Corrigé de l'épreuve Mathématiques, X/ENS 2023, filière PSI.

Laurent Bonavero - Lycée Champollion (Grenoble)

Avertissements : *ceci n'est pas LE corrigé mais UN corrigé.*
Il y a dans tous mes corrigés des erreurs potentielles ou des choses qui ne sembleront pas claires...me contacter le cas échéant !

Commentaire personnel (qui par définition n'engage que moi) : après un sujet 2022 catastrophique, le sujet 2023 est nettement plus satisfaisant sur le fond. En revanche, il contient de trop nombreuses omissions, coquilles et maladresses (que certains qualifieront même d'erreurs d'énoncé) qui témoignent d'un grand manque de soin dans sa relecture finale. J'en ai listé quelques unes au fil du corrigé mais la liste n'est pas exhaustive. On est encore très loin de la qualité de sujet attendue pour un concours de ce niveau.

Partie I : Convexité et points selles

(1) (a) Soient $x, y \in C$, $t \in [0,1]$. On a par convexité de f et g :

$$\begin{array}{rcl}(f+g)((1-t)x+ty) & = & f((1-t)x+ty)+g((1-t)x+ty)\\ & \leq & (1-t)f(x)+tf(y)+(1-t)g(x)+tg(y)\\ & = & (1-t)(f(x)+g(x))+t(f(y)+g(y))\\ & = & (1-t)(f+g)(x)+t(f+g)(y).\end{array}$$

Ceci montre que $\boxed{f+g \text{ est convexe}}$.

Si l'on suppose par exemple de plus que g est strictement convexe, que $x \neq y$ et $t \in]0,1[$, on a

$$\begin{array}{rcl}(f+g)((1-t)x+ty) & = & f((1-t)x+ty)+g((1-t)x+ty)\\ & < & (1-t)f(x)+tf(y)+(1-t)g(x)+tg(y)\\ & = & (1-t)(f(x)+g(x))+t(f(y)+g(y))\\ & = & (1-t)(f+g)(x)+t(f+g)(y)\end{array}$$

donc $\boxed{f+g \text{ est strictement convexe}}$.

(b) Supposons par l'absurde que le minimum de f sur C soit atteint en deux points $x \neq y$. Notons $m = f(x) = f(y)$. On a alors par stricte convexité de f :

$$m \leq f((x+y)/2) < (f(x)+f(y))/2 = m,$$

ce qui donne la contradiction.

Ainsi, $\boxed{\text{le minimum de } f \text{ sur } C\text{, s'il existe, est atteint en au plus un point}}$.

(2) *L'énoncé est regrettablement imprécis ici : il n'est pas précisé que le produit scalaire considéré est le produit scalaire canonique sur $\mathbb{R}^n$ et $\mathbb{R}^m$.*

(a) On a

$$\boxed{\langle Ax, \nu\rangle_{\mathbb{R}^m} = (Ax)^T\nu = x^TA^T\nu = x^T(A^T\nu) = \langle x, A^T\nu\rangle_{\mathbb{R}^n}}.$$

(b) *Il est maladroit d'utiliser la notation E ici alors que E est un espace vectoriel fixé au début de la partie I.*

Soient $x \in \ker(A)$ et $y \in \mathrm{Im}(A^T)$. Il existe $\nu \in \mathbb{R}^m$ tel que $y = A^T\nu$ et on a alors

$$\langle x, y\rangle_{\mathbb{R}^n} = \langle x, A^T\nu\rangle_{\mathbb{R}^n} \underset{(a)}{=} \langle Ax, \nu\rangle_{\mathbb{R}^m} \underset{x\in\ker(A)}{=} 0$$

ce qui montre que $x \in (\mathrm{Im}(A^T))^\perp$ et donc que $\boxed{\ker(A) \subset (\mathrm{Im}(A^T))^\perp}$.

