第三章 典型优化问题

修贤超

https://xianchaoxiu.github.io

目录

凸优化问题

- 标准形式的凸优化问题

$$
\begin{aligned}
\min_{x \in \mathbb{R}^n} \quad & f_0(x) \\
\text{s.t.} \quad & f_i(x) \le 0, \quad i = 1, \cdots, m \\
& a_i^\top x = b_i, \quad i = 1, \cdots, p
\end{aligned}
$$

  - $f_0, f_1, \ldots, f_m$ 为凸函数
  - $a_i^\top x = b_i$ 为线性等式约束

- 经常写成

$$
\begin{aligned}
\min_{x \in \mathbb{R}^n} \quad & f_0(x) \\
\text{s.t.} \quad & f_i(x) \le 0, \quad i = 1, \cdots, m \\
& Ax = b
\end{aligned}
$$

应用举例

- 考虑

$$
\begin{aligned}
\min_{x_1,x_2} \quad & f_0(x) = x_1^2 + x_2^2 \\
\text{s.t.} \quad & f_1(x) = x_1/(1+x_2^2) \le 0 \\
& h_1(x) = (x_1+x_2)^2 = 0
\end{aligned}
$$

  - f_0 为凸函数, 可行集 $\{(x_1,x_2) \mid x_1 = -x_2 \le 0\}$ 为凸集
  - f_1 非凸, h_1 不是线性函数

- 不是凸问题，但可转化为凸优化问题

$$
\begin{aligned}
\min_{x_1,x_2} \quad & x_1^2 + x_2^2 \\
\text{s.t.} \quad & x_1 \le 0 \\
& x_1 + x_2 = 0
\end{aligned}
$$

局部和全局极小

- **凸优化问题的任意局部极小点都是全局最优**

证明 设 x 是局部极小解，y 是全局最优解且 $f_0(y) < f_0(x)$. 存在 $R > 0$ 使

$$z\text{可行}, \quad \|z - x\|_2 \le R \quad \Rightarrow \quad f_0(z) \ge f_0(x)$$

考虑 $z = \theta y + (1 - \theta)x$ 且 $\theta = R/(2\|y - x\|_2)$

- $\|y - x\|_2 > R$, 则 $0 < \theta < 1/2$
- z 是两个可行点的凸组合，则也可行
- $\|z - x\|_2 = R/2$，并且

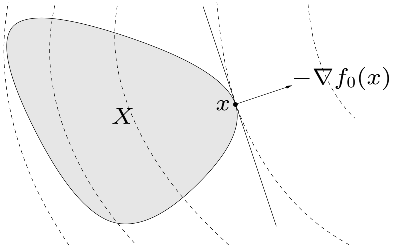
$$f_0(z) \le \theta f_0(x) + (1 - \theta) f_0(y) < f_0(x)$$

这与 x 是局部极小的假设矛盾

可微凸优化问题的最优性条件

- 设 x 是凸优化问题 $\min_{x \in X} f_0(x)$ 的最优解当且仅当 x 可行且满足

$$\nabla f_0(x)^\top (y - x) \geq 0, \quad \forall y \in X$$



- 如果 $\nabla f_0(x)$ 非零，它定义了可行集 X 在 x 处的支撑超平面

具体含义

- **无约束优化** x 是最优解当且仅当

$$x \in \mathrm{dom}\ f_0, \quad \nabla f_0(x) = 0$$

- **等式约束优化问题**

$$\min \quad f_0(x) \qquad \text{s.t.} \quad Ax = b$$

x 是最优解当且仅当存在 v 使得

$$x \in \mathrm{dom}\ f_0, \quad Ax = b, \quad \nabla f_0(x) + A^\top v = 0$$

- **非负约束优化问题**

$$\min \quad f_0(x) \qquad \text{s.t.} \quad x \geq 0$$

x 是最优解当且仅当

$$x \in \mathrm{dom}\ f_0, \quad x \geq 0, \quad \begin{cases} \nabla f_0(x)_i \geq 0, & x_i = 0 \\ \nabla f_0(x)_i = 0, & x_i > 0 \end{cases}$$

线性规划基本形式

- 线性规划问题的一般形式

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & c^\top x \\ \text{s.t.} \quad & Ax = b \\ & Gx \le e \end{aligned}$$

- 线性规划问题的标准形式

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & c^\top x \\ \text{s.t.} \quad & Ax = b \\ & x \ge 0 \end{aligned}$$

