最优化理论与方法

（QQ号：982419291 ）

主讲： 白敏茹教授

Email: minru-bai@hnu.edu.cn

助教：杨琨璟

教　　材：

1. 《数值最优化算法与理论》(第二版),李董辉 童小娇 万中 编,科学出版社，2009

2. 《Numerical Optimization》，J. Nocedal, S. J. Wright, Springer，2006

参考教材：

1. 《最优化理论与方法》，袁亚湘，孙文瑜，科学出版社

2. 《最优化：建模、算法与理论》，刘浩洋, 户将,李勇锋,文再文，高等教育出版社，2020

3. 《Convex Optimization》, Stephen Boyd, Lieven Vandenberghe, Cambridge University Press,2004

考核方式

- 平时成绩：20%。主要考核对每堂课知识点的复习、理解和掌握程度,主要包括考勤、作业（含大作业）、课堂测验。
- 编程试验和测验成绩：25%。主要考核编程实现掌握程度，编程实现掌握程度的测验分为平时程序练习测验，以及最后总测验1次。
- 小测验：15%。主要考核对知识点的理解和掌握程度，举行2次测验，分别对无约束优化收敛性理论、算法和约束优化理论进行考核。时间节点大约分别为第9周、第13周.
- 期末考试成绩：40%。主要考核最优化基本概念、基本计算方法的掌握程度。采用闭卷考试形式，在考试周内进行。题型为1、填空题；2、证明题；3、计算题等。

最优化理论和方法

主要介绍非线性规划的理论与求解方法

一、基本概念与基础知识

(第一章、第二章)

二、最优性条件

(第二章、第九章)

三、各类算法及收敛性理论

(其它各章)

教学内容

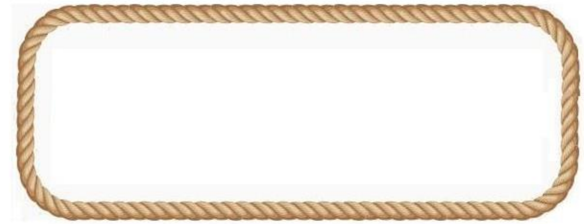
- 凸分析理论
- 无约束问题的最优性条件和下降算法收敛性理论
- 下降算法：最速下降法、牛顿法、拟牛顿法 、共轭梯度法
- 信赖域算法
- 非线性最小二乘问题 、非线性方程组问题
- 大规模无约束优化算法

教学内容

- 约束优化问题的最优性条件
- 对偶性理论
- 罚函数法、增广拉格朗日法、ADMM
- 线性规划*、二次规划*（时间允许时讲）

现实生活中的优化问题

问题： 建筑高宽成何比例时，最美观？



问题：一根绳子围成何种形状，包围的面积最大？

长度固定，如何设计外形，才能使飞机飞行、高铁运行时阻力最小？

由于宇宙组成是最完美也是最聪明造物主之产物，宇宙间万物都遵循某种最大或最小准则

——— 欧拉

Für, da das Gewebe des Universums am vollkommensten und die Arbeit eines klügsten ist Schöpfers, nichts an findet im Universum statt, in dem irgendeine Richtlinie des Maximums oder des Minimums nicht erscheint

Leonhard Euler (1707-1783)

- 任何存在/需要决策的问题都是优化问题

应 用

1990年诺贝尔经济学奖得主马尔柯维茨提出了投资组合问题的均值—方差二次规划模型

- **经济金融**：投资组合
- **交通运输**：列车运行、物流运输
- **信息科学**：数据挖掘、信号处理、机器学习、模式识别
- **生命医学**：医学图像、DNA序列、蛋白质折叠
- **人工智能**：无人驾驶中的智能控制
- **航空航天**：结构优化设计

飞机设计中的颤振问题模型就是一非线性优化问题

二、模型与分类

数学模型

$$\min \quad f(\boldsymbol{x})$$
$$\text{s.t. } \boldsymbol{x} \in \Omega$$

目标函数

决策集

subject to

决策变量

$$\boxed{\max} \quad f(\boldsymbol{x}) \qquad \longrightarrow \qquad \boxed{\min \quad -f(\boldsymbol{x})}$$
$$\text{s.t. } \boldsymbol{x} \in \Omega \qquad\qquad\qquad \text{s.t. } \boldsymbol{x} \in \Omega$$

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

决策变量

$$\Omega = \{\boldsymbol{x} \in \mathbb{R}^n \mid c_i(\boldsymbol{x}) = 0,\ i \in \mathcal{E}; \quad c_i(\boldsymbol{x}) \geqslant 0,\ i \in \mathcal{I}\}$$

等式约束　等式约束指标集　不等式约束　不等式约束指标集

最优化问题分类

1、根据约束划分

无约束优化 $\min_{\boldsymbol{x}\in\mathbb{R}^n} \quad f(\boldsymbol{x})$

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

约束优化问题一般比无约束优化问题难解，但有时可以相互转化

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} \mid \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = 1\} \Longleftrightarrow \min_{\boldsymbol{x}\in\mathbb{R}^n} \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}$$

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}) \Longleftrightarrow \min_{\boldsymbol{x}\in\mathbb{R}^n, t\in\mathbb{R}} \{t \mid t - f(\boldsymbol{x}) \geqslant 0\}$$

2、根据函数的线性度划分

线性规划：目标函数及约束函数都是线性

$$\begin{aligned} \min\ & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \\ \text{s.t.}\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{x} \geqslant \boldsymbol{0} \end{aligned}$$

非线性规划：目标函数或约束中含有非线性函数

$$\begin{aligned} \min\ & -3x_1 + \boxed{\ln(x_2^2 + x_3^2)} \\ \text{s.t.}\ & x_1 + x_2 + \boxed{e^{x_3}} \leqslant 20 \\ & -x_1 + 2x_2 + \boxed{x_3^2} = 0 \end{aligned}$$

非线性项

非线性项

非线性项

一类常用的非线性规划

二次规划：目标函数二次，约束函数线性

$$
\begin{aligned}
\min\ & Q(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{x} + \boldsymbol{g}^{\mathrm{T}}\boldsymbol{x} \\
\text{s.t.}\ & \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x} = b_i,\ i \in \mathcal{E} \\
& \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x} \geqslant b_i,\ i \in \mathcal{I}
\end{aligned}
$$

线性规划和二次规划问题是两类特殊的最优化问题。目前已有比较完善的理论和求解方法。

3、根据目标函数和可行域的凸性划分：

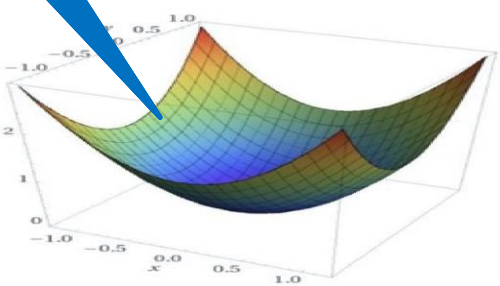
凸规划：目标函数为凸函数，可行域为闭凸集。

凸函数

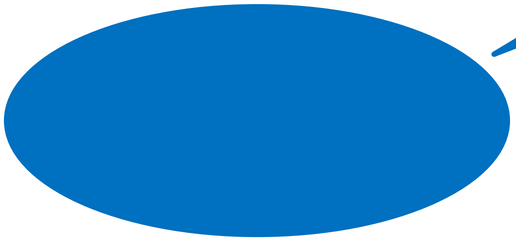
$$
\begin{aligned}
&\min \quad f(\boldsymbol{x}) \\
&\text{s.t.} \ \boldsymbol{x} \in \Omega
\end{aligned}
$$

闭凸集

凸函数



闭凸集



3、根据目标函数和可行域的凸性划分：

凸规划：目标函数为凸函数，可行域为闭凸集。

非凸优化：目标函数非凸或可行域非凸

性质：凸规划问题的局部最优解就是全局最优解，求解相对容易。

非凸优化问题难于求解，特别是全局最优解。

4、根据函数的解析性质划分

光滑优化：所有函数都连续可微, 如多项式优化

$$
\begin{aligned}
\min \ & \left(x_1-\frac{4}{9}\right)^2+(x_2-2)^2 \\
\text{s.t. } & -x_1^2+x_2 \geqslant 0 \\
& x_1+x_2 \leqslant 6 \\
& x_1, x_2 \geqslant 0
\end{aligned}
$$

非光滑优化：含不可微函数

$$
\min_{\boldsymbol{x}\in\mathbb{R}^n} \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x}-b\|^2+\underbrace{\lambda\|\boldsymbol{x}\|_1}_{\text{非光滑项}}
$$

5、根据可行域中可行点的个数划分

连续优化：可行域含有无穷多个不可数的点且可行域中的点连续变化。

离散优化：可行域含有有限(可数)个点, 即在由有限个点或可数个点组成的可行域中寻求最优解。

多数情况下, 该问题的可行域是通过某些元素的排列组合产生的. 因此又称组合优化。

简单的组合优化问题：

烧水泡茶：洗水壶，洗茶壶，洗茶杯，烧水，放茶叶。

如何安排工序才能在最短的时间喝上茶？

（华罗庚，《统筹方法》）

经典的组合优化问题

装箱问题、 旅行商问题(销售商问题)、中国邮路问题。

有多个结构相同、大小相等的箱子及多个大小不同的物品。如何用最少的箱子将全部物品装入箱内，或就现有的箱子装入尽可能多的物品.