(c) On a à l'aide de ce qui précède :

$$\dim\ker(A) \leq \dim(\mathrm{Im}(A^T))^\perp = n - \dim\mathrm{Im}(A^T).$$

Comme le rang de A est égal à celui de A^T, on a à l'aide du théorème du rang appliqué à A

$$\dim(\mathrm{Im}(A^T))^\perp = n - \dim\mathrm{Im}(A^T) == n - \dim\mathrm{Im}(A) = \dim\ker(A).$$

Ainsi,

$$\ker(A) \subset (\mathrm{Im}(A^T))^\perp \text{ et } \dim\ker(A) = \dim(\mathrm{Im}(A^T))^\perp$$

donc

$$\boxed{\ker(A) = (\mathrm{Im}(A^T))^\perp}.$$

(3) *L'énoncé est légèrement imprécis ici : on imagine que comme dans la question (2), $A \in M_{m,n}(\mathbb{R})$.*

(a) Soit $u \in \ker(A)$. Pour tout t suffisamment petit, on a $x = x_* + tu \in V_b$ donc

$$h(x_* + tu) \geq h(x_*).$$

D'après la formule de Taylor-Young à l'ordre 1, on a

$$h(x_*) + t\langle \nabla h(x_*), u\rangle_{\mathbb{R}^n} + o(t) \geq h(x_*).$$

En divisant par t pour $t > 0$ puis en faisant tendre vers 0, il vient $\langle \nabla h(x_*), u\rangle_{\mathbb{R}^n} \geq 0$.
En divisant par t pour $t < 0$ puis en faisant tendre vers 0, il vient $\langle \nabla h(x_*), u\rangle_{\mathbb{R}^n} \leq 0$.
Et donc

$$\boxed{\langle \nabla h(x_*), u\rangle_{\mathbb{R}^n} = 0}.$$

(b) On vient de montrer que

$$\nabla h(x_*) \in (\ker(A))^\perp \underset{(2c)}{=} \left((\mathrm{Im}(A^T))^\perp\right)^\perp = \mathrm{Im}(A^T)$$

donc il existe $\nu_* \in \mathbb{R}^m$ tel que $\boxed{\nabla h(x_*) = A^T\nu_*}$.

(c) On a

$$L(x,\nu) = h(x) - \langle A^T\nu, x\rangle_{\mathbb{R}^n} - \langle \nu, b\rangle_{\mathbb{R}^m}.$$

Donc,

$$\frac{\partial L}{\partial x_k}(x,\nu) = \frac{\partial h}{\partial x_k}(x) - (A^T\nu)_k = (\nabla h(x))_k - (A^T\nu)_k.$$

Au point (x_*, ν_*), il vient

$$\boxed{\frac{\partial L}{\partial x_k}(x_*,\nu_*) = (\nabla h(x_*) - A^T\nu_*)_k \underset{(b)}{=} 0}.$$

(d) Comme $x_* \in V_b$, la fonction $\nu \mapsto L(x_*,\nu)$ est constante égale à $h(x_*)$ sur $\mathbb{R}^m$. On a en particulier, *mais la formulation de l'énoncé est un peu ridicule*,

$$\boxed{\forall \nu \in \mathbb{R}^m,\ L(x_*,\nu) \leq L(x_*,\nu_*)}.$$

Pour l'autre inégalité, on remarque que la fonction

$$\varphi\ : x \mapsto L(x,\nu_*) = h(x) - \langle \nu_*, Ax + b\rangle_{\mathbb{R}^m}$$

est convexe sur U comme somme de deux fonctions convexes (la deuxième est affine !) et d'après (c) possède un point critique en x_*.
Montrons que ce point critique est un minimum global de φ sur U. Soit $x \in U$ fixé quelconque et soit $f\ : t \mapsto \varphi((1-t)x_* + tx)$. La fonction f est de classe $\mathcal{C}^1$ sur un intervalle de la forme $]-\varepsilon, 1+\varepsilon[$ pour $\varepsilon > 0$ suffisamment petit, est convexe (car φ l'est) et est de dérivée nulle en $t = 0$ par dérivation de fonctions composées et utilisation de (c) :

$$f'(t) = \langle \nabla\varphi((1-t)x_* + tx), x - x_*\rangle_{\mathbb{R}^n} \text{ donc } f'(0) = \langle \nabla\varphi(x_*, x - x_*\rangle_{\mathbb{R}^n} = 0.$$

