最优化方法

苏延旭
email: yanxu.su@ahu.edu.cn

2023 年 9 月 24 日

课程安排

- 第 1 周至第 9 周，共 36 课时
- 周一第 6、7 节课/周三第 3、4 节课，博学南楼 B401（1、2 班）
- 周一/三第 1、2 节课，博学南楼 B209（3、4 班）
- 考核形式：课堂测验 + 大作业 + 期末考试
- 课堂测验：30%
  大作业：20%
  期末考试：50%

参考书目

[1] Stephen Boyd and Lieven Vandenberghe. Convex Optimization. Cambridge University Press, 2004.

[2] Yurii Nesterov. Lectures on Convex Optimization. Springer, 2018.

[3] Jorge Nocedal and Stephen J Wright. Numerical Optimization. Springer, 1999.

[4] Dimitri P Bertsekas. Nonlinear Programming. Athena Scientific, 1999.

[5] 袁亚湘，孙文瑜. 最优化理论与方法. 科学出版社, 1997.

总目录

章节目录

优化/数学规划

优化/数学规划 (Optimization/Mathematical Programming)：从一个可行解的集合中，寻找最优的元素。

优化/数学规划

优化/数学规划 (Optimization/Mathematical Programming): 从一个可行解的集合中，寻找最优的元素。
从数学上看，任意一个优化问题都可以写成这样的形式，

$$\begin{array}{rl} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq b_i, i = 1, \ldots, m \end{array} \tag{1}$$

其中

$x = [x_1, \ldots, x_n]^\top$	优化变量 Optimization variables
$f_0 : \mathbb{R}^n \to \mathbb{R}$	目标函数 Objective function
$f_i : \mathbb{R}^n \to \mathbb{R}$	不等式约束 Inequality constraints

优化/数学规划

优化/数学规划 (Optimization/Mathematical Programming): 从一个可行解的集合中，寻找最优的元素。
从数学上看，任意一个优化问题都可以写成这样的形式，

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \le b_i,\ i = 1, \ldots, m \end{aligned} \tag{1}$$

其中

$x = [x_1, \ldots, x_n]^\top$	优化变量 Optimization variables
$f_0 : \mathbb{R}^n \to \mathbb{R}$	目标函数 Objective function
$f_i : \mathbb{R}^n \to \mathbb{R}$	不等式约束 Inequality constraints

1. $x^\star$ 最优 $\Leftrightarrow \forall z, z \in \Omega = \{f_i(z) \le b_i, i = 1, \ldots, m\}$(可行解集 feasible set)
2. 对于任意 z 满足 $f_i(z) \le b_i$ 都有 $f_0(z) \ge f_0(x^\star)$

全局/局部最优解

1. Global Minimizer: 如果对任意的 $x \in \Omega$，都有 $f(x^\star) \leq f(x), x^\star \in \Omega$ 成立，那么 $x^\star$ 就是函数的全局最小值。

全局/局部最优解

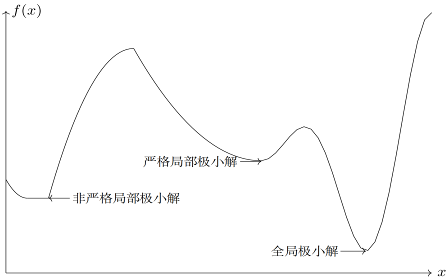
1. Global Minimizer: 如果对任意的 $x \in \Omega$，都有 $f(x^\star) \leq f(x), x^\star \in \Omega$ 成立，那么 $x^\star$ 就是函数的全局最小值。
2. (Strict) Local Minimizer: 如果对任意的 $x \in \Omega$，都有 $f(x^\star) \leq f(x)$, $x^\star \in \mathcal{N}_{x^\star} \subset \Omega$ 成立，那么 $x^\star$ 就是函数的局部最小值。如果严格不取等，那么就称它为严格局部最小值。这里 $\mathcal{N}_{x^\star}$ 称为 $x^\star$ 的邻域。

全局/局部最优解

1. Global Minimizer: 如果对任意的 $x \in \Omega$，都有 $f(x^\star) \le f(x), x^\star \in \Omega$ 成立，那么 $x^\star$ 就是函数的全局最小值。
2. (Strict) Local Minimizer: 如果对任意的 $x \in \Omega$，都有 $f(x^\star) \le f(x)$, $x^\star \in \mathcal{N}_{x^\star} \subset \Omega$ 成立，那么 $x^\star$ 就是函数的局部最小值。如果严格不取等，那么就称它为严格局部最小值。这里 $\mathcal{N}_{x^\star}$ 称为 $x^\star$ 的邻域。
3. Isolated Local Minimizer: 如果 $x^\star$ 是 $f(x)$ 的唯一局部最小值，则称它是孤立局部最小值。

全局/局部最优解

1. Global Minimizer：如果对任意的 $x \in \Omega$，都有 $f(x^\star) \le f(x), x^\star \in \Omega$ 成立，那么 $x^\star$ 就是函数的全局最小值。
2. (Strict) Local Minimizer：如果对任意的 $x \in \Omega$，都有 $f(x^\star) \le f(x)$，$x^\star \in \mathcal{N}_{x^\star} \subset \Omega$ 成立，那么 $x^\star$ 就是函数的局部最小值。如果严格不取等，那么就称它为严格局部最小值。这里 $\mathcal{N}_{x^\star}$ 称为 $x^\star$ 的邻域。
3. Isolated Local Minimizer：如果 $x^\star$ 是 $f(x)$ 的唯一局部最小值，则称它是孤立局部最小值。



优化问题的分类

最优化问题的具体形式非常丰富，可以按照目标函数、约束函数以及解的性质将其分类。

优化问题的分类

最优化问题的具体形式非常丰富，可以按照目标函数、约束函数以及解的性质将其分类。

- 当目标函数和约束函数均为线性函数时，问题称为**线性规划**

优化问题的分类

最优化问题的具体形式非常丰富，可以按照目标函数、约束函数以及解的性质将其分类。

- 当目标函数和约束函数均为线性函数时，问题称为**线性规划**
- 当目标函数和约束函数中至少有一个为非线性函数时，相应的问题称为**非线性规划**

优化问题的分类

最优化问题的具体形式非常丰富，可以按照目标函数、约束函数以及解的性质将其分类。

- 当目标函数和约束函数均为线性函数时，问题称为**线性规划**
- 当目标函数和约束函数中至少有一个为非线性函数时，相应的问题称为**非线性规划**
- 如果目标函数是二次函数而约束函数是线性函数则称为**二次规划**

优化问题的分类

最优化问题的具体形式非常丰富，可以按照目标函数、约束函数以及解的性质将其分类。

- 当目标函数和约束函数均为线性函数时，问题称为**线性规划**
- 当目标函数和约束函数中至少有一个为非线性函数时，相应的问题称为**非线性规划**
- 如果目标函数是二次函数而约束函数是线性函数则称为**二次规划**
- 包含非光滑函数的问题称为**非光滑优化**

优化问题的分类

最优化问题的具体形式非常丰富，可以按照目标函数、约束函数以及解的性质将其分类。

- 当目标函数和约束函数均为线性函数时，问题称为**线性规划**
- 当目标函数和约束函数中至少有一个为非线性函数时，相应的问题称为**非线性规划**
- 如果目标函数是二次函数而约束函数是线性函数则称为**二次规划**
- 包含非光滑函数的问题称为**非光滑优化**
- 不能直接求导数的问题称为**无导数优化**

优化问题的分类

最优化问题的具体形式非常丰富，可以按照目标函数、约束函数以及解的性质将其分类。

- 当目标函数和约束函数均为线性函数时，问题称为**线性规划**
- 当目标函数和约束函数中至少有一个为非线性函数时，相应的问题称为**非线性规划**
- 如果目标函数是二次函数而约束函数是线性函数则称为**二次规划**
- 包含非光滑函数的问题称为**非光滑优化**
- 不能直接求导数的问题称为**无导数优化**
- 变量只能取整数的问题称为**整数规划**

优化问题的分类

最优化问题的具体形式非常丰富，可以按照目标函数、约束函数以及解的性质将其分类。

- 当目标函数和约束函数均为线性函数时，问题称为**线性规划**
- 当目标函数和约束函数中至少有一个为非线性函数时，相应的问题称为**非线性规划**
- 如果目标函数是二次函数而约束函数是线性函数则称为**二次规划**
- 包含非光滑函数的问题称为**非光滑优化**
- 不能直接求导数的问题称为**无导数优化**
- 变量只能取整数的问题称为**整数规划**
- 在线性约束下极小化关于半正定矩阵的线性函数的问题称为**半定规划**，其广义形式为锥规划

优化问题的分类

- 最优解只有少量非零元素的问题称为**稀疏优化**

优化问题的分类

- 最优解只有少量非零元素的问题称为**稀疏优化**
- 最优解是低秩矩阵的问题称为**低秩矩阵优化**

优化问题的分类

- 最优解只有少量非零元素的问题称为**稀疏优化**
- 最优解是低秩矩阵的问题称为**低秩矩阵优化**
- 此外还有几何优化、二次锥规划、张量优化、鲁棒优化、全局优化、组合优化、网络规划、随机优化、动态规划、带微分方程约束优化、微分流形约束优化、分布式优化等

优化问题的分类

- 最优解只有少量非零元素的问题称为**稀疏优化**
- 最优解是低秩矩阵的问题称为**低秩矩阵优化**
- 此外还有几何优化、二次锥规划、张量优化、鲁棒优化、全局优化、组合优化、网络规划、随机优化、动态规划、带微分方程约束优化、微分流形约束优化、分布式优化等
- 就具体应用而言，问题 (1) 可涵盖统计学习、压缩感知、最优运输、信号处理、图像处理、机器学习、强化学习、模式识别、金融工程、电力系统等领域的优化模型

最优化问题的应用——投资组合优化

数学建模很容易给出应用问题的不同模型，可以对应性质很不相同的问题，其求解难度和需要的算法也将差别很大。在投资组合优化中，人们希望通过寻求最优的投资组合以降低风险、提高收益。

- 这时，决策变量 x_i 表示在第 i 项资产上的投资额，向量 $x \in \mathbb{R}^n$ 表示整体的投资分配
- 约束条件可能为总资金数、每项资产的最大 (最小) 投资额、最低收益等
- 目标函数通常是某种风险度量
- 如果是极小化收益的方差，则该问题是典型的二次规划
- 如果是极小化风险价值 (value at risk) 函数，则该问题是混合整数规划
- 如果极小化条件风险价值 (conditional value at risk) 函数，则该问题是非光滑优化，也可以进一步化成线性规划

最优化问题的应用——投资组合优化

- r_i：随机变量，投资第 i 项资产的回报率
- x_i：在第 i 项资产上的投资额
- 回报：$r = r_1x_1 + r_2x_2 + \cdots + r_nx_n$
- 期望回报：$R = E(r) = \sum E(r_i)x_i = \sum \mu_i x_i$
- 风险：$V = Var(r) = \sum_{i,j} \lambda_{i,j} x_i x_j = x^\top \Lambda x$

$$\begin{aligned} \min \quad & \frac{1}{2} x^\top \Lambda x \\ \text{s.t.} \quad & \sum \mu_i x_i \geq r_0 \\ & \sum x_i = 1 \\ & x_i \geq 0 \end{aligned}$$

$$\begin{aligned} \min \quad & \text{risk measure} \\ \text{s.t.} \quad & \sum \mu_i x_i \geq r_0 \\ & \sum x_i = 1 \\ & x_i \geq 0 \end{aligned}$$

最优化问题的应用——稀疏优化

给定 $b \in \mathbb{R}^m$，矩阵 $A \in \mathbb{R}^{m\times n}$，且向量 b 的维数远小于向量 x 的维数，即 $m \ll n$。考虑线性方程组求解问题：

$$Ax = b \tag{2}$$

- 注意到由于 $m \ll n$，方程组 (2) 是欠定的，因此存在无穷多个解，重构出原始信号看似很难。
- 这些解当中大部分是不重要的，真正有用的解释所谓的"稀疏解"，即原始信号中有较多的零元素。
- 如果加上稀疏性这一先验信息，且矩阵 A 以及原问题的解 u 满足某些条件，那么可以通过求解稀疏优化问题把 u 与方程组 (2) 的其他解区分开。
- 这类技术广泛应用于压缩感知 (Compressive sensing) 问题，即通过部分信息恢复全部信息的解决方案。

最优化问题的应用——稀疏优化

$$(\ell_0)\begin{cases}\min\limits_{x\in\mathbb{R}^n} & \|x\|_0\\ \text{s.t.} & Ax=b\end{cases}\quad(\ell_1)\begin{cases}\min\limits_{x\in\mathbb{R}^n} & \|x\|_1\\ \text{s.t.} & Ax=b\end{cases}\quad(\ell_2)\begin{cases}\min\limits_{x\in\mathbb{R}^n} & \|x\|_2\\ \text{s.t.} & Ax=b\end{cases}\tag{3}$$

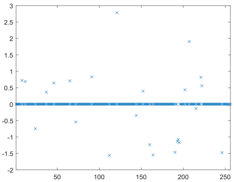
- 其中 $\|x\|_0$ 是指 x 中非零元素的个数。由于 $\|x\|_0$ 是不连续的函数，且取值只可能是整数，ℓ_0 问题实际上是 NP 难的，求解起来非常困难。
- 若定义 ℓ_1 范数：$\|x\|_1=\sum_{i=1}^n|x_i|$，则得到了另一个形式上非常相似的问题，又称 ℓ_1 范数优化问题，或基追踪问题。
- ℓ_2 范数：$\|x\|_2=(\sum_{i=1}^n x_i^2)^{\frac{1}{2}}$

最优化问题的应用——稀疏优化

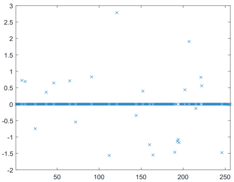
在 Matlab 环境中构造 A, u, b:

```
% 构造A, u, b
m = 128; n = 256;
A = randn(m,n);
u = sprandn(n,1,0.1);
b = A*u;
```

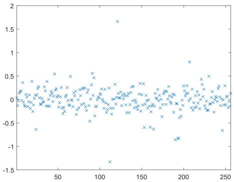
构造一个 128×256 的矩阵 A，它的每个元素都服从高斯随机分布。精确解只有 10% 的元素非零，每一个非零元素也服从高斯分布



(a) 精确解 u



(b) ℓ_1 问题的解



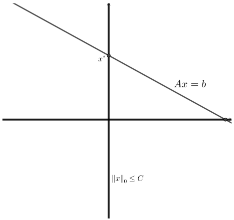
(c) ℓ_2 问题的解

最优化问题的应用——稀疏优化

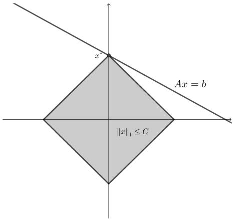
- 可以从理论上证明：若 A, b 满足一定的条件 (例如使用前面随机产生的 A 和 b)，此时向量 u 也是公式 (3) 中 ℓ_1 范数优化问题的唯一最优解。这一发现的重要之处在于，虽然 ℓ_1 范数优化仍没有显式解，但与 ℓ_0 范数优化相比，难度已经大大降低。
- 前面提到的 ℓ_0 范数优化问题是 NP 难问题，但 ℓ_1 范数优化问题的解可以非常容易地通过现有优化算法得到。从这个例子不难发现，优化学科的研究能够极大程度上帮助我们攻克现有的困难问题。但如果如此简单地把 ℓ_1 范数修改为 ℓ_2 范数，二者是否相等仍是一个问题。

最优化问题的应用——稀疏优化

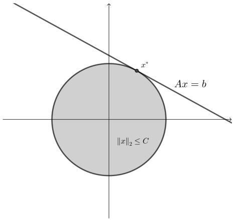
- ℓ_2 范数优化是原点到仿射集 $Ax = b$ 的投影，可以看出她的显式表达式。但 u 并不是该问题的解，也可以从下图看出该结论。
- 下图可说明 ℓ_1 范数问题的解具有稀疏性，而 ℓ_2 范数问题的解不具有该性质。



(a) ℓ_0 范数



(b) ℓ_1 范数



(c) ℓ_2 范数

最优化问题的应用——稀疏优化

- ℓ_1 范数优化的理论与算法研究在 2006 年带来了革命性的影响。理论上研究的课题包括什么条件下该问题具有稀疏性，如何改进这些条件，如何推广这些条件到其他应用。
- 常见的数据矩阵 A 一般由离散余弦变换、小波变换、傅里叶变换等生成。虽然这些矩阵本身并没有稀疏性，但通常具有很好的分析性质，保证稀疏解的存在性。
- 注意到绝对值函数在零点处不可微，ℓ_1 范数优化是非光滑优化问题。虽然它可以等价于线性规划问题，但是数据矩阵 A 通常是稠密矩阵，甚至 A 的元素未知或者不能直接存储，只能提供 Ax 或 $A^\top y$ 等运算结果。在这些特殊情况下，线性规划经典的单纯形法和内点法通常不太适用于求解大规模的问题。
- 需要强调的是，ℓ_1 范数优化的主要特点是其最优解是稀疏向量，它是稀疏优化的一种典型形式。

最优化问题的应用——LASSO 问题

考虑带 ℓ_1 范数正则项的优化问题

$$\min_{x\in\mathbb{R}^n} \mu\|x\|_1 + \frac{1}{2}\|Ax-b\|_2^2 \tag{4}$$

其中 $\mu>0$ 是给定的正则化参数。

- 问题 (4) 又称为 LASSO(Least Absolute Shrinkage and Selection Operator)，该问题可以看成是公式 (3) 中 ℓ_1 范数优化问题的二次罚函数形式。
- 由于问题 (4) 是无约束优化问题，形式上看起来比公式 (3) 中 ℓ_1 范数优化问题简单。全面掌握这两类问题的求解方法是掌握基本最优化方法的一个标志。

最优化问题的应用——低秩矩阵恢复

- 某视频网站提供了约 48 万用户对 1 万 7 千多部电影的上亿条评级数据，希望对用户的电影评级进行预测，从而改进用户电影推荐系统，为每个用户更有针对性地推荐影片。
- 显然每一个用户不可能看过所有的电影，每一部电影也不可能收到全部用户的评级。电影评级由用户打分 1 星到 5 星表示，记为 1~5 的整数。我们将电影评级放在一个矩阵 M 中，矩阵 M 的每一行表示不同用户，每一列表示不同电影。由于用户只对看过的电影给出自己的评价，矩阵 M 中很多元素是未知的。

$$
\begin{array}{c}
\\
\text{用户 1} \\ \text{用户 2} \\ \text{用户 3} \\ \text{用户 4} \\ \vdots \\ \text{用户 n}
\end{array}
\begin{array}{c}
\begin{array}{cccccc}
\text{电影 1} & \text{电影 2} & \text{电影 3} & \text{电影 4} & \cdots & \text{电影 n}
\end{array} \\
\begin{pmatrix}
4 & ? & ? & 3 & \cdots & ? \\
? & 2 & 4 & ? & \cdots & ? \\
3 & ? & ? & ? & \cdots & ? \\
2 & ? & 5 & ? & \cdots & ? \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
? & 3 & ? & 4 & \cdots & ?
\end{pmatrix}
\end{array}
$$

最优化问题的应用——低秩矩阵恢复

该问题在推荐系统、图像处理等方面有着广泛的应用。

- 由于用户对电影的偏好可进行分类，按年龄可分为：年轻人、中年人、老年人；且电影也能分为不同的题材：战争片、悬疑片、言情片等。因此这类问题隐含的假设为补全后的矩阵应为低秩的。即矩阵的行与列会有“合作”的特性，因此该问题又称为“collaborative filtering”。
- 除此之外，由于低秩矩阵可分解为两个低秩矩阵的乘积，所以低秩限制下的矩阵补全问题是比较实用的，这样利于储存且有更好的诠释性。
- 有些用户的打分可能不为自身真实情况，对评分矩阵有影响，所以原矩阵有可能是有噪声的。

最优化问题的应用——低秩矩阵恢复

由上述分析可以引出该问题：

- 令 Ω 是矩阵 M 中所有已知评级元素的下标的集合，则该问题可以初步描述为构造一个矩阵 X，使得在给定位置的元素等于已知评级元素，即满足 $X_{ij} = M_{ij}, (i,j) \in \Omega$。
- 低秩矩阵恢复 (low rank matrix completion)

$$\begin{aligned} \min_{X \in \mathbb{R}^{m \times n}} \quad & \text{rank}(X) \\ \text{s.t.} \quad & X_{ij} = M_{ij}, (i,j) \in \Omega \end{aligned} \tag{5}$$

rank(X) 正好是矩阵 X 所有非零奇异值的个数。

- 矩阵 X 的核范数 (nuclear norm) 为矩阵所有奇异值的和，即 $\|X\|_* = \sum_i \sigma_i(X) = \text{tr}(\sqrt{X^\top X})$：

$$\begin{aligned} \min_{X \in \mathbb{R}^{m \times n}} \quad & \|X\|_* \\ \text{s.t.} \quad & X_{ij} = M_{ij}, (i,j) \in \Omega \end{aligned} \tag{6}$$

最优化问题的应用——低秩矩阵恢复

- 可以证明问题 (6) 是一个凸优化问题，并且在一定条件下它与问题 (5) 等价。
- 也可以将问题 (6) 转换为一个半定规划问题，但是目前半定规划算法所能有效求解的问题规模限制了这种技术的实际应用。
- 考虑到观测可能出现误差，对于给定的参数 $\mu > 0$，给出该问题的二次罚函数形式：

$$\min_{X \in \mathbb{R}^{m \times n}} \quad \mu \|X\|_* + \frac{1}{2} \sum_{(i,j) \in \Omega} (X_{ij} - M_{ij})^2$$

最优化问题的应用——低秩矩阵恢复

- 秩 r 情形：$X = LR^\top$，其中 $L \in \mathbb{R}^{m\times r}$，$R \in \mathbb{R}^{n\times r}$，并且 $r \ll \min(m,n)$。则可将问题写为

$$\min_{L,R} \sum_{(i,j)\in\Omega} \left([LR^\top]_{ij} - M_{ij}\right)^2 + \alpha\|L\|_F^2 + \beta\|R\|_F^2$$

- 在该问题中，矩阵 X 在定义中已为秩 r 矩阵，所以没有必要再加上秩约束正则项。α, β 为正则化参数，这里正则化的作用是消除解 L, R 在缩放意义下的不唯一性。
- 此时 L, R 矩阵中的数字之和为 $(m+n)r$，远小于 np，不过此时问题是非凸的。
- 尽管这个问题是非凸的，但在某种意义上它是一个可处理问题的近似：如果对 X 有一个完整的观察，那么秩 r 近似可以通过 X 的奇异值分解来找到，并根据 r 导出的左奇异向量和右奇异向量定义 L 和 R。

最优化问题的应用——深度学习

深度学习 (deep learning) 是机器学习的一个子领域，它采用了一个特定的模型：一族通过某种方式连接起来的简单函数。由于这类模型的结构是受到人类大脑结构的启发而创造出来的，因此我们通常把它们称为神经网络 (neural networks)。神经网络中的函数链条能够将复杂的概念分解为多个层次的更简单的概念，这就是深度学习的核心思想。

- 深度学习的起源可以追溯到 20 世纪 40 年代，其雏形出现在控制论中。近十年来深度学习又重新走入了人们的视野，深度学习问题和算法的研究也经历了一次新的浪潮。
- 虽然卷积网络的设计受到了生物学和神经科学的启发，但深度学习目前的发展早已超越了机器学习模型中的神经科学观点。它用相对简单的函数来表达复杂的表示，从低层特征概括到更加抽象的高层特征，让计算机从经验中挖掘隐含的信息和价值。

