

Connexions entre Concepts

Levitt — Algèbre Linéaire Avancée & Numérique — L3 Paris-Saclay

Analyse des exercices et des connexions conceptuelles profondes

1. Architecture globale du cours

Le cours repose sur **trois piliers** et leurs interactions:

1. **Spectre** (valeurs propres, valeurs singulières): caractère qualitatif de A^n , e^{tA} , $f(A)$; min-max; HDP.
2. **Stabilité/quantitatif** (normes, conditionnement, SVD): robustesse des calculs, perturbations, convergence.
3. **Itération** (Richardson, Krylov, puissance): algorithmes efficaces exploitant la structure sparse.

La **connexion centrale**: \leq conditionnement $\kappa(A) = \frac{s_1}{s_r}$ contrôle à la fois la stabilité numérique (section 4) et la vitesse de convergence des méthodes itératives (sections 8-10). C'est \leq fil rouge \geq du cours.

2. Connexions entre grands concepts

2.1. Produit scalaire et algèbre linéaire

Le produit scalaire $\langle x, y \rangle = x^*y$ est omniprésent et structure toute la théorie:

- **Adjoint**: $\langle x, Ay \rangle = \langle A^*x, y \rangle$ — définit A^* de façon intrinsèque.
- **Hermitien**: $A = A^* \Leftrightarrow \langle x, Ay \rangle = \langle Ax, y \rangle \Rightarrow \lambda \in \mathbb{R}$ et vecteurs propres \perp .
- **Orthogonalité**: Gram-Schmidt \rightarrow QR; complément orthogonal \rightarrow Fredholm; bases orthonormées \rightarrow théorème spectral.
- **Quotient de Rayleigh**: $R(x) = x^*A\frac{x}{x^*x}$ — relie valeurs propres (Courant-Fischer), preuve du th. spectral (min sur sphère compacte), et convergence du gradient conjugué (A-norme).

Fil rouge: \leq produit scalaire induit la structure spectrale.

Toute matrice hermitienne agit “diagonalement” dans une base orthonormée adaptée. Les matrices normales sont précisément celles qui partagent cette propriété dans \mathbb{C} . Les matrices non-normales peuvent avoir un spectre trompeur (section 4: $\|A\| \gg \rho(A)$).

2.2. Factorisations et leurs usages

Les factorisations ne sont pas des fins en soi mais des outils pour différents problèmes:

Factorisation	Résout	Hypothèse clé
LU pivoté	$Ax = b$ direct	Carré inversible
Cholesky	$Ax = b$, test HDP	HDP
QR	Moindres carrés	Rang plein
SVD	Moindres carrés, rang, κ	Quelconque
Schur	Valeurs propres	Quelconque

Connexion Cholesky-Gram-Schmidt (exercice 2.4): si $A^*A = LL^*$ (Cholesky de la matrice de Gram), alors $A(L^*)^{-1}$ est unitaire — c'est exactement la factorisation QR vue différemment.

Connexion SVD-LU-QR: la SVD généralise toutes les autres (U, V orthogonales, S diagonale). Elle révèle les 4 sous-espaces fondamentaux de Strang.

2.3. Schur, Jordan, et robustesse

Le cours évite **délibérément** Jordan et \leq polynôme caractéristique au profit de Schur et SVD.

Pourquoi? Jordan est instable: une petite perturbation peut fusionner ou séparer des blocs. En revanche, Schur et SVD sont **stables** (valeurs singulières sont Lipschitz-continues en A , exercice 4.9).

Connexion Schur-calcul fonctionnel: la décomposition de Schur permet de définir $f(A)$ par $f(T)$ (triangulaire sup.) sans passer par les vecteurs propres, d'où la stabilité du calcul fonctionnel holomorphe.

2.4. SVD et conditionnement: \leq coeur quantitatif

La SVD révèle la **géométrie** de A : l'ima \geq de la boule unité est une ellipsoïde de demi-axes s_k .

