机器学习中的优化算法作业1　(1.3, 1.4, 3.3, 3.7, 3.8, 2.12, 2.13, 2.15, 5.1, 5.6(a)
5.7(b)(d), 5.16)

谢[illegible] 2020012544

1.3 (a) Q-超线性收敛

$$\lim_{k\to\infty}\frac{\|x^{k+1}-x^*\|}{\|x^k-x^*\|}=\lim_{k\to\infty}\frac{1}{|k+1|}=0$$

(b) Q-超线性收敛

由 $x^*=0$，若 k 为奇数则有 $\lim_{k\to\infty}\frac{\|x^{k+1}-x^*\|}{\|x^k-x^*\|}=$

$$\lim_{k\to\infty} k\cdot\left|\frac{x^{k+1}}{x^{k-1}}\right|=\lim_{k\to\infty}\frac{k}{64^{2^{k-1}}}=0$$

若 k 为偶数，$\lim_{k\to\infty}\frac{\|x^{k+1}-x^*\|}{\|x^k-x^*\|}=\lim_{k\to\infty}\frac{1}{(k+1)}\cdot\frac{x^k}{x^k}=0$

1.4 (a) 该点列收敛，且 $f(x^k)=(1+\frac{1}{2^k})^2$

$k\to\infty$ 时 $f\to1$　$f(x^k)-1=\frac{1}{2^k}(2+\frac{1}{2^k})$

因 $\lim_{k\to\infty}\frac{\|f(x^{k+1})-f^*\|}{\|f(x^k)-f^*\|}=\frac{2+\frac{1}{2^{k+1}}}{2(2+\frac{1}{2^k})}\to\frac{1}{2}$

则为 Q-线性收敛

(b) 不收敛

sin k. cos k 不收敛 而 $1+\frac{1}{2^k}\to1$

2.12 (a) $f(x)=\sum_i x_i \ln x_i$

$$f^*(y)=\sup_x y^T x - f(x) = \sum_i \sup_{x_i \geq 0} y_i x_i - x_i \ln x_i = \sum_i e^{y_i - 1}$$

(b) $f(X) = -\ln \det(X) = \ln\det(X^{-1})$

$$f^*(y) = \sup_{X \succ 0} \operatorname{tr}(Y^T X) + \ln \det(X)$$

若Y非负定　取Y的一特征向量 v s.t. $\|v\|_2 = 1$ 且 $\lambda \geq 0$

取 $X = I + t v v^T$ 则 $\operatorname{Tr}(Y^T X) + \ln\det(X) = \operatorname{Tr}(Y) + \lambda t + \ln(1+t) \to \infty$

$(t \to \infty)$

若Y负定 则 $\nabla(\operatorname{Tr}(Y^T X) + \ln\det(X)) = Y + X^{-1} = 0$

其中 $X = -Y^{-1}$ 为正定阵

$f^*(Y) = -\ln\det(-Y) - n$

$$\Rightarrow \quad f^*(Y) = \begin{cases} -\ln\det(-Y) - n & (Y\text{负定}) \\ +\infty & (\text{otherwise}) \end{cases}$$

(c) $f(x) = \max_i x_i$

$$f^*(y)=\sup_x\{y^Tx-\max_i x_i\}=\sup_x\{\sum_i y_ix_i-\max_i x_i\}$$

若$\exists i$, s.t $y_i<0$. 则对应 $x_i=-t$ $(t>0)$ 且其余 $x_j=0$ $(j\neq i)$

则 $t\to\infty$ 时 $y^Tx-\max_i x_i\to\infty$

$\Rightarrow y\notin \mathrm{dom}\, f^*$

若 $y\geq 0$ 且 $y^T\cdot 1>1$ 令 $x=t1$ $(t\to\infty)$

$\Rightarrow y^Tx-\max_i x_i = ty^T1-t=t(y^T1-1)\to\infty$

$\Rightarrow y\notin \mathrm{dom}\, f^*$

若 $y\geq 0$ 且 $y^T\cdot 1=1$

$y^Tx-\max_i x_i\leq 0$. 且在 $x=0$ 时取等.

