Convex Optimizations – Homework 1

Q1.

1. Given:

$f(x)$ is strongly Convex and twice differentiable with bounded Hessian $mI \leq \nabla^2 f(x) \leq MI$, $\kappa = \frac{M}{m}$

Since, $f(x)$ is convex and M-strongly smooth,

$$f(x+p) \leq f(x) + \nabla f(x)^T p + \frac{M}{2}\|p\|^2 \tag{1}$$

For gradient descent optimization,

$$f(x_{t+1}) = \min_{\eta} f\big(x_t - \eta \nabla f(x_t)\big)$$

Using the property (1) of strong smoothness, and $p = -\eta \nabla f(x_t)$, we get

$$f(x_{t+1}) \leq \min_{\eta} f(x_t) + \nabla f(x_t)^T(-\eta \nabla f(x_t) + \frac{M}{2}\|\eta \nabla f(x_t)\|^2)$$

$$\leq \min_{\eta} f(x_t) + \|\nabla f(x_t)\|^2(-\eta + \tfrac{M}{2}\eta^2)$$

Since, $\eta = \frac{1}{M}$, we get

$$f(x_{t+1}) = f(x_t) - \frac{1}{2M}\|\nabla f(x_t)\|^2 \tag{2}$$

Since $f(x)$ is m-strongly convex,

We have,

$$f(x_t) - p^* \leq \tfrac{1}{2m}\|\nabla f(x_t)\|^2 \tag{3}$$

Using (2) and (3), we have,

$$f(x_{t+1}) \leq f(x_t) - \frac{2m}{2M}(f(x_t) - p^*)$$

$$f(x_{t+1}) - p^* \leq (f(x_t) - p^*)(1 - \frac{m}{M})$$

$$\leq \left(1 - \frac{m}{M}\right)^{t+1} (f(x_0) - p^*) \leq \epsilon$$

∴ The function will be ϵ sub-optimal after at most T iterations of the algorithm, where T is

$$T = \frac{1}{\log(\frac{\kappa}{\kappa - 1})} \log(\frac{f(x_0) - p^*}{\epsilon})$$

Since the step size $\eta = \frac{1}{M}$ here, which is a constant, we don't have to find the best step size in every iteration. Hence, no function evaluations are required at this point.

Once we have the step size, we need to perform the gradient evaluation once for every step.

∴ We need to perform T gradient evaluations and 0 function evaluations for this algorithm.

2. Consider the function
   $\min_x f(x) = 2x^2 - 5$
   with $\eta = 2$ and $x_0 = 1$

$\therefore \nabla f(x) = 4x$ and direction $\Delta x = -\nabla f(x) = -4x$

Assume $\epsilon = 0$ and we stop when $\nabla f(x) = 0$

$$x_1 = x_0 + \eta \Delta x$$

$$\therefore x_1 = 1 + 2 * (-4 * 1) = -7 \ ; f(x_1) = -3$$

$\because \nabla f(x_1) = 4 \neq 0$, we perform the next iteration

$$\therefore x_2 = -7 + 2 * (-4 * -7) = 49 \ ; f(x_2) = 93$$

$$\nabla f(x_2) = -343 \neq 0$$

We know $x^* = 0 \ and \ f(x^*) = -5$, since the algorithm directly jumped from one side of the minima to the other side of minima and η is constant at $\eta = 2$, we can conclude that it won't converge at $x_t = x^*$.

We know that if a function is M-Strongly smooth, $\eta \leq \frac{1}{M}$ and the best step size is $\eta = \frac{1}{M}$

$\therefore$ If we want to decide a fixed step size for all steps, it depends on the Hessian of the function.

2. Newton's Method

1. To Prove: $\Delta x = A\Delta y$

Given

$x = Ay + b$

$g(y) = f(Ay + b)$, Δx and Δy are the Newton Steps of $f(x)$ and $g(y)$.

We know that, the Newton step for $f(x)$ is given by,

$$\Delta x = -[\nabla^2 f(x)]^{-1}\, \nabla f(x) \tag{1}$$

$$\begin{aligned}
\therefore \Delta y &= -[\nabla^2 g(y)]^{-1}\, \nabla g(y) \\
&= -[\mathrm{A}^{\mathrm{T}}\nabla^2 f(Ay + b)A]^{-1} \,.\, A^T\, \nabla f(Ay + b) \\
&= -[\mathrm{A}^{\mathrm{T}}\, \nabla^2 f(x)A]^{-1}\ .\, A^T \nabla f(x) \\
&= -A^{-1}[\nabla^2 f(x)]^{-1}(A^T)^{-1}A^T\nabla f(x) \\
&= -A^{-1}[\nabla^2 f(x)]^{-1}\nabla f(x) \\
&= A^{-1}(\Delta x) \qquad\qquad \text{... from (1)}
\end{aligned}$$

$$\therefore \Delta x = A\Delta y \tag{2}$$

Hence, proved.