最优化问题的应用——深度学习

很多机器学习中的问题可以写为：

$$\min_{x \in W} \quad \sum_{i=1}^{N} \frac{1}{2} \left\| a_i^\top x - b_i \right\|_2^2 + \mu \varphi(x) \qquad \text{线性回归}$$

$$\min_{x \in W} \quad \frac{1}{N} \sum_{i=1}^{N} \log\left(1 + \exp\left(-b_i a_i^\top x\right)\right) + \mu \varphi(x) \qquad \text{逻辑回归}$$

$$\min_{x \in W} \quad \frac{1}{N} \sum_{i=1}^{N} \ell_i \left(f(a_i, x), b_i\right) + \mu \varphi(x) \qquad \text{一般形式}$$

- a_i, b_i 是给定的数据对，b_i 是数据 a_i 对应的标签
- $\ell_i(\cdot)$：度量模型拟合数据点 i 的程度 (避免拟合不足)
- $\varphi(x)$：避免过拟合的正则项，如 $\|x\|_2^2$ 或者 $\|x\|_1$ 等等
- $f(a, x)$：线性函数或者由深度神经网络构造的模型

最优化问题的应用——多层感知机

多层感知机 (multi-layer perceptron, MLP) 也叫作深度前馈网络 (deep feedforward network) 或前馈神经网络 (feedforward neural network)，它通过已有的信息或者知识来对未知事物进行预测。

- 在神经网络中，已知的信息通常用数据集来表示。数据集一般分为训练集和测试集：训练集是用来训练神经网络，从而使得神经网络能够掌握训练集上的信息；测试集是用来测试训练完的神经网络的预测准确性。
- 一个常见的任务是分类问题．假设我们有一个猫和狗的图片集，将其划分成训练集和测试集 (保证集合中猫和狗图片要有一定的比例)。神经网络是想逼近一个从图片到 $\{0,1\}$ 的函数，这里 0 表示猫，1 表示狗．
- 因为神经网络本身的结构和大量的训练集信息，训练得到的函数与真实结果具有非常高的吻合性。

最优化问题的应用——多层感知机

给定训练集 $D=\{\{a_1,b_1\},\{a_2,b_2\},\cdots,\{a_m,b_m\}\}$。

- 假设数据 $a_i\in\mathbb{R}^p, b_i\in\mathbb{R}^q$。$a_{i1}=1$。图中给出了一种由 p 个输入单元和 q 个输出单元构成的 $(L+2)$ 层感知机，其含有一个输入层，一个输出层，和 L 个隐藏层。该感知机的第 l 个隐藏层共有 $m^{(l)}$ 个神经元。为了方便，用 $l=0$ 表示输入层，$l=L+1$ 表示输出层，并定义 $m^{(0)}=p$ 和 $m^{(L+1)}=q$。设 $y^{(l)}\in\mathbb{R}^{m^{(l)}}$ 为第 l 层的所有神经元，令 $y_1^{(l)}=1, 0\le l\le L$
- 其余的元素则是通过上一层的神经元的值进行加权求和得到。令参数 $x=(x^{(1)},x^{(2)},\ldots,x^{(L+1)})$ 表示网络中所有层之间的权重，其中 $x_{i,k}^{(l)}$ 是第 $(l-1)$ 隐藏层的第 k 个单元连接到第 l 隐藏层的第 i 个单元对应的权重，则在第 l 隐藏层中，第 i 个单元 ($i>1$，当 $l=L+1$ 时可取为 $i\ge 1$) 计算输出信息 $y_i^{(l)}$ 为

$$y_i^{(l)}=t\left(z_i^{(l)}\right),\quad z_i^{(l)}=\sum_{k=1}^{m^{(l-1)}}x_{i,k}^{(l)}y_k^{(l-1)}$$

这里函数 $t(\cdot)$ 称为激活函数。

最优化问题的应用——多层感知机

常见的激活函数类型有 Sigmoid 函数

$$t(z) = \frac{1}{1+\exp(-z)}$$

Heaviside 函数

$$t(z) = \begin{cases} 1, & z \geq 0 \\ 0, & z < 0 \end{cases}$$

以及 ReLU 函数

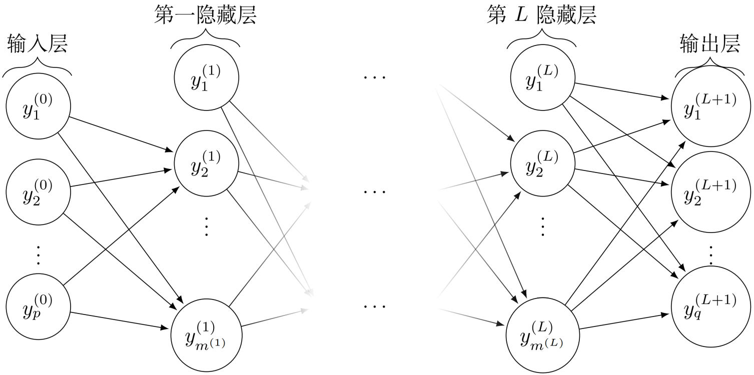
$$t(z) = \max\{0, z\}$$

整个过程可以描述为

$$y^{(0)} \xrightarrow{x^{(1)}} z^{(1)} \xrightarrow{t} y^{(1)} \xrightarrow{x^{(2)}} \cdots \xrightarrow{t} y^{(L+1)}$$

最优化问题的应用——卷积神经网络 (CNN)

上述过程可用下图表示：



最优化问题的应用——卷积神经网络（CNN）

- 给定二维图像 $\mathbf{I} \in \mathbb{R}^{n\times n}$ 和卷积核 $\mathbf{K} \in \mathbb{R}^{k\times k}$，定义卷积操作 $\mathbf{S} = \mathbf{I} * \mathbf{K}$，它的元素是

$$\mathbf{S}_{i,j} = \langle \mathbf{I}(i:i+k-1, j:j+k-1), \mathbf{K} \rangle$$

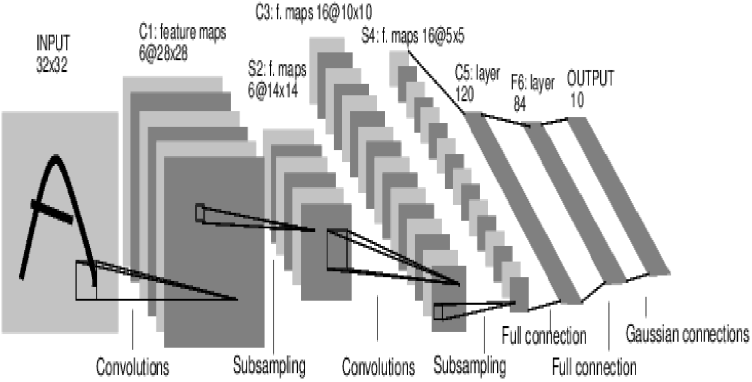
其中两个矩阵的内积是它们相应元素乘积之和

- 生成的结果 $\mathbf{S}$ 可以根据卷积核的维数、$\mathbf{I}$ 的边界是否填充、卷积操作时滑动的大小等相应变化

$$\underbrace{\begin{array}{|c|c|c|c|c|c|c|}\hline 1 & 2 & 2 & 1_{\times 1} & 2_{\times 0} & 1_{\times 1} & 1 \\ \hline 0 & 0 & 0 & 1_{\times 0} & 2_{\times 1} & 0_{\times 0} & 0 \\ \hline 2 & 0 & 1 & 2_{\times 1} & 0_{\times 0} & 1_{\times 1} & 1 \\ \hline 2 & 1 & 1 & 0 & 1 & 0 & 1 \\ \hline 1 & 0 & 2 & 0 & 1 & 2 & 2 \\ \hline 1 & 2 & 1 & 0 & 1 & 0 & 2 \\ \hline 0 & 2 & 2 & 1 & 1 & 0 & 0 \\ \hline \end{array}}_{\mathbf{I}} * \underbrace{\begin{array}{|c|c|c|}\hline 1 & 0 & 1 \\ \hline 0 & 1 & 0 \\ \hline 1 & 0 & 1 \\ \hline \end{array}}_{\mathbf{K}} = \underbrace{\begin{array}{|c|c|c|c|c|}\hline 6 & 5 & 6 & \boxed{7} & 4 \\ \hline 3 & 3 & 6 & 1 & 5 \\ \hline 7 & 3 & 4 & 6 & 4 \\ \hline 5 & 5 & 4 & 1 & 7 \\ \hline 7 & 4 & 6 & 4 & 4 \\ \hline \end{array}}_{\mathbf{S} = \mathbf{I} * \mathbf{K}}$$

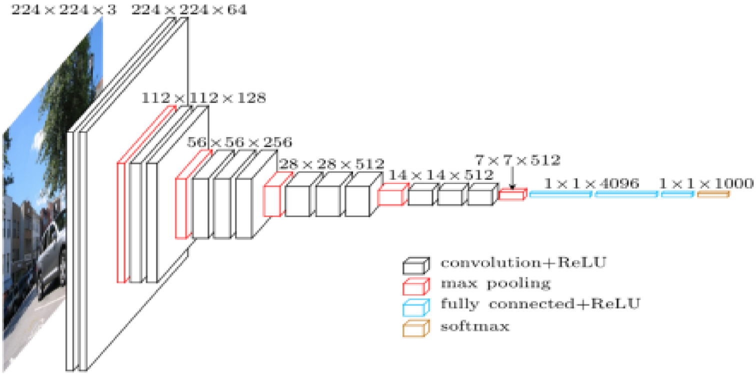
最优化问题的应用——卷积神经网络（CNN）

LeCun 等人开创性地建立了数字分类的神经网络。几家银行使用它来识别支票上的手写数字。



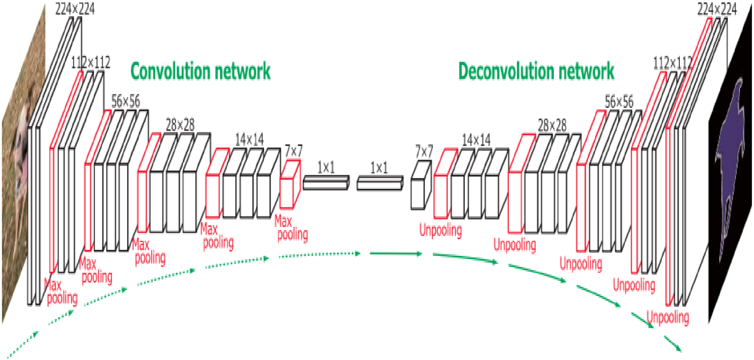
最优化问题的应用——VGGNet

VGG 小组 (牛津) 给出的 19 层卷积神经网络比 AlexNet 更简单层数更深。AlexNet 中的所有大型卷积核都被 3×3 卷积核的级联所取代 (中间是非线性的 ReLU 函数)。在进行两次或三次卷积后进行最大池化操作，每次池化之后，卷积核的数量总是增加一倍。



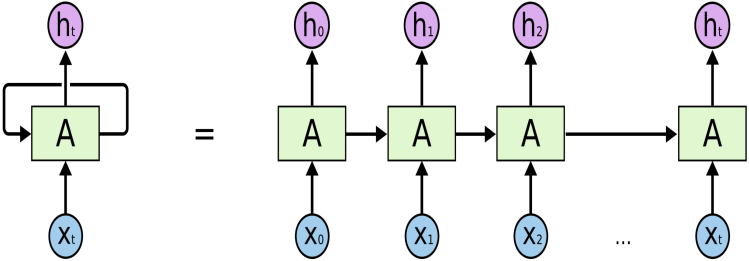
最优化问题的应用——反卷积网络

生成网络是一种特殊的卷积网络，它使用转置卷积，也称为反卷积层。



最优化问题的应用——递归神经网络 (RNN)

RNN 建立在与前馈神经网络相同的计算单元上。RNN 不必分层组织，并且允许定向循环。这样一来，他们就可以拥有内部存储器，从而可以处理顺序数据。如图所示，可以将 RNN 按顺序“展开”，从而将 RNN 转换为常规前馈神经网络。



最优化问题的应用——深度学习中的优化算法

随机梯度类算法

- pytorch/caffe2 里实现的算法有 Adadelta，Adagrad，Adam，Nesterov，RMSProp，YellowFin
- pytorch/torch 里有：SGD，ASGD，Adagrad，RMSProp，Adadelta，Adam，Adamax
- tensorflow 实现的算法有：Adadelta，AdagradDA，Adagrad，ProximalAdagrad，Ftrl，Momentum，adam，Momentum，CenteredRMSProp

最优化问题的应用——深度学习中的随机优化算法

考虑问题 $\min\limits_{x\in\mathbb{R}^d} \frac{1}{n}\sum\limits_{i=1}^{n} f_i(x)$

参考：Chapter 8 in http://www.deeplearningbook.org/

- 梯度下降 GD

$$x^{t+1} = x^t - \frac{\alpha^t}{n}\sum_{i=1}^{n}\nabla f_i(x^t)$$

- 随机梯度下降 SGD

$$x^{t+1} = x^t - \alpha^t\nabla f_i(x^t)$$

- 带动量项的 SGD

$$v^{t+1} = \mu^t v^t - \alpha^t\nabla f_i(x^t)$$
$$x^{t+1} = x^t + v^{t+1}$$

什么是凸优化问题

广义定义：凸目标函数，凸约束集

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & x \in \Omega \end{aligned}$$

其中 $f(x)$ 是凸函数，Ω 是凸集。

什么是凸优化问题

广义定义：凸目标函数，凸约束集

$$\begin{aligned} \text{min} \quad & f(x) \\ \text{s.t.} \quad & x \in \Omega \end{aligned}$$

其中 $f(x)$ 是凸函数，Ω 是凸集。
狭义定义：

$$\begin{aligned} \text{min} \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, i = 1, \ldots, m \\ & h_j(x) = 0, j = 1, \ldots, p \end{aligned}$$

其中 $f_i(x), i = 0, \ldots, m$ 是凸函数，$h_j(x), j = 1, \ldots, p$ 是仿射函数。

优化算法的收敛性

由于实际问题往往是没有办法显示求解的，因此常采用迭代算法。

优化算法的收敛性

由于实际问题往往是没有办法显示求解的，因此常采用迭代算法。

- 迭代算法的基本思想是：从一个初始点 x_0 出发，按照某种给定的规则进行迭代，得到一个序列 $\{x_k\}$。如果迭代在有限步内终止，那么最后一个点就是优化问题的解。如果迭代点列是无穷集合，那么希望该序列的极限点（或者聚点）则为优化问题的解。

优化算法的收敛性

由于实际问题往往是没有办法显示求解的，因此常采用迭代算法。

- 迭代算法的基本思想是：从一个初始点 x_0 出发，按照某种给定的规则进行迭代，得到一个序列 $\{x_k\}$。如果迭代在有限步内终止，那么最后一个点就是优化问题的解。如果迭代点列是无穷集合，那么希望该序列的极限点（或者聚点）则为优化问题的解。
- 在算法设计中，还需考虑算法产生的点列是否收敛到优化问题的解。给定初始点 x_0，记算法迭代产生的点列为 $\{x_k\}$。如果 $\{x_k\}$ 在某种范数 $\|\cdot\|$ 的意义下满足 $\lim_{k\to\infty}\|x_k - x^\star\| = 0$，且收敛的点 $x^\star$ 为一个局部（全局）极小解，那么我们称该点列收敛到局部（全局）极小解，相应的算法称为是**依点列收敛到局部（全局）极小解**的。

优化算法的收敛性

- 进一步地，如果从任意初始点 x_0 出发，算法都是依点列收敛到局部（全局）极小解的，我们称该算法是**全局依点列收敛到局部（全局）极小解**的。记对应的函数值序列 $\{f(x_k)\}$，可以定义算法的**（全局）依函数值收敛到局部（全局）极小值**的概念

优化算法的收敛性

- 进一步地，如果从任意初始点 x_0 出发，算法都是依点列收敛到局部（全局）极小解的，我们称该算法是**全局依点列收敛到局部（全局）极小解**的。记对应的函数值序列 $\{f(x_k)\}$，可以定义算法的**（全局）依函数值收敛到局部（全局）极小值**的概念
- 对于凸优化问题，因为其任何局部最优解都为全局最优解，算法的收敛性都是相对于其全局极小而言的。

优化算法的收敛性

- 进一步地，如果从任意初始点 x_0 出发，算法都是依点列收敛到局部（全局）极小解的，我们称该算法是**全局依点列收敛到局部（全局）极小解**的。记对应的函数值序列 $\{f(x_k)\}$，可以定义算法的**（全局）依函数值收敛到局部（全局）极小值**的概念
- 对于凸优化问题，因为其任何局部最优解都为全局最优解，算法的收敛性都是相对于其全局极小而言的。
- 除了点列和函数值的收敛外，实际中常用的还有每个迭代点的最优性条件（如无约束优化问题中的梯度范数，约束优化问题中的最优性条件违反度等等）的收敛.

优化算法的收敛性

- 进一步地，如果从任意初始点 x_0 出发，算法都是依点列收敛到局部（全局）极小解的，我们称该算法是**全局依点列收敛到局部（全局）极小解**的。记对应的函数值序列 $\{f(x_k)\}$，可以定义算法的**（全局）依函数值收敛到局部（全局）极小值**的概念
- 对于凸优化问题，因为其任何局部最优解都为全局最优解，算法的收敛性都是相对于其全局极小而言的。
- 除了点列和函数值的收敛外，实际中常用的还有每个迭代点的最优性条件（如无约束优化问题中的梯度范数，约束优化问题中的最优性条件违反度等等）的收敛.
- 对于带约束的情形，给定初始点 x_0，算法产生的点列 $\{x_k\}$ 不一定是可行的（即 $x_k \in \mathcal{X}$ 未必对任意 k 成立）。考虑到约束违反的情形，我们需要保证 $\{x_k\}$ 在收敛到 $x^\star$ 的时候，其违反度是可接受的。除此要求之外，算法的收敛性的定义和无约束情形相同。

算法的收敛速度

设 $\{x_k\}$ 为算法产生的迭代点列且收敛于 $x^\star$

- 算法（点列）Q-**线性收敛**：对充分大的 k 有

$$\frac{\|x_{k+1}-x^\star\|}{\|x_k-x^\star\|}\le a, \quad a\in(0,1)$$

算法的收敛速度

设 $\{x_k\}$ 为算法产生的迭代点列且收敛于 $x^\star$

- 算法（点列）Q-**线性收敛**：对充分大的 k 有

$$\frac{\|x_{k+1}-x^\star\|}{\|x_k-x^\star\|} \le a, \quad a \in (0,1)$$

- 算法（点列）Q-**超线性收敛**：对充分大的 k 有

$$\lim_{k\to\infty} \frac{\|x_{k+1}-x^\star\|}{\|x_k-x^\star\|} = 0,$$

算法的收敛速度

设 $\{x_k\}$ 为算法产生的迭代点列且收敛于 $x^\star$

- 算法（点列）Q-**线性收敛**：对充分大的 k 有

$$\frac{\|x_{k+1}-x^\star\|}{\|x_k-x^\star\|}\le a,\qquad a\in(0,1)$$

- 算法（点列）Q-**超线性收敛**：对充分大的 k 有

$$\lim_{k\to\infty}\frac{\|x_{k+1}-x^\star\|}{\|x_k-x^\star\|}=0,$$

- 算法（点列）Q-**次线性收敛**：对充分大的 k 有

$$\lim_{k\to\infty}\frac{\|x_{k+1}-x^\star\|}{\|x_k-x^\star\|}=1,$$

算法的收敛速度

设 $\{x_k\}$ 为算法产生的迭代点列且收敛于 $x^\star$

- 算法（点列）Q-**线性收敛**：对充分大的 k 有

$$\frac{\|x_{k+1}-x^\star\|}{\|x_k-x^\star\|}\le a,\quad a\in(0,1)$$

- 算法（点列）Q-**超线性收敛**：对充分大的 k 有

$$\lim_{k\to\infty}\frac{\|x_{k+1}-x^\star\|}{\|x_k-x^\star\|}=0,$$

- 算法（点列）Q-**次线性收敛**：对充分大的 k 有

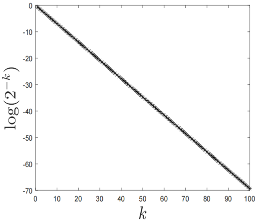
$$\lim_{k\to\infty}\frac{\|x_{k+1}-x^\star\|}{\|x_k-x^\star\|}=1,$$

- 算法（点列）Q-**二次收敛**：对充分大的 k 有

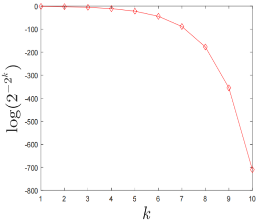
$$\frac{\|x_{k+1}-x^\star\|}{\|x_k-x^\star\|^2}\le a,\quad a>0$$

算法的收敛速度

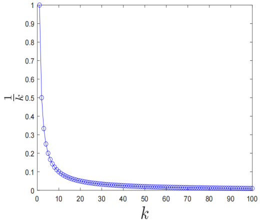
我们举例来更直观地展示不同的 Q-收敛速度，见下图（图中对所考虑的点列作了适当的变换）。点列 $\{2^{-k}\}$ 是 Q-线性收敛的，点列 $\{2^{-2^k}\}$ 是 Q-二次收敛的（也是 Q-超线性收敛的），点列 $\{\frac{1}{k}\}$ 是 Q-次线性收敛的。一般来说，具有 Q-超线性收敛速度和 Q-二次收敛速度的算法是收敛较快的。



(a) Q-线性收敛



(b) Q-二次收敛



(c) Q-次线性收敛

算法的收敛速度

- 算法（点列）R-**线性收敛**：设 $\{x_k\}$ 为算法产生的迭代点列且收敛于 $x^\star$，若存在 Q-线性收敛于 0 的非负序列 $\{t_k\}$ 并且

$$\|x_k - x^\star\| \le t_k,$$

对任意的 k 成立。类似地，可定义 R-**超线性收敛**和 R-**二次收敛**等收敛速度。从 R-收敛速度的定义可以看出序列 $\{\|x_k - x^\star\|\}$ 被另一趋于 0 的序列 $\{t_k\}$ 控制。当知道 t_k 的形式时，我们也称算法（点列）的收敛速度为 $\mathcal{O}(t_k)$。

算法的收敛速度

- 算法（点列）R-**线性收敛**：设 $\{x_k\}$ 为算法产生的迭代点列且收敛于 $x^\star$，若存在 Q-线性收敛于 0 的非负序列 $\{t_k\}$ 并且

$$\|x_k - x^\star\| \leq t_k,$$

对任意的 k 成立。类似地，可定义 R-**超线性收敛**和 R-**二次收敛**等收敛速度。从 R-收敛速度的定义可以看出序列 $\{\|x_k - x^\star\|\}$ 被另一趋于 0 的序列 $\{t_k\}$ 控制。当知道 t_k 的形式时，我们也称算法（点列）的收敛速度为 $\mathcal{O}(t_k)$。

- **算法复杂度**：设 $x^\star$ 为全局极小点，某一算法产生的迭代序列 $\{x_k\}$ 满足

$$f(x_k) - f(x^\star) \leq \frac{c}{\sqrt{k}}, \quad \forall k > 0$$

其中 $c > 0$ 为常数。如果需要计算算法满足精度 $f(x_k) - f(x^\star) \leq \varepsilon$ 所需的迭代次数，只需令 $\frac{c}{\sqrt{k}} \leq \varepsilon$ 则得到 $k \geq \frac{c^2}{\varepsilon^2}$，因此该优化算法对应的（迭代次数）复杂度为 $N(\varepsilon) = \mathcal{O}\left(\frac{1}{\varepsilon^2}\right)$。