$\Rightarrow f^*(y)=0$

综上, $f^*(y)=\begin{cases}0 & y\geq 0 \text{ 且 } y^T1=1\\ +\infty & \text{otherwise}\end{cases}$

(d) $f(x,t)=-\ln(t^2-x^Tx)$

$\mathrm{dom}\, f=\{(x,t)\in\mathbb{R}^n\times\mathbb{R}\mid \|x\|_2<t\}$ $(t>0)$

$$f^*(y,s)=\sup_{(x,t)\in\mathrm{dom} f}\{x^Ty+ts+\ln(t^2-x^Tx)\}$$

对 $(y,s)\in\{(y,s)\mid \|y\|_2\geq -s\}$

对于 $p>1$. 取 $x=py$ $\quad t=p(\|x\|_2+1)=p^2(\|y\|_2+1)>p\|y\|_2\geq -ps$

$\Rightarrow\ y^Tx+ts>p\|y\|_2^2-ps^2=p(\|y\|_2^2-s^2)\geq 0$

则 $p\to\infty$ 时. $y^Tx+ts\to\infty$

同时 $\ln(t^2-x^Tx)\to\infty$, 此时 $y^Tx+ts+\ln(t^2-x^Tx)$ 无上界

对 $(y,s)\in\{(y,s)\mid \|y\|_2<-s\}$, $\nabla(y^Tx+ts+\ln(t^2-x^Tx))=0$

此时 $x=\frac{2y}{s^2-y^Ty}$, $t=-\frac{2s}{s^2-y^Ty}$

$\Rightarrow f^*(y,s)=\frac{2y^Ty}{s^2-y^Ty}-\frac{2s^2}{s^2-y^Ty}+\ln 4-\ln(s^2-y^Ty)$

$=-2+\ln 4-\ln(s^2-y^Ty)$

2.13. (a) $f(x)=\|Ax-b\|_2+\|x\|_2$

$f_1(x)=\|Ax-b\|_2$. $f_2(x)=\|x\|_2$. 均为凸函数

$$\partial f(x)=\partial f_1(x)+\partial f_2(x)=\begin{cases} A^T\frac{Ax-b}{\|Ax-b\|_2}+\frac{x}{\|x\|_2} & Ax-b\neq 0,\ x\neq 0 \\ \frac{x}{\|x\|_2} & Ax-b=0,\ x\neq 0 \\ \frac{A^T(Ax-b)}{\|Ax-b\|_2} & Ax-b\neq 0.\ x=0 \\ 0 & Ax-b=0,\ x=0 \end{cases}$$

2.15. $f(x)$ 为 m-强凸 $\Rightarrow g(x) = f(x) - \frac{m}{2}\|x\|_2^2$ 为凸函数

由凸函数运算：$\partial g(x) + mx = \partial f(x)$

则对 $\forall \tilde{g} \in \partial f(x)$，$f(y) \geq f(x) + \frac{m}{2}\|y\|_2^2 - \frac{m}{2}\|x\|^2 + \langle \tilde{g} - mx, y-x \rangle$

$(\forall y \in \mathrm{dom} f)$

两边同时对 y 求 inf：$\inf\limits_{y \in \mathrm{dom} f} f(y) \geq f(x) - \frac{\|\tilde{g}\|_2^2}{2m} \quad (\forall \tilde{g} \in \partial f(x))$

$$\Rightarrow \inf_{y \in \mathrm{dom} f} f(y) \geq \sup_{\tilde{g} \in \partial f(x)} f(x) - \frac{\|\tilde{g}\|_2^2}{2m} = f(x) - \inf_{\tilde{g} \in \partial f(x)} \frac{\|\tilde{g}\|_2^2}{2m}$$

即得：$\inf\limits_{y \in \mathrm{dom} f} f(y) \geq f(x) - \frac{1}{2m} \mathrm{dist}^2(0, \partial f(x))$

3.3 模型 3.2.4.

