POM-HW5-Jasenv@CC98

Solution 1:

$$\min_{x \in R^n} f(x) = \frac{1}{2} r(x)^{\mathrm{T}} r(x) = \frac{1}{2} \sum_{i=1}^{m} [r_i(x)]^2$$

By derivation,we get the gradient and hessian matrix of $f(x)$.

$$\nabla f(x) = \nabla r(x)^{\mathrm{T}} r(x) = \sum_{i=1}^{m} r_i(x) \nabla r_i(x)$$

$$\nabla^2 f(x) = \nabla r(x)^{\mathrm{T}} \nabla r(x) + \sum_{i=1}^{m} r_i(x) \nabla^2 r_i(x) = M(x) + S(x)$$

So equivalently,this problem reduces to prove

$$S(x) = \sum_{i=1}^{m} r_i(x) \nabla^2 r_i(x) \to O, x \to x^*$$

For this specific problem:

$$r_1(x) = x_2 - x_1^2, r_2(x) = 1 - x_2$$

then,

$$\nabla^2 r_1(x) = \begin{pmatrix} -2 & 0 \\ 0 & 0 \end{pmatrix}, \nabla^2 r_2(x) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

so,

$$\begin{aligned} S(x) &= \left(x_2 - x_1^2\right) \cdot \begin{pmatrix} -2 & 0 \\ 0 & 0 \end{pmatrix} + (1 - x_2) \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} -2\left(x_2 - x_1^2\right) & 0 \\ 0 & 0 \end{pmatrix} \to O, x \to x^* = (1,1)^T \end{aligned}$$

Solution 2:

Define two functions:

$$x(\mu) = -\left(A^T A + \mu I\right)^{-1} A^T r, \quad \delta(x) = \|Ax + r\|^2$$

For $x(\mu)$,

$$x'(\mu) = \left(A^T A + \mu I\right)^{-2} A^T r$$

For $\delta(x)$, use the known condition: $A^T Ax = -A^T r - \mu x$

$$\delta(x) = (Ax + r)^T (Ax + r) = x^T A^T Ax + 2r^T Ax + r^T r$$

$$\nabla\delta(x) = 2\left(A^T Ax + A^T r\right) = 2\left(-A^T r - \mu x + A^T r\right) = -2\mu x$$

By the chain principle of composition function's derivation, $\mu > 0$

$$\begin{aligned}\delta'(x(\mu)) = (\nabla\delta(x))^T x'(\mu) &= -2\mu x^T \left(A^T A + \mu I\right)^{-2} A^T r \\ &= 2\mu x^T \left(A^T A + \mu I\right)^{-1} \left[-\left(A^T A + \mu I\right)^{-1} A^T r\right] \\ &= 2\mu x^T \left(A^T A + \mu I\right)^{-1} x > 0\end{aligned}$$

So, $\delta(x(\mu))$ is the increasing function of μ, $\mu_1 > \mu_2 > 0$,
$\delta\left(x\left(\mu_2\right)\right) < \delta\left(x\left(\mu_1\right)\right)$, $i.e.$ $\|Ax_2 + r\|^2 < \|Ax_1 + r\|^2$.