Le graphe de f étant situé au-dessus de ses tangentes, on a en particulier

$$f(1) \geq f(0) + 1 \times f'(0) = f(0)$$

et donc

$$\boxed{\forall x \in U,\ L(x,\nu_*) \geq L(x_*,\nu_*)}.$$

Partie II : Entropie et codage

Pourquoi noter log *ce que tout le monde ou presque en mathématiques note* ln *??*

(4) (a) Par croissances comparées, $\boxed{\varphi(0) = 1}$.

(b) La fonction φ est de classe $\mathcal{C}^2$ sur $\mathbb{R}^{+*}$ et

$$\forall x > 0,\ \varphi'(x) = \log(x) \text{ et } \varphi''(x) = \frac{1}{x} > 0.$$

On en déduit que $\boxed{\varphi \text{ est strictement convexe}}$, possède un unique minimum en $x = 1$ et comme $\varphi(1) = 0$, on a bien $\boxed{\varphi \geq 0 \text{ sur } \mathbb{R}^+ \text{ et } \varphi(x) = 0 \Leftrightarrow x = 1}$.

(c) La convexité de Q_χ est immédiate en revenant à la définition.
La stricte convexité et la positivité de $q \mapsto \mathrm{KL}(q, q')$ découlent de celles de φ et de la stricte positivité des q'.
Enfin,

$$\mathrm{KL}(q, q') = 0 \Leftrightarrow \forall x \in \chi,\ \varphi(q_x/q'_x)q'_x = 0 \Leftrightarrow \forall x \in \chi,\ q_x/q'_x = 1 \Leftrightarrow q = q'.$$

(5) (a) Supposons par l'absurde que $\bar{c}(x) = \bar{c}(y)$. Comme $c(x)_1 = c(y)_1$, ceci implique que $c(x) = c(y)$, contredisant l'injectivité de c.
Supposons à nouveau par l'absurde que $\bar{c}(x)$ est un préfixe de $\bar{c}(y)$. Comme $c(x)_1 = c(y)_1$, ceci implique que $c(x)$ est un préfixe de $c(y)$, contredisant le fait que c est un code préfixe.

(b) Vérifions que la restriction de $\bar{c}$ à χ_a est un code binaire préfixe. Il s'agît en particulier de vérifier que

$$\forall x \in \chi_a,\ \bar{c}(x) \neq \epsilon.$$

Par l'absurde, si $\bar{c}(x) = \epsilon$, alors $c(x) = a$ et comme χ_a contient un deuxième élément $y \neq x$, on en déduit que $c(x)$ est un préfixe de $c(y)$, ce qui donne la contradiction.
Le reste de la vérification est immédiat à l'aide de (a) puisque pour tout $x \neq y \in \chi_a$,

$$c(x)_1 = c(y)_1 = a.$$

Je remercie Michel Henri qui m'a signalé une erreur dans une première version de ce corrigé.

(c) *Cette inégalité est due à Kraft (1949).*
On raisonne donc par récurrence forte sur $\max_{x\in\chi} |c(x)|$.
Si $\max_{x\in\chi} |c(x)| = 1$, alors l'injectivité de c implique que χ possède au plus deux éléments puis que

$$\sum_{x\in\chi} 2^{-|c(x)|} = \frac{1}{2} \times \mathrm{card}(\chi) \leq 1.$$

Soit $n \in \mathbb{N}^*$. Supposons l'énoncé démontré pour $\max_{x\in\chi} |c(x)| \leq n$ et supposons $\max_{x\in\chi} |c(x)| = n + 1$. En décomposant χ en $\chi_0 \cup \chi_1$, on a

