CORRIGÉ DU CONTRÔLE CLASSANT DE MATHÉMATIQUES APPLIQUÉES
«Analyse numérique et optimisation»
Mardi 1^{er} Juillet 2003

Durée: 4 heures

Sujet proposé par Stéphane Gaubert et Claude Le Bris

Corrigé du Problème I: Méthodes de pénalisation pour le problème de Stokes

1. Méthode de pénalisation au niveau continu

Réponse à la Question 1.1 La fonctionnelle J^ε est la somme de la fonctionnelle fortement convexe $v \longrightarrow \dfrac{1}{2}\displaystyle\int_\Omega |\nabla v|^2$ sur $(H_0^1(\Omega))^3$ (à cause de l'inégalité de Poincaré), de la fonctionnelle quadratique positive (donc convexe) $v \longrightarrow \dfrac{1}{2\varepsilon}\displaystyle\int_\Omega |\operatorname{div} v|^2$ et de la fonctionnelle linéaire $v \longrightarrow -\displaystyle\int_\Omega f \,.\, v$. Elle est donc fortement convexe. Pour la même raison, elle est continue. Il s'ensuit, par le Théorème 9.2.6 du cours, qu'elle admet un minimiseur unique sur $(H_0^1(\Omega))^3$.

Réponse à la Question 1.2

C'est un simple calcul de voir que la fonctionnelle J^ε est en fait différentiable sur $(H_0^1(\Omega))^3$. L'équation d'optimalité vérifiée par son minimiseur u^ε sur l'espace vectoriel $(H_0^1(\Omega))^3$ exprime l'annulation de la différentielle première, c'est-à-dire

$$\int_\Omega \nabla u^\varepsilon(x).\nabla v(x)\,dx + \frac{1}{\varepsilon}\int_\Omega \operatorname{div} u^\varepsilon(x)\operatorname{div} v(x)\,dx = \int_\Omega f(x)v(x), \qquad \forall v \in (H_0^1(\Omega))^3. \tag{A.1}$$

Cette équation se met bien sous la forme (I.4) où

$$a^\varepsilon(w,v) = \int_\Omega \nabla w(x).\nabla v(x)\,dx + \frac{1}{\varepsilon}\int_\Omega \operatorname{div} w(x)\operatorname{div} v(x)\,dx,$$

$$L(w) = \int_\Omega f(x)v(x).$$

Réponse à la Question 1.3

La fonctionnelle J^ε est fortement convexe donc strictement convexe, et différentiable: tout point qui annule sa différentielle est en fait son unique point critique. Cela montre l'unicité de la solution de (I.4).

Mais en fait on peut procéder autrement. Comme la forme a^ε est bilinéaire bicontinue coercive (à cause de l'inégalité de Poincaré), et la forme L est linéaire continue, on sait par le Théorème de Lax-Milgram qu'il existe une solution unique w à la formulation variationnelle *Trouver* $w \in (H_0^1(\Omega))^3$ *vérifiant*

$$a^\varepsilon(w,v) = L(v), \text{pour tout } v \in (H_0^1(\Omega))^3.$$

Comme $w = u^\varepsilon$ convient, c'est donc la solution.

Réponse à la Question 1.4

Si $h \in H^1(\Omega)$, alors (voir la Définition 4.2.6 et le Lemme 4.2.7 du cours), la divergence faible de la fonction ∇h, est une fonction $L^2(\Omega)$ qui vérifie

$$\int_\Omega (-\operatorname{div}(\nabla h))\,v = \int_\Omega \nabla h(x)\ .\nabla v(x)\,dx,$$

pour toute fonction $v \in C_c^\infty(\Omega)$ En appliquant cela composante par composante pour la fonction u^ε, on obtient la définition du Laplacien au sens faible de u^ε

$$\int_\Omega (-\Delta u^\varepsilon)\,.\,v = \int_\Omega \nabla u^\varepsilon(x)\ .\nabla v(x)\,dx$$

pour tout $v \in (C_c^\infty(\Omega))^3$. En procédant de même, on a

$$\int_\Omega (-\nabla(\operatorname{div}(u^\varepsilon))\,.\,v = \int \operatorname{div}(u^\varepsilon)(x)\operatorname{div}(v)(x)\,dx,$$

pour tout $v \in (C_c^\infty(\Omega))^3$. La formulation variationnelle (I.4) s'écrit alors

$$-\Delta u^\varepsilon - \frac{1}{\varepsilon}\nabla(\mathrm{div}(u^\varepsilon)) = f, \tag{A.2}$$

au sens faible.

Réponse à la Question 1.5 Formellement, si on pose $p = p^\varepsilon = -\frac{1}{\varepsilon}\mathrm{div}\,(u^\varepsilon)$ on retrouve (I.5) à partir de (I.1).

Réponse à la Question 1.6
On prend $v = u_\varepsilon$ dans (A.1), et on obtient :

$$\begin{aligned}
\int_\Omega |\nabla u^\varepsilon(x)|^2\,dx + \frac{1}{\varepsilon}\int_\Omega |\mathrm{div}\,u^\varepsilon|^2(x)\,dx &= \int_\Omega f(x)u^\varepsilon(x) \\
&\le \left(\int_\Omega |f|^2\right)^{1/2}\left(\int_\Omega (u^\varepsilon)^2\right)^{1/2} \\
&\quad \text{par l'inégalité de Cauchy-Schwartz} \\
&\le \sqrt{C}\left(\int_\Omega |f|^2\right)^{1/2}\left(\int_\Omega |\nabla u^\varepsilon(x)|^2\,dx\right)^{1/2}
\end{aligned} \tag{A.3}$$

où C est la constante de l'inégalité de Poincaré (Lemme 3.3.6 du cours). D'où en particulier

$$\int_\Omega |\nabla u^\varepsilon(x)|^2\,dx \le \sqrt{C}\left(\int_\Omega |f|^2\right)^{1/2}\left(\int_\Omega |\nabla u^\varepsilon(x)|^2\right)^{1/2}$$

et donc en simplifiant, (I.6) avec $C_1 = C$. En injectant alors cette information (I.6) dans le membre de droite de (A.3), on obtient, cette fois en négligeant le premier terme du membre de gauche, l'estimation (I.7) avec $C_2 = C_1 = C$.

