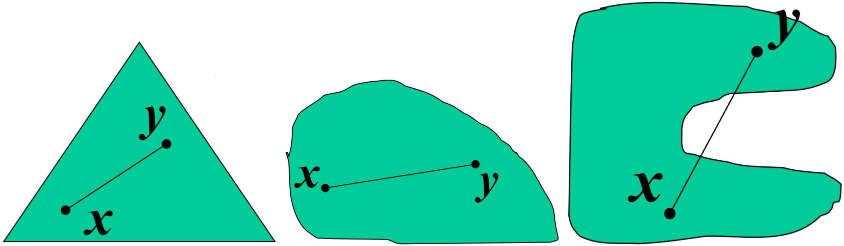
$$\lambda \boldsymbol{x} + (1-\lambda)\,\boldsymbol{y} \in \boldsymbol{\Omega}$$

则称集合$\boldsymbol{\Omega}$为凸集。

几何直观：任意两点的连线都在其中。

常见凸集：空集，单点集，整个欧氏空间$\mathbb{R}^n$。

凸集



凸集

例 满足线性规划问题的约束条件$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b},\boldsymbol{x}\geq\boldsymbol{0}$的一切$\boldsymbol{x}$所组成的集合$\boldsymbol{S}$是一个凸集，即证$\boldsymbol{S}=\{\boldsymbol{x}\in\mathbb{R}^n|\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b},\boldsymbol{x}\geq\boldsymbol{0}\}$是一个凸集.

例 满足线性规划问题的约束条件$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b},\boldsymbol{x}\geq\boldsymbol{0}$ 的一切$\boldsymbol{x}$所组成的集合$\boldsymbol{S}$是一个凸集，即证$\boldsymbol{S}=\{\boldsymbol{x}\in\mathbb{R}^n|\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b},\boldsymbol{x}\geq\boldsymbol{0}\}$是一个凸集.

证明：
对任意的$\boldsymbol{x},\boldsymbol{y}\in\boldsymbol{S}$，有$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b},\boldsymbol{x}\geq\boldsymbol{0};\boldsymbol{A}\boldsymbol{y}=\boldsymbol{b},\boldsymbol{y}\geq\boldsymbol{0}$.
所以，对任意的 $0\leq\lambda\leq1$，有: $\lambda\boldsymbol{x}+(1-\lambda)\boldsymbol{y}\geq\boldsymbol{0}$,
并且

$$\begin{aligned}\boldsymbol{A}[\lambda\boldsymbol{x}+(1-\lambda)\boldsymbol{y}]&=\lambda\boldsymbol{A}\boldsymbol{x}+(1-\lambda)\boldsymbol{A}\boldsymbol{y}\\&=\lambda\boldsymbol{b}+(1-\lambda)\boldsymbol{b}=\boldsymbol{b}\end{aligned}$$

即点$\lambda\boldsymbol{x}+(1-\lambda)\boldsymbol{y}\in\boldsymbol{S}$，故$\boldsymbol{S}$是凸集。

同理可证如下超平面也是凸集：

$$\boldsymbol{H}=\left\{\boldsymbol{x}\in R^n|\boldsymbol{c}^T\boldsymbol{x}=b\right\},$$

其中$\boldsymbol{c}\neq0$ 是n维向量，$b\in R$.

凸集

例 证明超球$\{\boldsymbol{x} \in R^n \mid \|\boldsymbol{x}\|_2 \leq r\}$为凸集.

凸集

例 证明超球$\{\boldsymbol{x} \in R^n | \, \|\boldsymbol{x}\|_2 \le r\}$为凸集.

证明：设$\boldsymbol{x}, \boldsymbol{y}$ 为超球中的任意两点， $\lambda \in [0, 1]$，

则：

$$
\begin{aligned}
\|\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}\|_2 &\le \lambda \|\boldsymbol{x}\|_2 + (1-\lambda) \|\boldsymbol{y}\|_2 \\
&\le \lambda r + (1-\lambda) r = r
\end{aligned}
$$

即点 $\lambda \boldsymbol{x} + (1-\lambda) \boldsymbol{y}$属于超球，故超球为凸集.

