PROPOSITION DE CORRECTION DU CONTRÔLE CLASSANT DE MATHÉMATIQUES APPLIQUÉES
Analyse numérique et optimisation (MAP431)
Lundi 30 juin 2008

Problème 1 - Estimation d'erreur a posteriori

1) La formulation variationnelle consiste à déterminer

$$u \in X := \{v \in H^1(\Omega) \ : \ v = 0 \text{ sur } \Gamma_D\}$$

tel que quel que soit $v \in X$,

$$\int_\Omega \nabla u \cdot \nabla v \, dx = \int_\Omega f v \, dx + \int_{\Gamma_N} gv \, ds.$$

D'après le théorème de Lax-Milgram, cette formulation variationnelle admet une solution unique. La coercivité de la forme bilinéaire découle de l'inégalité de type Poincaré rappelée dans l'énoncé.

2)

a. La formulation variationnelle est identique à celle obtenue précédemment, quitte à remplacer f par $f(h)$ et g par $g(h)$.

b. Pour tout $v \in X$, on a

$$\int_\Omega (\nabla u - \nabla u(h)) \cdot \nabla v \, dx = \int_\Omega (f - f(h)) v \, dx + \int_{\Gamma_N} (g - g(h)) v \, ds.$$

En choisissant $v = u - u(h)$, on en déduit en utilisant l'inégalité de Cauchy-Schwarz et le Théorème de trace que

$$\int_\Omega |\nabla u - \nabla u(h)|^2 \, dx \leq C \|u - u(h)\|_{H^1(\Omega)} (\|f - f(h)\|_{L^2(\Omega)} + \|g - g(h)\|_{L^2(\Gamma_n)}).$$

D'après l'inégalité de type Poincaré rappelé dans l'énoncé, il vient

$$\|u - u(h)\|_{H^1(\Omega)} \leq C(\|f - f(h)\|_{L^2(\Omega)} + \|g - g(h)\|_{L^2(\Gamma_n)}).$$

c. Soit T un triangle du maillage. On note a_0, a_1 et a_2 ses sommets et A_T la matrice $(a_1 - a_0 | a_2 - a_0)$. L'application affine $F_T(x) = A_T x + a_0$ est telle que $T = F_T(T_0)$.
On utilise la norme matricielle subordonnée à la norme euclidienne. On a

$$\|A_T\| = \sup_{\|x\| \leq 1} \|A_T x\|.$$

Soit B_0 la boule inscrite dans T_0 de rayon ρ_0 et de centre x_0. On a

$$\|A_T\| = \rho_0^{-1} \sup_{\|x\| \leq \rho_0} \|A_T x\| = \rho_0^{-1} \sup_{\|x - x_0\| \leq \rho_0} \|A_T x - A_T x_0\| = \rho_0^{-1} \sup_{x \in B_0} \|F_T(x) - F_T(x_0)\|.$$

Comme B_0 est incluse dans T_0, on en déduit que pour tout $x \in B_0$, $F_T(x)$ appartient à T. Ainsi $\|F_T(x) - F_T(x_0\| \leq h_T$ et

$$\|A_T\| \leq h_T/\rho_0.$$

Réciproquement, on a

$$\|A_T^{-1}\| \leq h_{T_0}/\rho_T,$$

où ρ_T est le diamètre du cercle inscrit dans T. Le maillage étant régulier, le diamètre du cercle inscrit dans T est du même ordre de grandeur que le diamètre du triangle T. Ainsi,

$$\|A_T^{-1}\| \leq Ch_T^{-1},$$

d. D'après l'inégalité de Poincaré-Wirtinger, pour toute fonction $\varphi \in H^1(T_0)$, on a

$$\int_{T_0} |\varphi - m(\varphi)|^2 \, dx \leq \int_{T_0} |\nabla\varphi|^2 \, dx;$$

On pose $\psi = \varphi \circ F_T^{-1}$. On a $\nabla\psi(y) = A_T^{-1}\nabla\varphi$. En effectuant un changement de variable dans l'inégalité précédente, on en déduit que

$$\int_T |\psi - m(\psi)|^2 |\det A_T|^{-1} \, dy \leq \int_T |A_T\nabla\psi|^2 |\det A_T|^{-1} \, dy,$$

soit

$$\int_T |\psi - m(\psi)|^2 \, dy \leq \int_T |A_T\nabla\psi|^2 \, dy.$$

D'après la question précédente, $\|A_T\| \leq Ch_T$ (où C est une constante indépendante de h_T) et

$$\int_T |\psi - m(\psi)|^2 \, dy \leq |h_T|^2 \int_T |\nabla\psi|^2 \, dy.$$

En appliquant cette inégalité à $\psi = f$, on en déduit que

$$\int_T |f - f(h)|^2 \, dx \leq |h_T|^2 \int_T |\nabla f|^2 \, dx.$$

e. Un raisonnement similaire nous permet d'établir que

$$\int_E |g - g(h)|^2 \leq Ch_E^2 \int_E |\nabla g|^2.$$

En combinant cette inégalité avec celles obtenues aux questions **b** et **d**, il vient

$$\|u - u(h)\|_{H^1(\Omega)} \leq Ch(\|\nabla f\|_{L^2(\Omega)} + \|\nabla g\|_{L^2(\Gamma_N)}).$$

3) Il suffit d'appliquer le Théorème de Lax-Milgram pour obtenir l'existence d'une unique solution au problème discrétisé (3).

4)

a. Un élément fini $\mathbb{P}_1$ est déterminé de manière unique par ses valeurs aux nœuds du maillage. La définition de I_h détermine pour tout φ un élément fini $\mathbb{P}_1$ $I_h(\varphi)$, car elle précise sa valeur en tous les nœuds du maillage. De plus, comme $I_h(\varphi)$ s'annule sur Γ_D, il appartient à X_h.

b. La famille de maillages $\mathcal{T}_h$ étant supposée régulière, l'angle formé par deux cotés d'un triangle $T \in \mathcal{T}_h$ est borné inférieurement, indépendamment de h. Notons $\theta_{\min} > 0$ cette borne inférieure. Un sommet x du maillage ne peut pas appartenir à plus de $2\pi/\theta_{\min}$ triangles du maillage et

$$|\operatorname{Card}\{T \in \mathcal{T}_h \ ; \ T \subset \omega_x\}| \leq 2\pi/\theta_{\min}$$

L'ouvert ω_x est inclus dans la boule centrée en x de rayon $\max_{T' \subset \omega_x} h_{T'}$. Ainsi, il suffit de prouver que pour tout triangle T du maillage inclus dans ω_x, on a

$$\max_{T' \subset \omega_x} h_{T'} \leq C h_T.$$

En fait, il suffit de montrer que le rapport h_{T_1}/h_{T_2} entre deux triangles possédant une arête commune est borné par une constante c_0. On en déduit alors, en raisonnant de proche en proche que

$$\max_{T' \subset \omega_x} h_{T'} \leq (c_0)^{m-1} h_T,$$

où $m = \sup_x |\operatorname{Card}\{T \in \mathcal{T}_h \ ; \ T \subset \omega_x\}|$. Soit E l'arête commune à T_1 et T_2 le maillage étant régulier, h_{T_1}/h_E et h_E/h_{T_2} sont tous deux majorés indépendamment de h. Il en est donc de même pour h_{T_1}/h_{T_2}.

c. Nous allons prouver que pour tout entier n il existe une constante C_n telle que pour tout sommet x du maillage appartenant à n triangles distincts du maillage (c'est à dire tel que $n = |\operatorname{Card}\{T \in \mathcal{T}_h \ ; \ T \subset \omega_x\}|$), on a

$$\|v - \pi_x v\|_{L^2(\omega_x)} \leq C_n (\operatorname{diam}(\omega_x)) \|\nabla v\|_{L^2(\omega_x)}.$$

Étant donné que n est majoré par m, on en déduit alors que

$$\|v - \pi_x v\|_{L^2(\omega_x)} \leq \sup_{n \leq m} C_n (\operatorname{diam}(\omega_x)) \|\nabla v\|_{L^2(\omega_x)}.$$

On suppose pour simplifier que x n'est pas un élément du bord du maillage (dans le cas contraire, on peut aisément adapter le raisonnement qui suit).

