典型优化问题

宋晓良

大连理工大学数学科学学院

教材《最优化：建模、算法与理论》配套电子教案

http://bicmr.pku.edu.cn/~wenzw/optbook.html

典型优化问题简介

先回顾一下最优化问题的一般形式:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & c_i(x) \leq 0,\ i = 1, \cdots, m \\ & c_i(x) = 0,\ i = m+1, \cdots, m+l. \end{aligned}$$

- 按照目标和约束函数的简易程度分，可以分为线性规划和非线性规划．线性规划是指所有的目标函数和约束函数都是线性的，非线性规划是指目标函数和约束函数中至少有一个是非线性的
- 另外也可按照问题最优解的性质，分为凸优化问题与非凸优化问题．凸优化问题的任何稳定点都是全局极小点，非凸优化问题的稳定点则可能是局部极小点，全局极小点甚至是鞍点．

提纲

标准形式的凸优化问题

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

- $f_0, f_1, ..., f_m$ 凸函数；线性等式约束
- 拟凸问题：如果f_0 是拟凸(且$f_1, ..., f_m$ 为凸函数)

经常写成：

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

重要性质：凸问题的可行集为凸集.

例子

$$\begin{array}{ll} \min & f_0(x) = x_1^2 + x_2^2 \\ \text{s.t.} & f_1(x) = x_1/(1+x_2^2) \leq 0 \\ & h_1(x) = (x_1+x_2)^2 = 0 \end{array}$$

- f_0 凸函数; 可行集$\{(x_1, x_2) | x_1 = -x_2 \leq 0\}$ 为凸集
- 根据我们定义不是凸问题: f_1 非凸, h_1 不是线性函数
- 等价于(但不完全相等) 凸优化问题：

$$\begin{array}{ll} \min & x_1^2 + x_2^2 \\ \text{s.t.} & x_1 \leq 0 \\ & x_1 + x_2 = 0 \end{array}$$

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：假设x是局部极小，y全局最优且$f_0(y) < f_0(x)$.
x 局部最优意味着存在$R > 0$使得

$$z\text{可行}, \quad \|z - x\|_2 \leq R \quad \Longrightarrow \quad f_0(z) \geq f_0(x).$$

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

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

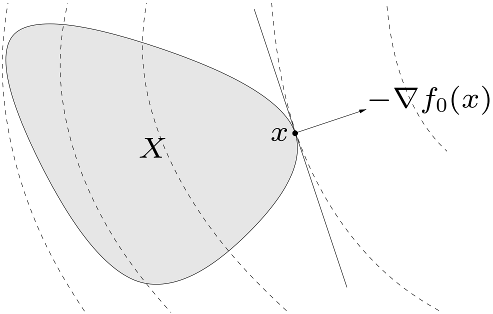
$$f_0(z) \leq \theta f_0(x) + (1 - \theta) f_0(y) < f_0(x),$$

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

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

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



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

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

- 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υ使得

$$x \in \operatorname{dom} f_0, \quad Ax = b, \quad \nabla f_0(x) + A^T \upsilon = 0$$

- 非负约束优化问题

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

x 是最优解当且仅当

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

提纲

线性规划基本形式

线性规划问题的一般形式如下：

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & c^\mathrm{T}x, \\ \text{s.t.} \quad & Ax = b, \\ & Gx \le e, \end{aligned} \tag{1}$$

其中$c \in \mathbb{R}^n, A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m, G \in \mathbb{R}^{p\times n}$ 和$e \in \mathbb{R}^p$ 是给定的矩阵和向量，$x \in \mathbb{R}^n$ 是决策变量．在实际中，由于其他形式都可进行转化，故考虑问题(1) 的两种特殊形式：标准形式

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & c^\mathrm{T}x, \\ \text{s.t.} \quad & Ax = b, \\ & x \ge 0, \end{aligned} \tag{2}$$

以及不等式形式

$$\begin{aligned} \max_{y\in\mathbb{R}^n} \quad & b^\mathrm{T}y, \\ \text{s.t.} \quad & A^\mathrm{T}y \le c. \end{aligned} \tag{3}$$