- **Norme:** $\|A\| = s_1$ (grand axe).
- **Conditionnement:** $\kappa(A) = \frac{s_1}{s_r}$ (excentricité).
- **Rang:** r = nombre de valeurs singulières non nulles.
- **Meilleure approximation:** troncation SVD à rang k (Eckart-Young).
- **Moindres carrés:** $x = VS^{-1}U^*b$ (stable même pour rang déficient via pseudo-inverse).

Connexion fondamentale: conditionnement \leftrightarrow vitesse de convergence.

Richardson: $n_{\text{iter}} \approx \kappa$ iters. Gradient conjugué: $n_{\text{iter}} \approx \sqrt{\kappa}$ iters. Le conditionnement $\kappa = \frac{s_1}{s_r}$ est la quantité qui gouverne la difficulté numérique d'un problème.

2.5. Analyse complexe et calcul fonctionnel

La formule de Cauchy $f(\lambda) = \frac{1}{2\pi i} \oint_C \frac{f(z)}{z-\lambda} dz$ est étendue au cas matriciel pour définir $f(A)$.

Pourquoi ce détour complexe?

- Définir $f(A)$ pour f quelconque holomorphe (pas seulement séries entières convergentes).
- Preuve de Gelfand: $A^n = \frac{1}{2\pi i} \oint_C z^n (z-A)^{-1} dz \rightarrow 0$ car $z^n \rightarrow 0$ uniformément sur \leq contour.
- Preuve qu'une matrice a toujours une valeur propre (exercice 6.6 + Liouville).
- Cayley-Hamilton via la formule de Cramer.

Connexion: résolvante et spectre.

$(z-A)^{-1}$ est holomorphe **exactement** hors de $\sigma(A)$. Les singularités de la résolvante **sont** \leq spectre. Cela unifie: valeurs propres (pôles de la résolvante), calcul fonctionnel (intégrale de contour évitant les pôles), et stabilité spectrale (les valeurs propres se déplacent peu si A est perturbée hermitienne — exercice 4.9).

2.6. Méthodes itératives: Richardson \rightarrow Krylov \rightarrow gradient conjugué

Hiérarchie des méthodes, du pire au meilleur:

1. **Richardson stationnaire** (α fixe): $n_{\text{iter}} = O(\kappa)$. Simple, souvent lent.
2. **Richardson adaptatif** (α_n variable): espace de Krylov K_n engendré. Meilleur, mais choix de α difficile.
3. **Gradient conjugué:** minimise $\|e\|_A$ dans $K_n \rightarrow$ optimal dans cet espace. $n_{\text{iter}} = O(\sqrt{\kappa})$.

La transition 2 \rightarrow 3 est la clé: on remplace la question "quel α ?" par "quel élément de K_n minimise l'erreur?". C'est la méthode de Galerkin avec $X = Y = K_n$.

3. Analyse des exercices par chapitre

3.1. Chapitre 1: Algèbre élémentaire

3.1.1. Ex. 1 (Formes quadratiques)

Concepts utilisés: forme quadratique $q_{A(x)} = x^T A x$; définition positive; méthode du complément du carré; diagonalisation.

Connexions: trois méthodes pour HDP — (a) majorer $|xy|$ par AM-GM (élémentaire), (b) compléter \leq carré (complément de Schur!), (c) diagonaliser et regarder les VP. La méthode (b) anticipe la section 1.2.3: compléter \leq carré = éliminer par complément de Schur.

Pour $q = x^2 + 2xy + y^2 = (x+y)^2$: SDP mais pas HDP (noyau = $\text{Span}((1, -1)^T)$). La matrice $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ a VP 0 et 2.

3.1.2. Ex. 2 (Calcul différentiel)

Concepts: gradient; développement de $\|x+h\|$.

$\|x + h\|^2 = \|x\|^2 + 2 \operatorname{Re}(x^*h) + \|h\|^2$, donc $\nabla\|x\| = \frac{x}{\|x\|}$ pour $x \neq 0$.

Connexion: calcul du gradient de formes quadratiques (section 1.1.4). Ce calcul revient partout: gradient de $\|Ax - b\|^2$ pour les moindres carrés, gradient du quotient de Rayleigh pour \leq th. spectral.