Soit ω_n un polygône régulier de diamètre 1, centré à l'origine. On décompose ω_n en n triangles notés T_i $(i = 1, \cdots, n)$. Soit F un difféomorphisme, de déterminant positif, de ω_n dans ω_x, continue, dont la restriction à chacun des triangles T_i est affine, tel que $F(0) = x$ et tels que $F(T_i)$ soit un triangle du maillage $\mathcal{T}_h$ (une telle application existe et est définie à une rotation de $2k\pi/n$ près $(k = 1, \cdots, n)$ du polygône ω_n). D'après l'inégalité (6) de l'énoncé, il existe C_n tel que pour tout $v \in H^1(\omega_x)$,

$$\|v - \pi_x v\|^2_{L^2(\omega_x)} \leq C_n \frac{\max \det F}{\min \det F} \|\nabla F\|^2 \|\nabla \varphi\|^2_{L^2(\omega_x)}.$$

Or $\max \det F = \max_{K \in \omega_x}(n|K|/|\omega_n|)$, $\min \det F = \min_{K \in \omega_x}(n|K|/|\omega_n|)$ et

$$\|\nabla F\| = \max_{\|y\| = \rho_n} \|Fy\|/\rho_n \leq h_T/\rho_n,$$

où ρ_n est le diamètre du cercle circonscrit des triangles T_i. On obtient donc la majoration recherchée suivante

$$\|v - \pi_x v\|^2_{L^2(\omega_x)} \leq C_n \frac{\max |K|}{\min |K|} h_T^2/\rho_n \|\nabla \varphi\|^2_{L^2(\omega_x)}.$$

d. L'ensemble des polynômes de degré inférieur ou égale à un étant un espace de dimension fini, les normes $\|p\|_{L^\infty(T_0)}$ et $\|p\|_{L^2(T_0)}$ sont équivalentes. Il existe donc des constantes C_1 et C_2 telles que

$$\|p\|_{L^\infty(T_0)} \leq C_1 \|p\|_{L^2(T_0)}$$

et

$$\|p\|_{L^2(T_0)} \leq C_1 \|p\|_{L^\infty(T_0)}.$$

Soit T un triangle quelconque de $\mathbb{R}^2$ et F une application affine, difféomorphisme de T_0 vers T. Pour tout polynôme p de degré au plus un, on pose $\hat{p} = p \circ F^{-1}$. $\hat{p}$ est également un polynôme de degré au plus un sur $\mathbb{R}^2$, et d'après les inégalités précédente,

$$\|\hat{p}\|_{L^\infty(T_0)} \leq C_1 \|\hat{p}\|_{L^2(T_0)} \tag{1}$$

et

$$\|\hat{p}\|_{L^2(T_0)} \leq C_2 \|\hat{p}\|_{L^\infty(T_0)}. \tag{2}$$

Or

$$\|\hat{p}\|_{L^\infty(T_0)} = \|p\|_{L^\infty(T)}$$

et

$$\|\hat{p}\|^2_{L^2(T_0)} = \int_T |p|^2 |\det \nabla F|^{-1} = |T|^{-1} \|p\|^2_{L^2(T)}/2.$$

Ainsi, d'après (1), on obtient

$$\|p\|_{L^\infty(T)} \leq C_1 |T|^{-1/2} \|p\|_{L^2(T)}/\sqrt{2}$$

et d'après (1),

$$\|p\|_{L^2(T)} \leq C_2 \sqrt{2} |T|^{1/2} \|p\|_{L^\infty(T)}.$$

e. Soit $\varphi \in X$, on a

$$(\varphi - I_h \varphi)_T = \sum_{x \in \mathcal{N}(T)} (\varphi - I_h(\varphi)(x))\phi_x. = \sum_{x \in \mathcal{N}(T)} (\varphi - \pi_x \varphi)\phi_x.$$

Ainsi,

$$\|\varphi - I_h \varphi\|_{L^2(T)} \leq \sum_{x \in \mathcal{N}(T)} \|\varphi - \pi_x \varphi\|_{L^2(T)}$$

Ainsi, d'après la question **4.c.**, il vient

$$\|\varphi - I_h \varphi\|_{L^2(T)} \leq C \sum_{x \in \mathcal{N}(T)} (\operatorname{diam} \omega_x) \|\nabla \varphi\|_{L^2(\omega_x)} \leq C h_T \|\nabla \varphi\|_{L^2(\widetilde{\omega}_T)}. \tag{3}$$

(Pour obtenir cette dernière inégalité, on utilise le fait que, le maillage étant régulier, le diamètre des ouvert ω_x est du même ordre de grandeur que h_T).

On effectue un raisonnement similaire pour majorer $\|\varphi - I_h \varphi\|_{L^2(E)}$. Tout d'abord, on montre, en effectuant un raisonnement identique à celui effectué à la question **4.c.** qu'il existe un constante C telle que pour toute arête E du maillage, tout sommet x de E et toute fonction $v \in H^1(\omega_x)$, on a

$$\|v - \pi_x v\|_{L^2(E)} \leq C h_E^{1/2} \|\nabla \varphi\|_{L^2(\omega_x)}.$$

$$\|\varphi - I_h\varphi\|_{L^2(E)} \leq \sum_{x\in\mathcal{N}(E)} |\varphi(x) - \pi_x\varphi| \|\phi_x\|_{L^2(E)}. \tag{4}$$

Soit $\varphi \in X$, on a

$$(\varphi - I_h\varphi)_E = \sum_{x\in\mathcal{N}(E)} (\varphi - I_h(\varphi)(x))\phi_x. = \sum_{x\in\mathcal{N}(E)} (\varphi - \pi_x\varphi)\phi_x.$$

Ainsi,

$$\|\varphi - I_h\varphi\|_{L^2(E)} \leq \sum_{x\in\mathcal{N}(E)} \|\varphi - \pi_x\varphi\|_{L^2(E)}$$

Ainsi, d'après (4), il vient

$$\|\varphi - I_h\varphi\|_{L^2(E)} \leq C \sum_{x\in\mathcal{N}(E)} h_E^{1/2} \|\nabla\varphi\|_{L^2(\omega_x)} \leq C h_E^{1/2} \|\nabla\varphi\|_{L^2(\widetilde{\omega}_E)}. \tag{5}$$