给定多个城市和每两城市间的距离。寻求一条最短线路，使得每座城市访问一次并最终回到起始城市

给定多条街道(村庄)及两两之间的距离。寻求一条最短路，从邮局出发，走遍所有街道（村庄），并最终回到邮局(数学家管梅谷1950s）

求解算法的差异：

- **连续优化：** 可借助函数的梯度信息建立梯度类算法。
- **离散优化：** 不能有效利用函数的梯度信息. 若将变量的整数约束松弛为实数，则求解后者得到的最优解无论通过什么方式取整都不能保证它是原问题的最优解.
- **问题联系：** 有些离散优化问题可通过松弛为连续优化问题求解；连续优化问题的优化技术，如对偶，可移植到离散优化问题的研究中.

根据变量的特殊要求，离散优化又分离出

整数规划： 所有变量都取整数（离散优化）

0-1规划： 所有变量取0或1（离散优化）

混合整数规划: 部分变量为整数变量, 其余变量为连续变量

混合0-1规划： 部分变量取0或1, 其余变量为连续变量

稀疏优化： 要求最优解稀疏，即非零个数尽可能少, 即

$$\|\boldsymbol{x}\|_0 \leqslant k$$

6、根据模型参数的确定性划分

确定规划：所有参数都是确定的

不确定规划：含有不确定参数

随机规划：不确定参数服从某种概率分布

模糊规划：源自对事物的不确定性判断，如高矮、黑白、胖瘦、好坏等。（**研究工具：**隶属度）

一些特殊规划：

几何优化、分布式优化、锥规划、多项式优化、鲁棒优化、矩阵优化、动态规划等

三、最优解的概念

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

可行解： $\boldsymbol{x} \in \Omega$

最优解： $\boldsymbol{x}^* \in \Omega$，对任意的 $\boldsymbol{x} \in \Omega$ 都有 $f(\boldsymbol{x}^*) \leqslant f(\boldsymbol{x})$

$$
\boldsymbol{x}^* = \arg\min_{\boldsymbol{x} \in \Omega} f(\boldsymbol{x})
$$

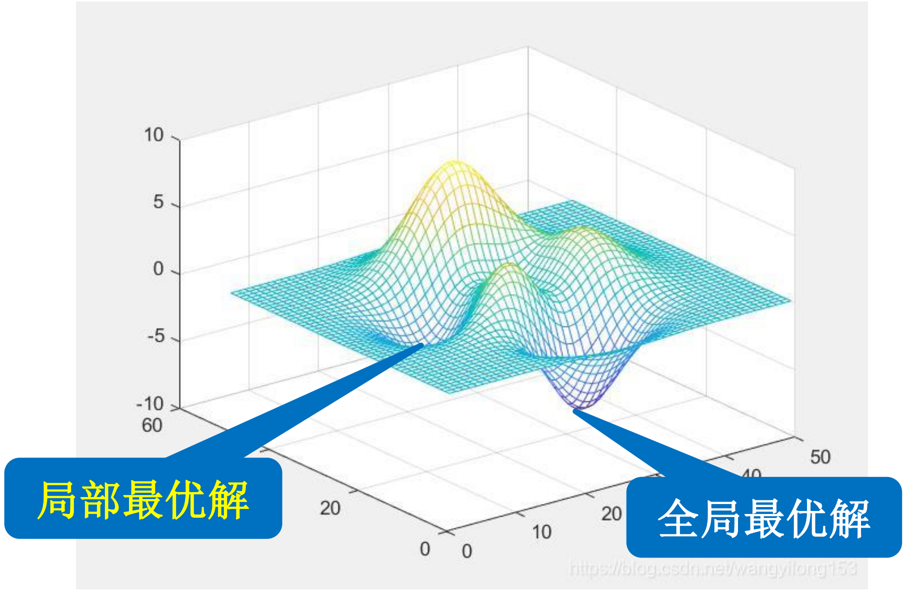
the argument of minimum

局部最优解： $\boldsymbol{x}^* \in \Omega$且存在邻域 $N(\boldsymbol{x}^*, \delta)$, 使对任意的 $\boldsymbol{x} \in N(\boldsymbol{x}^*, \delta) \cap \Omega$,

均有 $f(\boldsymbol{x}^*) \leqslant f(\boldsymbol{x})$

严格局部最优解：对任意的 $\boldsymbol{x} \in N(\boldsymbol{x}^*, \delta) \cap \Omega$, $\boldsymbol{x} \neq \boldsymbol{x}^*$，

都有 $f(\boldsymbol{x}^*) < f(\boldsymbol{x})$



最优值：目标函数在最优解处的值。

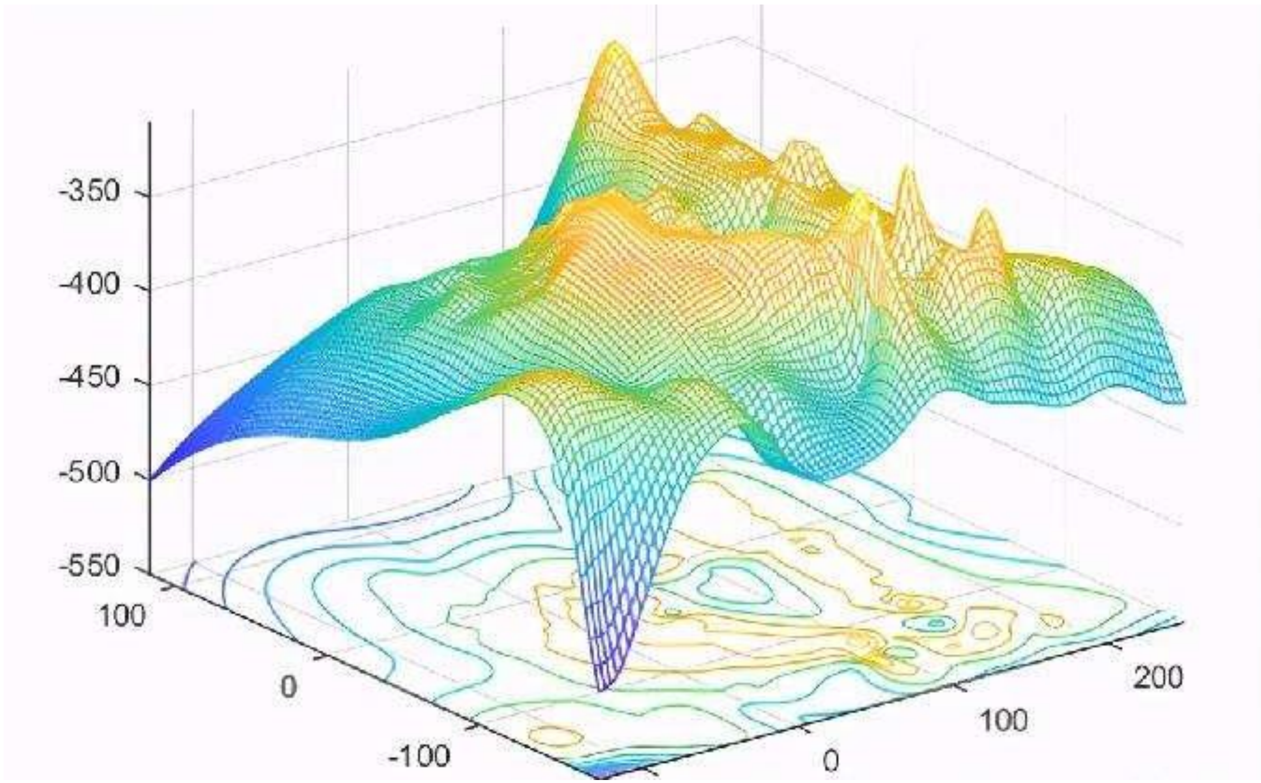
若目标函数在可行域上有下界，但无最优解，则目标函数在可行域上的**下确界**称为优化问题的最优值。

$$\inf\{f(\boldsymbol{x}) \mid \boldsymbol{x} \in \Omega\}$$

二元函数 $f(x) = x_1^2 + (1 - x_1x_2)^2$ 在二维空间中的最优值为0，但只在 $x_1 = 1/x_2$ 且 $x_2 \to \infty$ 时达到。