- 线性规划问题的不等式形式

$$\begin{aligned} \max_{y\in\mathbb{R}^n} \quad & b^\top y \\ \text{s.t.} \quad & A^\top y \le c \end{aligned}$$

应用举例: 基追踪问题

- 基追踪问题是压缩感知中的一个基本问题，可以写为

$$
\begin{aligned}
\min_{x\in\mathbb{R}^n} \quad & \|x\|_1 \\
\text{s.t.} \quad & Ax = b
\end{aligned}
$$

- 对每个 $|x_i|$ 引入一个新的变量 z_i，可以转化为

$$
\begin{aligned}
\min_{z\in\mathbb{R}^n} \quad & \sum_{i=1}^{n} z_i \\
\text{s.t.} \quad & Ax = b \\
& -z_i \le x_i \le z_i, \quad i = 1, \cdots, n
\end{aligned}
$$

应用举例: 数据拟合

- 最小 ℓ_∞ 范数模型

$$\min_{x\in\mathbb{R}^n} \quad \|Ax-b\|_\infty$$

- 令 $t=\|Ax-b\|_\infty$，得到等价问题

$$\begin{aligned}\min_{x\in\mathbb{R}^n,\, t\in\mathbb{R}} \quad & t \\ \text{s.t.} \quad & \|Ax-b\|_\infty \le t\end{aligned}$$

- 利用 ℓ_∞ 范数的定义，可以进一步写为

$$\begin{aligned}\min_{x\in\mathbb{R}^n,\, t\in\mathbb{R}} \quad & t \\ \text{s.t.} \quad & -t\mathbf{1} \le Ax-b \le t\mathbf{1}\end{aligned}$$

目录

最小二乘问题

- 最小二乘问题的一般形式如下

$$\min_{x \in \mathbb{R}^n} \quad \sum_{i=1}^{m} r_i^2(x)$$

- 如果所有的 $r_i : \mathbb{R}^n \to \mathbb{R}$ 都是线性函数，则称线性最小二乘问题，否则称为非线性最小二乘问题
- 如果噪声服从高斯分布，最小二乘问题的解对应于原问题的最大似然解
- 1801 年，24 岁的高斯计算出小行星的运动轨道

应用举例: 线性最小二乘问题

- 线性最小二乘问题是回归分析中的一个基本模型，它可以表示为

$$\min_{x\in\mathbb{R}^n} \quad \sum_{i=1}^{m}(a_i^\top x - b_i)^2$$

- 记 $A = [a_1, a_2, \cdots, a_m]^\top$，上式可以等价地写成

$$\min_{x\in\mathbb{R}^n} \quad f(x) = \frac{1}{2}\|Ax - b\|_2^2$$

- $x \in \mathbb{R}^n$ 为其全局极小解当且仅当 x 满足

$$\nabla f(x) = A^\top(Ax - b) = 0$$

应用举例: 数据插值

- 给定数据集 $\{a_i \in \mathbb{R}^p, b_i \in \mathbb{R}^q, i = 1, 2, \cdots, m\}$，**插值**是求一个映射 f，使得

$$b_i = f(a_i), \quad i = 1, 2, \cdots, m$$

- 利用线性函数 $f(a) = Xa + y$ 逼近，可以建立如下最小二乘问题

$$\min_{X \in \mathbb{R}^{q \times p}} \quad \sum_{i=1}^{m} \|Xa_i + y - b_i\|^2$$

- 设 $\{\phi_i(a)\}_{i=1}^n (n \leq m)$ 为插值空间的一组基，数据插值可以写成

$$b_j = f(a_j) = \sum_{i=1}^{n} x_i \phi_i(a_j), \quad j = 1, 2, \cdots, m$$

应用举例: 数据插值

- 设非线性向量函数 $\phi_i(\theta):\mathbb{R}^q \to \mathbb{R}^q$，并构造如下复合函数

$$f(\theta)=\phi_n(X_n\phi_{n-1}(X_{n-1}\cdots\phi_1(X_1\theta+y_1)\cdots+y_{n-1})+y_n)$$