$$\sum_{x\in\chi} 2^{-|c(x)|} = \sum_{x\in\chi_0} 2^{-|c(x)|} + \sum_{x\in\chi_1} 2^{-|c(x)|} = \sum_{x\in\chi_0} 2^{-1-|\bar{c}(x)|} + \sum_{x\in\chi_1} 2^{-1-|\bar{c}(x)|} = \frac{1}{2}\left(\sum_{x\in\chi_0} 2^{-|\bar{c}(x)|} + \sum_{x\in\chi_1} 2^{-|\bar{c}(x)|}\right).$$

Comme les restrictions de $\bar{c}$ à χ_0 et χ_1 sont des codes préfixes et comme

$$\max_{x\in\chi_0} |\bar{c}(x)| \leq n \text{ et } \max_{x\in\chi_1} |\bar{c}(x)| \leq n,$$

on peut appliquer l'hypothèse de récurrence :

$$\sum_{x\in\chi_0} 2^{-|\bar{c}(x)|} + \sum_{x\in\chi_1} 2^{-|\bar{c}(x)|} \leq 1 + 1 = 2.$$

Finalement,

$$\sum_{x\in\chi} 2^{-|c(x)|} \leq \frac{1}{2} \times 2 = 1.$$

Ceci achève l'hérédité et la récurrence.

(6) *Il y a deux coquilles d'énoncé : la somme porte sur les $x \in \chi$ et non sur les $x \in X$ dans (a) et il faut changer x en X dans l'indication pour (b)...Et à ce stade du sujet, le concepteur a un regret et repasse à la notation usuelle* ln...

(a) Par le théorème de transfert,

$$E(|c(X)|) = \sum_{x\in\chi} |c(x)| P(X=x) = \sum_{x\in\chi} -\frac{\ln(q_x)}{\ln(2)} p_x$$

et donc

$$\boxed{\ln(2)E(|c(X)|) = -\sum_{x\in\chi} p_x \ln(q_x)}.$$

(b) On a

$$\begin{aligned}
\ln(2)E(|c(X)|) &= -\sum_{x\in\chi} p_x \ln(q_x) \\
&= -\sum_{x\in\chi} p_x \ln(p_x) + \sum_{x\in\chi} p_x \ln(p_x/q_x) \\
&= H(p) + \sum_{x\in\chi} q_x (p_x/q_x) \ln(p_x/q_x) \\
&= H(p) + \sum_{x\in\chi} q_x (\varphi(p_x/q_x) + p_x/q_x - 1) \\
&= H(p) + \mathrm{KL}(p,q) + \sum_{x\in\chi} p_x - \sum_{x\in\chi} q_x \\
&= H(p) + \mathrm{KL}(p,q) + 1 - \sum_{x\in\chi} 2^{-|c(x)|}.
\end{aligned}$$

D'après les questions précédentes, $\mathrm{KL}(p,q) \geq 0$ et $1 - \sum_{x\in\chi} 2^{-|c(x)|} \geq 0$. On en déduit que $\ln(2)E(|c(X)|) \geq H(p)$ et donc que

$$\boxed{E(|c(X)|) \geq \frac{H(p)}{\ln(2)}}.$$

Ce résultat est un théorème dû à Shannon (1948), connu sous le nom de "théorème du codage de source".

Partie III : Transport régularisé

(7) L'ensemble Q est convexe de façon immédiate en revenant à la définition.
L'ensemble $F(\alpha,\beta)$ est convexe comme intersection de Q avec des sous-espaces affines.

