优化方法第四章作业

1. 总结等式约束凸优化问题的求解方法及相互联系

1. 解:

对于等式限制凸优化问题:

$minf(x)$

$s.t.Ax = b$

1. 消除方法: 消除等式约束, 转化为等价的无约束问题

仿射可行解集 $\{x|Ax = b\} = \{Fz + \hat{x}|z \in \mathbb{R}^{n-p}\} where \hat{x}$ 是满足 $A\hat{x} = b$ 的任意特解

消除矩阵 $F \in \mathbb{R}^{nx(n-p)}$ 是值域 $R(F)$ 为 A 的零空间的任意矩阵

消除等式约束后的优化问题: $\min\limits_{z\in\mathbb{R}^{n-p}} \hat{f}(Fz + \hat{x}) \Leftrightarrow min\{f(x)|s.t.Ax = b\}$

最优解 z^* 确定原等式约束问题的最优解 $x^* = Fz^* + \hat{x}$

牛顿方向:$d_z = -(\nabla^2\widetilde{f}(z))^{-1}\nabla\widetilde{f}(z) = -(F^T\nabla^2 f(x)F)^{-1}F^T\nabla f(x)$

与对偶函数联系: $Fd_z = d_x, \hat{f}$ 在 z 处的牛顿减少量 $\hat{\lambda}(z)$ 与 f 在 x 处的牛顿减少量 $\lambda(x)$ 相等

2. 对偶方法: 采用无约束优化方法求解对偶问题, 然后从对偶问题中复原等式约束优化问题的解

对偶函数: $g(v) = -b^Tv + \inf\limits_x(f(x) + v^TAx) = -b^Tv - \sup\limits_x((-A^Tv)^Tx - f(x)) = -b^Tv - f^*(-A^Tv)$

对偶问题: $\max\limits_x -b^Tv - f^*(-A^Tv)$ 该问题存在最优解, Slater 条件满足, 从而强对偶性成立,

故 $g(v^*) = p^*$

这样就将问题转化为了无约束最优化问题:

$\min\limits_x f(x) + (v^*)^T(Ax - b) \Leftrightarrow \nabla f(x) + A^Tv^* = 0$

$s.t.A(x + v) = b$

故 $v^* = -(AA^T)^{-1}A\nabla f(x^*)$ 为原等式约束问题的一个最优对偶变量

采用无约束凸优化问题牛顿方法求解

目标函数 $f(x)$ 在 x 附近的二阶 Taylor 近似模型: $\min\limits_v \hat{f}(x+v) = f(x) + \nabla f(x)^Tv + \frac{1}{2}v^T\nabla^2 f(x)v$

$Ax^* - b, \nabla f(x^*) + A^Tv^* = 0$, 因为 $x^* = x + d_x, v^* = w$ 故

$Ad_x = 0, \nabla^2 f(x) d_x + A^T w = -\nabla f(x)$ 即

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} d_x \\ w \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix}$$

d_x 为 x 处的牛顿方向

等式约束优化问题牛顿减少量 $\lambda(x) = (d_x^T \nabla^2 f(x) d_x)^{\frac{1}{2}}$

函数 f 在 x 处的二次展开 $\hat{f}(x+v) = f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f(x) v$

$f(x) - p^* \approx f(x) - inf(\hat{f}(x+v) : A(x+v) = b) = \frac{1}{2}(\lambda(x))^2$

f 沿着 d_x 的方向导数：$-\lambda(x)^2$

等式约束的牛顿方法:

初始化可行点 $x^0, Ax^0 = b,$

确定牛顿方向和牛顿减少量:

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} d_x \\ w \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix}$$

$\lambda(x)^2 = d_x^T \nabla^2 f(x) d_x$

停止准则 $\frac{1}{2}\lambda(x^k)^2 \leq \epsilon$

回溯直线搜索 $0 < \alpha < 0.5, 0 < \beta < 1, t^k = 1$

$while(f(x^k + t^k d_x^k) > f(x^k) + \alpha t^k \nabla f(x^k)^T d_x^k = f(x^k) - \alpha t^k \lambda(x^k)^2)$

$t^k = \beta t^k$

不可行初始点的牛顿方向:

找到一个在当前不可行点出牛顿方向 d 使得 $x + d$ 近似满足最优性条件，即 $x + d \approx x^*$

由 $Ax^* = b, \nabla f(x^*) + A^T v^* = 0$ 得 $A(x+d) = b, \nabla f(x^*) + \nabla f^2(x) d + A^T w = 0$

即

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} d \\ w \end{bmatrix} = - \begin{bmatrix} \nabla f(x) \\ Ax - b \end{bmatrix}$$

原对偶残差:

$$Let \quad y = \begin{bmatrix} x \\ v \end{bmatrix}$$

$$Then \quad r(y) = \begin{bmatrix} \nabla f(x) + A^T v \\ Ax - b \end{bmatrix} \triangleq \begin{bmatrix} r_{dual}(y) \\ r_{pri}(y) \end{bmatrix}$$