线性规划的应用：运输问题

- 线性规划最先在第二次世界大战中用于最大化资源的利用效率.
- 1947 年著名的单纯形方法被提出，使得线性规划问题可以被有效地求解. 之后，线性规划用到了更多其他领域当中，如农业、石油、钢铁、运输、通信和运筹学等. 线性规划的有效应用节省了大量的人力、物力和财力. 随着计算机以及求解算法的快速发展，我们可以求解更大规模的线性规划问题，保证了线性规划问题的应用前景.
- 运输问题的假设如下：假设有I 个港口$P_1, P_2, \cdots, P_I$，提供某种商品. 有J 个市场$M_1, M_2, \cdots, M_J$需要这种商品. 假设港口P_i 有s_i 单位的这种商品($i = 1, \cdots, I$)，市场M_j 需要r_j 单位的这种商品，且总供应与总需求相等，即$\sum_{i=1}^{I} s_i = \sum_{j=1}^{J} r_j$. 令$b_{ij}$ 为从港口P_i 运输单位数量商品到市场M_j 的成本. 运输问题是在满足市场需求下使得运输成本最低.

- 令x_{ij} 为从港口P_i 运输到市场M_j 的商品数量，总的运输代价为

$$\sum_{i=1}^{I}\sum_{j=1}^{J} x_{ij}b_{ij}. \tag{4}$$

- 港口P_i 总输出量为$\sum_{j=1}^{J} x_{ij}$，因为港口P_i 存有的商量总量为s_i，所以

$$\sum_{j=1}^{J} x_{ij} = s_i, \quad i = 1, 2, \cdots, I. \tag{5}$$

- 市场M_j 总输入量为$\sum_{i=1}^{I} x_{ij}$，因为市场M_j 的需求量为r_j，所以

$$\sum_{i=1}^{I} x_{ij} = r_j, \quad j = 1, 2, \cdots, J. \tag{6}$$

- 因为运输量是非负的，所以

$$x_{ij} \geq 0, \quad i = 1, 2, \cdots, I, \quad j = 1, 2, \cdots, J. \tag{7}$$

因此，想要在约束(5)—(7) 成立的情况下极小化(4)式．针对决策变量的$I\times J$ 矩阵(x_{ij})，可以得到如下线性规划问题：

$$
\begin{aligned}
\min_{x} \quad & \sum_{i=1}^{I}\sum_{j=1}^{J} x_{ij}b_{ij}, \\
\text{s.t.} \quad & \sum_{j=1}^{J} x_{ij} = s_i, \quad i = 1, 2, \cdots, I, \\
& \sum_{i=1}^{I} x_{ij} = r_j, \quad j = 1, 2, \cdots, J, \\
& x_{ij} \geq 0, \quad i = 1, 2, \cdots, I, \quad j = 1, 2, \cdots, J.
\end{aligned}
$$

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

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & \|x\|_1, \\ \text{s.t.} \quad & Ax = b. \end{aligned} \tag{8}$$

对每个$|x_i|$ 引入一个新的变量z_i，可以将问题(8) 转化为

$$\begin{aligned} \min_{z\in\mathbb{R}^n} \quad & \sum_{i=1}^n z_i, \\ \text{s.t.} \quad & Ax = b, \\ & -z_i \leq x_i \leq z_i, \quad i = 1, 2, \cdots, n, \end{aligned} \tag{9}$$

这是一个线性规划问题.

基追踪问题

- 除此之外，也可以引入x_i的正部和负部,其中$x_i^+ = \max\{x_i, 0\}$, $x_i^- = \max\{-x_i, 0\}$..
- 利用$x_i = x_i^+ - x_i^-, |x_i| = x_i^+ + x_i^-$，则问题(8) 转化为的另外一种等价的线性规划形式可以写成

$$
\begin{aligned}
\min_{x^+, x^- \in \mathbb{R}^n} \quad & \sum_{i=1}^{n} (x_i^+ + x_i^-), \\
\text{s.t.} \quad & Ax^+ - Ax^- = b, \\
& x^+, x^- \geq 0.
\end{aligned}
$$

可以看出这也是一个线性规划问题，且与原问题等价.

数据拟合

在数据拟合中，除了常用的最小二乘模型外，还有最小ℓ_1 范数模型

$$\min_{x\in\mathbb{R}^n} \quad \|Ax-b\|_1, \tag{10}$$

和最小ℓ_∞ 范数模型

$$\min_{x\in\mathbb{R}^n} \quad \|Ax-b\|_\infty. \tag{11}$$

这两个问题都可以转化成线性规划的形式.

- 对于问题(10)，通过引入变量$y=Ax-b$，可以得到如下等价问题：

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

- 利用基追踪问题中类似的技巧，可以将上述绝对值优化问题转化成线性规划问题.