优化算法的收敛准则

为了使算法能在有限步内终止，一般会通过一些收敛准则来保证迭代停在问题的一定精度逼近解上。

- 对于无约束优化问题，常用的收敛准则有

$$\frac{f(x_k) - f^\star}{\max\{|f^\star|, 1\}} \leq \varepsilon_1, \quad \|\nabla f(x_k)\| \leq \varepsilon_2$$

其中 $\varepsilon_1, \varepsilon_2$ 为给定的很小的正数，$\|\cdot\|$ 表示某种范数（这里可以简单理解为 ℓ_2 范数：$\|x\|_2 = (\sum_{i=1}^n x_i^2)^{\frac{1}{2}}$，$f^\star$ 为函数 f 的最小值，$\nabla f(x_k)$ 表示函数 f 在点 x 处的梯度。

优化算法的收敛准则

为了使算法能在有限步内终止，一般会通过一些收敛准则来保证迭代停在问题的一定精度逼近解上。

- 对于无约束优化问题，常用的收敛准则有

$$\frac{f(x_k) - f^\star}{\max\{|f^\star|, 1\}} \le \varepsilon_1, \quad \|\nabla f(x_k)\| \le \varepsilon_2$$

其中 $\varepsilon_1, \varepsilon_2$ 为给定的很小的正数，$\|\cdot\|$ 表示某种范数（这里可以简单理解为 ℓ_2 范数：$\|x\|_2 = (\sum_{i=1}^n x_i^2)^{\frac{1}{2}}$，$f^\star$ 为函数 f 的最小值，$\nabla f(x_k)$ 表示函数 f 在点 x 处的梯度。

- 对于约束优化问题，还需要考虑约束违反度。

优化算法的收敛准则

为了使算法能在有限步内终止，一般会通过一些收敛准则来保证迭代停在问题的一定精度逼近解上。

- 对于无约束优化问题，常用的收敛准则有

$$\frac{f(x_k) - f^\star}{\max\{|f^\star|, 1\}} \le \varepsilon_1, \quad \|\nabla f(x_k)\| \le \varepsilon_2$$

  其中 $\varepsilon_1, \varepsilon_2$ 为给定的很小的正数，$\|\cdot\|$ 表示某种范数（这里可以简单理解为 ℓ_2 范数：$\|x\|_2 = (\sum_{i=1}^n x_i^2)^{\frac{1}{2}}$，$f^\star$ 为函数 f 的最小值，$\nabla f(x_k)$ 表示函数 f 在点 x 处的梯度。
- 对于约束优化问题，还需要考虑约束违反度。
- 除了约束违反度，还需要考虑 x_k 与最优解之间的距离，如函数值与最优值的相对误差。但由于一般情况下事先无法获取最优解，在最优解唯一的情况下一般使用某种基准算法来得到 $x^\star$ 的一个估计，之后计算其与 x_k 的距离以评价算法的性能。

优化算法的收敛准则

- 因为约束的存在，不能简单地用目标函数的梯度来判断最优性，实际中采用的判别准则是点的最优性条件的违反度。

优化算法的收敛准则

- 因为约束的存在，不能简单地用目标函数的梯度来判断最优性，实际中采用的判别准则是点的最优性条件的违反度。
- 对于具体的算法，根据其设计的出发点，不一定能得到一个高精度的逼近解。为了避免无用的计算开销，需要一些停机准则来及时停止算法迭代。常用的停止准则有

$$\frac{\|x_{k+1}-x_k\|}{\max\{\|x_k\|,1|\}} \leq \varepsilon_3, \quad \frac{f(x_{k+1})-f(x_k)}{\max\{|f(x_k)|,1\}} \leq \varepsilon_4$$

其分别表示相邻迭代点和其对应目标函数值的相对误差很小。

优化算法的收敛准则

- 因为约束的存在，不能简单地用目标函数的梯度来判断最优性，实际中采用的判别准则是点的最优性条件的违反度。
- 对于具体的算法，根据其设计的出发点，不一定能得到一个高精度的逼近解。为了避免无用的计算开销，需要一些停机准则来及时停止算法迭代。常用的停止准则有

$$\frac{\|x_{k+1}-x_k\|}{\max\{\|x_k\|,1\|\}} \le \varepsilon_3, \quad \frac{f(x_{k+1})-f(x_k)}{\max\{|f(x_k)|,1\}} \le \varepsilon_4$$

  其分别表示相邻迭代点和其对应目标函数值的相对误差很小。
- 在算法设计中，这两个条件往往只能反映迭代点列接近收敛，但不能代表收敛到优化问题的最优解。

凸优化的主要内容

1. 凸集、凸函数、凸优化问题定义；

凸优化的主要内容

1. 凸集、凸函数、凸优化问题定义；
2. 凸优化理论知识：KKT 条件、对偶性 Duality、鞍点 Saddle Point、函数共轭 Conjugate 等理论工具；

凸优化的主要内容

1. 凸集、凸函数、凸优化问题定义；
2. 凸优化理论知识：KKT 条件、对偶性 Duality、鞍点 Saddle Point、函数共轭 Conjugate 等理论工具；
3. 若干凸优化算法。

章节目录

向量求导

- Gradient：$f \in C^1$，定义 $f(x)$ 的梯度为一个 n 维向量，且对应的第 i 个分量为 $\frac{\partial f}{\partial \xi_i}(x)$。
- Hessian：$f \in C^2$，定义 $f(x)$ 的海塞矩阵为一个 $n \times n$ 维的矩阵，且对应的第 (i,j) 个元素为 $\frac{\partial^2 f}{\partial \xi_i \partial \xi_j}(x)$。

向量求导

- Gradient：$f \in C^1$，定义 $f(x)$ 的梯度为一个 n 维向量，且对应的第 i 个分量为 $\frac{\partial f}{\partial \xi_i}(x)$。
- Hessian：$f \in C^2$，定义 $f(x)$ 的海塞矩阵为一个 $n \times n$ 维的矩阵，且对应的第 (i, j) 个元素为 $\frac{\partial^2 f}{\partial \xi_i \partial \xi_j}(x)$。

例：考虑 $f(x) = x^\top x, x \in \mathbb{R}^n$，求 $\nabla f(x)$
由于 $x^\top x = \sum_{i=1}^n x_i^2$，所以 $\frac{\partial f}{\partial x_i} = 2x_i$，因此，$\nabla f(x) = 2x$

向量范数的定义

令记号 $\|\cdot\|:\mathbb{R}^n\rightarrow\mathbb{R}^+$ 是一种非负函数，如果它满足：

- 正定性：对于一切定义域内的 $v\in\mathbb{R}^n$，有 $\|v\|\geq 0$，且 $\|v\|=0\Leftrightarrow v=0_{n\times 1}$；
- 齐次性：对于一切 $v\in\mathbb{R}^n$ 和 $\alpha\in\mathbb{R}$，有 $\|\alpha v\|=|\alpha|\|v\|$；
- 三角不等式：对于一切 $v,w\in\mathbb{R}^n$, 均成立 $\|v+w\|\leq\|v\|+\|w\|$。

则称 $\|\cdot\|$ 是定义在向量空间 $\mathbb{R}^n$ 上的向量范数。

向量范数的定义

令记号 $\|\cdot\|:\mathbb{R}^n\to\mathbb{R}^+$ 是一种非负函数，如果它满足：

- 正定性：对于一切定义域内的 $v\in\mathbb{R}^n$，有 $\|v\|\geq 0$，且 $\|v\|=0\Leftrightarrow v=0_{n\times 1}$；
- 齐次性：对于一切 $v\in\mathbb{R}^n$ 和 $\alpha\in\mathbb{R}$，有 $\|\alpha v\|=|\alpha|\|v\|$；
- 三角不等式：对于一切 $v,w\in\mathbb{R}^n$, 均成立 $\|v+w\|\leq\|v\|+\|w\|$。

则称 $\|\cdot\|$ 是定义在向量空间 $\mathbb{R}^n$ 上的向量范数。

最常用的向量范数即我们熟知的 ℓ_p 范数（其中 $p\geq 1$）：

$$\|v\|_p=\left(\sum_{i=1}^{n}|v_i|^p\right)^{\frac{1}{p}}$$

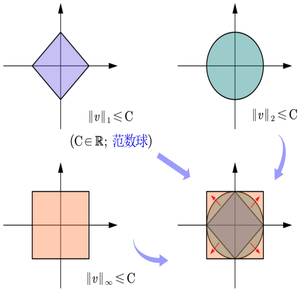
，这也是 ℓ^p 空间中最常用的距离范数。

当 $p=\infty$ 时，距离范数定义为向量所有分量的最大值，即

$$\|v\|_\infty=\max_{1\leq j\leq n}|v_{(j)}|$$

。

向量范数的定义



- 容易看出, $p=\infty$ 时, 有关"最大值"的定义要求向量的分量是有限的. 在一般化的空间中, 这一要求很可能不成立, 此时我们只需将"最大值"更换成"上确界"即可。
- 向量范数度量的是 v 与零点之间的距离. 在实际应用时, 我们通常使用 $p=1,2,\infty$ 的情形, 即分别使用 $\|v\|_1$, $\|v\|_2$, $\|v\|_\infty$ 度量 v 在不同意义下的距离, 这是因为它们具有鲜明的度量特征.
- 左图是它们各自的范数球实例, 请想一想不同范数所度量的距离分别具有怎样的特征? 这些特征分别适用于度量什么情形?

加权的向量范数

给定任意的向量范数 $\|v\|$，可以定义基于矩阵 $W \in \mathbb{R}^{n\times n}$ 的加权范数

$$\|v\|_W = \|Wv\|$$

，特别地，若 W 是对角矩阵，则 W 的对角元分别对应向量各分量的权系数。

加权的向量范数

给定任意的向量范数 $\|v\|$，可以定义基于矩阵 $W \in \mathbb{R}^{n\times n}$ 的加权范数

$$\|v\|_W = \|Wv\|$$

，特别地，若 W 是对角矩阵，则 W 的对角元分别对应向量各分量的权系数。

P-范数

P-范数是一类由正定矩阵诱导的向量范数。

给定任意的向量范数 $\|v\|$，可以基于正定矩阵 $P \in \mathbb{R}^{n\times n}$ 诱导

$$\|v\|_P = \sqrt{v^\top P v}$$

这即为 v 的 P-范数，记为 $\|v\|_P$。

P-范数

P-范数是一类由正定矩阵诱导的向量范数。

给定任意的向量范数 $\|v\|$，可以基于正定矩阵 $P \in \mathbb{R}^{n\times n}$ 诱导

$$\|v\|_P = \sqrt{v^\top P v}$$

这即为 v 的 P-范数，记为 $\|v\|_P$。

根据正定矩阵的定义，容易验证 P-范数是合法的向量范数。请验证 P-范数是关于" 三角不等式" 合法的。
事实上, 若规定 P 仍是对称的, 则 P-范数是一类以对称正定矩阵加权的向量范数:

$$\|v\|_P = \sqrt{v^\top P v} = \sqrt{v^\top P^{\frac{1}{2}} P^{\frac{1}{2}} v} = \sqrt{\left(P^{\frac{1}{2}} v\right)^\top \left(P^{\frac{1}{2}} v\right)} = \|v\|_{2,P^{\frac{1}{2}}}$$

Cauchy 不等式

设 $a, b \in \mathbb{R}^n$，则

$$|a^\top b| \leq \|a\|_2 \|b\|_2$$

且等号成立的条件是 a 与 b 线性相关。

Cauchy 不等式

设 $a, b \in \mathbb{R}^n$，则

$$|a^\top b| \le \|a\|_2 \|b\|_2$$

且等号成立的条件是 a 与 b 线性相关。

Cauchy 不等式在估值和放缩等不同的情形下具有重要的作用，在本课程后续的算法分析中有所体现。我们下面简证其成立。

矩阵范数的定义

矩阵范数的定义是将向量范数的定义推广至 $\mathbb{R}^{m\times n}$ 所直接形成的

令记号 $\|\cdot\|:\mathbb{R}^{m\times n}\to\mathbb{R}^{+}$ 是一种非负函数，如果它满足：

- 正定性：对于一切定义域内的 $A\in\mathbb{R}^{m\times n}$，有 $\|A\|\geq 0$，且 $\|A\|=0\Leftrightarrow A=0_{m\times n}$;
- 齐次性：对于一切 $A\in\mathbb{R}^{m\times n}$ 和 $\alpha\in\mathbb{R}$，有 $\|\alpha A\|=|\alpha|\|A\|$;
- 三角不等式：对于一切 $A,B\in\mathbb{R}^{m\times n}$, 均成立 $\|A+B\|\leq\|A\|+\|B\|$。

则称 $\|\cdot\|$ 是定义在向量空间 $\mathbb{R}^{m\times n}$ 上的矩阵范数。

矩阵范数的定义

矩阵范数的定义是将向量范数的定义推广至 $\mathbb{R}^{m\times n}$ 所直接形成的

令记号 $\|\cdot\|:\mathbb{R}^{m\times n}\to\mathbb{R}^{+}$ 是一种非负函数，如果它满足：

- 正定性：对于一切定义域内的 $A\in\mathbb{R}^{m\times n}$，有 $\|A\|\geq 0$，且 $\|A\|=0\Leftrightarrow A=0_{m\times n}$；
- 齐次性：对于一切 $A\in\mathbb{R}^{m\times n}$ 和 $\alpha\in\mathbb{R}$，有 $\|\alpha A\|=|\alpha|\|A\|$；
- 三角不等式：对于一切 $A,B\in\mathbb{R}^{m\times n}$，均成立 $\|A+B\|\leq\|A\|+\|B\|$。

则称 $\|\cdot\|$ 是定义在向量空间 $\mathbb{R}^{m\times n}$ 上的矩阵范数。

矩阵的 ℓ_p 范数

和向量的 ℓ_p 范数类似，矩阵的 ℓ_p 范数也与矩阵中各分量的信息密切相关。在接下来的定义中，我们认为 a_{ij} 是矩阵 A 的第 i 行 j 列的分量。

- $p=1$ 时，$\|A\|_1=\sum_{i=1}^m\sum_{j=1}^n|a_{ij}|$；
- $p=2$ 时，又称此时定义的“2-范数”为 Frobenius 范数（F-范数），记为 $\|A\|_F$，具体形式为 $\|A\|_F=\sqrt{\mathrm{Tr}(AA^\top)}=\sqrt{\sum_{i=1}^m\sum_{j=1}^n a_{ij}^2}$。

F-范数具有正交不变性。

对于任意的正交矩阵 $\mathcal{U}\in\mathbb{R}^{m\times m},\mathcal{V}\in\mathbb{R}^{n\times n}$，$\|\mathcal{U}A\mathcal{V}\|_F^2=\|A\|_F^2$ 成立。

矩阵的 ℓ_p 范数

和向量的 ℓ_p 范数类似，矩阵的 ℓ_p 范数也与矩阵中各分量的信息密切相关。在接下来的定义中，我们认为 a_{ij} 是矩阵 A 的第 i 行 j 列的分量。

- $p=1$ 时，$\|A\|_1=\sum_{i=1}^m\sum_{j=1}^n|a_{ij}|$;
- $p=2$ 时，又称此时定义的“2-范数”为 Frobenius 范数（F-范数），记为 $\|A\|_F$，具体形式为 $\|A\|_F=\sqrt{\mathrm{Tr}(AA^\top)}=\sqrt{\sum_{i=1}^m\sum_{j=1}^n a_{ij}^2}$。

F-范数具有正交不变性。

对于任意的正交矩阵 $\mathcal{U}\in\mathbb{R}^{m\times m},\mathcal{V}\in\mathbb{R}^{n\times n}$，$\|\mathcal{U}A\mathcal{V}\|_F^2=\|A\|_F^2$ 成立。

请自行证明上述定理。
提示：$\mathrm{Tr}(AB)=\mathrm{Tr}(BA)$。

核范数

上面介绍的范数是我们在数值分析或泛函分析中最常见的范数形式. 在最优化理论分析中, 我们另外规定核范数以衡量矩阵的秩的大小。

给定矩阵 $A \in \mathbb{R}^{m\times n}$，其核范数定义为

$$\|A\|_* = \sum_{i=1}^{r} \sigma_i$$

其中 $\sigma_i(i=1,\dots,r)$ 为 A 的所有非零奇异值，$r = \mathrm{rank}(A)$。

核范数

上面介绍的范数是我们在数值分析或泛函分析中最常见的范数形式. 在最优化理论分析中, 我们另外规定核范数以衡量矩阵的秩的大小。

给定矩阵 $A \in \mathbb{R}^{m\times n}$，其核范数定义为

$$\|A\|_* = \sum_{i=1}^{r} \sigma_i$$

其中 $\sigma_i(i=1,\ldots,r)$ 为 A 的所有非零奇异值，$r = \mathrm{rank}(A)$。

请自行证明如上规定的核范数是满足三角不等式的。

直线和线段

直线:

$$x_1 \neq x_2 \in \mathbb{R}^n, \theta \in \mathbb{R} \Rightarrow y = \theta x_1 + (1-\theta)x_2 = x_2 + \theta(x_1 - x_2)$$

线段:

$$x_1 \neq x_2 \in \mathbb{R}^n, \theta \in [0,1] \Rightarrow y = \theta x_1 + (1-\theta)x_2$$

仿射集

仿射集：一个集合 C 是仿射集，若 $\forall x_1, x_2 \in C$，则连接 x_1 和 x_2 的直线也在集合内。

仿射集

仿射集：一个集合 C 是仿射集，若 $\forall x_1, x_2 \in C$，则连接 x_1 和 x_2 的直线也在集合内。

直线、$\mathbb{R}^n$ 是仿射集；线段、闭合图形不是仿射集。

仿射集

仿射集：一个集合 C 是仿射集，若 $\forall x_1, x_2 \in C$，则连接 x_1 和 x_2 的直线也在集合内。

直线、$\mathbb{R}^n$ 是仿射集；线段、闭合图形不是仿射集。

仿射组合：设 $x_1, \ldots, x_k \in C$, $\theta_1, \ldots, \theta_k \in \mathbb{R}$, $\theta_1 + \cdots + \theta_k = 1$, 那么

$$\theta_1 x_1 + \cdots + \theta_k x_k$$

是点 $x_1, \ldots, x_k$ 的仿射组合。如果 C 是仿射集，那么点

$$\theta_1 x_1 + \cdots + \theta_k x_k \in C.$$

仿射集

仿射集：一个集合 C 是仿射集，若 $\forall x_1, x_2 \in C$，则连接 x_1 和 x_2 的直线也在集合内。

直线、$\mathbb{R}^n$ 是仿射集；线段、闭合图形不是仿射集。

仿射组合：设 $x_1, \ldots, x_k \in C$, $\theta_1, \ldots, \theta_k \in \mathbb{R}$, $\theta_1 + \cdots + \theta_k = 1$, 那么

$$\theta_1 x_1 + \cdots + \theta_k x_k$$

是点 $x_1, \ldots, x_k$ 的仿射组合。如果 C 是仿射集，那么点

$$\theta_1 x_1 + \cdots + \theta_k x_k \in C.$$

仿射包：$\mathrm{aff}\, C = \{\theta_1 x_1 + \cdots + \theta_k x_k \mid \forall x_1, \ldots, x_k \in C, \theta_1, \ldots, \theta_k \in \mathbb{R}, \theta_1 + \cdots + \theta_k = 1\}$

凸集

凸集 (Convex sets)：一个集合 C 是凸集，当任意两点之间的线段仍在 C 内。

凸集

凸集 (Convex sets)：一个集合 C 是凸集，当任意两点之间的线段仍在 C 内。

C 为凸集 $\Leftrightarrow \forall x_1, x_2 \in C, \forall\theta, \theta \in [0,1], \theta x_1 + (1-\theta)x_2 \in C$

凸集

凸集 (Convex sets)：一个集合 C 是凸集，当任意两点之间的线段仍在 C 内。

C 为凸集 $\Leftrightarrow \forall x_1, x_2 \in C, \forall\theta, \theta \in [0,1], \theta x_1 + (1-\theta)x_2 \in C$

凸组合：
$\theta_1 x_1 + \cdots + \theta_k x_k, \theta_1, \ldots, \theta_k \in \mathbb{R}, \theta_1 + \cdots + \theta_k = 1, \theta_1, \ldots, \theta_k \in [0,1]$

凸集

凸集 (Convex sets)：一个集合 C 是凸集，当任意两点之间的线段仍在 C 内。

C 为凸集 $\Leftrightarrow \forall x_1, x_2 \in C, \forall\theta, \theta \in [0,1], \theta x_1 + (1-\theta)x_2 \in C$

凸组合：
$\theta_1 x_1 + \cdots + \theta_k x_k, \theta_1, \ldots, \theta_k \in \mathbb{R}, \theta_1 + \cdots + \theta_k = 1, \theta_1, \ldots, \theta_k \in [0,1]$

C 为凸集 $\Leftrightarrow$ 任意元素的凸组合 $\in C$

凸集

凸集 (Convex sets)：一个集合 C 是凸集，当任意两点之间的线段仍在 C 内。

C 为凸集 $\Leftrightarrow \forall x_1, x_2 \in C, \forall\theta, \theta \in [0,1], \theta x_1 + (1-\theta)x_2 \in C$

凸组合：
$\theta_1 x_1 + \cdots + \theta_k x_k, \theta_1, \ldots, \theta_k \in \mathbb{R}, \theta_1 + \cdots + \theta_k = 1, \theta_1, \ldots, \theta_k \in [0,1]$

C 为凸集 $\Leftrightarrow$ 任意元素的凸组合 $\in C$

凸包：$C \in \mathbb{R}^n$, $\operatorname{Conv} C = \{\theta_1 x_1 + \cdots + \theta_k x_k \mid \forall x_1, \ldots, x_k \in C, \forall\theta_1, \ldots, \theta_k \in [0,1], \theta_1 + \cdots + \theta_k = 1\}$

凸锥

C 为锥 $\Leftrightarrow \forall x \in C, \forall \theta \geq 0$, 有 $\theta x \in C$

凸锥

C 为**锥** $\Leftrightarrow \forall x \in C, \forall \theta \geq 0$, 有 $\theta x \in C$

C 为**凸锥** $\Leftrightarrow \forall x_1, x_2 \in C, \forall \theta_1, \theta_2 \geq 0$, 有 $\theta_1 x_1 + \theta_2 x_2 \in C$

凸锥

C **为锥** $\Leftrightarrow \forall x \in C, \forall \theta \geq 0$, 有 $\theta x \in C$

C **为凸锥** $\Leftrightarrow \forall x_1, x_2 \in C, \forall \theta_1, \theta_2 \geq 0$, 有 $\theta_1 x_1 + \theta_2 x_2 \in C$

凸锥组合: $\theta_1 x_1 + \cdots + \theta_k x_k, \theta_1, \ldots, \theta_k \geq 0$

凸锥

C 为锥 $\Leftrightarrow \forall x \in C, \forall \theta \geq 0$，有 $\theta x \in C$

C 为**凸锥** $\Leftrightarrow \forall x_1, x_2 \in C, \forall \theta_1, \theta_2 \geq 0$，有 $\theta_1 x_1 + \theta_2 x_2 \in C$

凸锥组合：$\theta_1 x_1 + \cdots + \theta_k x_k, \theta_1, \ldots, \theta_k \geq 0$

凸锥包：$\{\theta_1 x_1 + \cdots + \theta_k x_k \mid x_1, \ldots, x_k \in C, \theta_1, \ldots, \theta_k \geq 0\}$

仿射、凸和凸锥组合的关系

$\theta_1 x_1 + \cdots + \theta_k x_k$

1. 仿射组合：$\forall \theta_1, \ldots, \theta_k,\ \theta_1 + \cdots + \theta_k = 1$
2. 凸组合：$\forall \theta_1, \ldots, \theta_k,\ \theta_1 + \cdots + \theta_k = 1,\ \theta_1, \ldots, \theta_k \in [0, 1]$
3. 凸锥组合：$\forall \theta_1, \ldots, \theta_k,\ \theta_1, \ldots, \theta_k \geq 0$