5) Soit $v \in X$. On a

$$\int_\Omega \nabla u_h \cdot \nabla v = \sum_{T\in\mathcal{T}_h} \int_T \nabla u_h \cdot \nabla v$$

Sur chaque triangle T, u_h est régulier, on peut donc effectuer l'intégration par partie

$$\int_T \nabla u_h \cdot \nabla v = -\int_T \Delta u_h v + \int_{\partial T} (\nabla u_h|_T \cdot n_T) v,$$

où n_T est la normale extérieure qu triangle T. La restriction de u_h à tout triangle T du maillage étant affine, $\Delta u_h = 0$ sur T. On en déduit que

$$\int_\Omega \nabla u_h \cdot \nabla v = \sum_{E\in\mathcal{E}_{h,N}} \int_E n_E \cdot \nabla u_h v + \sum_{E\in\mathcal{E}_{h,\Omega}} \int_E (\nabla u_h|_{T_1} \cdot n_{T_1} + \nabla u_h|_{T_2} \cdot n_{T_2}) v.$$

Sans perte de généralité, on peut supposer que $n_E = n_{T_1}$. On a alors Or $\nabla u_h|_{T_1} = \lim_{t\to 0^+} \nabla u_h(x - tn_E)$ et $\nabla u_h|_{T_1} = \lim_{t\to 0^+} \nabla u_h(x + tn_E)$. Ainsi,

$$\nabla u_h|_{T_1} \cdot n_{T_1} + \nabla u_h|_{T_2} \cdot n_{T_2} = \lim_{t\to 0^+} \nabla u_h(x - tn_E) \cdot n_E - \nabla u_h(x + tn_E) \cdot n_E = -[\nabla u_h \cdot n_E]_E.$$

Ainsi,

$$\int_\Omega \nabla u_h \cdot \nabla v = \sum_{E\in\mathcal{E}_{h,N}} \int_E n_E \cdot \nabla u_h v - \sum_{E\in\mathcal{E}_{h,\Omega}} \int_E [\nabla u_h \cdot n_E]_E v$$

et

$$\int_\Omega f(h) v + \int_{\Gamma_N} g(h) v - \int_\Omega \nabla u_h \cdot \nabla v$$
$$= \sum_{T\in\mathcal{T}_h} \int_T f(h) v + \sum_{E\in\mathcal{E}_{h,N}} \int_E (g(h) - n_E \cdot \nabla u_h) v + \sum_{E\in\mathcal{E}_{h,\Omega}} [n_E \cdot \nabla u_h] v. \tag{6}$$

6) Pour tout $v_h \in X_h$, on a

$$\int_\Omega f(h)v_h + \int_{\Gamma_N} g(h)v_h = \int_\Omega \nabla u_h \cdot \nabla v_h.$$

Ainsi, pour tout $v \in X$, on a

$$\int_\Omega f(h)v + \int_{\Gamma_N} g(h)v - \int_\Omega \nabla u_h \cdot \nabla v = \int_\Omega f(h)(v - v_h) + \int_{\Gamma_N} g(h)(v - v_h) - \int_\Omega \nabla u_h \cdot \nabla(v - v_h).$$

D'après l'égalité établie à la question précédente, suite à l'application de l'inégalité de Cauchy-Schwarz, il vient

$$\int_\Omega f(h)v + \int_{\Gamma_N} g(h)v - \int_\Omega \nabla u_h \cdot \nabla v \leq \sum_{T\in\mathcal{T}_h} \|f(h)\|_{L^2(T)} \|v - v_h\|_{L^2(T)}$$
$$+ \sum_{E\in\mathcal{E}_{h,N}} \|g(h) - n_E \cdot \nabla u_h\|_{L^2(E)} \|v - v_h\|_{L^2(E)} + \sum_{E\in\mathcal{E}_{h,\Omega}} \|[n_E \cdot \nabla u_h]_E\|_{L^2(E)} \|v - v_h\|_{L^2(E)}.$$

En appliquant cette majoration à $v_h = I_h(v)$ ainsi que les estimations (3) et (5), on obtient

$$\int_\Omega f(h)v + \int_{\Gamma_N} g(h)v - \int_\Omega \nabla u_h \cdot \nabla v \leq C\Bigg(\sum_{T\in\mathcal{T}_h} \|f(h)\|_{L^2(T)} h_T \|v\|_{H^1(\widetilde{\omega}_T)}$$
$$+ \sum_{E\in\mathcal{E}_{h,N}} \|g(h) - n_E \cdot \nabla u_h\|_{L^2(E)} h_E^{1/2} \|v\|_{H^1(\widetilde{\omega}_E)} + \sum_{E\in\mathcal{E}_{h,\Omega}} \|[n_E \cdot \nabla u_h]_E\|_{L^2(E)} h_E^{1/2} \|v\|_{H^1(\widetilde{\omega}_E)}\Bigg)$$

et

$$\int_\Omega f(h)v + \int_{\Gamma_N} g(h)v - \int_\Omega \nabla u_h \cdot \nabla v \leq C\left(\sum_{T\in\mathcal{T}_h} \|v\|_{H^1(\widetilde{\omega}_T)} 2 + \sum_{E\in\mathcal{E}_{h,N}} \|v\|^2_{H^1(\widetilde{\omega}_E)} + \sum_{E\in\mathcal{E}_{h,\Omega}} \|v\|^2_{H^1(\widetilde{\omega}_E)} \right)^{1/2}$$
$$\left(\sum_{T\in\mathcal{T}_h} \|f(h)\|^2_{L^2(T)} h_T^2 + \sum_{E\in\mathcal{E}_{h,N}} \|g(h) - n_E \cdot \nabla u_h\|^2_{L^2(E)} h_E + \sum_{E\in\mathcal{E}_{h,\Omega}} \|[n_E \cdot \nabla u_h]_E\|^2_{L^2(E)} h_E \right)^{1/2}$$

Enfin, comme chaque triangle T du maillage appartient à un nombre uniformément borné (par rapport à h) d'ouverts $\widetilde{\omega}_T$ et $\widetilde{\omega}_E$, on en déduit que

$$\int_\Omega f(h)v + \int_{\Gamma_N} g(h)v - \int_\Omega \nabla u_h \cdot \nabla v \leq C\|v\|_{H^1(\Omega)} \Bigg(\sum_{T\in\mathcal{T}_h} \|f(h)\|^2_{L^2(T)} h_T^2$$
$$+ \sum_{E\in\mathcal{E}_{h,N}} \|g(h) - n_E \cdot \nabla u_h\|^2_{L^2(E)} h_E + \sum_{E\in\mathcal{E}_{h,\Omega}} \|[n_E \cdot \nabla u_h]^2_{L^2(E)} h_E \Bigg)^{1/2}.$$

7) Il suffit de constater que pour tout $v \in X$, on a

$$\int_\Omega f(h)v + \int_{\Gamma_N} g(h)v = \int_\Omega \nabla u(h) \cdot \nabla v.$$

