Exercice I.1

1. paramètres $\theta = (\mu, \Sigma)$;

log-vraisemblance : $l_n(\theta) = \ln\left(\prod_{i=1}^n \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(X_i-\mu)^T\Sigma^{-1}(X_i-\mu)}\right) =$

$= -\frac{d}{2} n \ln(2\pi) - \frac{n}{2}\ln(|\Sigma|) - \frac{1}{2}\sum_{i=1}^n (X_i-\mu)^T\Sigma^{-1}(X_i-\mu)$

La fct $\mu \mapsto l_n(\mu)$ est str. concave, $-\infty$ à l'infini donc admet un unique max là où sa dérivée s'annulle :

$$l_n'(\mu) = -\sum_{i=1}^n (X_i-\mu)^T \Sigma^{-1} \stackrel{!}{=} 0 \overset{\Sigma^{-1}\neq 0}{\Longleftrightarrow} \sum_{i=1}^n X_i = n\mu \Leftrightarrow \mu = \frac{1}{n}\sum_{i=1}^n X_i$$

Donc $\hat{\mu}$ est un EMV de μ^*

2. Considérons le scalaire $\underbrace{(X_i-\mu)^T}_{1\times n}\underbrace{\Sigma^{-1}}_{n\times n}\underbrace{(X_i-\mu)}_{n\times 1}$ comme la trace d'une matrice de taille 1×1.

On peut alors utiliser l'identité $tr(AB) = tr(BA)$, ce qui donne

$$l_n(\Sigma) = cste - \frac{n}{2}\ln(|\Sigma|) - \frac{1}{2}\sum_{i=1}^n tr\left(\Sigma^{-1}(X_i-\mu)(X_i-\mu)^T\right)$$

Ainsi, en dérivant / à Σ^{-1} :

$$l_n'(\Sigma^{-1}) = -\frac{n}{2}\left(\Sigma^{-1}\right)^T + \frac{1}{2}\sum_{i=1}^n (X_i-\mu)(X_i-\mu)^T$$

d'où l'on déduit $\quad \Sigma = \frac{1}{n}\sum_{i=1}^n (X_i-\mu)(X_i-\mu)^T$

Puisque $\forall \Sigma$ fixé, $l(\theta)$ est maximale en $\mu = \hat{\mu}$, $\hat{\Sigma}$ est un EMV de Σ^*

Exercice I.3

g^* est 1 si $\ln(\pi_1) - 1/2\ln(|\Sigma|) - 1/2(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)$
$> \ln(\pi_{-1}) - 1/2\ln(|\Sigma|) - 1/2(x-\mu_{-1})^T\Sigma^{-1}(x-\mu_{-1})$

$$\Leftrightarrow \ln\left(\frac{\pi_1}{\pi_{-1}}\right) - \frac{1}{2}\Big[\cancel{x^T\Sigma^{-1}x} - \mu_1^T\Sigma^{-1}x - x\Sigma^{-1}\mu_1 + \mu_1^T\Sigma^{-1}\mu_1$$
$$\cancel{-x^T\Sigma^{-1}x} + \mu_{-1}^T\Sigma^{-1}x + \mu_{-1}^T\Sigma^{-1}x - \mu_{-1}^T\Sigma^{-1}\mu_{-1}\Big] > 0$$

$$\Leftrightarrow \underbrace{\ln\left(\frac{\pi_1}{\pi_{-1}}\right) + 1/2\left(\mu_{-1}^T\Sigma^{-1}\mu_{-1} - \mu_1^T\Sigma^{-1}\mu_1\right)}_{b} + \underbrace{(\mu_1-\mu_{-1})^T\Sigma^{-1}}_{w^T}x > 0$$

Exercice I.4

Si $\Sigma_1 \neq \Sigma_{-1}$, les calculs précédents donnent que g^* est 1 si

$$b + w^T x + \frac{1}{2}x^T\underbrace{\left(\Sigma_{-1}^{-1} - \Sigma_1^{-1}\right)}_{= Q}x > 0$$