- 常用的有 ReLU，即

$$\phi_i(\theta)=(\mathrm{ReLU}(\theta_1),\mathrm{ReLU}(\theta_2),\cdots,\mathrm{ReLU}(\theta_q))^\top,\quad i=1,2,\cdots,n$$

$$\mathrm{ReLU}(t)=\begin{cases}t, & t\geq 0\\ 0, & \text{其他}\end{cases}$$

- 更多未知的非线性，可能在更大的函数空间中得到一个更好的逼近

应用举例: 带微分方程约束优化问题

- 当约束中含微分方程时，称为带微分方程约束的优化问题
- 考虑瓦斯油催化裂解生成气体和其他副产物的反应过程

$$\begin{cases} \dot{y}_1 = -(\theta_1 + \theta_3) y_1^2 \\ \dot{y}_2 = \theta_1 y_1^2 - \theta_2 y_2 \end{cases}$$

- 转化为最小二乘问题

$$\begin{aligned} \min_{\theta \in \mathbb{R}^3} \quad & \sum_{j=1}^{n} \|y(\tau_j;\theta) - z_j\|^2 \\ \text{s.t.} \quad & y(\tau;\theta) \text{满足上述方程组} \end{aligned}$$

其中 z_j 是在时刻 τ_j 的 y 的测量值，n 为测量的时刻数量

目录

复合优化问题

- 复合优化问题一般可以表示为

$$\min_{x\in\mathbb{R}^n} \quad \psi(x) = f(x) + h(x)$$

  - $f(x)$ 是光滑函数, 如数据拟合项
  - $h(x)$ 可能是非光滑的, 如 ℓ_1 范数正则项， 约束集合的示性函数

- 常用的优化算法有
  - 次梯度法
  - 近似点梯度法
  - Nesterov 加速法
  - 交替方向乘子法

应用举例: 信号处理

- ℓ_1 范数正则化回归分析问题

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

- 矩阵分离问题

$$\begin{aligned} \min_{X,S\in\mathbb{R}^{m\times n}} \quad & \|X\|_*+\mu\|S\|_1 \\ \text{s.t.} \quad & X+S=M \end{aligned}$$

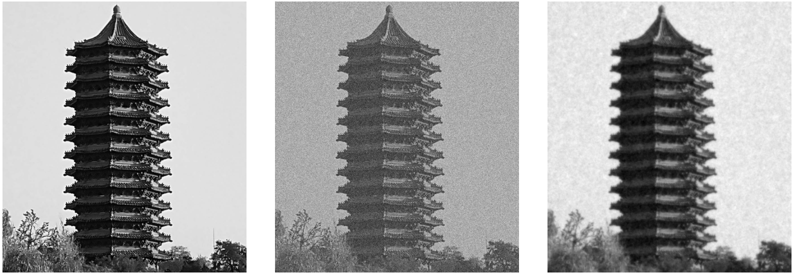
- 字典学习问题

$$\begin{aligned} \min_{X,D\in\mathbb{R}^{m\times n}} \quad & \frac{1}{2n}\|DX-A\|_F^2+\lambda\|X\|_1 \\ \text{s.t.} \quad & \|D\|_F\leqslant 1 \end{aligned}$$

应用举例：图像去噪

- 图像去噪是指从一个带噪声的图像中恢复出不带噪声的原图
- 由全变差模型，去噪问题可表示为

$$\min_{x\in\mathbb{R}^{n\times n}} \quad \frac{1}{2}\|x-y\|_F^2+\lambda\|x\|_{TV}$$



应用举例：盲反卷积

- 反卷积是从一个模糊的图像恢复出原来清晰的图像，也称为去模糊
- 反卷积问题的模型

$$y = a * x + \varepsilon$$

- 设噪声为高斯噪声，可转化为

$$\min_{a,x} \quad \frac{1}{2}\|y - a * x\|_2^2$$

- 设原始图像信号在小波变换下是稀疏的，进一步得到

$$\min_{a,x} \quad \frac{1}{2}\|y - a * x\|_2^2 + \|\lambda \odot (Wx)\|_1$$

其中 W 是小波框架，$\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_m)^\top$ 用来控制稀疏度