3.1.3. Ex. 5 (Valeurs propres du Laplacien discret)

Concepts: Laplacien discret L (différences finies); valeurs propres explicites $\lambda_k = 2N^2(1 - \cos(\frac{\pi k}{N}))$.

Connexion DFT: les vecteurs propres $\sin(\pi k \frac{n}{N})$ sont la partie imaginaire de $(v_k)_j = e^{i\pi k \frac{j}{N}}$ — relié aux matrices circulantes et à la DFT (section 3.6). Les VP de L sont les valeurs de la transformée de Fourier du filtre $(-1, 2, -1)$ évaluées aux fréquences $\frac{k}{N}$.

Connexion conditionnement: $\lambda_1 \approx (\frac{\pi}{N})^2$, $\lambda_N \approx 4N^2$ donc $\kappa(L) \approx N^4$. C'est l'origine du mauvais conditionnement des systèmes de différences finies (exercice 4.14).

Connexion méthodes itératives: $n_{\text{iter}} \approx \kappa^{\{\frac{1}{2}\}} = N^2$ pour \leq gradient conjugué, soit $O(N^3)$ total pour résoudre $Lu = f$ — identique au pivot de Gauss! Le préconditionnement est donc essentiel.

3.1.4. Ex. 10 (Estimation optimale)

Concepts: moindres carrés pondérés; Lagrange; complément de Schur pour l'inversibilité.

Minimiser $\operatorname{Var}(\sum x_i X_i) = x^T C x$ sous $\sum x_i = 1$, i.e. $\mathbf{1}^T x = 1$.

Système KKT: $\begin{pmatrix} 2C & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

Connexion Schur: la matrice 2×2 par blocs est inversible car son complément de Schur $-\mathbf{1}^T(2C)^{-1}\mathbf{1} = -\mathbf{1}^T C^{-1} \frac{1}{2} < 0$ est non nul (C HDP donc C^{-1} HDP donc $\mathbf{1}^T C^{-1} \mathbf{1} > 0$).

Connexion HDP: C HDP est utilisé deux fois — pour l'existence du minimum (Hessienne $2C > 0$) et pour l'inversibilité du système.

3.2. Chapitre 2: Systèmes linéaires

3.2.1. Ex. 2.4 (Cholesky pour l'orthogonalisation)

Connexion fondamentale: Gram-Schmidt \equiv Cholesky.

Si $A^*A = LL^*$, alors $(A(L^*)^{-1})^*(A(L^*)^{-1}) = L^{\{-*\}}A^*AL^{-1} = L^{\{-*\}}LL^*L^{-1} = I$: c'est la factorisation QR. Les deux algorithmes produisent la même chose par deux chemins différents.

Stabilité: Gram-Schmidt est numériquement instable (perte d'orthogonalité); Gram-Schmidt modifié et Givens sont préférés.

3.2.2. Ex. 2.5 (Pivot par blocs)

Concepts: pivot de Gauss par blocs; complément de Schur.

Appliquer l'élimination à $\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$: éliminer A_{21} en soustrayant $A_{21}A_{11}^{-1}$ fois la première ligne. Le résidu $A_{22} - A_{21}A_{11}^{-1}A_{12}$ est **exactement** \leq complément de Schur.

Connexion: la preuve de la proposition 1.1 (complément de Schur) et la preuve du pivot de Gauss (section 2.2) sont la même chose vue différemment.

3.3. Chapitre 3: Théorème spectral

3.3.1. Ex. 3.4 (AB et BA ont même spectre non nul)

Connexion résolvante: via séries entières, $(z - \varepsilon AB)^{-1} = (\frac{1}{z}) \sum (\varepsilon A \frac{B}{z})^n$. Puis formule $(z - \varepsilon AB)^{-1} = z^{-1}(I + A(z - \varepsilon BA)^{-1}\varepsilon B)$ (exercice 6.8). Valeur en $\varepsilon = 1$: $\sigma(AB) \setminus \{0\} = \sigma(BA) \setminus \{0\}$.