- 对于问题(11)，令$t=\|Ax-b\|_\infty$，则得到等价问题

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

- 利用ℓ_∞ 范数的定义，可以进一步写为

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

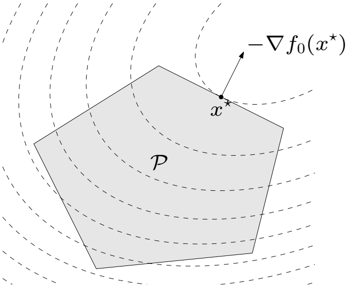
这是一个线性规划问题.

提纲

二次规划问题

$$
\begin{aligned}
\min \quad & \frac{1}{2}x^T P x + q^T x + r \\
\text{s.t.} \quad & Gx \leq h \\
& Ax = b
\end{aligned}
\tag{12}
$$

- $P \in \mathcal{S}_+^n$,故目标函数是二次的
- 在一个多面体内最小化一个二次凸问题



二次规划的应用

- 最小二乘问题：
$$\min \|Ax - b\|_2^2 \tag{13}$$
  - 该问题的解析解为$x^* = A^\dagger b$ (其中$A^\dagger$ 为广义逆)
- 随机线性规划
$$\begin{aligned} \min \quad & \bar{c}^T x + \gamma x^T \Sigma x = \mathbf{E} c^T x + \gamma \mathrm{var}(c^T x) \\ \text{s.t.} \quad & Gx \leq h, \quad Ax = b \end{aligned}$$
- c 是随机向量并且均值为$\bar{c}$，方差为Σ
- $c^T x$ 均值为$\bar{c}^T x$,方差为$x^T \Sigma x$
- $\gamma > 0$ 为风险参数,控制预期成本与风险.

带有二次约束的二次规划问题(QCQP)

考虑带有二次约束的二次规划问题：

$$
\begin{aligned}
\min \quad & (1/2)x^T P_0 x + q_0^T x + r_0 \\
\text{s.t.} \quad & (1/2)x^T P_i x + q_i^T x + r_i \le 0, \quad i = 1, \dots, m \\
& Ax = b
\end{aligned}
$$

- $P_i \in \mathbb{S}_+^n$;即目标函数与约束均为二次凸的
- 如果$P_1, \dots, P_m \in \mathbb{S}_{++}^n$, 则可行域为$m$个椭球与一个仿射集的交集.

提纲

半定规划

半定规划问题的一般形式如下：

$$
\begin{aligned}
\min \quad & c^{\mathrm{T}} x, \\
\text{s.t.} \quad & x_1 A_1 + x_2 A_2 + \cdots + x_n A_n + B \preceq 0, \\
& Gx = h,
\end{aligned} \tag{14}
$$

其中$c \in \mathbb{R}^n, A_i \in \mathcal{S}^m, i = 1, 2, \cdots, m, B \in \mathcal{S}^m, G \in \mathbb{R}^{p \times n}, h \in \mathbb{R}^p$ 为已知的向量和矩阵，$x = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$是自变量．

- 半定规划（SDP）是线性规划在矩阵空间中的一种推广．它的目标函数和等式约束均为关于矩阵的线性函数，而它与线性规划不同的地方是其自变量取值于半正定矩阵空间．
- 作为一种特殊的矩阵优化问题，半定规划在某些结构上和线性规划非常相似，很多研究线性规划的方法都可以作为研究半定规划的基础．由于半定规划地位的特殊性，我们将在本节中单独讨论半定规划的形式和应用．

类似于线性规划问题，我们考虑半定规划的标准形式

$$
\begin{aligned}
\min \quad & \langle C, X \rangle, \\
\text{s.t.} \quad & \langle A_1, X \rangle = b_1, \\
& \cdots \\
& \langle A_m, X \rangle = b_m, \\
& X \succeq 0,
\end{aligned}
\tag{15}
$$

和对偶形式：

$$
\begin{aligned}
\min \quad & -b^{\mathrm{T}} y, \\
\text{s.t.} \quad & y_1 A_1 + y_2 A_2 + \cdots + y_n A_n \preceq C.
\end{aligned}
\tag{16}
$$

形如(14) 式的优化问题都可以转化成(15) 式或者(16) 式的形式．

- 考虑二次约束二次规划问题

$$\begin{aligned}\min_{x\in\mathbb{R}^n} \quad & x^{\mathrm{T}}A_0x+2b_0^{\mathrm{T}}x+c_0,\\ \text{s.t.} \quad & x^{\mathrm{T}}A_ix+2b_i^{\mathrm{T}}x+c_i\le 0, \quad i=1,2,\cdots,m,\end{aligned} \tag{17}$$