目录

随机优化问题

- 随机优化问题可以表示为

$$\min_{x\in\mathcal{X}} \quad \mathbb{E}_{\xi}[F(x,\xi)] + h(x)$$

  - $F(x,\xi)$ 表示样本 ξ 上的损失或奖励
  - $h(x)$ 用来保证解的某种性质

- 设有 N 个样本 $\xi_1,\xi_2,\cdots,\xi_N$，令 $f_i(x)=F(x,\xi_i)$，得到经验风险极小化问题

$$\min_{x\in\mathcal{X}} \quad f(x)=\frac{1}{N}\sum_{i=1}^{N} f_i(x) + h(x)$$

- 样本数 N 比较多，可行域所在空间维数 n 比较大，导致计算困难

应用举例: 随机主成分分析

- 如果样本点 ξ 服从某个零均值分布 $\mathcal{D}$，则随机主成分分析可以写成

$$\max_{X\in\mathbb{R}^{p\times d}} \quad \mathrm{Tr}(X^\top AA^\top X) \qquad \text{s.t.} \quad X^\top X = I$$

$$\Downarrow$$

$$\max_{X\in\mathbb{R}^{p\times d}} \quad \mathrm{Tr}(X^\top \mathbb{E}_{\xi\sim\mathcal{D}}[\xi\xi^\top]X) \qquad \text{s.t.} \quad X^\top X = I$$

$$\Downarrow$$

$$\max_{X\in\mathbb{R}^{p\times d}} \quad \frac{1}{N}\sum_{i=1}^{N}\mathrm{Tr}(X^\top A_iA_i^\top X) \qquad \text{s.t.} \quad X^\top X = I$$

应用举例: 分布式鲁棒优化

- 为了提高深度学习预测器的泛化能力，考虑

$$\min_{h} \quad \mathbb{E}_z[F(h,z)]$$

$$\Downarrow$$

$$\min_{h} \quad \max_{\hat{z}\in \Gamma} \mathbb{E}_{\hat{z}}[F(h,\hat{z})]$$

- 集合 Γ 中随机变量的分布与真实数据的分布在一定意义下非常接近
- Wasserstein 距离可以改变原来经验分布的支撑集

应用举例: 分布式鲁棒优化

Generative Adversarial Nets

Ian J. Goodfellow[*], **Jean Pouget-Abadie**[†], **Mehdi Mirza**, **Bing Xu**, **David Warde-Farley**, **Sherjil Ozair**[‡], **Aaron Courville**, **Yoshua Bengio**[§]

The adversarial modeling framework is most straightforward to apply when the models are both multilayer perceptrons. To learn the generator's distribution p_g over data $\boldsymbol{x}$, we define a prior on input noise variables $p_{\boldsymbol{z}}(\boldsymbol{z})$, then represent a mapping to data space as $G(\boldsymbol{z};\theta_g)$, where G is a differentiable function represented by a multilayer perceptron with parameters θ_g. We also define a second multilayer perceptron $D(\boldsymbol{x};\theta_d)$ that outputs a single scalar. $D(\boldsymbol{x})$ represents the probability that $\boldsymbol{x}$ came from the data rather than p_g. We train D to maximize the probability of assigning the correct label to both training examples and samples from G. We simultaneously train G to minimize $\log(1 - D(G(\boldsymbol{z})))$. In other words, D and G play the following two-player minimax game with value function $V(G, D)$:

$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))]. \tag{1}$$

目录

半定规划

- 半定规划（SDP）是线性规划在矩阵空间中的一种推广
- 半定规划问题的标准形式

$$
\begin{aligned}
\min \quad & \langle C, X \rangle \\
\text{s.t.} \quad & \langle A_1, X \rangle = b_1 \\
& \cdots \\
& \langle A_m, X \rangle = b_m \\
& X \succeq 0
\end{aligned}
$$

- 对偶形式

$$
\begin{aligned}
\min \quad & -b^\top y \\
\text{s.t.} \quad & y_1 A_1 + y_2 A_2 + \cdots + y_n A_n \preceq C
\end{aligned}
$$

LP, SOCP 与 SDP 的比较

- LP 与 SDP

LP
$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & Ax \le b \end{aligned}$$

SDP
$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & \operatorname{diag}(Ax - b) \preceq 0 \end{aligned}$$

- SOCP 与 SDP

SOCP
$$\begin{aligned} \min \quad & f^\top x \\ \text{s.t.} \quad & \|A_i x + b_i\|_2 \le c_i^\top x + d_i, \quad i = 1, \cdots, m \end{aligned}$$

SDP
$$\begin{aligned} \min \quad & f^\top x \\ \text{s.t.} \quad & \begin{bmatrix} (c_i^\top x + d_i) I & A_i x + b_i \\ (A_i x + b_i)^\top & c_i^\top x + d_i \end{bmatrix} \succeq 0, \quad i = 1, \cdots, m \end{aligned}$$

