Exercices lecture 6

0.1 Exercise 1

$$
\begin{aligned}
\theta_t &= \theta_{t-1} - \gamma \nabla_f(\theta_{t-1}) \\
&= \theta_{t-1} - \gamma \frac{1}{n} X^T (X\theta_{t-1} - y) \\
&= \theta_{t-1} - \gamma (H\theta_{t-1} - \frac{1}{n} X^T y) \\
&= \theta_{t-1} - \gamma (H\theta_{t-1} - H\eta^* \\
&= \theta_{t-1} - \gamma H(\theta_{t-1} - \eta^*)
\end{aligned}
\tag{1}
$$

0.2 Exercise 3

Let X be an eigenvector of H with eigenvalue λ. Then,

$$
\begin{aligned}
(I - \gamma H)^{2t} X &= (I - \gamma H)^{2t-1} (I - \gamma H) X \\
&= (I - \gamma H)^{2t-1} (X - \gamma H X) \\
&= (I - \gamma H)^{2t-1} (1 - \gamma \lambda) X \\
&= (1 - \gamma \lambda)(I - \gamma H)^{2t-1} X \\
&= (1 - \gamma \lambda)^{2t} X
\end{aligned}
\tag{2}
$$

Hence $(1 - \gamma\lambda)^{2t}$ is an eigenvalue of $(I - \gamma H)^{2t}$. However this does not show the inverse property. It is better to exploit the fact that H is symmetric and real, hence there exists $P \in GL_n(\mathbb{R})$ and a diagonal matrix D containing the eigenvalues of H, such that

$$
H = PDP^{-1} \tag{3}
$$

Hence

$$
\begin{aligned}
I - \gamma H &= I - \gamma PDP^{1} \\
&= P(I - \gamma D)P^{-1}
\end{aligned}
\tag{4}
$$

and

$$
\begin{aligned}
(I - \gamma H)^{2t} &= \left(P(I - \gamma D)P^{-1}\right)^{2t} \\
&= P(I - \gamma D)P^{-1} P(I - \gamma D)P^{-1} \dots P(I - \gamma D)P^{-1} \\
&= P(I - \gamma D)^{2t} P^{-1}
\end{aligned}
\tag{5}
$$

But $(I - \gamma D)^{2t}$ is a diagonal matrix with values of the form $(1 - \gamma\lambda)^{2t}$ on the diagonal. We can conclude that the eigenvalues of $(I - \gamma D)^{2t}$ are exactly the $(1 - \gamma\lambda)^{2t}$.

0.3 Exercise 4

If λ is an eigenvalue of H, then

$$\mu \leqslant \lambda \leqslant L \tag{6}$$

Hence

$$1-\gamma L \leqslant 1-\gamma\lambda \leqslant 1-\gamma\mu \tag{7}$$

And

$$-(1-\gamma\mu) \leqslant -(1-\gamma\lambda) \leqslant -(1-\gamma L) \tag{8}$$

We have $|1-\gamma\lambda| = \max\Big((1-\gamma\lambda), -(1-\gamma\lambda)\Big)$.
With 7,

$$(1-\gamma\lambda) \leqslant 1-\gamma\mu \leqslant |1-\gamma\mu| \tag{9}$$

With 8,

$$-(1-\gamma\lambda) \leqslant 1-\gamma L \leqslant |1-\gamma L| \tag{10}$$

Finally,

$$|1-\gamma\lambda| \leqslant \max\left(|1-\gamma\mu|, |1-\gamma L|\right) \tag{11}$$

With $\gamma = \frac{1}{L}$,

- $|1-\gamma\mu| = |1-\frac{\mu}{L}| = (1-\frac{\mu}{L})$
- $|1-\gamma L| = 0$.

and

$$\max_{\lambda\in[\mu,L]} |1-\gamma\lambda| \leqslant (1-\frac{\mu}{L}) = (1-\frac{1}{\kappa}) \tag{12}$$

0.4 Exercise 6

Let $g(\alpha) = \alpha\exp(-\alpha)$. We can differentiate g and $g'(\alpha) = e^{-\alpha}(1-\alpha)$. Thus g is increasing on $[0,1]$ and decreasing on $[1,+\infty]$ and thus maximu is attained at $\alpha = 1$, with $g(1) = \frac{1}{e} \leqslant \frac{1}{2}$.

0.5 Exercise 7

We know that for all θ,

$$\nabla_\theta f = \frac{1}{n} X^T(X\theta - y) \tag{13}$$

Hence,

$$\begin{aligned} ||\nabla_\theta f - \nabla_{\theta'} f|| &= ||\frac{1}{n}(X^T(X\theta-y) - X^T(X\theta'-y))|| \\ &= \frac{1}{n}||X^T X(\theta-\theta')|| \\ &= \frac{1}{n}|||X^T X||| \times ||(\theta-\theta')|| \end{aligned} \tag{14}$$

0.6 Exercise 3