(8) *La notation* $\sum_{ij}$ *est éventuellement acceptable de la part d'un étudiant négligent lors d'un oral...*

(a) On a de suite

$$\boxed{\sum_{i\in I}\sum_{j\in J} q_{ij} = \sum_{i\in I} \alpha_i = 1.}$$

(b) On a pour tout i,

$$P(X_1 = i) = \sum_{j\in J} P(X_1 = i, X_2 = j) = \sum_{j\in J} q_{ij} = \alpha_i$$

donc $\boxed{X_1 \text{ suit la loi } \alpha \text{ sur } I}$. Et de même, $\boxed{X_2 \text{ suit la loi } \beta \text{ sur } J}$.

(c) Si $q = p$, on a

$$q_{ij} = P(X_1 = i, X_2 = j) = \alpha_i\beta_j = P(X_1 = i)P(X_2 = j)$$

donc $\boxed{X_1 \text{ et } X_2 \text{ sont indépendantes}}$.

(9) La fonction $q \mapsto \sum_{(i,j)} q_{ij}C_{ij} = \langle q, C\rangle_{\mathbb{R}^{I\times J}}$ est linéaire donc convexe. Comme $\epsilon > 0$, on en déduit que $\boxed{J_\epsilon \text{ est strictement convexe}}$ comme somme d'une fonction convexe et d'une fonction strictement convexe d'après (1a) et (4c).

(10) (a) L'ensemble $\boxed{F(\alpha,\beta) \text{ est fermé}}$ comme intersection d'images réciproques de singletons ou d'intervalles fermés par des applications continues. $\boxed{\text{Il est borné}}$ car pour tout $q \in F(\alpha,\beta)$, on a

$$\forall (i,j) \in I \times J,\, 0 \le q_{ij} \le 1.$$

(b) Comme J_ϵ est continue sur le fermé borné $F(\alpha,\beta)$, elle possède un minimum d'après le théorème des bornes atteintes.

Comme J_ϵ est strictement convexe, $\boxed{\text{ce minimum est atteint en un unique point}}$ d'après (1b).

(c) Si on pose $C = 0$, alors $\boxed{J_0 \text{ est constante égale à } 0}$, ce qui fournit un contre-exemple pour l'unicité.

(11) (a) Par l'absurde, supposons que

$$A = \{(i,j) \mid q(\epsilon)_{ij} = 0\} \neq \emptyset.$$

Soit

$$B = \{(i,j) \mid q(\epsilon)_{ij} > 0\}.$$

On a

$$\begin{aligned} J_\epsilon(q(\epsilon,t)) &= \langle q, C\rangle + t\langle p-q, C\rangle + \epsilon \sum_{(i,j)\in B} \varphi\big(q(\epsilon)_{ij}/p_{ij} + t(1-q(\epsilon)_{ij}/p_{ij})\big)p_{ij} + \epsilon \sum_{(i,j)\in A} \varphi(t)p_{ij} \\ &= \langle q, C\rangle + t\langle p-q, C\rangle + \epsilon \sum_{(i,j)\in B} \varphi\big(q(\epsilon)_{ij}/p_{ij} + t(1-q(\epsilon)_{ij}/p_{ij})\big)p_{ij} + \epsilon \sum_{(i,j)\in A} p_{ij} \\ &\quad +(\varphi(t)-1)\epsilon \sum_{(i,j)\in A} p_{ij}. \end{aligned}$$

Comme φ est dérivable sur $\mathbb{R}^{+*}$, on peut écrire

$$J_\epsilon(q(\epsilon,t)) = J_\epsilon(q(\epsilon)) + \lambda t + o(t) + (\varphi(t)-1)\,\epsilon \sum_{(i,j)\in A} p_{ij}$$

pour un réel λ. De là,

$$J_\epsilon(q(\epsilon,t)) = J_\epsilon(q(\epsilon)) + t\left(\lambda + o(1) + \frac{\varphi(t)-1}{t}\,\epsilon \sum_{(i,j)\in A} p_{ij}\right).$$

Or,

$$\lim_{t\to 0^+} \left(\lambda + o(1) + \frac{\varphi(t)-1}{t}\,\epsilon \sum_{(i,j)\in A} p_{ij}\right) = \lim_{t\to 0^+} \left(\lambda + o(1) + (\ln(t)-1)\,\epsilon \sum_{(i,j)\in A} p_{ij}\right) = -\infty.$$

On en déduit que

$$J_\epsilon(q(\epsilon,t)) - J_\epsilon(q(\epsilon)) < 0$$

pour $t > 0$ suffisamment petit, ce qui donne $\boxed{\text{la contradiction}}$.