Ainsi, d'après la question précédente,

$$\int_\Omega \nabla(u(h) - u_h) \cdot \nabla v \leq C\|v\|_{H^1(\Omega)} \left(\sum_{T\in\mathcal{T}_h} \eta_T^2\right)^{1/2}$$

où

$$\eta_T = \left\{ h_T^2 \|f_T\|_{L^2(T)}^2 + \frac{1}{2} \sum_{E\in\mathcal{E}_{h,\Omega}\cap\mathcal{E}(T)} h_E \|[n_E \cdot \nabla u_h]_E\|_{L^2(E)}^2 \right.$$
$$\left. + \sum_{E\in\mathcal{E}_{h,N}\cap\mathcal{E}(T)} h_E \|g_E - n_E \cdot \nabla u_h\|_{L^2(E)}^2 \right\}^{1/2}, \quad (7)$$

et $\mathcal{E}(T)$ est l'ensemble des arêtes du triangle T. En appliquant cette estimation à $v = u(h) - u_h$ et en utilisant l'inégalité de type Poincaré

$$\|w\|_{H^1(\Omega)}^2 \leq C \int_\Omega |\nabla w|^2,$$

pour tout $w \in X$, on obtient

$$\|u(h) - u_h\|_{H^1(\Omega)} \leq C \left(\sum_{T\in\mathcal{T}_h} \eta_T^2\right)^{1/2}.$$

8)

a. La fonction b_T est continue, car $\lambda_1(x)\lambda_2(x)\lambda_3(x)$ est nul sur le bord du triangle T. On a

$$\max_{x\in T} b_T(x) = \max_{\lambda\in K} 27\lambda_1\lambda_2\lambda_3,$$

où

$$K = \{\lambda \in \mathbb{R}^3 \ : \ \lambda_i \geq 0 \text{ et } \lambda_1 + \lambda_2 + \lambda_3 = 1\}.$$

Le maximum de b_T ne peut être atteint en un point tel que l'une des coordonnées de λ s'annule. D'après les conditions d'optimalité du premier ordre, il existe donc un réel μ tel que

$$27\lambda_2\lambda_3 = 27\lambda_1\lambda_2 = 27\lambda_1\lambda_3 = \mu$$

(les contraintes de type $\lambda_i \geq 0$ sont inactives). Il s'en suit que $\lambda_1 = \lambda_2 = \lambda_3 = 1/3$ et

$$\max_{x\in T} b_T(x) = 1.$$

b. D'après la formule de quadrature (6.43) du cours,

$$\int_T \lambda_1\lambda_2\lambda_3 = |T|1!1!1!2!/5! = |T|/60.$$

Ainsi,

$$\int_T b_T = \frac{27}{60}|T| = \frac{9}{20}|T|.$$

Le maillage étant régulier, $|T|$ est du même ordre de grandeur que h_T^2.

c. Soit F_T une application affine telle que $F_T(T_0) = T$. On a $b_T \circ F_T = b_{T_0}$ et

$$\|b_T\|^2_{L^2(T_0)} = \int_{T_0} |b_{T_0}|^2 |\det \nabla F_T| = |\det \nabla F_T| \|b_{T_0}\|^2_{L^2(T_0)} \tag{8}$$

De plus, $\nabla b_T = \nabla F_T^{-1} \nabla b_{T_0}$ et

$$\|\nabla b_T\|^2_{L^2(T)} = \int_{T_0} |\nabla F_T^{-1} \nabla b_{T_0}|^2 |\det \nabla F_T| \leq |\det \nabla F_T| \|\nabla F_T^{-1}\| \|\nabla b_{T_0}\|^2_{L^2(T_0)} \tag{9}$$

Ainsi,

$$\|\nabla b_T\|_{L^2(T)} \leq C \|\nabla F_T^{-1}\| \|b_T\|_{L^2(T)}.$$

On conclut en notant que $\|\nabla F_T^{-1}\| = \max_{\|Fx\|=\rho_T} \|x\|/\rho(T) \leq h_{T_0}/\rho_T \leq C h_T^{-1}$

9)

a. Par intégration par partie, la restriction de u_h à T étant affine et b_T étant nul sur le bord de T, on a

$$\int_T \nabla u_h \cdot \nabla b_T = 0.$$

De plus, d'après le problème variationnel vérifié par $u(h)$, on a

$$\int_T \nabla u(h) \cdot \nabla b_T = \int_T f_T b_T.$$

Ainsi,

$$\int_T \nabla (u(h) - u_h) \cdot \nabla b_T = \int_T f_T b_T.$$

b. Comme $\int_T b_T = 9|T|/20$, on a

$$9|T||f_T|/20 = \int_T \nabla(u(h) - u_h) \cdot \nabla b_T \leq \|\nabla(u(h) - u_h)\|_{L^2(T)} \|\nabla b_T\|_{L^2(T)}.$$

Or $\|\nabla b_T\|_{L^2(T)} \leq C h_T^{-1} \|b_T\|_{L^2(T)} \leq C h_T^{-1} |T|^{1/2}$. Ainsi,

$$|T||f_T| \leq C \|\nabla(u(h) - u_h)\|_{L^2(T)} h_T^{-1} |T|^{1/2}.$$

et

$$|T|^{1/2} |f_T| \leq C \|\nabla(u(h) - u_h)\|_{L^2(T)} h_T^{-1}.$$

On en déduit que

$$\|f_T\|^2_{L^2} = |f_T|^2 |T| \leq C \|\nabla(u(h) - u_h)\|^2_{L^2(T)} h_T^{-2}.$$

10)

a. La fonction b_E est continue sur chaque triangle. De plus, on a continuité le long de chaque arête: b_E nulle sur le bord de $T_1 \cap T_2$ égale à $4\lambda_1(x)\lambda_2(x)$ sur l'arête E. Elle est donc continue. On a

$$\max_{x \in T_1 \cup T_2} b_E(x) = \max_{0 \leq \lambda \leq 1} 4\lambda(1-\lambda) = 1.$$

b. D'après la formule (6.43) du cours, pour tout $T \in \omega_E$,

$$\int_T b_E = \int_T 4\lambda_1\lambda_2 = 4|T|2!/4! = |T|/3.$$

Enfin, comme le maillage est régulier, $|T|$ est du même ordre de grandeur que h_E^2.

$$\int_E b_E = \int_0^1 4s(1-s)h_E ds = 4h_E(1/2 - 1/3) = 2h_E/3.$$

c. Le même raisonnement que celui effectué à la question **8.c** s'applique et permet d'obtenir que

$$\|\nabla b_E\|_{L^2(T)} \leq C h_E^{-1} \|b_E\|_{L^2(T)}.$$

11)

a. Par intégration par partie,

$$\int_{\omega_E} \nabla u_h \cdot \nabla b_E = \sum_{T \subset \omega_E} \int_T \nabla u_h \cdot \nabla b_E = -\int_E [n_E \cdot \nabla u_h]_E b_E.$$

De plus, d'après le problème variationnel vérifié par $u(h)$,

$$\sum_{T \subset \omega_E} \int_T f(h) b_E = \int_{\omega_E} \nabla u(h) \cdot \nabla b_E.$$