$\min_x \frac{1}{2}\|Ax-b\|_2^2$, $x^* = A^{-1}b$. $b=(0,0,0)^T$ 时 $x^*=(0,0,0)^T$

$b=(10^{-4},0,0)^T$ 时. $x^*=(100,0,0)^T$

模型 3.2.6

$\min_x \frac{1}{2}\|Ax-b\|_2^2 + \mu\|x\|_2^2$, $x^* = (A^TA+2\mu I)^{-1}A^Tb$

$b=(0,0,0)^T$ 时. $x^*=(0,0,0)^T$

$b=(10^{-4},0,0)^T$ 时. $x^* = (\frac{10^2}{1+2\times10^2\mu}, 0, 0)^T$

模型 (3.2.4) 中矩阵 A 病态，故 b 的轻微扰动对 x 有较大影响

模型 (3.2.6) 增加了 l_2 正则项 使两次解极为接近

3.7 (a) $L(x) = \sum_{i=1}^m \ln P(b_i - a_i^Tx) = -\sum_{i=1}^m \left(\ln\sqrt{2\pi}\sigma + \frac{(b_i-a_i^Tx)^2}{2\sigma^2}\right)$

$\max_x L(x) \Leftrightarrow \min_x \sum_{i=1}^m (b_i - a_i^Tx)^2$

(b) $L(x) = \sum_{i=1}^m \ln P(b_i - a_i^Tx) = -\sum_{i=1}^m \left(\ln 2a + \frac{|b_i-a_i^Tx|}{a}\right)$

$\max_x L(x) \Leftrightarrow \min_x \sum_{i=1}^m |b_i - a_i^Tx|$

(c) $L(x) = \prod_{i=1}^m P(b_i - a_i^Tx) = \begin{cases} \left(\frac{1}{2a}\right)^m & \max_{1\le i\le m} |b_i - a_i^Tx| \le a \\ 0 & \text{otherwise} \end{cases}$

故优化问题可构造为 $\min_x \max_{1\le i\le m} |b_i - a_i^Tx|$

3.8. $\min_{x} \sum_{i=1}^{m} \ln\left(\frac{1+|a_i^T x|}{1+|a_i^T x|+b_i a_i^T x}\right)$

该模型得到的优化问题是非凸的

5.1 满足性质：A正定，$b \in \mathbb{R}(A)$

5.6. (a) 考虑方程 $Ax=b$ 的解：

① 无解 即 $b \notin \mathbb{R}(A)$ 此时 $\min_{Ax=b} c^T x = +\infty$

② 有解. 且 $c \perp \ker(A)$，设 $c = A^T\lambda + \hat{c}$ 其中 $A\hat{c}=0$

若 $c \in \mathbb{R}(A^T)$ 则有 $\hat{c}=0$，此时 $c^T x = \lambda^T A x + \hat{c}^T x = b^T\lambda$

$\Rightarrow \min_{Ax=b} c^T x = b^T\lambda$，且可行域中所有解均为最优解

③ 有解. 且 $c \notin \mathbb{R}(A^T)$ 此时 $\hat{c} \neq 0$

设 x_0 为一可行点，令 $x = x_0 - t\hat{c}$，则有 x 也可行

$$c^T x = (A^T\lambda + \hat{c})^T (x_0 - t\hat{c}) = \lambda^T A x_0 - t\lambda^T A\hat{c} + \hat{c}x_0 - t\|c\|_2^2$$

$$= \lambda^T b + \hat{c}x_0 - t\|c\|_2^2$$

$\lim\limits_{t\to\infty} c^T x \to -\infty$，此时最优值为 $-\infty$

综上，最优值 $p = \begin{cases} +\infty & b \notin \mathbb{R}(A) \\ \lambda^T b & c \in \mathbb{R}(A^T)\ (\text{存在}\lambda\text{使}\ c = A^T\lambda) \\ -\infty & \text{otherwise} \end{cases}$