这样，原等式约束凸优化问题优化过程等价于：给定任意 y, 找到一个下降方向 d_y, 使得 $y + d_y$ 满足最优性条件，即 $r(y + d_y) = 0$

原对偶残差在当前估计点 y 附近的一阶 $Taylor$ 近似为:$r(y+z) \approx r(y) + Dr(y)z$

$$Dr(y) = \begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix}$$

$r(y) + Dr(y)dy = 0 \Rightarrow$

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} d_x \\ d_v \end{bmatrix} = - \begin{bmatrix} \nabla f(x) + A^T v \\ Ax - b \end{bmatrix}$$

于是有:$d_y = -Dr(y)^{-1} r(y)$

不可行初始点的牛顿方向

$$\begin{bmatrix} d \\ w \end{bmatrix} = \begin{bmatrix} d_x \\ v + d_v \end{bmatrix}$$

不可行初始点牛顿下降回溯直线搜索方法:

初始化可行点 $x^0 \in domf, v^0$,

停止准则 $Ax^k = b \quad and \quad \|r(x^k, v^k)\|_2 \leq \epsilon$

确定原对偶牛顿方向:

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} d_x^k \\ d_v^k \end{bmatrix} = -\begin{bmatrix} -\nabla f(x^k) + A^T v^k \\ Ax^k - b \end{bmatrix}$$

对原对偶残差 $\|r\|_2$ 进行回溯，确定步长 t^k

$while(\|r(x^k + t^k d_x^k, v^k + t^k d_v^k)\|_2 > (1 - \alpha t^k)\|r(x^k, v^k)\|_2)$

$t^k = \beta t^k$

$x^{k+1} = x^k + t^k d_x^k, v^{k+1} = v^k + t^k d_v^k$ 返回检测停止准则

为止初始点牛顿方法的收敛性分析:

阻尼牛顿阶段: $\|r(y^k)\|_2 > \frac{1}{K^2L}$, $\|r(y + td_y)\|_2 \leq (1 - \alpha t)\|r(y)\|_2$

二次收敛阶段: $\|r(y^k)\|_2 \leq \frac{1}{K^2L}$, $\frac{K^2L\|r(y^{l+k})\|_2}{2} \leq (\frac{K^2L\|r(y^l)\|_2}{2})^{2^k} \leq (\frac{1}{2})^{2^k}$