同理可证如下$\boldsymbol{x}^0$的$\delta-$邻域：

$$
N_\delta(\boldsymbol{x}^0) = \left\{\boldsymbol{x} \in R^n | \, \left\|\boldsymbol{x} - \boldsymbol{x}^0\right\|_2 < \delta\right\}
$$

是凸集

凸集的性质

性质:

任意一组凸集的交集为凸集.

证明：设 $\boldsymbol{A}, \boldsymbol{B}$ 是凸集，证 $\boldsymbol{A} \bigcap \boldsymbol{B}$ 是凸集.

对任意的 $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{A} \bigcap \boldsymbol{B}$， $0 \leq \lambda \leq 1$，

有 $\boldsymbol{x} \in \boldsymbol{A} \bigcap \boldsymbol{B}$ 且 $\boldsymbol{y} \in \boldsymbol{A} \bigcap \boldsymbol{B}$

由于 $\boldsymbol{A}, \boldsymbol{B}$ 是凸集，则

$$\lambda \boldsymbol{x} + (1 - \lambda) \boldsymbol{y} \in \boldsymbol{A}, \lambda \boldsymbol{x} + (1 - \lambda) \boldsymbol{y} \in \boldsymbol{B}$$

即 $\lambda \boldsymbol{x} + (1 - \lambda) \boldsymbol{y} \in A \bigcap B$.

可见 $\boldsymbol{A} \bigcap \boldsymbol{B}$ 是凸集.

凸函数

定义:

设函数 $f(\boldsymbol{x})$定义在凸集 $\boldsymbol{\Omega} \subseteq \mathbb{R}^n$上。

若对任意的$\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{\Omega}$，及任意的 $\lambda \in [0,1]$都有

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

则称函数 $f(\boldsymbol{x})$ 为凸集合 $\boldsymbol{\Omega}$ 上的凸函数。

注:

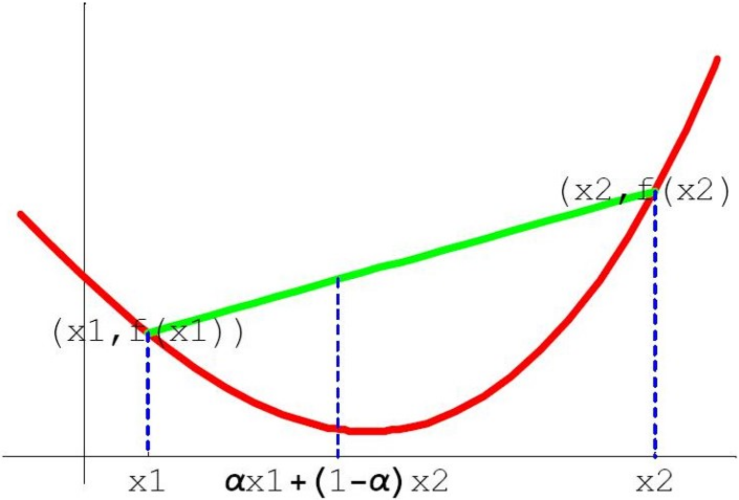
1. 将上述定义中的不等式取严格不等号，则是严格凸函数。
2. 高数教材上常称之为下凸函数。

凸函数的几何意义

对一元函数 $y=f(x)$，在几何上 $\lambda f(x_1)+(1-\lambda)f(x_2)$ $(0\le\lambda\le 1)$ 表示连接 $(x_1,f(x_1))$，$(x_2,f(x_2))$ 的线段，$f(\lambda x_1+(1-\lambda)x_2)$ 表示曲线 $y=f(x)$ 上横坐标为 $\lambda x_1+(1-\lambda)x_2$ 的点的纵坐标。