EXERCICE 1.2 (*)

$(X_i^j)_{1\le i\le n_j}$ iid

$$\mathbb{E}\left[\frac{1}{n_j}\sum_{i=1}^{n_j}(X_i^j-\hat{\mu}_j)(X_i^j-\hat{\mu}_j)^T\right] = \mathbb{E}[(X_1^j-\hat{\mu}_j)(X_1^j-\hat{\mu}_j)^T] =$$

$$= \mathbb{E}[X_1^j X_1^{jT}] - 2\,\mathbb{E}[\hat{\mu}_j X_1^{jT}] + \mathbb{E}[\hat{\mu}_j\hat{\mu}_j^T]$$

- $\mathbb{E}[\hat{\mu}_j X_1^{jT}] = \mathbb{E}\left[\frac{1}{n_j}\sum_{i=1}^{n_j} X_i^j X_1^{jT}\right] = \frac{1}{n_j}\mathbb{E}\left[X_1^j X_1^{jT} + \sum_{i=2}^{n_j} X_i^j X_1^{jT}\right] =$

$$= \frac{1}{n_j}\left(Var(X_1^j) + \mu_j\mu_j^T + (n_j-1)\mu_j\mu_j^T\right) =$$

$$= \frac{\Sigma}{n_j} + \mu_j\mu_j^T$$

- $\mathbb{E}[\hat{\mu}_j\hat{\mu}_j^T] = Var(\hat{\mu}_j) + \mu_j\mu_j^T = \frac{1}{n_j}\Sigma + \mu_j\mu_j^T$

$\hookrightarrow Var(\hat{\mu}_j) = Var\left(\frac{1}{n_j}\sum_{i=1}^{n_j} X_i^j\right) = \frac{1}{n_j}\Sigma$

- $\mathbb{E}[X_1^j X_1^{jT}] = Var(X_1^j) + \mu_j\mu_j^T = \Sigma + \mu_j\mu_j^T$

En rassemblant : (*) $= \Sigma + \mu_j\mu_j^T - \frac{2\Sigma}{n_j} - 2\mu_j\mu_j^T + \frac{\Sigma}{n_j} + \mu_j\mu_j^T =$

$$= \frac{n_j-1}{n_j}\Sigma$$

De plus, $\mathbb{E}[\hat{\Sigma}] = \mathbb{E}\left[\frac{1}{\sum_{j=1}^{C} n_j - C}\sum_{j=1}^{C}\sum_{i=1}^{n_j}(X_i^j-\hat{\mu}_j)(X_i^j-\hat{\mu}_j)^T\right] =$

$$= \frac{1}{\sum_{j=1}^{C} n_j - C}\sum_{j=1}^{C} n_j\,\mathbb{E}\left[\frac{1}{n_j}\sum_{i=1}^{n_j}(X_i^j-\hat{\mu}_j)(X_i^j-\hat{\mu}_j)^T\right] =$$

$$= \frac{1}{\sum_{j=1}^{C} n_j - C}\sum_{j=1}^{C}\cancel{n_j}\cdot\frac{n_j-1}{\cancel{n_j}}\Sigma =$$

$$= \frac{\sum_{j=1}^{C} n_j - C}{\sum_{j=1}^{C} n_j - C}\Sigma = \Sigma$$

EXERCICE 1.5

1. remember that $\forall x \in \mathbb{R}, \mathbb{1}_{x \leq 0} \leq e^{-x}$, thus

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{Y_i \frac{f_T(X_i)}{\|w\|_{\ell_1}} - \gamma < 0} \leq \frac{1}{n} \sum_{i=1}^{n} \exp\{- Y_i \cdot \frac{f_T(X_i)}{\underbrace{\|w\|_{\ell_1}}_{=1}} + \gamma \} =$$

$$= \frac{1}{n} \sum_{i=1}^{n} n \left(\prod_{t=1}^{T} Z_t\right) D_{T+1}(i) e^{\gamma} = \prod_{t=1}^{T} Z_t e^{\gamma}$$

$$= \prod_{t=1}^{T} 2\sqrt{\varepsilon_t(1-\varepsilon_t)}\, e^{\gamma} = 2^T \prod_{t=1}^{T} \sqrt{e^{2\gamma} \varepsilon_t (1-\varepsilon_t)}$$