5.7

(b) $\min\limits_{x\in\mathbb{R}^n} \|Ax-b\|$，令 $y = Ax-b$. 考虑 $\min\limits_{Ax-b-y=0} \|y\|_1$

$$L(x,y,\lambda) = \|y\|_1 - \langle \lambda, Ax-b-y \rangle$$

$$g(\lambda) = \inf_{x,y} \{\|y\|_1 - \langle \lambda, Ax \rangle + \langle \lambda, y \rangle + \langle \lambda, b \rangle\}$$

$$= -\sup_{x,y} \{\langle -\lambda, y \rangle - \|y\|_1 + \langle A^T\lambda, x \rangle\} - \langle \lambda, b \rangle$$

$\Rightarrow$ 当 $A^T\lambda \neq 0$ 时 $g(\lambda) = -\infty$

设 $f(y) = \|y\|_1$ 则 $f^*(y) = \begin{cases} 0 & \text{if } \|\lambda\|_\infty \leq 1 \\ +\infty & \text{otherwise} \end{cases}$

综上，$g(\lambda)=\begin{cases}-\langle\lambda,b\rangle & \text{if } \|\lambda\|_\infty\le 1,\ A^T\lambda=0\\ -\infty & \text{otherwise}\end{cases}$

$\Rightarrow$ 对偶问题：$\max\limits_{\|\lambda\|_\infty\le 1,\ A^T\lambda=0}\ \langle -\lambda,b\rangle$

(d) $\min\limits_{x\in\mathbb{R}^n,\ \|x\|_2^2\le 1}\ x^TAx+2b^Tx$ （A正定）

$L(x,\lambda)=x^TAx+2b^Tx-\lambda(1-\|x\|_2^2)$

$g(\lambda)=\inf\limits_{x}\ \{x^TAx+2b^Tx+\lambda\|x\|_2^2-\lambda\}$

$=\inf\limits_{x}\ \{x^T(A+\lambda I)x+2b^Tx-\lambda\}$

$=-b^T(A+\lambda I)^{-1}b-\lambda$

故对偶问题为 $\max\limits_{\lambda\ge 0}\ -b^T(A+\lambda I)^{-1}b-\lambda$

5.16 $\max\limits_{x,\xi}\ \frac{1}{2}\|x\|_2^2+u\sum\limits_{i=1}^{m}\xi_i$ such that $b_ia_i^Tx\ge 1-\xi_i$ ($\xi_i\ge 0$)

$(i=1,2,\cdots,m)$

$L(x,\xi,\lambda,\beta)=\frac{1}{2}\|x\|_2^2+u\sum\limits_{i=1}^{m}\xi_i-\sum\limits_{i}\xi_i\beta_i-\sum\limits_{i}\lambda_i(b_ia_i^Tx+\xi_i-1)$

$=\frac{1}{2}\|x\|_2^2-\sum\limits_{i}\lambda_ib_ia_i^Tx+\sum\limits_{i}\lambda_i+\sum\limits_{i=1}^{m}\xi_i(u-\beta_i-\lambda_i)$

$g(\lambda,\beta)=\inf\limits_{x,\xi}\ \{\frac{1}{2}\|x\|_2^2-\sum\limits_{i}\lambda_ib_ia_i^Tx+\sum\limits_{i=1}^{m}\xi_i(u-\beta_i-\lambda_i)+\sum\limits_{i}\lambda_i\}$

若 $u-\beta_i-\lambda_i\ne 0$，则 $g(\lambda,\beta)=-\infty$

$$
\Rightarrow \quad g(\lambda, \beta) = \begin{cases} -\frac{1}{2} \left\| \sum_i \lambda_i b_i a_i \right\|_2^2 + \sum_i \lambda_i & \text{if } u - \beta_i - \lambda_i = 0 \quad \forall i = 1, 2, \cdots, m \\ -\infty & \text{otherwise} \end{cases}
$$

故对偶问题为

$$
\max_{\lambda_i, \beta_i} \; -\frac{1}{2} \left\| \sum_i \lambda_i \beta_i a_i \right\|_2^2 + \sum_i \lambda_i
$$

$$
\begin{aligned}
\text{s.t.} \quad & u - \beta_i - \lambda_i = 0 \\
& \lambda_i \geq 0 \\
& \beta_i \geq 0 \\
& i = 1, 2, \cdots, m
\end{aligned}
$$