Ainsi,

$$\int_E [n_E \cdot \nabla u_h]_E b_E = \int_{\omega_E} \nabla(u(h) - u_h) \cdot \nabla b_E - \sum_{T \subset \omega_E} \int_T f(h) b_E.$$

b. D'après la question précédente, on a

$$|[n_E \cdot \nabla u_h]_E| \int_E b_E \leq \sum_{T \in \omega_E} \|f(h)\|_{L^2(T)} \|b_E\|_{L^2(T)} + \|\nabla(u(h) - u_h)\|_{L^2(T)} \|\nabla b_E\|_{L^2(T)}$$

Et d'après **10.c**,

$$|[n_E \cdot \nabla u_h]_E| \int_E b_E \leq \sum_{T \in \omega_E} \left(\|f(h)\|_{L^2(T)} + C h_E^{-1} \|\nabla(u(h) - u_h)\|_{L^2(T)} \right) \|b_E\|_{L^2(T)}$$

Or $\|b_E\|_{L^2(T)} \leq C h_E$ et $\int_E b_E = 2h_E/3$. Ainsi,

$$|[n_E \cdot \nabla u_h]_E| h_E \leq C \sum_{T \in \omega_E} \left(\|f(h)\|_{L^2(T)} + h_E^{-1} \|\nabla(u(h) - u_h)\|_{L^2(T)} \right) h_E.$$

On obtient la majoration souhaitée en divisant cette inégalité par $h_E^{1/2}$.

c. On a prouvé à la question **9.a** que pour tout triangle T,

$$h_T \|f_T\|_{L^2(T)} \leq C\|u(u) - u_h\|_{H^1(T)}$$

Comme h_T et h_E sont du même ordre de grandeur, il s'en suit que

$$h_E^{1/2} \|f_T\|_{L^2(T)} \leq C h_E^{-1/2} \|u(u) - u_h\|_{H^1(T)}.$$

D'après la question **11.b** , on a donc

$$\|[n_E \cdot \nabla u_h]_E\|_{L^2(E)} = h_E^{1/2} |[n_E \cdot \nabla u_h]_E| \leq C \sum_{T \in \omega_E} h_E^{-1/2} \|\nabla(u(h) - u_h)\|_{L^2(T)}.$$

12)

a. Par intégration par partie,

$$\int_{\omega_E} \nabla u_h \cdot \nabla b_E = \int_E n_E \cdot \nabla u_h b_E.$$

En utilisant la formulation vérifiée par $u(h)$, on en déduit que

$$\int_E (g_E - n_E \cdot \nabla u_h) b_E = \int_{\omega_E} \nabla(u(h) - u_h) \cdot \nabla b_E - \int_{\omega_E} f(h) b_E.$$

En effectuant exactement le même calcul que lors des questions **11.b** et **11.c** (quitte à remplacer $-[n_E \cdot \nabla u_h]_E$ par $g_E - n_E \cdot \nabla u_h$), on obtient

$$\|g_E - n_E \cdot \nabla u_h\|_{L^2(E)} \leq C h_E^{1/2} \|u(h) - u_h\|_{H^1(\omega_E)}.$$

13) En utilisant les résultats obtenus aux questions **10.c**, **11.c** et **12.a**, on en déduit que

$$\eta_T^2 \leq C\|u(h) - u_h\|_{H^1(\omega_T)}^2,$$

d'où

$$\sum_{T \in \mathcal{T}_h} \eta_T^2 \leq C\|u(h) - u_h\|_{H^1(\Omega)}^2.$$

L'estimation a posteriori est donc efficace.

Problème 2 - Dynamique de Hellmann-Feynmann

Question 1.

1.a Comme E et $y \mapsto y(q)$ sont de classe C^1, il est en de même de la fonction $q \mapsto V(q) = E(q, y(q))$ et on a

$$\frac{\partial V}{\partial q_i}(q) = \frac{\partial E}{\partial q_i}(q, y(q)) + \langle E'_y(q, y(q)), \frac{\partial y}{\partial q_i}(q)\rangle.$$

où E'_y désigne la différentielle par rapport à y de la fonction E. Comme $c'(y(q)) \neq 0$, on peut appliquer le théorème 10.2.8 : il existe $\lambda \in \mathbb{R}$ tel que

$$E'_y(q, y(q)) + \lambda c'(y(q)) = 0$$

Par ailleurs, $c(y(q)) = 0$ pour tout $q \in \mathbb{R}^d$. En dérivant par rapport à q_i, il vient

$$\langle c'(y(q)), \frac{\partial y}{\partial q_i}(q)\rangle = 0.$$

En conséquence,

$$\frac{\partial V}{\partial q_i}(q) = \frac{\partial E}{\partial q_i}(q, y(q)) + \langle E'_y(q, y(q)), \frac{\partial y}{\partial q_i}(q)\rangle = \frac{\partial E}{\partial q_i}(q, y(q)) - \lambda\langle c'(y(q)), \frac{\partial y}{\partial q_i}(q)\rangle = \frac{\partial E}{\partial q_i}(q, y(q)).$$

D'où (14).

1.b La formule (14) permet de calculer les composantes de la force $-\nabla V(q)$ à partir de $y(q)$ sans avoir à calculer les dérivées de $y(q)$, ce que nécessiterait l'utilisation naïve de la règle de la chaîne.

Question 2.

2.a Pour $x \in \mathbb{R}^n \setminus \{0\}$, on pose $\phi_x(M) = \frac{x^* M x}{|x|^2}$. Il est clair que ϕ_x est une forme linéaire sur l'espace vectoriel $\mathcal{M}_S(n)$. De plus l'application trace est également une forme linéaire sur $\mathcal{M}_S(n)$. Comme la condition $0 \leq M \leq I_n$ équivaut à $0 \leq \phi_x(M) \leq 1$ pour tout $x \in \mathbb{R}^n \setminus \{0\}$, il vient

$$\mathcal{K} = \left(\bigcap_{x \in \mathbb{R}^n} \phi_x^{-1}([0, 1]) \right) \bigcap \mathrm{Tr}^{-1}(\{p\}).$$

Les ensembles $\phi_x^{-1}([0, 1])$ et $\mathrm{Tr}^{-1}(\{p\})$ étant fermés (en tant qu'images réciproques d'ensembles fermés par une application continue) et une intersection quelconque d'ensembles fermés étant fermée, il en résulte que $\mathcal{K}$ est fermé.

Soit $M \in \mathcal{K}$, λ une valeur propre de M et $x \in \mathbb{R}^n \setminus \{0\}$ un vecteur propre associé. Comme $0 \leq M \leq I_n$, on a $0 \leq x^* M x \leq |x|^2$. Or $x^* M x = \lambda |x|^2$. Donc $0 \leq \lambda \leq 1$, ce qui implique que $\|M\|_2 \leq 1$. Donc $\mathcal{K}$ est borné.

La matrice diagonale par bloc $M = \begin{pmatrix} I_p & 0 \\ 0 & 0_{n-p} \end{pmatrix}$ est dans $\mathcal{K}$. Donc $\mathcal{K}$ est non vide.