应用举例: 二次约束二次规划问题的半定规划松弛

- 设A_i 为 $n \times n$ 对称矩阵, 考虑二次约束二次规划问题

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & x^\top A_0 x + 2b_0^\top x + c_0 \\ \text{s.t.} \quad & x^\top A_i x + 2b_i^\top x + c_i \le 0, \quad i = 1, 2, \cdots, m \end{aligned}$$

- 对任意 $x \in \mathbb{R}^n$ 以及 $A \in \mathcal{S}^n$，有恒等式

$$x^\top A x = \mathrm{Tr}(x^\top A x) = \mathrm{Tr}(A x x^\top) = \langle A, x x^\top \rangle$$

- 原始问题等价于

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & \langle A_0, X \rangle + 2b_0^\top x + c_0 \\ \text{s.t.} \quad & \langle A_i, X \rangle + 2b_i^\top x + c_i \le 0, \quad i = 1, 2, \cdots, m \\ & X = x x^\top \end{aligned}$$

应用举例：二次约束二次规划问题的半定规划松弛

- 进一步地

$$\begin{aligned} x^\top A_i x + 2b_i^\top x + c_i &= \left\langle \begin{pmatrix} A_i & b_i \\ b_i^\top & c_i \end{pmatrix}, \begin{pmatrix} X & x \\ x^\top & 1 \end{pmatrix} \right\rangle \\ &= \langle \overline{A_i}, \overline{X} \rangle, \quad i = 0, 1, \cdots, m \end{aligned}$$

- 半定规划松弛可以写成

$$\begin{aligned} \min \quad & \langle \overline{A_0}, \overline{X} \rangle \\ \text{s.t.} \quad & \langle \overline{A_i}, \overline{X} \rangle \leq 0, \quad i = 1, 2, \cdots, m \\ & \overline{X} \succeq 0 \\ & \overline{X}_{n+1,n+1} = 1 \end{aligned}$$

- 约束 $X = xx^\top$ 松弛成半正定约束 $X \succeq xx^\top$（等价于 $\overline{X} \succeq 0$）

应用举例：最大割问题的半定规划松弛

- 最大割问题是找到节点集合 V 的一个子集 S 使得 S 与它的补集 $\overline{S}$ 之间相连边的权重之和最大化
- 令 $x_j = 1,\ j \in S$ 和 $x_j = -1,\ j \in \overline{S}$，则

$$\begin{aligned} \max \quad & \frac{1}{2}\sum_{i<j}(1 - x_i x_j) w_{ij} \\ \text{s.t.} \quad & x_j \in \{-1, 1\},\ j = 1, 2, \cdots, n \end{aligned}$$

- 只有当 x_i 与 x_j 不同时，目标函数中 w_{ij} 的系数非零
- 最大割问题是一个离散优化问题，很难在多项式时间内找到最优解

应用举例: 最大割问题的半定规划松弛

- 令 $W = (w_{ij}) \in \mathcal{S}^n$，并定义 $C = -\frac{1}{4}(\mathrm{diag}(W\mathbf{1}) - W)$，得到

$$\begin{aligned} \min \quad & x^\top C x \\ \text{s.t.} \quad & x_i^2 = 1,\ i = 1, 2, \cdots, n \end{aligned}$$

- 令 $X = xx^\top$，则最大割问题可以转化为

$$\begin{aligned} \min \quad & \langle C, X \rangle \\ \text{s.t.} \quad & X_{ii} = 1,\ i = 1, 2, \cdots, n \\ & X \succeq 0,\ \mathrm{rank}(X) = 1 \end{aligned}$$

  - $x_i^2 = 1$ 意味着矩阵 X 对角线元素 $X_{ii} = 1$
  - $X = xx^\top$ 可以用约束 $X \succeq 0$ 和 $\mathrm{rank}(X) = 1$ 等价刻画

应用举例: 极小化最大特征值

- 极小化最大特征值问题可表示为

$$\min \quad \lambda_{\max}(A_0 + \sum_i x_i A_i)$$

- 由于 $\lambda_{\max}(A) \le t \Leftrightarrow A \preceq tI$，则极小化最大特征值可以转化为

SDP 形式

$$\begin{aligned} \min \quad & z \\ \text{s.t.} \quad & zI - \sum_i x_i A_i \succeq A_0 \end{aligned}$$

对偶问题形式

$$\begin{aligned} \max \quad & \langle A_0, Y \rangle \\ \text{s.t.} \quad & \langle A_i, Y \rangle = 0 \\ & \langle I, Y \rangle = 1 \\ & Y \succeq 0 \end{aligned}$$