所以，一元严格凸函数表示连接函数图形上任意两点的线段总是位于曲线弧的上方。

凸函数的几何意义



凸函数

例 试证线性函数$f(\boldsymbol{x}) = \boldsymbol{c}^T\boldsymbol{x} = c_1x_1 + c_2x_2 + \cdots + c_nx_n$是 $\mathbb{R}^n$上的凸函数

凸函数

例 试证线性函数$f(\boldsymbol{x}) = \boldsymbol{c}^T\boldsymbol{x} = c_1x_1 + c_2x_2 + \cdots + c_nx_n$是 $\mathbb{R}^n$上的凸函数

证明：设 $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$， $\lambda \in [0, 1]$，则

$$\begin{aligned} f(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}) &= \boldsymbol{c}^T(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \\ &= \lambda\boldsymbol{c}^T\boldsymbol{x} + (1-\lambda)\boldsymbol{c}^T\boldsymbol{y} \\ &= \lambda f(\boldsymbol{x}) + (1-\lambda)f(\boldsymbol{y}) \end{aligned}$$

所以 $f(\boldsymbol{x})$ 是凸函数。

凸函数的性质

(1) 设 $f(\boldsymbol{x})$ 是凸集 $\boldsymbol{\Omega} \subseteq \mathbb{R}^n$ 上的凸函数，实数 $k \geq 0$，则 $kf(\boldsymbol{x})$ 也是 $\boldsymbol{\Omega}$ 上的凸函数。

证明：任取 $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{\Omega}$， $\lambda \in [0, 1]$, 由 f 的凸性，则

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

注意到 $k \geq 0$，故有

$$kf(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leq k[\lambda f(\boldsymbol{x}) + (1-\lambda)f(\boldsymbol{y})] = \lambda kf(\boldsymbol{x}) + (1-\lambda)kf(\boldsymbol{y})$$

可见 $kf(\boldsymbol{x})$， $k \geq 0$ 是凸函数。

凸函数的性质

(2) 设 $f_1(\boldsymbol{x})$，$f_2(\boldsymbol{x})$ 是凸集 $\boldsymbol{\Omega} \subseteq \mathbb{R}^n$ 上的凸函数，则 $f_1(\boldsymbol{x}) + f_2(\boldsymbol{x})$ 也是 $\boldsymbol{\Omega}$ 上的凸函数。

证明：任取 $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{\Omega}$，$\lambda \in [0,1]$，有

$$
\begin{aligned}
f_1(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}) + f_2(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \\
&\leq \lambda f_1(\boldsymbol{x}) + (1-\lambda) f_1(\boldsymbol{y}) + \lambda f_2(\boldsymbol{x}) + (1-\lambda) f_2(\boldsymbol{y}) \\
&= \lambda[f_1(\boldsymbol{x}) + f_2(\boldsymbol{x})] + (1-\lambda)[f_1(\boldsymbol{y}) + f_2(\boldsymbol{y})]
\end{aligned}
$$

可见 $f_1(\boldsymbol{x}) + f_2(\boldsymbol{x})$ 也是 $\boldsymbol{\Omega}$ 上的凸函数。

凸函数的性质

(3) 设 $f_1(\boldsymbol{x}), \cdots, f_m(\boldsymbol{x})$ 是凸集 $\boldsymbol{\Omega} \subseteq \mathbb{R}^n$ 上的凸函数，实数 $w_i \geq 0$，则 $\sum_{i=1}^{m} w_i f_i(\boldsymbol{x})$ 也是 $\boldsymbol{\Omega}$ 上的凸函数。

证明：结论(3)是上面结论(1)和(2)的推论。

(4) 设 $f(\boldsymbol{x})$ 是凸集 $\boldsymbol{\Omega} \subseteq \mathbb{R}^n$ 上的凸函数，c 是实数，则水平集 $\boldsymbol{\Omega}_c = \{\boldsymbol{x} \in \boldsymbol{\Omega} \mid f(\boldsymbol{x}) \leq c\}$ 是凸集。

