南昌大学
NANCHANG UNIVERSITY

大数据优化：理论、算法及其应用

——来源于《最优化计算方法》（高教社）

★★★ 主讲人：彭振华 ★★★

联系方式：zhenhuapeng@ncu.edu.cn
zhenhuapeng@whu.edu.cn
15870605317（微信同号）

研究兴趣：1. 非凸非光滑优化算法与理论
2. 智能决策
3. 智能计算与机器学习

数学与计算机学院

2026年

课件资源：https://zhenhuapeng.github.io//coursematerials/

目录

基础知识

第 二 部分

- 范数与导数
- 凸规划：凸集与凸函数
- 共轭函数
- 次梯度

- 思考: 数之间可以比较大小, 那向量和矩阵咋比较呢?
- 受距离定义启发, 给出范数的定义如下:

范数: 称一个从向量空间R^n到实数域R的非负函数$||\cdot||$为范数, 如果

I. 正定性: 对于所有的$v \in R^n$, 有$||v|| \geq 0$, 且$||v|| = 0$当且仅当$v = 0$;

II. 齐次性: 对于所有的$v \in R^n$和$\alpha \in R$, 有$||\alpha v|| = |\alpha|||v||$;

III. 三角不等式: 对于所有的$v,w \in R^n$, 有$||v + w|| \leq ||v|| + ||w||$.

你能联想到实数和复数中哪些相关概念?

范数与导数

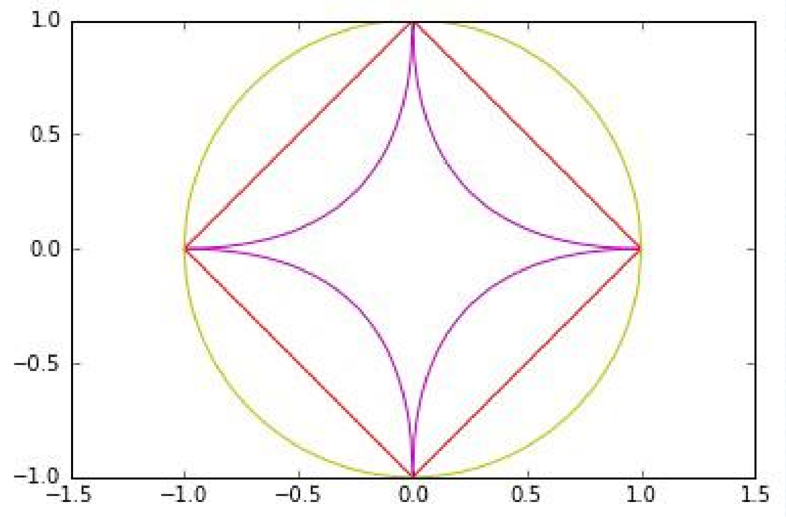
➢ 常用的向量范数

I. $||v||_0$: 向量v 中非0元素的个数

II. $||v||_p = (|v_1|^p + \cdots + |v_n|^p)^{1/p}$ (p ≥ 1)

III. $||v||_\infty = \max_i |v_i|$

IV. 由正定矩阵A诱导的范数$||v||_A = \sqrt{x^T A x}$



- $||v||_1 = |v_1| + \cdots + |v_n|$
- $||v|| = (|v_1|^2 + \cdots + |v_n|^2)^{1/2}$
- 柯西（Cauchy）不等式：$|a^T b| \le ||a||\ ||b||$
- 练习: 求向量$v = (0, 0, 1, 1, 2, 0, 0, 1, 1)^T$的各个范数

练习：$\|x\|_\infty \le \|x\|_2 \le \|x\|_1 \le n^{1/2}\|x\|_2 \le n\|x\|_\infty$

*课程思政：范数也是稀疏优化和机器学习领域中常用的正则项, 用来防止过拟合, 已被成功应用于金融、医疗和生物科学和图像识别和处理等领域中. 引导学生关注范数在实际应用中的运用和立志投身相关研究中，为祖国发展贡献力量.