  其中A_i 为$n\times n$ 对称矩阵．当部分A_i为对称不定矩阵时，问题(17)是NP 难的非凸优化问题．

- 写出问题(17) 的半定规划松弛问题．对任意$x\in\mathbb{R}^n$以及$A\in\mathcal{S}^n$，有恒等式

$$x^{\mathrm{T}}Ax=trx^{\mathrm{T}}Ax=trAxx^{\mathrm{T}}=\left\langle A,xx^{\mathrm{T}}\right\rangle,$$

  因此问题(17) 中所有的二次项均可用下面的方式进行等价刻画：

$$x^{\mathrm{T}}A_ix+2b_i^{\mathrm{T}}x+c_i=\left\langle A_i,xx^{\mathrm{T}}\right\rangle+2b_i^{\mathrm{T}}x+c_i.$$

二次约束二次规划问题的半定规划松弛

由上述分析，原始问题等价于

$$
\begin{aligned}
\min_{x\in\mathbb{R}^n} \quad & \langle A_0, X\rangle + 2b_0^{\mathrm{T}}x + c_0 \\
\text{s.t.} \quad & \langle A_i, X\rangle + 2b_i^{\mathrm{T}}x + c_i \le 0, \quad i = 1, 2, \cdots, m, \\
& X = xx^{\mathrm{T}}.
\end{aligned} \tag{18}
$$

进一步地，

$$
\begin{aligned}
x^{\mathrm{T}}A_i x + 2b_i^{\mathrm{T}}x + c_i &= \left\langle \begin{pmatrix} A_i & b_i \\ b_i^{\mathrm{T}} & c_i \end{pmatrix}, \begin{pmatrix} X & x \\ x^{\mathrm{T}} & 1 \end{pmatrix} \right\rangle, \\
&\overset{\text{def}}{=\!=} \langle \overline{A_i}, \overline{X} \rangle, \quad i = 0, 1, \cdots, m.
\end{aligned}
$$

二次约束二次规划问题的半定规划松弛

接下来将等价问题(18) 松弛为半定规划问题.

- 在问题(18) 中，唯一的非线性部分是约束$X = xx^{\mathrm{T}}$，我们将其松弛成半正定约束$X \succeq xx^{\mathrm{T}}$. 可以证明，$\overline{X} \succeq 0$ 与$X \succeq xx^{\mathrm{T}}$ 是等价的.
- 因此这个问题的半定规划松弛可以写成

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

其中“松弛”来源于我们将$X = xx^{\mathrm{T}}$ 替换成了$X \succeq xx^{\mathrm{T}}$.

- 令G 为一个无向图，其节点集合为$V = \{1, 2, \cdots, n\}$ 和边的集合为E. 令$w_{ij} = w_{ji}$ 为边$(i, j) \in E$ 上的权重，并假设$w_{ij} \geq 0$, $(i, j) \in E$. 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.
- 可以将最大割问题写成如下整数规划的形式：令$x_j = 1$, $j \in S$ 和$x_j = -1$, $j \in \overline{S}$，则

$$\begin{aligned} \max \quad & \frac{1}{2} \sum_{i<j} (1 - x_i x_j) w_{ij} \\ \text{s.t.} \quad & x_j \in \{-1, 1\}, \ j = 1, 2, \cdots, n. \end{aligned} \tag{19}$$

- 在问题(19) 中，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它的最优解.

二次约束二次规划问题的半定规划松弛

接下来介绍如何将问题(19) 松弛成一个半定规划问题.

- 令$W=(w_{ij})\in \mathcal{S}^n$，并定义$C=-\frac{1}{4}(\mathrm{Diag}(W\mathbf{1})-W)$为图$G$的拉普拉斯矩阵的$-\frac{1}{4}$倍，则问题(19) 可以等价地写为

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

  由于目标函数是关于x的二次函数，可将其等价替换为$\langle C, xx^{\mathrm{T}}\rangle$.

- 接下来令$X=xx^{\mathrm{T}}$，注意到约束$x_i^2=1$，这意味着矩阵X对角线元素$X_{ii}=1$. 因此利用矩阵形式我们将最大割问题转化为

$$\begin{aligned} \min \quad & \langle C, X\rangle, \\ \text{s.t.} \quad & X_{ii} = 1,\ i = 1, 2, \cdots, n, \\ & X \succeq 0,\ \mathrm{rank}(X) = 1. \end{aligned} \tag{20}$$

- 问题(20) 和(19) 是等价的，这是因为$X=xx^{\mathrm{T}}$ 可以用约束$X\succeq 0$和$\mathrm{rank}(X)=1$等价刻画.

