# Homework 3

Yujun Kim

IE539 Convex Optimization

Nov 2023

## 1 Smoothness and Strong Convexity of a Quadratic Function

$\nabla f(x) = Ax + b, \nabla^2 f(x) = A$ for all $x$. Let $\alpha, \beta$ be the smallest and largest eigenvalue of $A$ respectively. $\beta \geq \alpha \geq 0$ as $A$ is PSD. If $A$ is positive definite, $\beta \geq \alpha > 0$. Note that for any $v$, $\alpha||v||^2 \leq v^T Av \leq \beta||v||^2$. This is clear from the eigen-decomposition of $A$: Let $A = V\Lambda V^T$ for orthonormal $V$ and diagonal $\Lambda$ with diagonal entries $\beta = \lambda_n \geq \lambda_{n-1} \geq \cdots \geq \lambda_1 = \alpha$. Let $w = V^T v$ so that $v = Vw$. Then, $||w|| = ||V^T v|| = ||v||$ and

$$v^T Av = w^T V^T V \Lambda V^T V w = w^T \Lambda w$$

As $\alpha||v||^2 = \alpha||w||^2 \leq w^T \Lambda w = \sum \lambda_t w_i^2 \leq \beta||w||^2 = \beta||v||^2$, we obtain

$$\alpha||v||^2 = v^T Av \leq \beta||v||^2$$

Now, We prove that $f$ is $\beta$-smooth and $\alpha$-strongly convex.

**(Step 1)** *$f$ is $\beta$-smooth.* By theorem11.1, it suffices to prove $g(x) = \frac{\beta}{2}||x||^2 - f(x)$ is convex.

$$\nabla^2 g(x) = \beta I - \nabla^2 f(x) = \beta I - A$$

For any $v$, $v^T \nabla^2 g(x) v = v^T(\beta I - A)v = \beta||v||^2 - v^T Av \geq \beta||v||^2 - \beta||v||^2 = 0$. The last inequality is due to $v^T Av \leq \beta||v||^2$. This shows $\nabla^2 g(x) \succeq 0$ and so $g$ is convex.

**(Step 2)** *$f$ is $\alpha$-strongly convex.* By definition we have to show $h(x) = f(x) - \frac{\alpha}{2}||x||^2$ is convex.

$$\nabla^2 h(x) = \nabla^2 f(x) - \alpha I = A - \alpha I$$

For any $v$, $v^T \nabla^2 h(x) v = v^T(A - \alpha I)v = v^T Av - \alpha||v||^2 \geq \alpha||v||^2 - \alpha||v||^2 = 0$. The last inequality is due to $\alpha||v||^2 \leq v^T Av$. This shows $\nabla^2 h(x) \geq 0$ and so $h$ is convex.

## 2 $\mathcal{O}(\frac{1}{T})$ convergence of PGD under Strong Convexity

We first state and prove two lemmas. These two lemmas extend the results held for differentiable function to non differentiable function by using subgradient.

**Lemma 1** If $f$ is $L$-Lipshictz, then for any $g \in \partial f(x)$, $||g|| \leq L$.
**(proof)** For any $y$, $f(y) \geq f(x) + g^T(y-x)$ so that $g^T(y-x) \leq f(y) - f(x) \leq |f(y) - f(x)| \leq L||x-y||$. Take $y = x + g$. Then, the inequality becomes $||g||^2 \leq L||g||$ so that $||g|| \leq L$ for $g \neq 0$(The result is trivial if $g = 0$).

**Lemma 2** If $f$ is $\alpha$-strongly convex and $g \in \partial f(x)$, then

$$f(y) \geq f(x) + g^T(y-x) + \frac{\alpha}{2}||y-x||^2$$

**(proof)** $x \mapsto f(x) - \frac{\alpha}{2}||x||^2$ is convex. Any subgradient of this function at $x$ is of the form $g - \alpha x$ for some $g \in \partial f(x)$. Thus,

$$f(y) - \frac{\alpha}{2}||y||^2 \geq f(x) - \frac{\alpha}{2}||x||^2 + (g - \alpha x)^T(y-x)$$

Rearranging the terms gives the result.