应用举例: 极小化二范数问题

- 令 $A_i \in \mathbb{R}^{m\times n}$, 极小化 $A(x) = A_0 + \sum_i x_i A_i$ 的二范数

$$\min_{x} \quad \|A(x)\|_2$$

- SDP 形式

$$\begin{aligned} \min_{x,t} \quad & t \\ \text{s.t.} \quad & \begin{pmatrix} tI & A(x) \\ A(x)^\top & tI \end{pmatrix} \succeq 0 \end{aligned}$$

- 约束形式来源于

$$\begin{aligned} \|A\|_2 \le t \Leftrightarrow \quad & A^\top A \preceq t^2 I, \quad t \ge 0 \\ \Leftrightarrow \quad & \begin{pmatrix} tI & A(x) \\ A(x)^\top & tI \end{pmatrix} \succeq 0 \end{aligned}$$

目录

矩阵优化的基本形式

- 矩阵优化问题的形式

$$\min_{X \in \mathcal{X}} \quad \psi(X)$$

  - $\mathcal{X}$ 为特定的矩阵空间
  - $\psi(X) : \mathcal{X} \to \mathbb{R}$ 为给定的函数，可能是非光滑的
- 和向量相比，矩阵有许多新的性质, 如秩、特征值等
- 广泛地出现在组合数学、材料科学、机器学习和统计学等

矩阵优化的基本形式

- 低秩矩阵恢复问题

$$\min_{X\in\mathbb{R}^{m\times n}} \quad \frac{1}{2}\|X-M\|_F^2+\mu\|X\|_*$$

考虑函数 $h(X)=\|X\|_*$ 的次微分

$$\partial h(X)=\{UV^\top+W \mid \|W\|_2\le 1,\ U^\top W=0,\ WV=0\}$$

- 主成分分析问题

$$\min_{X\in\mathbb{R}^{p\times d}} \quad \psi(X)=-\mathrm{Tr}(X^\top AA^\top X) \qquad \text{s.t.} \quad X^\top X=I_d$$

考虑目标函数的微分

$$\nabla\psi(X)=-2AA^\top X$$

应用举例：非负矩阵分解

- 给定矩阵 $A = [a_1, a_2, \cdots, a_n] \in \mathbb{R}^{d\times n}$，将其分解成非负基矩阵 $X \in \mathbb{R}^{d\times p}$ 和非负系数矩阵 $Y \in \mathbb{R}^{p\times n}$ 的乘积，即

$$A = XY$$

- 由于观测含有噪声，原始数据矩阵 A 和分解 XY 不会完全吻合，应考虑

$$\begin{aligned} \min_{X\in\mathbb{R}^{d\times p},\, Y\in\mathbb{R}^{p\times n}} \quad & \frac{1}{2}\|A - XY\|_F^2 \\ \text{s.t.} \quad & X \geq 0,\ Y \geq 0 \end{aligned}$$

- 本质上是将高维空间中的数据在一个低维空间中表示
- 和主成分分析模型类似，但会得到比主成分分析模型更有实际意义的解

应用举例：非负矩阵分解

- 根据具体应用的不同，还可以考虑带正则项的非负矩阵分解模型

$$
\begin{aligned}
\min_{X\in\mathbb{R}^{d\times p},\, Y\in\mathbb{R}^{p\times n}} \quad & \frac{1}{2}\|A - XY\|_F^2 + \alpha r_1(X) + \beta r_2(Y) \\
\text{s.t.} \quad & X \geq 0, \quad Y \geq 0
\end{aligned}
$$

  - $r_1(X)$ 和 $r_2(Y)$ 是正则项
  - $\alpha, \beta > 0$ 是用来权衡拟合项和正则项的正则化参数

- 如果基向量的线性无关性，取 $r_1(X) = \|X^\top X - I\|_F^2$
- 如果每一个观测值都可以用少数几个基向量来表示，取 $r_2(Y) = \|Y\|_1$

应用举例：非负矩阵分解

Algorithms for non-negative matrix factorization

D Lee, HS Seung - Advances in neural information …, 2000 - proceedings

… nonnegativity is a useful constraint for matrix factorization that can lear
data [4, 5]. The nonnegative … for learning the optimal nonnegative factor

☆ 保存 99 引用 被引用次数：12260 相关文章 所有 36 个版本 »