证明：任取 $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{\Omega}_c$， $\lambda \in [0,1]$，
则 $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{\Omega}$， $f(\boldsymbol{x}) \leq c$， $f(\boldsymbol{y}) \leq c$
由 $\boldsymbol{\Omega}$ 和 $f(\boldsymbol{x})$ 的凸性，得 $\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y} \in \boldsymbol{\Omega}$
并且 $f(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leq \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y}) \leq \lambda c + (1-\lambda)c = c$

这说明 $\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y} \in \boldsymbol{\Omega}_c$，即 $\boldsymbol{\Omega}_c$ 是凸集得证。

凸函数的性质

(5) 设 $f_i(\boldsymbol{x})(i=1,2,\cdots,p)$ 为凸集 $\boldsymbol{\Omega} \subseteq \mathbb{R}^n$ 上的凸函数，实数$c_i(i=1,2,\cdots,p)$ 为实常数，则下面的集合

$$\widehat{\boldsymbol{\Omega}} = \{\boldsymbol{x} \mid \boldsymbol{x} \in \boldsymbol{\Omega}, f_1(\boldsymbol{x}) \leq c_1, \cdots, f_p(\boldsymbol{x}) \leq c_p\}$$

也是凸集。

证明：结论(5)是上面结论(4)和凸集的性质的推论。

判断凸函数的一阶条件

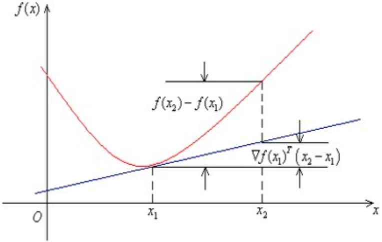
定理:

设在凸集 $\boldsymbol{\Omega} \subseteq \mathbb{R}^n$ 上 $f(\boldsymbol{x})$ 可微，则 $f(\boldsymbol{x})$ 在 $\boldsymbol{\Omega}$ 上为凸函数的充要条件是对任意的 $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \boldsymbol{\Omega}$，都有：

$$f(\boldsymbol{x}_2) \geq f(\boldsymbol{x}_1) + \nabla f(\boldsymbol{x}_1)^T(\boldsymbol{x}_2 - \boldsymbol{x}_1)$$

定理:

凸函数(充要条件)：可微函数为凸函数的充要条件是在其定义域凸集中任一点处的切平面(切线)都不在曲面(曲线)的上方。



判断凸函数的二阶条件

定理:

设在开凸集$\Omega \subseteq \mathbb{R}^n$ 内 $f(\boldsymbol{x})$二阶连续可微，则$f(\boldsymbol{x})$是 Ω内的凸函数的充要条件为：在Ω 内任一点$\boldsymbol{x}$处,$f(\boldsymbol{x})$的Hessian阵半正定，

其中:

$$\nabla^2 f\left(\boldsymbol{x}\right)=\begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

矩阵半正定的判定(补充)

设$\boldsymbol{A}$是n阶实对称阵，则$\boldsymbol{A}$半正定
$\Leftrightarrow$ $\boldsymbol{A}$的所有主子式大于等于零.
$\Leftrightarrow$ $\boldsymbol{A}$的所有特征值都大于等于零.

注意：$\boldsymbol{A}$的顺序主子式与主子式不同.

$\boldsymbol{A}$的k阶主子式是取任意k行和相应k列，位于他们交叉位置的元素构成的k阶行列式.

定理:

设在开凸集$\mathbf{\Omega} \subseteq \mathbb{R}^n$内$f(\boldsymbol{x})$二阶可微，若在$\mathbf{\Omega}$内$\nabla^2 f(\boldsymbol{x})$正定，则$f(\boldsymbol{x})$在$\mathbf{\Omega}$内是严格凸函数。

注： 反之不成立.