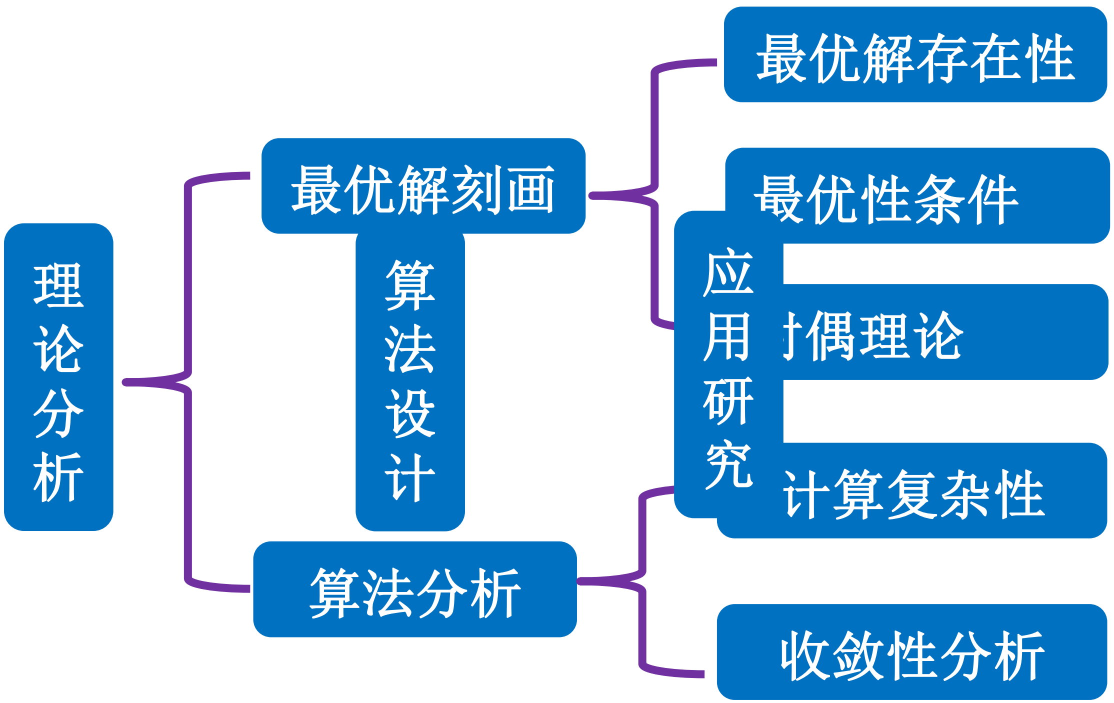
最优值记为

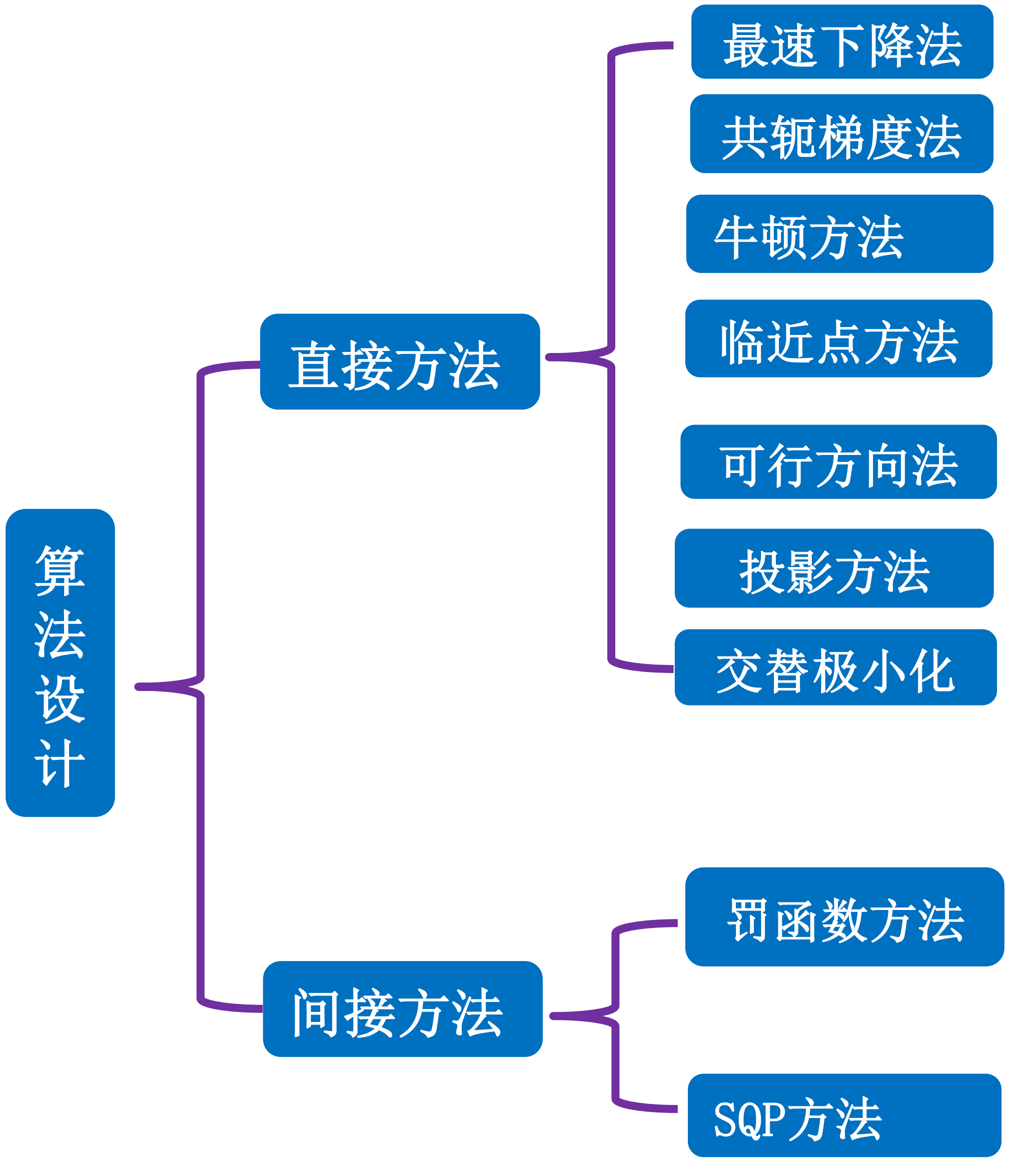
$$f^* = \inf_{x \in \Omega} f(x)$$



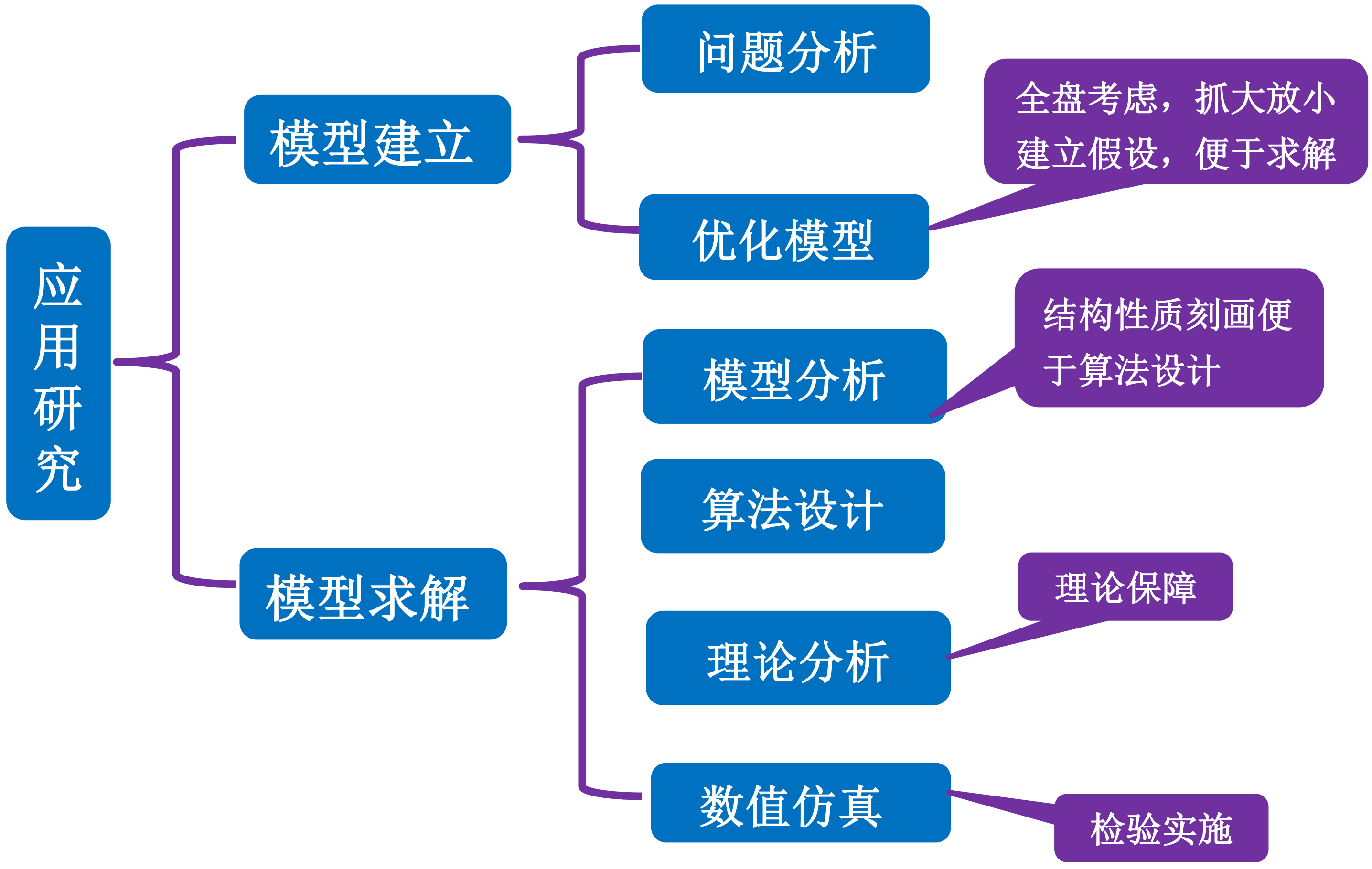
目标函数结构极其复杂，一般得到局部最优解就可以了

主要研究内容





算法设计
直接方法
最速下降法
共轭梯度法
牛顿方法
临近点方法
可行方向法
投影方法
交替极小化
间接方法
罚函数方法
SQP方法


应用研究
模型建立
问题分析
优化模型
全盘考虑，抓大放小
建立假设，便于求解
模型求解
模型分析
结构性质刻画便
于算法设计
算法设计
理论分析
理论保障
数值仿真
检验实施