范数与导数

➢ 矩阵范数

I. $||A||_1 = \sum\sum |a_{ij}|$ (一般指最大列绝对值和)

II. $||A||_2 = \lambda_{max}(A)$

III. $||A||_F = \sqrt{Tr(AA^T)} = \sqrt{\sum\sum (a_{ij})^2}$

IV. 核范数：$||A||_* = \sum\delta_i$ (非0奇异值求和)

练习：(1) $\|A\|_2 \le ||A||_F$

(2) 计算如下矩阵的各种范数A=[0, 1; 1, 1]

□ 矩阵F范数具有正交不变性：对任意的正交矩阵U, V，有

$$(||UAV||_F)^2 = Tr(UAVV^TA^TU^T) = Tr(UAA^TU^T) = Tr(AA^TU^TU) = Tr(AA^T) = (||A||_F)^2$$

□ 矩阵范数的相容性：$||Ax||_2 \le ||A||_2||x||_2$

□ 矩阵范数的柯西不等式：$|<A,B>|\ (= |Tr(AB^T)| = |\sum\sum a_{ij}b_{ij}|) \le ||A||_F||B||_F$

范数与导数

➤导入模块库

```
import numpy as np
import cvxpy as cp
```

➤计算向量范数

```
x = np.array([0, 0, 1, 1, 2, 0, 0, 1, 1])
print(np.linalg.norm(x, ord = 1))
print(np.linalg.norm(x, ord = 2))
print(np.linalg.norm(x,ord = np.inf))
print(cp.norm(x,1).value)
print(cp.norm(x,2).value)
print(cp.norm(x,"inf").value)
```

➤计算矩阵范数

```
A = np.array([[0,1],[1,1]])
print(np.linalg.norm(A, ord = 1))
print(np.linalg.norm(A, ord = 2))
print(np.linalg.norm(A, ord = np.inf))
print(np.linalg.norm(A, 'fro'))
print(cp.norm(A,1).value)
print(cp.norm(A,2).value)
print(cp.norm(A,"inf").value)
print(cp.norm(A,"fro").value)
```

范数与导数

➢导数定义

$$f'(x)=\lim_{\Delta x\to 0}\frac{f(x+\Delta x)-f(x)}{\Delta x}$$

➢例题：求如下函数的导数 *f(x) = x² + x/2 + xsinx - sinx/x*

$$f'(x)=2x+\frac{1}{2}+x\cos x+\sin x+\frac{\sin x}{x^2}-\frac{\cos x}{x}$$

➢一阶泰勒展开式

$$f(x+\Delta x)=f(x)+f'(x)\Delta x+o(\Delta x)$$

思考：如何推广至多元函数呢？

范数与导数

- 设 $f(x) = f(x_1, x_2, \ldots, x_n)$, 梯度的计算 $\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \ldots, \frac{\partial f(x)}{\partial x_n} \right)^T$
- 梯度的定义

$$\lim_{p \to 0} \frac{f(x+p) - f(x) - g^T p}{\|p\|} = 0$$

- 取 $p = \varepsilon\ e_i$, 有

$$\lim_{\varepsilon \to 0} \frac{f(x_1, \ldots, x_i + \varepsilon, \ldots, x_n) - f(x_1, \ldots, x_i, \ldots, x_n) - g_i \varepsilon}{\varepsilon} = 0$$

$$\Downarrow$$

$$\lim_{\varepsilon \to 0} \frac{f(x_1, \ldots, x_i + \varepsilon, \ldots, x_n) - f(x_1, \ldots, x_i, \ldots, x_n)}{\varepsilon} = g_i$$

- **海瑟Hessian矩阵：**

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

对称矩阵

- **雅可比Jacobi矩阵：**

$$\boldsymbol{f(x) = (f_1(x), f_2(x), \ldots, f_l(x))}$$

$$J(x) = \begin{pmatrix} \frac{\partial f_1(x)}{\partial x_1} & \frac{\partial f_1(x)}{\partial x_2} & \cdots & \frac{\partial f_1(x)}{\partial x_n} \\ \frac{\partial f_2(x)}{\partial x_1} & \frac{\partial f_2(x)}{\partial x_2} & \cdots & \frac{\partial f_2(x)}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_l(x)}{\partial x_1} & \frac{\partial f_l(x)}{\partial x_2} & \cdots & \frac{\partial f_l(x)}{\partial x_n} \end{pmatrix}$$