Réponse à la Question 1.7
Il suffit d'écrire les notions de divergence, gradient et Laplacien au sens faible du cours (voir Chapitre 4 et ce qui a été fait ci-dessus en réponse à la Question 1.4).

Réponse à la Question 1.8

On utilise $v = u_\varepsilon - u$ comme fonction test dans les deux formulations (I.4) (c'est-à-dire (A.1) ci-dessus) et (I.8), puis on soustrait l'une à l'autre. Nous obtenons

$$\int_\Omega |\nabla(u_\varepsilon - u)|^2 + \frac{1}{\varepsilon}\int_\Omega |\mathrm{div}\,(u_\varepsilon - u)|^2 + \int_\Omega p\,\mathrm{div}\,(u_\varepsilon - u) = 0,$$

puisque $\mathrm{div}(u) = 0$. Nous en déduisons

$$\int_\Omega |\nabla(u_\varepsilon - u)|^2 + \frac{1}{\varepsilon}\int_\Omega |\mathrm{div}\,(u_\varepsilon - u)|^2 \le \|p\|_{L^2(\Omega)^N}\|\mathrm{div}\,(u_\varepsilon - u)\|_{L^2(\Omega)}, \tag{A.4}$$

et donc en particulier

$$\frac{1}{\varepsilon}\int_\Omega |\mathrm{div}\,(u_\varepsilon - u)|^2 \le \|p\|_{L^2(\Omega)^N}\|\mathrm{div}\,(u_\varepsilon - u)\|_{L^2(\Omega)},$$

d'où

$$\|\mathrm{div}\,(u_\varepsilon - u)\|_{L^2(\Omega)} = O(\varepsilon). \tag{A.5}$$

En injectant cette information dans (A.4), on trouve alors

$$\int_\Omega |\nabla(u_\varepsilon - u)|^2 = O(\varepsilon) \longrightarrow 0, \tag{A.6}$$

quand ε tend vers 0, et donc, par l'inégalité de Poincaré, la convergence dans $H_0^1(\Omega)^3$ de u_ε vers u. Ceci montre que la résolution du problème pénalisé (I.4) conduit asymptotiquement, quand $\varepsilon \longrightarrow 0$, à la solution vitesse du problème de Stokes.

2. Discrétisation de cette méthode par éléments finis

Réponse à la Question 2.1

Il suffit de décomposer u_h^ε selon les ϕ_i et de tester successivement (I.12) sur chaque $v_h = \phi_j$, $1 \leq j \leq N$. On trouve ainsi le système (I.13) avec

$$(A_h)_{ij} = \int_{\Omega_h} \nabla\phi_i \,.\, \nabla\phi_j,$$

$$(C_h)_{ij} = \int_{\Omega_h} \mathrm{div}\,(\phi_i)\,\mathrm{div}\,(\phi_j),$$

$$(F_h)_i = \int_{\Omega_h} f \,.\, \phi_i.$$

Réponse à la Question 2.2

La matrice A_h est symétrique définie positive puisque

$$(A_h U, U) = \int_{\Omega_h} \left|\nabla\Big(\sum_{i=1}^N U_i\phi_i\Big)\right|^2,$$

et que cette quantité n'est nulle, en vertu de l'inégalité de Poincaré, que quand $U = 0$. Il est de plus immédiat de voir que la matrice C_h est symétrique et positive ou nulle, puisque

$$(C_h U, U) = \int_{\Omega_h} \left|\mathrm{div}\Big(\sum_{i=1}^N U_i\phi_i\Big)\right|^2.$$

La matrice $A_h + \dfrac{1}{\varepsilon} C_h$ est donc aussi symétrique définie positive, et donc inversible. D'où l'existence et l'unicité de U_h^ε.

Réponse à la Question 2.3

Pour une matrice symétrique B, on a

$$\lambda_{\max}(B) = \max_{u\neq 0} \frac{u^* B u}{u^* u} \quad .$$

Donc

$$\begin{aligned}\lambda_{\max}(A+C) &= \max_{u\neq 0} \frac{u^*(A+C)u}{u^*u} = \max_{u\neq 0} \frac{u^*Au + u^*Cu}{u^*u} \\ &\leq \max_{u\neq 0} \frac{u^*Au}{u^*u} + \max_{v\neq 0} \frac{v^*Cv}{v^*v} = \lambda_{\max}(A) + \lambda_{\max}(C) \quad .\end{aligned}$$

On a montré la seconde inégalité dans (I.14). En utilisant la seconde inégalité de (I.14), on obtient

$$\lambda_{\max}(C) = \lambda_{\max}((A+C) + (-A)) \leq \lambda_{\max}(A+C) + \lambda_{\max}(-A) = \lambda_{\max}(A+C) - \lambda_{\min}(A)$$

ce qui montre la première inégalité de (I.14). En appliquant (I.14) à $-A$ et $-C$, on obtient (I.15).

Réponse à la Question 2.4

Il résulte de (I.14) que

$$\lambda_{\min}(A) + \frac{1}{\epsilon}\lambda_{\max}(C) \leq \lambda_{\max}(A + \frac{1}{\epsilon}C) \leq \lambda_{\max}(A) + \frac{1}{\epsilon}\lambda_{\max}(C) \tag{A.7}$$

On note que $\lambda_{\min}(C) = 0$, car C est supposée symétrique positive non-inversible. En appliquant (I.15) on obtient donc

$$\lambda_{\max}(A) \geq \lambda_{\min}(A + \frac{1}{\epsilon}C) \geq \lambda_{\min}(A) \quad . \tag{A.8}$$

Réunissant (A.7) et (A.8), nous obtenons

$$\lambda_{\max}(A)^{-1}(\lambda_{\min}(A) + \frac{1}{\epsilon}\lambda_{\max}(C)) \leq \mathrm{cond}(A + \frac{1}{\epsilon}C) \leq \lambda_{\min}(A)^{-1}(\lambda_{\max}(A) + \frac{1}{\epsilon}\lambda_{\max}(C)) \quad . \tag{A.9}$$

Comme C est symétrique positive, et $C \neq 0$, $\lambda_{\max}(C) > 0$. Les bornes inf et sup dans (A.9) sont donc bien d'un ordre de grandeur $\dfrac{1}{\epsilon}$ quand $\epsilon \to 0$, ce qui montre que $\mathrm{cond}(A + \dfrac{1}{\epsilon}C)$ est d'ordre $\dfrac{1}{\epsilon}$ quand $\epsilon \to 0$.