四、历史与现状

最优化的起源：春秋战国时期的田忌赛马

	齐威王		田忌		齐威王		田忌
上	A	>	B		A	>	F
中	C	>	D		C	<	B
下	E	>	F		E	<	D
	3	:	0	→	1	:	2

以弱胜强

运筹策帷幄之中，决胜于千里之外

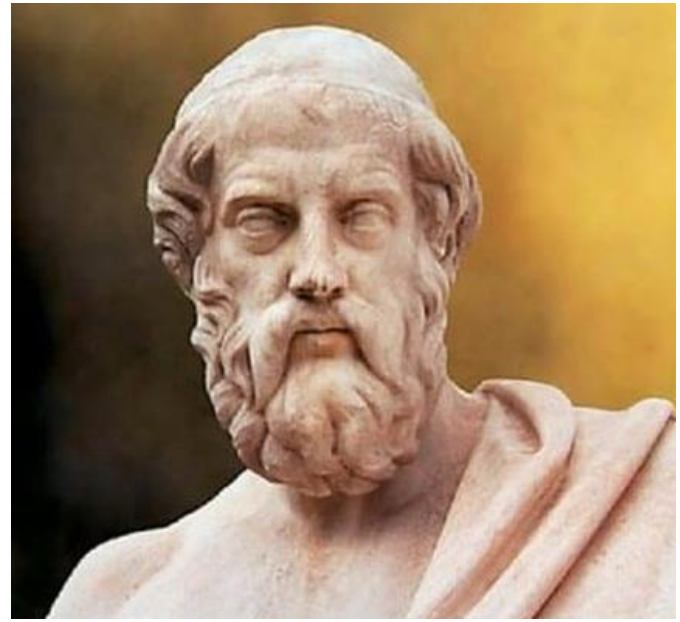
——《史记. 高祖本纪》

军事运筹学： 以弱胜强，

以最小的代价夺取最大的胜利

Operations Research — 运筹学

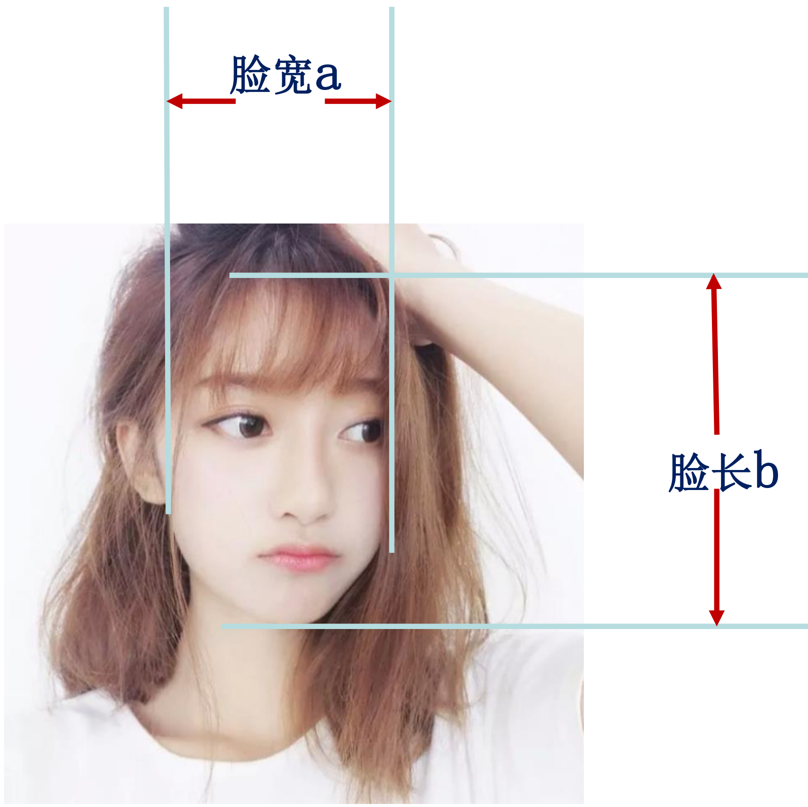
张良　　韩信　　诸葛亮



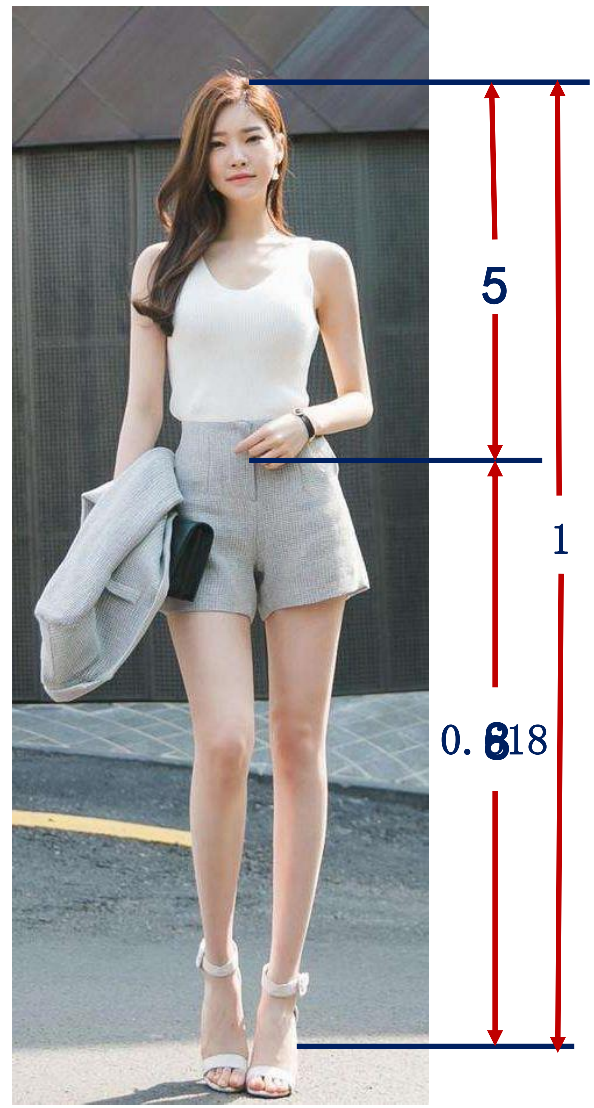
柏拉图

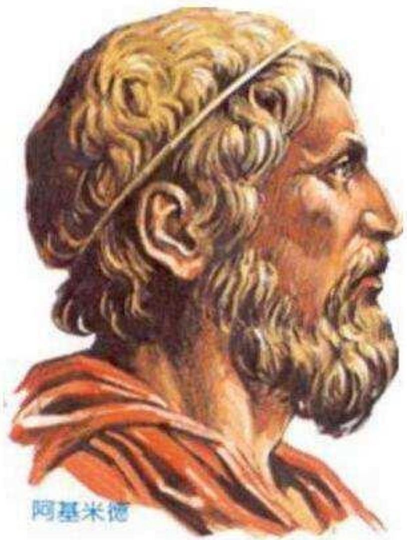
能引起事物美感的最佳比例—0.618

建筑的高宽比大致为0.6(黄金分割点0.618)时最赏心悦目。



a/b=0. 618





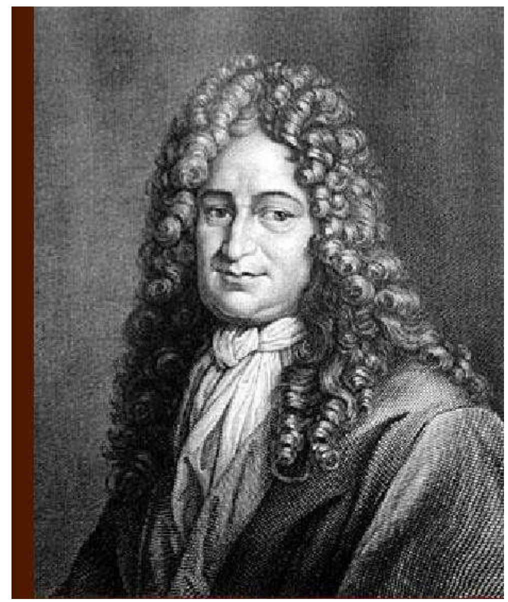
古希腊数学家 **阿基米德**

结论：给定周长，圆所包围的面积为最大。

欧洲城堡几乎都建成圆形

17世纪，牛顿和莱布尼兹发明微积分后，提出了实值函数的极值方法。维尔斯特拉斯对函数变量的函数极值问题，建立了优化问题的变分法——古典最优化方法。

牛顿



莱布尼兹

18世纪，欧拉、拉格朗日、高斯等展开对力学、天文学中极值问题的研究，如最小二乘问题

欧拉

拉格朗日

高斯

二次世界大战之中和之后，军事需要、科学技术的进步和生产的发展，促进了近代最优化方法的产生。

年代	研究问题	主要贡献	贡献者	国籍
1940’	线性规划	单纯形方法	康托罗维奇 Dantzig	苏联数学家 美国数学家
1950’	非线性规划	最优性条件	Kuhn Tucker	美国数学家
1950’	动态规划		贝尔曼	美国数学家
1950’	优化控制	极值原理	庞特里亚金	苏联数学家
1950’	旅行商问题	整数规划	Dantzig Fulkerson Johnson	美国数学家
1960’	多目标规划		查尔斯 库柏	美国数学家