例：$f(\boldsymbol{x}) = x^4$是严格凸的,
但在点$\boldsymbol{x} = \mathbf{0}$处$\nabla^2 f(\mathbf{0})$不是正定的.

判断凸函数的二阶条件

矩阵正定的判定(补充)

设$\boldsymbol{A}$是n阶实对称阵，则$\boldsymbol{A}$正定
$\Leftrightarrow$ $\boldsymbol{A}$的所有顺序主子式大于零.
$\Leftrightarrow$ $\boldsymbol{A}$的n个特征值都大于零.

重要结论

定理: (特殊情形)

若函数$f(\boldsymbol{x})$定义在$\mathbb{R}^n$上的凸函数，且具有一阶连续偏导数，则$\boldsymbol{x}^*$是$f(\boldsymbol{x})$的极小值点的充要条件是:

$$\nabla f\left(\boldsymbol{x}^*\right)=0.$$

定理: (极值点的二阶充分条件)

若函数$f(\boldsymbol{x})$定义在$\boldsymbol{\Omega} \subseteq \mathbb{R}^n$并具有二阶连续偏导数，且$\boldsymbol{x}^*$是$\boldsymbol{\Omega}$内的一点，$\nabla f\left(\boldsymbol{x}^*\right)=\mathbf{0}, \nabla^2 f\left(\boldsymbol{x}^*\right)$正定，则 $\boldsymbol{x}^*$是$f(\boldsymbol{x})$的极小值点

重要结论

例 设函数$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x} + c$，其中$\boldsymbol{A} \in \mathbb{R}^{n\times n}$为实对称矩阵，$\boldsymbol{b} \in \mathbb{R}^n, c \in \mathbb{R}$ **证明：**

(1)当$\boldsymbol{A}$半正定时，$f(\boldsymbol{x})$是凸函数.

(2)当$\boldsymbol{A}$正定时，$f(\boldsymbol{x})$是严格凸函数.

重要结论

例 设函数$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x} + c$，其中$\boldsymbol{A} \in \mathbb{R}^{n\times n}$为实对称矩阵，$\boldsymbol{b} \in \mathbb{R}^n, c \in \mathbb{R}$ 证明：
(1)当$\boldsymbol{A}$半正定时，$f(\boldsymbol{x})$是凸函数.
(2)当$\boldsymbol{A}$正定时，$f(\boldsymbol{x})$是严格凸函数.

证明：由上一节的例3知：

$$\nabla^2 f(\boldsymbol{x}) = \boldsymbol{A}$$

结合定理3.3和3.4即可得到上述结论.

重要结论

例 证明函数

$$f(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{x} = x_1^2 + x_2^2 + \cdots + x_n^2$$

是$\mathbb{R}^n$上的严格凸函数。

例 证明函数

$$f(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{x} = x_1^2 + x_2^2 + \cdots + x_n^2$$

是$\mathbb{R}^n$上的严格凸函数。

证明: 易看出$\nabla^2 f(\boldsymbol{x}) = 2\boldsymbol{I}_n$为数量矩阵，正定。

由定理3.4，即得所证。

重要结论

例 讨论函数$f(\boldsymbol{x}) = 5x_1^2 - 6x_1x_2 + 5x_2^2$的凸性

例 讨论函数$f(\boldsymbol{x}) = 5x_1^2 - 6x_1x_2 + 5x_2^2$的凸性

解：容易求得$\nabla^2 f(\boldsymbol{x}) = \begin{pmatrix} 10 & -6 \\ -6 & 10 \end{pmatrix}$

由于各阶顺序主子式

$$10 > 0, \quad \begin{vmatrix} 10 & -6 \\ -6 & 10 \end{vmatrix} = 64 > 0$$

即Hessian矩阵正定。

因此$f(\boldsymbol{x})$是严格凸函数.