仿射、凸和凸锥组合的关系

$\theta_1 x_1 + \cdots + \theta_k x_k$

1. 仿射组合：$\forall \theta_1, \ldots, \theta_k,\ \theta_1 + \cdots + \theta_k = 1$
2. 凸组合：$\forall \theta_1, \ldots, \theta_k,\ \theta_1 + \cdots + \theta_k = 1,\ \theta_1, \ldots, \theta_k \in [0, 1]$
3. 凸锥组合：$\forall \theta_1, \ldots, \theta_k,\ \theta_1, \ldots, \theta_k \geq 0$

$C = \{x\},\ \theta_1 x + \theta_2 x = x$ 仿射集; $\emptyset$ 也是仿射集

仿射、凸和凸锥组合的关系

$\theta_1 x_1 + \cdots + \theta_k x_k$

1. 仿射组合：$\forall \theta_1, \ldots, \theta_k,\ \theta_1 + \cdots + \theta_k = 1$
2. 凸组合：$\forall \theta_1, \ldots, \theta_k,\ \theta_1 + \cdots + \theta_k = 1,\ \theta_1, \ldots, \theta_k \in [0, 1]$
3. 凸锥组合：$\forall \theta_1, \ldots, \theta_k,\ \theta_1, \ldots, \theta_k \geq 0$

$C = \{x\}$, $\theta_1 x + \theta_2 x = x$ 仿射集; $\emptyset$ 也是仿射集

$\emptyset$ 是仿射集、凸集、凸锥

几种重要的凸集

- 超平面：$\{x \mid a^\top x = b\}$, 其中 $x, a \in \mathbb{R}^n, b \in \mathbb{R}, a \neq 0$

几种重要的凸集

- 超平面：$\{x \mid a^\top x = b\}$, 其中 $x, a \in \mathbb{R}^n, b \in \mathbb{R}, a \neq 0$
- 半空间：$\{x \mid a^\top x \leq b\}$, 其中 $x, a \in \mathbb{R}^n, b \in \mathbb{R}, a \neq 0$

几种重要的凸集

- 超平面：$\{x \mid a^\top x = b\}$，其中 $x, a \in \mathbb{R}^n, b \in \mathbb{R}, a \neq 0$
- 半空间：$\{x \mid a^\top x \leq b\}$，其中 $x, a \in \mathbb{R}^n, b \in \mathbb{R}, a \neq 0$
- 球：$B(x_c, r) = \{x \mid \|x - x_c\|_2 \leq r\} = \{x \mid \sqrt{(x - x_c)^\top (x - x_c)} \leq r\}$

几种重要的凸集

- 超平面：$\{x \mid a^\top x = b\}$，其中 $x, a \in \mathbb{R}^n, b \in \mathbb{R}, a \neq 0$
- 半空间：$\{x \mid a^\top x \leq b\}$，其中 $x, a \in \mathbb{R}^n, b \in \mathbb{R}, a \neq 0$

- 球：$B(x_c, r) = \{x \mid \|x - x_c\|_2 \leq r\} = \{x \mid \sqrt{(x - x_c)^\top (x - x_c)} \leq r\}$
- 椭球：$\epsilon(x_c, P) = \{x \mid (x - x_c)^\top P^{-1}(x - x_c) \leq 1\}$，其中 $P \in S^n_{++}(n \times n$ 对称正定矩阵)

几种重要的凸集

- 超平面：$\{x \mid a^\top x = b\}$，其中 $x, a \in \mathbb{R}^n, b \in \mathbb{R}, a \neq 0$
- 半空间：$\{x \mid a^\top x \leq b\}$，其中 $x, a \in \mathbb{R}^n, b \in \mathbb{R}, a \neq 0$

- 球：$B(x_c, r) = \{x \mid \|x - x_c\|_2 \leq r\} = \{x \mid \sqrt{(x - x_c)^\top (x - x_c)} \leq r\}$
- 椭球：$\epsilon(x_c, P) = \{x \mid (x - x_c)^\top P^{-1} (x - x_c) \leq 1\}$，其中 $P \in S^n_{++}$($n \times n$ 对称正定矩阵)

- 多面体 (Polyhedron)：
  $\mathcal{P} = \{x \mid a_j^\top x \leq b_j, j = 1, \ldots, m, c_j^\top x = d_j, j = 1, \ldots, p\}$

几种重要的凸集

- 超平面：$\{x \mid a^\top x = b\}$，其中 $x, a \in \mathbb{R}^n, b \in \mathbb{R}, a \neq 0$
- 半空间：$\{x \mid a^\top x \leq b\}$，其中 $x, a \in \mathbb{R}^n, b \in \mathbb{R}, a \neq 0$

- 球：$B(x_c, r) = \{x \mid \|x - x_c\|_2 \leq r\} = \{x \mid \sqrt{(x - x_c)^\top (x - x_c)} \leq r\}$
- 椭球：$\epsilon(x_c, P) = \{x \mid (x - x_c)^\top P^{-1} (x - x_c) \leq 1\}$，其中 $P \in S^n_{++}$($n \times n$ 对称正定矩阵)

- 多面体 (Polyhedron)：
  $\mathcal{P} = \{x \mid a_j^\top x \leq b_j, j = 1, \ldots, m, c_j^\top x = d_j, j = 1, \ldots, p\}$

- 单纯形 (Simplex)：
  $C = \mathrm{conv}\{v_0, \ldots, v_k\} = \{\theta_0 v_0 + \ldots + \theta_k v_k \mid \theta \succeq 0, \mathbf{1}^\top \theta = 1\}$

几种重要的凸集

- 超平面：$\{x \mid a^\top x = b\}$，其中 $x, a \in \mathbb{R}^n, b \in \mathbb{R}, a \neq 0$
- 半空间：$\{x \mid a^\top x \leq b\}$，其中 $x, a \in \mathbb{R}^n, b \in \mathbb{R}, a \neq 0$

- 球：$B(x_c, r) = \{x \mid \|x - x_c\|_2 \leq r\} = \{x \mid \sqrt{(x - x_c)^\top (x - x_c)} \leq r\}$
- 椭球：$\epsilon(x_c, P) = \{x \mid (x - x_c)^\top P^{-1} (x - x_c) \leq 1\}$，其中 $P \in S^n_{++}$($n \times n$ 对称正定矩阵)

- 多面体 (Polyhedron)：
  $\mathcal{P} = \{x \mid a_j^\top x \leq b_j, j = 1, \ldots, m, c_j^\top x = d_j, j = 1, \ldots, p\}$

- 单纯形 (Simplex)：
  $C = \mathrm{conv}\{v_0, \ldots, v_k\} = \{\theta_0 v_0 + \ldots + \theta_k v_k \mid \theta \succeq 0, \mathbf{1}^\top \theta = 1\}$

- 正半定锥：
  - 对称矩阵集合 $S^n = \{X \in \mathbb{R}^{n \times n} \mid X = X^\top\}$ 是凸锥

几种重要的凸集

- 超平面：$\{x \mid a^\top x = b\}$，其中 $x, a \in \mathbb{R}^n, b \in \mathbb{R}, a \neq 0$
- 半空间：$\{x \mid a^\top x \leq b\}$，其中 $x, a \in \mathbb{R}^n, b \in \mathbb{R}, a \neq 0$

- 球：$B(x_c, r) = \{x \mid \|x - x_c\|_2 \leq r\} = \{x \mid \sqrt{(x - x_c)^\top (x - x_c)} \leq r\}$
- 椭球：$\epsilon(x_c, P) = \{x \mid (x - x_c)^\top P^{-1}(x - x_c) \leq 1\}$，其中 $P \in S^n_{++}$($n \times n$ 对称正定矩阵)

- 多面体 (Polyhedron)：
  $\mathcal{P} = \{x \mid a_j^\top x \leq b_j, j = 1, \ldots, m, c_j^\top x = d_j, j = 1, \ldots, p\}$

- 单纯形 (Simplex)：
  $C = \mathrm{conv}\{v_0, \ldots, v_k\} = \{\theta_0 v_0 + \ldots + \theta_k v_k \mid \theta \succeq 0, \mathbf{1}^\top \theta = 1\}$

- 正半定锥：
  - 对称矩阵集合 $S^n = \{X \in \mathbb{R}^{n \times n} \mid X = X^\top\}$ 是凸锥
  - 对称半正定矩阵集合 $S^n_+ = \{X \in S^n \mid X = X^\top, X \succeq 0\}$ 是凸集，也是凸锥

几种重要的凸集

- 超平面：$\{x \mid a^\top x = b\}$，其中 $x, a \in \mathbb{R}^n, b \in \mathbb{R}, a \neq 0$
- 半空间：$\{x \mid a^\top x \leq b\}$，其中 $x, a \in \mathbb{R}^n, b \in \mathbb{R}, a \neq 0$

- 球：$B(x_c, r) = \{x \mid \|x - x_c\|_2 \leq r\} = \{x \mid \sqrt{(x-x_c)^\top (x-x_c)} \leq r\}$
- 椭球：$\epsilon(x_c, P) = \{x \mid (x-x_c)^\top P^{-1} (x-x_c) \leq 1\}$，其中 $P \in S^n_{++}$($n \times n$ 对称正定矩阵)

- 多面体 (Polyhedron)：
  $\mathcal{P} = \{x \mid a_j^\top x \leq b_j, j = 1, \ldots, m, c_j^\top x = d_j, j = 1, \ldots, p\}$

- 单纯形 (Simplex)：
  $C = \operatorname{conv}\{v_0, \ldots, v_k\} = \{\theta_0 v_0 + \ldots + \theta_k v_k \mid \theta \succeq 0, \mathbf{1}^\top \theta = 1\}$

- 正半定锥：
  - 对称矩阵集合 $S^n = \{X \in \mathbb{R}^{n \times n} \mid X = X^\top\}$ 是凸锥
  - 对称半正定矩阵集合 $S^n_+ = \{X \in S^n \mid X = X^\top, X \succeq 0\}$ 是凸集，也是凸锥
  - 对称正定矩阵集合 $S^n_{++} = \{X \in S^n \mid X = X^\top, X \succ 0\}$ 不是凸锥

保凸变换

交集：若 S_1, S_2 为凸集，则 $S_1 \cap S_2$ 也是凸集。

保凸变换

交集：若 S_1, S_2 为凸集，则 $S_1 \cap S_2$ 也是凸集。
推论：若 S_α 为凸集 $(\forall \alpha \in \mathcal{A})$，则 $\bigcap_{\alpha \in \mathcal{A}} S_\alpha$ 也是凸集。

保凸变换

交集：若 S_1, S_2 为凸集，则 $S_1 \cap S_2$ 也是凸集。
推论：若 S_α 为凸集 $(\forall \alpha \in \mathcal{A})$，则 $\bigcap_{\alpha \in \mathcal{A}} S_\alpha$ 也是凸集。

仿射函数 (线性变换)：
函数 $f: \mathbb{R}^n \to \mathbb{R}^m$ 是仿射的，当 $f(x) = Ax + b$, 其中 $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$。

保凸变换

交集：若 S_1, S_2 为凸集，则 $S_1 \cap S_2$ 也是凸集。
推论：若 S_α 为凸集 $(\forall \alpha \in \mathcal{A})$，则 $\bigcap_{\alpha\in\mathcal{A}} S_\alpha$ 也是凸集。

仿射函数 (线性变换)：
函数 $f : \mathbb{R}^n \to \mathbb{R}^m$ 是仿射的，当 $f(x) = Ax + b$, 其中 $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$。
若 $S \in \mathbb{R}^n$ 为凸集，$f : \mathbb{R}^n \to \mathbb{R}^m$ 是仿射的，则 $f(S) = \{f(x) \mid x \in S\}$ 为凸集。

保凸变换

交集：若 S_1, S_2 为凸集，则 $S_1 \cap S_2$ 也是凸集。
推论：若 S_α 为凸集 $(\forall \alpha \in \mathcal{A})$，则 $\bigcap_{\alpha \in \mathcal{A}} S_\alpha$ 也是凸集。

仿射函数 (线性变换)：
函数 $f : \mathbb{R}^n \to \mathbb{R}^m$ 是仿射的，当 $f(x) = Ax + b$, 其中
$A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$。
若 $S \in \mathbb{R}^n$ 为凸集，$f : \mathbb{R}^n \to \mathbb{R}^m$ 是仿射的，则 $f(S) = \{f(x) \mid x \in S\}$ 为凸集。

缩放和位移：$\alpha S = \{\alpha x \mid x \in S\}$, $S + a = \{x + a \mid x \in S\}$

保凸变换

交集：若 S_1, S_2 为凸集，则 $S_1 \cap S_2$ 也是凸集。
推论：若 S_α 为凸集 $(\forall \alpha \in \mathcal{A})$，则 $\bigcap_{\alpha\in\mathcal{A}} S_\alpha$ 也是凸集。

仿射函数 (线性变换)：
函数 $f : \mathbb{R}^n \to \mathbb{R}^m$ 是仿射的，当 $f(x) = Ax + b$, 其中 $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$。
若 $S \in \mathbb{R}^n$ 为凸集，$f : \mathbb{R}^n \to \mathbb{R}^m$ 是仿射的，则 $f(S) = \{f(x) \mid x \in S\}$ 为凸集。

缩放和位移：$\alpha S = \{\alpha x \mid x \in S\}$, $S + a = \{x + a \mid x \in S\}$

和与笛卡尔积：$S_1 + S_2 = \{x_1 + x_2 \mid x_1 \in S_1, x_2 \in S_2\}$,
$S_1 \times S_2 = \{(x_1, x_2) \mid x_1 \in S_1, x_2 \in S_2\}$

保凸变换

交集：若 S_1, S_2 为凸集，则 $S_1 \cap S_2$ 也是凸集。
推论：若 S_α 为凸集 $(\forall \alpha \in \mathcal{A})$，则 $\bigcap_{\alpha \in \mathcal{A}} S_\alpha$ 也是凸集。

仿射函数 (线性变换)：
函数 $f : \mathbb{R}^n \to \mathbb{R}^m$ 是仿射的，当 $f(x) = Ax + b$, 其中 $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$。
若 $S \in \mathbb{R}^n$ 为凸集，$f : \mathbb{R}^n \to \mathbb{R}^m$ 是仿射的，则 $f(S) = \{f(x) \mid x \in S\}$ 为凸集。

缩放和位移：$\alpha S = \{\alpha x \mid x \in S\}$, $S + a = \{x + a \mid x \in S\}$

和与笛卡尔积：$S_1 + S_2 = \{x_1 + x_2 \mid x_1 \in S_1, x_2 \in S_2\}$,
$S_1 \times S_2 = \{(x_1, x_2) \mid x_1 \in S_1, x_2 \in S_2\}$

线性矩阵不等式 (Linear Matrix Inequalities, LMIs)：
$A(x) = x_1 A_1 + \cdots + x_n A_n \preceq B$, 其中 $x_i, A_i, B \in \mathbb{S}^m$, 则 $\{x \mid A(x) \preceq B\}$ 为凸集。

保凸变换

透视函数： $P:\mathbb{R}^{n+1}\to\mathbb{R}^n$, $\operatorname{dom}P=\mathbb{R}^n\times\mathbb{R}_{++}$,

$$P(z,t)=\frac{z}{t},\ z\in\mathbb{R}^n,\ t\in\mathbb{R}_{++}$$

保凸变换

透视函数：$P : \mathbb{R}^{n+1} \to \mathbb{R}^n$, $\mathrm{dom}\, P = \mathbb{R}^n \times \mathbb{R}_{++}$,

$$P(z, t) = \frac{z}{t}, z \in \mathbb{R}^n, t \in \mathbb{R}_{++}$$

任意凸集的反透视映射仍是凸集。

保凸变换

透视函数： $P:\mathbb{R}^{n+1}\to\mathbb{R}^n$, $\mathrm{dom}\,P=\mathbb{R}^n\times\mathbb{R}_{++}$,

$$P(z,t)=\frac{z}{t},\, z\in\mathbb{R}^n,\, t\in\mathbb{R}_{++}$$

任意凸集的反透视映射仍是凸集。
若 $C\subseteq\mathbb{R}^n$ 为凸集，则 $P^{-1}(C)=\{(x,t)\in\mathbb{R}^{n+1}\mid \frac{x}{t}\in C, t>0\}$。

保凸变换

透视函数： $P:\mathbb{R}^{n+1}\to\mathbb{R}^n$, $\mathrm{dom}\,P=\mathbb{R}^n\times\mathbb{R}_{++}$,

$$P(z,t)=\frac{z}{t},\ z\in\mathbb{R}^n,\ t\in\mathbb{R}_{++}$$

任意凸集的反透视映射仍是凸集。
若 $C\subseteq\mathbb{R}^n$ 为凸集，则 $P^{-1}(C)=\{(x,t)\in\mathbb{R}^{n+1}\mid \frac{x}{t}\in C, t>0\}$。

线性分数函数： 线性分数函数 f(由点 x 先做仿射变换，再做透视得到)

保凸变换

透视函数：$P:\mathbb{R}^{n+1}\to\mathbb{R}^n$, $\mathrm{dom}\,P=\mathbb{R}^n\times\mathbb{R}_{++}$,

$$P(z,t)=\frac{z}{t},\ z\in\mathbb{R}^n,\ t\in\mathbb{R}_{++}$$

任意凸集的反透视映射仍是凸集。
若 $C\subseteq\mathbb{R}^n$ 为凸集，则 $P^{-1}(C)=\{(x,t)\in\mathbb{R}^{n+1}\mid \frac{x}{t}\in C, t>0\}$。

线性分数函数：线性分数函数 f(由点 x 先做仿射变换，再做透视得到)
$g:\mathbb{R}^n\to\mathbb{R}^{m+1}$ 为仿射映射，$g(x)=\begin{bmatrix}A\\C^\top\end{bmatrix}x+\begin{bmatrix}b\\d\end{bmatrix}$, 其中 $A\in\mathbb{R}^{m\times n}$,
$C\in\mathbb{R}^n$, $b\in\mathbb{R}^m$, $d\in\mathbb{R}$
$P:\mathbb{R}^{m+1}\to\mathbb{R}^m$ 为透视变换
$f:\mathbb{R}^n\to\mathbb{R}^m\triangleq P\circ g$ 为透视变换

$$f(x)=\frac{Ax+b}{C^\top x+d},\quad \mathrm{dom}f=\{x\mid C^\top x+d>0\}$$

习题

2.1, 2.2, 2.5, 2.7, 2.10, 2.16, 2.18, 2.19

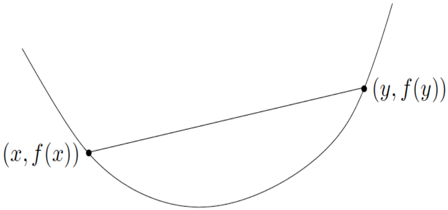
章节目录

凸函数的定义

凸函数 (一)：函数 $f : \mathbb{R}^n \to \mathbb{R}$ 是凸函数，若 dom$f$ 为凸，且对于 $\forall x, y \in$ dom$f$，$0 \leq \theta \leq 1$，都有 $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$。

凸函数的定义

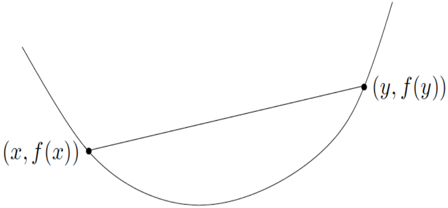
凸函数 (一)：函数 $f : \mathbb{R}^n \to \mathbb{R}$ 是凸函数，若 dom$f$ 为凸，且对于 $\forall x, y \in$ dom$f$，$0 \leq \theta \leq 1$，都有 $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$。



任意两点之间的弦 (线段) 都在函数图像上。

凸函数的定义

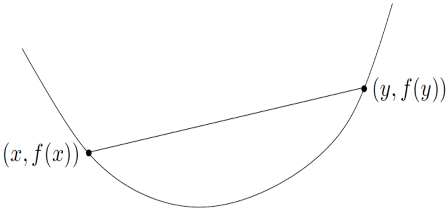
凸函数 (一)：函数 $f:\mathbb{R}^n \to \mathbb{R}$ 是凸函数，若 dom$f$ 为凸，且对于 $\forall x, y \in$ dom$f$，$0 \le \theta \le 1$，都有 $f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta)f(y)$。



任意两点之间的弦 (线段) 都在函数图像上。
凸函数 (二)：函数 $f:\mathbb{R}^n \to \mathbb{R}$ 是凸的，当且仅当其在与其定义域相交的任何直线上都是凸的。(切西瓜)

凸函数的定义

凸函数 (一)：函数 $f: \mathbb{R}^n \to \mathbb{R}$ 是凸函数，若 $\mathrm{dom} f$ 为凸，且对于 $\forall x, y \in \mathrm{dom} f$，$0 \le \theta \le 1$，都有 $f(\theta x + (1-\theta) y) \le \theta f(x) + (1-\theta) f(y)$。



任意两点之间的弦 (线段) 都在函数图像上。
凸函数 (二)：函数 $f: \mathbb{R}^n \to \mathbb{R}$ 是凸的，当且仅当其在与其定义域相交的任何直线上都是凸的。(切西瓜)
$f: \mathbb{R}^n \to \mathbb{R}$ 是凸函数 $\Leftrightarrow \forall x \in \mathrm{dom} f, \forall v,\ g(t) = f(x + tv)$ 为凸。

凸函数的定义

凸函数的扩展：$f:\mathbb{R}^n\to\mathbb{R}$ 为凸函数，$\mathrm{dom}f=C\subseteq\mathbb{R}^n$

$$\tilde{f}=\begin{cases} f(x), & x\in \mathrm{dom}f \\ \infty, & x\notin \mathrm{dom}f \end{cases}$$

凸函数的定义

凸函数的扩展：$f:\mathbb{R}^n\to\mathbb{R}$ 为凸函数，$\mathsf{dom}f=C\subseteq\mathbb{R}^n$

$$\tilde{f}=\begin{cases} f(x), & x\in \mathsf{dom}f \\ \infty, & x\notin \mathsf{dom}f \end{cases}$$

示性函数是凸函数。

凸函数的定义

凸函数的扩展：$f:\mathbb{R}^n\to\mathbb{R}$ 为凸函数，$\mathsf{dom}f=C\subseteq\mathbb{R}^n$

$$\tilde{f}=\begin{cases}f(x), & x\in \mathsf{dom}f\\ \infty, & x\notin \mathsf{dom}f\end{cases}$$

示性函数是凸函数。

凸集 $C\subseteq\mathbb{R}^n$，

$$f_C(x)=\begin{cases}\text{无定义}, & x\notin C\\ 0, & x\in C\end{cases}\quad I_C(x)=\begin{cases}\infty, & x\notin C\\ 0, & x\in C\end{cases}\quad J_C(x)=\begin{cases}1, & x\notin C\\ 0, & x\in C\end{cases}$$

凸函数的定义

凸函数的扩展：$f:\mathbb{R}^n \to \mathbb{R}$ 为凸函数，$\mathrm{dom} f = C \subseteq \mathbb{R}^n$

$$\tilde{f} = \begin{cases} f(x), & x \in \mathrm{dom} f \\ \infty, & x \notin \mathrm{dom} f \end{cases}$$

示性函数是凸函数。

凸集 $C \subseteq \mathbb{R}^n$，

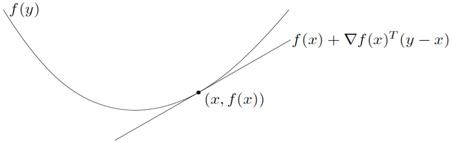
$$f_C(x) = \begin{cases} \text{无定义}, & x \notin C \\ 0, & x \in C \end{cases} \quad I_C(x) = \begin{cases} \infty, & x \notin C \\ 0, & x \in C \end{cases} \quad J_C(x) = \begin{cases} 1, & x \notin C \\ 0, & x \in C \end{cases}$$