Now using the two lemmas, we derive a common iteration-wise recursion for both (a), (b). By strong convexity and lemma 2, $f(x^*) \geq f(x_t) + g_t^T(x^* - x_t) + \frac{\alpha}{2}||x^* - x_t||^2$. Thus, $g_t^T(x_t - x^*) \geq f(x_t) - f(x^*) + \frac{\alpha}{2}||x_t - x^*||^2$. By fact that a projection mapping is a contraction mapping(HW2 Problem 4 - (a)), and lemma 1,2 that follows from the Lipschitsness of $f$ and the strong convexity of $f$, we have the following.

$$\begin{aligned}
||x_{t+1} - x^*||^2 &\leq ||x_t - \eta g_t - x^*||^2 \\
&= ||x_t - x^*||^2 - 2\eta_t g_t^T(x_t - x^*) + \eta_t^2||g_t||^2 \\
&\leq ||x_t - x^*||^2 - 2\eta_t(f(x_t) - f(x^*) + \frac{\alpha}{2}||x_t - x^*||^2) + \eta_t^2 L^2 \\
&= (1 - \alpha\eta_t)||x_t - x^*||^2 + \eta_t^2 L^2 - 2\eta_t(f(x_t) - f(x^*))
\end{aligned}$$

**(a)** Thus, for $\eta_t = \frac{2}{\alpha(t+1)}$, by rearranging the recursion and multiplying $\frac{2t}{2\eta_t}$ on both side,

$$\begin{aligned}
2t(f(x_t) - f(x^*)) &\leq \frac{2t}{2\eta_t}((1-\alpha\eta_t)||x_t - x^*||^2 - ||x_{t+1} - x^*||^2 + \eta_t^2 L^2) \\
&= \frac{\alpha t(t+1)}{2}\left(\frac{t-1}{t+1}||x_t - x^*||^2 - ||x_{t+1} - x^*||^2\right) + t\eta_t L^2 \\
&\leq \frac{\alpha}{2}((t-1)t||x_t - x^*||^2 - t(t+1)||x_{t+1} - x^*||^2) + \frac{2L^2}{\alpha}
\end{aligned}$$

Sum this equation over $t = 1, \cdots, T$ to obtain

$$\sum_{t=1}^{T} 2t(f(x_t) - f(x^*)) \leq \frac{2L^2 T}{\alpha}$$

Divide each side by $T(T+1)$. Then

$$\frac{1}{T(T+1)}\sum_{t=1}^{T} 2t(f(x_t) - f(x^*)) \le \frac{2L^2}{\alpha(T+1)}$$

By the convexity of $f$ and $\frac{1}{T(T+1)}\sum_{t=1}^{T} 2t = 1$, we have

$$f\left(\sum_{t=1}^{T}\frac{2t}{T(T+1)}x_t\right) - f(x^*) \le \frac{1}{T(T+1)}\sum_{t=1}^{T} 2t(f(x_t) - f(x^*)) \le \frac{2L^2}{\alpha(T+1)}$$

**(b)** On the other hand, for $\eta_t = \frac{1}{\alpha t}$, by rearranging the recursion and multiplying $\frac{1}{\eta_t} = \alpha t$ on both side,

$$2(f(x_t) - f(x^*)) \le \alpha(t-1)||x_t - x^*||^2 - \alpha t||x_{t+1} - x^*||^2 + \eta_t L^2$$

Sum this euqation over $t = 1, \cdots, T$ to obtain

$$\sum_{t=1}^{T} 2(f(x_t) - f(x^*)) \le L^2\sum_{t=1}^{T}\eta_t = \frac{L^2}{\alpha}\sum_{t=1}^{T}\frac{1}{t} \le \frac{L^2(1+log(T))}{\alpha}$$

Note that the last inequality is due to $u \mapsto \frac{1}{u}$ is decreasing on $\mathbb{R}_{++}$ so that $\sum_{t=2}^{T}\frac{1}{t} \le \int_{u=1}^{T}\frac{1}{u}du = log(T)$. Divide both side by $2T$ and use convexity of $f$ to obtain

$$f(\frac{1}{T}\sum_{t=1}^{T} x_t) - f(x^*) \le \frac{1}{T}\sum_{t=1}^{T}(f(x_t) - f(x^*)) \le \frac{L^2(1+log(T))}{2\alpha T}$$

# 3 Online PGD under Strong Convexity and Lipschitzness

**(a)** By strong convexity of $f_t$ and lemma 2 from Problem 2,

$$f_t(x^*) \ge f_t(x_t) + g_t^T(x^* - x_t) + \frac{\alpha}{2}||x_t - x^*||^2$$

so that $f_t(x_t) - f_t(x^*) + \frac{\alpha}{2}||x_t - x^*||^2 \le g_t^T(x_t - x^*)$. Using projection is a contraction mapping, lemma 1 from Problem 2, we have

$$\begin{aligned}
||x_{t+1} - x^*||^2 &= ||proj_C\{x_t - \eta_t g_t\} - proj_C\{x^*\}||^2 \\
&\le ||x_t - \eta_t g_t - x^*||^2 \\
&= ||x_t - x^*||^2 - 2\eta_t g_t^T(x_t - x^*) + \eta_t^2||g_t||^2 \\
&\le ||x_t - x^*||^2 - 2\eta_t(f_t(x_t) - f_t(x^*) + \frac{\alpha}{2}||x_t - x^*||^2) + \eta_t^2||g_t||^2 \\
&= (1 - \alpha\eta_t)||x_t - x^*||^2 - 2\eta_t(f_t(x_t) - f_t(x^*)) + \eta_t^2||g_t||^2
\end{aligned}$$