Dans notre cas, la matrice C_h est susceptible d'avoir un noyau non nul car il existe dans V_{0h} des fonctions à divergence nulle, ou presque (asymptotiquement quand $h \longrightarrow 0$ l'espace V_{0h} approche tout l'espace $(H_0^1(\Omega))^3$). Pour approcher avec une bonne précision la solution u du problème de Stokes, on aura envie, au vu de la Section 1, de prendre ε tout petit, d'où, en appliquant le raisonnement précédent, un conditionnement grand pour $A_h + \dfrac{1}{\varepsilon}C_h$.

Or, on sait que le conditionnement d'une matrice mesure la sensibilité du problème de la résolution du système linéaire associé. Si le conditionnement est grand, la résolution sera d'autant plus sensible aux erreurs et arrondis numériques. Pire, si l'algorithme envisagé est un algorithme itératif, comme le gradient conjugué, la vitesse de convergence de l'algorithme sera d'autant moins bonne que le conditionnement est grand. Dans ce dernier cas, il faudra notamment avoir recours à des techniques de préconditionnement si on ne veut pas un temps de calcul trop long. La méthode décrite dans cette section risque donc de nécessiter de telles techniques.

3. Méthode de pénalisation au niveau discret

Réponse à la Question 3.1 Pour $\varepsilon = 0$, (I.17) est évidemment l'approximation discrète de (I.8).

Réponse à la Question 3.2 En décomposant $\tilde{u}_h^\varepsilon$ sur la base des ϕ_i et $\tilde{p}_h^\varepsilon$ sur la base des ψ_i, et en testant (I.17) sur chaque $v_h = \phi_i$,$1 \leq j \leq N$, $q_h = \psi_j$, $1 \leq j \leq M$, on trouve le système (I.18) avec

$$(A_h)_{ij} = \int_{\Omega_h} \nabla\phi_i \,.\, \nabla\phi_j,\ 1 \leq j \leq N$$

$$(B_h)_{ij} = -\int_{\Omega_h} \psi_i \,\mathrm{div}\,(\phi_j),\ 1 \leq i \leq M,\ 1 \leq j \leq N$$

$$(M_h)_{ij} = \int_{\Omega_h} \psi_i\, \psi_j,\ 1 \leq i \leq M,\ 1 \leq j \leq M$$

$$(F_h)_i = \int_{\Omega_h} f \,.\, \phi_i\ \ 1 \leq j \leq N.$$

Réponse à la Question 3.3 On suit l'indication du texte en écrivant

$$P_h^\varepsilon = \frac{1}{\varepsilon}\,(M_h)^{-1}\,B_h\,\tilde{U}_h^\varepsilon,$$

(la matrice M_h est effectivement inversible puisque symétrique définie positive par construction, les ψ_i, $1 \leq i \leq M$, formant une base de Q_h). En injectant cette expression de P_h^ε dans la première “ligne” de (I.18), on obtient

$$(A_h + \frac{1}{\varepsilon}\,B_h^*\,(M_h)^{-1}\,B_h)\,\tilde{U}_h^\varepsilon = F_h. \tag{A.10}$$

Or M_h est symétrique définie positive, donc $(M_h)^{-1}$ aussi, donc

$$(B_h^*\,(M_h)^{-1}\,B_h U,U) = ((M_h)^{-1}\,(B_h U),(B_h U)) \geq 0$$

pour tout $U \in \mathbb{R}^N$, donc la matrice $B_h^*\,(M_h)^{-1}\,B_h$ est symétrique, positive ou nulle. Comme A_h est symétrique définie positive, par somme on sait donc que la matrice $A_h + \dfrac{1}{\varepsilon}\,B_h^*\,(M_h)^{-1}\,B_h$ est symétrique définie positive, donc inversible. Il s'ensuit que la solution de (A.10), et donc de (I.18), existe et de façon unique.

Réponse à la Question 3.4 Le système (I.18) conduit à l'équation en $\tilde{U}_h^\varepsilon$ (A.10) à savoir

$$(A_h + \frac{1}{\varepsilon}\,B_h^*\,(M_h)^{-1}\,B_h)\,\tilde{U}_h^\varepsilon = F_h,$$

laquelle est différente de l'équation (I.13) puisque $C_h \neq B_h^* (M_h)^{-1} B_h$ en général (C_h ne dépend pas des ψ_i). Les deux approches consistant à discrétiser la fonctionnelle pénalisée (approche suivie à la Section 1 puis 2), ou à pénaliser le système discret (I.17) pour $\varepsilon = 0$ en le transformant en (I.17) pour $\varepsilon > 0$ ne sont donc pas équivalentes.

4. Méthode de Lagrangien augmenté

Réponse à la Question 4.1 On montre que la matrice carrée

$$\mathcal{A} = \begin{pmatrix} A & B^* \\ B & 0 \end{pmatrix}$$

est injective pour prouver que le système (I.19) admet une solution, et ce de façon unique. Si $\mathcal{A} \begin{pmatrix} U \\ P \end{pmatrix} = 0$, on a $BA^{-1}B^*P = 0$, donc $(BA^{-1}B^*P,P)_M = (A^{-1}B^*P,B^*P)_N = 0$, ce qui, comme A^{-1} est symétrique définie positive, impose $B^*P = 0$. Or le noyau de B^* est réduit à $\{0\}$ car B est de rang maximal. Ceci entraîne $P = 0$, d'où $U = -A^{-1}B^*P = 0$ et l'unicité.

