

FTML Exercices 5

Pour le 4 mai 2023

1 HEAVY-BALL

Exercice 1 :

We have that

$$\begin{aligned}\theta_{t+1} - \eta^* &= \theta_t - \eta^* - \gamma \nabla_{\theta} f(\theta_t) + \beta(\theta_t - \theta_{t-1}) \\ &= \theta_t - \eta^* - \gamma \left(H\theta_t - \frac{1}{n} X^T y \right) + \beta(\theta_t - \theta_{t-1}) \\ &= \theta_t - \eta^* - \gamma(H\theta_t - H\eta^*) + \beta(\theta_t - \theta_{t-1}) \\ &= (I_d - \gamma H)(\theta_t - \eta^*) + \beta((\theta_t - \eta^*) - (\theta_{t-1} - \eta^*))\end{aligned}\tag{1}$$

Hence,

$$\begin{aligned}\langle \theta_{t+1} - \eta^*, u_{\lambda} \rangle &= \langle (I_d - \gamma H)(\theta_t - \eta^*) + \beta((\theta_t - \eta^*) - (\theta_{t-1} - \eta^*)), u_{\lambda} \rangle \\ &= \langle (I_d - \gamma H)(\theta_t - \eta^*), u_{\lambda} \rangle + \beta \langle \theta_t - \eta^*, u_{\lambda} \rangle - \beta \langle \theta_{t-1} - \eta^*, u_{\lambda} \rangle \\ &= \langle \theta_t - \eta^*, (I_d - \gamma H)u_{\lambda} \rangle + \beta \langle \theta_t - \eta^*, u_{\lambda} \rangle - \beta \langle \theta_{t-1} - \eta^*, u_{\lambda} \rangle \\ &= \langle \theta_t - \eta^*, (1 - \gamma\lambda)u_{\lambda} \rangle + \beta \langle \theta_t - \eta^*, u_{\lambda} \rangle - \beta \langle \theta_{t-1} - \eta^*, u_{\lambda} \rangle \\ &= (1 - \gamma\lambda + \beta) \langle \theta_t - \eta^*, u_{\lambda} \rangle - \beta \langle \theta_{t-1} - \eta^*, u_{\lambda} \rangle\end{aligned}\tag{2}$$

which is the expected result. We have used that $I_d - \gamma H$ is symmetric and that u_{λ} is an eigenvector of H .

Exercice 2 : The characteristic polynomial is

$$P(X) = X^2 - (1 - \gamma\lambda + \beta)X + \beta\tag{3}$$

In order to find the values of the sequence $(a_t)_{t \in \mathbb{R}n}$ we need to know the roots of the equation

$$X^2 - (1 - \gamma\lambda + \beta)X + \beta = 0\tag{4}$$

The discriminant writes

$$\begin{aligned}\Delta &= (1 - \gamma\lambda + \beta)^2 - 4\beta \\ &= 1 + (\gamma\lambda)^2 + \beta^2 - 2\gamma\lambda + 2\beta - 2\gamma\lambda\beta - 4\beta \\ &= 1 + (\gamma\lambda)^2 + \beta^2 - 2\gamma\lambda - 2\gamma\lambda\beta - 2\beta \\ &= 1 + \left(\frac{4}{(\sqrt{L} + \sqrt{\mu})^2} \lambda \right)^2 + \left(\frac{\sqrt{L} - \sqrt{\mu}}{\sqrt{L} + \sqrt{\mu}} \right)^4 \\ &\quad - 2 \frac{4}{(\sqrt{L} + \sqrt{\mu})^2} \lambda - 2 \frac{4}{(\sqrt{L} + \sqrt{\mu})^2} \lambda \left(\frac{\sqrt{L} - \sqrt{\mu}}{\sqrt{L} + \sqrt{\mu}} \right)^2 - 2 \left(\frac{\sqrt{L} - \sqrt{\mu}}{\sqrt{L} + \sqrt{\mu}} \right)^2\end{aligned}\tag{5}$$