极小化最大特征值

- 问题的形式可表示为：$\lambda_{\max}(A_0+\sum_i x_iA_i)$:

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

SDP形式：

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

对偶问题形式：

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

- 等价形式来源于：

$$\lambda_{\max}(A)\le t \Longleftrightarrow A \preceq tI$$

提纲

最小二乘问题

- 最小二乘问题的一般形式如下：

$$\min_{x\in\mathbb{R}^n} \quad \sum_{i=1}^{m} r_i^2(x), \tag{21}$$

  其中$r_i:\mathbb{R}^n\to\mathbb{R}$为实值函数．如果所有的$r_i$都是线性函数，则称问题(21) 为线性最小二乘问题，否则称其为非线性最小二乘问题．最小二乘问题是线性回归和非线性回归的基础．

- 最小二乘问题也常用于线性（非线性）方程组问题当中．当线性（非线性）观测带有噪声时，我们一般会基于该线性（非线性）系统建立最小二乘模型．特别地，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．

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

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

即$r_i(x)=a_i^{\mathrm{T}}x-b_i,\ i=1,2,\cdots,m$.

- 记$A=[a_1,a_2,\cdots,a_m]^{\mathrm{T}}$，那么线性最小二乘问题可以等价地写成如下形式：

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

这是一个无约束二次目标函数的优化问题.

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

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

- 在实际中，出于计算上的可行性，我们一般会限制在一个特定函数空间上来求f的一个逼近解．如果利用线性函数逼近，即$f(a) = Xa + y$，其中$X \in \mathbb{R}^{q \times p}$，$y \in \mathbb{R}^q$，则为了求解$X, y$，可以建立如下最小二乘问题：

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

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

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

其中x_i 为待定系数．这是关于x 的线性方程组．

除了这种基函数的和的方式，深度学习也通过一些简单函数的复合来逼近原未知函数.

- 具体地，假设有一些简单的非线性向量函数$\phi_i(\theta): \mathbb{R}^q \to \mathbb{R}^q$，并构造如下复合函数：

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

- 在实际中常用的简单非线性函数有ReLU，即

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

且

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

- 这样的做法往往会带来更多未知的非线性，因而可能在更大的函数空间中得到未知函数的一个更好的逼近．将点a_i 处的取值代入，则得到如下非线性方程组：

$$\begin{aligned} f(a_i)-b_i &= \phi_n\left(X_n\phi_{n-1}(X_{n-1}\cdots\phi_1(X_1a_i+y_1)\cdots+y_{n-1}\right)+y_n\right)-b_i \\ &= 0, \quad i=1,2,\cdots,n. \end{aligned}$$

- 需要求解的是关于$X_1\in\mathbb{R}^{q\times p}, X_i\in\mathbb{R}^{q\times q}$, $y_i\in\mathbb{R}^q, i=1,2,\cdots,n$ 的非线性方程组．我们一般考虑替代的最小二乘问题

$$\min_{\{X_i,y_i\}} \sum_{i=1}^{m}\|f(a_i)-b_i\|^2.$$

带微分方程约束优化问题

- 当约束中含微分方程时，称相应的优化问题为带微分方程约束的优化问题．其在最优控制、形状优化等各种领域中有着广泛应用．求解瓦斯油催化裂解生成气体和其他副产物的反应系数的问题反应过程可以由如下非线性常微分方程组表示：

$$\begin{cases} \dot{y}_1 = -(\theta_1 + \theta_3) y_1^2, \\ \dot{y}_2 = \theta_1 y_1^2 - \theta_2 y_2, \end{cases} \tag{22}$$

其中系数$\theta_i \geq 0$, $i = 1, 2, 3$，且y_1, y_2 的初值条件是已知的．

- 我们考虑的问题是

$$\begin{aligned} \min_{\theta \in \mathbb{R}^3} \quad & \sum_{j=1}^{n} \|y(\tau_j; \theta) - z_j\|^2, \\ \text{s.t.} \quad & y(\tau; \theta) \text{满足方程组(22)} \end{aligned}$$

这里z_j 是在时刻τ_j 的y 的测量值，n 为测量的时刻数量．

提纲

复合优化问题

- 复合优化问题一般可以表示为如下形式：

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

其中$f(x)$ 是光滑函数（比如数据拟合项），$h(x)$ 可能是非光滑的（比如ℓ_1范数正则项，约束集合的示性函数，或它们的线性组合）.