范数与导数

- **例题: 对实对称矩阵**A, **求函数**$f_1(x) = d^T x$**和**$f_2(x) = 1/2x^TAx + b^T x + c$**的梯度和Hessian矩阵?**
- $\nabla f_1(x) = d$ $\qquad \nabla^2 f_1(x) = 0$
- $\nabla f_2(x) = Ax + b$ $\qquad \nabla^2 f_2(x) = A$
- **练习：求解函数**$f_3(x) = ||x||^2$**和**$f_4(x) = 1/2||Cx - d||^2$**的梯度和Hessian矩阵?**
- $\nabla f_3(x) = 2x$ $\qquad \nabla^2 f_3(x) = 2I$
- $\nabla f_4(x) = C^T(Cx - d)$ $\qquad \nabla^2 f_4(x) = C^TC$

范数与导数

- **拉格朗日中值定理**
- **如果函数$f(x)$满足在闭区间$[a,b]$上连续，在开区间(a,b)内可导，那么在(a,b)内至少有一点z，使等式$f(b) - f(a) = f'(z)(b-a)$成立。**
- **积分中值定理**
- **如果函数$f(x)$满足在闭区间$[a,b]$上连续，那么在(a,b)内至少有一点z，使得**

$$\int_a^b f(x)dx = f(z)(b-a)$$

范数与导数

➢ 一阶泰勒展开式

$$f(x+p)=f(x)+\nabla f(x)^T p+o(\|p\|)$$
$$=f(x)+\int_0^1 \nabla f(x+tp)^T p\,dt$$

➢ 二阶泰勒展开式

$$\nabla f(x+tp)=\nabla f(x)+\int_0^1 \nabla^2 f(x+tp)p\,dt$$

$$f(x+p)=f(x)+\nabla f(x)^T p+\frac{1}{2}p^T\nabla^2 f(x+tp)\,p$$

其中$t\in(0,1)$

范数与导数

➢ **梯度L-利普希茨连续(L-光滑)：**

$$\left\|\nabla f\left(x\right)-\nabla f\left(y\right)\right\| \le L\left\|x-y\right\|, \forall x, y \in dom\ f.$$

➢ **二次上界：** $f(y) \le f(x)+\nabla f(x)^T(y-x)+\dfrac{L}{2}\left\|y-x\right\|^2, \forall x, y \in dom\ f.$

$$f(y)-f(x)-\nabla f(x)^T(y-x)=\int_0^1 \nabla f(x+t(y-x))^T(y-x)dt-\nabla f(x)^T(y-x)$$

$$=\int_0^1 \left(\nabla f(x+t(y-x))-\nabla f(x)\right)^T(y-x)dt$$

$$\le \int_0^1 \left\|\nabla f(x+t(y-x))-\nabla f(x)\right\|\left\|y-x\right\|dt$$

$$\le \int_0^1 Lt\left\|y-x\right\|^2 dt=\frac{L}{2}\left\|y-x\right\|^2$$

范数与导数

➢ 设可微函数$f(x)$的定义域为R^n 且存在一个全局极小点x^*，若$f(x)$为梯度L- 利普希茨连续的, 则对任意的x有

$$\frac{1}{2L}\left\|\nabla f(x)\right\|^2 \le f(x) - f\left(x^*\right)$$

$$f\left(x^*\right) \le f(y) \le f(x) + \nabla f(x)^T (y-x) + \frac{L}{2}\left\|y-x\right\|^2$$

$$\inf_y f\left(x^*\right) \le \inf_y f(x) + \nabla f(x)^T (y-x) + \frac{L}{2}\left\|y-x\right\|^2$$

$$y = x - \frac{\nabla f(x)}{L}$$

$$\longrightarrow \quad f\left(x^*\right) \le f(x) - \frac{1}{2L}\left\|\nabla f(x)\right\|^2$$