Application: $\sigma(AB) = \sigma(B^{\frac{1}{2}}AB^{\frac{1}{2}})$ si B HDP, ce qui montre que AB est diagonalisable à valeurs propres réelles (exercice 3.4).

3.3.2. Ex. 3.7 (Interlacement des valeurs propres)

Concepts: formules min-max; sous-matrices.

Si B est une restriction de A à un sous-espace de dimension $k < N$: par Courant-Fischer, $\lambda_{i(A)} \leq \lambda_{i(B)} \leq \lambda_{\{i+1\}}(A)$ (interlacement).

Connexion: même argument que la preuve du min-max. Utilisé en pratique pour borner les valeurs propres d'une matrice modifiée.

3.4. Chapitre 4: Algèbre quantitative

3.4.1. Ex. 4.3 (Conditionnement, déterminant)

Concepts: $\kappa(A) = \frac{s_1}{s_n}$; $\det(A) = \prod_i \lambda_i$.

Matrice $\text{diag}(\frac{1}{100}, 100)$: $\kappa = 10^4$, $\det = 1$. Matrice $\text{diag}(100, 100)$: $\kappa = 1$, $\det = 10^4$.

Morale: \leq déterminant est un mauvais indicateur du conditionnement. Une matrice peut avoir $\det = 1$ et être mal conditionnée (ou \det énorme et être bien conditionnée).

3.4.2. Ex. 4.9 (Stabilité des valeurs propres)

Connexion min-max et perturbations: pour A hermitienne, $\lambda_{i(A+B)} \leq \lambda_{i(A)} + \|B\|$ par min-max (en prenant \leq même sous-espace S pour $A+B$ que pour A , et en utilisant $\max_S \frac{x^*(A+B)x}{x^*x} \leq \max_S x^*Ax + \|B\|$).

Même argument pour les valeurs singulières: $|s_{i(A+B)} - s_{i(A)}| \leq \|B\|$. C'est la robustesse fondamentale de la SVD.

Contraste: les valeurs propres d'une matrice non-hermitienne peuvent être instables (exercice 4.6: matrice $\begin{pmatrix} 0 & 1 \\ \varepsilon & \varepsilon \end{pmatrix}$ avec $\kappa(P) \approx \frac{1}{\varepsilon}$).

3.4.3. Ex. 4.13 (Sherman-Morrison)

Concepts: série de Neumann; perturbation de rang 1.

$$(A + \varepsilon uu^*)^{-1} = A^{-1} - \varepsilon A^{-1} u u^* \frac{A^{-1}}{1 + \varepsilon u^* A^{-1} u}.$$

Connexion Schur: peut se déduire du complément de Schur appliqué à la matrice augmentée $\begin{pmatrix} A & u \\ u^* & -\frac{1}{\varepsilon} \end{pmatrix}$.

Application: mise à jour efficace d'un inverse quand on ajoute une observation (statistiques, optimisation).

3.5. Chapitre 6: Calcul fonctionnel

3.5.1. Ex. 6.1 (Blocs de Jordan et calcul fonctionnel)

Connexion: pour $A = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} = \lambda I + N$ (N nilpotent), $N^2 = 0$ donc $f(A) = f(\lambda)I + f'(\lambda)N$. D'où: $e^{tA} = e^{t\lambda}(I + tN) = \begin{pmatrix} e^{t\lambda} & te^{t\lambda} \\ 0 & e^{t\lambda} \end{pmatrix}$.

Conséquence: solution de $x'' = \alpha x' + \beta x$ (racine double λ): $x(t) = (c_1 + c_2 t)e^{\lambda t}$. Justification rigoureuse de la méthode des tentatives.

Connexion Gelfand: $\|A^n\| \approx n \cdot |\lambda|^{n-1}$ — \leq facteur n vient du bloc Jordan et est capturé dans la borne $\|A^n\| \leq C(\rho(A) + \varepsilon)^n$ mais pas $C\rho(A)^n$.