- 从已经介绍的各种各样的应用问题不难发现，复合优化问题在实际中有着重要的应用，并且其中的函数$h(x)$一般都是凸的.
- 由于应用问题的驱动，复合优化问题的算法近年来得到了大量的研究，比如次梯度法，近似点梯度法，Nesterov加速算法和交替方向乘子法，等等.

考虑带有ℓ_1 范数正则项的优化问题：

$$\min_{x} \quad f(x) + \mu\|x\|_1, \tag{23}$$

这里$\mu > 0$ 为给定的参数. 这个问题广泛存在于各种各样的应用中：

- ℓ_1 范数正则化回归分析问题：

$$\min_{x\in\mathbb{R}^n} \quad \mu\|x\|_1 + \frac{1}{2}\|Ax - b\|_2^2, \tag{24}$$

  其中$\mu > 0$是给定的正则化参数. 该问题可以看成是问题(23) 的特殊形式,问题(24) 又称为LASSO（least absolute shrinkage and selection operator）.其中$f(x) = \frac{1}{2}\|Ax - b\|_2^2, h(x) = \mu\|x\|_1$.

复合优化问题：应用举例

- 矩阵分离问题：

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

  其中$f(x) = \|X\|_*, h(x) = \mu\|S\|_1$.

- 字典学习问题：

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

  其中$f(D,X) = \frac{1}{2n}\|DX - A\|_F^2$.

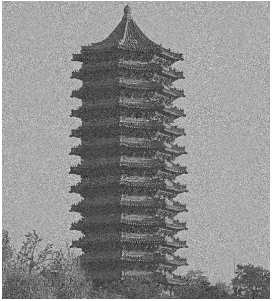
复合优化问题：图像去噪

- 图像去噪问题是指从一个带噪声的图像中恢复出不带噪声的原图．记带噪声的图像为y，噪声为ε，那么

$$y = x + \varepsilon,$$

其中x 为要恢复的真实图像．

(a) 原始图像u



(b) 添加噪声后的图像

(c) 恢复后的图像

Figure: 图像去噪的例子

复合优化问题：图像去噪

- 由全变差模型，去噪问题可表示为

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

这里，离散的线性算子为单位矩阵.

- 也可以利用小波框架. 小波变换可以很好地保护信号尖峰和突变信号，并且噪声对应的小波系数往往很小. 因此，去噪问题的小波分解模型可以写为

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

其中 $\lambda=(\lambda_1,\lambda_2,\cdots,\lambda_m)^{\mathrm{T}}$ 是给定的.

复合优化问题：盲反卷积

盲反卷积是图像处理中的一个基本问题，其目的是从一个模糊的图像恢复出原来清晰的图像.

- 为了简化问题，假设模糊是线性的以及空间不变的. 线性且空间不变的模糊可以表示成一个卷积. 令x 为原始的清晰图像，a 为未知的卷积核对应的矩阵，y 为观测到的模糊图像以及ε 为观测噪声. 盲反卷积问题可以表示成

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

  其中$*$ 为卷积算子. 假设噪声为高斯噪声，则转化为求解优化问题

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

- 再假设原始图像信号在小波变换下是稀疏的，进一步得到如下复合优化问题：

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

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

提纲

矩阵优化的基本形式

矩阵优化问题具有如下形式：

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

其中$\mathcal{X}$ 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．对于矩阵优化问题，如果决策变量为一个$n\times n$ 矩阵，那么我们可能需要确定n^2 个元素．因此，决策变量的维数过大往往是矩阵优化问题难以快速求解的一个重要原因．

- 矩阵优化是在近几十年发展起来的一类变量含有矩阵的优化问题．它广泛地出现在组合数学、材料科学、机器学习和统计学等各种各样的应用当中．
- 和向量相比，矩阵有许多新的性质：例如秩、特征值等．所以矩阵优化问题的求解通常要困难一些．

矩阵优化问题

- 半定规划问题(15)：半定规划是一类特殊的矩阵优化问题，它的目标函数和约束均为线性函数，自变量X取值于半正定矩阵空间中．
- 低秩矩阵恢复问题：

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

考虑函数$h(X)=\|X\|_*$，其次微分为

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

其中$X=U\Sigma V^{\mathrm{T}}$为X的约化奇异值分解．对于函数

$$f(X)=\frac{1}{2}\sum_{(i,j)\in\Omega}(X_{ij}-M_{ij})^2,$$

令矩阵$P\in\mathbb{R}^{m\times n}$为

$$P_{ij}=\begin{cases}1, & (i,j)\in\Omega,\\ 0, & \text{其他},\end{cases}$$

矩阵优化问题

- 主成分分析问题：

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

通过简单计算，我们有

$$\nabla\psi(X)=-2AA^{\mathrm{T}}X.$$

- 矩阵分离问题：

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

在这里自变量X与S均为矩阵，$\psi(X,S)$关于X和S均为不可微函数.