范数与导数

➢ 对于以$m \times n$矩阵X为自变量的函数$f(X)$，若存在矩阵$G \in R^{m\times n}$满足

$$\lim_{V \to 0} \frac{f(X+V)-f(X)-\langle G, V\rangle}{\|V\|}=0$$

称G为f在Fréchet可微意义下的梯度。类似于向量情形，矩阵变量函数$f(X)$的梯度为

$$\nabla f(X)=\begin{pmatrix} \frac{\partial f}{\partial x_{11}} & \frac{\partial f}{\partial x_{12}} & \cdots & \frac{\partial f}{\partial x_{1n}} \\ \frac{\partial f}{\partial x_{21}} & \frac{\partial f}{\partial x_{22}} & \cdots & \frac{\partial f}{\partial x_{2n}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f}{\partial x_{m1}} & \frac{\partial f}{\partial x_{m2}} & \cdots & \frac{\partial f}{\partial x_{mn}} \end{pmatrix}$$

在实际应用中，矩阵Fréchet可微的定义和使用往往比较繁琐

- **Gâteaux 可微**：设$f(X)$为矩阵变量函数，如果对任意方向$V \in R^{m\times n}$，存在矩阵$G \in R^{m\times n}$满足

$$\lim_{t\to 0}\frac{f(X+tV)-f(X)-t\langle G, V\rangle}{t}=0$$

称G为f在X处在Gâteaux可微意义下的梯度.

当f是Fréchet 可微函数时，f也是Gâteaux可微的，且这两种意义下的梯度相等.

- $\nabla_X(\mathrm{Tr}(AX^TB)) = \mathrm{BA}$ $\nabla_X(\mathrm{Tr}(X^TAX)) = (A+A^T)\mathrm{X}$ $\nabla_X(\mathrm{Tr}(XAX^T)) = \mathrm{X}(A+A^T)$
- $\nabla_X(\mathrm{Tr}(\mathrm{X}AX^TB)) = \mathrm{BXA} + \mathrm{B}^T\mathrm{XA}^T$ $\nabla_X(\det(X)) = \det(X)(\mathrm{X}^{-1})^T$
- $\nabla_X(1/2\|XY - A\|^2) = (XY - A)Y^T$ $\nabla_Y(1/2\|XY - A\|^2) = X^T(XY - A)$

凸规划: 凸集与凸函数

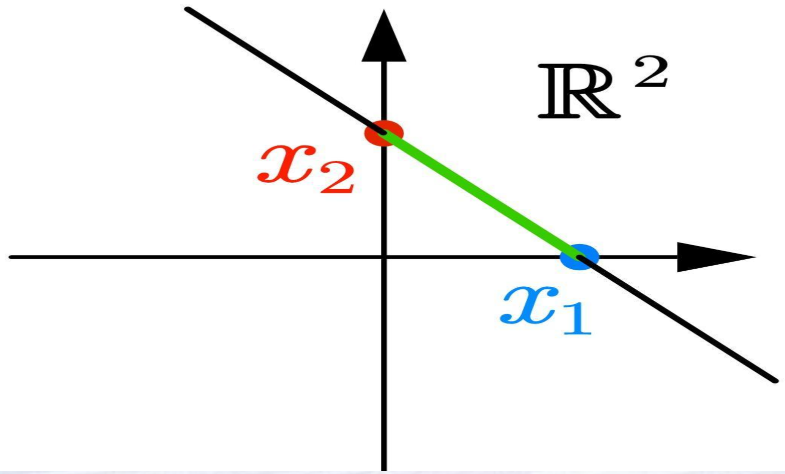
- **凸与非凸优化**

$$\min_{x} f(x) \quad s.t. \; x \in \chi.$$

I. 凸优化：目标函数是凸函数+可行域为凸集

II. 非凸优化：目标函数不是凸函数或可行域不为凸集

怎样判断凸集和凸函数?

