Date. Page.

Assignment 11

1. $\nabla f(x) = \beta \|x\|_2^{\beta-2} x$

$$\nabla^2 f(x) = \beta(\beta-2)\|x\|_2^{\beta-4} x x^T + \beta \|x\|_2^{\beta-2} I$$

由Sherman-Morrison-Woodbury公式得

$$[\nabla^2 f(x)]^{-1} = \frac{1}{\beta\|x\|_2^{\beta-2}} I - \frac{\beta-2}{\beta(\beta-1)\|x\|_2^{\beta}} \cdot x x^T$$

则 Newton direction $= -\nabla^2 f(x)^{-1} \nabla f(x)$

$$= -x + \frac{(\beta-2)x}{\beta-1}$$

$$= -\frac{1}{\beta-1} x$$

故 $x_{k+1} = x_k - \frac{1}{\beta-1} x_k = \frac{\beta-2}{\beta-1} x_k$

显然问题的最优解为 $x^* = 0$

$$\|x_{k+1} - x^*\|_2 = \|x_{k+1}\|_2 = \left|\frac{\beta-2}{\beta-1}\right| \cdot \|x_k\|_2$$

$$\therefore \frac{\|x_{k+1}\|_2}{\|x_k\|_2} = \left|\frac{\beta-2}{\beta-1}\right|$$

(a) 若 $\beta > 1$ 且 $\beta \neq 2$,

当 $\left|\frac{\beta-2}{\beta-1}\right| < 1$ 时, $\beta > \frac{3}{2}$ 且 $\beta \neq 2$

由此可知, 若 $\beta > \frac{3}{2}$ 且 $\beta \neq 2$, 则收敛是线性的

而若 $1 < \beta \leq \frac{3}{2}$, $\left|\frac{\beta-2}{\beta-1}\right| \geq 1$, 是发散的 (这里我怀疑题目出了问题, 于是编写了matlab代码进行验证, 验证后确实发现在 $1 < \beta \leq \frac{3}{2}$ 时发散, 而非题目所说的收敛).

Date.　　　　Page.

令 $x=[a,0,0,\cdots,0]^T,(a>0)$

则 $\nabla^2 f(x)=diag\{\beta(\beta-1)a^{\beta-2},\beta a^{\beta-2},\beta a^{\beta-2},\cdots,\beta a^{\beta-2}\}$

取 $m>0$

$$\nabla^2 f(x)-mI=diag\{\beta(\beta-1)a^{\beta-2}-m,\beta a^{\beta-2}-m,\cdots,\beta a^{\beta-2}-m\}$$

当 $\frac{3}{2}<\beta<2$ 时，对 $\forall m>0$

取 $a>\left[\frac{m}{\beta(\beta-1)}\right]^{\frac{1}{\beta-2}}$　则 $\beta(\beta-1)a^{\beta-2}-m<0$

此时 $\nabla^2 f(x)-mI$ 非半正定

当 $\beta>2$ 时，取 $a<\left[\frac{m}{\beta(\beta-1)}\right]^{\frac{1}{\beta-2}}$，则 $\beta(\beta-1)a^{\beta-2}-m<0$

此时 $\nabla^2 f(x)-mI$ 非半正定

$\therefore$ 不存在 $m>0$，使得对于 $\forall x$，$\nabla^2 f(x)-mI\succeq 0$

$\therefore$ $f(x)$ 不是强凸函数

$\therefore$ 没有局部二次收敛

(b) 若 $0<\beta<1$，则 $\left|\frac{\beta-2}{\beta-1}\right|>1$　$\therefore$ 发散

2. (a)　$f'(x)=\alpha x^{\alpha-1}-e^x$　　$f''(x)=\alpha(\alpha-1)x^{\alpha-2}-e^x$

$$x_{k+1}=x_k-\frac{\alpha x_k^{\alpha-1}-e^{x_k}}{\alpha(\alpha-1)x_k^{\alpha-2}-e^{x_k}}$$

(b) 取 $\alpha=2$，初始点 $x_0=1$

则 $x_{k+1}=x_k-\frac{2x_k-e^{x_k}}{2-e^{x_k}}=\frac{(1-x_k)e^{x_k}}{2-e^{x_k}}$

$x_1=0$，$x_2=1$，$x_{2k}=1$；$x_{2k+1}=0$，$k=1,2,\cdots$

$\therefore$ $\{x_k\}$ 并不像工程师所想那样发散至无穷

(c) 由于 Hessian 矩阵 $\alpha(\alpha-1)x^{\alpha-2}-e^{x}$ 显然不是半正定的，故目标函数 $f(x)$ 不是凸的。牛顿法的收敛性是局部的，这意味着往往会收敛到离初始点最近的局部极小值点，而不是工程师所认为的会收敛到一个非凸函数的全局最小值点（即发散到 ∞）

（非凸函数的局部极小值点不一定是全局最小值点）