If we use the common denominator $(\sqrt{L} + \sqrt{\mu})^4$, the numerator N writes

$$N = (\sqrt{L} + \sqrt{\mu})^4 + 16\lambda^2 + (\sqrt{L} - \sqrt{\mu})^4 - 8\lambda(\sqrt{L} + \sqrt{\mu})^2 - 8\lambda(\sqrt{L} - \sqrt{\mu})^2 - 2(\sqrt{L} + \sqrt{\mu})^2(\sqrt{L} - \sqrt{\mu})^2 \quad (6)$$

But

$$\begin{aligned} (\sqrt{L} + \sqrt{\mu})^4 + (\sqrt{L} - \sqrt{\mu})^4 - 2(\sqrt{L} + \sqrt{\mu})^2(\sqrt{L} - \sqrt{\mu})^2 &= \left((\sqrt{L} + \sqrt{\mu})^2 - (\sqrt{L} - \sqrt{\mu})^2 \right)^2 \\ &= \left((L + \mu + 2\sqrt{L\mu}) - (L + \mu - 2\sqrt{L\mu}) \right)^2 \\ &= \left(4\sqrt{L\mu} \right)^2 \\ &= 16L\mu \end{aligned} \quad (7)$$

And

$$\begin{aligned} -8\lambda(\sqrt{L} + \sqrt{\mu})^2 - 8\lambda(\sqrt{L} - \sqrt{\mu})^2 &= -8\lambda \left((\sqrt{L} + \sqrt{\mu})^2 + (\sqrt{L} - \sqrt{\mu})^2 \right) \\ &= -8\lambda \left(2(L + \mu) \right) \\ &= -16\lambda(L + \mu) \end{aligned} \quad (8)$$

Finally,

$$\begin{aligned} N &= 16\lambda^2 + 16\lambda\mu - 16\lambda(L + \mu) \\ &= 16(L - \lambda)(\mu - \lambda) \\ &\leq 0 \end{aligned} \quad (9)$$

As $\Delta \leq 0$, the roots of equation 4 are $a \pm ib$, with $a, b \in \mathbb{R}$.

- their module ρ verifies $\rho^2 = a^2 + b^2$
- their arguments are y and $-y \in \mathbb{R}$.

The sequence $(a_t)_{t \in \mathbb{R}_n}$ is of the form

$$\rho^t (A \cos(ty) + B \sin(ty)) \quad (10)$$

with A and B real constants. The convergence rate of a_t is thus fully determined by ρ . It is thus sufficient to show that $\rho < 1$. We know that

$$b = \frac{1}{2} \sqrt{-\Delta} \quad (11)$$

Hence

$$b^2 = -\frac{1}{4} \Delta \quad (12)$$

We also have that

$$a = -\frac{1}{2} (1 - \gamma\lambda + \beta) \quad (13)$$

Hence

$$a^2 = \frac{1}{4} (1 - \gamma\lambda + \beta)^2 \quad (14)$$

We also have that

$$\Delta = (1 - \gamma\lambda + \beta)^2 - 4\beta \quad (15)$$

Finally

$$\begin{aligned}
 \rho &= \sqrt{a^2 + b^2} \\
 &= \sqrt{\frac{1}{4} \left((1 - \gamma\lambda + \beta)^2 - \Delta \right)} \\
 &= \sqrt{\beta}
 \end{aligned} \tag{16}$$

Exercise 3 :

We have that

$$\begin{aligned}
 \sqrt{\beta}^{2t} &\leq \left(1 - \frac{1}{\sqrt{\kappa}}\right)^{2t} \\
 &\leq \exp\left(-\frac{2t}{\sqrt{\kappa}}\right)
 \end{aligned} \tag{17}$$

By a decomposition in a basis of eigenvectors of H (which exists as H is symmetric and real), we obtain that $\theta_t - \eta^*$ converges to 0 at least at an exponential rate of characteristic time $\sqrt{\kappa}$.