Exercice 2

1. $$e^{-m_\theta} = e^{-\min_i Y_i X_i^T \theta}$$
$$= \frac{1}{n}\sum_{j=1}^{n} e^{\underbrace{-\min_i Y_i X_i^T\theta}_{\geq -Y_j X_j^T \theta \quad \forall j \in [\![1,n]\!]}}$$
$$\geq \underbrace{\frac{1}{n}\sum_{i=1}^{n} e^{-Y_i X_i^T\theta}}_{= e^{-m_\theta} + \underbrace{\sum_{\substack{i=1 \\ i \neq \arg\min_j Y_j X_j^T\theta}}^{n} e^{-Y_i X_i^T \theta}}_{>0}}$$
$$\geq \frac{1}{n} e^{-m_\theta}$$

Donc $e^{-m_\theta} \geq \frac{1}{n}\sum_{i=1}^{n} e^{-Y_i X_i^T\theta} \geq \frac{1}{n} e^{-m_\theta}$

en passant au log,
$$-\min_i Y_i X_i^T \theta \geq \underbrace{\log\left(\frac{1}{n}\sum_{i=1}^{n} e^{-Y_i X_i^T\theta}\right)}_{-F(\theta)} \geq -\log(n) - \min_i Y_i X_i^T\theta$$
$$\min_i Y_i X_i^T \theta \leq F(\theta) \leq \min_i Y_i X_i^T\theta + \log(n)$$

2. a) $\gamma := \max_{\|\theta\|_2 \leq 1} \min_{1\leq i \leq n} Y_i X_i^T \theta$

- Δ_n est un n-simplexe, donc pour un $p \in \Delta_n$ donné $\begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$ - i-ème indice, $p^T Z = Y_i X_i^T$. Donc trouver $j \in [\![1,n]\!]$ tq. $Y_j X_j^T \theta$ minimise $Y_i X_i^T \theta$ revient à trouver $p' \in \Delta_n$ tq. $p'^T Z$ minimise $p^T Z$.
Donc $\gamma = \max_{\|\theta\|_2 \leq 1} \min_{p \in \Delta_n} p^T Z \theta$.

- $\max_{\|\theta\|_2\leq 1} \min_{p\in\Delta_n} p^T Z\theta = \min_{p\in\Delta_n} \max_{\|\theta\|_2 \leq 1} p^T Z \theta$ découle directement du thm. minimax dans le cas particulier des fonctions bilinéaires.

b) $\gamma := \max_{\|\theta\|_2 \leq 1} \min_i Y_i X_i^T \theta = e_i^T Z\theta$

$= \max_{\|\theta\|_2\leq 1} \min_{p \in \Delta_n} p^T Z \theta$ (since Δ_n is the convex hull of $\{e_i\}_{i=1,\dots,n}$)

$= \min_{p\in\Delta_n} \max_{\|\theta\|_2 \leq 1} p^T Z\theta$ by the minimax thm. for bilinear fcts in game theory.

Therefore, $\gamma = \min_{p \in \Delta_n} \max_{\|\theta\|_2\leq 1} p^T Z \theta = \langle Z^T p, \theta \rangle$

$$= \min_{p\in\Delta_n} \langle Z^T p, \frac{Z^T p}{\|Z^T p\|_2} \rangle$$

$$= \min_{p \in \Delta_n} \|Z^T p\|_2 .$$

c) $F(\theta) = \log\left(\sum_{i=1}^{n} \exp(-Y_i X_i^T \theta)\right)$

Therefore, $\forall \theta$
$$\nabla F(\theta) = \frac{1}{\sum_{i=1}^{n} \exp(-Y_i X_i^T\theta)} \sum_{i=1}^{n} \exp(-Y_i X_i^T \theta) Y_i X_i$$

Thus $\nabla F(\theta) = Z^T p^{(\theta)}$ with $p^{(\theta)} = (p_i^{(\theta)})_{i=1,\dots,n}$, $p_i^{(\theta)} = \frac{\exp(-Y_i X_i^T\theta)}{\sum_j \exp(-Y_j X_j^T \theta)}$

$p^{(\theta)} \in \Delta_n$, it follows that
$$\|\nabla F(\theta)\|_2 \geq \min_{p\in\Delta_n} \|Z^T p\|_2 = \gamma$$

Exercice 1

1. Minimizers of the least squares pb are solutions to $\frac{2}{n} X^T(X\theta - Y) = 0 \iff X\theta - Y = 0 \iff Y = X\theta$.

2. $\theta^* = X^T(XX^T)^{-1}Y = X^+ Y$

3. $\theta = \theta^* + v$, where $v \in \operatorname{Ker} X$

4.

5. By stability of the GD updates in $\operatorname{Im}(X^T)$, starting from θ_0,

$\theta_\infty = x + (I-P)\theta_0$ for $x \in \operatorname{Im}(X^T)$

and $\theta_\infty = \theta^* + v$ for $v \in \operatorname{Ker} X$

$\theta_\infty = \theta^* + (I-P)\theta_0$

6. $\operatorname{proj}_{sol}(\theta_0) = \operatorname{argmin}_\theta \frac{1}{2}\|\theta - \theta_0\|_2^2$ s.t. $Y = X\theta$

$\mathcal{L}(\theta, \lambda) = \frac{1}{2}\|\theta - \theta_0\|_2^2 + \langle \lambda, X\theta - Y \rangle$

KKT: $\begin{cases} \theta - \theta_0 + X^T\lambda = 0 \\ X\theta - Y = 0 \end{cases}$ $\Rightarrow X\theta - X\theta_0 + XX^T\lambda = X\theta - Y \Rightarrow XX^T\lambda = X\theta_0 - Y$

$\Rightarrow \lambda = (XX^T)^{-1}X\theta_0 - (XX^T)^{-1}Y$

donc

$\operatorname{proj}_{sol}(\theta_0) = \theta_0 - X^T(XX^T)^{-1}X\theta_0 + X^T(XX^T)^{-1}Y$

$= (I - \underbrace{X^T(XX^T)^{-1}X}_{P})\theta_0 + \theta^*$

$= \theta$