凸　　　　凸　　　　既不凸又不凹

凸函数的定义

一阶条件：设 $f:\mathbb{R}^n \to \mathbb{R}$ 可微，即梯度 ∇f 在 dom$f$ 上均存在，则 f 为凸，等价于：

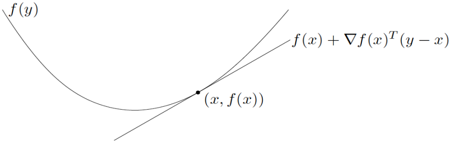
1. dom$f$ 为凸集
2. $f(y) \geq f(x) + \nabla f^{\top}(x)(y-x), \forall x, y \in$ dom$f$



凸函数的定义

一阶条件：设 $f:\mathbb{R}^n \to \mathbb{R}$ 可微，即梯度 ∇f 在 dom$f$ 上均存在，则 f 为凸，等价于：

1. dom$f$ 为凸集
2. $f(y) \geq f(x) + \nabla f^\top(x)(y-x), \forall x, y \in$ dom$f$



二阶条件：设 $f:\mathbb{R}^n \to \mathbb{R}$ 二阶可微，则 f 为凸函数等价于：

1. dom$f$ 为凸集
2. $\nabla^2 f(x) \succeq 0, \forall x \in$ dom$f$，其中 $\nabla^2 f(x)$ 为 Hessian 矩阵

常见的凸函数

- 二次函数
  $f:\mathbb{R}^n \to \mathbb{R}, \mathrm{dom} f=\mathbb{R}^n,\ f(x)=\frac{1}{2}X^\top PX+q^\top X+r,$
  $P\in S^n, q\in\mathbb{R}^n, r\in\mathbb{R}$
  $\nabla^2 f(x)=P$ 正定 $\Leftrightarrow$ 凸函数

常见的凸函数

- 二次函数
  $f:\mathbb{R}^n \to \mathbb{R}, \mathrm{dom} f = \mathbb{R}^n,\ f(x) = \frac{1}{2}X^\top PX + q^\top X + r,$
  $P \in S^n, q \in \mathbb{R}^n, r \in \mathbb{R}$
  $\nabla^2 f(x) = P$ 正定 $\Leftrightarrow$ 凸函数
- 仿射函数
  $f(x) = Ax + b,\ \ \nabla^2 f(x) = 0_{n\times n} \succeq 0$ 既凸又凹

常见的凸函数

- 二次函数
  $f: \mathbb{R}^n \to \mathbb{R}, \mathrm{dom} f = \mathbb{R}^n$, $f(x) = \frac{1}{2} X^\top P X + q^\top X + r$,
  $P \in S^n, q \in \mathbb{R}^n, r \in \mathbb{R}$
  $\nabla^2 f(x) = P$ 正定 $\Leftrightarrow$ 凸函数
- 仿射函数
  $f(x) = Ax + b, \quad \nabla^2 f(x) = 0_{n\times n} \succeq 0$ 既凸又凹
- 指数函数
  $f(x) = e^{ax}, x \in \mathbb{R}, f'(x) = ae^{ax}, f''(x) = a^2 e^{ax} \geq 0$

常见的凸函数

- 二次函数

  $f:\mathbb{R}^n\to\mathbb{R}, \mathrm{dom} f=\mathbb{R}^n,\ f(x)=\frac{1}{2}X^\top PX+q^\top X+r,$
  $P\in S^n, q\in\mathbb{R}^n, r\in\mathbb{R}$
  $\nabla^2 f(x)=P$ 正定 $\Leftrightarrow$ 凸函数
- 仿射函数

  $f(x)=Ax+b,\ \ \nabla^2 f(x)=0_{n\times n}\succeq 0$ 既凸又凹
- 指数函数

  $f(x)=e^{ax}, x\in\mathbb{R}, f'(x)=ae^{ax}, f''(x)=a^2e^{ax}\geq 0$
- 幂函数

  $f(x)=x^a, x\in\mathbb{R}_{++}, f'(x)=ax^{a-1}, f''(x)=$
  $$a(a-1)x^{a-2}\begin{cases}\geq 0, & a\geq 1\text{或}a\leq 0 \quad \text{凸}\\ \leq 0, & 0\leq a\leq 1 \quad \text{凹}\end{cases}$$

常见的凸函数

- 绝对值的幂函数

$$f(x) = \|x\|^p, x \in \mathbb{R}, f'(x) = \begin{cases} px^{p-1}, & x \geq 0 \\ -p(-x)^{p-1}, & x < 0 \end{cases},$$

$$f''(x) = \begin{cases} p(p-1)x^{p-2}, & x \geq 0 \\ p(p-1)(-x)^{p-2}, & x < 0 \end{cases}$$

常见的凸函数

- 对数函数
  $f(x) = \log x, x \in \mathbb{R}_{++}, f'(x) = \frac{1}{x}, f''(x) = -\frac{1}{x^2} < 0$ 严格凹函数

常见的凸函数

- 对数函数
  $f(x) = \log x, x \in \mathbb{R}_{++}, f'(x) = \frac{1}{x}, f''(x) = -\frac{1}{x^2} < 0$ 严格凹函数
- 负熵
  $f(x) = x \log x, x \in \mathbb{R}_{++}, f'(x) = 1 + \log x, f''(x) = \frac{1}{x} > 0$ 严格凸函数

常见的凸函数

- 对数函数
  $f(x) = \log x, x \in \mathbb{R}_{++}, f'(x) = \frac{1}{x}, f''(x) = -\frac{1}{x^2} < 0$ 严格凹函数
- 负熵
  $f(x) = x\log x, x \in \mathbb{R}_{++}, f'(x) = 1 + \log x, f''(x) = \frac{1}{x} > 0$ 严格凸函数
- 范数
  $\mathbb{R}^n$ 空间的范数 $P(x), x \in \mathbb{R}^n$ 满足
  1. $P(ax) = \|a\| P(x)$
  2. $P(x + y) \le P(x) + P(y)$
  3. $P(x) = 0 \Leftrightarrow x = 0$

常见的凸函数

- 对数函数
  $f(x) = \log x, x \in \mathbb{R}_{++}, f'(x) = \frac{1}{x}, f''(x) = -\frac{1}{x^2} < 0$ 严格凹函数
- 负熵
  $f(x) = x\log x, x \in \mathbb{R}_{++}, f'(x) = 1 + \log x, f''(x) = \frac{1}{x} > 0$ 严格凸函数
- 范数
  $\mathbb{R}^n$ 空间的范数 $P(x), x \in \mathbb{R}^n$ 满足
  1. $P(ax) = \|a\|P(x)$
  2. $P(x+y) \leq P(x) + P(y)$
  3. $P(x) = 0 \Leftrightarrow x = 0$

  根据凸函数的定义 (一)，$\forall x, y \in \mathbb{R}^n, \forall 0 \leq \theta \leq 1$,
  $P(\theta x + (1-\theta)y) \leq P(\theta x) + P((1-\theta)y) = \theta P(x) + (1-\theta)P(y)$，是凸函数。

常见的凸函数

- 对数函数
  $f(x) = \log x, x \in \mathbb{R}_{++}, f'(x) = \frac{1}{x}, f''(x) = -\frac{1}{x^2} < 0$ 严格凹函数
- 负熵
  $f(x) = x \log x, x \in \mathbb{R}_{++}, f'(x) = 1 + \log x, f''(x) = \frac{1}{x} > 0$ 严格凸函数
- 范数
  $\mathbb{R}^n$ 空间的范数 $P(x), x \in \mathbb{R}^n$ 满足
  1. $P(ax) = \|a\| P(x)$
  2. $P(x + y) \leq P(x) + P(y)$
  3. $P(x) = 0 \Leftrightarrow x = 0$

  根据凸函数的定义 (一)，$\forall x, y \in \mathbb{R}^n, \forall 0 \leq \theta \leq 1$,
  $P(\theta x + (1-\theta)y) \leq P(\theta x) + P((1-\theta)y) = \theta P(x) + (1-\theta)P(y)$，是凸函数。
- 零范数
  $\|x\|_0 = x$ 中非零元素的个数，不是范数，也不是凸函数

常见的凸函数

- 极大值函数
  $f(x) = \max\{x_1, \ldots, x_n\}, x \in \mathbb{R}^n,$

$$
\begin{aligned}
f(\theta x + (1-\theta)y) &= \max\{\theta x_i + (1-\theta)y_i, i = 1, \ldots, n\} \\
&\le \theta \max\{x_i, i = 1, \ldots, n\} + (1-\theta)\max\{y_i, i = 1, \ldots, n\} \\
&= \theta f(x) + (1-\theta) f(y)
\end{aligned}
$$

常见的凸函数

- 极大值函数
  $f(x) = \max\{x_1, \ldots, x_n\}, x \in \mathbb{R}^n,$

$$\begin{aligned} f(\theta x + (1-\theta)y) &= \max\{\theta x_i + (1-\theta)y_i, i = 1, \ldots, n\} \\ &\le \theta \max\{x_i, i = 1, \ldots, n\} + (1-\theta)\max\{y_i, i = 1, \ldots, n\} \\ &= \theta f(x) + (1-\theta)f(y) \end{aligned}$$

- log-sum-exp(极大值函数的解析逼近)
  $f(x) = \log(e^{x_1} + \cdots + e^{x_n}), x \in \mathbb{R}^n,$
  $\max\{x_1, \ldots, x_n\} \le f(x) \le \max\{x_1 + \cdots + x_n\} + \log n$

$$\frac{\partial f}{\partial x_i} = \frac{e^{x_i}}{e^{x_1} + \cdots + e^{x_n}}, \qquad H = [H_{ij}]$$

常见的凸函数

当 $i \neq j$ 时，

$$H_{ij} = \frac{\partial^2 f}{\partial x_i \partial y_i} = \frac{-e^{x_i} e^{x_j}}{(e^{x_1} + \cdots + e^{x_n})^2} = \frac{-e^{x_i} e^{x_j}}{(\mathbf{1} \| z \|)^2}$$

当 $i = j$ 时，

$$H_{ii} = \frac{\partial^2 f}{\partial x_i^2} = \frac{-e^{x_i} e^{x_i} + e^{x_i}(e^{x_1} + \cdots + e^{x_n})}{(e^{x_1} + \cdots + e^{x_n})^2} = \frac{-e^{x_i} e^{x_i} + e^{x_i} \mathbf{1}^\top z}{(\mathbf{1} \| z \|)^2}$$

其中 $z = [e^{x_1}, \ldots, e^{x_n}]^\top$。

$$H = \frac{1}{(\mathbf{1}^\top z)^2} ((\mathbf{1}^\top z) \mathrm{diag}\{z\} - z z^\top)$$

令 $K = (\mathbf{1}^\top z)\mathrm{diag}\{z\} - z z^\top$, $\forall v \in \mathbb{R}^n$,

$$v^\top K v = (\mathbf{1}^\top z) v^\top \mathrm{diag}\{z\} v - v^\top z z^\top v = (\sum_i z_i)(\sum_i v_i^2 z_i) - (\sum_i v_i z_i)^2$$

根据 Cauchy-Schwartz 不等式， $v^\top K v \geq 0 \Rightarrow$log-sum-exp 函数为凸函数

常见的凸函数

- 几何平均
  $f(x) = (x_1 \cdots x_n)^{\frac{1}{n}}, x \in \mathbb{R}_{++}$ 为凹函数。

常见的凸函数

- 几何平均
  $f(x) = (x_1 \cdots x_n)^{\frac{1}{n}}, x \in \mathbb{R}_{++}$ 为凹函数。
- 对称半正定矩阵的行列式的对数
  $f(x) = \log\det(x), \mathrm{dom} f = S^n_{++}$,
  当 $n = 1$ 时，是凹函数；

常见的凸函数

- 几何平均
  $f(x) = (x_1 \cdots x_n)^{\frac{1}{n}}, x \in \mathbb{R}_{++}$ 为凹函数。
- 对称半正定矩阵的行列式的对数
  $f(x) = \log\det(x), \mathrm{dom} f = S^n_{++}$,
  当 $n = 1$ 时，是凹函数；
  当 $n > 1$ 时，$\forall z \in S^n_{++}, \forall t \in \mathbb{R}, \forall v \in \mathbb{R}^{n\times n}$, $z + tv \in S^n_{++} = \mathrm{dom} f$,
  因此 $v \in S^n$。

常见的凸函数

- 几何平均
  $f(x)=(x_1\cdots x_n)^{\frac{1}{n}}, x\in\mathbb{R}_{++}$ 为凹函数。
- 对称半正定矩阵的行列式的对数
  $f(x)=\log\det(x), \mathrm{dom}f=S^n_{++}$,
  当 $n=1$ 时，是凹函数；
  当 $n>1$ 时，$\forall z\in S^n_{++}, \forall t\in\mathbb{R}, \forall v\in\mathbb{R}^{n\times n}$，$z+tv\in S^n_{++}=\mathrm{dom}f$,
  因此 $v\in S^n$。

$$
\begin{aligned}
g(t)=f(z+tv)&=\log\det(z+tv)\\
&=\log\det(z^{\frac{1}{2}}(I+tz^{-\frac{1}{2}}vz^{-\frac{1}{2}})z^{\frac{1}{2}})\\
&=\log\det(z)+\log\det(I+tz^{-\frac{1}{2}}vz^{-\frac{1}{2}})\\
&=\log\det(z)+\sum_{i=1}^{n}\log(1+t\lambda_i)
\end{aligned}
$$

常见的凸函数

令 $z^{-\frac{1}{2}}vz^{-\frac{1}{2}} = Q\Lambda Q^\top$，其中 $QQ^\top = I$

$$\begin{aligned}\det(I + tz^{-\frac{1}{2}}vz^{-\frac{1}{2}}) &= \det(QQ^\top + tQ\Lambda Q^\top) \\ &= \det(Q)\det(I + t\Lambda)\det(Q^\top) \\ &= \det(QQ^\top)\det(I + t\Lambda)\end{aligned}$$

$$g'(t) = \sum_i \frac{\lambda_i}{1 + t\lambda_i}$$

$$g''(t) = \sum_i \frac{-\lambda_i^2}{(1 + t\lambda_i)^2} \leq 0$$

因此，$g(t)$ 为凹函数。

保凸运算

非负加权和：

$f_1, \ldots, f_n$ 为凸，则 $f = \sum_{i=1}^{n} w_i f_i$ 为凸，其中 $w_i \geq 0, \forall i$。
定义域 f_i 为凸，$\mathrm{dom} f = \bigcap_{i=1}^{n} \mathrm{dom} f_i$ 为凸。

保凸运算

非负加权和:

$f_1, \ldots, f_n$ 为凸，则 $f = \sum_{i=1}^{n} w_i f_i$ 为凸，其中 $w_i \geq 0, \forall i$。

定义域 f_i 为凸，$\mathrm{dom} f = \bigcap_{i=1}^{n} \mathrm{dom} f_i$ 为凸。

写成积分形式:

若 $f(x, y)$ 对任意 $y \in \mathcal{A}$ 均为凸，则称其对 (x, y) 为 jointly convex。

保凸运算

非负加权和:

$f_1, \ldots, f_n$ 为凸，则 $f = \sum_{i=1}^n w_i f_i$ 为凸，其中 $w_i \geq 0, \forall i$。

定义域 f_i 为凸，$\mathrm{dom} f = \bigcap_{i=1}^n \mathrm{dom} f_i$ 为凸。

写成积分形式:

若 $f(x, y)$ 对任意 $y \in \mathcal{A}$ 均为凸，则称其对 (x, y) 为 jointly convex。

设 $w(y) \geq 0$, $\forall y \in \mathcal{A}$, $g(x) = \int_{y \in \mathcal{A}} w(y) f(x, y) dy$ 为凸。

保凸运算

非负加权和:
$f_1, \ldots, f_n$ 为凸，则 $f = \sum_{i=1}^n w_i f_i$ 为凸，其中 $w_i \geq 0, \forall i$。
定义域 f_i 为凸，$\mathrm{dom} f = \bigcap_{i=1}^n \mathrm{dom} f_i$ 为凸。
写成积分形式:
若 $f(x, y)$ 对任意 $y \in \mathcal{A}$ 均为凸，则称其对 (x, y) 为 jointly convex。
设 $w(y) \geq 0,\ \forall y \in \mathcal{A},\ g(x) = \int_{y \in \mathcal{A}} w(y) f(x, y)\, dy$ 为凸。

仿射映射:
$f : \mathbb{R}^n \to \mathbb{R},\ A \in \mathbb{R}^{n \times m},\ b \in \mathbb{R}^n,\ g(x) = f(Ax + b)$,
$\mathrm{dom}\, g = \{x \mid Ax + b \in \mathrm{dom} f\}$。若 f 为凸，那么 g 也是凸函数；若 f 为凹，那么 g 也是凹函数。

保凸运算

非负加权和:
$f_1, \ldots, f_n$ 为凸，则 $f = \sum_{i=1}^{n} w_i f_i$ 为凸，其中 $w_i \geq 0, \forall i$。
定义域 f_i 为凸，$\mathrm{dom} f = \bigcap_{i=1}^{n} \mathrm{dom} f_i$ 为凸。
写成积分形式:
若 $f(x, y)$ 对任意 $y \in \mathcal{A}$ 均为凸，则称其对 (x, y) 为 jointly convex。
设 $w(y) \geq 0, \forall y \in \mathcal{A}$, $g(x) = \int_{y \in \mathcal{A}} w(y) f(x, y) dy$ 为凸。

仿射映射:
$f : \mathbb{R}^n \to \mathbb{R}$, $A \in \mathbb{R}^{n \times m}$, $b \in \mathbb{R}^n$, $g(x) = f(Ax + b)$,
$\mathrm{dom}\, g = \{x \mid Ax + b \in \mathrm{dom} f\}$。若 f 为凸，那么 g 也是凸函数；若 f 为凹，那么 g 也是凹函数。
$f_i : \mathbb{R}^n \to \mathbb{R}, i = 1, \ldots, m$ 为凸, $A \in \mathbb{R}^m$, $b \in \mathbb{R}$,
$g(x) = A^\top [f_1(x) \cdots f_m(x)]^\top + b$, 若 A 非负，则 g 为凸函数。

保凸运算

非负加权和:
$f_1, \dots, f_n$ 为凸，则 $f = \sum_{i=1}^n w_i f_i$ 为凸，其中 $w_i \geq 0, \forall i$。
定义域 f_i 为凸，$\mathrm{dom} f = \bigcap_{i=1}^n \mathrm{dom} f_i$ 为凸。
写成积分形式:
若 $f(x, y)$ 对任意 $y \in \mathcal{A}$ 均为凸，则称其对 (x, y) 为 jointly convex。
设 $w(y) \geq 0, \ \forall y \in \mathcal{A}, \ g(x) = \int_{y \in \mathcal{A}} w(y) f(x, y) dy$ 为凸。

仿射映射:
$f : \mathbb{R}^n \to \mathbb{R}, \ A \in \mathbb{R}^{n \times m}, \ b \in \mathbb{R}^n, \ g(x) = f(Ax + b)$,
$\mathrm{dom} g = \{x \mid Ax + b \in \mathrm{dom} f\}$。若 f 为凸，那么 g 也是凸函数；若 f 为凹，那么 g 也是凹函数。
$f_i : \mathbb{R}^n \to \mathbb{R}, i = 1, \dots, m$ 为凸, $A \in \mathbb{R}^m, \ b \in \mathbb{R}$,
$g(x) = A^\top [f_1(x) \cdots f_m(x)]^\top + b$, 若 A 非负，则 g 为凸函数。
两个函数的极大值函数：f_1, f_2 是凸函数，定义
$f(x) = \max\{f_1(x), f_2(x)\}, \ \mathrm{dom} f = \mathrm{dom} f_1 \cap \mathrm{dom} f_2$

保凸运算

函数的组合： $h:\mathbb{R}^k \to \mathbb{R},\ g:\mathbb{R}^n \to \mathbb{R}^k,\ f = h \circ g:\mathbb{R}^n \to \mathbb{R}$

$f(x) = h(g(x))$ $\mathrm{dom} f = \{x \in \mathrm{dom}\, g \mid g(x) \in \mathrm{dom}\, h\}$

保凸运算

函数的组合： $h:\mathbb{R}^k\to\mathbb{R}$, $g:\mathbb{R}^n\to\mathbb{R}^k$, $f=h\circ g:\mathbb{R}^n\to\mathbb{R}$
$f(x)=h(g(x))$ $\mathrm{dom}f=\{x\in\mathrm{dom}g\mid g(x)\in\mathrm{dom}h\}$

一维： $k=n=1$, $\mathrm{dom}g=\mathrm{dom}h=\mathrm{dom}f=\mathbb{R}$, h 和 g 均为二阶可微

保凸运算

函数的组合： $h:\mathbb{R}^k \to \mathbb{R},\ g:\mathbb{R}^n \to \mathbb{R}^k,\ f = h\circ g:\mathbb{R}^n \to \mathbb{R}$
$f(x) = h(g(x))\ \mathrm{dom} f = \{x \in \mathrm{dom} g \mid g(x) \in \mathrm{dom} h\}$

一维： $k = n = 1,\ \mathrm{dom} g = \mathrm{dom} h = \mathrm{dom} f = \mathbb{R},\ h$ 和 g 均为二阶可微
f 为凸 $\Leftrightarrow f''(x) \geq 0$

保凸运算

函数的组合： $h:\mathbb{R}^k \to \mathbb{R}$, $g:\mathbb{R}^n \to \mathbb{R}^k$, $f = h\circ g:\mathbb{R}^n \to \mathbb{R}$

$f(x) = h(g(x))$ $\mathrm{dom} f = \{x \in \mathrm{dom}\, g \mid g(x) \in \mathrm{dom}\, h\}$

一维： $k = n = 1$, $\mathrm{dom}\, g = \mathrm{dom}\, h = \mathrm{dom} f = \mathbb{R}$, h 和 g 均为二阶可微

f 为凸 $\Leftrightarrow f''(x) \geq 0$

$f'(x) = h'(g(x))g'(x)$, $f''(x) = h''(g(x))(g'(x))^2 + h'(g(x))g''(x)$

保凸运算

函数的组合： $h:\mathbb{R}^k \to \mathbb{R}$, $g:\mathbb{R}^n \to \mathbb{R}^k$, $f = h \circ g:\mathbb{R}^n \to \mathbb{R}$

$f(x) = h(g(x))$ $\mathrm{dom} f = \{x \in \mathrm{dom}\, g \mid g(x) \in \mathrm{dom}\, h\}$

一维： $k = n = 1$, $\mathrm{dom}\, g = \mathrm{dom}\, h = \mathrm{dom} f = \mathbb{R}$, h 和 g 均为二阶可微

f 为凸 $\Leftrightarrow f''(x) \geq 0$

$f'(x) = h'(g(x))g'(x)$, $f''(x) = h''(g(x))(g'(x))^2 + h'(g(x))g''(x)$

$h''(g(x))$ 反映 h 的凸性，$h'(g(x))$ 反映 h 的增减性，$g''(x)$ 反映 g 的凸性

保凸运算

函数的组合： $h: \mathbb{R}^k \to \mathbb{R}$, $g: \mathbb{R}^n \to \mathbb{R}^k$, $f = h \circ g: \mathbb{R}^n \to \mathbb{R}$
$f(x) = h(g(x))$ $\mathrm{dom} f = \{x \in \mathrm{dom}\, g \mid g(x) \in \mathrm{dom}\, h\}$