2. examine $f(x) := \ln(x^{1-\gamma}(1+x)^{1+\gamma})$ for $x \in [0, 1/2]$:

$f(x) \xrightarrow[x \to -\infty]{} -\infty$, ~~$f(1/2)$~~ f est une fonction strictement $\nearrow$

donc sur $[0, 1/2]$, ~~$\max f(x) = f(1/2) = \ln(1/2 - \gamma)^{1-\gamma}(\frac{1}{2}+\gamma)^{1+\gamma}$~~
$\max f(x) = f(1/2)$

Ainsi, $\sqrt{\varepsilon_t^{1-\gamma}(1-\varepsilon_t)^{1+\gamma}} \underset{\uparrow}{\leq} \sqrt{(1/2-\gamma)^{1-\gamma}(1/2+\gamma)^{1+\gamma}} \quad \forall t$

$\varepsilon_t \in]0, 1/2 - \gamma[$

d'où le résultat

3. $$2^T \left[(1/2-\gamma)^{1-\gamma}(1/2+\gamma)^{1+\gamma}\right]^{T/2} = \left[2(1/2-\gamma)^{1-\gamma} 2 (1/2+\gamma)^{1+\gamma}\right]^{T/2} =$$

$$\underset{\uparrow}{=} \left[(1-2\gamma)^{1-\gamma}(1+2\gamma)^{1+\gamma}\right]^{T/2}$$

Car $2 \cdot 2 = 2 \cdot 2^{-\gamma} \cdot 2 \cdot 2^{\gamma} = 2^{1-\gamma} 2^{1+\gamma}$

EXERCICE 1.6

$$\varepsilon_{t+1} = \sum_{i=1}^{n} D_{t+1}(i) \mathbb{1}_{Y_i \neq g_t(X_i)}$$

with $w_{t+1} = 1/2 \ln\left(\frac{1-\varepsilon_{t+1}}{\varepsilon_{t+1}}\right)$
~~and D_{t+1}~~
and $D_{t+1}(i) = \frac{D_t(i) e^{-w_t Y_i g_t(X_i)}}{Z_t}$

$$\varepsilon_{t+1} = \sum_{Y_i g_t(X_i)=1} \cancel{\mathbb{1}_{Y_i \neq g_t(X_i)}}^{0} D_{t+1}(i) + \sum_{Y_i g_t(X_i)=-1} D_{t+1}(i) \mathbb{1}_{Y_i \neq g_t(X_i)} =$$

$$= \sum_{Y_i g_t(X_i)=-1} D_{t+1}(i) = \sum_{Y_i g_t(X_i)=-1} \frac{D_t(i) e^{+w_t}}{Z_t} = \frac{\sum_{Y_i g_t(X_i)=-1} D_t(i)}{\sum_{i=1}^{n} D_t(i) e^{-w_t Y_i g_t(X_i)} / e^{w_t}}$$

$$= \frac{\sum_{y_i g_t(x_i)=-1} D_t(i)}{\sum_{y_i g_t=-1} D_t(i) e^{w_t}/e^{w_t} + \sum_{y_i g_t=1} D_t(i) e^{-w_t}/e^{w_t}} =$$

$$= \frac{\sum_{g_t y_i=-1} D_t(i)}{\sum_{y_i g_t=-1} D_t(i) + e^{-2w_t} \sum_{y_i g_t=1} D_t(i)} \quad (*)$$

