运筹学基础：课后练习 1 答案

1. (a) 因为 $\lim_{k\to\infty}\frac{1}{k!}=0$ 设 $x^*=0$ 。
$\lim_{k\to\infty}\frac{\left|x^{k+1}\right|}{\left|x^k\right|}=\lim_{k\to\infty}\frac{\frac{1}{(k+1)!}}{\frac{1}{k!}}=\lim_{k\to\infty}\frac{1}{k+1}=0$。因为 $\lim_{k\to\infty}\frac{\left|x^{k+1}\right|}{\left|x^k\right|}=0$，所以点列 $\{x^k\}$ 超线性收敛。

   (b) 当 k 为奇数时：$k=2m$，$x^{2m}=(\frac{1}{4})^{2^{2m}}$；
当 k 为偶数时：$k=2m+1$，$x^{2m+1}=\frac{x^{2m}}{2m+1}$。
因此，$\lim_{k\to\infty}x^k=0$ 设 $x^*=0$ 。
对于偶数 $k=2m$，$\frac{|x^{2m+1}|}{|x^{2m}|}=\frac{1}{2m+1}$，
对于奇数 $k=2m+1$，$\frac{|x^{2m+2}|}{|x^{2m+1}|}=\frac{(\frac{1}{4})^{2^{2m+2}}}{\frac{(\frac{1}{4})^{2^{2m}}}{2m+1}}=(2m+1)(\frac{1}{4})^{2^{2m+2}-2^{2m}}$ 。
$\lim_{k\to\infty}\frac{|x^{k+1}|}{|x^k|}=0$，所以 $\{x^k\}$ 超线性收敛。

2. (a) 已知 $f(x)=x_1^2+x_2^2$，$x^k=(1+\frac{1}{2^k})(\cos k,\sin k)^T$，
则 $f(x^k)=(1+\frac{1}{2^k})^2(\cos^2 k+\sin^2 k)=(1+\frac{1}{2^k})^2$。$\lim_{k\to\infty}f(x^k)=1$。
$\frac{|f(x^{k+1})-1|}{|f(x^k)-1|}=\frac{(1+\frac{1}{2^{k+1}})^2-1}{(1+\frac{1}{2^k})^2-1}=\frac{1+\frac{1}{2^k}+\frac{1}{2^{2k+2}}-1}{1+\frac{1}{2^{k-1}}+\frac{1}{2^{2k}}-1}=\frac{\frac{1}{2^k}+\frac{1}{2^{2k+2}}}{\frac{1}{2^{k-1}}+\frac{1}{2^{2k}}}\lim_{k\to\infty}\frac{|f(x^{k+1})-1|}{|f(x^k)-1|}=\frac{1}{2}$，
所以 $\{f(x^{k+1})\}$ 线性收敛（Q - 线性收敛）。

   (b) 由于 $\lim_{k\to\infty}x^k=\lim_{k\to\infty}(1+\frac{1}{2^k})(\cos k,\sin k)^T$，由于 $\cos k$ 和 $\sin k$ 是周期函数，所以 $\{x^{k+1}\}$ 不收敛。

3. 矩阵 A 的奇异值分解为 $A=U\Sigma V^T$，其中 U 和 V 是正交矩阵，$\Sigma=\mathrm{diag}(\sigma_1,\sigma_2,\cdots,\sigma_r,0,\cdots,0)$，$\sigma_1\geq\sigma_2\geq\cdots\geq\sigma_r>0$。
对于任意 $\|x\|_2=1$，有 $\|Ax\|_2=\left\|U\Sigma V^Tx\right\|_2=\left\|\Sigma V^Tx\right\|_2$。
令 $y=V^Tx$，则 $\|y\|_2=1$，$\|\Sigma y\|_2=\sqrt{\sum_{i=1}^n\sigma_i^2y_i^2}\leq\sigma_1$。
当 $y=(1,0,\cdots,0)^T$ 时，$\|\Sigma y\|_2=\sigma_1$。
所以 $\sigma_1(A)=\max_{\|x\|_2=1}\|Ax\|_2$。

4. 证明：

(a)

$$\|AB\|_F^2 = \sum_{i=1}^{n} \|Ab_i\|_2^2$$

因为 $\|Ab_i\|_2 \leq \|A\|_2\|b_i\|_2$，所以：

$$\|AB\|_F^2 \leq \sum_{i=1}^{n} \|A\|_2^2\|b_i\|_2^2 = \|A\|_2^2 \sum_{i=1}^{n} \|b_i\|_2^2 = \|A\|_2^2\|B\|_F^2$$

两边开平方，得到：

$$\|AB\|_F \leq \|A\|_2\|B\|_F$$

(b) 对矩阵 A 和 B 进行奇异值分解：

$$A = U\Sigma V^T, \quad B = X\Gamma Y^T$$

代入内积：

$$\langle A, B\rangle = \mathrm{tr}(A^T B) = \mathrm{tr}(V\Sigma U^T X\Gamma Y^T) = \mathrm{tr}(\Gamma Y^T V\Sigma U^T X)$$

