

IOE 611: Homework 6

Yulun Zhuang

December 9, 2024

Problem 1

Trust region Newton method. If $\nabla^2 f(x)$ is singular (or very ill-conditioned), the Newton step $\Delta x_{\text{nt}} = -\nabla^2 f(x)^{-1} \nabla f(x)$ is not well defined. Instead we can define a search direction Δx_{tr} as the solution of

$$\begin{aligned} & \text{minimize} && \frac{1}{2} v^\top H v + g^\top v \\ & \text{subject to} && \|v\|_2 \leq \gamma, \end{aligned}$$

where $H = \nabla^2 f(x)$, $g = \nabla f(x)$, and γ is a positive constant. The point $x + \Delta x_{\text{tr}}$ minimizes the second-order approximation of f at x , subject to the constraint that $\|(x + \Delta x_{\text{tr}}) - x\|_2 \leq \gamma$. The set $\{v \mid \|v\|_2 \leq \gamma\}$ is called the *trust region*. The parameter γ , the size of the trust region, reflects our confidence in the second-order model.

Show that Δx_{tr} minimizes

$$\frac{1}{2} v^\top H v + g^\top v + \hat{\beta} \|v\|_2^2,$$

for some $\hat{\beta}$. This quadratic function can be interpreted as a regularized quadratic model for f around x .

Proof. The Lagrangian of the original optimization is

$$\mathcal{L}(v, \lambda) = \frac{1}{2} v^\top H v + g^\top v + \lambda (\|v\|_2 - \gamma)$$

The KKT conditions are

$$\lambda \geq \mathbf{0} \tag{1}$$

$$\|v\|_2 \leq \gamma \tag{2}$$

$$\lambda (\|v\|_2 - \gamma) = \mathbf{0} \tag{3}$$

$$\nabla \mathcal{L} = (H + 2\lambda I)v - g = \mathbf{0} \tag{4}$$

Case 1: $H \succ 0$.

Given $H + 2\lambda I > 0$ for all $\lambda \succ 0$, we can solve for $v^* = (H + 2\lambda I)^{-1}g$.

Note that $\|v^*\|_2$ is decreasing with respect to λ , we know that $\min_{\lambda \geq 0} \|v^*\|_2 = \|H^{-1}g\|_2$.

If $\|H^{-1}g\|_2 > \gamma$, we must have $\lambda^* > 0$ to satisfy (2), therefore by (3) we have $\|(H + 2\lambda^*I)^{-1}g\|_2 = \gamma$.

Let $\hat{\beta} = \lambda^*$, we know that v^* minimizes $\frac{1}{2}v^\top H v + g^\top v + \hat{\beta}\|v\|^2$ since

$$\frac{\partial}{\partial v} \left(\frac{1}{2}v^\top H v + g^\top v + \hat{\beta}\|v\|^2 \right) = (H + 2\hat{\beta}I)v + g = (H + 2\lambda^*I)v + g.$$

If $\|H^{-1}g\| \leq \gamma$, then (3) will be satisfied if and only if $\lambda = 0$. Similarly we know that v^* minimizes $\frac{1}{2}v^\top H v + g^\top v$.

Case 2: H is singular.

If $\lambda^* > 0$, then $(H + 2\lambda^*I) \succ 0$, and $\gamma = \|(H + 2\lambda^*I)^{-1}g\|_2$.

Let $\hat{\beta}$ be the solution of the above equation, then v^* minimizes $\frac{1}{2}v^\top H v + g^\top v + \hat{\beta}\|v\|^2$.

If $\lambda^* = 0$, then by (4) we have $Hv^* = g$, i.e. $g \in \text{Range}(H)$, $v^* = H^\dagger g$ and by (2) we have $r \geq \|v^*\|_2 = \|H^\dagger g\|_2$.

Therefore, v^* minimizes $\frac{1}{2}v^\top H v + g^\top v$.

□

Problem 2

Equality constrained entropy maximization. Consider the equality constrained entropy maximization problem

$$\begin{aligned} & \text{minimize} && f(x) = \sum_{i=1}^n x_i \log x_i \\ & \text{subject to} && Ax = b, \end{aligned}$$

with $\text{dom } f = \mathbb{R}_{++}^n$ and $A \in \mathbb{R}^{p \times n}$, with $p < n$.

Generate a problem instance with $n = 100$ and $p = 30$ by choosing A randomly (checking that it has full rank), choosing \hat{x} as a random positive vector (e.g., with entries uniformly distributed on $[0, 1]$) and then setting $b = A\hat{x}$. (Thus, \hat{x} is feasible.)

Compute the solution of the problem using the following methods.