凸规划

考虑如下约束数学规划模型

$$\begin{aligned} \min\ f(\boldsymbol{x}), &\quad \boldsymbol{x} \in \mathbb{R}^n \\ s.t.\ h_i(\boldsymbol{x}) &= 0, i = 1, 2, \cdots, l \\ g_j(\boldsymbol{x}) &\le 0, j = 1, 2, \cdots, m \end{aligned}$$

定义:

若$\boldsymbol{D} = \{\boldsymbol{x} \in \mathbb{R}^n \,|h_i(\boldsymbol{x}) = 0, g_j(\boldsymbol{x}) \le 0, i = 1, \cdots, l; j = 1, \cdots, m\}$ 是凸集，目标函数$f(\boldsymbol{x})$是可行域$\boldsymbol{D}$上的凸函数则称问题是凸规划。

凸规划

定理

考虑如下约束非线性数学规划模型

$$\begin{aligned} \min\ & f(\boldsymbol{x}), \boldsymbol{x} \in \mathbb{R}^n \\ s.t.\ & h_i(\boldsymbol{x}) = 0, i = 1, 2, \cdots, l \\ & g_j(\boldsymbol{x}) \leq 0, j = 1, 2, \cdots, m \end{aligned}$$

若$f(\boldsymbol{x}), g_j(\boldsymbol{x})(j = 1, 2, \cdots, m)$均是可行域$\boldsymbol{D}$上的凸函数，$h_i(\boldsymbol{x})(i = 1, 2, \cdots$
是线性函数，则该问题是凸规划。

凸规划

证明：由定义知，只需证

$$\boldsymbol{D} = \{\boldsymbol{x} \in \mathbb{R}^n \,|h_i(\boldsymbol{x}) = 0, g_j(\boldsymbol{x}) \le 0, i = 1, \cdots, l; j = 1, \cdots, m\}$$

是凸集即可。

记

$$\boldsymbol{D}_1 = \{\boldsymbol{x} \in \mathbb{R}^n \,|h_i(\boldsymbol{x}) = 0, i = 1, \cdots, l\};$$

$$\boldsymbol{D}_2 = \{\boldsymbol{x} \in \mathbb{R}^n \,|g_j(\boldsymbol{x}) \le 0, j = 1, \cdots, m\}$$

则$\boldsymbol{D} = \boldsymbol{D}_1 \bigcap \boldsymbol{D}_2$

凸规划

下面分别证 D_1和$\boldsymbol{D}_2$都是凸集。

1. 由于$h_i(\boldsymbol{x})$皆是线性函数，由例3.1和凸集性质可得 D_1是凸集.
2. 对于 $\boldsymbol{D}_2 = \{\boldsymbol{x} \in \mathbb{R}^n \mid g_j(\boldsymbol{x}) \leq 0, j = 1, \cdots, m\}$, 任取$\boldsymbol{x}_1, \boldsymbol{x}_2 \in \boldsymbol{D}_2$, $0 \leq \lambda \leq 1$，由于$g_j(\boldsymbol{x})$是凸函数，则

$$g_j\left(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2\right) \leq \lambda g_j(\boldsymbol{x}_1) + (1-\lambda)\, g_j(\boldsymbol{x}_2) \leq 0$$

这说明$\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 \in \boldsymbol{D}_2$ 因此，$\boldsymbol{D}_2$是凸集。

综上可知，$\boldsymbol{D} = \boldsymbol{D}_1 \bigcap \boldsymbol{D}_2$是凸集。

凸规划

定理

1. 凸规划问题的任一局部极小点都是整体极小点，全体极小点构成的集合是凸集。
2. 若$f(\boldsymbol{x})$是凸集$\boldsymbol{D} \subseteq \mathbb{R}^n$上的严格凸函数，且严格凸规划问题 $\min_{\boldsymbol{x} \in \boldsymbol{D}} f(\boldsymbol{x})$ 的整体极小点存在，则整体极小点是唯一的。