一维： $k = n = 1$, $\mathrm{dom}\, g = \mathrm{dom}\, h = \mathrm{dom} f = \mathbb{R}$, h 和 g 均为二阶可微
f 为凸 $\Leftrightarrow f''(x) \geq 0$
$f'(x) = h'(g(x))g'(x)$, $f''(x) = h''(g(x))(g'(x))^2 + h'(g(x))g''(x)$
$h''(g(x))$ 反映 h 的凸性，$h'(g(x))$ 反映 h 的增减性，$g''(x)$ 反映 g 的凸性

1. h 为凸，不降，g 为凸，则 f 为凸
2. h 为凸，不增，g 为凹，则 f 为凸
3. h 为凹，不降，g 为凹，则 f 为凹
4. h 为凹，不增，g 为凸，则 f 为凹

保凸运算

高维：$n, k \geq 1$，$\mathrm{dom}\, g, \mathrm{dom}\, h, \mathrm{dom}\, f \neq \mathbb{R}^n, \mathbb{R}^k, \mathbb{R}^n$，$h$ 和 g 均不二阶可微

保凸运算

高维： $n, k \geq 1$，$\mathrm{dom} g, \mathrm{dom} h, \mathrm{dom} f \neq \mathbb{R}^n, \mathbb{R}^k, \mathbb{R}^n$，$h$ 和 g 均不二阶可微

类似于一维的情况，我们可以给出以下四条结论：

1. h 为凸，$\tilde{h}$ 不降，g 为凸，则 f 为凸
2. h 为凸，$\tilde{h}$ 不增，g 为凹，则 f 为凸
3. h 为凹，$\tilde{h}$ 不降，g 为凹，则 f 为凹
4. h 为凹，$\tilde{h}$ 不增，g 为凸，则 f 为凹

保凸运算

透视函数：$P:\mathbb{R}^{n+1}\to\mathbb{R}^n$, $\mathrm{dom}\,P=\mathbb{R}^n\times\mathbb{R}_{++}$,

$$P(z,t)=\frac{z}{t}, z\in\mathbb{R}^n, t\in\mathbb{R}_{++}$$

保凸运算

透视函数： $P : \mathbb{R}^{n+1} \to \mathbb{R}^n$, $\mathrm{dom}P = \mathbb{R}^n \times \mathbb{R}_{++}$,

$$P(z,t) = \frac{z}{t}, z \in \mathbb{R}^n, t \in \mathbb{R}_{++}$$

函数的透视： $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \times \mathbb{R}_{++} \to \mathbb{R}$

$$g(x,t) = tf\left(\frac{x}{t}\right)$$

函数 g 为 f 的透视，$\mathrm{dom}g = \{(x,t) \mid t > 0, \frac{x}{t} \in \mathrm{dom}f\}$。

保凸运算

透视函数： $P : \mathbb{R}^{n+1} \to \mathbb{R}^n$, $\mathrm{dom}P = \mathbb{R}^n \times \mathbb{R}_{++}$,

$$P(z, t) = \frac{z}{t}, z \in \mathbb{R}^n, t \in \mathbb{R}_{++}$$

函数的透视： $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \times \mathbb{R}_{++} \to \mathbb{R}$

$$g(x, t) = tf\left(\frac{x}{t}\right)$$

函数 g 为 f 的透视， $\mathrm{dom}g = \{(x, t) \mid t > 0, \frac{x}{t} \in \mathrm{dom}f\}$。
f 为凸时， g 为凸； f 为凹时， g 为凹。

保凸运算

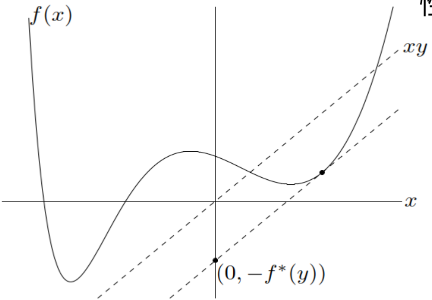
函数的共轭 (Conjugate)： $f : \mathbb{R}^n \to \mathbb{R}$, $f^* : \mathbb{R}^n \to \mathbb{R}$

$$f^*(y) = \sup_{x \in \mathrm{dom} f} \left(y^\top x - f(x) \right)$$

保凸运算

函数的共轭 (Conjugate)： $f:\mathbb{R}^n\rightarrow\mathbb{R}$, $f^*:\mathbb{R}^n\rightarrow\mathbb{R}$

$$f^*(y)=\sup_{x\in\text{dom}f}\left(y^\top x-f(x)\right)$$



性质：

- $f(x)$ 若可微，则 $f^*(y)$ 对应的 x 必是 $f'(x)=y$ 的一点，$(y^\top x-f(x))'=y-f'(x)=0$。
- 函数的共轭一定是凸的。

拟凸函数

α-**sublevel set**：若 $f: \mathbb{R}^n \to \mathbb{R}$，定义其 α-sublevel set 为 $C_\alpha = \{x \in \mathrm{dom} f \mid f(x) \leq \alpha\}$。

拟凸函数

α-sublevel set：若 $f : \mathbb{R}^n \to \mathbb{R}$，定义其 α-sublevel set 为 $C_\alpha = \{x \in \mathrm{dom} f \mid f(x) \le \alpha\}$。

凸函数的所有 α-sublevel set 都是凸集。
若函数的所有 α-sublevel set 都是凸集，则该函数不一定是凸函数。

拟凸函数

α-sublevel set：若 $f: \mathbb{R}^n \to \mathbb{R}$，定义其 α-sublevel set 为 $C_\alpha = \{x \in \mathrm{dom} f \mid f(x) \le \alpha\}$。

凸函数的所有 α-sublevel set 都是凸集。
若函数的所有 α-sublevel set 都是凸集，则该函数不一定是凸函数。

拟凸函数 (Quasiconvex functions)：$f: \mathbb{R}^n \to \mathbb{R}$

- Quasiconvex：$S_\alpha = \{x \in \mathrm{dom} f \mid f(x) \le \alpha\}$
- Quasiconcave：$S'_\alpha = \{x \in \mathrm{dom} f \mid f(x) \ge \alpha\}$
- Quasilinear：$S''_\alpha = \{x \in \mathrm{dom} f \mid f(x) = \alpha\}$

拟凸函数

可微拟凸函数的一阶条件：

- 凸：dom$f$ 为凸，$\forall x, y \in \mathrm{dom} f,\ f(y) \geq f(x) + \nabla f^{\top}(x)(y - x)$
- 拟凸：dom$f$ 为凸，$\forall x, y \in \mathrm{dom} f,\ f(y) \leq f(x) \Rightarrow \nabla f^{\top}(x)(y - x) \leq 0$

拟凸函数

可微拟凸函数的一阶条件：

- 凸：dom$f$ 为凸，$\forall x, y \in \text{dom} f,\ f(y) \geq f(x) + \nabla f^{\top}(x)(y-x)$
- 拟凸：dom$f$ 为凸，$\forall x, y \in \text{dom} f,\ f(y) \leq f(x) \Rightarrow \nabla f^{\top}(x)(y-x) \leq 0$

- 凸：若 $\nabla f(x) = 0$，则 $\forall y \in \text{dom} f,\ f(y) \geq f(x)$
- 拟凸：若 $\nabla f(x) = 0$，则 $\forall y \in \text{dom} f,\ f(y) \leq f(x) \Rightarrow 0 \leq 0$ 没意义

拟凸函数

可微拟凸函数的一阶条件:

- 凸: dom$f$ 为凸, $\forall x, y \in \mathrm{dom} f$, $f(y) \geq f(x) + \nabla f^{\top}(x)(y-x)$
- 拟凸: dom$f$ 为凸, $\forall x, y \in \mathrm{dom} f$, $f(y) \leq f(x) \Rightarrow \nabla f^{\top}(x)(y-x) \leq 0$

- 凸: 若 $\nabla f(x) = 0$, 则 $\forall y \in \mathrm{dom} f$, $f(y) \geq f(x)$
- 拟凸: 若 $\nabla f(x) = 0$, 则 $\forall y \in \mathrm{dom} f$, $f(y) \leq f(x) \Rightarrow 0 \leq 0$ 没意义

可微拟凸函数的二阶条件:

- 凸: dom$f$ 为凸, $\forall x \in \mathrm{dom} f$, $\nabla^2 f(x) \succeq 0$
- 拟凸: dom$f$ 为凸, $\forall x \in \mathrm{dom} f, \forall y \in \mathbb{R}^n, y \neq 0$,
  $y^{\top} \nabla f(x) = 0 \Rightarrow y^{\top} \nabla^2 f(x) y \geq 0$

习题

3.1, 3.2, 3.5, 3.13, 3.18, 3.21, 3.32, 3.33, 3.36, 3.43

章节目录

优化问题

一般优化问题的描述：

$$
\begin{aligned}
\text{minimize (min)} \quad & f_0(x) \\
\text{subject to (s.t.)} \quad & f_i(x) \le 0,\, i = 1, \ldots, m \\
& h_i(x) = 0,\, i = 1, \ldots, p
\end{aligned}
$$

优化问题

一般优化问题的描述：

$$
\begin{aligned}
\text{minimize (min)} \quad & f_0(x) \\
\text{subject to (s.t.)} \quad & f_i(x) \leq 0, i = 1, \ldots, m \\
& h_i(x) = 0, i = 1, \ldots, p
\end{aligned}
$$

其中

$x = [x_1, \ldots, x_n]^\top \in \mathbb{R}^n$ 优化变量

优化问题

一般优化问题的描述：

$$
\begin{aligned}
\text{minimize (min)} \quad & f_0(x) \\
\text{subject to (s.t.)} \quad & f_i(x) \le 0, i = 1, \ldots, m \\
& h_i(x) = 0, i = 1, \ldots, p
\end{aligned}
$$

其中

$x = [x_1, \ldots, x_n]^\top \in \mathbb{R}^n$　优化变量

$f_0 : \mathbb{R}^n \to \mathbb{R}$　目标函数/损失函数

优化问题

一般优化问题的描述：

$$
\begin{array}{rl}
\text{minimize (min)} & f_0(x) \\
\text{subject to (s.t.)} & f_i(x) \le 0, i = 1, \dots, m \\
 & h_i(x) = 0, i = 1, \dots, p
\end{array}
$$

其中

$x = [x_1, \dots, x_n]^\top \in \mathbb{R}^n$	优化变量
$f_0 : \mathbb{R}^n \to \mathbb{R}$	目标函数/损失函数
$f_i(x) \le 0, i = 1, \dots, m$	不等式约束

优化问题

一般优化问题的描述：

$$
\begin{array}{rl}
\text{minimize (min)} & f_0(x) \\
\text{subject to (s.t.)} & f_i(x) \leq 0, i = 1, \ldots, m \\
 & h_i(x) = 0, i = 1, \ldots, p
\end{array}
$$

其中

$x = [x_1, \ldots, x_n]^\top \in \mathbb{R}^n$	优化变量
$f_0 : \mathbb{R}^n \to \mathbb{R}$	目标函数/损失函数
$f_i(x) \leq 0, i = 1, \ldots, m$	不等式约束
$f_i : \mathbb{R}^n \to \mathbb{R}$	不等式约束函数

优化问题

一般优化问题的描述：

$$\begin{array}{rl} \text{minimize (min)} & f_0(x) \\ \text{subject to (s.t.)} & f_i(x) \leq 0, i = 1, \ldots, m \\ & h_i(x) = 0, i = 1, \ldots, p \end{array}$$

其中

$x = [x_1, \ldots, x_n]^\top \in \mathbb{R}^n$	优化变量
$f_0 : \mathbb{R}^n \to \mathbb{R}$	目标函数/损失函数
$f_i(x) \leq 0, i = 1, \ldots, m$	不等式约束
$f_i : \mathbb{R}^n \to \mathbb{R}$	不等式约束函数
$h_i(x) = 0, i = 1, \ldots, p$	等式约束

优化问题

一般优化问题的描述:

$$
\begin{array}{rl}
\text{minimize (min)} & f_0(x) \\
\text{subject to (s.t.)} & f_i(x) \le 0, i = 1, \dots, m \\
 & h_i(x) = 0, i = 1, \dots, p
\end{array}
$$

其中

$x = [x_1, \dots, x_n]^\top \in \mathbb{R}^n$	优化变量
$f_0 : \mathbb{R}^n \to \mathbb{R}$	目标函数/损失函数
$f_i(x) \le 0, i = 1, \dots, m$	不等式约束
$f_i : \mathbb{R}^n \to \mathbb{R}$	不等式约束函数
$h_i(x) = 0, i = 1, \dots, p$	等式约束
$h_i : \mathbb{R}^n \to \mathbb{R}$	等式约束函数

优化问题

一般优化问题的描述：

$$
\begin{array}{rl}
\text{minimize (min)} & f_0(x) \\
\text{subject to (s.t.)} & f_i(x) \leq 0,\, i = 1, \ldots, m \\
 & h_i(x) = 0,\, i = 1, \ldots, p
\end{array}
$$

其中

$x = [x_1, \ldots, x_n]^\top \in \mathbb{R}^n$	优化变量
$f_0 : \mathbb{R}^n \to \mathbb{R}$	目标函数/损失函数
$f_i(x) \leq 0,\, i = 1, \ldots, m$	不等式约束
$f_i : \mathbb{R}^n \to \mathbb{R}$	不等式约束函数
$h_i(x) = 0,\, i = 1, \ldots, p$	等式约束
$h_i : \mathbb{R}^n \to \mathbb{R}$	等式约束函数
$m = p = 0$	无约束

优化问题

优化问题的域 (domain): $\mathcal{D} = \bigcap_{i=0}^{m} \mathrm{dom} f_i \cap \bigcap_{i=1}^{p} \mathrm{dom} h_i$

优化问题

优化问题的域 (domain): $\mathcal{D} = \bigcap_{i=0}^{m} \mathrm{dom} f_i \cap \bigcap_{i=1}^{p} \mathrm{dom} h_i$

可行解 (feasible solution): $x \in \mathcal{D}$ 为可行解，若 $f_i(x) \leq 0, i = 1, \ldots, m$; $h_i(x) = 0, i = 1, \ldots, p$

优化问题

优化问题的域 (domain): $\mathcal{D} = \bigcap_{i=0}^{m} \mathrm{dom} f_i \cap \bigcap_{i=1}^{p} \mathrm{dom} h_i$

可行解 (feasible solution): $x \in \mathcal{D}$ 为可行解，若 $f_i(x) \leq 0, i = 1, \ldots, m$; $h_i(x) = 0, i = 1, \ldots, p$

可行解集 (feasible set): $X_f = \{x \text{为可行解}\}$

优化问题

优化问题的域 (domain): $\mathcal{D} = \bigcap_{i=0}^{m} \mathrm{dom} f_i \cap \bigcap_{i=1}^{p} \mathrm{dom} h_i$
可行解 (feasible solution): $x \in \mathcal{D}$ 为可行解，若 $f_i(x) \leq 0, i = 1, \ldots, m$; $h_i(x) = 0, i = 1, \ldots, p$
可行解集 (feasible set): $X_f = \{x\text{为可行解}\}$
问题的最优值 (optimal value): $p^\star = \inf\{f_0(x) \mid x \in X_f\}$

优化问题

优化问题的域 (domain): $\mathcal{D} = \bigcap_{i=0}^{m} \mathrm{dom} f_i \cap \bigcap_{i=1}^{p} \mathrm{dom} h_i$

可行解 (feasible solution): $x \in \mathcal{D}$ 为可行解，若 $f_i(x) \le 0, i = 1, \dots, m$; $h_i(x) = 0, i = 1, \dots, p$

可行解集 (feasible set): $X_f = \{x\text{为可行解}\}$

问题的最优值 (optimal value): $p^\star = \inf\{f_0(x) \mid x \in X_f\}$

若 X_f 为空集，则 $p^\star = +\infty$

优化问题

优化问题的域 (domain): $\mathcal{D} = \bigcap_{i=0}^{m} \mathrm{dom} f_i \cap \bigcap_{i=1}^{p} \mathrm{dom} h_i$

可行解 (feasible solution): $x \in \mathcal{D}$ 为可行解，若 $f_i(x) \leq 0, i = 1, \ldots, m$;
$h_i(x) = 0, i = 1, \ldots, p$

可行解集 (feasible set): $X_f = \{x\text{为可行解}\}$

问题的最优值 (optimal value): $p^\star = \inf\{f_0(x) \mid x \in X_f\}$

若 X_f 为空集，则 $p^\star = +\infty$

最优解 (optimal point/solution)：若 $x^\star$ 可行，且 $f(x^\star) = p^\star$

优化问题

优化问题的域 (domain): $\mathcal{D} = \bigcap_{i=0}^{m} \text{dom} f_i \cap \bigcap_{i=1}^{p} \text{dom} h_i$

可行解 (feasible solution): $x \in \mathcal{D}$ 为可行解，若 $f_i(x) \le 0, i = 1, \ldots, m$;
$h_i(x) = 0, i = 1, \ldots, p$

可行解集 (feasible set): $X_f = \{x\text{为可行解}\}$

问题的最优值 (optimal value): $p^\star = \inf\{f_0(x) \mid x \in X_f\}$

若 X_f 为空集，则 $p^\star = +\infty$

最优解 (optimal point/solution)：若 $x^\star$ 可行，且 $f(x^\star) = p^\star$

最优解集 (optimal set)：$X_{\text{opt}} = \{x \mid x \in X_f, f_0(x) = p^\star\}$

优化问题

优化问题的域 (domain): $\mathcal{D} = \bigcap_{i=0}^{m} \mathrm{dom} f_i \cap \bigcap_{i=1}^{p} \mathrm{dom} h_i$

可行解 (feasible solution): $x \in \mathcal{D}$ 为可行解，若 $f_i(x) \leq 0, i = 1, \ldots, m$；
$h_i(x) = 0, i = 1, \ldots, p$

可行解集 (feasible set): $X_f = \{x\text{为可行解}\}$

问题的最优值 (optimal value): $p^\star = \inf\{f_0(x) \mid x \in X_f\}$

若 X_f 为空集，则 $p^\star = +\infty$

最优解 (optimal point/solution)：若 $x^\star$ 可行，且 $f(x^\star) = p^\star$

最优解集 (optimal set)：$X_{\mathrm{opt}} = \{x \mid x \in X_f,\ f_0(x) = p^\star\}$

ϵ-次优解集 (ϵ-suboptimal set)：$X_\epsilon = \{x \mid x \in X_f,\ f_0(x) = p^\star + \epsilon\}$

优化问题

优化问题的域 (domain): $\mathcal{D}=\bigcap_{i=0}^{m}\mathrm{dom}f_i\cap\bigcap_{i=1}^{p}\mathrm{dom}h_i$

可行解 (feasible solution): $x\in\mathcal{D}$ 为可行解，若 $f_i(x)\le 0, i=1,\dots,m$; $h_i(x)=0, i=1,\dots,p$

可行解集 (feasible set): $X_f=\{x\text{为可行解}\}$

问题的最优值 (optimal value): $p^\star=\inf\{f_0(x)\mid x\in X_f\}$

若 X_f 为空集，则 $p^\star=+\infty$

最优解 (optimal point/solution)：若 $x^\star$ 可行，且 $f(x^\star)=p^\star$

最优解集 (optimal set)： $X_{\mathrm{opt}}=\{x\mid x\in X_f,\ f_0(x)=p^\star\}$

ϵ-次优解集 (ϵ-suboptimal set)： $X_\epsilon=\{x\mid x\in X_f,\ f_0(x)=p^\star+\epsilon\}$

局部最优解 (locally optimal)：$\exists R>0$ 使 x 局部最优化

$$
\begin{aligned}
f_0(x)=\inf\{f_0(z)\ |f_i(x)\le 0, i=1,\dots,m,\\
h_i(x)=0, i=1,\dots,p,\\
\|z-x\|_2\le R\}
\end{aligned}
$$

优化问题

考虑不等式约束

$$
\begin{aligned}
f_i(x) &\leq 0, \quad i = 1, \ldots, m \\
f_i(x) &< 0, \quad i = 1, \ldots, m
\end{aligned}
$$

优化问题

考虑不等式约束

$$
\begin{aligned}
f_i(x) &\le 0, \quad i=1,\dots,m \\
f_i(x) &< 0, \quad i=1,\dots,m
\end{aligned}
$$

结论：若 $x \in X_f$, $f_i(x) = 0$, 则 $f_i(x) \le 0$ 为活动约束 (active)；
$f_i(x) < 0$, 则 $f_i(x) \le 0$ 为不活动约束 (inactive)。

优化问题

考虑不等式约束

$$
\begin{aligned}
f_i(x) &\le 0, \quad i = 1, \ldots, m \\
f_i(x) &< 0, \quad i = 1, \ldots, m
\end{aligned}
$$

结论：若 $x \in X_f$, $f_i(x) = 0$, 则 $f_i(x) \le 0$ 为活动约束 (active)；
$f_i(x) < 0$, 则 $f_i(x) \le 0$ 为不活动约束 (inactive)。

$$
\begin{aligned}
\min \quad & \text{utility} \\
\text{s.t.} \quad & \text{money} < 100 \rightarrow \text{money} \le 100 \\
& \quad \textbf{or} \; -|\log(100 - \text{money})| \le 0
\end{aligned}
$$

优化问题

可行性优化问题 (feasibility problem)：

$$
\begin{aligned}
\text{find} \quad & x \to \min \quad 0 \\
\text{s.t.} \quad & f_i(x) \leq 0, i = 1, \dots, m \\
& h_i(x) = 0, i = 1, \dots, p
\end{aligned}
$$

优化问题

可行性优化问题 (feasibility problem)：

$$
\begin{aligned}
\text{find} \quad & x \to \min \quad 0 \\
\text{s.t.} \quad & f_i(x) \leq 0, i = 1, \ldots, m \\
& h_i(x) = 0, i = 1, \ldots, p
\end{aligned}
$$

等价问题：将优化问题写成一般形式。

凸优化问题

优化问题的一般形式为：

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, i = 1, \ldots, m \\ & h_i(x) = 0, i = 1, \ldots, p \end{aligned}$$

凸优化问题

优化问题的一般形式为：

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & f_i(x) \le 0, i = 1, \ldots, m \\ & h_i(x) = 0, i = 1, \ldots, p \end{array}$$

凸优化的一般形式 (狭义) 为：

$$\begin{array}{lll} \min & f_0(x) & \text{凸} \\ \text{s.t.} & f_i(x) \le 0, i = 1, \ldots, m & \text{凸} \\ & a_i^\top x = b_i, i = 1, \ldots, p & \text{仿射} \end{array}$$

凸优化问题

优化问题的一般形式为：

$$
\begin{array}{ll}
\min & f_0(x) \\
\text{s.t.} & f_i(x) \leq 0, i = 1, \ldots, m \\
 & h_i(x) = 0, i = 1, \ldots, p
\end{array}
$$

凸优化的一般形式 (狭义) 为：

$$
\begin{array}{lll}
\min & f_0(x) & \text{凸} \\
\text{s.t.} & f_i(x) \leq 0, i = 1, \ldots, m & \text{凸} \\
 & a_i^\top x = b_i, i = 1, \ldots, p & \text{仿射}
\end{array}
$$

Convex/Quasiconvex optimization

凸优化问题

优化问题的一般形式为：

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, i = 1, \ldots, m \\ & h_i(x) = 0, i = 1, \ldots, p \end{aligned}$$

凸优化的一般形式 (狭义) 为：

$$\begin{aligned} \min \quad & f_0(x) && \text{凸} \\ \text{s.t.} \quad & f_i(x) \leq 0, i = 1, \ldots, m && \text{凸} \\ & a_i^\top x = b_i, i = 1, \ldots, p && \text{仿射} \end{aligned}$$

Convex/Quasiconvex optimization
Nonconvex optimization/Concave optimization

凸优化问题

凸优化问题的重要性质

$$\text{局部最优} = \text{全局最优}$$

证明: 局部最优: $\exists R > 0,\ f_0(x) = \inf\{f_0(z) \mid x, z\text{可行}, \|x - z\|_2 \leq R\}$