Enfin, si M_1 et M_2 sont dans $\mathcal{K}$ et si $\theta \in [0,1]$, alors la matrice $\theta M_1 + (1-\theta)M_2$ est symétrique, $\mathrm{Tr}\,(\theta M_1 + (1-\theta)M_2) = \theta \mathrm{Tr}\,(M_1) + (1-\theta)\mathrm{Tr}\,(M_2) = p$, et

$$\forall x \in \mathbb{R}^n \setminus \{0\}, \quad \phi_x(\theta M_1 + (1-\theta)M_2) = \theta\phi_x(M_1) + (1-\theta)\phi_x(M_2) \in [0,1].$$

Donc $\theta M_1 + (1-\theta)M_2 \in \mathcal{K}$, ce qui prouve que $\mathcal{K}$ est convexe.

2.b Posons $F_q(M) = E(q, M)$. On a

$$F_q(M+P) = \mathrm{Tr}\,(H(q)(M+P)) + \frac{1}{2}\|M+P\|_{\mathrm{F}}^2 = F_q(M) + \mathrm{Tr}\,((H(q)+M)P) + \frac{1}{2}\|P\|_{\mathrm{F}}^2.$$

Donc $\langle F_q'(M), P\rangle = \mathrm{Tr}\,((H(q)+M)P)$ et $F_q''(M)(P,P) = \|P\|_{\mathrm{F}}^2$.

Il en résulte

- que F_q est α-convexe sur $\mathcal{M}_S(n)$ avec $\alpha = 1$ (voir proposition 10.1.5 et exercice 10.1.9 du cours), et qu'en conséquence (15) admet un unique point de minimum $M(q)$ sur $\mathcal{K}$ (cf. théorème 9.2.6) ;
- que $M(q)$ vérifie l'inéquation d'Euler (cf. théorème 10.2.1)

$$\forall M \in \mathcal{K}, \quad \langle F_q'(M(q)), M - M(q)\rangle \geq 0,$$

ce qui s'écrit aussi

$$\forall M \in \mathcal{K}, \quad \mathrm{Tr}\,((H(q)+M(q))M) \geq \mathrm{Tr}\,((H(q)+M(q))M(q)).$$

2.c Comme $\nabla F_q(M) = H(q) + M$, on reconnaît dans

$$M_{k+1} = \Pi_{\mathcal{K}}\left(M_k - \mu(H(q) + M_k)\right)$$

l'algorithme du gradient à pas fixe avec projection. Or on sait que F_q est α-convexe avec $\alpha = 1$. De plus F_q' est Lipschizienne de constante de Lipschitz égale à 1. En effet

$$\|\nabla F_q(M_1) - \nabla F_q(M_2)\|_{\mathrm{F}} = \|M_1 - M_2\|_{\mathrm{F}}.$$

Il résulte donc du théorème 10.5.8 que l'algorithme du gradient à pas fixe avec projection converge pour tout $0 < \mu < 2$.

Question 3.

3.a Posons $V = \mathbb{R}^n$ et $J(x) = \sum_{i=1}^n |x_i - y_i|^2$. L'ensemble X_{ad} est un compact, convexe non vide de V et la fonction J est α-convexe sur V. Le problème (18) admet donc un unique point de minimum x^0.

On est dans le cadre d'application du théorème 10.2.19 avec $N = 1$, $G(x) = \sum_{i=1}^n x_i - p$, $M = 2n$, $F_i(x) = -x_i$, $F_{i+n}(x) = x_i - 1$. Les conditions d'optimalités s'écrivent donc

$$x_i^0 - y_i + \mu - \lambda_i + \lambda_{i+N} = 0, \quad \lambda_j \geq 0, \quad 0 \leq x_i^0 \leq 1, \quad \lambda_i x_i^0 = \lambda_{i+N}(x_i^0 - 1) = 0.$$

Pour chaque $1 \leq i \leq N$, on a donc trois possibilités :

1. $x_i^0 = y_i - \mu$

2. $x_i^0 = 0$, auquel cas $\lambda_i \geq 0$, $\lambda_{i+N} = 0$, et $0 = x_i^0 = y_i - \mu + \lambda_i \geq y_i - \mu$
3. $x_i^0 = 1$, auquel cas $\lambda_i = 0$, $\lambda_{i+N} \geq 0$, et $1 = x_i^0 = y_i - \mu - \lambda_{i+N} \leq y_i - \mu$.

Il en résulte que
$$x_i^0 = \max(0, \min(y_i - \mu, 1)).$$
Il reste à déterminer la valeur du multiplicateur de Lagrange μ. Comme $\sum_{i=1}^n x_i^0 = p$, μ est nécessairement solution de l'équation $f_y(\mu) = p$.

Notons que la fonction f_y est continue et décroissante sur $\mathbb{R}$, et telle que $f_y(-\infty) = n$, $f_y(+\infty) = 0$. L'équation $f_y(\mu) = p$ admet donc au moins une solution (ce qu'on savait déjà puisque (18) a un point de minimum). Il est possible que $f_y^{-1}(\{p\})$ soit un intervalle compact de $\mathbb{R}$, mais dans ce cas chacun des x_i^0 est égaal soit à 0, soit à 1, et la valeur de x^0 est la même quel que soit le choix de μ dans cet intervalle. On peut vérifier cela soit directement, soit en invoquant le théorème de Kuhn et Tucker, qui s'applique ici car toutes les contraintes sont affines.

3.b Posons $\widetilde{\mathcal{P}} = \left\{ N = \mathrm{diag}(N_{11}, \cdots, N_{nn}), \quad 0 \leq N_{ii} \leq 1, \quad \sum_{i=1}^n N_{ii} = p \right\}$.

Soit $N \in \mathcal{P}$. Comme $N \in \Delta(n)$, $N = \mathrm{diag}(N_{11}, \cdots, N_{nn})$ et la condition $\mathrm{Tr}\,(N) = p$ s'écrit $\sum_{i=1}^n N_{ii} = p$. On a montré par ailleurs précédemment que toutes les valeurs propres d'un élément de $\mathcal{K}$ étaient dans l'intervalle $[0, 1]$. Donc $0 \leq N_{ii} \leq 1$ pour tout i. Ceci prouve que $N \in \widetilde{\mathcal{P}}$.

Réciproquement, soit $N \in \widetilde{\mathcal{P}}$. Il est clair que $N \in \Delta(n)$ et que $\mathrm{Tr}\,(N) = p$. Pour tout $x \in \mathbb{R}^n \setminus \{0\}$,
$$\phi_x(N) = \frac{\sum_{i=1}^n N_{ii}|x_i|^2}{\sum_{i=1}^n |x_i|^2}.$$
D'où $0 \leq \min_{1 \leq i \leq n} N_{ii} \leq \phi_x(N) \leq \max_{1 \leq i \leq n} N_{ii} \leq 1$, ce qui prouve que $0 \leq N \leq I_n$. Donc $N \in \Delta(n) \cap \mathcal{K} = \mathcal{P}$.

3.c Soit $M \in \mathcal{M}_S(n)$ et $N \in \mathcal{M}_S(n)$ définie par $N_{ij} = M_{ij}\delta_{ij}$. Il est clair que $N \in \Delta(n)$. De plus $M - N = M_{ij}(1 - \delta_{ij}) \in \Delta(n)^\perp$. Donc $N = \Pi_{\Delta(n)}(M)$.