- 字典学习问题：

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

令$f(X,D) = \frac{1}{2n}\|DX - A\|_F^2$，我们有

$$\begin{aligned} \nabla_X f &= \frac{1}{n} D^{\mathrm{T}}(DX - A), \\ \nabla_D f &= \frac{1}{n}(DX - A)X^{\mathrm{T}}. \end{aligned}$$

在这里需要注意$f(X,D)$关于两个变量分别求梯度的形式的区别.

假设a 为d 维空间中的非负随机向量，它的n 个观测值为$\{a_i\}_{i=1}^n$．并记矩阵$A=[a_1,a_2,\cdots,a_n]\in\mathbb{R}^{d\times n}$，非负矩阵分解问题是指将高维矩阵$A$ 分解成非负$d\times p$ 基矩阵$X=[x_1,x_2,\cdots,x_p]$和非负$p\times n$ 系数矩阵$Y=[y_1,y_2,\cdots,y_n]$的乘积，即

$$A=XY.$$

- 从上面的表达式可以看出，y_j 为观测点a_j 在基矩阵X 上的权重系数．也就是说，非负矩阵分解把数据分成基向量的线性组合．
- 通常选取$p\ll d$，那么得到的基矩阵X 的列张成了原数据空间的一个子空间．这本质上是将高维空间中的数据在一个低维空间中表示．当数据点的内蕴结构完全被基矩阵X 包含时，我们就得到了一个很好的低维表示．

应用举例：非负矩阵分解

- 一般情况下，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．在这种情况下我们应当寻找误差最小的解．利用矩阵的F范数可以定义相似性度量

$$\|A - XY\|_F^2,$$

我们考虑如下优化问题

$$\min_{X \in \mathbb{R}^{d \times p},\ Y \in \mathbb{R}^{p \times n}} \quad \|A - XY\|_F^2, \quad \text{s.t.} \quad X \geq 0,\ Y \geq 0, \tag{25}$$

其中"≥ 0"表示矩阵的每个元素是非负的．

- 从低维空间逼近的角度来看，非负矩阵分解模型和主成分分析模型类似．但在实际问题中，非负矩阵分解模型会得到比主成分分析模型更有实际意义的解．
- 比如，给定很多幅人脸图片（都可以用元素值为$0 \sim 255$的矩阵来表示其灰度图），我们想要提取脸部的特征．利用主成分分析得到的主成分可能包含负数像素值，这是不合理的．但是如果使用非负矩阵分解，则可以有效避免这类情形的发生．

- 我们称问题(25) 为基本的非负矩阵分解模型．根据具体应用的不同，有时还考虑带正则项的非负矩阵分解模型

$$\begin{aligned} \min_{X\in\mathbb{R}^{d\times p},\ Y\in\mathbb{R}^{p\times n}} \quad & \|A-XY\|_F^2+\alpha_1 r_1(X)+\beta r_2(Y), \\ \text{s.t.} \quad & X\geq 0, \quad Y\geq 0, \end{aligned} \tag{26}$$

- 其中$r_1(X)$ 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．比如，如果Y 的列是稀疏的，那么每一个观测值都可以用少数几个基向量来表示．相应地，我们可以惩罚Y 的每一列的ℓ_1 范数．为了保证基向量的线性无关性，往往还要求X 的列之间是相互正交的．
- 如果数据矩阵A 分布在一个低维的非线性流形上，则考虑流形或者图上的非负矩阵分解模型．

提纲

典型优化算法软件

前面介绍了各种各样的优化问题．对于每一类优化问题，我们都有相应的求解算法以及一些流行的算法软件包．

- SDPT3: 这个开源软件包的基本代码是用MATLAB 来写的，但是关键的子程序是用FORTRAN 和C 语言通过MEX 文件来完成的．它可以求解锥规划问题，其中锥可以是半定矩阵锥、二次锥和非负象限中的一个或者多个的乘积．这个软件主要实现的算法是一种原始– 对偶内点法．
- MOSEK: 这个商业软件包可以求解线性规划、二次锥规划、半定规划、二次规划等凸优化问题，以及混合整数线性规划和混合整数二次规划等．它的重点是求解大规模稀疏问题，尤其在求解线性规划、二次锥规划和半定规划的内点法设计上做得非常有效．除了内点法之外，MOSEK 还实现了线性规划问题的单纯形算法，特殊网络结构问题的网络单纯形算法以及针对混合整数规划问题的算法．它提供C, C#, Java, Python, MATLAB 和R 等接口．