(b) Le contre-exemple donné en (10c) fournit un contre-exemple pour cette question aussi.

Partie IV : Dualité

Nouvelle imprécision d'énoncé : il n'est pas précisé que l'on fixe $\epsilon > 0$ dans les questions (12) et (13) et il eut été plus cohérent de noter $\mathcal{L}_\epsilon$ au lieu de $\mathcal{L}$.

(12) (a) L'ensemble $\boxed{Q_{>0} \text{ est ouvert comme produit d'ouverts}}$.

(b) *Notation (f,g) un peu maladroite...et il aurait fallu écrire que $(q(\epsilon),(f(\epsilon),g(\epsilon))$ (et non $\mathcal{L}\big(q(\epsilon),(f(\epsilon),g(\epsilon))\big)$ est un point selle de $\mathcal{L}$!*

Pour $q \in \mathbb{R}^{I\times J}$, notons

$$a(q) = \Big(\big(\sum_{j\in J} q_{ij}\big)_{i\in I}, \big(\sum_{i\in I} q_{ij}\big)_{j\in J}\Big) \in \mathbb{R}^I \times \mathbb{R}^J$$

et

$$b = (-\alpha, -\beta) \in \mathbb{R}^I \times \mathbb{R}^J.$$

On peut alors ré-écrire $\mathcal{L}(q,(f,g))$ sous la forme

$$\mathcal{L}(q,(f,g)) = J_\epsilon(q) - \langle (f,g), a(q)+b\rangle_{\mathbb{R}^I\times\mathbb{R}^J}.$$

L'application a ainsi définie est linéaire et on peut écrire $a(q) = Aq$ où A est la matrice de a dans la base canonique. Ainsi

$$\mathcal{L}(q,(f,g)) = J_\epsilon(q) - \langle (f,g), Aq+b\rangle_{\mathbb{R}^I\times\mathbb{R}^J}.$$

D'autre part

$$V_b = \{q \in Q \mid Aq+b=0\} = F(\alpha,\beta).$$

D'après (10b) et (11b), J_ϵ possède un unique minimum $q(\epsilon)$ sur $V_b = F(\alpha,\beta)$. Les questions (3b) et (3d) assure l'existence de $(f(\epsilon),g(\epsilon)) \in \mathbb{R}^I\times\mathbb{R}^J$ tel que $\boxed{\big(q(\epsilon),(f(\epsilon),g(\epsilon))\big)}$ est un point selle de $\mathcal{L}$.

(13) (a) Considérons, $(f,g)\in\mathbb{R}^I\times\mathbb{R}^J$ étant fixé :

$$L : q \in Q_{>0} \mapsto \mathcal{L}(q,(f,g)) = \epsilon \sum_{(i,j)\in I\times J} \varphi(q_{ij}/p_{ij})p_{ij} + \sum_{(i,j)\in I\times J} C_{ij}q_{ij} + \sum_{i\in I} f_i(\alpha_i - \sum_{j\in J} q_{ij}) + \sum_{j\in J} g_j(\beta_j - \sum_{i\in I} q_{ij}).$$

Montrons que L possède un unique point critique sur $Q_{>0}$: on a en effet

$$\frac{\partial L}{\partial q_{ij}}(q) = \epsilon\varphi'(q_{ij}/p_{ij}) + C_{ij} - f_i - g_j = \epsilon \ln(q_{ij}/p_{ij}) + C_{ij} - f_i - g_j.$$