2. 等式约束熵极大化

等式约束熵极大化问题：$minf(x) = \sum_{i=1}^{n} x_i log x_i$

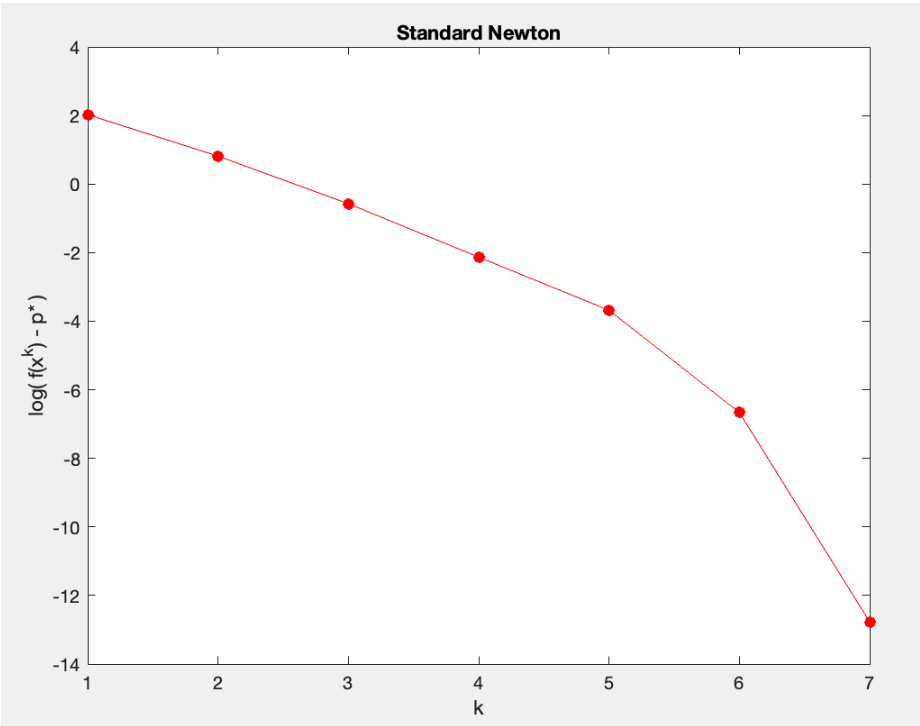
subjected to $Ax = b_i$

采用以下方法计算该问题的解：

(a) 标准 Newton 方法:

```
 Standard Newton: times = 7
res = -33.365147
```

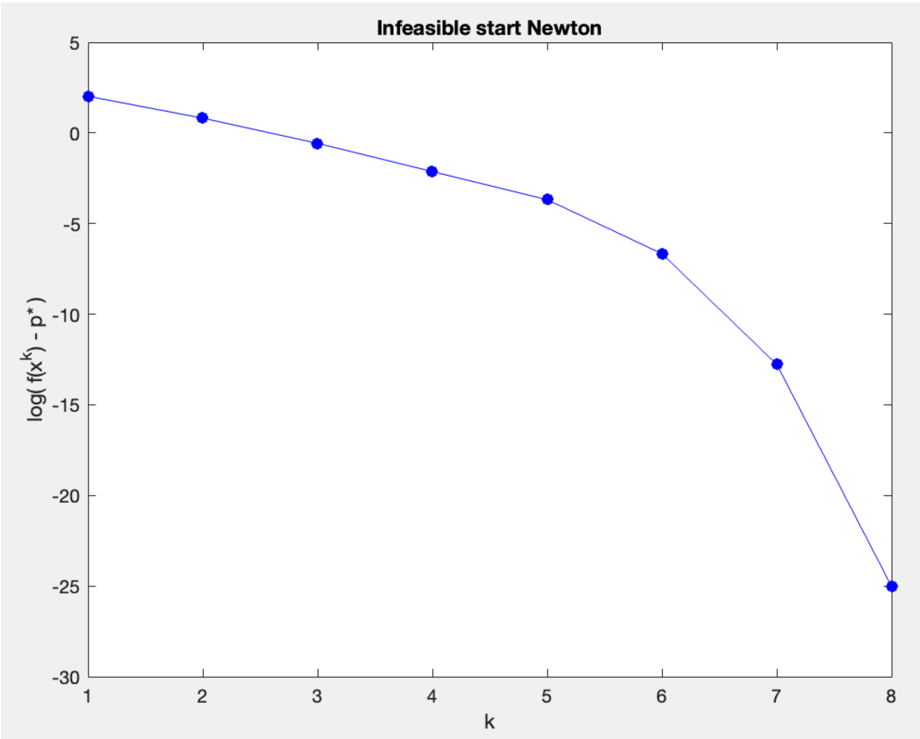
以 $log(f(x^k) - p^*)$ 为纵轴，k 为横轴画图：



(b) 不可行初始点 Newton 方法:

```
 Infeasible start Newton: times = 8
res = -33.365147
```

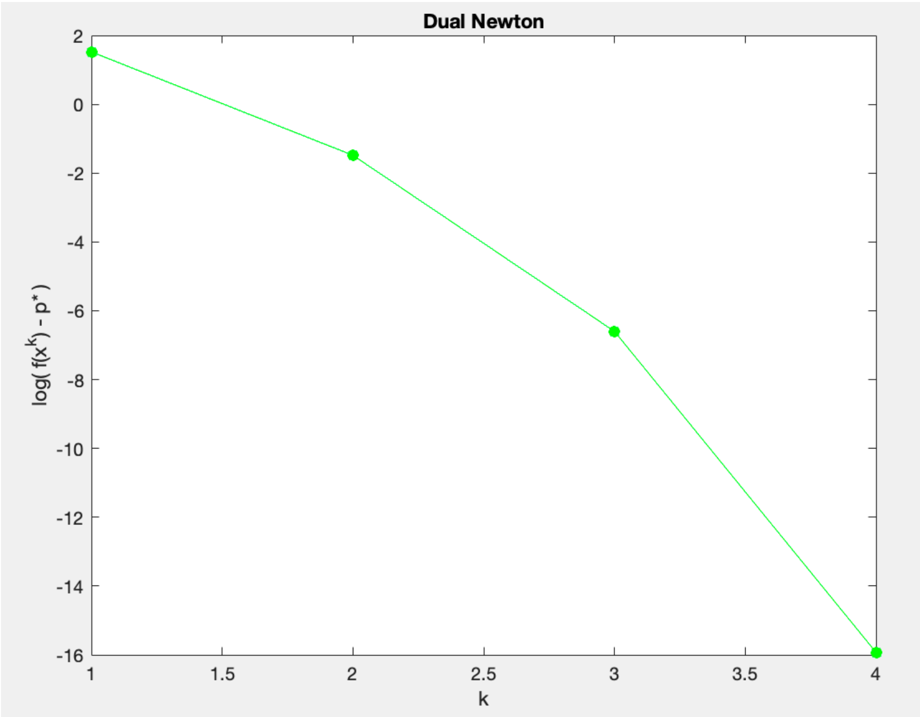
以 $log(f(x^k) - p^*)$ 为纵轴，k 为横轴画图：



(c) 对偶 newton 方法:

```
Dual Newton: times = 5
res = -33.365147
```

以 $log(L(v^k) - d^*)$ 为纵轴，k 为横轴画图：



由代码输出及图像得，起始点相同时，三种方法的结果相同。

在标准牛顿方法和不可信初始点牛顿方法中，计算系数矩阵

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix}$$

来解决问题。其中 $\nabla^2 f(x) = \mathbf{diag}(x)^{-1}$

这样系数矩阵就可以表示为 $A\mathbf{diag}(x)A^T$

在对偶牛顿方法中，计算系数矩阵 $-\nabla^2 g(v) = ADA^T$, 这里 $D_{ii} = e^{-a_i^T v - 1}$