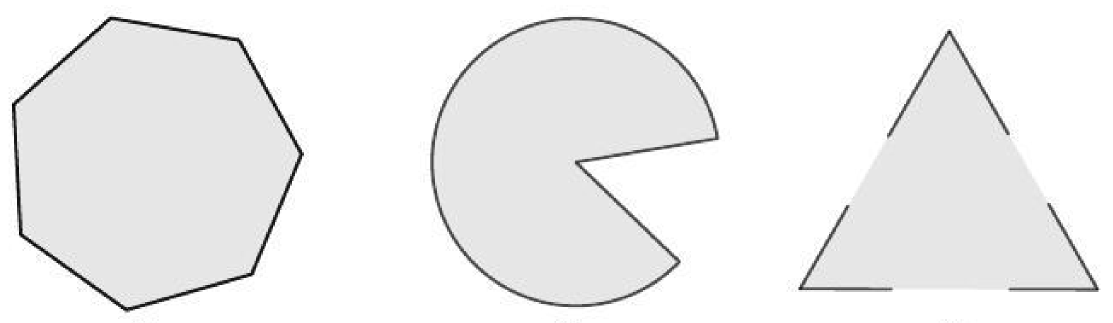


- $x_2 + \theta(x_1 - x_2) = \theta x_1 + (1 - \theta)x_2$
- **直线：** $\theta \in R$
- **线段：** $\theta \in [0,1]$
- **凸集：如果连接集合C中的任意两点的线段都在C内, 则称C为凸集, 即**

$$x_1, x_2 \in C \Rightarrow \theta x_1 + (1-\theta)x_2 \in C, \; \forall 0 \le \theta \le 1.$$

凸规划: 凸集与凸函数

- 列出了一些凸集、非凸集的例子. 其中(a) 为凸集, (b) 和(c)均为非凸集.



(a) (b) (c)

- 从凸集中可以引出凸组合和凸包的概念.

- $x_1, \ldots, x_k$ 的凸组合：$x = \theta_1 x_1 + \cdots + \theta_k x_k,\ \theta_1 + \cdots + \theta_k = 1,\ \ \theta_i \geq 0, i=1, \ldots, k$

- 凸包：集合S的所有点的凸组合构成的点集为S的凸包, 记为$conv\ S$.

若$\mathrm{conv}\ S \subseteq S$, 则$S$是凸集; 反之亦然.

凸规划: 凸集与凸函数

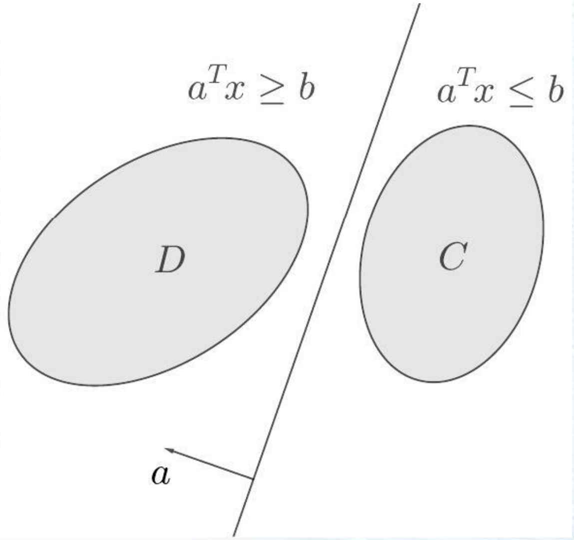
➤ 重要凸集例子

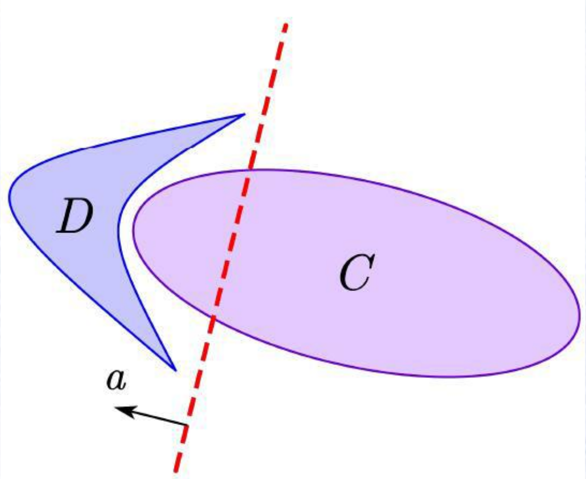
凸集有什么好处?