On a donc

$$\frac{\partial L}{\partial q_{ij}}(q) = 0 \Leftrightarrow q_{ij} = e^{(f_i+g_j-C_{ij})/\epsilon}p_{ij}.$$

Ainsi, $\boxed{L \text{ possède un unique point critique } q(f,g)}$ sur $Q_{>0}$ donné par

$$\boxed{\forall (i,j)\in I\times J,\ (q(f,g))_{ij} = e^{(f_i+g_j-C_{ij})/\epsilon}p_{ij}}.$$

La fonction L étant strictement convexe, le même argument que celui donné en (3d) montre alors que $\boxed{q(f,g) \text{ est un minimum global de } L}$: si $q \in Q_{>0}\setminus\{q(f,g)\}$, la fonction $t\mapsto L((1-t)q(f,g)+tq)$ est strictement convexe de dérivée nulle en $t=0$, donc $f(1) > f(0)$: $\boxed{L(q) > L(q(f,g))}$.

(b) On a, à l'aide d'un calcul peu passionnant et laissé au lecteur :

$$\boxed{G(f,g) = \epsilon - \epsilon \sum_{(i,j)\in I\times J} e^{(f_i+g_j-C_{ij})/\epsilon}p_{ij} + \sum_{i\in I} f_i\alpha_i + \sum_{j\in J} g_j\beta_j}.$$

(c) La concavité de G est conséquence de la convexité de

$$(x,y)\mapsto e^{x+y}$$

sur $\mathbb{R}^2$ que l'on démontre en revenant à la définition et en utilisant la convexité de l'exponentielle sur $\mathbb{R}$: pour $(x_1,y_1),(x_2,y_2)\in\mathbb{R}^2$ et $t\in[0,1]$:

$$(\star)\ : e^{(1-t)x_1+tx_2+(1-t)y_1+ty_2} = e^{(1-t)(x_1+y_1)+t(x_2+y_2)} \le (1-t)e^{x_1+y_1} + te^{x_2+y_2}.$$

La fonction $-G$ est alors combinaison linéaire à coefficients strictement positifs de telles fonctions et d'une fonction affine : elle est donc convexe et $\boxed{G \text{ est bien concave}}$.

(14) *Notons le retour grotesque du* log...

On a

$$\frac{\partial G}{\partial f_i}(f,g) = -\sum_{j\in J} e^{(f_i+g_j-C_{ij})/\epsilon}p_{ij} + \alpha_i = -e^{f_i/\epsilon}\sum_{j\in J} e^{(g_j-C_{ij})/\epsilon}p_{ij} + \alpha_i = \alpha_i\big(1 - e^{f_i/\epsilon}\sum_{j\in J} e^{(g_j-C_{ij})/\epsilon}\beta_j\big).$$

De là,

$$\frac{\partial G}{\partial f_i}(f_*(g),g) = \alpha_i\left(1 - \frac{1}{\sum_{j\in J} e^{(g_j-C_{ij})/\epsilon}\beta_j}\sum_{j\in J} e^{(g_j-C_{ij})/\epsilon}\beta_j\right) = 0.$$

Un calcul analogue donne $\boxed{\dfrac{\partial G}{\partial g_j}(f, g_*(f)) = 0}$.

(15) La fonction G est concave sur $\mathbb{R}^I \times \mathbb{R}^J$ donc sa restriction $f \mapsto G(f,g)$ à $\mathbb{R}^I \times \{g\}$ l'est aussi pour tout g. Et la question précédente montre que $f_*(g)$ est un point critique de cette restriction. Des arguments analogues à ceux de (3d) et (13a) montrent que $f_*(g)$ est un maximum global de $f \mapsto G(f,g)$. On a donc

$$\forall g \in \mathbb{R}^J, \forall f \in \mathbb{R}^I,\, G(f,g) \leq G(f_*(g), g)$$

et de même

$$\forall g \in \mathbb{R}^J, \forall f \in \mathbb{R}^I,\, G(f,g) \leq G(f, g_*(f)).$$

On en déduit que pour tout k,

$$G(f^{k+1}, g^{k+1}) = G(f_*(g^{k+1}), g^{k+1}) \geq G(f^k, g^{k+1}) = G(f^k, g_*(f^k)) \geq G(f^k, g^k)$$

ce qui montre que la suite $\boxed{(G(f^k, g^k))_{k\in\mathbb{N}} \text{ est croissante}}$.

