5.9 Strong convexity and smoothness

Recap :

- $\|w\|_1 \geq \|w\|_2 \geq \cdots \geq \|w\|_\infty$
- the dual norm of $\|\cdot\|$ is $\|x\|_* = \sup_{\|y\| \leq 1} (x \cdot y)$
- $\|\cdot\|_p$ and $\|\cdot\|_q$ are dual to each other when $\frac{1}{p} + \frac{1}{q} = 1$

[Definition 29] strong convexity | smoothness.

A function f is α-strongly convex w.r.t. a norm $\|\cdot\|$ iff for all w, u:

$$D_f(w \| u) \geq \frac{\alpha}{2} \|w-u\|^2$$

A function f is α-strongly smooth w.r.t. a norm $\|\cdot\|$ iff for all w, u:

$$D_f(w \| u) \leq \frac{\alpha}{2} \|w-u\|^2$$

[Lemma 8] strong convexity and strong smoothness.

The following two statements are equivalent.

① $\psi(w)$ is $1/\eta$-strongly convex w.r.t. $\|\cdot\|$.

② $\psi^*(\theta)$ is η-strongly smooth w.r.t. $\|\cdot\|_*$.

Pf: We only need to prove ① ⇒ ② : (since ② ⇒ ① is analogons)

Assume $D_\psi(w \| u) \geq \frac{1}{2\eta} \|w-u\|^2$ for $\forall w, u$

$$\Rightarrow \quad \psi(w) - \psi(u) - \nabla\psi(u) \cdot (w-u) \geq \frac{1}{2\eta} \|w-u\|^2 \qquad (\mathrm{I})$$

and $$\psi(u) - \psi(w) - \nabla\psi(w) \cdot (u-w) \geq \frac{1}{2\eta} \|u-w\|^2 \qquad (\mathrm{II})$$

(I) + (II) :

$$(\nabla\psi(w) - \nabla\psi(u)) \cdot (w-u) \geq \frac{1}{\eta} \|w-u\|^2 \qquad (\mathrm{III})$$

Let $\theta_1 = \nabla\psi(w)$, $\theta_2 = \nabla\psi(u)$.

then $w = \nabla\psi^*(\theta_1)$, $u = \nabla\psi^*(\theta_2)$. (Ⅲ) can be rewritten:

$$(\theta_1-\theta_2)\cdot(\nabla\psi^*(\theta_1)-\nabla\psi^*(\theta_2)) \geq \frac{1}{\eta}\|\nabla\psi^*(\theta_1)-\nabla\psi^*(\theta_2)\|^2$$

take L_1-norm, and divide both sides by $\frac{1}{\eta}\|\nabla\psi^*(\theta_1)-\nabla\psi^*(\theta_2)\|$

$$\|\nabla\psi^*(\theta_1)-\nabla\psi^*(\theta_2)\| \leq \eta\left|(\theta_1-\theta_2)\cdot\frac{\nabla\psi^*(\theta_1)-\nabla\psi^*(\theta_2)}{\|\nabla\psi^*(\theta_1)-\nabla\psi^*(\theta_2)\|}\right| \leq \eta\|\theta_1-\theta_2\|_* \quad (*)$$

Since ψ^* is convex, $\psi^*(\theta_1)-\psi^*(\theta_2)-\nabla\psi^*(\theta_2)\cdot(\theta_1-\theta_2)\geq 0$

By $(*)$ we have

$$\begin{aligned}
&\psi^*(\theta_1)-\psi^*(\theta_2)-\nabla\psi^*(\theta_2)\cdot(\theta_1-\theta_2)\\
&= \left|\int_0^1 \nabla\psi^*(\theta_2+t(\theta_1-\theta_2))\cdot(\theta_1-\theta_2)\,dt - \nabla\psi^*(\theta_2)\cdot(\theta_1-\theta_2)\right|\\
&\leq \int_0^1 \|\nabla\psi^*(\theta_2+t(\theta_1-\theta_2))-\nabla\psi^*(\theta_2)\|\cdot\|\theta_1-\theta_2\|_*\,dt \quad \text{(by duality)}\\
&\leq \int_0^1 \eta t\|\theta_1-\theta_2\|_*^2\,dt\\
&= \frac{\eta}{2}\|\theta_1-\theta_2\|_*^2
\end{aligned}$$

, which completes the proof $\square$

[Theorem 32] regret of OMG using norms

Suppose ψ is a $\frac{1}{\eta}$-strongly convex regularizer

$$\text{Regret}(u) \leq [\psi(u)-\psi(w_1)] + \frac{\eta}{2}\sum_{t=1}^{T}\|z_t\|_*^2$$

Pf: By Lemma 8, ψ^* is η-strongly smooth.

Note that $\theta_{t+1} = \theta_t - z_t$, we have

$$D_{\psi^*}(\theta_{t+1}\|\theta_t) \leq \frac{\eta}{2}\|z_t\|_*^2.$$

By Theorem 31, we have

$$\begin{aligned}
\text{Regret}(u) &\leq [\psi(u)-\psi(w_1)] + \sum_{t=1}^{T} D_{\psi^*}(\theta_{t+1}\|\theta_t)\\
&\leq [\psi(u)-\psi(w_1)] + \frac{\eta}{2}\sum_{t=1}^{T}\|z_t\|_*^2 \qquad \square
\end{aligned}$$

* Learning with expert advice.

Recall when using quadratic regularizer, we got

$$\text{Regret} \sim \sqrt{dT}$$

To reduce it, just use another norm $\|z\|_\infty \le 1$, now

$$\text{Regret} \sim \sqrt{T} \qquad (\text{by Th.32})$$

but this means ψ should be strongly convex w.r.t. L_1, which is harder since $(\|\cdot\|_1 \geq \|\cdot\|_2)$. $\psi(w) = \frac{1}{2\eta}\|w\|_2^2$ is only $\frac{1}{\eta d}$-strongly convex w.r.t. L_1 norm. [however, entropy is $\frac{1}{\eta}$-SC.].

[Example 36] exponentiated gradient (EG)

- $\psi(w) = \frac{1}{\eta}\sum_{j=1}^{d} w_j \log w_j$ for $w \in \Delta_d$
- Recall $\psi^*(w) = \frac{1}{\eta}\log\sum_{j=1}^{d} e^{\eta\theta_j}$ and $\nabla\psi^*(\theta_j) = \frac{e^{\eta\theta_j}}{\sum_{k=1}^{d} e^{\eta\theta_k}}$
- OMD updates : $w_{t,j} \propto e^{\eta\theta_{t,j}}$

The equivalent recursive formula:

$$w_{t+1,j} \propto w_{t,j}\, e^{-\eta z_{t,j}}$$

[Example 37] EG for learning with expert advise

- $f_t(w) = w \cdot z_t$
- $\|z_t\|_\infty \le 1$ (dual norm)
- $\max_w \psi(w) = 0$, $\min_w \psi(w) = \frac{\log(1/d)}{\eta}$

Then $\text{Regret} \le \frac{\log(d)}{\eta} + \frac{\eta T}{2} \le \sqrt{2\log(d)\,T}$