- (a) *Standard Newton method.* You can use initial point $x^{(0)} = \hat{x}$.
- (b) *Infeasible start Newton method.* You can use initial point $x^{(0)} = \hat{x}$ (to compare with the standard Newton method), and also the initial point $x^{(0)} = \mathbf{1}$.
- (c) *Dual Newton method,* i.e. the standard Newton method applied to the dual problem.

Verify that the three methods compute the same optimal point (and Lagrange multiplier). Compare the computational effort per step for the three methods, assuming relevant structure is exploited. (Your implementation, however, does not need to exploit structure to compute the Newton step.)

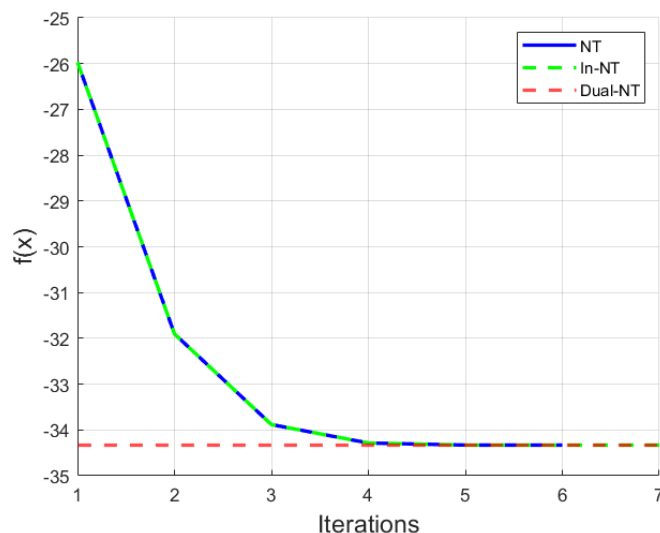


Figure 1: Comparisons of three Newton's methods

Solution. The computational effort is identical across all three methods. In both standard and infeasible start Newton methods, we solve equations with the coefficient matrix

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

where $\nabla^2 f(x) = \mathbf{diag}(x)^{-1}$. Using block elimination, these equations reduce to a system with the coefficient matrix $A \mathbf{diag}(x) A^\top$.

In the dual method, the equations involve a coefficient matrix of the form $-\nabla^2 g(\lambda) = ADA^\top$, where $D = \mathbf{diag}(\exp(-A^\top \lambda - 1))$.

Thus, in all three methods, the primary computational task in each iteration is solving a linear system of the form $A^\top D A z = -g$, where D is a diagonal matrix with positive diagonal elements.

□

Problem 3

Tangent to central path. This problem concerns $dx^*(t)/dt$, which gives the tangent to the central path at the point $x^*(t)$. For simplicity, we consider a problem without equality constraints; the results readily generalize to problems with equality constraints.

(a) Find an explicit expression for $dx^*(t)/dt$.

(b) Show that $f_0(x^*(t))$ decreases as t increases. Thus the objective value in the barrier method decreases, as the parameter t is increased. (We already know that the duality gap, which is m/t decreases as t increases.)

Solution. (a) Given the centrality equation

$$0 = t\nabla f_0(x^*) + \sum_{i=1}^m \frac{1}{-f_i(x^*)} \nabla f_i(x^*) + A^\top \hat{v},$$

Differentiate it with respect to t , we have

$$\begin{aligned} 0 = & \nabla f_0(x^*) + t\nabla^2 f_0(x^*) \frac{dx^*}{dt} \\ & + \sum_{i=1}^m \left[\frac{1}{f_i^2(x^*)} \nabla f_i(x^*) \nabla f_i(x^*)^\top \frac{dx^*}{dt} \right. \\ & \left. - \frac{1}{f_i(x^*)} \nabla^2 f_i(x^*) \frac{dx^*}{dt} \right], \end{aligned}$$

which gives the explicit expression

$$\frac{dx^*}{dt} = - \left[t\nabla^2 f_0(x^*) - \sum_{i=1}^m \left(\frac{1}{f_i(x^*)} \nabla^2 f_i(x^*) - \frac{1}{f_i^2(x^*)} \nabla f_i(x^*) \nabla f_i(x^*)^\top \right) \right]^{-1} \nabla f_0(x^*).$$

(b) To show that $f_0(x^*)$ is monotonically decreasing, we can show that $\frac{d}{dt}f_0(x^*) < 0$.

From (a), we know that $\frac{dx^*}{dt} = -\mathcal{G}^{-1}\nabla f_0(x^*)$, where $\mathcal{G} \succ 0$.

Therefore

$$\begin{aligned} \frac{d}{dt}f_0(x^*) &= \nabla f_0(x^*)^\top \frac{dx^*}{dt} \\ &= -\nabla f_0(x^*)^\top \mathcal{G}^{-1}\nabla f_0(x^*) \\ &< 0. \end{aligned}$$

□