凸优化问题

凸优化问题的重要性质

局部最优 $=$ 全局最优

证明: 局部最优：$\exists R > 0,\ f_0(x) = \inf\{f_0(z) \mid x, z\text{可行}, \|x - z\|_2 \leq R\}$
设 x 不是全局最优，即 $\exists y$ 可行，$f_0(y) < f_0(x)$

凸优化问题

凸优化问题的重要性质

局部最优 $=$ 全局最优

证明: 局部最优：$\exists R > 0,\ f_0(x) = \inf\{f_0(z) \mid x, z\text{可行}, \|x - z\|_2 \le R\}$
设 x 不是全局最优，即 $\exists y$ 可行，$f_0(y) < f_0(x)$
因为 x 为局部最优，那么 $\|y - x\|_2 > R$

凸优化问题

凸优化问题的重要性质

局部最优 = 全局最优

证明: 局部最优: $\exists R > 0,\ f_0(x) = \inf\{f_0(z) \mid x, z\text{可行}, \|x - z\|_2 \le R\}$
设 x 不是全局最优，即 $\exists y$ 可行，$f_0(y) < f_0(x)$
因为 x 为局部最优，那么 $\|y - x\|_2 > R$
$z = (1-\theta)x + \theta y$, 取 $\theta = \frac{R}{2\|y-x\|_2}$

凸优化问题

凸优化问题的重要性质

局部最优 = 全局最优

证明: 局部最优：$\exists R > 0,\ f_0(x) = \inf\{f_0(z) \mid x, z\text{可行}, \|x - z\|_2 \leq R\}$
设 x 不是全局最优，即 $\exists y$ 可行，$f_0(y) < f_0(x)$
因为 x 为局部最优，那么 $\|y - x\|_2 > R$
$z = (1-\theta)x + \theta y$, 取 $\theta = \frac{R}{2\|y-x\|_2}$
z 可行，且 $f_0(z) \leq (1-\theta)f_0(x) + \theta f_0(y)$

凸优化问题

凸优化问题的重要性质

局部最优 $=$ 全局最优

证明: 局部最优：$\exists R > 0,\ f_0(x) = \inf\{f_0(z) \mid x, z$可行$, \|x - z\|_2 \leq R\}$
设 x 不是全局最优，即 $\exists y$ 可行，$f_0(y) < f_0(x)$
因为 x 为局部最优，那么 $\|y - x\|_2 > R$
$z = (1 - \theta)x + \theta y$, 取 $\theta = \frac{R}{2\|y-x\|_2}$
z 可行，且 $f_0(z) \leq (1 - \theta)f_0(x) + \theta f_0(y)$
$\|z - x\|_2 = \theta\|x - y\|_2 = \frac{R}{2} < R$

凸优化问题

凸优化问题的重要性质

局部最优 = 全局最优

证明: 局部最优: $\exists R > 0,\ f_0(x) = \inf\{f_0(z) \mid x, z\text{可行}, \|x - z\|_2 \leq R\}$
设 x 不是全局最优, 即 $\exists y$ 可行, $f_0(y) < f_0(x)$
因为 x 为局部最优, 那么 $\|y - x\|_2 > R$
$z = (1-\theta)x + \theta y$, 取 $\theta = \frac{R}{2\|y-x\|_2}$
z 可行, 且 $f_0(z) \leq (1-\theta)f_0(x) + \theta f_0(y)$
$\|z - x\|_2 = \theta\|x - y\|_2 = \frac{R}{2} < R$
$f_0(x) \leq f_0(z)$

凸优化问题

凸优化问题的重要性质

局部最优 = 全局最优

证明: 局部最优: $\exists R > 0,\ f_0(x) = \inf\{f_0(z) \mid x, z\text{可行}, \|x - z\|_2 \le R\}$
设 x 不是全局最优，即 $\exists y$ 可行，$f_0(y) < f_0(x)$
因为 x 为局部最优，那么 $\|y - x\|_2 > R$
$z = (1-\theta)x + \theta y$, 取 $\theta = \frac{R}{2\|y-x\|_2}$
z 可行，且 $f_0(z) \le (1-\theta)f_0(x) + \theta f_0(y)$
$\|z - x\|_2 = \theta\|x - y\|_2 = \frac{R}{2} < R$
$f_0(x) \le f_0(z)$
所以 $f_0(y) < f_0(z)$，即 $f_0(z) < f_0(z)$，矛盾。
因此，对于凸优化问题，局部最优 = 全局最优。

凸优化问题

可微目标函数情况下的最优性：

若凸优化问题中的目标函数 f_0 可微，则对于 $\forall x, y \in \mathrm{dom} f_0$，存在

$$f_0(y) \geq f_0(x) + \nabla f_0^\top(x)(y - x)$$

凸优化问题

可微目标函数情况下的最优性：
若凸优化问题中的目标函数 f_0 可微，则对于 $\forall x, y \in \mathrm{dom} f_0$，存在

$$f_0(y) \geq f_0(x) + \nabla f_0^\top(x)(y - x)$$

凸优化问题的可行集为：

$$\begin{aligned} X_f = \{x \,|f_i(x) \leq 0, i = 1, \ldots, m, \\ h_i(x) = 0, i = 1, \ldots, p\} \cap \mathrm{dom} f_0 \end{aligned}$$

凸优化问题

可微目标函数情况下的最优性：
若凸优化问题中的目标函数 f_0 可微，则对于 $\forall x, y \in \mathrm{dom} f_0$，存在

$$f_0(y) \geq f_0(x) + \nabla f_0^\top(x)(y - x)$$

凸优化问题的可行集为：

$$X_f = \{x \,|f_i(x) \leq 0, i = 1, \ldots, m, \\ h_i(x) = 0, i = 1, \ldots, p\} \cap \mathrm{dom} f_0$$

$x^\star \in X_f$ 最优 $\Leftrightarrow \nabla f_0^\top(x^\star)(y - x^\star) \geq 0 \quad \forall y \in X_f$

凸优化问题

约束仅为等式约束的情况：

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & Ax = b \end{aligned}$$

凸优化问题

约束仅为等式约束的情况：

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & Ax = b \end{array}$$

若 $\exists x,\ Ax=b,\ x$ 最优 $\Leftrightarrow \forall y, Ay=b, \nabla f_0^\top(x)(y-x)\geq 0$
$Ax=b,\ Ay=b \Rightarrow y=x+v, \forall v\in\mathcal{N}(A)$
x 最优 $\Leftrightarrow \nabla f_0^\top(x)v\geq 0, \forall v\in\mathcal{N}(A)$

凸优化问题

约束仅为等式约束的情况：

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & Ax = b \end{array}$$

若 $\exists x,\ Ax = b,\ x$ 最优 $\Leftrightarrow \forall y, Ay = b, \nabla f_0^\top(x)(y-x) \geq 0$
$Ax = b,\ Ay = b \Rightarrow y = x + v, \forall v \in \mathcal{N}(A)$
x 最优 $\Leftrightarrow \nabla f_0^\top(x) v \geq 0, \forall v \in \mathcal{N}(A)$

如果一个线性函数在子空间中非负，则它在子空间上必恒等于零。因此，

- $v = 0,\ A,\ x = A^{-1}b$
- $\nabla f_0(x) \perp \mathcal{N}(A)$

凸优化问题

约束仅为非负约束的情况：

$$
\begin{aligned}
\min \quad & f_0(x) \\
\text{s.t.} \quad & x \succeq 0
\end{aligned}
$$

凸优化问题

约束仅为非负约束的情况：

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & x \succeq 0 \end{array}$$

若 $\exists x \geq 0$, x 最优 $\Leftrightarrow \forall y \geq 0, \nabla f_0^\top(x)(y-x) \geq 0$

凸优化问题

约束仅为非负约束的情况：

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & x \succeq 0 \end{aligned}$$

若 $\exists x \geq 0$, x 最优 $\Leftrightarrow \forall y \geq 0, \nabla f_0^\top(x)(y-x) \geq 0$
因此，可以得出以下结论：

- 若 $\exists i$ 使得 $(\nabla f_0(x))_i < 0$, 则 $\nabla f_0^\top(x) y$ 必可以取无穷小, 则必有 $\nabla f_0(x) \geq 0$

凸优化问题

约束仅为非负约束的情况：

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & x \succeq 0 \end{array}$$

若 $\exists x \geq 0$, x 最优 $\Leftrightarrow \forall y \geq 0, \nabla f_0^\top(x)(y-x) \geq 0$
因此，可以得出以下结论：

- 若 $\exists i$ 使得 $(\nabla f_0(x))_i < 0$, 则 $\nabla f_0^\top(x)y$ 必可以取无穷小, 则必有 $\nabla f_0(x) \geq 0$
- $\forall y$ 均有 $\nabla f_0^\top(x)(y-x) \geq 0 \Rightarrow \nabla f_0^\top(x)x \leq 0$

凸优化问题

约束仅为非负约束的情况：

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & x \succeq 0 \end{array}$$

若 $\exists x \geq 0$, x 最优 $\Leftrightarrow \forall y \geq 0, \nabla f_0^\top(x)(y-x) \geq 0$
因此，可以得出以下结论：

- 若 $\exists i$ 使得 $(\nabla f_0(x))_i < 0$, 则 $\nabla f_0^\top(x) y$ 必可以取无穷小, 则必有 $\nabla f_0(x) \geq 0$
- $\forall y$ 均有 $\nabla f_0^\top(x)(y-x) \geq 0 \Rightarrow \nabla f_0^\top(x) x \leq 0$
- $\nabla f_0(x) \succeq 0, x \succeq 0$ 且 $\nabla f_0^\top(x) x \leq 0 \Rightarrow \nabla f_0^\top(x) x = 0$

凸优化问题

约束仅为非负约束的情况：

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & x \succeq 0 \end{aligned}$$

若 $\exists x \geq 0$, x 最优 $\Leftrightarrow \forall y \geq 0, \nabla f_0^\top(x)(y-x) \geq 0$
因此，可以得出以下结论：

- 若 $\exists i$ 使得 $(\nabla f_0(x))_i < 0$, 则 $\nabla f_0^\top(x) y$ 必可以取无穷小, 则必有 $\nabla f_0(x) \geq 0$
- $\forall y$ 均有 $\nabla f_0^\top(x)(y-x) \geq 0 \Rightarrow \nabla f_0^\top(x) x \leq 0$
- $\nabla f_0(x) \succeq 0, x \succeq 0$ 且 $\nabla f_0^\top(x) x \leq 0 \Rightarrow \nabla f_0^\top(x) x = 0$

最优化条件

$x \succeq 0,\ \nabla f_0(x) \succeq 0,\ (\nabla f_0(x))_i x_i = 0(\text{Complementarity}), i = 1, \ldots, n$

线性规划

当目标函数和约束函数均为仿射时，称为**线性规划** (LP)。

$$\begin{aligned} \min \quad & C^\top x + d \\ \text{s.t.} \quad & Gx \preceq h \\ & Ax = b \end{aligned}$$

线性规划

当目标函数和约束函数均为仿射时，称为**线性规划** (LP)。

$$
\begin{aligned}
\min \quad & C^\top x + d \\
\text{s.t.} \quad & Gx \preceq h \\
& Ax = b
\end{aligned}
$$

其中 $C \in \mathbb{R}^n, d \in \mathbb{R}, G \in \mathbb{R}^{m\times n}, h \in \mathbb{R}^m, A \in \mathbb{R}^{k\times n}, b \in \mathbb{R}^k$。
等价变换：

$$
\begin{aligned}
min \quad & C^\top x + d \\
\text{s.t.} \quad & Gx + s = h \\
& Ax = b \\
& s \succeq 0
\end{aligned}
\quad
\overset{x^+ \succeq 0,\, x^- \succeq 0}{\underset{}{\overset{x = x^+ - x^-}{\Longleftrightarrow}}}
\quad
\begin{aligned}
min \quad & C^\top x^+ - C^\top x^- + d \\
\text{s.t.} \quad & Gx^+ - G^\top x^- + s = h \\
& Ax^+ - A^\top x^- = b \\
& s \succeq 0, x^+ \succeq 0, x^- \succeq 0
\end{aligned}
$$

线性分式规划

当目标函数和约束函数均为仿射时，称为**线性规划** (LP)。

$$
\begin{aligned}
\min \quad & f_0(x) = \frac{c^\top x + d}{e^\top x + f} \\
\text{s.t.} \quad & Gx \preceq h \\
& Ax = b
\end{aligned}
$$

其中 dom$f_0 = \{x \mid e^\top x + f > 0\}$。

线性分式规划

当目标函数和约束函数均为仿射时，称为**线性规划** (LP)。

$$\begin{aligned} \min \quad & f_0(x) = \frac{c^\top x + d}{e^\top x + f} \\ \text{s.t.} \quad & Gx \preceq h \\ & Ax = b \end{aligned}$$

其中 $\mathrm{dom} f_0 = \{x \mid e^\top x + f > 0\}$。
将其转化为等价的线性规划：

$$\begin{aligned} \min \quad & c^\top y + dz \\ \text{s.t.} \quad & Gy - hz \preceq 0 \\ & Ay - bz = 0 \\ & e^\top y + fz = 1 \\ & z \geq 0 \end{aligned}$$

二次规划

当目标函数为 (凸) 二次型，且约束函数为仿射时，称为**二次规划** (QP)。

$$\begin{array}{ll} \min & \frac{1}{2}x^\top Px + q^\top x + r \\ \text{s.t.} & Gx \preceq h \\ & Ax = b \end{array}$$

其中 $P \in S^n_+$。

二次规划

当目标函数为 (凸) 二次型，且约束函数为仿射时，称为**二次规划** (QP)。

$$
\begin{aligned}
\min \quad & \frac{1}{2}x^\top Px + q^\top x + r \\
\text{s.t.} \quad & Gx \preceq h \\
& Ax = b
\end{aligned}
$$

其中 $P \in S^n_+$。
如果不等式约束也为二次型，则称为**二次约束二次规划** (QCQP)。

$$
\begin{aligned}
\min \quad & \frac{1}{2}x^\top Px + q^\top x + r \\
\text{s.t.} \quad & \frac{1}{2}x^\top P_i x + q_i^\top x + r_i \le 0, i = 1, \dots, m \\
& Ax = b
\end{aligned}
$$

其中 $P \in S^n_+$，$P_i \in S^n_+$。。

多目标优化

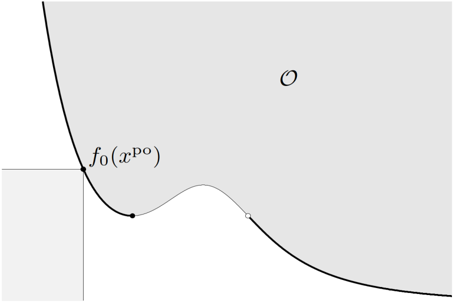
多目标优化与传统的单目标优化相对。多目标优化的概念是在某个情景中需要达到多个目标时，由于容易存在目标间的内在冲突，一个目标的优化是以其他目标劣化为代价，因此很难出现唯一最优解，取而代之的是在他们中间做出协调和折衷处理，使总体的目标尽可能的达到最优。

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \le 0, i = q, \ldots, m \\ & h_i(x) = 0, i = 1, \ldots, p \end{aligned}$$

其中 $f_0 : \mathbb{R}^n \to \mathbb{R}^q$, $f_i : \mathbb{R}^n \to \mathbb{R}$, $h_i : \mathbb{R}^n \to \mathbb{R}$。

多目标优化

Pareto Optimal Front：若找到另一解使之在某指标上更优，必在另外的某指标上变得更差。
Pareto Optimal Value：对应的函数值
Pareto Optimal Point：对应的策略



多目标优化

若 $\{f_0(x)\}$ 在 $\mathbb{R}^k$ 中为凸，$f_i(x)$ 为凸，$h_i(x)$ 为仿射，则必可通过下述方法求得 Pareto front 中的一点。

$$
\begin{aligned}
\min \quad & \sum_{i=1}^{q} \lambda_i f_{0i}(x) \\
\text{s.t.} \quad & f_i(x) \le 0, i = q, \ldots, m \\
& h_i(x) = 0, i = 1, \ldots, p
\end{aligned}
$$

其中 $\lambda_i \ge 0$. 通过遍历 $\{\lambda_i\}$ 可找出所有点。

多目标优化

若 $\{f_0(x)\}$ 在 $\mathbb{R}^k$ 中为凸，$f_i(x)$ 为凸，$h_i(x)$ 为仿射，则必可通过下述方法求得 Pareto front 中的一点。

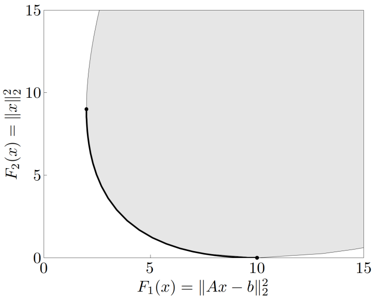
$$\begin{aligned} \min \quad & \sum_{i=1}^{q} \lambda_i f_{0i}(x) \\ \text{s.t.} \quad & f_i(x) \le 0, i = q, \ldots, m \\ & h_i(x) = 0, i = 1, \ldots, p \end{aligned}$$

其中 $\lambda_i \ge 0$. 通过遍历 $\{\lambda_i\}$ 可找出所有点。

解多目标优化问题，一般是将多目标变单目标，而这是通过将某些优化目标变为约束实现的。

多目标优化

岭回归 (Ridge Regression)
$b = Ax + e,\ A \in \mathbb{R}^{m\times n},\ b \in \mathbb{R}^m,\ e \in \mathbb{R}^m,\ x \in \mathbb{R}^n$。



$$\begin{array}{ll} \min & \|b - Ax\|_2^2 \\ \min & \|x\|_2^2 \\ & \Updownarrow \\ \min & \|b - Ax\|_2^2 + \lambda\|x\|_2^2 \\ & \Updownarrow \\ \min & \|b - Ax\|_2^2 \\ \text{s.t.} & \|x\|_2^2 \le \epsilon \end{array}$$

习题

4.3, 4.9, 4.22, 4.24, 4.59, 4.62

章节目录

拉格朗日对偶

考虑如下一般优化问题：

$$
\begin{aligned}
\text{min} \quad & f_0(x) \\
\text{s.t.} \quad & f_i(x) \le 0, i = 1, \dots, m \\
& h_i(x) = 0, i = 1, \dots, p
\end{aligned}
$$

拉格朗日对偶

考虑如下一般优化问题：

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \le 0, i = 1, \ldots, m \\ & h_i(x) = 0, i = 1, \ldots, p \end{aligned}$$

其中 $x \in \mathbb{R}^n$, $\mathcal{D} = \bigcap_{i=0}^{m} \mathrm{dom} f_i \cap \bigcap_{i=1}^{p} \mathrm{dom} h_i$，问题的最优值为 $p^\star$。

拉格朗日对偶

考虑如下一般优化问题：

$$\begin{aligned} \text{min} \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, i = 1, \ldots, m \\ & h_i(x) = 0, i = 1, \ldots, p \end{aligned}$$

其中 $x \in \mathbb{R}^n$, $\mathcal{D} = \bigcap_{i=0}^{m} \text{dom} f_i \cap \bigcap_{i=1}^{p} \text{dom} h_i$，问题的最优值为 $p^\star$。

Lagrangian function/Lagrangian:

$$L(x, \lambda, \mu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \mu_i h_i(x)$$

拉格朗日对偶

考虑如下一般优化问题：

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, i = 1, \ldots, m \\ & h_i(x) = 0, i = 1, \ldots, p \end{aligned}$$

其中 $x \in \mathbb{R}^n$, $\mathcal{D} = \bigcap_{i=0}^{m} \mathrm{dom} f_i \cap \bigcap_{i=1}^{p} \mathrm{dom} h_i$，问题的最优值为 $p^\star$。

Lagrangian function/Lagrangian:

$$L(x, \lambda, \mu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \mu_i h_i(x)$$

Lagrange multiplier: λ, μ

拉格朗日对偶

Lagrange dual function/dual function

$$g(\lambda, \mu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \mu) = \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \mu_i h_i(x) \right)$$

拉格朗日对偶

Lagrange dual function/dual function

$$g(\lambda, \mu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \mu) = \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \mu_i h_i(x) \right)$$

- 对偶函数为凹函数，但是对偶问题是凸优化问题

拉格朗日对偶

Lagrange dual function/dual function

$$g(\lambda,\mu)=\inf_{x\in\mathcal{D}} L(x,\lambda,\mu)=\inf_{x\in\mathcal{D}}\left(f_0(x)+\sum_{i=1}^{m}\lambda_i f_i(x)+\sum_{i=1}^{p}\mu_i h_i(x)\right)$$

- 对偶函数为凹函数，但是对偶问题是凸优化问题
- $\forall\lambda\geq 0,\ \forall\mu,\ g(\lambda,\mu)\leq p^\star$

拉格朗日对偶

Lagrange dual function/dual function

$$g(\lambda, \mu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \mu) = \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \mu_i h_i(x) \right)$$

- 对偶函数为凹函数，但是对偶问题是凸优化问题
- $\forall \lambda \geq 0,\ \forall \mu,\ g(\lambda, \mu) \leq p^\star$

证明：设 $x^\star$ 是原问题的最优化，则必可行，则满足 $f_i(x^\star) \leq 0$, $h_i(x^\star) = 0$。

拉格朗日对偶

Lagrange dual function/dual function

$$g(\lambda,\mu)=\inf_{x\in\mathcal{D}} L(x,\lambda,\mu)=\inf_{x\in\mathcal{D}}\left(f_0(x)+\sum_{i=1}^{m}\lambda_i f_i(x)+\sum_{i=1}^{p}\mu_i h_i(x)\right)$$

- 对偶函数为凹函数，但是对偶问题是凸优化问题
- $\forall\lambda\geq 0,\ \forall\mu,\ g(\lambda,\mu)\leq p^\star$

证明：设 $x^\star$ 是原问题的最优化，则必可行，则满足 $f_i(x^\star)\leq 0$, $h_i(x^\star)=0$。对 $\forall\lambda\geq 0$, $\forall\mu$, 都有 $\sum_{i=1}^{m}\lambda_i f_i(x^\star)+\sum_{i=1}^{p}\mu_i h_i(x^\star)\leq 0$。

拉格朗日对偶

Lagrange dual function/dual function

$$g(\lambda, \mu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \mu) = \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \mu_i h_i(x) \right)$$

- 对偶函数为凹函数，但是对偶问题是凸优化问题
- $\forall \lambda \geq 0,\ \forall \mu,\ g(\lambda, \mu) \leq p^\star$

证明：设 $x^\star$ 是原问题的最优化，则必可行，则满足 $f_i(x^\star) \leq 0$, $h_i(x^\star) = 0$。对 $\forall \lambda \geq 0,\ \forall \mu$, 都有 $\sum_{i=1}^{m} \lambda_i f_i(x^\star) + \sum_{i=1}^{p} \mu_i h_i(x^\star) \leq 0$。

$$L(x^\star, \lambda, \mu) = f_0(x^\star) + \sum_{i=1}^{m} \lambda_i f_i(x^\star) + \sum_{i=1}^{p} \mu_i h_i(x^\star) \leq p^\star$$

拉格朗日对偶

Lagrange dual function/dual function

$$g(\lambda,\mu)=\inf_{x\in\mathcal{D}} L(x,\lambda,\mu)=\inf_{x\in\mathcal{D}}\left(f_0(x)+\sum_{i=1}^{m}\lambda_i f_i(x)+\sum_{i=1}^{p}\mu_i h_i(x)\right)$$

- 对偶函数为凹函数，但是对偶问题是凸优化问题
- $\forall\lambda\geq 0,\ \forall\mu,\ g(\lambda,\mu)\leq p^\star$