Learning the parts of objects by non-negative matrix fact

DD Lee, HS Seung - nature, 1999 - nature.com

… an algorithm for non-negative matrix factorization that is able to … Non
distinguished from the … When non-negative matrix factorization is impler

☆ 保存 99 引用 被引用次数：16498 相关文章 所有 17 个版本

应用举例：相关系数矩阵估计

- 给定对称矩阵 $C \in \mathcal{S}^n$ 和非负对称权重矩阵 $H \in \mathcal{S}^n$，求解一个秩小于等于 p 的相关系数矩阵 X，使得在结合了权重矩阵的某种度量下最小化

$$
\begin{aligned}
\min_{X \succeq 0} \quad & \frac{1}{2}\|H \odot (X - C)\|_F^2 \\
\text{s.t.} \quad & X_{ii} = 1, \ i = 1, 2, \cdots, n \\
& \operatorname{rank}(X) \leq p
\end{aligned}
$$

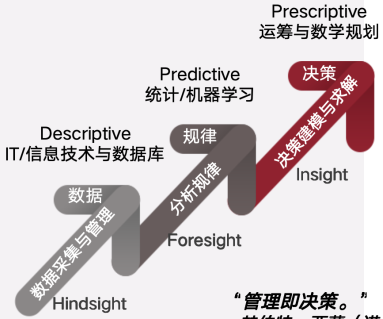
- 将 $\operatorname{rank}(X) \leq p$ 表示为 $X = V^\top V$, 其中 $V = [V_1, V_2, \cdots, V_n] \in \mathbb{R}^{p \times n}$，得到

$$
\begin{aligned}
\min_{V \in \mathbb{R}^{p \times n}} \quad & \frac{1}{2}\|H \odot (V^\top V - C)\|_F^2 \\
\text{s.t.} \quad & \|V_i\|_2 = 1, \ i = 1, 2, \cdots, n
\end{aligned}
$$

目录

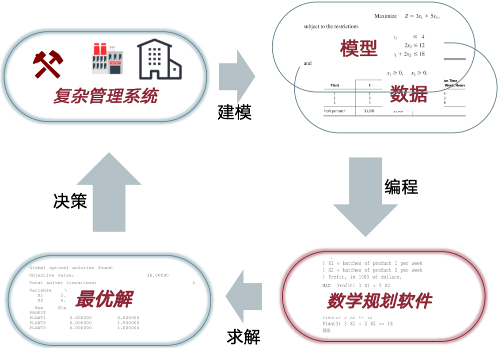
优化软件发展

- 现代复杂管理系统决策的主要特点：数据驱动，规模超大，多因素制约，多目标优化
- 数据驱动的决策流程：**数据-->规律-->决策**
- **数学规划：基于数据驱动的决策，问题建模与求解的核心环节**



“管理即决策。”

—赫伯特·西蒙（诺贝尔经济学奖）

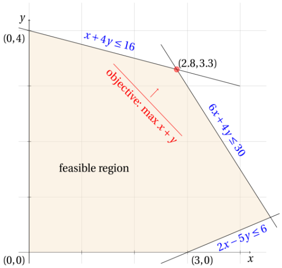


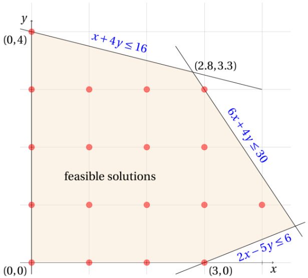
智能决策中两个经典的关键科学问题：

1，如何建模：将复杂决策问题翻译成**简洁**、**有效**的数学表达形式

2，如何求解：快速、**准确**地求解数学模型

优化软件发展





线性规划

- 问题目标：在可行域寻找最优解
- 规模较大：常有数百万变量和约束
- 稀疏性质：大量的稀疏矩阵运算
- 数值问题：受计算机精度影响

整数规划

- 可行域是离散的点
- 是经典的“NP-完全”问题
- 复杂多样的问题结构
- 需要求解多个线性规划问题

二阶锥规划

凸二次规划

凸二次约束规划

......