2. **To Prove:** For any $\eta > 0$, the exit condition for backtracking linesearch on $f(x)$ in direction of Δx will hold if and only if the exit condition holds for $g(y)$ for Δy

Consider the exit condition for $f(x)$

$$f(x + \eta\Delta x) \leq f(x) + \alpha\eta\left(\nabla f(x)\right)^T \Delta x$$

$\because x = Ay + b$ … [given] and $\Delta x = A\Delta y$ … [from (2)]

$$f(x + \eta\Delta x) \leq f(Ay + b) + \alpha\eta(\nabla f(Ay + b)^{\mathrm{T}})(\mathrm{A}\Delta y)$$

$$f((Ay + b) + \eta A\Delta y) \leq f(Ay + b) + \alpha\eta\nabla f(Ay + b)^{\mathrm{T}}(\mathrm{A}\Delta y)$$

$$f(A(y + \eta\Delta y) + b) \leq f(Ay + b) + \alpha\eta\nabla f(Ay + b)^{\mathrm{T}}(\mathrm{A}\Delta y)$$

$$\because g(y) = f(Ay + b) \; and \; \nabla g(y)^T = \nabla f(Ay + b)^T A$$

$$g(y + \eta\Delta y) \leq g(y) + \alpha\eta\nabla g(y)^T A^{-1}. A\Delta y$$

$$g(y + \eta\Delta y) \leq g(y) + \alpha\eta\nabla g(y)^T \Delta y$$

$\therefore$ The exit condition for $f(x)$ for Δx holds if and only if the exit condition holds for $g(y)$ for Δy.

3. Given

$$x^{(0)} = Ay^{(0)} + b$$

We need to prove that $x^{(k)} = Ay^{(k)} + b$ and $f\left(x^{(k)}\right) = g(y^{(k)})$

Let's look at it as an Induction problem,
∴ Let's assume hypothesis,

$$x^{(n)} = Ay^{(n)} + b \tag{3}$$

When we run the Newton Algorithm on $g(.)$ starting at $y^{(0)}$, we get next positions of y as

$$y^{(k)}, where\ k = \{1,2,3 \ldots n\}$$
$$Ay^{(n+1)} + b = A\left[y^{(n)} + \eta\Delta y^{(n)}\right] + b$$

From definition of Newton's step, $\Delta y = -\left(\nabla^2 g\left(y^{(n)}\right)\right)^{-1} \Delta \mathrm{g}\left(\mathrm{y}^{(\mathrm{n})}\right)$

$$\therefore Ay^{(n+1)} + b = Ay^{(n)} + \eta A\left[-\left(\nabla^2 g\left(y^{(n)}\right)\right)^{-1} \Delta \mathrm{g}\left(\mathrm{y}^{(\mathrm{n})}\right)\right] + b$$
$$Ay^{(n+1)} + b = Ay^{n} + \eta A\left[-A^{-1}\left(\nabla^2 f\left(Ay^{(n)} + b\right)\right)^{-1} (\mathrm{A}^{\mathrm{T}})^{-1}.\mathrm{A}^{\mathrm{T}}\Delta \mathrm{f}\left(A\mathrm{y}^{(\mathrm{n})} + b\right)\right] + b$$
$$= Ay^{(n)} - \eta A.A^{-1}\left(\nabla^2 f\left(x^{(n)}\right)\right)^{-1} \nabla f\left(x^{(n)}\right) + b \qquad \text{... [from (3)]}$$
$$= \left(Ay^{(n)} + b\right) - \eta A.A^{-1}\left(\nabla^2 f\left(x^{(n)}\right)\right)^{-1} \nabla f\left(x^{(n)}\right)$$
$$= x^{(n)} + \eta\Delta x^{(n)}$$
$$= x^{(n+1)}$$

$\therefore Ay^{(n+1)} + b = x^{(n+1)}$

$\therefore By\ Induction,$

$$x^{(k)} = Ay^{(k)} + b \tag{4}$$

$$\because g(y) = f(Ay + b)$$
$$g\left(y^{(k)}\right) = f\left(fAy^{(k)} + b\right)$$
$$from\ (4)\ we\ get,$$
$$\therefore \quad g\left(y^{(k)}\right) = f\left(x^{(k)}\right) \tag{5}$$

4. The Newton decrement for $f(x)$ is defined as

$$\lambda(x) = \left(\nabla f(x)^T [\nabla^2 f(x)]^{-1} \nabla f(x)\right)^{\frac{1}{2}}$$

$$\lambda^2(x) = \nabla f(x)^T [\nabla^2 f(x)]^{-1} \nabla f(x) \tag{6}$$

Similarly, for $g(y)$ we have,

$$\begin{aligned}
\lambda^2(y) &= \nabla g(y)^T [\nabla^2 g(y)]^{-1} \nabla g(y) \\
&= \nabla f(Ay+b)^T A[A^T \nabla^2 f(Ay+b)A]^{-1} \mathrm{A^T} \nabla f(Ay+b) && \text{... [given]} \\
&= \nabla f(x)^T A[A^T \nabla^2 f(x)A]^{-1} \mathrm{A^T} \nabla f(x) && \because x = Ay + b\ [given] \\
&= \nabla f(x)^T A. A^{-1} \left(\nabla^2 f(x)\right)^{-1} (A^T)^{-1} A^T \nabla f(x)
\end{aligned}$$

$$\begin{aligned}
\therefore \lambda^2(y) &= \nabla f(x)^T \left(\nabla^2 f(x)\right)^{-1} \nabla f(x) \\
&= \lambda^2(x) && \text{... [from (6)]}
\end{aligned}$$

$\therefore The\ Newton\ decrement\ for\ f(.)\ at\ x\ is\ equal\ to\ the\ Newton's\ decrement\ for\ g(.)\ at\ y$

Since the stopping criterion for the Newton's method is,

$$\frac{\lambda^2}{2} \le \epsilon$$

The stopping conditions are also identical.