3.5.2. Ex. 6.6 (Résolvante et valeurs propres)

Concepts: $(z - A)^{-1}$ bornée pour $|z| \gg 1$; Liouville.

$$\|(z - A)^{-1}\| \leq \frac{C}{|z|} \text{ pour } |z| \text{ grand (série de Neumann: } (z - A)^{-1} = \left(\frac{1}{z}\right)\left(I - \frac{A}{z}\right)^{-1}, \left\|\frac{A}{z}\right\| < 1 \text{ pour } |z| > \|A\|).$$

Si A n'avait pas de valeur propre, $(z - A)^{-1}$ serait entière et bornée donc constante (Liouville) — contradiction. Preuve analytique de l'existence des VP sans polynôme caractéristique.

3.6. Chapitre 7: Asymptotique

3.6.1. Ex. 7.3 (Pré-asymptotique)

Concepts: Gelfand; comportement transitoire non-normal.

Pour $A = \begin{pmatrix} 0.9 & 1000 \\ 0 & 0.8 \end{pmatrix}$: $\rho(A) = 0.9$, $A^n \rightarrow 0$. Mais \leq vecteur $(1, 0)^T$ donne $\|A^n e_1\| \approx 1000 \cdot \frac{0.9^n}{0.1}$ — croissance initiale massive avant décroissance.

Morale: l'asymptotique $\rho(A)^n$ ne dit rien sur \leq comportement transitoire. La non-normalité peut causer une amplification transiente énorme. C'est crucial en stabilité de fluides (transition à la turbulence).

Connexion conditionnement: $\|A^n\| \leq \|P\| \|P^{-1}\| \rho(A)^n = \kappa(P) \rho(A)^n$. Si A est mal conditionnée comme matrice de passage, \leq pré-facteur $\kappa(P)$ peut être gigantesque.

3.7. Chapitre 8: Richardson

3.7.1. Ex. 8.1 (Richardson = descente de gradient)

Connexion: $\min_x \frac{1}{2} x^T A x - b^T x$ (convexe si A HDP). Gradient: $Ax - b = -(b - Ax)$. Mise à jour gradient: $x_{n+1} = x_n - \alpha \nabla f(x_n) = x_n - \alpha(Ax_n - b)$ = Richardson.

Géométrie: les courbes de niveau sont des ellipses (axes = vecteurs propres, longueurs = $\frac{1}{\sqrt{\lambda_i}}$). Le gradient est \perp aux courbes de niveau. La convergence est lente si les ellipses sont très allongées (κ grand) car \leq gradient pointe "de travers".

Connexion conditionnement: \leq ratio $\frac{\lambda_N}{\lambda_1} = \kappa$ détermine l'élongation des ellipses et \leq taux de convergence.

3.7.2. Ex. 8.3 (Richardson pour l'équation de Poisson)

Connexion chapitres 1 et 8: l'équation $-u'' = f$ discrétisée donne la matrice du Laplacien (exercice 1.5) avec $\kappa = O(N^2)$ (exercice 4.14). Richardson converge donc en $O(N^2)$ itérations, chaque itération coûte $O(N)$ (matrice creuse tridiagonale), total $O(N^3)$ — identique au pivot de Gauss. Le gradient conjugué donne $O(N)$ itérations $\times O(N)$ coût = $O(N^2)$: gain d'un facteur N .

3.8. Chapitre 9: Krylov

3.8.1. Ex. 9.3 (Valeurs propres distinctes et terminaison)

Concepts: approximation polynomiale; spectre discret.

Si A a k valeurs propres distinctes, \leq polynôme minimal a degré k . Dans K_k , \leq gradient conjugué trouve la solution exacte car il peut construire \leq polynôme annulateur de A .

Connexion DFT: si A est une matrice circulante (N valeurs propres $\hat{a}_0, \dots, \hat{a}_{N-1}$, potentiellement avec répétitions), \leq gradient conjugué converge en au plus $k = |\{1\}|$ itérations.

3.8.2. Ex. 9.5 (Approximation polynomiale)

Connexion profonde: \leq gradient conjugué cherche $p_n \in \mathbb{C}_{n[X]}$ minimisant $\int \lambda (p_n(\lambda) - \frac{1}{\lambda})^2 d\mu(\lambda)$ où $\mu = \sum_i |\langle v_i, b \rangle|^2 \delta_{\{\lambda_i\}}$ (mesure spectrale de b).