D'ailleurs, $w_t = 1/2 \ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$ i.e. $e^{2w_t} = \frac{1-\varepsilon_t}{\varepsilon_t}$

$$\frac{\sum_{g_t y_i=-1} D_t(i)}{\sum_{g_t y_i=-1} D_t(i) + \sum_{g_t y_i=1} D_t(i)} = \varepsilon_t$$

donc $\sum_{g_t y_i=1} D_t(i) = \frac{\sum_{g_t y_i=-1} D_t(i)}{\varepsilon_t} - \sum_{g_t y_i=-1} D_t(i)$

plug into $(*)$:

$$= \frac{\sum_{g_t y_i=-1} D_t(i)}{\sum_{g_t y_i=-1} D_t(i) + e^{-2w_t}\left(\frac{\sum_{g_t y_i=-1} D_t(i)}{\varepsilon_t} - \sum_{g_t y_i=-1} D_t(i)\right)} =$$

$$= \frac{\sum_{g_t y_i=-1} D_t(i)}{\sum_{g_t y_i=-1} D_t(i) + \frac{\varepsilon_t}{1-\varepsilon_t}\left(\frac{\sum_{g_t y_i=-1} D_t(i)}{\varepsilon_t} - \sum_{g_t y_i=-1} D_t(i)\right)} =$$

$$= \frac{\sum_{g_t y_i=-1} D_t(i)}{\sum_{g_t y_i=-1} D_t(i) + \frac{\sum_{g_t y_i=-1} D_t(i)}{1-\varepsilon_t} - \frac{\varepsilon_t}{1-\varepsilon_t} \sum_{g_t y_i=-1} D_t(i)} =$$

$$= \frac{1}{1 + \frac{1}{1-\varepsilon_t} - \frac{\varepsilon_t}{1-\varepsilon_t}} = \frac{1}{1 + \frac{1-\varepsilon_t}{1-\varepsilon_t}} = \frac{1}{1+1} = 1/2$$

Exo 1.7

1. $(\nabla g(x)-\nabla g(x'))^T(x-x') = (Lx-\nabla f(x) - Lx' + \nabla f(x'))^T(x-x') =$

$= (L(x-x') - (\nabla f(x) - \nabla f(x')))^T(x-x') =$

$= L\|x-x'\|^2 - \underbrace{(\nabla f(x)-\nabla f(x'))^T(x-x')}_{\leq L\|x-x'\|^2} \geq 0$

(one-sided Lipschitz)

thus, g is convex.

2. Using the Taylor expansion of $f(x)$:

$$f(x') = f(x) + \int_{\tau=0}^{1} d\tau \, \langle \nabla f(x+\tau(x'-x)), (x'-x) \rangle =$$

$$= f(x) + \langle \nabla f(x), x'-x \rangle + \int_{\tau=0}^{1} d\tau \, \langle \nabla f(x+\tau(x'-x)) - \nabla f(x), (x'-x) \rangle$$

$$\leq f(x) + \langle \nabla f(x), x'-x \rangle + \int_{\tau=0}^{1} d\tau \, \|\nabla f(x+\tau(x'-x)) - \nabla f(x)\| \, \|x'-x\|$$

$$\leq f(x) + \langle \nabla f(x), x'-x \rangle + L\int_{\tau=0}^{1} d\tau \, \tau \|x'-x\|^2 =$$

$$= f(x) + \langle \nabla f(x), x'-x \rangle + \frac{L}{2}\|x'-x\|^2$$

3. Par (H2) on a

$$f(x)-f(x^*) \leq \langle \nabla f(x), x-x^* \rangle - \frac{\mu}{2}\|x^*-x\|_2^2 =$$

$$= -\frac{1}{2}\left\|\sqrt{\mu}\,(x-x^*) - \frac{1}{\sqrt{\mu}}\nabla f(x)\right\|_2^2 + \frac{1}{2\mu}\|\nabla f(x)\|_2^2$$

$$\leq \frac{1}{2\mu}\|\nabla f(x)\|_2^2$$

question 4 sur la prochaine page

5. $f(x_{t+1}) - f(x^*) \leq f(x_t) - f(x^*) - \frac{\gamma}{4L}\|\nabla f(x_t)\|^2 - \frac{1}{4L}\cdot\frac{1}{2}\|d_t\|^2 =$