证明：设 $x^\star$ 是原问题的最优化，则必可行，则满足 $f_i(x^\star)\leq 0$, $h_i(x^\star)=0$。对 $\forall\lambda\geq 0,\ \forall\mu$, 都有 $\sum_{i=1}^{m}\lambda_i f_i(x^\star)+\sum_{i=1}^{p}\mu_i h_i(x^\star)\leq 0$。

$$L(x^\star,\lambda,\mu)=f_0(x^\star)+\sum_{i=1}^{m}\lambda_i f_i(x^\star)+\sum_{i=1}^{p}\mu_i h_i(x^\star)\leq p^\star$$

$$g(\lambda,\mu)=\inf_{x\in\mathcal{D}} L(x,\lambda,\mu)\leq L(x^\star,\lambda,\mu)\leq p^\star$$

共轭函数

f^* 是 $f:\mathbb{R}^n\to\mathbb{R}$ 的共轭，若 $f^*(y)=\sup_{x\in\mathrm{dom}f}\left(y^\top x-f(x)\right)$。

共轭函数

f^* 是 $f: \mathbb{R}^n \to \mathbb{R}$ 的共轭，若 $f^*(y) = \sup_{x \in \mathrm{dom} f} \left(y^\top x - f(x) \right)$。

$$
\text{例:} \quad \begin{array}{ll} \min & f(x) \\ \text{s.t.} & x = 0 \end{array}
$$

共轭函数

f^* 是 $f:\mathbb{R}^n \to \mathbb{R}$ 的共轭，若 $f^*(y) = \sup_{x\in \mathrm{dom} f}\left(y^\top x - f(x)\right)$。

例:
$$
\begin{aligned}
\min \quad & f(x) \\
\text{s.t.} \quad & x = 0
\end{aligned}
$$

$$
\begin{aligned}
&\Rightarrow L(x,\mu) = f(x) + \mu^\top x,\ \mathrm{dom} L = \mathrm{dom} f \times \mathbb{R}^n \\
&\Rightarrow g(\mu) = \inf_{x\in \mathrm{dom} f}\left(f(x) + \mu^\top x\right) \\
&\qquad = -\sup_{x\in \mathrm{dom} f}\left(-\mu^\top x - f(x)\right) \\
&\qquad = -f^*(-\mu)
\end{aligned}
$$

共轭函数

$$
\text{例:}\quad \begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & Ax \preceq b \\ & Cx = d \end{aligned}
$$

共轭函数

$$
\text{例:}\quad \begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & Ax \preceq b \\ & Cx = d \end{aligned}
$$

$$
\begin{aligned}
\Rightarrow L(x,\lambda,\mu) &= f(x) + \lambda^\top (Ax - b) + \mu^\top (Cx - d) \\
&= f_0(x) + \left(\lambda^\top A + \mu^\top C\right) x - \lambda^\top b - \mu^\top d \\
\Rightarrow g(\lambda,\mu) &= \inf_{x \in \mathrm{dom} f} L(x,\lambda,\mu) \\
&= -f^*\left(-\left(\lambda^\top A + \mu^\top C\right)\right) - \lambda^\top b - \mu^\top d
\end{aligned}
$$

拉格朗日对偶问题

$$
(P)\quad \begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & f_i(x) \le 0, i = 1, \dots, m \\ & h_i(x) = 0, i = 1, \dots, p \end{array}
$$

$$
(D)\quad \begin{array}{ll} \max & g(\lambda, \mu) \\ \text{s.t.} & \lambda \ge 0 \end{array}
$$

拉格朗日对偶问题

$$
(P)\quad \begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & f_i(x) \le 0, i = 1, \ldots, m \\ & h_i(x) = 0, i = 1, \ldots, p \end{array}
$$

$$
(D)\quad \begin{array}{ll} \max & g(\lambda, \mu) \\ \text{s.t.} & \lambda \ge 0 \end{array}
$$

- $d^\star \le p^\star$(weak duality); $d^\star = p^\star$(strong duality)

拉格朗日对偶问题

$$
(P) \quad \begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & f_i(x) \leq 0, i = 1, \ldots, m \\ & h_i(x) = 0, i = 1, \ldots, p \end{array}
$$

$$
(D) \quad \begin{array}{ll} \max & g(\lambda, \mu) \\ \text{s.t.} & \lambda \geq 0 \end{array}
$$

- $d^\star \leq p^\star$(weak duality); $d^\star = p^\star$(strong duality)
- $p^\star - d^\star$: duality gap

拉格朗日对偶问题

$$
(P) \quad \begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & f_i(x) \le 0, i = 1, \dots, m \\ & h_i(x) = 0, i = 1, \dots, p \end{array}
$$

$$
(D) \quad \begin{array}{ll} \max & g(\lambda, \mu) \\ \text{s.t.} & \lambda \ge 0 \end{array}
$$

- $d^\star \le p^\star$(weak duality); $d^\star = p^\star$(strong duality)
- $p^\star - d^\star$: duality gap
- $\lambda^\star$, $\mu^\star$: optimal Lagrange multipliers

Slater's condition

集合 D 的 relative interior：

$\mathrm{relint} D = \{x \in D \mid B(x, r) \cap \mathrm{aff} D \leq D, \exists r > 0\}$

Slater's condition

集合 D 的 relative interior：
relint$D = \{x \in D \mid B(x, r) \cap \text{aff}D \leq D, \exists r > 0\}$
Slater's condition($p^\star = d^\star$，充分条件)
若有凸问题

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, i = 1, \ldots, m \\ & Ax = b \end{aligned}$$

Slater's condition

集合 D 的 relative interior：

relint$D = \{x \in D \mid B(x, r) \cap \text{aff}D \leq D, \exists r > 0\}$

Slater's condition($p^\star = d^\star$，充分条件)

若有凸问题

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, i = 1, \ldots, m \\ & Ax = b \end{aligned}$$

其中 $f_i(x)$ 为凸，$\forall i$，当 $\exists x \in$ relint$D$，使 $f_i(x) < 0, i = 1, \ldots, m, Ax = b$ 满足时，$p^\star = d^\star$。

A weaker Slater's condition

若不等式约束为仿射 ($Cx + d \leq 0$) 时，只要可行域非空，必有 $p^\star = d^\star$。

几何解释

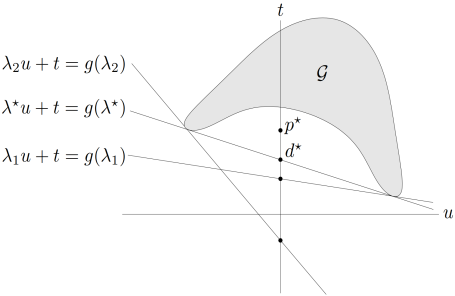
针对如下优化问题：

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & f_i(x) \leq 0, i = 1, \ldots, m \end{array}$$

几何解释

针对如下优化问题：

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \le 0, i = 1, \dots, m \end{aligned}$$



考虑单不等式约束的情况：

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_1(x) \le 0 \end{aligned}$$

$$\begin{aligned} G &= \{(f_1(x), f_0(x)) \mid x \in D\} \\ p^\star &= \inf\{t \mid (u, t) \in G, u \le 0\} \\ g(\lambda) &= \inf\{\lambda u + t \mid (u, t) \in G\} \end{aligned}$$

多目标优化的解释

考虑如下多目标优化问题：

$$\min\{f_0(x), f_1(x), \ldots, f_m(x)\}$$

多目标优化的解释

考虑如下多目标优化问题：

$$\min\{f_0(x), f_1(x), \ldots, f_m(x)\}$$

通过引入一组 $\{\lambda_i\}$ 可将上述优化问题转化为 $\min\ f_0(x) + \sum_i \lambda_i f_i(x)$。

多目标优化的解释

考虑如下多目标优化问题：

$$\min\{f_0(x), f_1(x), \ldots, f_m(x)\}$$

通过引入一组 $\{\lambda_i\}$ 可将上述优化问题转化为 $\min\ f_0(x) + \sum_i \lambda_i f_i(x)$。$L(x, \lambda) = f_0(x) + \sum_i \lambda_i f_i(x)$ 是如下优化问题的拉格朗日函数。

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \le 0, i = 1, \ldots, m \end{aligned}$$

多目标优化的解释

考虑如下多目标优化问题：

$$\min\{f_0(x), f_1(x), \ldots, f_m(x)\}$$

通过引入一组 $\{\lambda_i\}$ 可将上述优化问题转化为 $\min\ f_0(x) + \sum_i \lambda_i f_i(x)$。$L(x,\lambda) = f_0(x) + \sum_i \lambda_i f_i(x)$ 是如下优化问题的拉格朗日函数。

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & f_i(x) \le 0, i = 1, \ldots, m \end{array}$$

求解步骤：

1. 先给定 $\lambda \Rightarrow \min\limits_{x\in D}\ L(x,\lambda) \Rightarrow \tilde{x}$
2. 再 $\max\limits_{\lambda\ge 0}\ g(\lambda) \Rightarrow \tilde{\lambda}$
3. 然后 $\min\limits_{x\in D}\ L(x,\tilde{\lambda}) \Rightarrow x$

多目标优化的解释

考虑如下多目标优化问题：

$$\min\{f_0(x), f_1(x), \ldots, f_m(x)\}$$

通过引入一组 $\{\lambda_i\}$ 可将上述优化问题转化为 $\min\ f_0(x) + \sum_i \lambda_i f_i(x)$。$L(x,\lambda) = f_0(x) + \sum_i \lambda_i f_i(x)$ 是如下优化问题的拉格朗日函数。

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \le 0, i = 1, \ldots, m \end{aligned}$$

求解步骤：

1. 先给定 $\lambda \Rightarrow \min\limits_{x \in D} \ L(x,\lambda) \Rightarrow \tilde{x}$
2. 再 $\max\limits_{\lambda \ge 0} \ g(\lambda) \Rightarrow \tilde{\lambda}$
3. 然后 $\min\limits_{x \in D} \ L(x,\tilde{\lambda}) \Rightarrow x$

当对偶间隙为 0 时，该 x 为 $\min\ f_0(x) + \sum_i \tilde{\lambda} f_i(x)$ 的最优解。

鞍点的解释

原问题最优值： $p^\star = \inf_{x\in D} \sup_{\lambda\geq 0} L(x,\lambda)$ 上确界的最小值

对偶问题最优值： $d^\star = \sup_{\lambda\geq 0} \inf_{x\in D} L(x,\lambda)$ 下确界的最大值

弱对偶情况下：

$$
\begin{aligned}
d^\star &= \sup_{\lambda\geq 0} \inf_{x\in D} L(x,\lambda) = \inf_{x\in D} L(x,\lambda^\star) \\
&\leq L(x^\star,\lambda^\star) = f_0(x^\star) + \sum_i \lambda_i^\star f_i(x^\star) \\
&\leq f_0(x^\star) = p^\star
\end{aligned}
$$

鞍点的解释

鞍点：$(\tilde{w}, \tilde{z})$ 使不等式成立，则称为鞍点。

$$f(\tilde{w}, z) \leq f(\tilde{w}, \tilde{z}) \leq f(w, \tilde{z})$$

鞍点的解释

鞍点：$(\tilde{w}, \tilde{z})$ 使不等式成立，则称为鞍点。

$$f(\tilde{w}, z) \leq f(\tilde{w}, \tilde{z}) \leq f(w, \tilde{z})$$

鞍点定理：若 $(\tilde{x}, \tilde{\lambda})$ 为 $L(x, \lambda)$ 的鞍点 $\Leftrightarrow$ 强对偶存在，且 $\tilde{x}, \tilde{\lambda}$ 为原问题和对偶问题的最优解。

回顾原始对偶问题

回顾：

$$
(P)\quad \begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & f_i(x) \leq 0, i=1,\ldots,m \\ & h_i(x) = 0, i=1,\ldots,p \end{array}
$$

$$
(D)\quad \begin{array}{ll} \max & g(\lambda,\mu) \\ \text{s.t.} & \lambda \geq 0 \end{array}
$$

回顾原始对偶问题

回顾：

$$
(P)\quad \begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & f_i(x) \le 0, i = 1, \ldots, m \\ & h_i(x) = 0, i = 1, \ldots, p \end{array}
$$

$$
(D)\quad \begin{array}{ll} \max & g(\lambda, \mu) \\ \text{s.t.} & \lambda \ge 0 \end{array}
$$

其中对偶函数为：

$$
g(\lambda, \mu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \mu) \le L(x^\star, \lambda, \mu) \le p^\star
$$

KKT 条件

假设 $p^\star = d^\star$，且所有函数可微。则 $x^\star$, $\lambda^\star$, $\mu^\star$ 分别为原问题和对偶问题的最优解，满足如下**可行条件** (primal/dual feasibility)：

$$\begin{cases} f_i(x^\star) \le 0, i = 1, \dots, m \\ h_i(x^\star) = 0, i = 1, \dots, p \\ \lambda^\star \ge 0 \end{cases}$$

KKT 条件

假设 $p^\star = d^\star$，且所有函数可微。则 $x^\star$, $\lambda^\star$, $\mu^\star$ 分别为原问题和对偶问题的最优解，满足如下**可行条件** (primal/dual feasibility)：

$$\begin{cases} f_i(x^\star) \le 0, i = 1, \ldots, m \\ h_i(x^\star) = 0, i = 1, \ldots, p \\ \lambda^\star \ge 0 \end{cases}$$

弱对偶关系的不等式关系为：

$$\begin{aligned} d^\star &= \sup_{\lambda \ge 0, \mu} \inf_{x \in D} L(x, \lambda, \mu) = \inf_{x \in D} L(x, \lambda^\star, \mu^\star) \\ &\le L(x^\star, \lambda^\star, \mu^\star) = f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star f_i(x^\star) + \sum_{i=1}^p \mu_i^\star h_i(x^\star) \\ &\le f_0(x^\star) = p^\star \end{aligned}$$

KKT 条件

假设 $p^\star = d^\star$，且所有函数可微。则 $x^\star$, $\lambda^\star$, $\mu^\star$ 分别为原问题和对偶问题的最优解，满足如下**可行条件** (primal/dual feasibility)：

$$\begin{cases} f_i(x^\star) \leq 0, i = 1, \dots, m \\ h_i(x^\star) = 0, i = 1, \dots, p \\ \lambda^\star \geq 0 \end{cases}$$

弱对偶关系的不等式关系为：

$$\begin{aligned} d^\star &= \sup_{\lambda \geq 0, \mu} \inf_{x \in D} L(x, \lambda, \mu) = \inf_{x \in D} L(x, \lambda^\star, \mu^\star) \\ &\leq L(x^\star, \lambda^\star, \mu^\star) = f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star f_i(x^\star) + \sum_{i=1}^p \mu_i^\star h_i(x^\star) \\ &\leq f_0(x^\star) = p^\star \end{aligned}$$

当问题满足强对偶关系时，上述不等式都取等号。

KKT 条件

此时，$\sum_{i=1}^{m} \lambda_i^\star f_i(x^\star) = 0 \Leftrightarrow \lambda_i^\star f_i(x^\star) = 0$。

$$\begin{cases} \lambda_i^\star > 0\text{时}, f_i(x^\star) = 0 \\ f_i(x^\star) < 0\text{时}, \lambda_i^\star = 0 \end{cases}$$

KKT 条件

此时，$\sum_{i=1}^{m} \lambda_i^\star f_i(x^\star) = 0 \Leftrightarrow \lambda_i^\star f_i(x^\star) = 0$。

$$\begin{cases} \lambda_i^\star > 0\text{时}, f_i(x^\star) = 0 \\ f_i(x^\star) < 0\text{时}, \lambda_i^\star = 0 \end{cases}$$

这一性质称为**互补松弛** (Complementarity Slackness)。

KKT 条件

此时，$\sum_{i=1}^{m} \lambda_i^\star f_i(x^\star) = 0 \Leftrightarrow \lambda_i^\star f_i(x^\star) = 0$。

$$\begin{cases} \lambda_i^\star > 0\text{时}, f_i(x^\star) = 0 \\ f_i(x^\star) < 0\text{时}, \lambda_i^\star = 0 \end{cases}$$

这一性质称为**互补松弛** (Complementarity Slackness)。
第二行的不等式取等号，说明 $x^\star$ 是 $L(x, \lambda^\star, \mu^\star)$ 的极小值点，则这点处偏导数为 0。

$$\left.\frac{\partial L(x, \lambda^\star, \mu^\star)}{\partial x}\right|_{x=x^\star} = 0$$

即 $\nabla f_0(x^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla f_i(x^\star) + \sum_{i=1}^{p} \mu_i^\star \nabla h_i(x^\star) = 0$，这个条件称为**稳定点** (stationary)。

KKT 条件

以上三个条件 (可行条件、互补松弛、稳定点) 组合在一起，就是 KKT 条件 (Karush-Kuhn-Tucker condition)。

$$
\begin{cases}
f_i(x^\star) \le 0, i = 1, \dots, m \\
h_i(x^\star) = 0, i = 1, \dots, p \\
\lambda_i^\star \ge 0, i = 1, \dots, m \\
\lambda_i^\star f_i(x^\star) = 0, i = 1, \dots, m \\
\nabla f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star \nabla f_i(x^\star) + \sum_{i=1}^p \mu_i^\star \nabla h_i(x^\star) = 0
\end{cases}
$$

KKT 条件

以上三个条件 (可行条件、互补松弛、稳定点) 组合在一起，就是 KKT 条件 (Karush-Kuhn-Tucker condition)。

$$\begin{cases} f_i(x^\star) \le 0, i = 1, \ldots, m \\ h_i(x^\star) = 0, i = 1, \ldots, p \\ \lambda_i^\star \ge 0, i = 1, \ldots, m \\ \lambda_i^\star f_i(x^\star) = 0, i = 1, \ldots, m \\ \nabla f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star \nabla f_i(x^\star) + \sum_{i=1}^p \mu_i^\star \nabla h_i(x^\star) = 0 \end{cases}$$

当原问题为凸问题、各个函数可微，且满足强对偶关系时，KKT 条件是最优解的充分必要条件。

敏感性分析

$$
\text{(原问题)}\quad
\begin{array}{ll}
\min & f_0(x) \\
\text{s.t.} & f_i(x) \le 0,\ i = 1, \ldots, m \\
 & h_i(x) = 0,\ i = 1, \ldots, p
\end{array}
$$

$$
\text{(干扰问题)}\quad
\begin{array}{ll}
\min & f_0(x) \\
\text{s.t.} & f_i(x) \le u_i,\ i = 1, \ldots, m \\
 & h_i(x) = w_i,\ i = 1, \ldots, p
\end{array}
$$

敏感性分析

$$
(\text{原问题})\quad
\begin{aligned}
\min\quad & f_0(x) \\
\text{s.t.}\quad & f_i(x) \le 0,\ i = 1, \dots, m \\
& h_i(x) = 0,\ i = 1, \dots, p
\end{aligned}
$$

$$
(\text{干扰问题})\quad
\begin{aligned}
\min\quad & f_0(x) \\
\text{s.t.}\quad & f_i(x) \le u_i,\ i = 1, \dots, m \\
& h_i(x) = w_i,\ i = 1, \dots, p
\end{aligned}
$$

最优值变为 $p^\star(u, w)$，而 $p^\star(0, 0) = p^\star$。

敏感性分析

$$
(\text{原问题})\quad \begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & f_i(x) \leq 0, i=1,\ldots,m \\ & h_i(x)=0, i=1,\ldots,p \end{array}
$$

$$
(\text{干扰问题})\quad \begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & f_i(x) \leq u_i, i=1,\ldots,m \\ & h_i(x)=w_i, i=1,\ldots,p \end{array}
$$

最优值变为 $p^\star(u,w)$，而 $p^\star(0,0)=p^\star$。

- **性质 1**：若原问题为凸问题，则 $p^\star(u,w)$ 为 (u,w) 的凸函数

敏感性分析

$$
\text{(原问题)} \quad \begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & f_i(x) \le 0, i = 1, \ldots, m \\ & h_i(x) = 0, i = 1, \ldots, p \end{array}
$$

$$
\text{(干扰问题)} \quad \begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & f_i(x) \le u_i, i = 1, \ldots, m \\ & h_i(x) = w_i, i = 1, \ldots, p \end{array}
$$

最优值变为 $p^\star(u, w)$，而 $p^\star(0, 0) = p^\star$。

- **性质 1**：若原问题为凸问题，则 $p^\star(u, w)$ 为 (u, w) 的凸函数
- **性质 2**：若原问题为凸问题，且对偶间隙为 0，$\lambda^\star$、$\mu^\star$ 为对偶问题的最优解，则

$$
p^\star(u, w) \ge p^\star(0, 0) - (\lambda^\star)^\top u - (\mu^\star)^\top w
$$

敏感性分析

性质 2 的应用:

- 若 $\lambda_i^\star$ 很大，且加紧第 i 项不等式约束 $(u_i < 0)$，则 $p^\star(u, w)$ 急剧增加

敏感性分析

性质 2 的应用：

- 若 $\lambda_i^\star$ 很大，且加紧第 i 项不等式约束 $(u_i < 0)$，则 $p^\star(u,w)$ 急剧增加
- 若 $\mu_i^\star$ 为很大的正值，且 $w_i < 0$；或 μ_i 为绝对值很大的负值，且 $w_i > 0$，则 $p^\star(u,w)$ 急剧增加

敏感性分析

性质 2 的应用：

- 若 $\lambda_i^\star$ 很大，且加紧第 i 项不等式约束 $(u_i < 0)$，则 $p^\star(u,w)$ 急剧增加
- 若 $\mu_i^\star$ 为很大的正值，且 $w_i < 0$；或 μ_i 为绝对值很大的负值，且 $w_i > 0$，则 $p^\star(u,w)$ 急剧增加
- 若 $\lambda_i^\star$ 很小，且 $u_i > 0$，则 $p^\star(u,w)$ 下降不大

敏感性分析

性质 2 的应用：

- 若 $\lambda_i^\star$ 很大，且加紧第 i 项不等式约束 $(u_i < 0)$，则 $p^\star(u,w)$ 急剧增加
- 若 $\mu_i^\star$ 为很大的正值，且 $w_i < 0$；或 μ_i 为绝对值很大的负值，且 $w_i > 0$，则 $p^\star(u,w)$ 急剧增加
- 若 $\lambda_i^\star$ 很小，且 $u_i > 0$，则 $p^\star(u,w)$ 下降不大
- 若 $\mu_i^\star$ 为很小的正值，且 $w_i < 0$；或 μ_i 为绝对值很小的负值，且 $w_i > 0$，则 $p^\star(u,w)$ 几乎不变

敏感性分析

性质 2 的应用：

- 若 $\lambda_i^\star$ 很大，且加紧第 i 项不等式约束 ($u_i < 0$)，则 $p^\star(u,w)$ 急剧增加
- 若 $\mu_i^\star$ 为很大的正值，且 $w_i < 0$；或 μ_i 为绝对值很大的负值，且 $w_i > 0$，则 $p^\star(u,w)$ 急剧增加
- 若 $\lambda_i^\star$ 很小，且 $u_i > 0$，则 $p^\star(u,w)$ 下降不大
- 若 $\mu_i^\star$ 为很小的正值，且 $w_i < 0$；或 μ_i 为绝对值很小的负值，且 $w_i > 0$，则 $p^\star(u,w)$ 几乎不变

- **性质 3**：(局部敏感性) 若原问题为凸问题，且对偶间隙为 0，且 $p^\star(u,w)$ 在 $(u,w) = (0,0)$ 处可微，

$$\lambda_i^\star = -\left.\frac{\partial p^\star(u,w)}{\partial u_i}\right|_{(0,0)} \qquad \mu_i^\star = -\left.\frac{\partial p^\star(u,w)}{\partial w_i}\right|_{(0,0)}$$

习题

5.5, 5.20, 5.27

章节目录

无约束优化

章节目录

等式约束优化