中国的优化专家

将优选法和统筹法（“双法”）推广应用于我国的工农业生产，取得了显著效益。开辟了我国运筹学应用的先河。

中科院院士
著名数学家

华罗庚

华罗庚：同样容积，用料最少！

马蜂窝为什么是六棱柱结构？

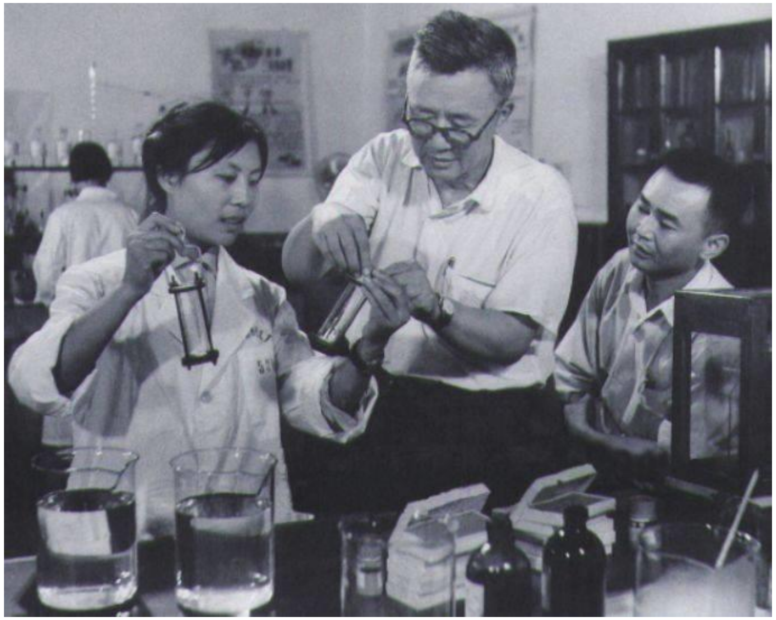
传统60^0　现52^0，45^0，35^0等

60^{0} ？ 52^{0}，45^{0}，35^{0}等

发酵时多加水，酒精度会降低。但酒中的高级脂肪酸乙酯会因溶解度降低而析出，导致酒体浑浊，影响外观，香气也会变差！

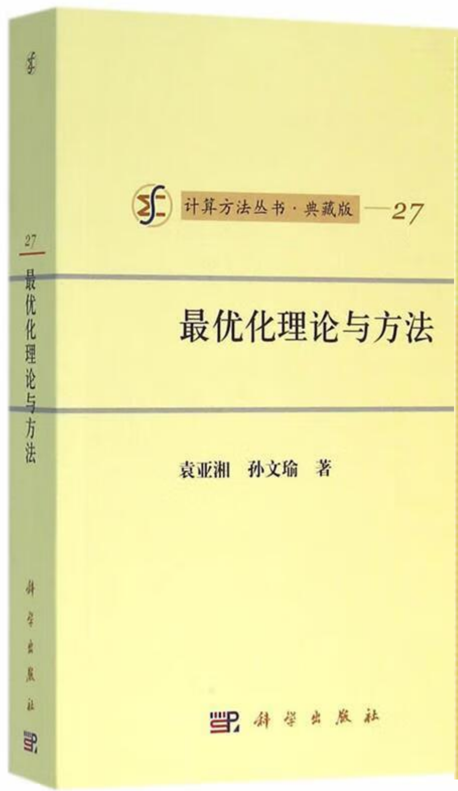
酒中高级脂肪酸乙酯的溶解与酒精度和温度有关。

如何找到合适的发酵温度，使得高级脂肪酸乙酯在低度酒中不溶解或溶解度最低？



华罗庚利用优选法指导五粮液的生产工艺

在非线性规划问题的拟牛顿算法和信赖域法算法等方面做出了突出贡献。



中科院院士

著名数学家

袁亚湘

最优化研究现状

最优化问题有很多研究分支，如多目标规划、向量优化、整数规划、组合优化、随机优化、变分不等式和互补问题等。它们在上世纪后半叶得到了快速发展，成为相对成熟和相对独立的研究分支。

近年来，锥优化、鲁棒优化、半定规划、随机优化、统计优化、稀疏优化、张量和多项式优化、机器学习中的最优化等成为最优化问题的研究热点。

最优化学科归属

最优化是运筹学的基础学科和重要分支，也是应用数学的一个重要研究分支。

最优化与统计和模拟仿真一起构成其三大基本方法和技术。

最优化广泛应用于经济、科学和工程等领域中。

五、机遇与挑战

1、压缩感知优化：华裔天才、著名数学家、菲尔兹奖得主陶哲轩与斯坦福大学教授Candès、美国科学院院士Donoho独立提出了压缩感知（Compressed Sensing)理论，建立了稀疏优化模型。

Terence Tao
陶泽轩

Emmanuel Candès

David Donoho

稀疏优化

$$Ax = b, \qquad A \in R^{m\times n}, b \in R^{m}, m << n$$

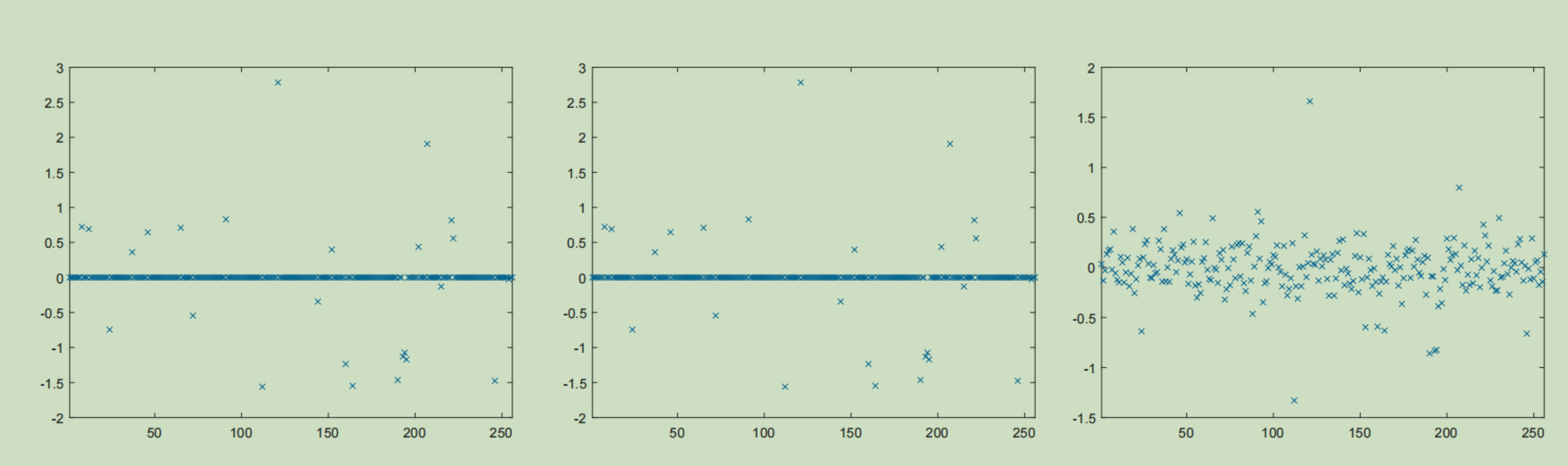
- 寻找最稀疏的解问题：压缩感知

$$\min \|x\|_0$$
$$\text{s.t. } Ax = b$$

菲尔兹奖得主陶哲轩和美国科学院院士**Candes**，以及美国科学院院士**Donoho**分别提出压缩感知理论：如果原始的高维信号具有稀疏性，则可以通过少量的观测信息得以恢复。被评为2007年度美国十大科技进展之一。

[1] Candès E J, Tao T. Decoding by linear programming [J]. *IEEE Transactions on Information Theory*, 2005, **51**(12): 4203-4215.
[2] Candès E J. Compressive sampling [C]//*Proceedings of the International Congress of Mathematicians*, 2006.
[3] Candès E J, Romberg J, Tao T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information [J]. *IEEE Transactions on Information Theory*, 2006, **52**(2): 489-509.
[4] Donoho D L. Compressed sensing [J]. *IEEE Transactions on Information Theory*, 2006, **52**(4): 1289-1306.

压缩感知理论的诞生，打破了香浓定律对采样频率的限制，实现了以较少的采样资源、较高的采样速度和较低的软硬件复杂度来获取原始信号，为高维数据处理提供了新思路，在信息工程领域产生了革命性的巨大影响，带动了大量跨学科领域工作的发展。