$$= f(x_t) - f(x^*) - \frac{\gamma}{4L}\underbrace{\|\nabla f(x_t)\|^2}_{\geq 2\mu(f(x_t)-f(x^*))} - \frac{1}{8L}\|d_t\|^2$$

$$\leq (f(x_t)-f(x^*))\left(1-\frac{\gamma\mu}{2L}\right)$$

6. immediate through recursion:

$$f(x_t)-f(x^*) \leq \left(1-\frac{\gamma\mu}{2L}\right)(f(x_{t-1})-f(x^*)) \leq \left(1-\frac{\gamma\mu}{2L}\right)^2(f(x_{t-2})-f(x^*))$$

$$\leq \ldots \leq \left(1-\frac{\gamma\mu}{2L}\right)^t (f(x_0)-f(x^*))$$

4.

$$f(x_{t+1}) = f(x_t - \eta d_t) \le f(x_t) - \langle \nabla f(x_t), \overbrace{x_t - x_{t+1}}^{=\eta d_t} \rangle + \frac{L}{2}\|\eta d_t\|^2 =$$

$$= f(x_t) - \eta \langle \nabla f(x_t), d_t \rangle + \frac{L\eta^2}{2}\|d_t\|^2 \quad (**)$$

we have $-\langle \nabla f(x_t), d_t \rangle = \frac{1}{2}\|\nabla f(x_t) - P(\nabla f(x_t))\|^2 - \frac{1}{2}\|\nabla f(x_t)\|^2 - \frac{1}{2}\|d_t\|^2$

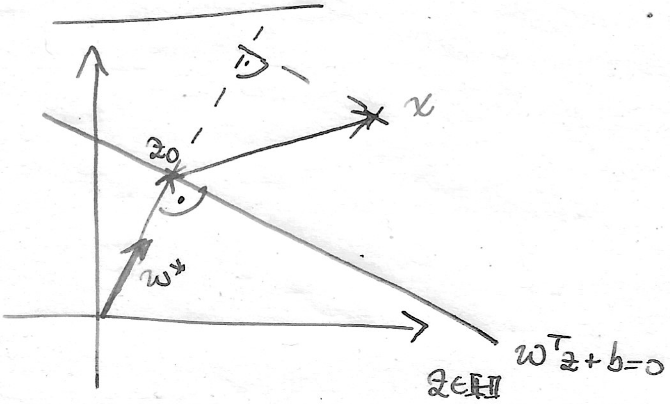
$$\underset{(H4)}{\le} \frac{1-\gamma}{2}\|\nabla f(x_t)\|^2 - \frac{1}{2}\|\nabla f(x_t)\|^2 - \frac{1}{2}\|d_t\|^2 =$$

$$= -\frac{\gamma}{2}\|\nabla f(x_t)\|^2 - \frac{1}{2}\|d_t\|^2$$

plugging into (**) $\le f(x_t) - \frac{\eta\gamma}{2}\|\nabla f(x_t)\|^2 - \frac{\eta}{2}\|d_t\|^2 + \frac{L\eta^2}{2}\|d_t\|^2 =$

$$= f(x_t) - \frac{\eta\gamma}{2}\|\nabla f(x_t)\|^2 - \frac{\eta}{2}(1 - L\eta)\|d_t\|^2$$

EXERCICE 1.8



1) $\forall z_1, z_2 \in \mathbb{H}$, $w^T(z_1 - z_2) = 0$,
hence $w^* = w/\|w\|$ is the vector normal to the surface of $\mathbb{H}$

2) $\forall z_0 \in \mathbb{H}$, $w^T z_0 = -b$

3) evidently, $d(\mathbb{H}, x) \stackrel{!}{=} \|\mathrm{proj}_w (x - z)\| = |w^{*T}(x - z)| =$

$$= \frac{1}{\|w\|} |w^T x + b|$$

EXERCICE 1.9