Réponse à la Question 4.2 Les fonctions $\mathcal{J}(V) = \frac{1}{2}(AV,V)_N - (F,V)_M$ et $\mathcal{F}(V) = BV$ sont continues et dérivables. En appliquant la Proposition 10.3.2, on sait donc que tout point-selle (V_0,Q_0) de $\mathcal{L}$ sur $\mathbb{R}^{\mathcal{N}} \times \mathbb{R}^M$ est solution de

$$AV_0 - F + B^*Q_0 = 0, \;\; BV_0 = 0,$$

ce qui est exactement le système (I.19). En fait, ici la fonction $\mathcal{J}(V)$ est convexe, et la fonction $\mathcal{F}(V)$ est affine, donc on peut appliquer le Théorème de Kuhn et Tücker (Théorème 10.3.4) et son extension aux contraintes d'égalité affines énoncée dans la Remarque 10.3.5 (voir le corrigé de la Question 2.6 du second problème pour plus de détails sur ce point). On a donc équivalence entre le fait d'être un point-selle et le fait de vérifier (I.19). Il s'ensuit que le seul point-selle de $\mathcal{L}$ sur $\mathbb{R}^N \times \mathbb{R}^M$ est $\begin{pmatrix} U \\ P \end{pmatrix}$ solution de (I.19).

Réponse à la Question 4.3

On applique le même raisonnement qu'à la Question 4.2 pour montrer l'existence et l'unicité du point-selle. Bien sûr, comme $BU^r = 0$, on a $(A + rB^*B)U^r = AU^r$, et donc $\begin{pmatrix} U \\ P \end{pmatrix} = \begin{pmatrix} U^r \\ P^r \end{pmatrix}$, puisque les systèmes admettent une unique solution.

Réponse à la Question 4.4

La matrice $A^{-1}B^*B$ est une matrice symétrique pour le produit scalaire

$$(x,y)_A = (Ax,y).$$

Elle est donc diagonalisable dans une base orthonormée pour ce produit scalaire et on vérifie facilement que si λ est une valeur propre de cette matrice associé au vecteur propre x, alors

$$\lambda(Ax,x) = (Bx,Bx),$$

ce qui impose, comme A est symétrique définie positive, que λ soit un réel positif ou nul. En suivant l'énoncé, notons désormais λ_i les valeurs propres de $A^{-1}B^*B$ sur son image. Il s'agit donc des valeurs propres strictement positives de $A^{-1}B^*B$. Soit Λ la plus grande d'entre elles.

Réponse à la Question 4.5

Posons

$$V_n = U_n - U, \qquad\qquad Q_n = P_n - P.$$

Il est facile de voir que les itérations (I.23) s'écrivent aussi :

$$\left\{ \begin{array}{l} (A + rB^*B)V^n = -B^*Q^n \\ Q^{n+1} = Q^n + \mu BV^n, \end{array} \right. \tag{A.11}$$

d'où, en éliminant V_n,

$$Q_{n+1} = (\mathrm{Id} - \mu B(A + rB^*B)^{-1}B^*)Q_n,$$

et en posant comme indiqué dans l'énoncé $Z_n = A^{-1}B^*Q_n$,

$$Z_{n+1} = \left(\mathrm{Id} - \mu A^{-1}B^*B(1 + rA^{-1}B^*B)^{-1}\right) Z_n. \tag{A.12}$$

On a donc $H = \mathrm{Id} - \mu A^{-1}B^*B(1 + rA^{-1}B^*B)^{-1}$.

Réponse à la Question 4.6

Comme B est de rang maximal M, et que pour tout n, $Z_n = A^{-1}B^*(P_n - P)$, on sait donc que, pour tout n, $Z_n \in \mathrm{Im}(A^{-1}B^*B)$. Il en résulte que Z_n est une combinaison linéaire des vecteurs propres associés aux valeurs propres de $A^{-1}B^*B$ restreinte à son image. Il s'agit donc précisément des valeurs propres strictement positives de $A^{-1}B^*B$. On a donc bien :

$$Z_n = \sum_{i=1}^{N_0} (Z_n)_i e_i$$

On voit que la relation (A.12) s'écrit pour tout i

$$(Z_{n+1})_i = \left(\frac{1 - (\mu - r)\lambda_i}{1 + r\lambda_i}\right) (Z_n)_i. \tag{A.13}$$

Réponse à la Question 4.7

Comme $P^n - P$ et $U^n - U = -(A + rB^*B)^{-1}B^*(P_n - P)$ s'expriment de façon univoque en fonction de Z_n, la convergence de l'algorithme est donc exactement équivalente à celle de Z_n vers zéro et donc à la condition

$$\rho(\mu,r) = \max_{1\leq i\leq N_0} \left|\frac{1 - (\mu - r)\lambda_i}{1 + r\lambda_i}\right| < 1. \tag{A.14}$$

Cette condition se récrit

$$0 < \mu < 2\left(r + \frac{1}{\Lambda}\right).$$

Si elle est remplie, la convergence est géométrique de rapport $\rho(\mu,r)$.

Réponse à la Question 4.8

On a

$$\rho(\mu = r,r) = \max_{1\leq i\leq N_0} \left|\frac{1}{1 + r\lambda_i}\right|$$

et comme tous les λ_i, $1 \leq i \leq N_0$, sont strictement positifs, on a $\rho(\mu = r,r) \longrightarrow 0$ quand $r \longrightarrow +\infty$, ce qui rend la convergence aussi rapide que voulu. Cependant, il faut prendre garde au fait que ceci mesure le "nombre" d'itérations, et que chaque itération de l'algorithme coûte en fait la résolution du système (I.23). Or, pour r grand, on peut anticiper, comme on l'a vu à la Question 2.3 que la résolution de ce système est difficile, car la matrice $A + rB^*B$ est susceptible d'être mal conditionnée. Il faudra donc faire un compromis et choisir un r "intermédiaire", ou alors recourir à des techniques de préconditionnement pour le système linéaire (I.23).