Soit $M \in \mathcal{K}$ et $N = \Pi_{\Delta(n)}(M)$. Comme $N_{ij} = M_{ij}\delta_{ij}$, on a déjà $N \in \Delta(n)$ et $\sum_{i=1}^n N_{ii} = \sum_{i=1}^n M_{ii} = \mathrm{Tr}\,(M) = p$. Soit e_i est le vecteur de $\mathbb{R}^n$ dont la i-ème composante vaut 1 et dont toutes les autres composantes sont nulles. Comme $0 \leq M \leq I_n$, on a $0 \leq e_i^* M e_i \leq |e_i|^2 = 1$. Or $N_{ii} = M_{ii} = e_i^* M e_i$. D'où $0 \leq N_{ii} \leq 1$ pour tout i, ce qui montre que $N \in \mathcal{P}$.

3.d Soit $N \in \Delta(n)$ une matrice diagonale. On a $(N - \Pi_{\Delta(n)}(\Pi_{\mathcal{K}}(N))) \in \Delta(n)$ et $(\Pi_{\mathcal{K}}(N) - \Pi_{\Delta(n)}(\Pi_{\mathcal{K}}(N))) \in \Delta(n)^\perp$. Donc
$$\begin{aligned} \|N - \Pi_{\mathcal{K}}(N)\|_{\mathrm{F}}^2 &= \left\| \left(N - \Pi_{\Delta(n)}(\Pi_{\mathcal{K}}(N))\right) + \left(\Pi_{\Delta(n)}(\Pi_{\mathcal{K}}(N)) - \Pi_{\mathcal{K}}(N)\right) \right\|_{\mathrm{F}}^2 \\ &= \left\| N - \Pi_{\Delta(n)}(\Pi_{\mathcal{K}}(N)) \right\|_{\mathrm{F}}^2 + \left\| \Pi_{\Delta(n)}(\Pi_{\mathcal{K}}(N)) - \Pi_{\mathcal{K}}(N) \right\|_{\mathrm{F}}^2. \end{aligned}$$
D'après la question précédente, $\Pi_{\Delta(n)}(\Pi_{\mathcal{K}}(N)) \in \mathcal{K}$. Or $\Pi_{\mathcal{K}}(N)$ est caractérisé par $\Pi_{\mathcal{K}}(N) \in \mathcal{K}$ et $\|N - \Pi_{\mathcal{K}}(N)\|_{\mathrm{F}}^2 = \min_{\widetilde{N} \in \mathcal{K}} \|N - \widetilde{N}\|_{\mathrm{F}}^2$. Il en résulte que $\Pi_{\mathcal{K}}(N) = \Pi_{\Delta(n)}(\Pi_{\mathcal{K}}(N))$, ce qui montre que $\Pi_{\mathcal{K}}(N) \in \Delta(n)$. Ainsi, $\Pi_{\mathcal{K}}(N) \in \Delta(n) \cap \mathcal{K} = \mathcal{P}$.

3.e Pour tout $\widetilde{N} = \mathrm{diag}(x_1, \cdots, x_n) \in \Delta(n)$,

$$\|N - \widetilde{N}\|_{\mathrm{F}}^2 = \sum_{i=1}^{n} |x_i - N_{ii}|^2.$$

Les éléments diagonaux de $\Pi_{\mathcal{K}}(N)$ peuvent donc être obtenus en résolvant le problème d'optimisation

$$\min \left\{ \sum_{i=1}^{n} |x_i - N_{ii}|^2, \quad 0 \leq x_i \leq 1, \quad \sum_{i=1}^{n} x_i = p \right\}.$$

Un algorithme de calcul de $\Pi_{\mathcal{K}}(N)$ est donc :

1. Chercher par dichotomie une solution μ_* de l'équation $f_N(\mu) = p$ où

$$f_N(\mu) = \sum_{i=1}^{n} \max(0, \min(N_{ii} - \mu, 1)).$$

   Cette étape ne pose pas de difficulté car f_N est continue et décroissante sur $\mathbb{R}$ et qu'on a $f_N(\min(N_{ii}) - 1) = n$ et $f_N(\max(N_{ii})) = 0$.

2. Poser

$$\Pi_{\mathcal{K}}(N) = \mathrm{diag}(\max(0, \min(N_{11} - \mu_*, 1)), \cdots \max(0, \min(N_{nn} - \mu_*, 1))).$$

3.f Soit $U \in \mathcal{M}(n)$ une matrice orthogonale et $\mathcal{K}_U = U^*\mathcal{K}U$. Soit $\widetilde{M} \in \mathcal{K}_U$ et $M \in \mathcal{K}$ tel que $\widetilde{M} = U^*MU$. On a $\widetilde{M}^* = (U^*MU)^* = U^*M^*U = U^*MU = \widetilde{M}$, Tr $(\widetilde{M}) =$ Tr $(U^*MU) =$ Tr $(UU^*M) =$ Tr $(M) = p$ et pour tout $x \in \mathbb{R}^n \setminus \{0\}$, $\phi_x(\widetilde{M}) = \frac{x^*(U^*MU)x}{|x|^2} = \frac{(Ux)^*M(Ux)}{|Ux|^2} = \phi_{Ux}(M)$. Donc $0 \leq \phi_x(\widetilde{M}) \leq 1$. On a donc bien $\widetilde{M} \in \mathcal{K}$ et donc $\mathcal{K}_U \subset \mathcal{K}$.

Ceci étant vrai pour tout matrice orthogonale U, on a en particulier $\mathcal{K}_{U^*} \subset \mathcal{K}$. Par suite $\mathcal{K} = U^*\mathcal{K}_{U^*}U \subset U^*\mathcal{K}U = \mathcal{K}_U \subset \mathcal{K}$. Finalement $\mathcal{K}_U = \mathcal{K}$.

3.g Si $M \in \mathcal{M}_S(n)$ et $\widetilde{M} \in \mathcal{M}_S(n)$, on a pour toute matrice unitaire $U \in \mathcal{M}(n)$,

$$\begin{aligned} \|U^*MU - U^*\widetilde{M}U\|_{\mathrm{F}}^2 &= \|U^*(M - \widetilde{M})U\|_{\mathrm{F}}^2 = \mathrm{Tr}\ (U^*(M - \widetilde{M})UU^*(M - \widetilde{M})U) \\ &= \mathrm{Tr}\ ((M - \widetilde{M})^2) = \|M - \widetilde{M}\|_{\mathrm{F}}^2. \end{aligned}$$

Comme $U^*\mathcal{K}U = \mathcal{K}$, on a $U^*\Pi_{\mathcal{K}}(M)U \in \mathcal{K}$ et

$$\begin{aligned} \|U^*MU - U^*\Pi_{\mathcal{K}}(M)U\|_{\mathrm{F}}^2 &= \|M - \Pi_{\mathcal{K}}(M)\|_{\mathrm{F}}^2 = \min_{\widetilde{M} \in \mathcal{K}} \|M - \widetilde{M}\|_{\mathrm{F}}^2 \\ &= \min_{\widetilde{M} \in \mathcal{K}} \|U^*MU - U^*\widetilde{M}U\|_{\mathrm{F}}^2 \\ &= \min_{Q \in U^*\mathcal{K}U} \|U^*MU - Q\|_{\mathrm{F}}^2 = \min_{Q \in \mathcal{K}} \|U^*MU - Q\|_{\mathrm{F}}^2. \end{aligned}$$

En conséquence $\Pi_{\mathcal{K}}(U^*MU) = U^*\Pi_{\mathcal{K}}(M)U$. Pour calculer $\Pi_{\mathcal{K}}(M)$, on peut donc

1. Diagonaliser M dans une base orthogonale, i.e. exhiber une matrice orthogonale $U \in \mathcal{M}(n)$ telle que $N = U^*MU \in \Delta(n)$.