令 $W = Y^T V$，$Z = U^T X$，则：

$$\langle A, B\rangle = \mathrm{tr}(\Gamma W\Sigma Z)$$

由于 W 和 Z 是正交矩阵的乘积，它们的元素满足 $|W_{ij}| \leq 1$ 和 $|Z_{ji}| \leq 1$。因此：

$$\langle A, B\rangle = \sum_{i,j} \sigma_i(B)\sigma_j(A)W_{ij}Z_{ji} \leq \sum_{i,j} \sigma_i(B)\sigma_j(A)$$

谱范数 $\|A\|_2$ 是矩阵 A 的最大奇异值，即 $\|A\|_2 = \sigma_1(A)$。核范数 $\|B\|_*$ 是矩阵 B 的奇异值之和，即 $\|B\|_* = \sum_{i=1}^{n} \sigma_i(B)$。

因此：

$$\langle A, B\rangle \leq \sum_{i,j} \sigma_i(B)\sigma_j(A) \leq \sigma_1(A) \sum_{i=1}^{n} \sigma_i(B) = \|A\|_2\|B\|_*$$

5. 先证明 A 可逆。分块矩阵的行列式公式：$\det(A) = \det(I - BB^T)$，因为 $\|B\|_2 < 1$，BB^T 的所有特征值 λ_i 满足 $|\lambda_i| \leq \|B\|_2^2 < 1$，所以 $\det(I - BB^T) \neq 0$，A 可逆。

然后求矩阵 A 的特征值。已知 $A=\begin{bmatrix} I & B \\ B^T & I \end{bmatrix}$，设 λ 是 A 的特征值，$z=\begin{bmatrix} x \\ y \end{bmatrix}$ 是对应的特征向量，其中 $x,y\in\mathbb{R}^n$。则 $A\begin{bmatrix} x \\ y \end{bmatrix}=\lambda\begin{bmatrix} x \\ y \end{bmatrix}$，即

$$\begin{cases} x+By=\lambda x \\ B^Tx+y=\lambda y \end{cases}.$$

由第一个方程可得 $By=(\lambda-1)x$，两边左乘 B^T，得到 $B^TBy=(\lambda-1)B^Tx$。由第二个方程可得 $B^Tx=(\lambda-1)y$，则有 $B^TBy=(\lambda-1)^2y$。

同理，可得 $BB^Tx=(\lambda-1)^2x$。

这表明 $(\lambda-1)^2$ 是 B^TB 和 BB^T 的特征值。

设 σ 是 B 的奇异值，则 σ^2 是 B^TB 的特征值，我们有 $(\lambda-1)^2=\sigma^2$，所以 $\lambda=1\pm\sigma$。

根据矩阵 ℓ_2 范数的定义 $\|A\|_2=\max_{\|z\|_2=1}\|Az\|_2$，且矩阵的 ℓ_2 范数等于其最大奇异值，而矩阵的奇异值是其特征值的绝对值。因为 $\|B\|_2$ 是 B 的最大奇异值，A 的特征值为 $\lambda=1\pm\sigma$（σ 是 B 的奇异值），要使 $|\lambda|$ 最大，当取 $\lambda=1+\|B\|_2$ 时（因为 $\sigma\geq 0$，且 $\|B\|_2$ 是 B 的最大奇异值），所以 $\|A\|_2=1+\|B\|_2$。

接着求 A^{-1}。先求 A 的逆矩阵，已知 $A=\begin{bmatrix} I & B \\ B^T & I \end{bmatrix}$，利用分块矩阵求逆公式

$$\begin{bmatrix} M & N \\ P & Q \end{bmatrix}^{-1}=\begin{bmatrix} M^{-1}+M^{-1}N(Q-PM^{-1}N)^{-1}PM^{-1} & -M^{-1}N(Q-PM^{-1}N)^{-1} \\ -(Q-PM^{-1}N)^{-1}PM^{-1} & (Q-PM^{-1}N)^{-1} \end{bmatrix}$$

对于 A，$M=I$，$N=B$，$P=B^T$，$Q=I$，则

$$A^{-1}=\begin{bmatrix} (I-BB^T)^{-1} & -(I-BB^T)^{-1}B \\ -B^T(I-BB^T)^{-1} & (I-B^TB)^{-1} \end{bmatrix}.$$

然后求 A^{-1} 的特征值，设 μ 是 A^{-1} 的特征值，由 A 的特征值 λ 与 A^{-1} 的特征值 μ 的关系得 $\mu=\frac{1}{\lambda}$。

已知 A 的特征值为 $\lambda=1\pm\sigma$，要使 $|\mu|$ 最大，因为 $\|B\|_2<1$，$\sigma\in[0,\|B\|_2]$，则 $|\mu|$ 最大时，λ 取最小的非零值 $1-\|B\|_2$（因为 $\lambda=1\pm\sigma$，当 $\sigma=\|B\|_2$ 且取 $\lambda=1-\|B\|_2$ 时，$\frac{1}{|\lambda|}$ 最大），所以 $\|A^{-1}\|_2=\frac{1}{1-\|B\|_2}$。