Réponse à la Question 4.9

L'algorithme de Uzawa appliqué directement au système (I.19) correspond bien sûr à prendre $r = 0$ dans l'analyse précédente, de sorte que la convergence arrive pour

$$0 < \mu < 2\,\frac{1}{\Lambda},$$

ce qui est plus restrictif, et le taux de convergence est

$$\rho(\mu,r = 0) = \max_{1\leq i\leq N_0} \left|1 - \mu\lambda_i\right|.$$

Le meilleur paramètre μ possible est donc celui qui minimise le problème

$$\inf_{\mu} \max_{1\leq i\leq N_0} \left|1 - \mu\lambda_i\right|,$$

à savoir $\mu = \dfrac{2}{\Lambda+\lambda}$ (où $\lambda = \min_{1\le i\le N_0}\lambda_i$) et conduit à un taux de convergence $\rho(\mu,r=0) = \dfrac{\Lambda-\lambda}{\Lambda+\lambda}$ qui n'est pas arbitrairement petit (dès que deux λ_i au moins sont différents). L'intérêt de la méthode du Lagrangien augmenté est claire.

Corrigé du Problème II: Réseaux électriques

1. Potentiel électrique de réseaux linéaires

Réponse à la Question 1.1. Si $Gv = 0$, alors $v_i = v_j$ pour tout $(i,j) \in \mathsf{A}$, et donc

$$v_i = v_j \text{ dès que les nœuds } i \text{ et } j \text{ sont les extrémités d'un même fil.} \tag{B.1}$$

Le réseau étant connexe, pour tous nœuds i et j, il existe une suite de nœuds $i_0,\dots,i_k$, avec $i_0 = i$, $i_k = j$, et $C_{i_0 i_1}\cdots C_{i_{k-1}i_k} > 0$. Comme i_0 et i_1 sont les extrémités d'un même fil, d'après (B.1), $v_{i_0} = v_{i_1}$. On conclut de même que $v_{i_1} = v_{i_2}$, ... , $v_{i_{k-1}} = v_{i_k}$. Ainsi, $v_i = v_j$, pour tous $i,j \in \mathsf{N}$, donc $v \in \mathbb{R}1_{\mathsf{N}}$. Réciproquement, si $v \in \mathbb{R}1_{\mathsf{N}}$, on a trivialement $Gv = 0$. (Autrement dit, le gradient -discret- de v est nul, si et seulement si v est constante.)

Réponse à la Question 1.2. On a, en posant $x_{ij} = (Gv)_{ij} = v_j - v_i$,

$$\mathcal{P}(v) = \frac{1}{2}\sum_{(i,j)\in\mathsf{A}} C_{ij}x_{ij}^2$$

et comme $(Dx)_{ij} = C_{ij}x_{ij}$,

$$\mathcal{P}(v) = \frac{1}{2}\sum_{(i,j)\in\mathsf{A}} (Dx)_{ij}x_{ij} = \frac{1}{2}(DGv,Gv)_{\mathsf{A}} = \frac{1}{2}(G^*DGv,v)_{\mathsf{N}} = \frac{1}{2}(-Lv,v)_{\mathsf{N}}$$

où $L = -G^*DG$. On a $L^* = L$ car $D^* = D$.

Réponse à la Question 1.3. Si $v \in \mathbb{R}1_{\mathsf{N}}$, on a $Gv = 0$ d'après la Question 1.1, donc $Lv = -G^*DGv = 0$ d'après la Question 1.2. Réciproquement, si $Lv = 0$, on a $(-Lv,v)_{\mathsf{N}} = (G^*DGv,v)_{\mathsf{N}} = (DGv,Gv)_{\mathsf{A}} = 0$, et comme la forme quadratique $x \mapsto (Dx,x)_{\mathsf{A}}$ est définie positive (D est diagonal à coefficients diagonaux strictement positifs), on a $Gv = 0$. En utilisant à nouveau la Question 1.1, on a $v \in \mathbb{R}1_{\mathsf{N}}$.

Réponse à la Question 1.4. La fonction $v \mapsto \mathcal{P}(v)$ est convexe, car $\mathcal{P}$ est une forme quadratique positive. La fonction $v \mapsto (-b,v)_{\mathsf{N}}$ est linéaire, et a fortiori convexe. Comme J est somme de deux fonctions convexes, J est convexe.

Réponse à la Question 1.5. La fonction $J(v) = \frac{1}{2}(-Lv,v)_{\mathsf{N}} - (b,v)_{\mathsf{N}}$ est différentiable, et son gradient au point v est $J'(v) = -Lv - b$. Pour une fonction convexe différentiable, $J'(v) = 0$ est une condition nécessaire et suffisante pour que v soit un point de minimum de J. Donc v est point de minimum de J si et seulement si $-Lv = b$.

Réponse à la Question 1.6. On a $\ker L = (\operatorname{im} L^*)^\perp = (\operatorname{im} L)^\perp$, car $L = L^*$. Comme $\ker L = \mathbb{R}1_{\mathsf{N}}$ d'après la Question 1.3, $\operatorname{im} L = (\mathbb{R}1_{\mathsf{N}})^\perp$, donc l'équation caractérisant l'optimalité de v, $-Lv = b$, admet une solution si, et seulement si, $(1_{\mathsf{N}},b)_{\mathsf{N}} = 0$. En outre, comme deux solutions de $-Lv = b$ différent d'un élément de $\ker L$, la solution de $-Lv = b$, quand elle existe, est unique à une constante additive près.