2. Calculer $\Pi_{\mathcal{K}}(N)$ en utilisant l'algorithme construit à la question 3.e.
3. Calculer $\Pi_{\mathcal{K}}(M)$ par la formule $\Pi_{\mathcal{K}}(M) = U\Pi_{\mathcal{K}}(N)U^*$.

Question 4.

4.a U est une matrice orthogonale et $M(q) \in \mathcal{K}$. Il découle donc du résultat de la question 3.f que $U^*M(q)U \in \mathcal{K}$. En utilisant le résultat de la question 3.c, on obtient $N(q) = \Pi_{\Delta(n)}(U^*M(q)U) \in \mathcal{P}$.

Comme $S(q) = UD(q)U^*$, l'inéquation d'Euler s'écrit

$$\forall M \in \mathcal{K}, \quad \mathrm{Tr}\,(UD(q)U^*M) \geq \mathrm{Tr}\,(UD(q)U^*M(q)),$$

soit

$$\forall M \in \mathcal{K}, \quad \mathrm{Tr}\,(D(q)U^*MU) \geq \mathrm{Tr}\,(D(q)U^*M(q)U)$$

soit encore

$$\forall \widetilde{M} \in \mathcal{K}_U = \mathcal{K}, \quad \mathrm{Tr}\,(D(q)\widetilde{M}) \geq \mathrm{Tr}\,(D(q)U^*M(q)U).$$

Etant donné que $\mathcal{P} \subset \mathcal{K}$, on a donc aussi

$$\forall N \in \mathcal{P}, \quad \mathrm{Tr}\,(D(q)N) \geq \mathrm{Tr}\,(D(q)U^*M(q)U).$$

En notant que $D(q) \in \Delta(n)$ et que $N(q) = \Pi_{\Delta(n)}(U^*M(q)U)$, on obtient

$$\mathrm{Tr}\,(D(q)U^*M(q)U) = (D(q), U^*M(q)U)_{\mathrm{F}} = (D(q), N(q))_{\mathrm{F}} = \mathrm{Tr}\,(D(q)N(q)).$$

Finalement,

$$\forall N \in \mathcal{P}, \quad \mathrm{Tr}\,(D(q)N) \geq \mathrm{Tr}\,(D(q)N(q)).$$

4.b L'ensemble $\mathcal{P}$ est un polyèdre puisque c'est l'intersection des $2n+2$ demi-espaces de $\Delta(n)$ définis par $N_{ii} \geq 0$, $1 - N_{ii} \geq 0$, $\sum_{i=1}^n N_{ii} - p \geq 0$ et $p - \sum_{i=1}^n N_{ii} \geq 0$.

Notons $\mathcal{P}_{\mathrm{e}}$ l'ensemble des éléments de $\mathcal{P}$ dont p coefficients diagonaux sont égaux à 1, les $n-p$ autres coefficients diagonaux étant égals à 0. Nous allons montrer que $\mathcal{P}_{\mathrm{ext}} = \mathcal{P}_{\mathrm{e}}$.

Soit $N \in \mathcal{P} \setminus \mathcal{P}_{\mathrm{e}}$. Tous les coefficients diagonaux de N ne sont pas dans l'ensemble $\{0, 1\}$. Sinon, la contrainte $\mathrm{Tr}\,(N) = p$ contredirait le fait que $N \notin \mathcal{P}_{\mathrm{e}}$. Comme par ailleurs, $p \in \mathbb{N}$, au moins deux coefficients diagonaux de N sont dans $]0,1[$. Soit donc $1 \leq i_1 < i_2 \leq n$ tel que $0 < N_{i_1i_1} < 1$ et $0 < N_{i_2,i_2} < 1$. Soit $N_\epsilon^\pm \in \mathbb{R}^n$ obtenu à partir de N en remplaçant $N_{i_1i_1}$ par $N_{i_1i_1} \pm \epsilon$ et $N_{i_2i_2}$ par $N_{i_2i_2} \mp \epsilon$. Il est clair que pour $\epsilon > 0$ assez petit $N_\epsilon^\pm \in \mathcal{P}$, $N_\epsilon^+ \neq N_\epsilon^-$ et $N = \frac{1}{2}(N_\epsilon^+ + N_\epsilon^-)$. Donc $N \notin \mathcal{P}_{\mathrm{ext}}$.

Réciproquement soit $N \in \mathcal{P}_{\mathrm{e}}$, $P \in \mathcal{P}$, $Q \in \mathcal{P}$ avec $P \neq Q$ et $\theta \in [0,1]$ tel que $N = \theta P + (1-\theta)Q$. Soit $1 \leq i \leq n$ tel que $P_{ii} \neq Q_{ii}$. Il vient $\theta P_{ii} + (1-\theta)Q_{ii} \in \{0,1\}$. Comme $0 \leq P_{ii} \leq 1$, $0 \leq Q_{ii} \leq 1$, $P_{ii} \neq Q_{ii}$ et $\theta \in [0,1]$, on a nécessairement $\theta \in \{0,1\}$. Donc $N \in \mathcal{P}_{\mathrm{ext}}$.

Le cardinal de l'ensemble $\mathcal{P}_{\mathrm{ext}}$ est donc égal à $\begin{pmatrix} n \\ p \end{pmatrix}$.

La fonction $N \mapsto \mathrm{Tr}\,(D(q)N)$ est linéaire. Son minimum sur le polyèdre convexe $\mathcal{P}$ est donc atteint en l'un des points extrémaux de $\mathcal{P}$. Donc

$$\mathrm{Tr}\,(D(q)N(q)) = \min_{N \in \mathcal{P}} \mathrm{Tr}\,(D(q)N) = \min_{N \in \mathcal{P}_{\mathrm{ext}}} \mathrm{Tr}\,(D(q)N) = \min_{1 \leq i_1 \leq \cdots \leq i_p \leq n} \sum_{j=1}^{p} \epsilon_{i_j}(q) = \sum_{j=1}^{p} \epsilon_j(q).$$

4.c Si $\epsilon_p(q) < \epsilon_{p+1}(q)$ l'unique point de minimum sur $\mathcal{P}$ de la fonction $N \mapsto \mathrm{Tr}\,(D(q)N)$ est le point extrémal

$$N_0 = \begin{pmatrix} I_p & 0 \\ 0 & 0_{n-p} \end{pmatrix}.$$

Donc $N(q) = N_0$. D'où

$$M(q) = UN(q)U^*.$$

On pourra remarquer que

$$M(q) = \sum_{i=1}^{p} x_i(q) x_i(q)^*$$

où $(x_1(q), \cdots, x_n(q))$ est une base orthonormale de vecteurs propres de $M(q)$ telle que pour tout $1 \leq i \leq n$, $S(q)x_i(q) = \epsilon_i(q)x_i(q)$.