(16) *A nouveau une (petite) négligence : il n'est pas précisé que les limites sont à prendre lorsque k tend vers $+\infty$. Et à nouveau une coquille malheureuse dans la définition de G_* : le sup est à prendre pour $(f,g) \in \mathbb{R}^I \times \mathbb{R}^J$ et non $I \times J$.*

(a) L'expression obtenue en (13b) montre que G est de classe $\mathcal{C}^1$. Comme G est concave, il suffit donc de montrer que (f^∞, g^∞) est un point critique de G pour en déduire que c'est un point où G atteint son maximum. Or, on a pour tout k d'après (15) :

$$\frac{\partial G}{\partial f_i}(f^{k+1}, g^{k+1}) = \frac{\partial G}{\partial f_i}(f_*(g^{k+1}), g^{k+1}) = 0 \text{ et } \frac{\partial G}{\partial g_j}(f^k, g^{k+1}) = \frac{\partial G}{\partial g_j}(f^k, g_*(f^k)) = 0.$$

Par continuité des dérivées partielles et passage à la limite quand k tend vers $+\infty$, il vient

$$\frac{\partial G}{\partial f_i}(f^\infty, g^\infty) = 0 \text{ et } \frac{\partial G}{\partial g_j}(f^\infty, g^\infty) = 0,$$

ce qui conclut !

(b) Par définition de $q(f,g)$ comme minimum de $q \mapsto \mathcal{L}(q,(f,g))$, on a

$$G(f,g) = \mathcal{L}(q(f,g),(f,g)) \leq \mathcal{L}(q(\epsilon),(f,g)).$$

Mais comme $(q(\epsilon),(f(\epsilon),g(\epsilon)))$ est un point-selle de $\mathcal{L}$, on a

$$L(q(\epsilon),(f,g)) \leq \mathcal{L}(q(\epsilon),(f(\epsilon),g(\epsilon))) = G(f(\epsilon),g(\epsilon)).$$

Finalement,

$$\forall (f,g),\, G(f,g) \leq G(f(\epsilon),g(\epsilon))$$

ce qui montre que $\boxed{G_* = G(f(\epsilon),g(\epsilon))}$.

(c) *Cette question me semble mal quantifiée : la constante a dépend du couple (i,j).*
Comme $G(f(\epsilon),g(\epsilon)) = G(f^\infty,g^\infty) = \max G$, il est naturel d'étudier d'où vient le manque de **stricte** concavité de G. En reprenant le raisonnement de (13c), l'exponentielle étant strictement convexe, on a égalité dans $(\star)$ si et seulement si $x_1 + y_1 = x_2 + y_2$.
On en déduit que $G(f(\epsilon),g(\epsilon)) = G(f^\infty,g^\infty)$ si et seulement si, pour tout couple (i,j) :

$$f_i^\infty + g_j^\infty = f(\epsilon)_i + g(\epsilon)_j$$

ce qui donne

$$\boxed{f_i^\infty - f(\epsilon)_i = g(\epsilon)_j - g_j^\infty \underset{\text{not}}{=} a_{ij} \in \mathbb{R}}.$$

(d) A l'aide de (13a), on en déduit que

$$q(\epsilon) = q(f^\infty, g^\infty)$$

et (13a) montrant aussi que $(f,g) \mapsto q(f,g)$ est continue, il vient finalement

$$\boxed{q(\epsilon) = q(f^\infty,g^\infty) = \lim_{k\to+\infty} q(f^k,g^k)}.$$