2. Dualité potentiels-courants pour les réseaux non-linéaires

Réponse à la Question 2.1. Afin de calculer G^*x, évaluons

$$\begin{aligned}(x,Gv)_{\mathsf{A}} &= \sum_{(i,j)\in\mathsf{A}} x_{ij}(v_j - v_i) = \sum_{(i,j)\in\mathsf{A}} x_{ij}v_j - x_{ij}v_i = \sum_{(j,i)\in\mathsf{A}} x_{ji}v_i - \sum_{(i,j)\in\mathsf{A}} x_{ij}v_i \\ &= \sum_{i\in\mathsf{N}}\left(\sum_{j\in\mathsf{N} \text{ tel que } (j,i)\in\mathsf{A}} x_{ji} - \sum_{j\in\mathsf{N} \text{ tel que } (i,j)\in\mathsf{A}} x_{ij}\right)v_i\end{aligned}$$

et en identifiant avec $(x,Gv)_{\mathsf{A}} = (G^*x,v)_{\mathsf{N}} = \sum_{i\in\mathsf{N}}(G^*x)_i v_i$, il vient

$$(G^*x)_i = \sum_{j\in\mathsf{N}\ \text{tel que}\ (j,i)\in\mathsf{A}} x_{ji} - \sum_{j\in\mathsf{N}\ \text{tel que}\ (i,j)\in\mathsf{A}} x_{ij} \ .$$

Réponse à la Question 2.2. La question précédente nous a permis d'écrire la loi des nœuds sous la forme $-G^*x = b$. Si $(1_{\mathsf{N}},b)_{\mathsf{N}} = 0$, alors d'après la Question 1.6, J admet un point de minimum v, et d'après la Question 1.5, $-Lv = b$. Comme $-Lv = -G^*DGv$, on a $b = -G^*x$ avec $x = DGv$. Réciproquement, si $b = -G^*x$, $(1_{\mathsf{N}},b)_{\mathsf{N}} = (1_{\mathsf{N}},-G^*x)_{\mathsf{N}} = (-G1_{\mathsf{N}},x)_{\mathsf{A}} = (0,x)_{\mathsf{A}} = 0$, d'après la Question 1.1.

Réponse à la Question 2.3. L'ensemble $\mathcal{K}_b$ est fermé (car défini par une équation affine), et non-vide. La fonction $\mathcal{Q}_q(x)$ est continue, elle est définie sur un espace de dimension finie, et elle vérifie $\lim_{\|x\|\to\infty} \mathcal{Q}_q(x) = +\infty$. On a donc existence d'un point de minimum.

Réponse à la Question 2.4. Comme $\mathcal{K}_b$ est convexe, afin de montrer que le minimum de $\mathcal{Q}_q$ sur $\mathcal{K}_b$ est unique, il suffit de voir que $\mathcal{Q}_q$ est strictement convexe. Comme $\mathcal{Q}_q(x) = \sum_{(i,j)\in\mathsf{A}} R_{ij} f(x_{ij})$, où f est la fonction de $\mathbb{R}\to\mathbb{R}$, telle que $f(t) = |t|^q/q$, et les R_{ij} sont strictements positifs, il suffit de montrer que f est strictement convexe, car l'inégalité de stricte convexité pour $\mathcal{Q}_q$ s'obtient aussitôt à partir de l'inégalité de stricte convexité de f. Notons que f est de classe C^1, avec

$$f'(t) = t|t|^{q-2} \ ,$$

et donc

$$s < t \implies f'(s) < f'(t) \ , \tag{B.2}$$

ce qui entraîne la stricte convexité de f par un résultat classique. (Rappelons que ce résultat s'obtient en prenant $a < c < b$, avec $c = \theta a + (1-\theta)b$, et en appliquant la formule des accroissements finis à f, entre a et c, d'une part, et entre c et b, d'autre part; on obtient ainsi $f(c) - f(a) = f'(\xi)(c-a)$, $f(b) - f(c) = f'(\eta)(b-c)$, avec $b > \eta > c > \xi > a$, et $f'(\eta) > f'(\xi)$, d'où l'on tire l'inégalité de stricte convexité $f(c) < \theta f(a) + (1-\theta) f(b)$.)

Réponse à la Question 2.5. Lorsque $q = 1$, introduisons pour tout $(i,j)\in\mathsf{A}$ une variable auxiliaire $t_{ij}\in\mathbb{R}$, et remarquons que

$$|x_{ij}| \le t_{ij} \iff (x_{ij} \le t_{ij}) \text{ et } (-x_{ij} \le t_{ij}) \ .$$

Introduisons donc la forme linéaire en le couple $(x,t)\in\mathbb{R}^{\mathsf{A}}\times\mathbb{R}^{\mathsf{A}}$:

$$\phi(t,x) = \sum_{(i,j)\in\mathsf{A}} R_{ij} t_{ij} \ ,$$

($\phi(t,x)$ ne dépend pas de x), ainsi que les contraintes affines

$$-G^*x = b, \text{ et } x_{ij} \le t_{ij},\ -x_{ij} \le t_{ij} \ ,\forall (i,j)\in\mathsf{A} \ . \tag{B.3}$$

En tenant compte de la positivité des R_{ij}, on a que le problème

$$\min_{x\in\mathcal{K}_b} \mathcal{Q}_q(x)$$

est équivalent au problème

$$\min_{\substack{(x,t)\,\in\,\mathbb{R}^{\mathsf{A}}\times\mathbb{R}^{\mathsf{A}} \\ \text{satisfaisant (B.3)}}} \phi(t,x)$$

qui est bien un programme linéaire.

Réponse à la Question 2.6. On a observé dans la réponse à la Question 2.4 que $\mathcal{Q}_q$ est convexe (et même strictement convexe), lorsque $q > 1$, et $\mathcal{Q}_q$ est clairement convexe lorsque $q = 1$. Le problème consistant à minimiser la fonction convexe $\mathcal{Q}_q$ sous la contrainte affine $b + G^*x = 0$, relève donc du théorème de Kuhn et Tucker.

Pour appliquer l'énoncé du polycopié, qui fait intervenir seulement des contraintes inégalités, réécrivons, conformément à la Remarque 10.3.5 du polycopié, le problème sous la forme

$$\min_{-G^* x \leq b,\ G^* x \leq -b} \mathcal{Q}_q(x) \quad . \tag{B.4}$$

Les contraintes étant affines, les contraintes sont qualifiées dès que l'ensemble des solutions admissibles est non-vide (appliquer la Remarque 10.2.22 du polycopié). C'est ici le cas car l'on a supposé que $(1_{\mathsf{N}},b)_{\mathsf{N}} = 0$, c'est-à-dire, $\mathcal{K}_b \neq \emptyset$, d'après la Question 2.2.