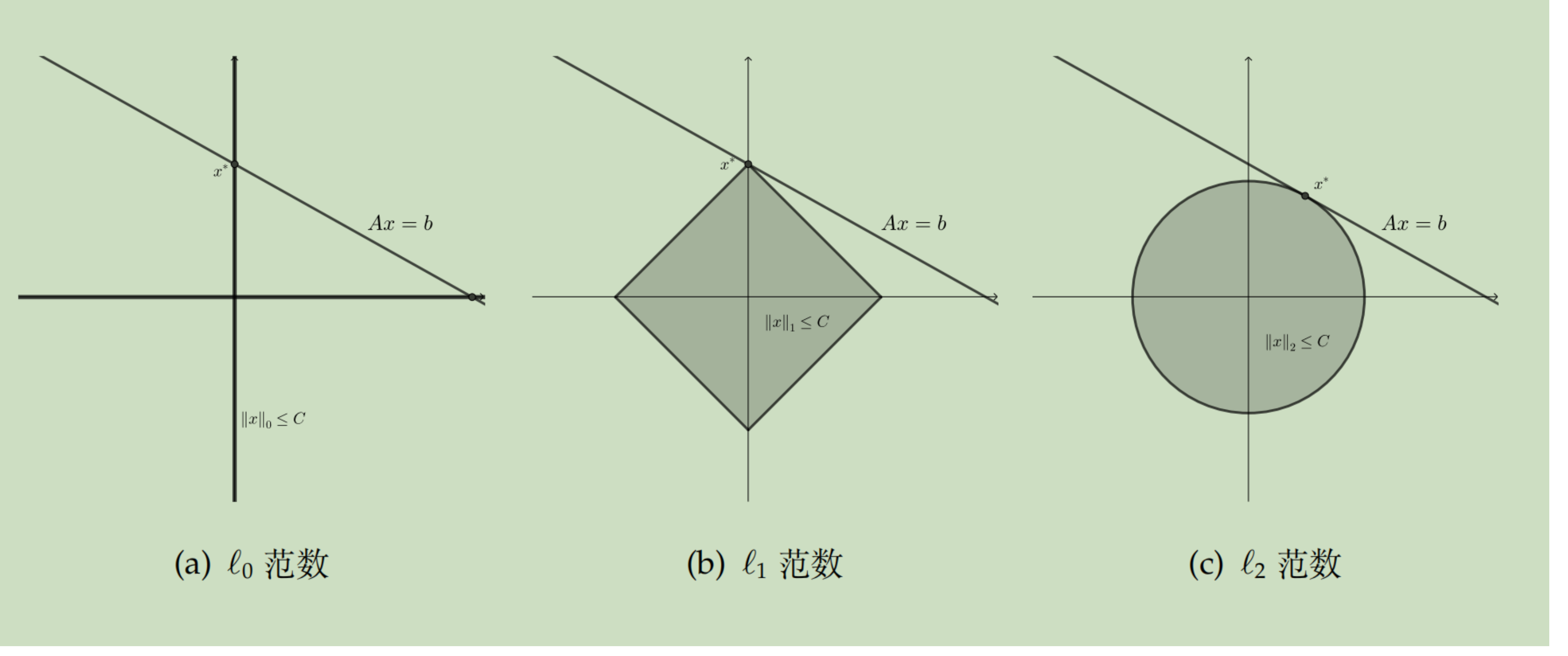


$$\min \|x\|_0 \quad \text{s.t. } Ax = b$$

$$\min \|x\|_1 \quad \text{s.t. } Ax = b$$

$$\min \|x\|_2 \quad \text{s.t. } Ax = b$$

三种范数优化问题求解示意图



$$\min \|x\|_0 \quad \text{s.t. } Ax = b$$

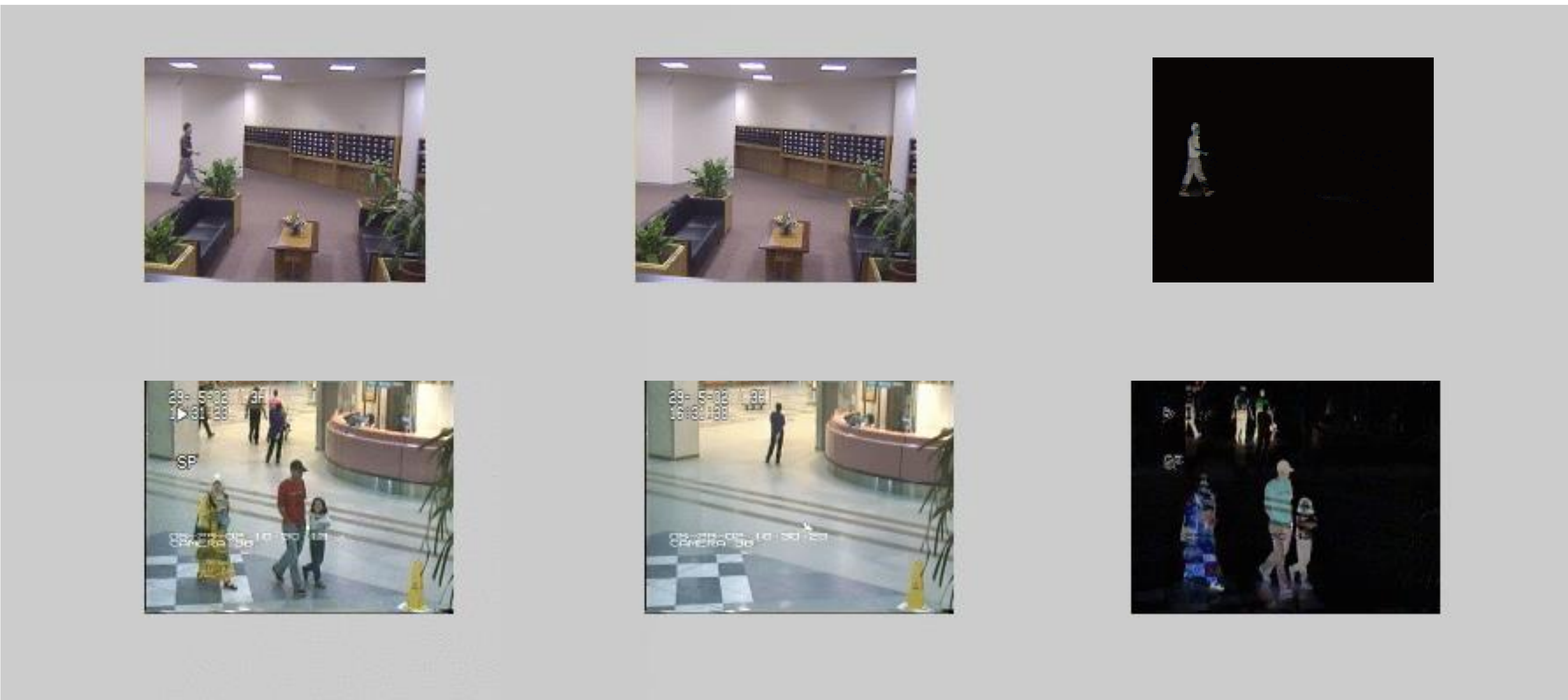
$$\min \|x\|_1 \quad \text{s.t. } Ax = b$$

$$\min \|x\|_2 \quad \text{s.t. } Ax = b$$

- **基于低秩矩阵分解的视频监控**

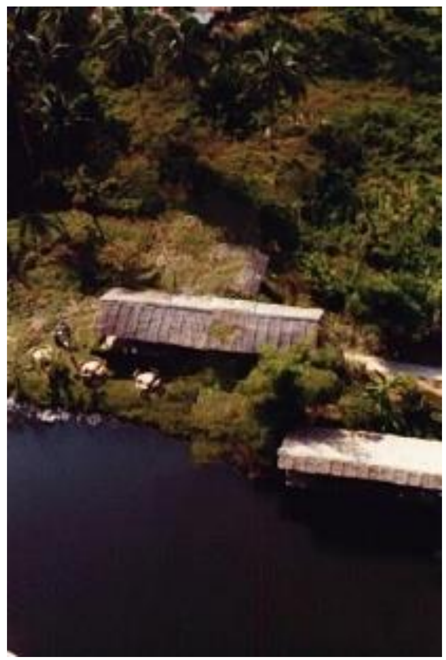
$$\min_{A,E} \|A\|_{1/2*}^{1/2} + \lambda \|E\|_1$$

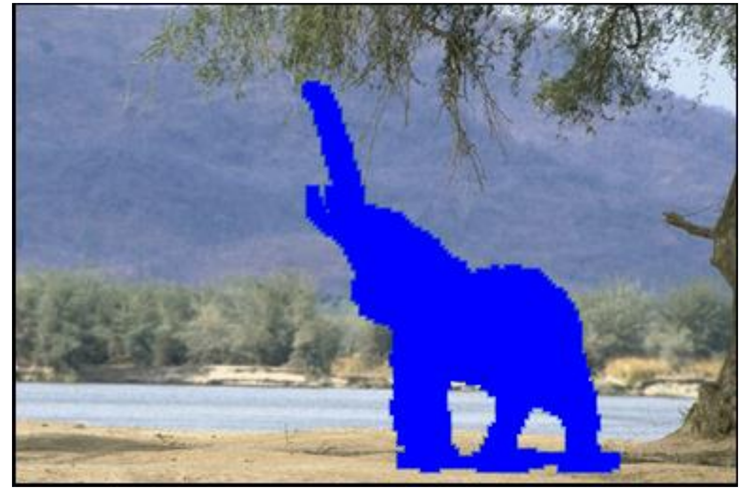
$$s.t. \quad A + E = D$$



- **基于稀疏正则化的图像填充**

$$\min\{\| P\sum_{n=1}^{N}\alpha_n B_n - P\Psi_p \|_2^2 + \lambda \| \vec{a} \|_{1/2}^{1/2}\}$$





• 基于低秩矩阵完整化的推荐系统

$$\min rank(X)$$
$$s.t.\ X_{ij} = A_{ij},\ (i, j)\ in\ \Omega$$

$\longrightarrow$

$$\min \| X \|_*$$
$$s.t.\ X_{ij} = A_{ij},\ (i, j)\ in\ \Omega$$

where $\| X \|_*$ is the nuclear norm

电影打分

观众	A	B	C	D	E	F
Movie 1	?	?	4	?	1	?
Movie 2	2	5	?	?	?	2
Movie 3	?	?	5	3	3	5
Movie 4	1	?	?	4	?	?
Movie 5	3	2	?	2	?	?