C'est une **projection orthogonale** de $\frac{1}{\lambda}$ sur les polynômes de degré n dans $L^2(\mu)$. La convergence est rapide si les VP sont groupées (peu de "support" de μ), même si κ est grand — c'est pourquoi \leq préconditionnement fonctionne.

3.9. Chapitre 10: Valeurs propres numériques

3.9.1. Ex. 10.3 (Puissance shiftée)

Connexion Richardson: choisir $\alpha = \frac{1}{\lambda}$ (shift) pour que Richardson converge vers le vecteur propre de λ_N . L'analogie du conditionnement pour les VP est $\frac{\lambda_N - \lambda^*}{\lambda_{N-1} - \lambda^*}$ où λ^* est la VP cible.

Connexion calcul fonctionnel: la puissance inverse avec shift $(A - \mu I)^{-1}$ converge vers le vecteur propre de la VP la plus proche de μ . La fonction $f(\lambda) = \frac{1}{\lambda - \mu}$ est amplifiée près de μ .

3.9.2. Ex. 10.6 (Puissance inverse + gradient conjugué)

Avantage: pas besoin de factoriser A complètement; $A^{-1}x_n$ est approché par gradient conjugué.

Connexion Krylov: si on résout $(A - \mu I)y = x_n$ par gradient conjugué jusqu'à précision ε , et qu'on utilise y comme nouveau x_{n+1} , on obtient une méthode combinée "shift-invert Lanczos" très efficace pour les grands problèmes.

Subtilité: la convergence du gradient conjugué pour $(A - \mu I)$ est rapide si μ est proche d'une VP (conditionnement local). Ironie: la puissance inverse conver<=> d'autant mieux que μ est proche d'une VP, et <=> gradient conjugué conver<=> d'autant plus mal. Les deux effets se compensent partiellement.

4. Connexions transversales majeures

4.1. Le triangle HDP — min-max — Gradient conjugué

HDP est la condition clé pour les méthodes itératives (Richardson, gradient conjugué convergent). Elle garantit:

1. Quotient de Rayleigh borné \Rightarrow VP réelles positives (th. spectral).
2. A^*A HDP si A rang plein \Rightarrow moindres carrés bien posé.
3. Formule min-max applicable \Rightarrow interlacement, stabilité.
4. A_n tridiagonale \Rightarrow Lanczos; A_n inversible \Rightarrow GC bien défini.
5. Minimiser $\frac{1}{2}x^T Ax - b^T x \equiv$ Richardson \equiv GC dans K_n .

4.2. Complément de Schur comme outil universel

Le complément de Schur $\tilde{A} = A_{22} - A_{21}A_{11}^{-1}A_{12}$ apparaît:

- **Preuve du pivot de Gauss** (section 2.2): A inversible $\Leftrightarrow \tilde{A}$ inversible.
- **Moindres carrés avec Lagrange** (ex. 1.10): système KKT inversible via Schur.
- **Estimation optimale** (ex. 1.10): inversibilité du système Lagrangien.
- **Gradient conjugué** (section 9.6): récurrence courte via Schur sur A_n tridiagonale.
- **Sherman-Morrison** (ex. 4.13): $(A + uv^*)^{-1}$ via matrice augmentée + Schur.
- **Théorème de Norton** (ex. 1.7): équivalence de circuits via Schur.

4.3. Séries de Neumann comme perturbation universelle

$(A - B)^{-1} = A^{-1} \sum (BA^{-1})^n$ apparaît dans:

- **Conditionnement** (section 4.5): stabilité de $Ax = b$ sous perturbation.
- **Résolvante** (section 6.3): $(z - A)^{-1}$ holomorphe, développement en z_0 .
- **Richardson** (section 8.3): la série de Neumann conver<=> \Leftrightarrow Richardson converge.
- **Sherman-Morrison** (ex. 4.13): perturbation de rang 1.
- **Gelfand** (section 7): borne sur $\|A^n\|$ via <=> contour d'intégration.