Formons le Lagrangien $\mathcal{L}(x;v^+,v^-) = \mathcal{Q}_q(x) + (v^+,(-G^*x - b))_{\mathsf{N}} + (v^-,G^*x + b)_{\mathsf{N}}$, avec $x \in \mathbb{R}^{\mathsf{A}}$ et $v^+,v^- \in \mathbb{R}_+^{\mathsf{N}}$. D'après le théorème de Kuhn et Tucker, on a que $x \in \mathcal{K}_b$ est solution de (B.4) si, et seulement si, il existe un couple de vecteurs de multiplicateurs de Lagrange $(\bar{v}^+,\bar{v}^-) \in \mathbb{R}_+^{\mathsf{N}}$, tel que $(\bar{x};\bar{v}^+,\bar{v}^-)$ soit point-selle de $\mathcal{L}$ sur $\mathbb{R}^{\mathsf{A}} \times (\mathbb{R}_+^{\mathsf{N}})^2$:

$$\mathcal{L}(\bar{x};v^+,v^-) \leq \mathcal{L}(\bar{x};\bar{v}^+,\bar{v}^-) \leq \mathcal{L}(x;\bar{v}^+,\bar{v}^-) \ , \quad \forall x \in \mathbb{R}^{\mathsf{A}}, \forall (v^+,v^-) \in (\mathbb{R}_+^{\mathsf{N}})^2 \ .$$

En posant $\mathcal{L}(x,v) = \mathcal{Q}_q(x) + (v,G^*x + b)_{\mathsf{N}}$, $v = v^- - v^+$, $\bar{v} = \bar{v}^- - \bar{v}^+$, on a donc que $\bar{x} \in \mathbb{R}^{\mathsf{A}}$ est minimum de $\mathcal{Q}_q$ sur $\mathcal{K}_b$ si et seulement si il existe $\bar{v} \in \mathbb{R}^{\mathsf{N}}$ tel que

$$\mathcal{L}(v,\bar{x}) \leq \mathcal{L}(\bar{v},\bar{x}) \leq \mathcal{L}(\bar{v},x) \ , \quad \forall x \in \mathbb{R}^{\mathsf{A}}, \forall v \in \mathbb{R}^{\mathsf{N}} \ .$$

Réponse à la Question 2.7. Les résultats du cours sur le point selle et le Théorème de Kuhn et Tucker montrent que

$$\min_{x \in \mathcal{K}_b} \mathcal{Q}_q(x) = \mathcal{Q}_q(\bar{x}) = \mathcal{L}(\bar{x},\bar{v}) = \min_{x \in \mathbb{R}^{\mathsf{A}}} \sup_{v \in \mathbb{R}^{\mathsf{N}}} \mathcal{L}(x,v) = \max_{v \in \mathbb{R}^{\mathsf{N}}} \inf_{x \in \mathbb{R}^{\mathsf{A}}} \mathcal{L}(x,v) \ . \tag{B.5}$$

Explicitons maintenant

$$\begin{aligned}
\inf_{x \in \mathbb{R}^{\mathsf{A}}} \mathcal{L}(x,v) &= \inf_{x \in \mathbb{R}^{\mathsf{A}}} \mathcal{Q}_q(x) + (v,G^*x)_{\mathsf{N}} + (v,b)_{\mathsf{N}} = \inf_{x \in \mathbb{R}^{\mathsf{A}}} \mathcal{Q}_q(x) + (Gv,x)_{\mathsf{A}} + (v,b)_{\mathsf{N}} \\
&= (v,b)_{\mathsf{N}} + \inf_{x \in \mathbb{R}^{\mathsf{A}}} \sum_{(i,j) \in \mathsf{A}} \left(\frac{R_{ij}}{q} |x_{ij}|^q + x_{ij}(v_j - v_i) \right) \qquad \text{(B.6)} \\
&= (v,b)_{\mathsf{N}} + \sum_{(i,j) \in \mathsf{A}} \inf_{x_{ij} \in \mathbb{R}} \left(\frac{R_{ij}}{q} |x_{ij}|^q + x_{ij}(v_j - v_i) \right) \ , \qquad \text{(B.7)}
\end{aligned}$$

car le problème (B.6) se décompose en sous-problèmes d'optimisation indépendants, en chaque variable x_{ij}. Considérons le sous-problème:

$$\inf_{x_{ij} \in \mathbb{R}} \left(\frac{R_{ij}}{q} |x_{ij}|^q + x_{ij}(v_j - v_i) \right)$$

En dérivant par rapport à x_{ij}, on trouve que les x_{ij} optimaux vérifient la loi d'Ohm non-linéaire

$$R_{ij} x_{ij} |x_{ij}|^{q-2} = v_i - v_j \tag{B.8}$$

(on retrouve la loi d'Ohm pour $q = 2$). Donc

$$\inf_{x_{ij} \in \mathbb{R}} \left(\frac{R_{ij}}{q} |x_{ij}|^q + x_{ij}(v_j - v_i) \right) = (\frac{1}{q} - 1) x_{ij}(v_i - v_j) = -\frac{R_{ij}}{p} \left| \frac{v_i - v_j}{R_{ij}} \right|^p \ , \tag{B.9}$$

et en reportant (B.9) dans (B.7), on a

$$\inf_{x \in \mathbb{R}^{\mathsf{A}}} \mathcal{L}(x,v) = (v,b)_{\mathsf{N}} - \mathcal{P}_p(v) \ ,$$

ce qui avec (B.5), donne

$$\min_{x \in \mathcal{K}_b} \mathcal{Q}_q(x) = \max_{v \in \mathbb{R}^{\mathsf{N}}} ((b,v)_{\mathsf{N}} - \mathcal{P}_p(v)) \ .$$

C'est là une version non-linéaire de la dualité entre courants et tensions: il est équivalent de minimiser la puissance $\mathcal{Q}_q(x)$ associée au courant x, ou la puissance $\mathcal{P}_p(v) - (b,v)_{\mathsf{N}}$ associée au potentiel v. Remarquons enfin (ce dernier point n'était pas demandé) que le potentiel optimal vérifie une équation non-linéaire analogue à $-Lv = b$. Cette équation s'obtient en dérivant $\mathcal{P}_p(v) - (b,v)_{\mathsf{N}}$:

$$\sum_{j \in \mathsf{N} \text{ tel que } (i,j) \in \mathsf{A}} \frac{v_i - v_j}{R_{ij}} \left| \frac{v_i - v_j}{R_{ij}} \right|^{p-2} = b_i \ , \forall i \in \mathsf{N} \ . \tag{B.10}$$