Thus,

$$f_t(x_t) - f(x^*) \le \left(\frac{1}{2\eta_t} - \frac{\alpha}{2}\right)||x_t - x^*||^2 - \frac{1}{2\eta_t}||x_{t+1} - x^*||^2 + \frac{\eta_t}{2}||g_t||^2$$

**(b)** From (a), take $\eta_t = \frac{1}{\alpha t}$ and multiply each side by 2.

$$2(f_t(x_t) - f(x^*)) \le \alpha(t-1)||x_t - x^*||^2 - \alpha t||x_{t+1} - x^*||^2 + \eta_t L^2$$

Thus,

$$\sum_{t=1}^{T} 2(f_t(x_t) - f(x^*)) \le L^2 \sum_{t=1}^{T} \eta_t = \frac{L^2}{\alpha} \sum_{t=1}^{T} \frac{1}{t} \le \frac{L^2(1 + log(T))}{\alpha}$$

Divide each side be 2 to obtain

$$\sum_{t=1}^{T} f_t(x_t) - \min_{x \in C} \sum_{t=1}^{T} f_t(x) = \sum_{t=1}^{T} (f_t(x_t) - f_t(x^*)) \le \frac{L^2(1 + log(T))}{2\alpha}$$

# 4 SGD with Strong Convexity and Lipschitzness

**(Setting)** First note that we obtain an unbiased estimator of subgradient at each step. i.e.

$$\mathbb{E}[\hat{g_{x_t}}|x_t] =: g_t \in \partial f(x_t)$$

and have update rule as $x_{t+1} = proj_C\{x_t - \eta_t \hat{g_{x_t}}\}$.

**(Step 1)** *Iteration-wise recursion.*

$$\begin{aligned} ||x_{t+1} - x^*||^2 &\le ||x_t - \eta_t \hat{g_{x_t}} - x^*||^2 \\ &= ||x_t - x^*||^2 - 2\eta_t \hat{g_{x_t}}^T (x_t - x^*) + \eta_t^2 ||\hat{g_{x_t}}||^2 \\ &\le ||x_t - x^*||^2 - 2\eta_t \hat{g_{x_t}}^T (x_t - x^*) + \eta_t^2 L^2 \end{aligned}$$

Take conditional expectation given $x_t$ on both side. Using lemma 2 from Problem 2 that follows from the strong convexity,

$$\begin{aligned} \mathbb{E}[||x_{t+1} - x^*||^2|x_t] &\le ||x_t - x^*||^2 - 2\eta_t g_t^T (x_t - x^*) + \eta_t^2 L^2 \\ &\le ||x_t - x^*||^2 - 2\eta_t (f(x_t) - f(x^*) + \frac{\alpha}{2}||x_t - x^*||^2) + \eta_t^2 L^2 \\ &= (1 - \alpha\eta_t)||x_t - x^*||^2 - 2\eta_t (f(x_t) - f(x^*)) + \eta_t^2 L^2 \end{aligned}$$

Rearrange the terms and multiply each side by $\frac{1}{\eta_t} = \alpha t$.

$$2(f(x_t) - f(x^*)) \le \alpha(t-1)||x_t - x^*||^2 - \alpha t \mathbb{E}[||x_{t+1} - x^*||^2|x_t] + \eta_t L^2$$

Take expectation over $x_t$ to obtain

$$2\mathbb{E}[f(x_t) - f(x^*)] \le \alpha(t-1)\mathbb{E}[||x_t - x^*||^2] - \alpha t \mathbb{E}[||x_{t+1} - x^*||^2] + \eta_t L^2$$

**(Step 2)** *Convergence bonund.* Sum the above equation over $t = 1, \cdots, T$ and divide each side by $2T$.

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[f(x_t) - f(x^*)] \leq \frac{L}{2T}\sum_{t=1}^{T}\eta_t = \frac{L}{2\alpha T}\sum_{t=1}^{T}\frac{1}{t} \leq \frac{L(1 + log(T))}{2\alpha T}$$

By convexity, $f(\sum_{t=1}^{T}\frac{1}{T}x_t) - f(x^*) \leq \frac{1}{T}\sum_{t=1}^{T}(f(x_t) - f(x^*)$. The inequality remains even if we take expectation on both side. Thus,

$$\mathbb{E}[f(\sum_{t=1}^{T}\frac{1}{T}x_t) - f(x^*)] \leq \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[f(x_t) - f(x^*)] \leq \frac{L(1 + log(T))}{2\alpha T}$$