优化软件发展

数学理论门槛高

- 理论储备不足，基本理论框架由西方发展起来
- 求解计算模块的相关理论在国内基本空白状态
- 国内理论研究主要聚焦在数学规划在各个领域的定制化应用

软件工程难度大

- 数学规划的软件工程工作量巨大
- 线性5万行，整数规划百万行
- 相关的系统开发的理论研究国内极少
- 参照MATLAB

领域知识积累难

- Domain Knowledge欠缺
- 例：热启动的30种备选该如何从数千篇文献选取?
- 欧美积累了40多年
- 论文里不会提及，国外求解器也不会公布的不传之秘
- 只能自己一步步摸索，不断试错

技术创新迭代快

- 例如Gurobi研发组每年需要看3到500篇论文，但是只有5到7个技巧显著有用
- 需要自己观察与发展新的理论研究
- 更需要与新一代人工智能算法结合，软硬件结合，寻求“弯道超车”可能性

困难：投入大，周期长，工作很多时候是对欧美的追赶，很难产生论文

优化软件发展

世界著名数学规划求解器研发历程

- 1939年，苏联诺贝尔经济学奖获得者Leonid Kantorovich发明线性规划
- 1979年芝加哥大学的Charge 发布Lingo
- 1983年英国爱丁堡大学Ashford创办了XPRESS
- 1987年莱斯大学Bixby创办了CPLEX公司
- 2000年COIN-OR成立，并发布CLP和CBC
- 2005年德国ZIB发布了开源整数规划SCIP
- 2008年Cplex团队Bixby等离职创办GUROBI
- 2017年，上财发布中国第一个开源数学规划求解器LEAVES，2018年中科院 CMIP
- 2019年，杉数科技发布中国第一个专业数学规化求解器COPT，此后阿里，华为等纷纷入局
- 2021年至今，谷歌，ORACLE，微软纷纷开始组建自己的数学规划求解器团队

世界著名数学规划求解器一览表

GUROBI
CPLEX
XPRESS
美国与英国三大求解器巨头，累计三十年以上研发历史和95%以上市场。线性，整数，非线性各个模块功能齐备

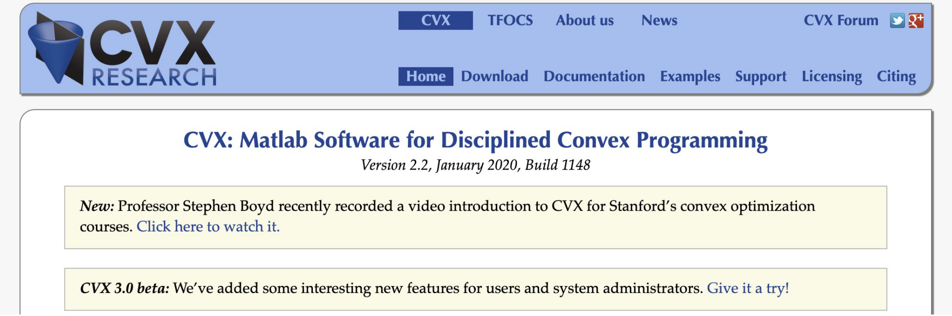
SAS ←世界最大商业统计软件（北卡）
CVX ←世界著名求解器建模平台(斯坦福大学)
IPOPT ←著名非线性规划开源软件(卡耐基梅隆)
Coin-OR ←世界最好开源线性规划（多组织维护）
Baron ←世界最好非线性规划（多组织维护）
SCIP ←世界最好整数规划(ZIB)

美国	德国	丹麦	俄罗斯	中国
NEOS	SCIP	MOSEK	GLPK	OPTV 华为
CBC	SOPLEX			MDOPT 阿里
				COPT 杉数
				CMIP 中科院
				LEAVES 上财

NEOS：世界最大免费服务平台（威斯康星大学）
CBC：世界最好开源组织（多组织维护）
MOSEK：世界最好锥规划（MOSEK）

优化模型语言

- CVX 以 MATLAB 为基础的优化模型语言，求解凸优化问题
  - 快速构造和识别凸性
  - 调用已有软件包求解变形后的凸优化问题
  - 包括免费软件 SDPT3 和 SeDuMi 以及商业软件 Gurobi 和 MOSEK 等



CVX

- 考虑优化问题

$$\begin{aligned} \min \quad & \|Ax - b\|_2 \\ \text{s.t.} \quad & Cx = d \\ & \|x\|_\infty \le e \end{aligned}$$

```
m = 20; n = 10; p = 4;
A = randn(m,n); b = randn(m,1);
C = randn(p,n); d = randn(p,1); e = rand;
cvx_begin
    variable x(n)
    minimize( norm( A * x - b, 2 ) )
    subject to
    C * x == d
    norm( x, Inf ) <= e
cvx_end
```

Q&A

Thank you!

感谢您的聆听和反馈