3. Analogue continu

Réponse à la Question 3.1. On peut écrire:

$$ph\mathcal{P}_p(v^h) = \sum_{\substack{0 \le i_1 \le k-1 \\ 0 \le i_2 \le k}} h^2 \left| \frac{v^h_{i_1+1,i_2} - v^h_{i_1 i_2}}{h} \right|^p + \sum_{\substack{0 \le i_1 \le k \\ 0 \le i_2 \le k-1}} h^2 \left| \frac{v^h_{i_1,i_2+1} - v^h_{i_1 i_2}}{h} \right|^p \ ,$$

et en utilisant la formule des accroissements finis,

$$ph\mathcal{P}_p(v^h) = \sum_{\substack{0 \le i_1 \le k-1 \\ 0 \le i_2 \le k}} h^2 \left| \frac{\partial v}{\partial x_1}(\xi_{i_1 i_2}) \right|^p + \sum_{\substack{0 \le i_1 \le k \\ 0 \le i_2 \le k-1}} h^2 \left| \frac{\partial v}{\partial x_2}(\eta_{i_1 i_2}) \right|^p \ ,$$

où $\xi_{i_1 i_2}$ et $\eta_{i_1 i_2}$ sont des points des segments ouverts joignant $(i_1 h, i_2 h)$ à $((i_1+1)h, i_2 h)$, et $(i_1 h, i_2 h)$ à $(i_1 h, (i_2+1)h))$, respectivement. Montrons que

$$\lim_{h \to 0} ph\mathcal{P}(v^h) = \|\nabla v\|^p_{L^p} \tag{B.11}$$

à l'aide du théorème de convergence dominée. Notons que les carrés

$$K_{i_1 i_2} = [i_1 h, (i_1+1)h[\times[i_2 h, (i_2+1)h[, \text{ avec } 0 \le i_1 \le k-1, 0 \le i_2 \le k-1 \ ,$$

forment une partition de $[0,1[\times[0,1[$, ce qui nous permet de définir la fonction $f^h : [0,1[\times[0,1[$, constante par morceaux, valant identiquement $|\frac{\partial v}{\partial x_1}(\xi_{i_1 i_2})|^p$ sur le carré $K_{i_1 i_2}$. On a

$$\sum_{\substack{0 \le i_1 \le k-1 \\ 0 \le i_2 \le k}} h^2 \left| \frac{\partial v}{\partial x_1}(\xi_{i_1 i_2}) \right|^p = \int_{[0,1]^2} f^h(x_1,x_2) dx_1 dx_2 \ . \tag{B.12}$$

Vérifions que

$$\lim_{h \to 0} \int_{[0,1[^2} f^h(x_1,x_2) dx_1 dx_2 = \int_{[0,1]^2} \left| \frac{\partial v}{\partial x_1} \right|^p dx_1 dx_2 \ . \tag{B.13}$$

La fonction $\frac{\partial v}{\partial x_1}$, qui est continue, est majorée par une constante sur $[0,1]^2$, donc $f^h = |f^h|$ est majorée par une fonction intégrable (constante) indépendamment de h. D'autre part, pour tout $(x_1,x_2) \in [0,1[\times[0,1[$, il existe un unique carré $K_{i_1 i_2}$ tel que $(x_1,x_2) \in K_{i_1 i_2}$, et comme le carré $K_{i_1 i_2}$ est de coté h, la distance entre (x_1,x_2) et $\xi_{i_1 i_2}$ tend vers 0 quand $h \to 0$. Comme $\frac{\partial v}{\partial x_1}$ est continue, et $f^h(x_1,x_2) = |\frac{\partial v}{\partial x_1}(\xi_{i_1 i_2})|^p$, cela entraîne que f^h converge simplement vers $\frac{\partial v}{\partial x_1}$ sur $[0,1[^2$, d'où l'on déduit (B.13) par application du théorème de convergence dominée. En combinant (B.13) avec (B.12), et en utilisant les mêmes arguments pour la dérivée en x_2, on obtient (B.11).

Réponse à la Question 3.2. Soit $J(u) = \|\nabla u\|^p_{L^p}$, et soit u un point de minimum de J dans l'ensemble admissible $\{u \in C^1([0,1]^2) \mid u = g \text{ sur } \Gamma\}$. Prenons une fonction quelconque $v \in C^1([0,1]^2)$, telle que $v = 0$ sur Γ. Comme $u + tv = g$ sur Γ, $u + tv$ est une solution admissible pour tout t, donc $J(u + tv) \ge J(u)$. En dérivant sous le signe somme, ce qui est justifié car u et v sont de classe C^1, et $s \mapsto |s|^p$ est C^1 pour $p > 1$, on obtient

$$\frac{\partial J(u+tv)}{\partial t}|_{t=0} = \int p \frac{\partial u}{\partial x_1} \left| \frac{\partial u}{\partial x_1} \right|^{p-2} \frac{\partial v}{\partial x_1} dx_1 dx_2 + \int p \frac{\partial u}{\partial x_2} \left| \frac{\partial u}{\partial x_2} \right|^{p-2} \frac{\partial v}{\partial x_2} dx_1 dx_2 = 0 \ .$$

Comme ceci est vrai pour toute fonction C^1 à support compact, on conclut que u est solution faible du problème aux limites

$$A(u) = 0, \ u = g \text{ sur } \Gamma \ , \tag{B.14}$$

avec

$$A(u) = \frac{\partial}{\partial x_1}\left(\frac{\partial u}{\partial x_1}\left|\frac{\partial u}{\partial x_1}\right|^{p-2}\right) + \frac{\partial}{\partial x_2}\left(\frac{\partial u}{\partial x_2}\left|\frac{\partial u}{\partial x_2}\right|^{p-2}\right) \ .$$

On retrouve évidemment le Laplacien pour $p = 2$. L'équation (B.14) est une version continue de (B.10).