推荐系统

- 猜出“？”所对应的数值，就是推荐系统要做的事情。如果我们猜到了这个人对某个产品的打分，我们就知道这个人对这个产品的喜好程度（虽然他并没有进行购买）。如果评分好，就可以推荐给他，他就很有可能去购买这个产品，从而提高公司的盈利。
- 现在推荐系统应用得非常广泛，在国外，比如亚马逊、Google、YouTube都有。国内的很多购物平台、视频网站、门户网站，都会在不同的地方推荐一些商品。这都是推荐系统应用的体现。

Netflix 百万美元奖

Netflix 问题

从1998年10月到2005年12月 搜集数据
（请客户给电影打分）

- 480, 189 用户
- 17, 770 电影
- 100,480,507 分数 （1到5）

问题： 如何填补没有打分的数据？


NETFLIX
2009
DATE 09.21.09
PAY TO THE ORDER OF BellKor's Pragmatic Chaos
$ 1,000,000 00
AMOUNT ONE MILLION 00/100
FOR The Netflix Prize
Reed Hastings

• 鲁棒张量填充问题

$$\min_{\mathcal{L},\mathcal{E}} Q(\mathcal{L},\mathcal{E}) := \|\mathcal{L}\|_{\mathrm{TNN}} + \lambda(\|\mathcal{P}_\Omega(\mathcal{E})\|_1 - \Psi_\gamma(\mathcal{P}_\Omega(\mathcal{E}))) + \frac{\mu}{2}\|\mathcal{L}+\mathcal{E}-\overline{\mathcal{X}}\|^2, \quad (3)$$

Table 3: PSNR(dB) values for restoring results of different methods for color image corrupted by sample losing and salt-and-pepper noise.

sample ratios	noise level	Pepper				Lena				House			
		RTCℓ_1	W-ST	Ho-S	NCRTC	RTCℓ_1	W-ST	Ho-S	NCRTC	RTCℓ_1	W-ST	Ho-S	NCRTC
80%	20%	29.08	27.48	28.43	**34.7**	29.5	25.69	28.26	**34.26**	30.08	25.68	28.04	**36.63**
	40%	23.56	22.12	23.22	**27.48**	24.96	21.46	24.12	**28.46**	24.6	21.18	23.42	**29.5**
	60%	18.89	17.49	18.63	**20.58**	21.09	17.89	20.19	**22.92**	19.99	17.35	19.21	**22.14**
70%	20%	27.85	26.75	27.43	**32.73**	28.43	24.53	27.41	**32.53**	28.86	25.06	27.09	**34.4**
	40%	22.71	21.6	22.54	**26.02**	24.3	21.05	23.52	**27.3**	23.7	20.69	22.7	**28.26**
	60%	18.25	16.96	18.04	**19.77**	20.56	17.54	19.76	**22.38**	19.39	16.65	18.54	**21.36**
60%	20%	26.56	25.93	26.27	**30.82**	27.34	24.53	26.5	**30.79**	27.57	24.36	26.09	**32.22**
	40%	21.82	21.04	21.72	**24.77**	23.57	20.6	22.82	**26.19**	22.75	20.14	22.01	**26.96**
	60%	17.79	16.35	17.3	**18.91**	19.93	17.09	19.13	**21.75**	18.79	16.07	17.82	**20.73**
50%	20%	25.15	24.95	24.95	**28.69**	26.2	23.84	25.5	**29.38**	26.14	23.53	24.91	**30.34**
	40%	20.86	20.32	20.77	**23.39**	22.75	20.03	21.97	**25.07**	21.84	19.52	21.26	**25.28**
	60%	16.79	15.59	16.29	**17.87**	19.18	16.56	18.35	**21**	18.05	15.41	16.98	**19.79**
40%	20%	23.61	23.42	23.48	**26.43**	24.97	22.97	24.32	**27.57**	24.61	22.58	23.7	**28.19**
	40%	19.75	19.39	19.7	**21.89**	21.82	19.38	21.05	**23.9**	20.86	18.77	20.28	**23.37**
	60%	15.87	14.51	15.24	**16.38**	18.26	15.96	17.52	**19.94**	17.2	14.7	15.69	**18.63**

(a) PSNR: 9.91 (b) PSNR: 8.94 (c) PSNR: 8.14 (d) PSNR: 7.46 (e) PSNR: 6.87

(f) PSNR: 27.48 (g) PSNR: 24.1 (h) PSNR:22.12 (i) PSNR: 19.67 (j) PSNR: 17.49

(k) PSNR:28.43 (l) PSNR:25.57 (m) PSNR:23.2 (n) PSNR:21.06 (o) PSNR:18.63

(p) PSNR:29.08 (q) PSNR: 26.09 (r) PSNR:23.56 (s) PSNR: 21.25 (t) PSNR: 18.89

(u) PSNR:34.7 (v) PSNR:31.04 (w) PSNR:27.48 (x) PSNR:23.93 (y) PSNR:20.58

2、大数据优化：大数据分析主要从海量和复杂数据中挖掘有用信息并发现其规律，以期发挥数据资源的最大效益。它是重要的最优化模型。

大数据优化问题规模大、决策变量多、载息量丰富、约束条件复杂、甚至没有清晰的目标函数、问题病态等，传统的优化理论和方法无法在合理时间内给出问题的最优解或近似解。

国际最优化协会前主席Wright、SIAM J Optim主编Necedal：大数据优化是未来的一个重要发展方向。

3、人工智能 近年来，人工智能得到快速发展和广泛应用，如AlphaGo、李世石和柯洁的围棋大战，无人驾驶等，最近很火的ChatGPT，Deepseek。

美国工程院院士和科学院院士Michael Jordan 和机器学习大师Tom Mitchell 认为，机器学习是人工智能和数据科学的核心。

4、机器学习、深度学习

机器学习目前存在很多局限性，如算法多是针对单一数据结构且不能保证全局最优，并在处理超大规模问题时往往需要超长的计算机运行时间。

数学大师、菲尔兹奖获得者丘成桐："现代以神经网络为代表的统计方法及机器学习在工程实践中取得了很大的成功，但其理论基础非常薄弱；人工智能需要一个可以被证明的理论作为基础。"

总 结

- 我们赶上好时代：好机遇，大挑战
- 优化问题越来越重要，越来越难
- 学习：理论和计算编程都要好，

  1) 上机：周二下午5-6，数学院一楼机房

  （第3、4、5、6、7、8、12、13周）

  2) 课堂学习和课后作业练习；

  3) 自我拓广学习，提高学习兴趣和综合应用能力。

范 数

按某种规则(或映射)

（一） 向量的范数

定义3.3.1 对于n维向量空间R^n中任意一个向量x，若存在惟一一个实数$\|x\| \in R$与x对应，且满足

(1) (正定性) $\|x\| \geq 0$, 且$\forall x \in R^n$, $\|x\| = 0 \Leftrightarrow x = 0$;

(2) (齐次性) $\|\alpha x\| = |\alpha| \cdot \|x\|$，$\forall x \in R^n, \alpha \in R$;

(3) (三角不等式) $\|x + y\| \leq \|x\| + \|y\|$，$\forall x, y \in R^n$.

则称$\|x\|$为向量x的范数(或模或长度).

对于复线性空间C^n中的向量范数可以类似定义

由(3)可推出不等式: $\big|\|x\| - \|y\|\big| \leq \|x - y\|$.

在向量空间$R^n(C^n)$中，设$x=(x_1,x_2,\cdots,x_n)^T$

常用的向量x的范数有：

$\|x\|_2=\sqrt{(x,x)}=\sqrt{x^{\mathrm{T}}\cdot x}$

$$\|x\|_2=(|x_1|^2+|x_2|^2+\cdots+|x_n|^2)^{1/2} \quad \text{--------(1)}$$

x的$2-$范数或欧氏范数

$$\|x\|_1=|x_1|+|x_2|+\cdots+|x_n| \quad \text{--------(2)}$$

x的$1-$范数或平均范数

$$\|x\|_\infty=\max_{1\le i\le n}|x_i| \quad \text{--------(3)}$$

x的$\infty-$范数或最大范数或切比雪夫范数

$$\|x\|_p=(|x_1|^p+|x_2|^p+\cdots+|x_n|^p)^{1/p} \quad \text{--------(4)}$$

x的$p-$范数，$p\ge 1$

显然由于$\|x\|_1$和$\|x\|_2$是$\|x\|_p$当$p=1$和$p=2$时的特例，并且由于

$$\max_{1\le i\le n}|x_i| \le (|x_1|^p+|x_2|^p+\cdots+|x_n|^p)^{1/p} \le (n\max_{1\le i\le n}|x_i|^p)^{1/p}$$
$$= n^{1/p}\max_{1\le i\le n}|x_i| \to \max_{1\le i\le n}|x_i| \quad (p\to\infty)$$

$\|x\|_p \to \|x\|_\infty \ (p\to\infty$时$)$，所以$\|x\|_\infty$也是$\|x\|_p$的特例

且 $\frac{1}{n}\|x\|_1 \le \|x\|_\infty \le \|x\|_2 \le \|x\|_1$

例3.3.1 求下列向量的各种常用范数

$$x=(1,4,3,-1)^T$$

解

$$\|x\|_1 = |x_1|+|x_2|+\cdots+|x_4| = 9$$

$$\|x\|_2 = (|x_1|^2+|x_2|^2+\cdots+|x_4|^2)^{1/2} = \sqrt{27} = 3\sqrt{3}$$

$$\|x\|_\infty = \max_{1\le i\le 4}|x_i| = 4$$

向量范数是其分量的连续函数，即有下述定理：

定理3.3.1（向量范数连续性定理）

设$\|\cdot\|$是向量空间R^n中的一种范数，则$\|x\|$是关于x的分量$x_1, x_2, \ldots, x_n$的连续函数。

定理3.3.1（**有限维向量空间的范数等价性定理**）

设$\|\cdot\|$与$\|\cdot\|'$为向量空间R^n中的两种范数，则存在正常数c和C，使得下面不等式成立：

$$c\|x\|' \leq \|x\| \leq C\|x\|', \qquad \forall\, x \in R^n.$$

❖ 向量序列的收敛性

设 $x \in R^n$，$\tilde{x}$为其近似向量，则

$$\|\tilde{x}-x\|, \qquad \frac{\|\tilde{x}-x\|}{\|x\|}$$

分别称为$\tilde{x}$的绝对误差和相对误差。

定义3.3.2 如果向量序列$\{x^{(k)}\} \subset \mathrm{R}^n$和向量 $x \in \mathrm{R}^n$满足

$$\lim_{k\to\infty} x_j^{(k)} = x_j, \quad j = 1, 2, \ldots, n,$$

则称向量序列$\{x^{(k)}\}$收敛于向量 x，记为

$$\lim_{k\to\infty} x^{(k)} = x \quad \text{或} \quad x^{(k)} \to x \ (k \to \infty).$$

显然，R^n 中向量序列$\{x^{(k)}\}$收敛于 x 的充分必要条件是

$$\lim_{k\to\infty}\left\|x^{(k)}-x\right\|_\infty=0.$$

由向量范数的等价性定理可得到结论：如果在一种范数意义下向量序列收敛时，则在任何一种范数意义下向量序列亦收敛，即

$$\lim_{k\to\infty}x^{(k)}=x \Longleftrightarrow \lim_{k\to\infty}\left\|x^{(k)}-x\right\|=0,$$

其中$\|\cdot\|$为向量的任意一种范数。

（二） 矩阵的范数

定义3.3.3 对于空间$R^{n\times n}$中任意一个矩阵A,

若存在惟一一个实数$\|A\|\in R$与A对应，且满足

(1) (正定性) $\|A\|\geq 0$, 且$\forall A\in R^{n\times n}$, $\|A\|=0 \Leftrightarrow A=0$;

(2) (齐次性) $\|\alpha A\|=|\alpha|\cdot\|A\|$， $\forall A\in R^{n\times n}, \alpha\in R$;

(3) (三角不等式) $\|A+B\|\leq\|A\|+\|B\|$， $\forall A,B\in R^{n\times n}$.

