Heavy Ball solution

1 HEAVY-BALL

Exercice 1 :
We have that

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

Hence,

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

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

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

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

$$\begin{aligned} N = &(\sqrt{L}+\sqrt{\mu})^4 + 16\lambda^2 + (\sqrt{L}-\sqrt{\mu})^4 \\ &- 8\lambda(\sqrt{L}+\sqrt{\mu})^2 - 8\lambda(\sqrt{L}-\sqrt{\mu})^2 - 2(\sqrt{L}+\sqrt{\mu})^2(\sqrt{L}-\sqrt{\mu})^2 \end{aligned} \tag{6}$$

But

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

And

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

Finally,

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

As $\Delta \leqslant 0$, 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

$$\rho^t (A\cos(ty) + B\sin(ty)) \tag{10}$$

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

$$b = \frac{1}{2}\sqrt{-\Delta} \tag{11}$$

Hence

$$b^2 = -\frac{1}{4}\Delta \tag{12}$$

We also have that

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

Hence

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

We also have that

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

Finally

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

Exercice 3 :
We have that

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

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 $\sqrt{\kappa}$.