- CPLEX: 这个商业软件可以求解整数规划问题，非常大规模的线性规划问题（使用单纯形方法或者内点法），凸和非凸二次规划问题，二次锥规划问题．它提供C++, C#, Java, Microsoft Excel和MATLAB 接口，并且提供一个独立的交互式优化器可执行文件，用于调试和其他目的．
- Gurobi: 这个商业软件可以求解线性规划（单纯形法和并行的内点法），二次规划（采用单纯形法和内点法），二次约束规划，混合整数线性规划，混合整数二次规划，混合整数二次约束规划．它提供C, C++, Java, .NET, Python, MATLAB 和R 等接口．

典型优化算法软件

- IPOPT: 这个开源软件可以求解大规模非线性规划问题，主要实现了原始–对偶内点法，并使用滤波（filter）方法代替线搜索．IPOPT 主要使用C++ 语言编写，并提供C, C++, FORTRAN, Java, MATLAB 和R 等接口．
- Knitro: 用来求解大规模非线性优化问题的商业软件．这个软件提供了四种不同的优化方法，两种内点型方法和两种积极集（active set）方法，可以用来求解一般非凸非线性规划问题，非线性方程组，线性规划，二次（线性）约束二次规划问题，线性（非线性）最小二乘问题，混合整数规划问题以及无导数优化问题，等等．Knitro 支持的编程语言有C, FORTRAN, C++, C#, Java, MATLAB, R, Python 等，以及模型语言AMPL，AIMMS，GAMS 和MPL 等．因其具有大量的用户友善的选项以及自动调试器，全局优化的并行多重启动策略，导数逼近和检查以及内部预分解器，在实际中被广泛采用．

优化模型语言

- 模型语言的发展开始于19世纪70年代后期，其主要动因是计算机的出现．在优化模型语言中，优化模型可以写成和数学表达式很类似的方式，以此给用户带来更便捷的服务．
- 模型的表达式形式是与求解器无关的，不同的求解器需要用优化模型语言将给定的模型和数据转为其求解的标准形式，然后再对其进行求解．这类工具有三个优点：一是将容易出错的转化步骤交给计算机完成，降低错误率；二是在模型和算法之间建立了一个清晰的界限；三是对于困难的问题，可以尝试不用的求解器，得到更好的结果．

- CVX是以MATLAB 为基础的优化模型语言，用来求解凸优化问题．它允许将优化问题的目标函数以及约束用MATLAB 语法写出．
- CVX 采用了一种快速构造和识别凸性的准则，服从这个准则的凸问题都可以很快地被识别出来．之后CVX 根据用户的选择调用已有软件包来求解变形后的凸优化问题，这些软件包括免费软件SDPT3 和SeDuMi 以及商业软件Gurobi 和MOSEK 等．除了一些典型问题外，CVX 还可以识别一些更复杂的凸优化问题，例如带ℓ_1范数的优化问题．目前CVX 还有Julia 语言版本和Python 语言版本CVXPY．
- 除CVX 外，还有很多发展成熟的优化模型语言可供我们使用，如AMPL，YALMIP 等．

CVX

考虑如下优化问题：

$$\begin{aligned} \min \quad & \|Ax-b\|_2, \\ \text{s.t.} \quad & Cx = d, \\ & \|x\|_\infty \le e, \end{aligned} \tag{27}$$

它可以写成:

```
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
```

CVX

- 代码中的前三行是关于A, b, C, d, e的构造．在调用CVX 求解的时候，对应的代码需要以`cvx_begin` 开始，并且以`cvx_end` 结尾．在这两行语句之间，我们需要定义
- 要求解的优化问题．在上面的例子中，`variable x(n)` 表示决策变量x 为n 维空间中的向量．目标函数$\|Ax-b\|_2$ 则用`norm( A * x - b, 2 )` 来表示，`minimize` 表示求解目标函数的极小值．最后以`subject to` 开始描述问题的约束，`C * x == d` 和`norm( x, Inf ) <= e` 分别表示约束$Cx = d$ 和$\|x\|_\infty \le e$．
- 执行上述代码，CVX 会选取默认的凸优化问题算法来返回上面问题的解．