综上，

$$\|A\|_2 \left\|A^{-1}\right\|_2 = \frac{1+\|B\|_2}{1-\|B\|_2}$$

6. 证明 $\langle A,B\rangle \geq 0$（$A,B$ 为半正定矩阵）因为 A 是半正定矩阵，存在正交矩阵 U 使得 $A=U\Lambda U^T$，其中 $\Lambda=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$，$\lambda_i\geq 0$，$i=1,\cdots,n$。$\langle A,B\rangle=\mathrm{Tr}(A^TB)=\mathrm{Tr}(U\Lambda U^TB)=\mathrm{Tr}(\Lambda U^TBU)$ 令 $C=U^TBU$，C 也是半正定矩阵，其对角线元素非负。$\langle A,B\rangle=\mathrm{Tr}(\Lambda C)=\sum_{i=1}^n\lambda_i c_{ii}\geq 0$

7. 具体推导见课件

   (a) $\nabla f(X)=ab^T$

   (b) $\nabla f(X)=(A+A^T)X$

   (c) $\nabla f(X)=X^{-T}$

8. (a) 1.-4. 套用定义即可得证。

   (b) 5. 设 $(x_1,t_1),(x_2,t_2)\in\mathcal{C}$，即 $t_1^2\geq x_1^Tx_1,t_1\geq 0$ 且 $t_2^2\geq x_2^Tx_2,t_2\geq 0$。对于任意 $\lambda\in[0,1]$，令 $(x,t)=\lambda(x_1,t_1)+(1-\lambda)(x_2,t_2)=(\lambda x_1+(1-\lambda)x_2,\lambda t_1+(1-\lambda)t_2)$。

$$\begin{aligned}t^2&=(\lambda t_1+(1-\lambda)t_2)^2\\&=\lambda^2t_1^2+2\lambda(1-\lambda)t_1t_2+(1-\lambda)^2t_2^2\\&\geq\lambda^2x_1^Tx_1+2\lambda(1-\lambda)\sqrt{x_1^Tx_1}\sqrt{x_2^Tx_2}+(1-\lambda)^2x_2^Tx_2\end{aligned}$$

   根据柯西 - 施瓦茨不等式 $(\lambda x_1+(1-\lambda)x_2)^T(\lambda x_1+(1-\lambda)x_2)\leq\lambda^2x_1^Tx_1+2\lambda(1-\lambda)\sqrt{x_1^Tx_1}\sqrt{x_2^Tx_2}+(1-\lambda)^2x_2^Tx_2$，所以 $t^2\geq x^Tx$ 且 $t=\lambda t_1+(1-\lambda)t_2\geq 0$。所以 $(x,t)\in\mathcal{C}$，故双曲锥是凸集。

9. (a) 套用定义即可得证。

   (b) 设 $x_1,x_2\in C'$，则 $x_1,x_2\in C$ 且

$$g^Tx_1+h=0,\ g^Tx_2+h=0.$$

   对于任意 $\theta\in[0,1]$，令 $x=\theta x_1+(1-\theta)x_2$

$$g^Tx+h=\theta(g^Tx_1+h)+(1-\theta)(g^Tx_2+h)=0$$

令 $A' = A + \lambda g g^T$，则 A' 半正定。定义 $q(x) = x^T A x + b^T + c$，则 $q(x_1) \leq 0$，$q(x_2) \leq 0$。有

$$
\begin{aligned}
q(x) =& \theta^2 x_1^T A x_1 + (1-\theta)^2 x_2^T A x_2 + 2\theta(1-\theta) x_1^T A x_2 + \theta b^T x_1 + (1-\theta) b^T x_2 + c \\
=& \theta q(x_1) + (1-\theta) q(x_2) + \theta(1-\theta)(x_1 - x_2)^T A (x_1 - x_2),
\end{aligned}
$$

$$
\begin{aligned}
(x_1 - x_2)^T A (x_1 - x_2) &= (x_1 - x_2)^T (A' - \lambda g g^T)(x_1 - x_2) \\
&= (x_1 - x_2)^T A' (x_1 - x_2) - \lambda (g^T (x_1 - x_2))^2
\end{aligned}
$$

由于 A' 半正定，$(x_1 - x_2)^T A' (x_1 - x_2) \geq 0$，且 $q(x_1) \leq 0$，$q(x_2) \leq 0$
所以 $q(x) \leq 0$，$x \in C'$，C' 是凸集。

10. 所有函数均为凸函数。

(a) $f^*(y) = \sum_{i=1}^n e^{y_i - 1}$。

(b) $f^*(Y) = \begin{cases} -n - \ln\det(-Y), & Y \prec 0 \\ +\infty, & \text{otherwise} \end{cases}$。

(c) $f^*(y) = \begin{cases} 0, & y_i \leq 0, \sum_{i=1}^n y_i = 1 \\ +\infty, & \text{otherwise} \end{cases}$。

(d) $f^*(y, s) = \begin{cases} -\ln(1 - y^T y - s^2) - 1, & y^T y + s^2 < 1 \\ +\infty, & \text{otherwise} \end{cases}$。