- □ 超平面：$\{x: a^T x = b\}$
- □ 半空间：$\{x: a^T x \leq b\}$
- □ 多面体：$\{x: Ax \leq b, Cx = d\}$
- □ 范数球：$\{x: \|x - x_0\| \leq r\}$
- □ 范数锥：$\{(x, t): \|x\| \leq t\}$

➤ 保凸运算

- □ 任意多个凸集的交为凸集
- □ S是凸集，则$\{Ax + b : x \in S\}$为凸集



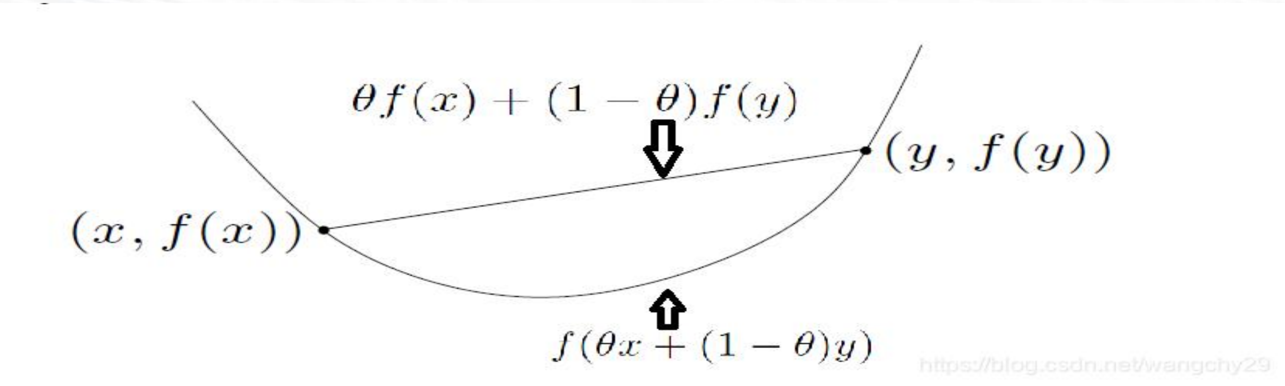


凸规划: 凸集与凸函数

➢ **凸函数**：$f : R^n \to R$ 为适当函数，如果$D \subset dom\, f$是凸集，且

$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta)f(y)$$

对所有$x, y \in D$, $0 \le \theta \le 1$ 都成立，则称函数f在集合D上是凸函数。



- 若f是凸函数，则$-f$是**凹函数**
- 若对所有$x, y \in D$, $x \neq y$, $0 < \theta < 1$，有$f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta)f(y)$

则称f在集合D上是**严格凸函数**

凸规划: 凸集与凸函数

➢ 一元凸函数的例子

- 仿射函数: 对任意$a \in R^n, b \in R$, $f(x) = a^T x + b$ 是R上的凸函数(凹函数)
- 指数函数: 对任意$a \in R$, e^{ax} 是R上的凸函数
- 幂函数: 对$\alpha \geq 1$ 或$\alpha \leq 0$, x^{α} 是R_{++}上的凸函数
- 绝对值的幂: 对$p \geq 1$, $|x|^p$ 是R上的凸函数
- 负熵: $x \log x$ 是R_{++}上的凸函数
- 幂函数: 对$0 \leq \alpha \leq 1$, x^{α} 是R_{++}上的凹函数
- 对数函数: $\log x$ 是R_{++}上的凹函数

凸规划: 凸集与凸函数

➢ 练习：ℓ -1/2、ℓ -1，ℓ -2，ℓ -∞范数哪些是凸（凹）函数?

凸优化：目标函数是凸函数+可行域为凸集

➢ 练习：你能判断如下优化问题哪些是凸优化?