(4) $\|AB\|\leq\|A\|\cdot\|B\|$， $\forall A,B\in R^{n\times n}$.

则称$\|A\|$为矩阵A的范数.

例3.3.2 设n阶方阵$A=(a_{ij})_{n\times n}$ 类似向量的 2-范数

设 $\|A\|_F=\left(\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}^2\right)^{1/2}$ --------(5)

不难验证其满足定义3.3.3的4个条件

因此 $\|A\|_F$ 是一种矩阵范数

称为Frobenius范数，简称F-范数

而且可以验证 $\|A\|_F=\left(tr(A^TA)\right)^{1/2}=\left(tr(AA^T)\right)^{1/2}$

tr为矩阵的迹 --------(6)

设$x \in R^n, A \in R^{n\times n}, \|\cdot\|_p$为一种向量范数

则$\dfrac{\|Ax\|_p}{\|x\|_p}$对所有的$x \neq 0$有最大值,令

$$\|A\|_p = \max_{x\neq 0}\frac{\|Ax\|_p}{\|x\|_p} = \max_{\|x\|_p=1}\|Ax\|_p . \quad \text{---------(7)}$$

可以验证$\|A\|_P$满足定义3.3.3的4个条件

定义3.3.4 由(7)式确定的$\|A\|_p$称为从属于给定向量范数$\|x\|_p$的矩阵范数。

简称为从属范数或导出范数或算子p-范数

由此可知，给出一种向量范数，就有一种相应的矩阵范数

显然，由定义不难推出

$$\|Ax\|_p \le \|A\|_p \|x\|_p \qquad \text{--------(8)}$$

定义3.3.5 对于给定的向量范数$\|\cdot\|_\upsilon$和矩阵范数$\|\cdot\|_\mu$，若$\forall x \in R^n, A \in R^{n\times n}$，都有

$$\|Ax\|_\upsilon \le \|A\|_\mu \|x\|_\upsilon \qquad \text{--------(9)}$$

则称所给的向量范数$\|\cdot\|_\upsilon$和矩阵范数$\|\cdot\|_\mu$相容.

由(8)式,可知算子范数和其对应的向量范数是相容的

根据向量的常用范数可以得到常用的矩阵算子范数

$$(1)\quad \|A\|_1 = \max_{x\neq 0}\left(\|Ax\|_1 \big/ \|x\|_1 \right) = \max_{1\le j\le n}\sum_{i=1}^{n}|a_{ij}| \qquad \text{--------}(10)$$

A的每列绝对值之和的最大值，称为A的列范数

$$(2)\quad \|A\|_\infty = \max_{x\neq 0}\left(\|Ax\|_\infty \big/ \|x\|_\infty \right) = \max_{1\le i\le n}\sum_{j=1}^{n}|a_{ij}| \qquad \text{--------}(11)$$

A的每行绝对值之和的最大值，称为A的行范数

$$(3)\quad \|A\|_2 = \max_{x\neq 0}\left(\|Ax\|_2 \big/ \|x\|_2 \right) = \sqrt{\lambda_{\max}(A^T A)} \qquad \text{--------}(12)$$

$\lambda_{\max}(A^T A)$为$A^T A$的特征值的绝对值的最大值，称为A的$2-$范数

$\|A\|_F$ 是不从属于任意向量范数 $\|\cdot\|_v$ 的矩阵范数

故而矩阵范数和算子范数并不完全是一回事

不过

$$\|A\|_F = \left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^2 \right)^{1/2} = \left(tr(A^T A) \right)^{1/2} = \left(tr(AA^T) \right)^{1/2}$$

$$\|A\|_2 = \sqrt{\lambda_{\max}(A^T A)} \leq \|A\|_F$$

$$\|Ax\|_2 \leq \|A\|_2 \|x\|_2 \leq \|A\|_F \|x\|_2$$

因此 $\|A\|_F$ 与 $\|x\|_2$ 相容

例3.3.4 求矩阵A的各种常用范数

$$A=\begin{array}{cc}\begin{pmatrix}1 & 2 & 0\\ -1 & 2 & -1\\ 0 & 1 & 1\end{pmatrix} & \begin{array}{c}3\\4\\2\end{array}\\ \begin{array}{ccc}2 & 5 & 2\end{array} & \end{array}$$

解

$$\|A\|_1=\max_{1\le j\le n}\sum_{i=1}^{n}|a_{ij}|=\max_{1\le j\le n}\{2,5,2\}=5$$

$$\|A\|_\infty=\max_{1\le i\le n}\sum_{j=1}^{n}|a_{ij}|=\max_{1\le i\le n}\{3,4,2\}=4$$

由于 $\|A\|_2=\sqrt{\lambda_{\max}(A^TA)}$

因此先求$A^T A$的特征值

$$A^T A = \begin{pmatrix} 1 & -1 & 0 \\ 2 & 2 & 1 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ -1 & 2 & -1 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 9 & -1 \\ 1 & -1 & 2 \end{pmatrix}$$

特征方程为

$$\det(\lambda I - A^T A) = \begin{pmatrix} \lambda-2 & 0 & -1 \\ 0 & \lambda-9 & 1 \\ -1 & 1 & \lambda-2 \end{pmatrix} = 0$$

可得$A^T A$的特征值为

$$\lambda_1 = 9.1428, \lambda_2 = 2.9211, \lambda_3 = 0.9361$$

$$\lambda_{\max}(A^T A) = 9.1428$$

$$\|A\|_2 = \sqrt{\lambda_{\max}(A^T A)} = 3.0237$$

$$\|A\|_F = \sqrt{tr(A^T A)} = \sqrt{2+9+2} = 3.6056$$

$\|A\|_1$ $\|A\|_\infty$ $\|A\|_2$ $\|A\|_F$

容易计算

计算较复杂

不是从属范数

对矩阵元素的变化比较敏感

性质较好

(理论上)使用最广泛

❖ 矩阵范数的等价性定理

定理3.3.3 设$\|\cdot\|$与$\|\cdot\|'$为$R^{n\times n}$中的两种范数，则存在正常数c和C，使得下面不等式成立：

$$c\|A\|' \le \|A\| \le C\|A\|', \qquad \forall\ A\in R^{n\times n}.$$

可以验证：

4种范数相互等价

（1）$\|A\|_2 \le \|A\|_F \le n^{1/2}\|A\|_2$;

（2）$n^{-1/2}\|A\|_\infty \le \|A\|_2 \le n^{1/2}\|A\|_\infty$;

（3）$n^{-1/2}\|A\|_1 \le \|A\|_2 \le n\|A\|_1$

❖ 矩阵序列的收敛性

设 $A\in R^{n\times n}$，$\tilde{A}$为其近似矩阵，则

$$\|\tilde{A}-A\|,\qquad \frac{\|\tilde{A}-A\|}{\|A\|}$$

分别称为$\tilde{A}$的绝对误差和相对误差。

定义3.3.6 如果n阶矩阵序列$\{A^{(k)}\}\subset \mathrm{R}^{n\times n}$和矩阵$A\in \mathrm{R}^{n\times n}$ 满足（其中$A^{(k)}=(a_{ij}{}^{(k)})_{n\times n}$，$A=(a_{ij})_{n\times n}$）

$$\lim_{k\to\infty} a_{ij}^{(k)}=a_{ij},\quad i,j=1,2,\ldots,n,$$

则称矩阵序列$\{A^{(k)}\}$收敛于矩阵 A，记为

$$\lim_{k\to\infty} A^{(k)}=A,\quad 或\quad A^{(k)}\to A\,(k\to\infty).$$

显然，$R^{n\times n}$ 中矩阵序列$\{A^{(k)}\}$ 收敛于矩阵A 的充分必要条件是

$$\boxed{\lim_{k\to\infty} A^{(k)} = A} \Longleftrightarrow \boxed{\lim_{k\to\infty}\left\|A^{(k)} - A\right\| = 0,}$$

其中$\|\cdot\|$为矩阵的任意一种范数。

摄动定理　设$A\in R^{n\times n}$可逆, 且$\left\|A^{-1}\right\|\le\alpha, C\in R^{n\times n}, \|A-C\|\le\beta,$ 且$\alpha\beta<1,$则C可逆， 且$\left\|C^{-1}\right\|\le\dfrac{\alpha}{1-\alpha\beta}$

小 结

- 最优化理论和方法课程简介
- 向量范数、矩阵范数、收敛性的刻画

参考复习：

《数值计算方法》（曾金平）第三章线性方程组的性态

《最优化方法》（袁亚湘）第一章