$(I)\min_x \sqrt{(x-x_0)^2+(ax+b-y_0)^2}$

$(II)\min_{\ell,d}\ \ell\times d \quad s.t.\ 2(\ell+d)=22.$

(III)线性规划、二次规划、约束优化

$(IV)\min_{a,b}\sum_i\left(a^T x^i+b-y^i\right)^2$

$(V)\min_a\sum_i\left(a_0+a_1x^i+a_2\left(x^i\right)^2\cdots-y^i\right)^2$

$(VI)\min_\theta\|X\theta-y\|^2$

(VII)硬(软)间隔SVM、SVR

$(VIII)$均值方差模型

(IX)交通网络模型

(X)深度神经网络模型

凸规划: 凸集与凸函数

$$\min_x \ f(x) \quad s.t. \ x \in \chi.$$

➢ **凸规划的局部最优解均是全局最优解**

□ 设x^*是凸规划的局部最优解，但不是其全局最优解，则必存在可行解x'，使得$f(x') < f(x^*)$. ***对充分小的正数*** θ***，有***

$$f\left(x^* + \theta\left(x' - x^*\right)\right) = f\left(\theta x' + (1-\theta)x^*\right)$$

由可行域为凸集得 $x^* + \theta(x' - x^*)$ 是凸规划的可行解

$$\le \theta f\left(x'\right) + (1-\theta) f\left(x^*\right)$$

$$< f\left(x^*\right)$$

□ 这与x^*是凸规划的局部最优解矛盾。

凸规划: 凸集与凸函数

$$\min_{x} \; f(x) \quad s.t. \; x \in \chi.$$

➢ **若凸规划解存在，且其目标函数是严格凸的，则其最优解唯一**

□ 设x^*和x'是凸规划的最优解，但不相等。不难发现$f(x') = f(x^*)$. 对

$\theta \in (0,1)$，有

$$f\left(\theta x' + (1-\theta) x^*\right) < \theta f\left(x'\right) + (1-\theta) f\left(x^*\right)$$
$$= f\left(x^*\right)$$

由可行域为凸集得 $\theta x' + (1-\theta) x^*$ $(\neq x^*)$ 是凸规划的可行解

□ 这与x^*是凸规划的最优解矛盾。

凸规划: 凸集与凸函数

➢ **强凸函数**：$f: R^n \rightarrow R$ 为适当函数，若存在常数$m > 0$，使得

$$g(x) = f(x) - m/2\|x\|^2$$

为凸函数，则称$f(x)$为强凸函数，其中m为强凸参数．为了方便我们也称$f(x)$为**m-强凸函数**

➢ **等价定义**：若存在常数$m > 0$，使得对所有$x, y \in dom\, f$, $0 < \theta < 1$ 都有

$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta)f(y) - m/2\theta(1-\theta)\|x-y\|^2$$

则称$f(x)$为强凸函数，其中m为强凸参数．

凸规划: 凸集与凸函数

- 练习：ℓ -1/2、ℓ -1，ℓ -2，ℓ -∞范数哪些是强（严格）凸函数?

- 练习：二次函数是否为强（严格）凸函数?

➢ 由凸函数诱导出的重要的不等式

- **基础Jensen 不等式:** 设函数f是凸函数，则对$0 \le \theta \le 1$，有

$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta)f(y)$$

- **Jensen 不等式:** 设函数f是凸函数，则对$\theta_1 + \cdots + \theta_k = 1,\ \theta_i \ge 0,$ $i=1, \ldots, k$，有

$$f(\theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k) \le \sum \theta_i f(x_k)$$

- **概率Jensen 不等式:** 设f是凸函数，则对任意随机变量z, 有

$$f(Ez) \le Ef(z)$$

课件资源：https://zhenhuapeng.github.io//coursematerials/

南昌大学
NANCHANG UNIVERSITY

感 谢 观 看

大数据优化：理论、算法及其应用

——来源于《最优化计算方法》（高教社）

★★★ 主讲人：彭振华 ★★★

联系方式：zhenhuapeng@ncu.edu.cn
zhenhuapeng@whu.edu.cn
15870605317（微信同号）

研究兴趣：1. 非凸非光滑优化算法与理论
2. 智能决策
3. 智能计算与机器学习

数学与计算机学院